Protein Synthesis and Ribosome Structure

Translating the Genome

Edited by
Knud H. Nierhaus and
Daniel N. Wilson

WILEY-VCH Verlag GmbH & Co. KGaA

Prof. Dr. Knud H. Nierhaus
Dr. Daniel N. Wilson
Max-Planck-Institute for Molecular Genetics
Research Group Ribosomes
Ihnestr. 73
14195 Berlin
Germany
nierhaus@molgen.mpg.de
wilson@molgen.mpg.de

Library of Congress Card No.: applied for.
British Library Cataloguing-in-Publication Data:
A catalogue record for this book is available from the British Library

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de.

Printed in the Federal Republic of Germany.

Printed on acid-free paper

Typesetting: Shiv e-Publishing Technologies Pvt. Ltd. Bangalore, India
Printing: Strauss GmbH, Mörlenbach
Bookbinding: Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN 3-527-30638-2

Cover
The cover shows the small and large subunit of the ribosome surrounded by various accessory factors involved in protein biosynthesis. Cover layout by Grafik-Design Schulz, Fußgönheim, based on an illustration by K. H. Nierhaus and D. N. Wilson.

Contents

Preface

A large part of this book is the offspring of a series of more than 30 lectures covering all the aspects of translation that excite us. The lectures have been offered at two Berlin universities every year for the past decades, and every time the students have enquired where they can read about the topics of lectures, since none of the classical textbooks are treating the subjects in depth. When we received the offer from Wiley-VCH to write a book about the translational machinery, we agreed immediately, as this gave us a chance to provide future students with an appropriate reference. Furthermore, we felt that the timing was perfect because of the recent advances in terms of ribosome structure contributed by the cryo-electron microscopy and crystallography groups, which have sharpened our view of ribosome function in a revolutionary way.

It was immediately clear to us that we could not fulfill this enormous task without the help of specialists in fields where we have limited expertise. This is particularly true for the topics concerning the history of ribosome research, synthetases, mRNA decay, recoding events and protein folding, as well as the highly complex areas of eukaryotic translation, namely, the assembly, initiation and regulation of eukaryotic ribosomes. We are therefore pleased, grateful and honoured that leading scientists in these fields accepted our invitation and provided such wonderful contributions.

We would also like to thank all the members of the Nierhaus group, both past and present, who have participated directly as authors or indirectly with stimulating and enthusiastic literature discussions every Friday afternoon. Last but not least, we appreciate the understanding, leniency and wonderful support of Dr. Frank Weinreich at Wiley-VCH, who has allowed this book to be as colorful as it is.

Berlin, July 2004

Knud H. Nierhaus and
Daniel N. Wilson

Contributors

Jean-Hervé Alix
Institut de Biologie Physico-Chimique
CNRS, UPR 9073
13 rue Pierre et Marie Curie
75005 Paris
France
jean-herve.alix@ibpc.fr

Gregor Blaha
Department of Molecular Biophysics and Biochemistry
HHMI, Yale University
266 Whitney Avenue / Bass 418
New Haven, CT 06536
USA
blaha@mail.csb.yale.edu

Agamemnon J. Carpousis
Laboratoire de Microbiologie et Génétique Moléculaire
CNRS, UMR 5100
Paul Sabietier Université
118 Route de Narbonne
31062 Toulouse
France
carpousi@ibcg.biotoul.fr

Sean R. Connell
Institut für medizinische Physik und Biophysik
Universitätsklinikum Charité
Humboldt-Universität Berlin
Ziegelstrasse 5-9
10117 Berlin
Germany
sean.connell@charite.de

Andrew G. Cridge
Dept. of Biochemistry
University of Otago
P.O. Box 56
Dunedin
New Zealand

Thomas E. Dever
Laboratory of Gene Regulation and Development
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, MD 20892
USA
tdever@box-t.nih.gov

Marc Dreyfus
Laboratoire de Génétique Moléculaire
CNRS, UMR 8541
Ecole Normale Superieure
46, rue d'Ulm
75230 Paris
France
mdreyfus@wotan.ens.fr

Alan G. Hinnebusch
Laboratory of Gene Regulation and Development,
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, MD 20892
USA
ahinnebusch@nih.gov

Madina Iskakova
Max-Planck-Institut für Molekulare Genetik
Arbeitsgruppe Ribosomen
Ihnestr. 73
14195 Berlin
Germany

Pavel Ivanov
Max-Planck-Institut für Molekulare Genetik
Arbeitsgruppe Ribosomen
Ihnestr. 73
14195 Berlin
Germany
ivanov@molgen.mpg.de

Denis L. J. Lafontaine
Institut de Biologie et de Médecine
Moléculaires
Université Libre de Bruxelles
Rue des Profs Jeener et Brachet 12
B-6041 Charleroi-Gosselies
Belgium
denis.lafontaine@ulb.ac.be

Louise L. Major
Center for Biomolecular Sciences
North Haugh
The University
St. Andrews, KY 16 9 ST
Scotland, UK

Viter Marquéz
Max-Planck-Institut für Molekulare Genetik
Arbeitsgruppe Ribosomen
Ihnestr. 73
14195 Berlin
Germany
viter@molgen.mpg.de

Knud H. Nierhaus
Max-Planck-Institut für Molekulare Genetik
Arbeitsgruppe Ribosomen
Ihnestr. 73
14195 Berlin
Germany
nierhaus@molgen.mpg.de

Hans-Jörg Rheinberger
Max Planck Institute for the History of Science
Wilhelmstr. 44
10117 Berlin
Germany
rheinbg@mpiwg-berlin.mpg.de

Lluís Ribas de Pouplana
Barcelona Institute for Biomedical Research
C/ Baldiri Reixach, 10-12
08028 Barcelona
Spain
lribasp@pcb.ub.es

Elizabeth S. Poole
Dept. of Biochemistry
University of Otago
P.O. Box 56
Dunedin
New Zealand

Paul Schimmel
Skaggs Institute for Chemical Biology
and Depts. of Molecular Biology and Chemistry
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, CA 92037
USA
schimmel@scripps.edu

Nahum Sonenberg
Dept. of Biochemistry
McGill University
3655 Promenade Sir William Osler
Montreal,
Quebec H3G 1Y6
Canada
nahum.sonenberg@mcgill.ca

Warren P. Tate
Dept. of Biochemistry
University of Otago
PO. Box 56
Dunedin
New Zealand
warren.tate@stonebow.otago.ac.nz

Daniel N. Wilson
Max-Planck-Institut für
Molekulare Genetik
Arbeitsgruppe Ribosomen
Ihnestr. 73
14195 Berlin
Germany
wilson@molgen.mpg.de

1
A History of Protein Biosynthesis and Ribosome Research

Hans-Jörg Rheinberger

1.1
Introduction

It is a challenge to write the history of protein synthesis, the structure and function of ribosomes and of the other components of translation (see Refs. [1, 11-18] for earlier accounts). Many researchers and research groups have been involved (see Refs. [2-10] for autobiographical accounts), and widely different experimental systems, methods, and traditions of skills have been involved. The efforts to elucidate the protein synthesis machinery were scattered all over the world. Nevertheless, a scientific community of surprising cohesion has developed over time, and a network of shared and standardized procedures has been established. Although the formal connection that kept it together was minimal, its meetings have been milestones of a vigorously ongoing process of investigation for several decades (the meetings took place at Cold Spring Harbor, New York 1969 and 1974, Madison, Wisconsin 1979, Port Aransas, Texas 1985, East Glacier Park, Montana 1989, Berlin, Germany 1992, Victoria, British Columbia, Canada 1995, and Helsingør, Denmark 1999; cf. Refs. [19–26]). Emerging to a considerable degree out of cancer research at its beginning, the field of protein synthesis research has only gradually become an integral part of molecular genetics. To trace the broader context of the emergence of the experimental culture of translation research is the aim of this introductory chapter. All those involved in the work of the period covered here but not mentioned will, if not excuse me, realize that I am aware of my limitations: selective reading, specific idiosyncrasies, and, above all, the structural constraints of writing the history of such a complex, empirically driven research field in such a compressed manner. My historical survey will mainly focus on the decades between 1940 and 1970. The more recent developments will only be summarized at the end, since they will be largely covered in the subsequent chapters of this book.

In May 1959, Paul Zamecnik, who can be regarded as the grand old man of protein synthesis research, had been invited to deliver one of the prestigious *Lectures* at the Harvey Society in New York. He chose to speak about "Historical and current aspects of protein synthesis", and he traced them back to "careful, patient studies" extending, as he said, "over half a century" [27, p. 256]. He then began with Franz Hofmeister [28] and Emil Fischer [29], who recognized the peptide bond structure of proteins;

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

went on to Henry Borsook [30], who realized that peptide bond formation was of an endergonic nature; to Fritz Lipmann [31], who postulated the participation of a high-energy phosphate intermediate in protein synthesis; to Max Bergmann [32], who determined the specificity of proteolytic enzymes; to Rudolf Schoenheimer [33] and David Rittenberg [34], who pioneered the use of radioactive tracer techniques in following metabolic pathways; to Torbjörn Caspersson [35] and Jean Brachet [36], who became aware of the possible role of RNA in protein synthesis; to Frederick Sanger [37], who unraveled the first primary structure of a protein, showing the specificity and uniqueness of the amino acid composition of insulin; and finally to George Palade [38], who gave visual evidence for the particulate structures in the cytoplasm acting as the cellular sites of protein synthesis. This is an impressive list of pioneers, who all, according to Zamecnik, "blazed the trail to the present scene", which in his retrospect inadvertently had assumed the character of a royal path to present knowledge. It was not until the very end of the lecture that he relativized this linear perspective: "From a historical vantage point", he said, "too simple a mechanistic view [has] been taken in the past. [The] details of the mechanisms at present unfolding were *largely unanticipated*". We may ask, then, how the unanticipated was brought into being. Zamecnik's answer was: "By the direct experimental approach of the foot soldiers at work in the field" ([27, p. 278], emphasis added). I hope that the following lines offer at least a trace of the history of quirks and breaks that mark protein synthesis research as a collective and multidisciplinary endeavor whose outcome, as with science in general, cannot be told in advance. Scientists usually tell their stories from the point of view of those selected insights that have made their career. No historian can escape this retrospective valuation either, but we should, at least, try to remain aware of its shortcomings.

1.2 The Archaeology of Protein Synthesis – The 1940s: Forgotten Paradigms

The early 1940s were the heydays of what Lily Kay [39] has aptly described as the 'protein paradigm of life'. The transformation experiments of Oswald Avery and his colleagues at the Rockefeller Institute notwithstanding [40], proteins for quite some time continued to be seen as the key substances, not only of biochemical function, but also of hereditary transmission (from Delbrück [41] to Haurowitz [42]). It is surprising then to learn that, despite this early focus on proteins, the mechanism of protein synthesis largely remained a black box throughout the 1940s. Thoughts on mechanism during that decade mainly centered around the conception, favored by eminent biochemists of the time such as Max Bergmann and Joseph Fruton, that the mechanism of protein synthesis might be based on a reversal of proteolysis [43, 32, 44]. Max Bergmann, then at the Rockefeller Institute in New York, investigated the specificity of proteolytic enzymes, and it was in his laboratory that Paul Zamecnik, as a postdoctoral fellow in 1941–1942, became interested in protein synthesis. The proteolysis concept, however, remained a controversial issue, especially since it could hardly be reconciled with the endergonic

nature of peptide bond formation that appeared to be evident from Henry Borsook's investigations at the California Institute of Technology in Pasadena. His measurements favored the idea that the formation of peptide bonds might involve some sort of activation of amino acids prior to their condensation, a topic on which Fritz Lipmann [31] as well as Herman Kalckar [45] had speculated as early as at the beginning of the 1940s (for an assessment of the 'multi-enzyme program' of protein synthesis, its neglect in the history of biochemistry and its resurrection in biotechnology, see Ref. [46]). These considerations, however, remained without conclusive experimental evidence for the next 15 years. Classical biochemistry alone did not provide a definite handle on the question of the cellular mechanisms of protein biosynthesis, despite the growing sophistication of experimental enzymology and of the structural, physical, and chemical analysis of proteins, including powerful new devices such as chromatography, electrophoresis, and X-ray crystallography (see Refs. [47, 48] for historical accounts).

Some observations on the part of cytochemistry were intriguing but also remained erratic for the time being. Around 1940, Torbjörn Caspersson from Stockholm and Jack Schultz from the Kerckhoff Laboratories in Pasadena had developed techniques for measuring the UV absorption of nucleic acids within cells as well as UV microscopy of cells [35]. With that, they were able to correlate growth, i.e., the production of proteins, with the increased presence of ribonucleic acids at certain nuclear and cytoplasmic locations. Around the same time, Jean Brachet and his colleagues Raymond Jeener and Hubert Chantrenne in Brussels reached similar conclusions on the basis of differential staining and *in situ* RNase digestion of tissues [36].

The elucidation of the particulate structure of the cytoplasm by means of high-speed centrifugation dates back to the 1930s and derives from still other lines of research. Normand Hoerr and Robert Bensley in Chicago had used centrifugation to isolate and characterize mitochondria [49]. Albert Claude, in James Murphy's Laboratory at the Rockefeller Institute, was working on the isolation of Peyton Rous' chicken sarcoma agent when he, around 1938, incidentally realized that the particles he was sedimenting from *infected* cells had exactly the same chemical constitution than those sedimented from *normal* chick embryo tissue and thus were cellular constituents [50]. Figure 1-1 shows early dark-field microscopic images (segments 1 and 5) of preparations of Claude's "small particles". (All the figures in this introductory historical chapter are reproduced from the original publications, and in three cases from handwritten laboratory notes.) After tentatively identifying his high-speed sediment with mitochondria or fragments thereof for some years, he, in 1943, came to the conclusion that his pellet did contain another class of cytoplasmic particles. They were definitely smaller than mitochondria, and Claude termed them "microsomes" [51] accordingly. In contrast with the mitochondria, these particles were particularly rich in ribonucleic acid – Claude estimated them to consist of 50% lipids, about 35% proteins, and some 15% nucleic acids. Speculating that they might be self-replicating nucleoproteins, he was tempted to place them in the category of 'plasmagenes', a notion associated with the idea – widely discussed at the time – of some form of cytoplasmic inheritance [52]. But although these particles were reported to carry

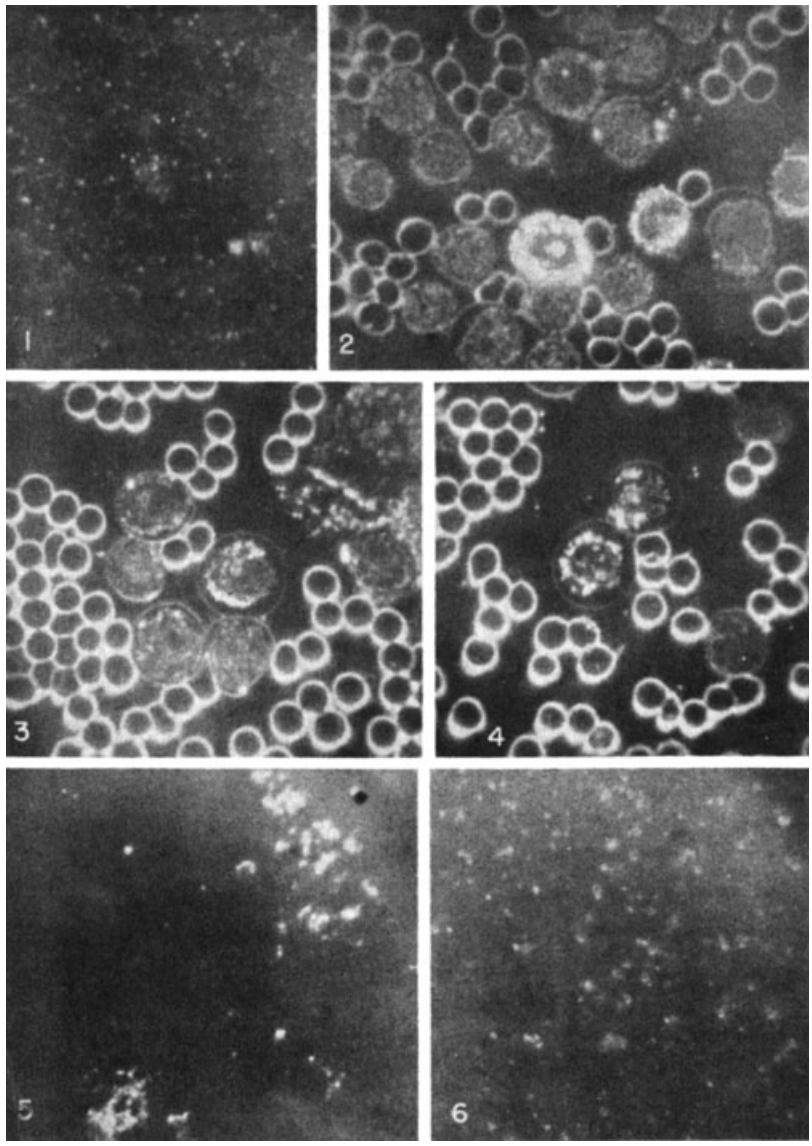


Figure 1-1 Cytoplasmic particles derived from differential high-speed centrifugation. 1–6, Dark-field photographs, magnifications 1000×. 1. Rat leukemia: 'small particles', purified, in neutral water; 2. rat leukemia: whole blood showing cytoplasmic granules in leukemic and normal cells; 3. rat leukemia: heparinated blood; cytoplasmic granules in lymphoid cells; 4. rat leukemia: heparinated blood; cytoplasmic granules in lymphoid cells; purified particles added to plasma; 5. Guinea pig liver: 'small particles' agglutinated at pH 6; phosphate buffer; 6. Guinea pig liver: purified 'Bensley' granules in neutral water. 'Small particles' are represented in segments 1 and 5 (Ref. [301], Figures 1–6).

varying amounts of oxidases and hydrolases [53], they tenaciously resisted all attempts by Claude and his collaborators, especially Walter Schneider and George Hogeboom, to correlate specific and unique enzymatic functions with them [54]. At that time, assigning functions essentially meant enzyme mapping. Unfortunately, this procedure did not work with microsomes. In contrast, however, the microsomes became preferential objects of ultracentrifugation. The centrifugation methods of Hubert Chantrenne [55] from Brachet's laboratory in Brussels, and of Cyrus Barnum and Robert Huseby [56] from the Division of Cancer Biology at the University of Minnesota in Minneapolis were more sophisticated than the Rockefeller method and pointed to a greatly varying size of the particles – if they had a definable size at all. Despite Brachet's recurrent claim of a close connection between microsomes and protein synthesis, no particular experimental efforts were made in all these studies to enforce this line of argument. Still, by the end of the 1940s, Albert Claude thought that microsomes were most probably involved in anaerobic glycolysis [57]. However, the various efforts of an *in vitro* characterization of the cytoplasm by means of ultracentrifugation resulted in a set of procedures for the gentle isolation of cytoplasmic fractions – especially centrifugation through sucrose solutions [58] – that soon proved very useful in a wide variety of other experimental contexts.

1.3 Basic Mechanisms – The 1950s

This situation was bound to change between 1945 and 1950 through still another approach to assess metabolic events. Right after World War II, low-energy radioactive tracers, especially ^{35}S, ^{32}P, ^{14}C, and ^{3}H, became available for research to a wider scientific public as a byproduct of expanding reactor technology. The ensuing new attack on the mechanism of protein synthesis by way of radioactive amino acids was embedded in a particular, historical conjuncture of interests that benefitted greatly from the vast resources made available for cancer research after the War [59], and from the efforts of the American Atomic Energy Commission to demonstrate the potentials of a peaceful use of radioactivity [60, 61]. In fact, cancer research programs provided the background for much of the protein synthesis research during those years. Cancer was related to abnormal growth, and growth was considered to be intimately linked with the metabolism of proteins. This constellation also explains why much of protein synthesis research during the decade between 1950 and 1960 was done on the basis of experimental systems derived from higher animals, especially rat liver, and not on bacteria, as might be expected from hindsight.

1.3.1 Steps toward an *in vitro* Protein Synthesis System

The first attempts at approaching protein synthesis via tracing consisted in administering radioactive amino acids to test animals and in following the incorporation of the label in to the proteins of different tissues. However, radioactively labeled amino acids were not yet commerically sold and were therefore available only in limited

amounts at that time. In addition, the tracing technique posed problems of controlling the experimental conditions. One of the biggest concerns of these early radioactive *in vivo* studies was to maintain control over the specific activity of the injected material. Consequently, researchers in the field attempted to establish test tube protein synthesizing systems from animal tissues. Among the first to use tissue slices – a kind of hybrid system between *in vivo* and *in vitro* – were Jacklyn Melchior and Harold Tarver [62], as well as Theodore Winnick, Felix Friedberg and David Greenberg [63], all from the University of California Medical School at Berkeley. Going one step further, attempts to incorporate amino acids into proteins of tissue homogenates were also made at that time by Melchior and Tarver [62], by Friedberg et al. [64], and by Henry Borsook's team at Caltech [65]. Initially, they all used different amino acids: sulfur-labeled cysteine and methionine (Tarver), carbon-labeled glycine (Greenberg and Winnick), and carbon-labeled lysine (Borsook). All these labels were incorporated, but some of the amino acid 'incorporations' in these early *in vitro* studies turned out to be due to amino acid turnover reactions that were not related to peptide bond formation. Granting that the experimental observation of amino acid 'uptake' indeed meant peptide bond formation became one of the biggest concerns of all those trying protein synthesis in the test tube between 1950 and 1955.

I cannot follow all these activities in detail here. Instead, I will organize my narrative around the efforts of one particular group, thereby illustrating the conjuncture of centrifugation and radioactive tracing through which microsomes became linked to protein biosynthesis. The group is Paul Zamecnik's at the Massachusetts General Hospital (MGH) in Boston, whose work can rightly be considered to have been at the cutting edge of the field for the decade between 1950 and 1960. Zamecnik started his work on protein synthesis in 1945. As a medical doctor, he had an interest in the action of carcinogenic agents. Protein metabolism seemed to him to be a suitable target for studying the differences between normal and neoplastic tissue. The choice of rat liver followed from this comparative interest; a standardized procedure of inducing hepatomas in rat belonged to the laboratory routines at MGH.

In 1948, Robert Loftfield, an organic chemist from the Radioactivity Center at MIT, joined the staff of the Massachusetts General Hospital as part of a collaboration of the Center with the Huntington Laboratories. In the preceding 2 years at MIT, he had worked out a suitable method for the synthesis of ^{14}C-alanine and glycine [66]. Together with Loftfield, Warren Miller, and Ivan Frantz, Zamecnik started to introduce radioactive amino acids into the livers of rats. Miller, from the Physics Department of MIT, had been involved in the development of a new method of radioactive carbon gas counting. Ivan Frantz, who belonged to the Huntington Laboratories, was an expert in the technique of incubating sliced livers.

Right in the first series of these liver slice experiments, cancer tissues proved to be considerably more active than normal liver in taking up radioactive amino acids. But the signal that redirected the research process came from a control. In the laboratory of Fritz Lipmann, who was a neighbor of Zamecnik's at MGH, William Loomis had just shown that dinitrophenol (DNP) specifically interfered with the process of phosphorylation [67]. When the Zamecnik group included DNP into one of their slice

experiments, it stopped all amino acid incorporation activity. The result suggested that, as Lipmann had assumed for a long time, protein synthesis was indeed coupled with the utilization of phosphate bond energy [68]. At that point, the research perspective of Zamecnik's group began to shift from the cancer-related problem of malignant growth to the bioenergetic aspects in the making of proteins.

There was no chance, however, to approach the problem by further manipulating liver slices. But to proceed along the lines of cell homogenization meant, as Zamecnik remarked, to enter a "biochemical bog" [69]. It was a largely unexplored experimental field, and the MGH group worked for 3 years, from 1948 to 1951, to arrive at something that could be taken as the 'incorporation' via peptide bond formation of radioactive amino acids into protein in the test tube. In 1951, Philip Siekevitz, who had joined Zamecnik's group in 1949, had achieved a preliminary fractionation of the liver homogenate by means of a regular Sorvall laboratory centrifuge [70]. His main fractions were a mitochondrial fraction, a fraction enriched in what was taken to be 'microsomes', and a supernatant fluid. None of the fractions was fully active when incubated alone. But when all of them were put together again, as can be seen in Fig. 1-2, the activity of the homogenate was restored, although the signal was extremely faint.

In these efforts, the combination of two methodologies had been instrumental: radioactive tracing and differential centrifugation. From a superposition of them, the system acquired dynamic capacities. In 1953, a tiny but decisive detail was incorporated into the system at MGH. It consisted of a slightly altered, gentle homogenization procedure [71]. 'Loose homogenization' enhanced the activity of the cell-free protein synthesis system by a factor of 10. During the same year, the laboratory centrifuge was replaced by a high-speed ultracentrifuge. The new instrument made a quantitative sedimentation of the microsomes possible, leaving behind a non-particulate, soluble enzyme supernatant. As shown in Fig. 1-3, incorporation activity was restored from these two fractions under the condition that the test tube was supplemented with ATP and an ATP-regenerating system [72, 73]. (The investigation of mitochondrial and chloroplast protein synthesis will not be pursued here. It was investigated in parallel. It should also not be forgotten that the nucleus, too, continued to be considered a site of protein synthesis throughout the 1950s; cf. e.g., Ref. [74]).

1.3.2
Amino Acid Activation and the Emergence of Soluble RNA

Towards the end of 1953, Mahlon Hoagland took up his work in Zamecnik's lab, after having spent a year with Lipmann. Figure 1-4 shows Hoagland, Zamecnik, and Mary Stephenson in their laboratory in the mid-1950s. In what later appeared to Hoagland as one of those "vagaries of fortune in science" [10, p. 71], he realized that he could use the technique of 'phosphate-ATP-exchange' developed in Lipmann's lab as a tool in Zamecnik's rat liver system. He proceeded to graft this technique onto the fractionated protein synthesis setup. Within a year, a first partial, molecular

Fraction	Protein per 100 mg. wet weight tissue	Per cent weight	O_2 per hr.	Qo_2	PO_4 esterified per 5 min.	α-Ketoglutarate disappearing per 60 min.	C.p.m. per mg. protein	
							Minus α-ketoglutarate	Plus α-ketoglutarate
	mg.		*microatoms*		*μM*	*μM*		
Homogenate	17.6	100	37.3	24	3.6	19.1	1.4	10.8*
Nuclei + cells	3.3	18	1.6	5	0.0		1.2	2.9
Mitochondria	2.5	15	7.3	36	3.4	4.0	0.9	1.3
Mixed fraction	2.2	14	1.1	6	0.0		1.7	1.1
Microsomes	1.5	12	0.1	1	0.0	0.0	1.6	1.0
Supernatant	7.8	40	1.2	1	0.0	0.0	0.1	0.4
Mitochondria + microsomes	4.0	24	13.8	39	3.0	6.6	1.1	4.2
" + supernatant	10.3	55	17.1	19	4.6	7.9	1.0	6.6
" + microsomes + supernatant	11.8	64	20.8	20	3.5	10.7	0.8	9.8
All fractions	17.3	98	38.9	25	3.4	18.8	0.8	10.5

* 0.012 μM of L-alanine per gm. of protein per 30 minutes.

Figure 1-2 Ability of various rat liver fractions to incorporate radioactive alanine into their proteins. Each sample contained 3.2 μM of adenosine-5-phosphate, 10 μM of MgCl2, 40 μM of α-ketoglutarate, 0.5 ml of phosphate-sucrose solution containing 15 μM of phosphate, and isotonic sucrose to a final volume of 2.0 ml. Each radioactivity vessel also contained 0.08 mg of [1-^{14}C]DL-alanine (760 000 c.p.m.) in addition. Incubation at 37°C. As seen from a comparison of the last two columns, the activity was dependent on the addition of α-ketoglutarate, an oxidative substrate for mitochondria (Ref. [70], Table 1).

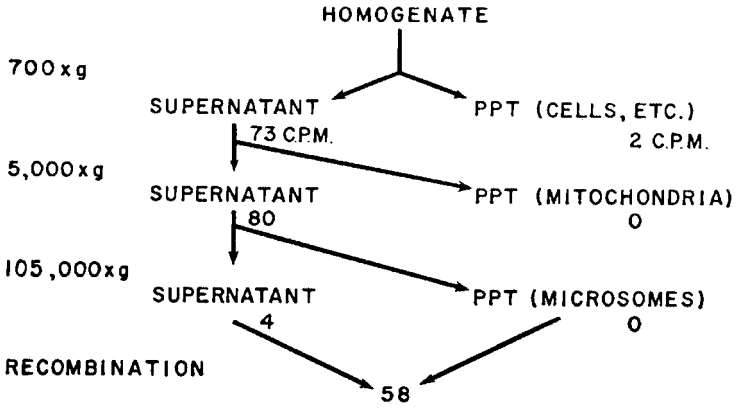


Figure 1-3 Fractionation and recombination of a rat liver extract. The homogenization medium was as follows: 0.004 M $MgCl_2$, 0.04 M potasium phosophate buffer (pH 7.4), 0.01 M HDP, and 0.25 M sucrose. All fractions were prepared from the same homogenization and were incubated simultaneously for 30 min at 37°C in 95% N_2–5% CO_2 (Ref. [103], Figure 2).

model of protein synthesis emerged [75]. The combination of the phosphate-exchange reaction with another model reaction, that of amino acids with hydroxylamine, suggested an activation by ATP of the amino acids as represented in Fig. 1-5. These experiments induced a major turn in the representation of the fractionated system. Its energy requirement became linked to a particular fraction. What until then had been the 'soluble fraction', or the '105 000 × *g* supernatant', or the 'pH 5 precipitate', became now viewed as a set of activating enzymes. With that, amino acid activation began to attract the attention of a larger scientific community.

Several other groups quickly added similar observations obtained in other systems. David Novelli, who had moved from Lipmann's lab to the Department of Microbiology at Case Western Reserve University in Cleveland, established an amino acid-dependent PP/ATP-exchange reaction with microbial extracts [76]. Paul Berg,

Figure 1-4 Left to right: Mahlon Hoagland, Paul Zamecnik and Mary Stephenson, about 1956 [302].

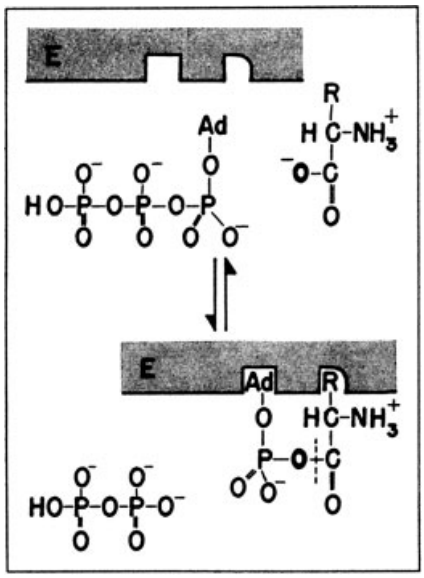


Figure 1-5 Schematic representation of amino acid carboxyl activation by ATP and the pH 5 fraction. Ad, adenosine. The heavily drawn O indicates the attacking carboxyl oxygen which would remain with the nucleotide moiety upon subsequent splitting of the activated compound (dashed line) (Ref. [79], Figure 5).

from Washington University School of Medicine in St. Louis, reported on the activation of methionine in yeast extracts [77, 78]. Lipmann's lab took up the task of isolating and purifying one of the amino acid-activating enzymes. Within half a year, the general character of the carboxyl-activation mechanism appeared to be established [79].

At this point, the participation of ribonucleic acids in protein synthesis still appeared as a black box conveniently termed 'ribonucleoprotein' (cf. e.g., the representation in Fig. 1-6). This black box now attracted the attention of both biochemists and geneticists. Microsomal RNA, by 1955, was generally assumed to play the role of an ordering device, jig, or 'template' for the assembly of the amino acids. The actual point of discussion at that time, however, to which Sol Spiegelman from the University of Illinois at Urbana and Ernest Gale from Cambridge repeatedly referred, was accumulating indirect evidence for a coupling of the synthesis of proteins with the actual synthesis of RNA [80–82]. Also in 1955, Marianne Grunberg-Manago, in Severo Ochoa's laboratory in New York, identified an enzyme which was able to synthesize RNA from nucleoside diphosphates [83]. For the first time, an RNA-synthesizing enzyme had been isolated.

Late in 1955, Zamecnik began to look for RNA synthesis activity in his fractionated protein synthesis system. He added radioactive ATP to a mixture of the

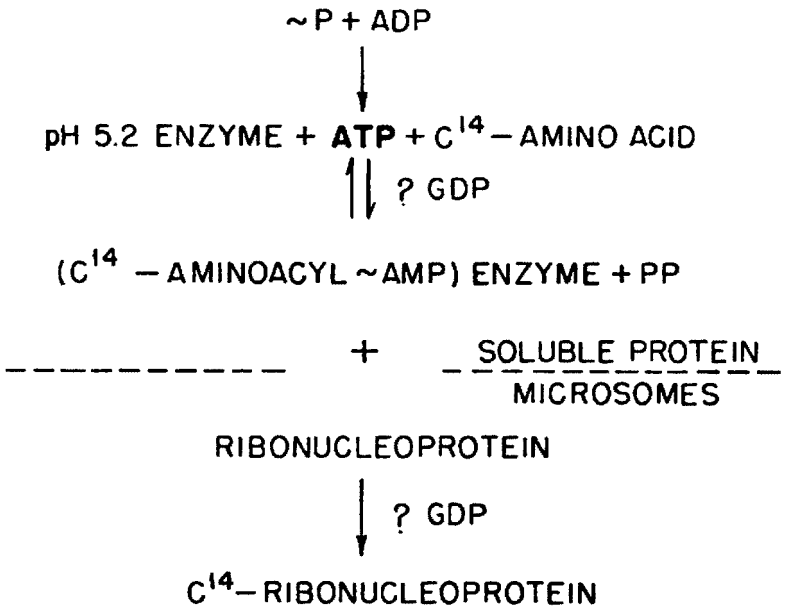


Figure 1-6 Intermediate steps in protein synthesis as seen in 1956 (Ref. [303], Figure 5).

enzyme supernatant and the microsomal fraction. To his astonishment, the nucleotide indeed labeled an RNA component of the system. But there was another, even more puzzling observation. In a parallel experiment, Zamecnik had incubated non-radioactive ATP and ^{14}C-labeled leucine instead of non-radioactive leucine and ^{14}C-labeled ATP together with the fractions. As Zamecnik recorded in his notebook (see Fig. 1-7), the assay suggested – quite contrary to his expectation – that radioactive leucine also became attached to the RNA. In fact, it took another year before Zamecnik, in collaboration with Mary Stephenson and Hoagland, became convinced of the significance of the finding and was ready to publish it [84]. Zamecnik had searched for hints of a *synthesis of RNA* on the microsomes. What he had found was an RNA in the soluble fraction to which *amino acids* were attached. For the time being, the new entity was termed 'soluble RNA'.

Soluble RNA immediately helped to focus research under way in a variety of other laboratories and in a variety of similar systems. The further differentiation of the cell-free protein synthesis system now became the working field for a growing protein synthesis 'industry'. In 1956, evidence for the presence of an RNA intermediate in protein synthesis was being gathered by Robert Holley, from Cornell University. He had found a ribonuclease-sensitive step in the alanine-dependent conversion of AMP into ATP [85]. Paul Berg, soon joined by James Ofengand, went ahead with studies on the amino acid incorporation into soluble RNA of *Escherichia coli* [86]. In 1956, Tore Hultin from the Wenner-Gren Institute in Stockholm had obtained independent evidence for an intermediate step in protein synthesis from kinetic isotope

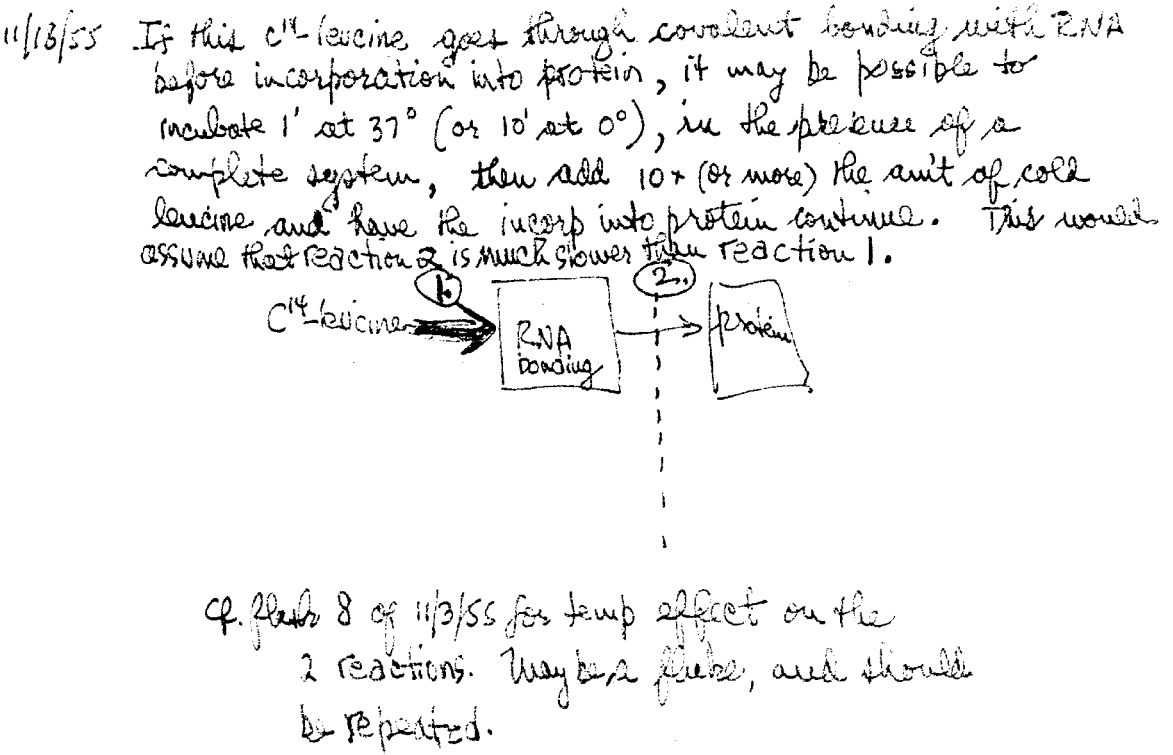


Figure 1-7 First representation of S-RNA as an 'intermediate' in protein synthesis (Zamecnik, laboratory notebook, 10 November 1955, with kind permission from the author).

dilution studies [87]. Kikuo Ogata and Hiroyoshi Nohara at the Niigata University School of Medicine in Japan also had collected hints for an RNA-connected intermediate in protein synthesis [88]. By the end of 1957, amino acid–oligonucleotide compounds were being investigated by at least three other research groups: Victor Koningsberger, Olav Van der Grinten, and Johannes Overbeek [89] at the Van't Hoff Laboratory in Utrecht; Richard Schweet, Freeman Bovard, Esther Allen, and Edward Glassman [90] at the Biological Division of Caltech; and Samuel Weiss, George Acs, and Fritz Lipmann [91], who had moved from the Massachusetts General Hospital to the Rockefeller Institute in New York. All of them joined the race for adding items to the list of what these molecules and their activating enzymes did and what they failed to do. In the process, what had emerged as a biochemical intermediate in protein synthesis soon turned into one of those big missing pieces within the flow scheme of the expression of molecular information. At Richard Schweet's suggestion, the molecule was later referred to as transfer RNA [92], and it became identified with what, based on considerations rooted in the double-helical structure of DNA, Francis Crick had postulated as an adaptor of some sort of the genetic code [93–95]. Figure 1-8 represents the interaction of soluble RNA and microsomal RNA as seen by Zamecnik at the end of the 1950s.

1.3.3
From Microsomes to Ribosomes

As we have seen, it was not until the beginning of the 1950s, and in a context quite different from their original characterization, that the 'small particles' or 'microsomes', operationally defined in terms of fractional sedimentation, optical inspection,

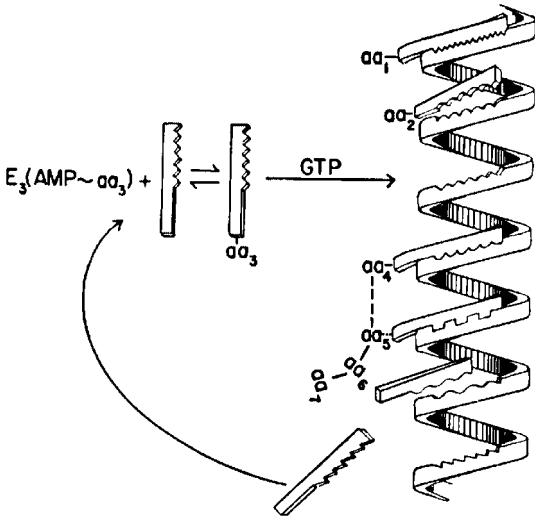


Figure 1-8 A scheme for the interaction of microsomal RNA and soluble RNA-amino acid (Ref. [27], Figure 5).

and chemical composition, became linked on experimental grounds to protein synthesis *in vivo* [97–102] and *in vitro* [70, 96, 102, 103]. Around this time, hints were also accumulating that eukaryotic microsomal material was quite heterogenous in size as well as in composition. It took another decade before the the isolation of active cytoplasmic particles through sucrose-gradient centrifugtion became a laboratory standard. To obtain 'purified' microsomes became one of *the* major issues in the development of cell-free protein synthesis around 1955 [104].

For purification, Zamecnik's colleague John Littlefield took advantage of the detergent sodium deoxycholate which solubilized the protein–lipid aggregates of the microsomal fraction. The RNA-to-protein content (1 : 1) of his particles corresponded to the value given by Howard Schachman from Wendell Stanley's Virus Laboratory in Berkeley for *Pseudomonas fluorescens* particles [105] and by Mary Petermann from the Sloan Kettering Institute in New York for rat liver and spleen particles [106].

Around the same time, George Palade [38], by using an ensemble of advanced specimen preparation techniques, was able to visualize small, electron dense particles on the surface of the endoplasmatic reticulum *in situ* by means of electron microscopy (see Fig. 1-9). Philip Siekevitz had joined Palade in 1954. He added his biochemical expertise to the work at the Rockefeller Institute which aimed at a correlation of the "cytochemical concepts" of microsomal particles and "morphological concepts" derived from electron microscopy [107 pp. 171–172].

Besides electron microscopy, the calibration of these 'macromolecules' involved velocity sedimentation and electrophoretic mobility [106, 108–110]. These structures became a synonym for cytoplasmic RNA, although the postmicrosomal supernatant invariably also contained RNA – approximately 10% of the cell's total RNA [107]. From analytical ultracentrifugation, a sedimentation pattern emerged, and a sedimentation coefficient of the particles could be calculated. Littlefield's rat liver particles appeared as a major 47S peak in the optical record, similar to the main macromolecular component already described by Petermann and their co-workers between 1952 and 1954 (see Fig. 1-10 for Petermann's pattern). A broader peak running ahead of the 47S particle disappeared upon treatment of the material with deoxycholate. However, there was also an additional smaller peak running behind the 47S particle which was not deoxycholate-sensitive. Thus, the suspicion was reinforced that the particulate portion of the microsomal fraction might be in itself heterogenous.

The ribonucleoprotein particles gradually took shape by a comparison of the representations delivered by different biophysical and biochemical techniques applied in different laboratories. The main problem was that the material was no longer active in the test tube after the different isolation procedures. This meant that for the time being there was no functional reference available for comparison. The 'deoxycholate particle', for instance, entered the field of *in vitro* protein synthesis around 1953, and around 1956 it disappeared again from the scene because nobody had succeeded in rendering it functionally active. Preparation procedures played a dominant role, and the terminology faithfully reflected their operational character. Successively, the cellular component at issue had changed from a sedimentable entity no longer visible

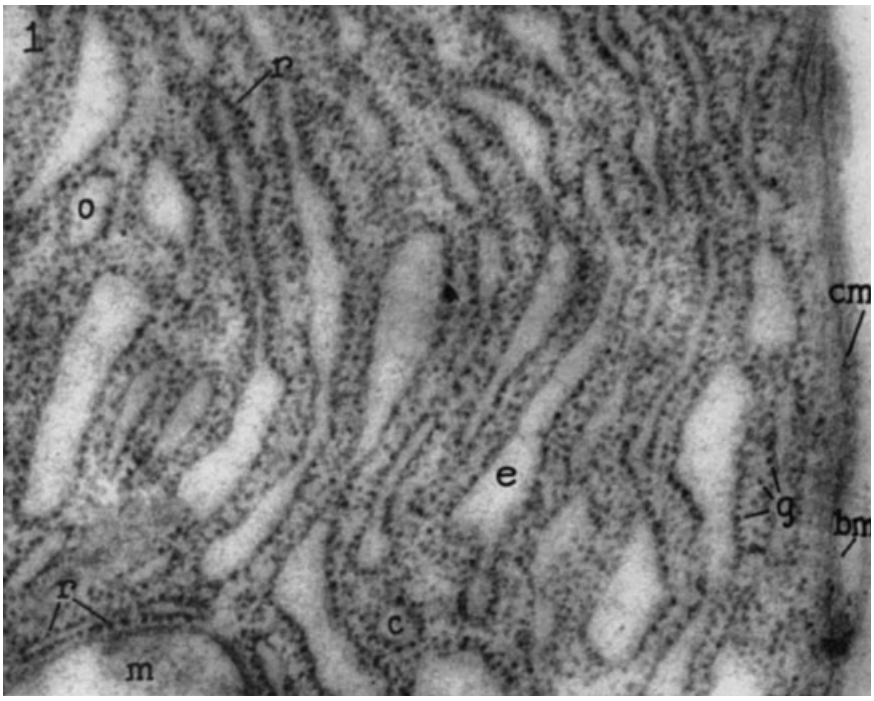


Figure 1-9 Electron micrograph of a limited field in the basal region of an acinar cell of the pancreas (rat). The cell membrane (cm) is coated towards the exterior by a poorly defined layer of dense material (bm), which may be the equivalent of a basement membrane. Part of a mitochondrial profile appears at m. The rest of the field is taken up by elongated (e), oval (o), and circular (c) profiles of the endoplasmic reticulum. Note that in the matrix there are numerous small and dense granules (g) which appear to have particular affinity for the membrane limiting the cavities of the endoplasmic reticulum. The outside surface of this membrane is actually covered by many such particles which in a few places (r) appear to be more or less regularly disposed in rows. Magnification 73,000 × (Ref. [38], Figure 1).

under the light microscope, the 'microsome', to a granular cytoplasmic constituent which was 'deoxycholate-insoluble', and finally to a 'ribonucleoprotein particle' presumably involved in amino acid incorporation into protein, consisting of half protein and half RNA, and visible under the electron microscope. Following a lingering trajectory, the different means and modes of representation eventually produced particles that became firmly linked with subcellular morphology, in particular the endoplasmic reticulum, and to the biochemistry of protein synthesis. The match was, however, hardly perfect.

In the course of the 1950s, RNA-containing particles had attracted more and more attention. Around 1955, their RNA was generally assumed to provide the template upon which the amino acids were assembled into protein threads. In 1958, Howard Dintzis coined the term 'ribosome' for purified microsomes devoid of membrane fragments (Wim Möller, pers. comm.; see also Refs. [111, 112]). During the following

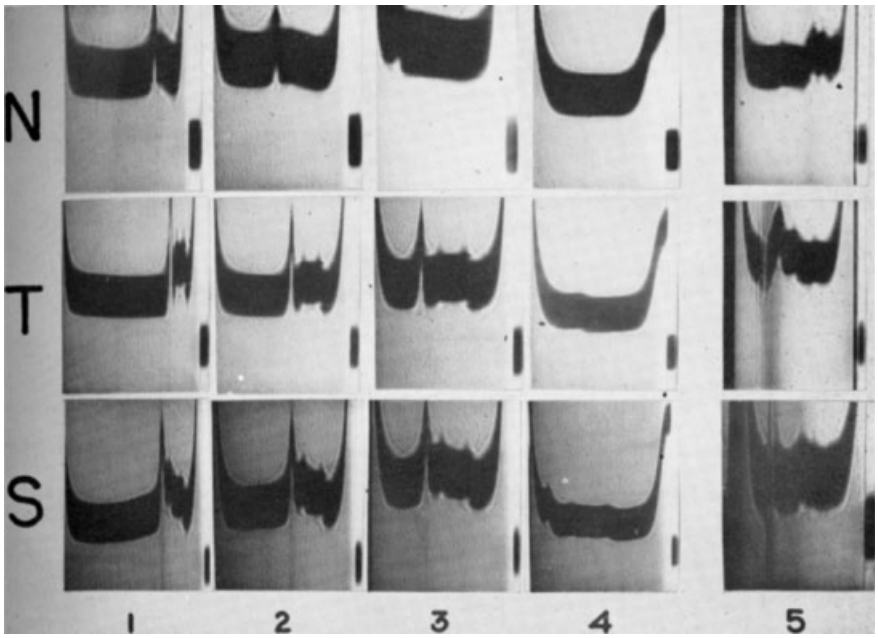


Figure 1-10 Sedimentation patterns of normal and leukemic spleen particles. The direction of sedimentation is to the left: N, normal spleen; T, transplanted leukemia; S, spontaneous leukemia (Ref. [108], Figure 1).

years, this neologism made its way into the laboratories and into the literature. The reason for changing the name was the presumed role of the particle's RNA. The new designation no longer reflected a mere technical representation, but a biological function. Like 'transfer RNA', the 'ribosome' began to relocate protein synthesis from biochemistry to molecular genetics, transforming it into an integral part of what Crick, apparently without minding about the theological connotations of the term, had called the "central dogma" of molecular biology [113, p. 153]. It codified the notion that the genetic information makes its way from DNA to RNA to protein and that, once in the protein, it cannot get back into DNA. The central dogma subsumed the process of protein synthesis as the final, translational, step in the overarching process of gene expression.

With respect to their physical parameters, the protein synthesizing particles considerably changed their appearance between 1955 and 1960. Around 1956 and after many trials, Schachman had found yeast microsomes sedimenting with a velocity constant (*S*) of 80 and to dissociate reproducibly into two unequal portions of 60S and 40S [114]. In a similar manner, Petermann and co-workers were able to separate 78S liver ribosomes into 62S and 46S particles [115]. Alfred Tissières and James Watson, at Harvard, had started to work with *E. coli* ribosomes and had their bacterial particles sediment with 70S. Most interestingly, they could dissociate them

reversibly into a 50S and a 30S component [116, 117]. Gradually, in a decade of painstaking isolation attempts, in which sucrose-gradient centrifugation came to occupy a central place, the confusion about the size of the RNP particles cleared up, and it was realized that the secret of stabilization lay chiefly in the concentration of divalent Mg^{2+} ions. Work on a variety of particles from other sources began to converge on two distinguishing features: bacterial particles (roughly 70S) were consistently smaller than their eukaryotic counterparts (roughly 80S), but both could be separated into something that began to be recognized as a small and a large ribosomal subunit.

1.3.4
Models

The state-of-the-art of protein synthesis, as a process of translation of genetic information, was conceptually re-framed by Francis Crick and his colleagues, especially Sidney Brenner, between 1955 and 1957, and summarized by Crick in his seminal paper of 1958. After years of theorizing from template models, starting with, among others, Hans Friedrich-Freksa [118] and Max Delbrück [41], and continuing with Hubert Chantrenne [119], Felix Haurowitz [42], Alexander Dounce [120], Victor Koningsberger and Johannes Overbeek [121], Fritz Lipmann [122], George Gamow [123], Henry Borsook [124] and Robert Loftfield [125], Crick had come up with a new proposal. During the 1940s, models of autocatalytic protein replication were at the forefront (cf., e.g., Delbrück's scheme [41] and Haurowitz' [42]), as seen in Fig. 1-11. At that time, nucleic acids were still considered, if at all, as structural scaffolds facilitating protein replication. Gene duplication thus meant protein duplication. Friedrich-Freksa [118] had envisaged a protein copying process whereby nucleic acid bases

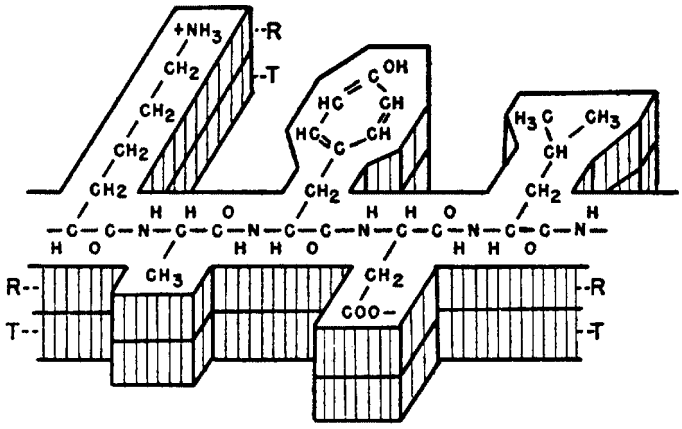


Figure 1-11 Model of protein as template. Replication of a peptide chain formed by lysine, alanine, tyrosine, aspartic acid, and leucine: T, template; R, replica (Ref. [42], Figure 2).

served as a kind of intermediate 'mirror-image'. Later models (such as that of Koningsberger and Overbeek [121] seen in Fig. 1-12) conceived the process of molecular information transfer in terms of a physicochemical interaction between ribonucleic acids and amino acids involving covalent bonding. In the aftermath of the Watson and Crick [126] seminal model of the DNA double helix, Gamow [123] proposed an interaction between DNA and amino acids based on the geometrical shape of holes in the double helix (cf. Fig. 1-13). Crick, thinking of the complementarity features of the DNA double helix, now envisaged what he called "adaptation", i.e., a specific base-pairing interaction between an amino acid-carrying nucleic acid adaptor exposing a signature complementary to the code of a template nucleic acid (cf. Fig. 1-14). It is interesting to note that at the time Crick launched his adaptor hypothesis, he

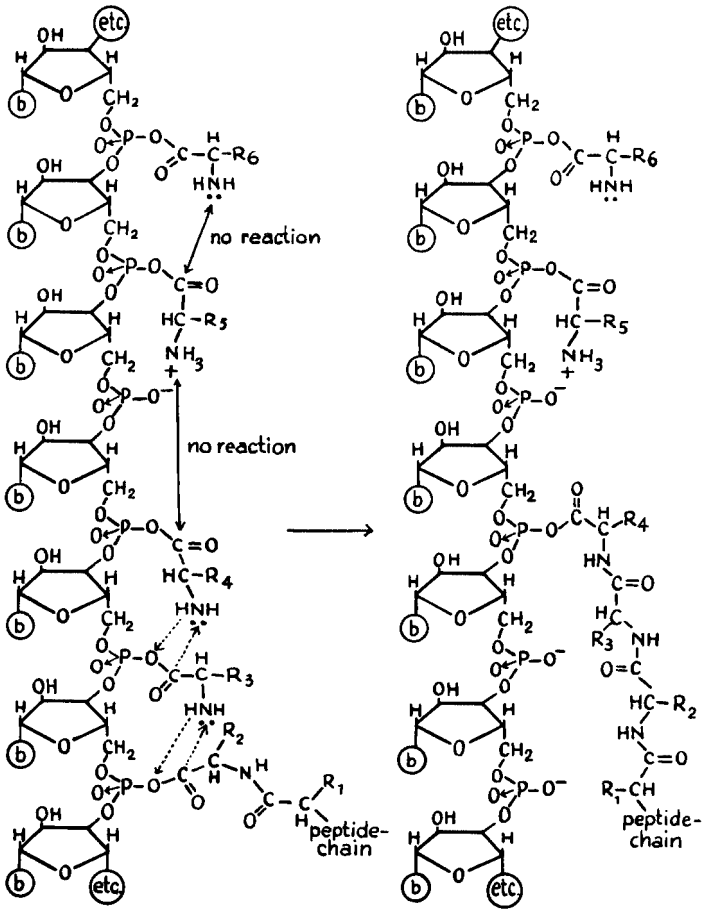


Figure 1-12 RNA as template. Schematic outline of the synthesis of a polypeptide chain starting from an amino acid–nucleic acid compound: b, purine and pyrimidine derivatives (Ref. [121], Figure 2).

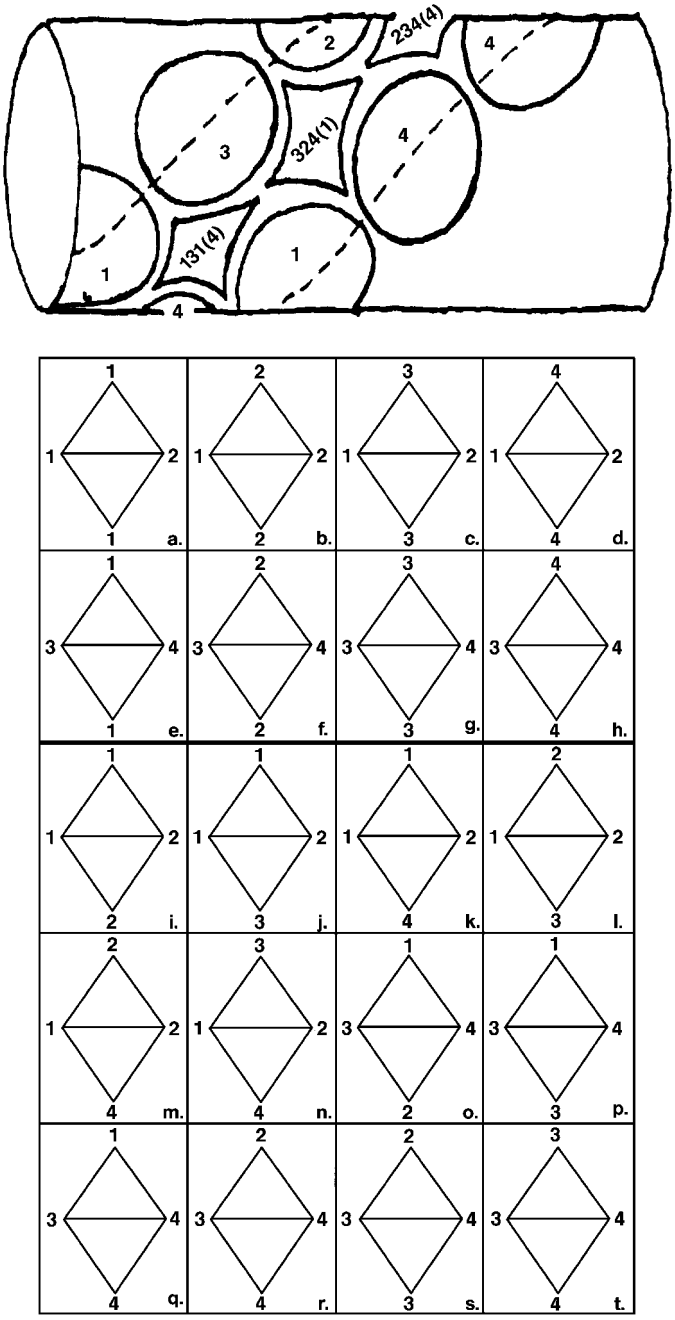


Figure 1-13 DNA as template. Diamond-shaped holes in the double helix of DNA, and the coding scheme for protein synthesis. The rhombs represent the 20 possibilities of arranging the four bases (numbered 1–4) to form niches for the 20 amino acids [123].

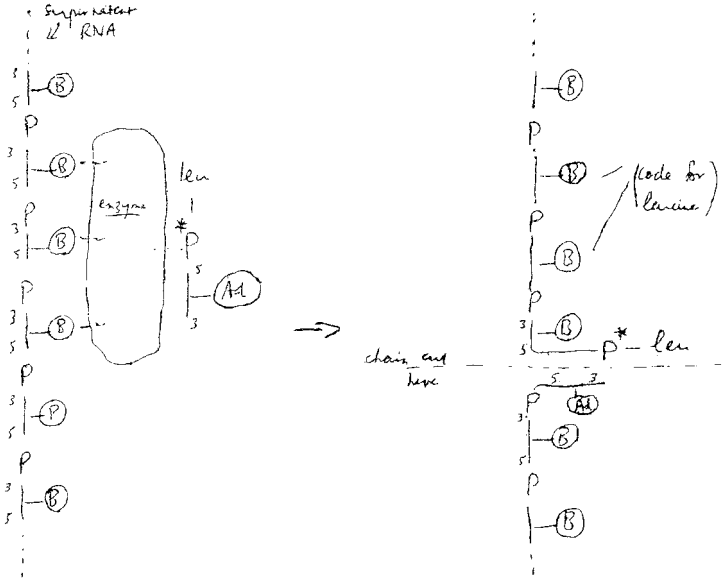


Figure 1-14 Crick's scheme of chopping soluble RNA into trinucleotide adaptors (reprinted from a letter of Crick to Hoagland, January 1957, with kind permission from the author).

obviously did not judge it important enough to be published. It was only its linkage to soluble RNA that made it a prominent concept and a prophecy as seen in hindsight. According to Crick, information-carrying nucleic acids, or templates, and the corresponding proteins were, first, co-linear and second, characterized by strict sequence specificity. But the code itself, the correlation between the building blocks of nucleic acids and those of proteins, remained elusive.

With the surprising emergence of soluble, amino acid-carrying RNAs, the attractive possibility of cracking the code seemed to appear on the horizon. Immediately after Crick had heard the news on soluble RNA from the Massachusetts General Hospital, he invited Hoagland to spend a year with him at Cambridge to isolate an individual S-RNA molecule and to determine its 'signature'. These efforts remained without success. While protein synthesis research during the previous years had certainly not been guided by the theoretical coding discussion, the first attempt to solve the code on the basis of a molecule involved in protein synthesis also failed.

Meanwhile, Zamecnik and Liza Hecht had established as a common feature of all S-RNAs a common 3′-end to which the amino acids became attached: an invariable -CCA trinucleotide [127]. This was anything but a distinct code! Hoagland had hoped to have, with transfer RNA, the "Rosetta Stone" for deciphering the code in his hands [128, p. 61]. But trying to obtain the code through transfer RNA with a direct experimental approach led only to a dead end. Ernest Gale and Joan Folkes at Cambridge, who were analyzing the relation between protein synthesis and the synthesis of nucleic acids in a staphylococcal *in vitro* system, also got stuck with the

characterization of their 'incorporation factors', i.e., nucleic acid fractions that stimulated the incorporation of amino acids into protein [129–131]. And Robert Holley, who since 1957 had put all his efforts into isolating, purifying and sequencing the S-RNA specific for alanine from yeast, took many years and a massive crew of co-workers to arrive at the primary sequence of the first transfer RNA [132]. When he presented the sequence, the code had already been solved by following a completely different experimental track.

1.4 The Golden Age of Translation – The 1960s

The genetic code was solved between 1961 and 1965 with a breathtaking velocity that nobody would have dared to predict even a year before the decisive events. The 1960s also saw the emergence of messenger RNA, the dissection of the ribosome into its components, and the resolution of the translational process into partial functions. Through transfer RNA, messenger RNA, and the code, the biochemistry of protein synthesis merged and for a while even tended to become synonymous with molecular biology, a situation that had been unimaginable a decade earlier when a gap still loomed large between those who considered themselves to be the avantgarde of molecular biology and those who did the messy work of experimentally draining the 'bog' of nucleic acid or protein biochemistry and metabolism [10].

In vitro systems remained central to the field, but the procedures changed. The main strategies of the 1950s had been grounded in what might be called a pursuit of 'integral requirements'. As long as virtually all fractions of the translational system remained black boxes, it would have been deleterious to reduce the system, since this way one could lose essential information. During the 1960s, however, an opposite strategy of 'minimal requirements' became feasible. It was based on the preliminary partitioning of the translational machinery that had been achieved around 1960, the historical stages of which are depicted in Fig. 1-15. It was greatly facilitated through the transition from mammalian systems to bacterial, especially *E. coli* systems of protein synthesis (for the further development of this reductive type of system, see the historical review of Spirin [9]). *E. coli*, so central as a *genetic model* throughout the 1950s, was not yet a *model of translation* during this decade. It was only around 1960 that molecular genetics and protein synthesis research joined forces on the basis of one single model organism. With that, the stage was set for the characterization of some fundamental features of the translational apparatus that still constitute present-day textbook knowledge (see Ref. [133] for an overview of the field around 1960).

1.4.1 From Enzymatic Adaptation to Gene Regulation: Messenger RNA

Towards the end of the 1950s, the work of Jacques Monod and François Jacob at the Pasteur Institute in Paris acquired a new and unforeseen direction and resulted in a

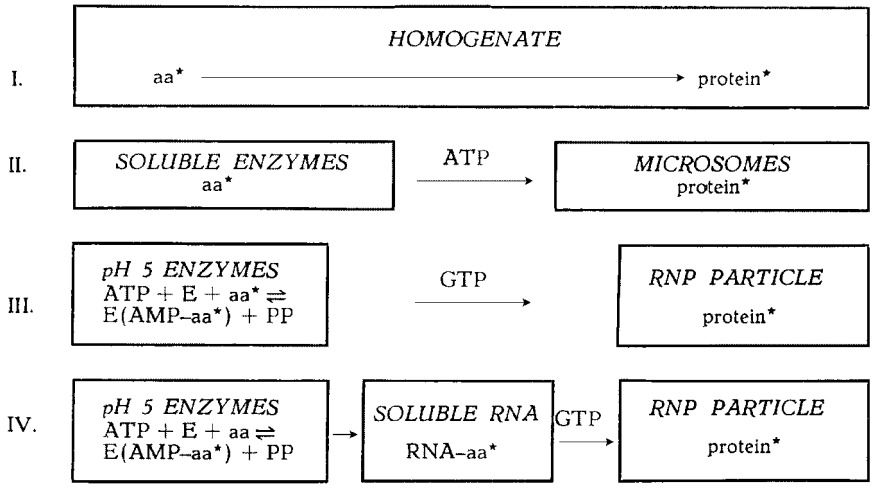


Figure 1-15 Historical stages in the dissection of the rat-liver cell-free system for the incorporation of [^{14}C]-amino acids (aa*) into protein: I, end of the 1940s; II, ca. 1952; III, ca. 1955; IV, end of the 1950s (Ref. [304], Figure 1).

major contribution to understanding protein synthesis and its regulation. Since the beginning of the 1940s, Monod had studied 'enzyme adaptation' in *E. coli*, i.e., the enzymatic response of these bacteria to changes in nutritional conditions. Using a range of mutants, he concentrated on the lactose complex as a model system. At the beginning, Monod's ideas on the subject were shaped by the contemporary theories of immunological instruction. (According to the 'instructional' view, it was the antigen which imprinted the appropriate three-dimensional conformation onto the antibody.) Monod gradually swtiched, at the beginning of the 1950s, to the idea of a genetic control of enzyme synthesis. Conceptually, this resulted in the transition from 'enzyme adaptation' to 'enzyme induction' [134]. Around 1955, Monod and his co-worker Georges Cohen distinguished three genes: the y-gene specifying a permease responsible for the import of lactose into the bacterial cell, the z-gene responsible for the sugar-decomposing ß-galactosidase, and an i-factor responsible for the induction of the system.

François Jacob had started his work on the viral phenomenon of lysogeny in the laboratory of André Lwoff at the Pasteur Institute in 1950. Soon he developed a tight cooperation with Elie Wollman, who had returned from Caltech where he had worked on phage infection in the laboratory of Max Delbrück. Around that time, decisive developments in bacterial genetics were about to take shape. William Hayes in London and Luca Cavalli-Sforza in Milan found hints for a sexual differentiation in *E. coli* bacteria and learned to distinguish between donor and recipient cells during conjugation. In 1951, Joshua Lederberg and Norton Zinder described the phenomenon of viral 'transduction', and Esther Lederberg observed lysogeny in *E. coli* K12. The phage involved in the process was termed 'lambda'. In 1953, Hayes

characterized a high-frequency recombinant donor variant of K12 (Hfr). Soon thereafter, Wollman and Jacob started to work with this lysogenic system. In the process of doing recombination kinetics with multiple mutants of K12, they invented a trick that proved to be highly consequential: If the process of conjugation was interrupted at certain time intervals by mechanical agitation in a mixer, the transfer of different characters could be resolved in a linear fashion. 'Mapping by mating' became a clue to the genetic mapping of bacterial chromosomes [135]. The gene for ß-galactosidase turned out to be in the vicinity of the insertion site of the phage lambda. It was precisely this proximity that allowed an efficient coupling of the systems of Monod and Jacob, respectively.

The collaboration began in 1957 and included Arthur Pardee from the virus laboratory of the University of California at Berkeley. It resulted in the famous series of 'Pa-Ja-Mo' experiments which led to the operon model of gene expression. The experiments suggested that Cohen and Monod's 'i-factor' was responsible for the production of a cytoplasmic substance influencing the structural gene or its product. It was for this special, regulatory substance that Pardee, Jacob and Monod used, for the first time, the term 'cytoplasmic messenger'. 'Messenger' therefore was a concept that originated in the framework of regulatory phenomena and not in a framework of genetic information transfer to begin with. Additional observations pointed to a functionally "*unstable* intermediate" responsible for the expression of the structural genes as well [136, p. 224].

It took some time before Jacob and Sidney Brenner arrived at drawing a parallel between these experiments and the observation of Lazarus Astrachan and Elliot Volkin [137] from the Oak Ridge National Laboratory of a quickly metabolizing RNA that appeared after infection of their bacteria with T2 phages. The question became whether this unstable intermediate was some sort of an "information carrying RNA" [136, p. 225], which transiently combined with existing microsomes, thus inducing the immediate synthesis of a specific protein [6].

There had been hints in the literature pointing towards quickly metabolizing RNAs for quite some time, but obviously they had not been taken into serious account, either by the Pasteur group or by Crick and his molecular biologist colleagues in Cambridge and elsewhere. As early as 1955, microbiologist Ernest Gale, who was a neighbor of Crick in Cambridge, had claimed that in inducible systems, protein synthesis is accompanied by or even dependent upon RNA synthesis [138]. In addition, Sol Spiegelman, who also worked on enzyme induction, had assumed that the RNA templates of induced enzymes are unstable [139].

The concept of microsomes had emerged from eukaryotic *in vitro* systems with reduced metabolic activity, and as it had gained currency towards the end of the 1950s, it was clearly at odds with these observations on bacterial metabolism. Microsomal RNA appeared to be inert, and for all those working on cells from higher organisms, the ribosome represented "a stable factory", already containing an RNA transcript of DNA [10, p. 107], or to use Crick's words at that time: "'Template RNA' is located inside the microsomal particles" [113, p. 157]. Implicit in this vision was a kind of 'one-microsome-one-enzyme-hypothesis', meaning that a particular

ribosome was engaged in the fabrication of one specific protein or set of proteins. Moreover, bacterial *in vitro* systems had a bad reputation in the leading circles of protein synthesis workers in the late 1950s. They were considered 'dirty' systems which were difficult to control [125].

The decisive experiment establishing the role of messenger RNA came from a joint effort of Jacob, Brenner and Matthew Meselson at Caltech: They grew bacteria on heavy isotopes to tag the ribosomes and infected the *E. coli* cells with a virulent phage in the presence of radioactive isotopes. What they found was that newly synthesized radioactive phage RNA indeed became associated with pre-existing heavy ribosomes [140]. 'Messenger RNA' [141] now assumed the general meaning of a molecular information transmitter whose transcription was controlled by feedback loops according to the operon model (see Fig. 1-16). Around the same time, Masayasu Nomura and Benjamin Hall, in Spiegelman's laboratory at Urbana, had characterized a 'soluble' form of RNA synthesized in *E. coli* after bacteriophage T2 infection. It became associated with ribosomes in the presence of high magnesium concentrations [142]. They, however, drew no conclusions with respect to its function. As Nomura recalls, he was "unaware of the new developments, both experimental and conceptual, that were taking place in Cambridge, England, as well as in Paris" [8, p. 5]. And François Gros, Walter Gilbert, and Chuck Kurland, in the laboratory of Watson at Harvard, showed that unstable

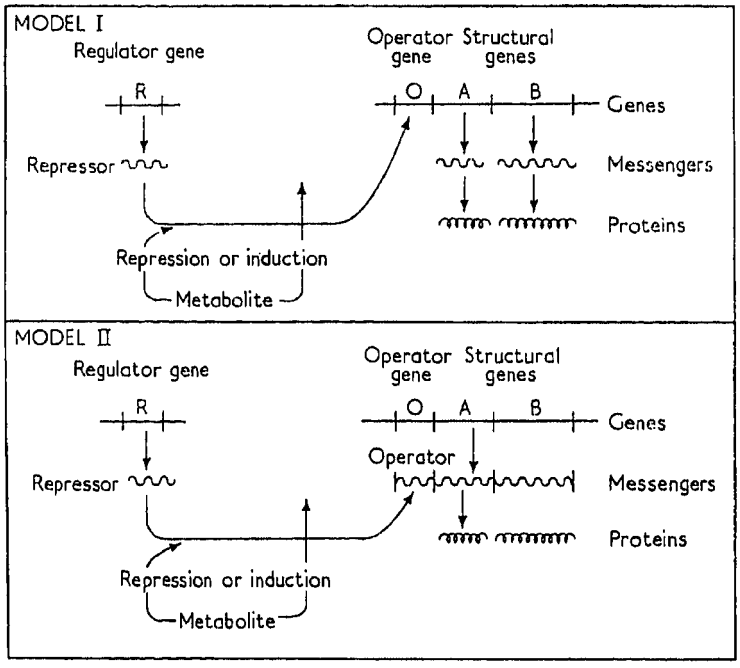


Figure 1-16 Operon models of the regulation of protein synthesis (Ref. [141], Figure 6).

'messenger RNA templates' (cf. the sedimentation pattern in Fig. 1-17) also belonged to the metabolic makeup of uninfected *E. coli* cells [143].

1.4.2
A Bacterial *in vitro* System of Protein Synthesis and the Cracking of the Genetic Code

The differentiation of reliable bacterial *in vitro* systems occurred in parallel, but independent of the experimental context of enzyme induction. The first to report on a system based on *E. coli* were Dietrich Schachtschabel and Wolfram Zillig at the Max Planck Institute for Biochemistry in Munich [144]. Published in German, this paper was ignored by most of the protein synthesis community. In 1958, Marvin Lamborg, a postdoctoral Fellow of the National Cancer Institute from NIH, had come to work with Zamecnik. Zamecnik had tried to obtain a reliable protein synthesizing system from broken *E. coli* cells as early as 1951, but had failed to clean it sufficiently from intact bacteria. Lamborg finally managed to establish a cell-free protein synthesis

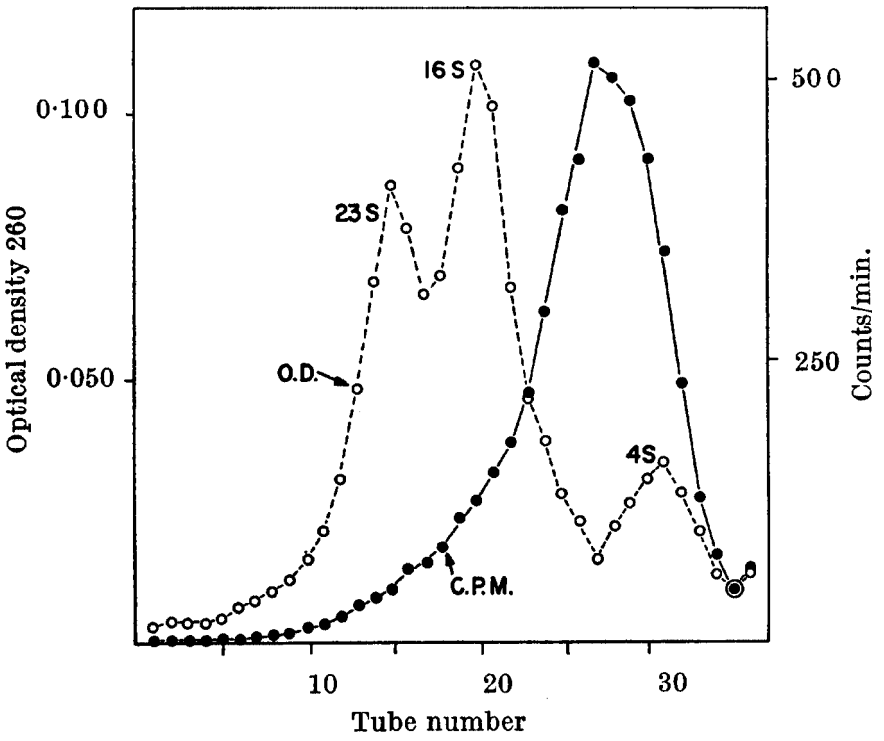


Figure 1-17 Sedimentation of [^{14}C]uracil pulse-labeled (unstable) RNA of *E. coli*. The RNA was run on a sucrose gradient for 10 h at 25 000 rpm: O.D., optical density; 23S, RNA of the large ribosomal subunit; 16S, RNA of the small ribosomal subunit; 4S, soluble RNA (Ref. [143], Figure 8).

system based on *E. coli* extracts [145]. In a rapid dissemination, the Lamborg–Zamecnik type of system made its way into other laboratories and soon became a leading model system for protein synthesis research. Besides Tissières in Watson's lab [146], among the first to use such a system were David Novelli at the Oak Ridge National Laboratory [147], Daniel Nathans and Fritz Lipmann [148] at the Rockefeller Institute in New York, Kenichi Matsubara and Itaru Watanabe [149] at the University of Tokyo and at Kyoto University, and James Ofengand, then on a fellowship at the Medical Research Council Unit for Molecular Biology in Cambridge [150]. In 1962, there were no less than six reports from five laboratories using the *E. coli* system in a rapid publication journal such as *Biochem. Biophys. Res. Commun.*, and seven reports from five laboratories in the biochemical *Fed. Proc.* In Watson's group, with Tissières, Schlessinger, Kurland, Gros, and Gilbert, the structure and function of bacterial ribosomes and messenger RNA had moved to the center of attention. But the *E. coli* system was also being introduced at the National Institutes of Health in Bethesda. The days of the rat-liver system as a pace-maker for unprecedented events were over. Its role was displaced from representation to demonstration: it became marginal. In contrast, the *E. coli* system shifted in the opposite direction. It allowed investigators to bring the genetic code into the realm of experimental manipulation, in a surprising move which left behind all those who had tried to tackle the code through procedures based on virus and phage mutation.

Marshall Nirenberg, at NIH, had just started to establish a cell-free *E. coli* system when Heinrich Matthaei joined him in the fall of 1960. Nirenberg had set himself the task of investigating the steps that connect DNA, RNA and proteins, and synthesizing, in a cell-free system, a specific protein [151]. Despite many efforts (cf., e.g., Ref. [152]), the synthesis of a defined and complete protein *in vitro* had remained a challenge – and a dream – for all those concerned with protein synthesis ever since the end of the 1940s.

If the system was to express a specific protein, conditions had to be found under which it responded to specific templates. This appears to have been the crucial clue in the Nirenberg and Matthaei advances. With respect to the initial phase of the work, there is every reason to assume that Nirenberg and Matthaei proceeded well within the context of the prevailing picture of the ribosome, its RNA still being assumed to play the role of a template. A minor, but finally decisive procedure set the stage for their accomplishment: the preincubation of the bacterial cell extract. Matthaei and Nirenberg put the system to work until its endogenous activity came to a halt. Only then did they add the exogenous RNA. First they found a small, but specific effect with superadded ribosomal RNA. Then, according to a principle of variation, they introduced additional RNAs into the system, such as viral RNA, and finally artificial homopolymers and heteropolymers. It was a lucky coincidence that the synthesis of RNA fell into the special expertise of Leon Heppel, who was the director of the laboratory in which Nirenberg and Matthaei were working. With these polymers at their disposal, they needed only a few months until they, by systematically varying their radioactive amino acids, had deciphered the first code word: The homopolymer polyuridylic acid coded for the artificial protein poly-phenylalanine [18, 153]. Figure 1-18 shows the decisive experiment from Matthaei's laboratory notebook.

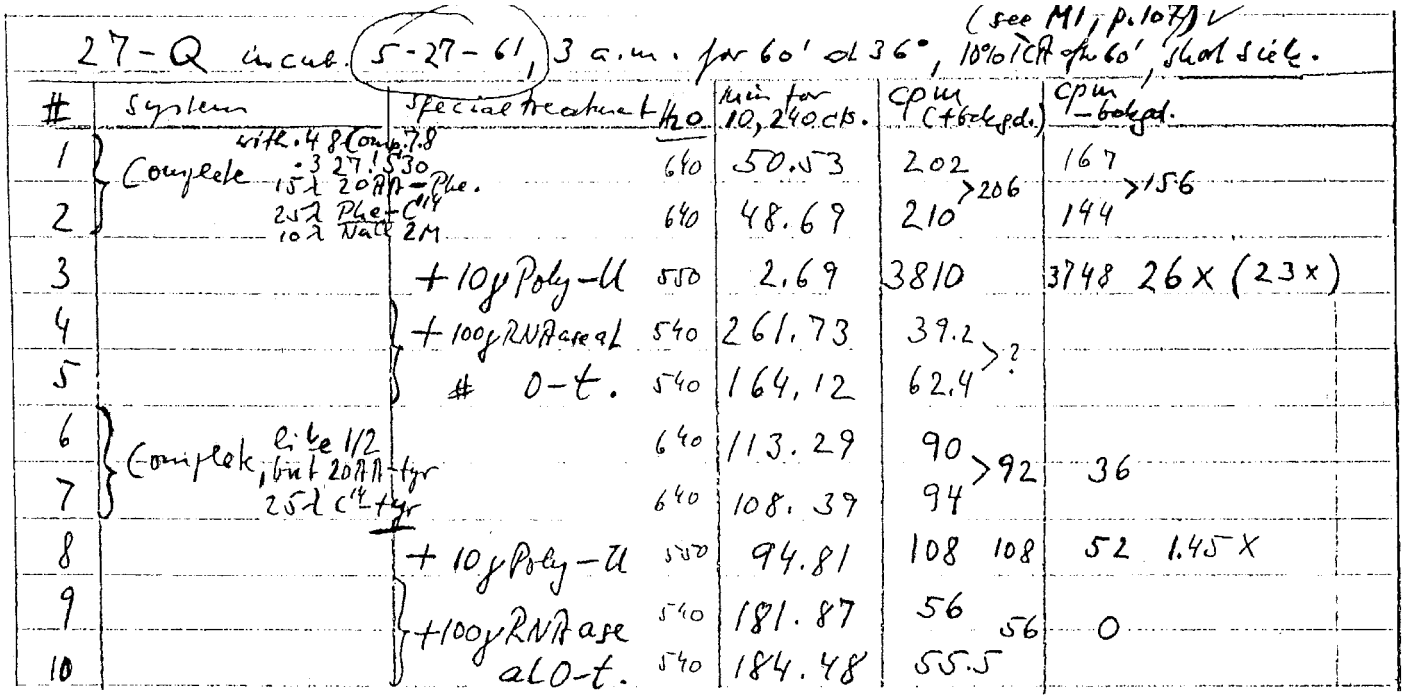


Figure 1-18 First poly(U) assay identifying phenylalanine as the corresponding amino acid (Matthaei, Notebook M2, p. 54, Experiment 27Q from 27 May 1961, with kind permission from the author).

Both Matthaei and Nirenberg, like Nomura, claim to have been unaware of the news from Paris concerning a cytoplasmic messenger at that time [12]. Thus, we have to assume that the concept of messenger arose at least twice in the history of molecular biology. It emerged from two experimental contexts that could not have been more different: from a delicate, genetically triggered *in vivo* system of enzyme induction, and from a comparatively modest, fractionated *in vitro* system of protein synthesis. Despite these radical breakthroughs, microsomal RNA continued to be considered for quite a while as a possible template, at least for ribosomal proteins. We still find this idea, e.g., in the first edition of Watson's *Molecular Biology of the Gene* in 1965 [154].

After the Fifth International Congress of Biochemistry in August 1961 in Moscow, where Nirenberg reported the findings from his laboratory, the other attempts at deciphering the code by genetic and chemical microanalysis of phage mutants in Cambridge and of tobacco mosaic virus mutants in Berkeley and Tübingen could be dropped (see, e.g., Refs. [155–158]). The subsequent hunt for the different code words became a matter of refining the experimental conditions of the *E. coli* system. The triplet binding assay of Philip Leder was one of the key accomplishments in the years to come [159]. Besides Nirenberg, it was mainly Severo Ochoa and his co-workers in New York and Gobind Khorana in Wisconsin who, on the basis of their experience with polymer synthesis, were able to join the race ([160], see Ref. [161] for a review). An initiation codon and the corresponding, formylated initiator tRNA [162, 163] as well as special codons functioning as stop signals were soon identified genetically [164, 165] and biochemically [166–168]. By 1967, the complete code was in place. For the next 10 years, the new findings on translation resulted, along the lines of ever new twists, quirks, and refinements, from the dissection of bacterial systems. After the initial technical difficulties in handling bacterial extracts had been mastered, these systems proved less complex, easier to take apart and simpler to entertain. The relation between eukaryotic and procaryotic systems was reversed. At the end of the decade, it was self-evident that a volume on *The Mechanism of Protein Synthesis* would deal primarily with bacteria, devoting just one special section of 116 pages out of a total of 855 to "Mammalian Systems" [19].

1.4.3 The Functional Dissection of Translation

With the isolation of ribosomes, the purification of specific transfer RNAs and their corresponding synthetases, and the beginning of a deliberate manipulation of viral and synthetic messengers, the stage was set for the dissection of ribosomal function [169, 170]. From the first observations onwards [171], one of the big riddles concerning the energy turnover of peptide elongation had been the involvement of GTP in the process. Around 1960, it had become clear that GTP was not involved in the amino acid-activation reaction *per se*. In a manner still not understood, GTP did interfere with the amino acid transfer mechanism (see discussion in Ref. [172]). The transfer depended on a partial fraction of the pH 5 enzyme supernatant [173]. But attempts to clarify the role of GTP as a co-factor for a presumed 'transfer factor'

had failed so far [148]. It took another 3 years until Jorge Allende and Robin Monro in Lipmann's lab identified an enzyme fraction in *E. coli* whose transfer activity overlapped with a GTPase activity [174] and was termed 'G factor' ([175, 176], see Ref. [3] for a review). At the same time, a complementary 'T factor' was resolved into a temperature-stable component Ts and an unstable component Tu [177]. In bacteria, they became known as elongation factors EF-G and EF-Tu/EF-Ts (EF2 and EF1A/EF1B, respectively, in eukaryotes). In a reticulocyte system, Boyd Hardesty and Richard Schweet, a few years earlier, had already identified two fractions, TF-1 and TF-2, that were involved in the GTP-dependent interaction of Phe-tRNA with poly(U)-programmed ribosomes [178]. The identification of three factors required for the initiation [179, 180], and of factors required for the termination of the translation process soon followed [181]. The characterization, in terms of function and primary as well as tertiary structure, of these transient ribosomal-binding factors continued well into the next two decades. They became model proteins for the study of RNA–protein interactions.

Transfer RNA binding to ribosomes and to their subunits became a major subfield for studying ribosomal function. At the beginning, these studies were still done with eukaryotic microsomes. Among the pioneers were Tore Hultin in Stockholm and Leendert Bosch in Leiden [182–185]. Around the same time, Hoagland, in the process of performing one of the first experiments in which a doubly labeled tRNA was used, observed a "background" phenomenon which he found "difficult to reduce" [186]. The binding of transfer RNA to the rat liver microsomes occurred at zero time, and it took place prior to the amino acid incorporation reaction. A control experiment revealed that uncharged S-RNA bound to the microsomes as well as did S-RNA charged with amino acids. This finding opened the door for detailed studies of the interaction between tRNA and ribosomes.

The majority of the ensuing tRNA binding studies was done in bacterial systems, where the poly(U)-dependent Phe-tRNA binding assay became by far the most prominent. Soon, Walter Gilbert showed that the tRNA carrying the growing polypeptide is associated with the 50S subunit [187], whereas the binding of poly(U) apparently involved the small subunit [188], and the binding of transfer RNA in general depended on the presence of a messenger [189]. Jonathan Warner and Alex Rich found active reticulocyte ribosomes carrying two transfer RNAs [190].

Quantification in these early binding studies was difficult: too many parameters were not yet standardized, and stoichiometric relations could only be estimated. In this situation, a functional and clearcut distinction between two different binding sites of charged tRNAs on the ribosome would be of considerable advantage. Robert Traut and Robin Monro [191] provided it with the puromycin-peptidyltransferase assay which allowed investigators to distinguish a puromycin-sensitive (B-site, later named P-site) and a puromycin-insensitive binding state (A-site) of aminoacylated tRNA (see Fig. 1-19). Based on this observation, the two-site model of ribosomal elongation as shown in Fig. 1-20 became codified by Watson [192] and continued to serve as a reference system for research on ribosomal function well into the 1980s. Many features of translational initiation [193], elongation [194–196]

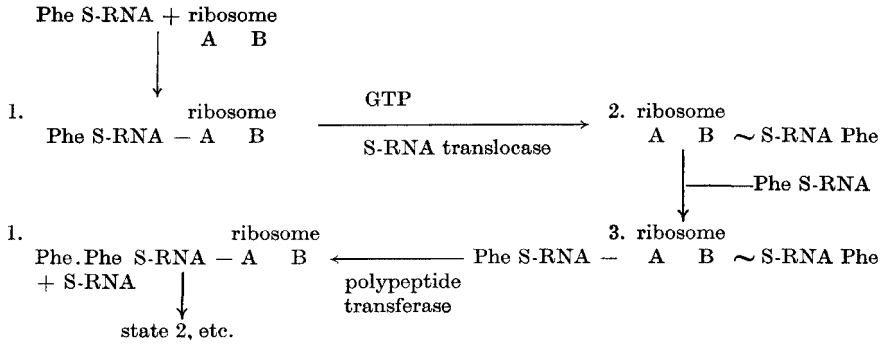


Figure 1-19 Model of the ribosomal elongation cycle as derived from the reaction with puromycin identifying two ribosomal tRNA binding sites, B and A, respectively [191].

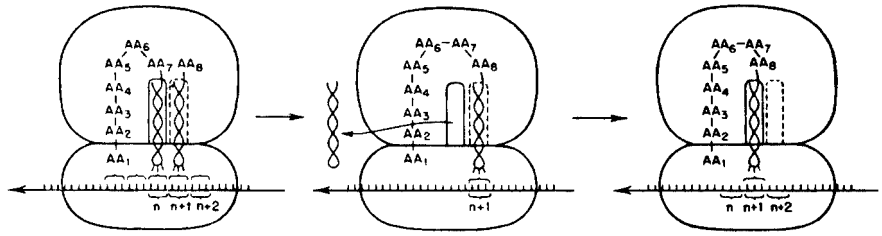


Figure 1-20 Model of the ribosomal elongation cycle comprising two tRNA binding sites (Ref. [192], Figure 20).

and termination [197] were outlined in more and more sophisticated and reduced partial *in vitro* systems (see Ref. [9] for a more recent survey), with acetylated Phe-tRNA (AcPhe-tRNA) becoming a generally accepted analogue for an initiator tRNA and a peptidyl-tRNA analog in the poly(U) system [198].

Antibiotics revealed themselves to be invaluable tools for the dissection of partial ribosomal functions as well as for the ongoing *in vivo* studies concerning regulation, speed, and accuracy of protein synthesis. Among the prominent drugs were puromycin as an elongation terminating agent (see Refs. [199–201] for early studies); chloramphenicol as a specific inhibitor of bacterial peptidyltransferase [202–204]; fusidic acid as interfering with the translocation factor EF-G [205, 206]; and streptomycin as inducing misreading [207, 208]. One of the earliest realistic measurements concerning the accuracy of the process of polypeptide formation came from Robert Loftfield [209]. (For more details about antibiotic effects on ribosomes see Chap. 12).

In the context of pursuing ribosomal function, and after the mRNA concept had been established, gentle isolation of messenger–ribosome complexes became a matter of priority in the early 1960s. Particles larger than 70S or 80S appeared on sucrose-gradient patterns and electron microscopic images (see Fig. 1-21 as an

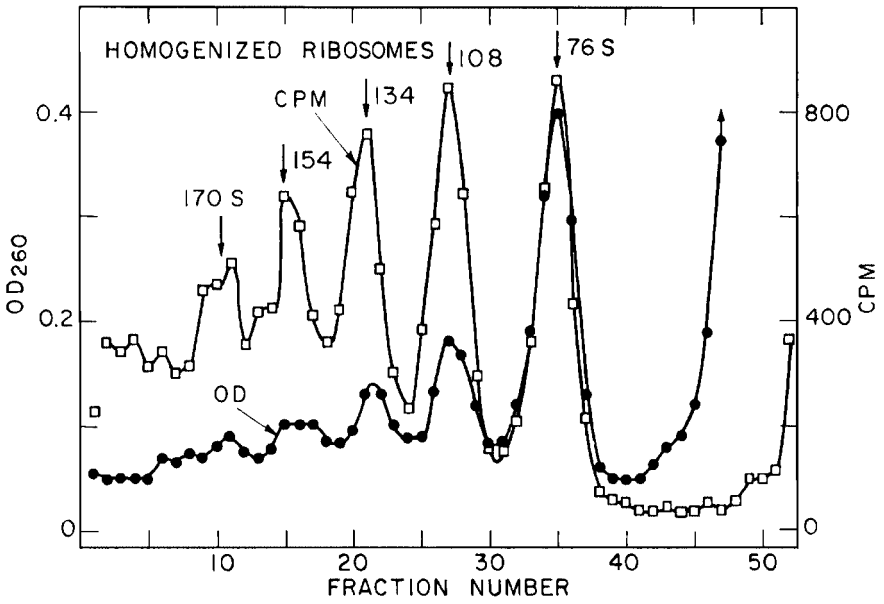


Figure 1-21 Polysomes. Sucrose gradient of lysed reticulocytes after incubation with [^{3}H]leucine. After lysis and low-speed centrifugation, the ribosomes were pelleted three times at 28 000 rpm and resuspended with a homogenizer. A 5–20% sucrose gradient was used. The numbers next to the arrows represent the sedimentation constants associated wih each peak (Ref. [214], Figure 3).

example). They were variously termed 'ribosomal clusters' [210], 'active complexes' [211], 'ergosomes' [212], or 'aggregated ribosomes' [213], before Warner and Rich coined the term 'polysomes' [214] which quickly came into general use. Polysomes appeared to consist of strings of ribosomes occupying a particular messenger RNA. Special isolation procedures were required to prevent them from breaking down to monosomes during fractionation. On the other hand, *in vivo* and *in vitro* evidence grew that ribosomes dissociated and reassociated during their functional cycle [215, 216], and that initiation started on the 30S subunit [217].

Around the same time, Peter Traub, together with Nomura, found the right temperature and ionic conditions for reconstituting the small ribosomal subunit of *E. coli* in the test tube [218] from its RNA and protein moieties, respectively. After the much simpler, symmetric TMV in the early 1950s, the highly asymmetric ribosome became the emblem of molecular self-assembly in the late 1960s and early 1970s (see Fig. 1-22 for Nomura's assembly map of the 30S subunit).

In the following years, multiple attempts to repeat the procedure for total reconstitution of the large 50S subunit from *E. coli* remained without success. A reason for these failures was suspected to be the high activation energy necessary for the 50S

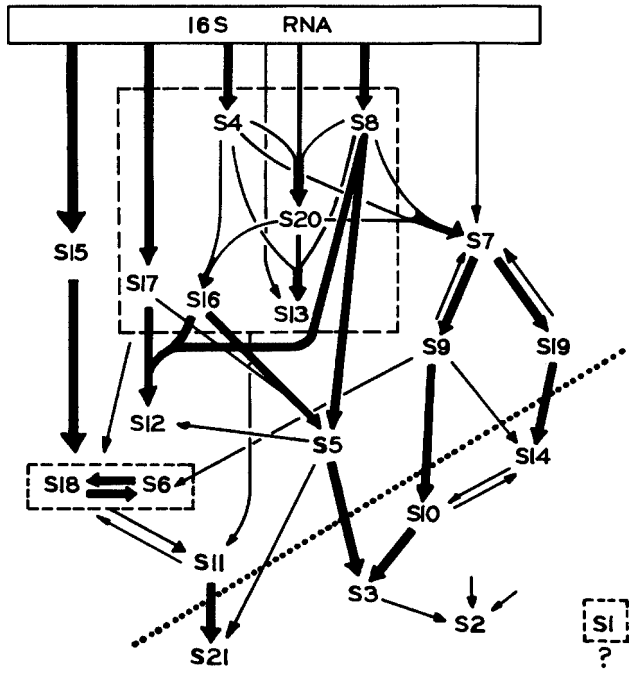


Figure 1-22 Assembly map of *E. coli* 30S ribosomal proteins. Arrows between proteins indicate the facilitating effect of one protein on the binding of another – a thick arrow indicates a major facilitating effect (Ref. [305], Figure 1).

assembly *in vitro*. Accordingly, Nomura and his coworkers shifted their interest to the thermophilic bacterium *Bacillus stearothermophilus*. In 1970, they achieved the first total reconstitution of a large ribosomal subunit [219]. In view of the widespread use of the *E. coli* ribosome as a model organelle, however, the search for a way to reconstitute the *E. coli* 50S ribosomal subunit continued. Four years later, Knud Nierhaus and Ferdinand Dohme succeeded in this task. They developed an alternative, two-step reconstitution method thus obviating the need for incubation at high temperatures above 50°C that would have been required in the previous one-step procedure [220].

The possibility of *in vitro* ribosome assembly opened the field for a multiplicity of structure–function correlation studies at a previously unknown level, including the construction of assembly pathways and maps, detailed interactions between ribosomal proteins and ribosomal RNA, and functional reconstitution experiments where one or more components were omitted (see Chap. 3.1 for details). The hope, however, that a particular ribosomal protein might be singled out as responsible for the peptidyl transferase reaction did not materialize (see Chap. 8.4).

1.4.4 The Structural Dissection of the Ribosome

Throughout the 1950s, the macromolecular composition of microsomes and ribonucleoprotein particles had remained obscure. The original assumption of Watson at Harvard, Schachman in Berkeley, and others who started to analyze bacterial particles, had been that their structure might be analogous to that of RNA viruses: an RNA moiety wrapped with multiple copies of a coat protein. The virus analogy dates back all the way to the beginning of microsome research. Although it had lost its early connotation of a cytoplasmic replicator, the analogy continued to play the role of an obstacle rather than that of a research promoting conceptual tool. It had certainly not been favorable either to the emergence of the concept of messenger RNA, or to the emergence of the view of an asymmetric particle consisting of many different proteins. Resisting the viral analogy, neither ribosomes nor their protein subunits seemed regular enough to form crystals, as had been the case with, e.g., tobacco mosaic virus.

In view of the complex protein make-up of ribosomes, it is not surprising that RNA was the first ribosomal component to be characterized physically and chemically in terms of sedimentation behavior, molecular weight, and overall base composition. As for the ribosomes, so for rRNA, too, sucrose-gradient centrifugation was crucial. Around 1960, there was still considerable uncertainty about the identity of ribosomal RNA. As long as it was considered to represent the template(s) for protein synthesis, there had been, understandably, no reason to assume that rRNA might be homogenous and well defined. In contrast, credit was given to the idea that ribosomal RNA might be composed of smaller templates that became linked within the particle later on, either non-covalently or covalently. Before RNase-free strains of bacteria became available [221, 222], the problem of RNA breakdown during preparation could hardly be mastered. Yet the introduction of the separation of RNA from cellular protein by phenol extraction had already greatly facilitated laboratory manipulation of RNA. This method came into quick and general use soon after its publication [223, 224]. In 1959, Paul Ts'o [225] separated rRNA from pea seedlings and rabbit reticulocytes into two major 28S and 18S peaks. A series of careful studies on *E. coli* ribosomes in Watson's laboratory led Kurland to propose that ribosomal RNA came in two large species, 16S and 23S, respectively [226]. Alexander Spirin in Moscow had reached basically the same conclusion [227]. The question however whether this represented the 'native' state of ribosomal RNA, whether originally they were made up from smaller fragments or derived from a large precursor, continued to be a matter of debate for several years [228]. The controversy eventually came to a satisfactory end when it became evident that mature ribosomal RNA originated from a large transcript that was processed in the event of ribosome formation, and that, indeed, a small defined RNA, 5S RNA, was part of the 50S subunit [229]. Subsequently, 5S rRNA became the first ribosomal RNA molecule to be completely sequenced in 1968 [230]. This breakthrough had been made possible through Sanger's 2D fractionation procedure for radioactive nucleotides [231]. It took 3 years

to determine its 120 nucleotides. In comparison, sequencing the first transfer RNA (yeast $tRNA^{Ala}$) with slightly more than half the number of nucleotides had taken Holley and his co-workers (see Fig. 1-23 for its sequence and putative secondary

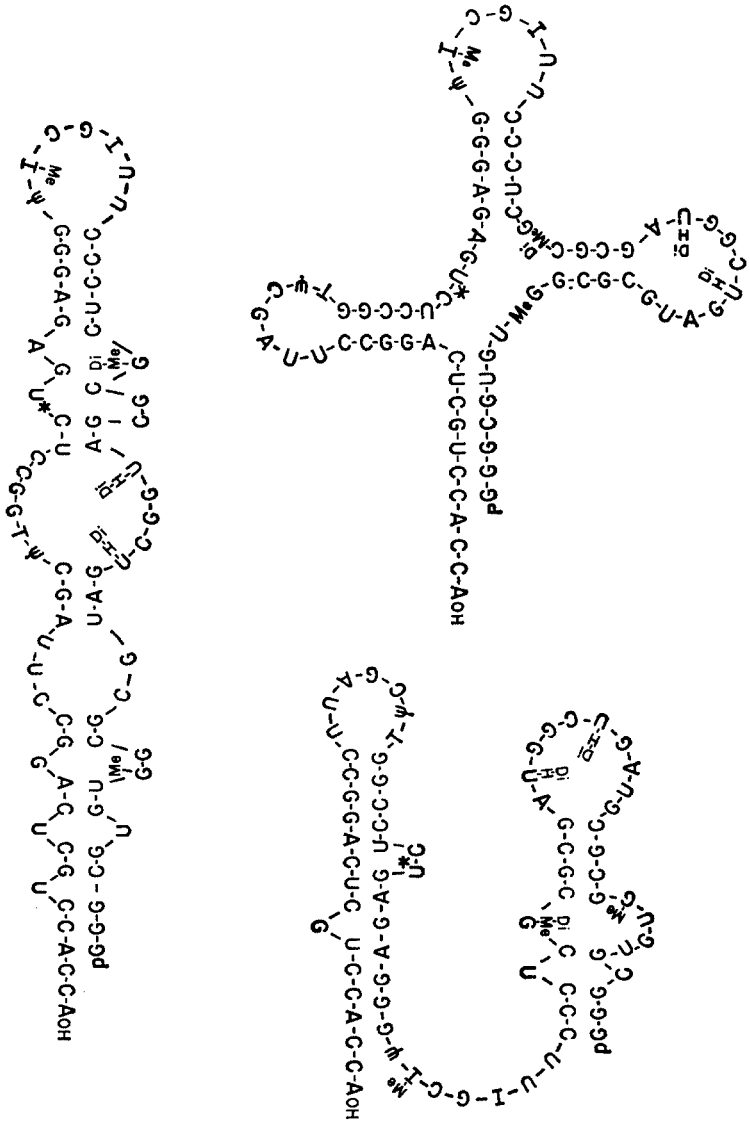


Figure 1-23 Proposed secondary structures for the first tRNA sequenced – $tRNA^{Ala}$ from yeast (Ref. [132], Figure 2). The cloverleaf structure was proposed by Elizabeth Keller who had moved from Zamecnik's laboratory to Holley's (cf. Ref. [11, pp. 282–285]).

structures) some eight years [132]. Other groups soon followed with other tRNA species [232, 233]. The detailed functional elucidation of these molecules, however, had to await further studies. Their crystallization proved to be a major prerequisite for moving forward in this direction (see Refs. [234, 235] among others).

Whether the protein moiety of ribosomes was made up of multiple copies of a single species or of many different proteins, and whether all ribosomes had the same protein composition, was still an open question at the beginning of the 1960s. Serious analysis, on the basis of starch-gel electrophoresis, of the protein composition of ribosomes goes back to the work of Jean-Pierre Waller [236] and to the fractionation studies of David Elson [237] and Pnina Spitnik-Elson [238]. One of the first ribosomal proteins to be characterized individually was the acidic A-protein of the large subunit studied by Wim Moeller and later known as L7 (L12) [239]. Major efforts to develop methods for separating and purifying individual proteins came, among others, from Heinz Günter Wittmann and Brigitte Wittmann-Liebold's laboratory in Berlin [240], Tissière's in Geneva [241], and Kurland's in Wisconsin [242]. A prominent achievement in this endeavor was the separation of all ribosomal proteins by 2D polyacrylamidegel electrophoresis [243] as shown in Fig. 1-24. It served as an efficient and economizing standardization vehicle in the field of ribosomal protein identification.

1.5 1970–1990s: A Brief Synopsis

The survey of the following three decades from the 1970s to the 1990s will be very brief. There is no need to go into the details of an ongoing research in this historical introduction, since the major events during these decades will be extensively dealt with in the following chapters. The 1970s can be regarded as the period of the elucidation of the primary structure of the components of the translational apparatus. Indeed, around the turn of the decade, the ribosome of *E. coli* became the first cellular organelle whose RNA [244–246] and protein components [247] were completely sequenced. Sequencing the complete ribosomal RNA became a feasible task only after the new sequencing methods of Maxam and Gilbert [248], and of Sanger [249] had been introduced.

The emergent recombinant DNA technologies helped to construct a detailed genetic map of the components involved in protein biosynthesis. The ribosomal RNA genes, however, were mapped before the era of recombinant DNA technology. A dozen years had elapsed between their first identification in 1962 [250] and their precise mapping [251]. Knowledge about ribosomal protein genes and operons rapidly accumulated after the subsequent isolation of protein gene-transducing lambda phages [252]. Another source of information was provided by the systematic work with ribosomal protein mutants [253, 254].

Molecular details of ribosomal function also became available, such as the interaction of mRNA with 16S RNA during initiation [255, 256], and the mechanisms by which ribosomes achieve their accuracy [257, 258]. The regulation of ribosome biosynthesis, starting with the early findings on the genetics of RNA synthesis [259],

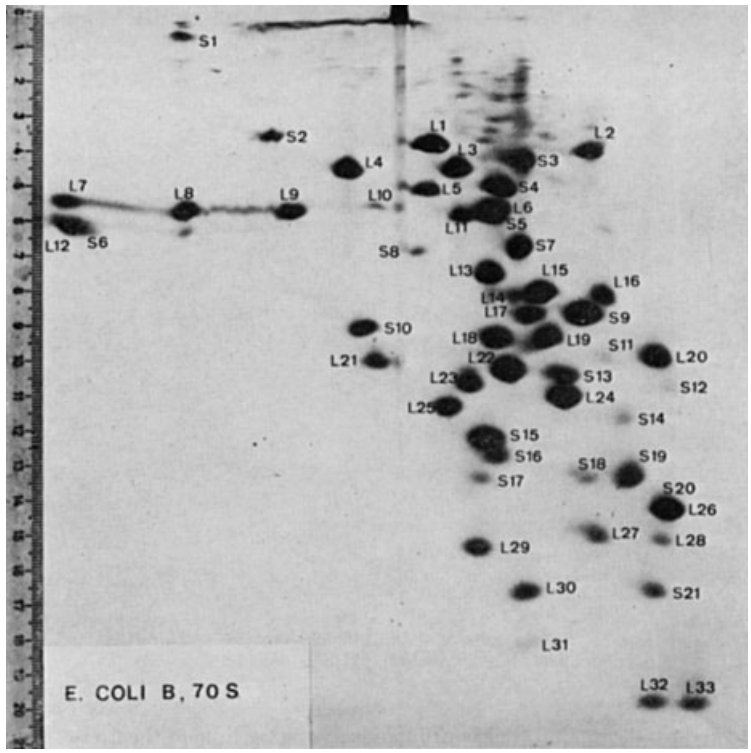


Figure 1-24 2-D electrophoretogram of 70S proteins of *E. coli* B: First dimension, 4% acrylamide, pH 8.6; second dimension, 18% acrylamide, pH 4.6 (Ref. [243], Figure 4).

also became a major field of investigation during the 1970s [260–262]. Over the years, a detailed view, first of transcriptional, then of translational feedback regulation mechanisms emerged. Since Monod and Jacob's work on the lac operon, transcriptional control had been the leading paradigm. The shift of interest from transcriptional to translational regulation was indeed an unprecedented turn. The major events in this area have both been initiated and reviewed some time ago by Nomura [8].

During the 1970s, ribosome research became a focus for the development and application of numerous advanced biochemical, biophysical and biological techniques. *In vitro* reconstitution of ribosomes [263] and *in situ* localization of ribosomal components via immunoelectron microscopy [264, 265], scattering studies [266], cross-linking [267] and affinity labeling [268] led to early insights into the quaternary structure of the protein synthesizing organelle and its functional characteristics such as factor binding and the constitution of the peptidyltransferase center.

The 1980s, on the one hand, were characterized by an increasing backshift of emphasis towards eukaryotic systems (see Ref. [269] for a contemporary overview).

Ira Wool in Chicago had pioneered mammalian ribosomal proteins during the era of *E. coli* (see Ref. [270] for a review), Rudi Planta in Amsterdan had done much of the genetic and structural work on yeast ribosomal RNA (see Ref. [271] for a review). On the other hand, after a lag period, the tedious and time-consuming task of secondary, tertiary, and quaternary structure modelling came to fruition and became linked to ribosomal function. Protein–protein crosslinking [272], protein–RNA crosslinking [273, 274], protection and modification studies [275], neutron scattering [276, 277], electron microscopy [278], and ribosome crystallization [279] figure prominently among the methods involved in this continuing endeavor. On the functional side, exhaustive tRNA-binding studies led to new model conceptions of the elongation cycle involving a third tRNA binding site [280–283, 275]. Peter Moore has judged on this topic: "The two-site model for the ribosome, which the world has accepted for a generation is dead. The existence of a third site for tRNA binding, the exit site, is now established beyond reasonable doubt. This is unquestionably the most significant advance in our understanding of the ribosomal events of protein synthesis in many years" [284].

Finally, the 1990s were dominated by major efforts to carry the structural analysis of ribosomes to atomic resolution. The availability of suitable crystals of ribosomes and ribosomal subunits, particularly from thermophilic and halophilic sources, and the solution of the phasing problem led to a proliferation of X-ray crystallographic studies to which Wittmann and Ada Yonath [285] in Berlin and the group at Pushchino [286] had laid the ground with their ribosome crystallization initiatives in the 1980s. After almost 20 years of continued efforts, atomic resolution has now been achieved for the large ribosomal subunit from *Haloarcula marismortui* [287] and *Deinococcus radiodurans* [288], the small subunit from *Thermus thermophilus* [289, 290], and near-atomic resolution for the 70S-tRNA–mRNA complex [291]. In parallel, the development of cryoelectron microscopic image reconstruction has helped to refine the overall 3D shape of ribosomal particles, in particular as related to specific functional states [292, 293]. Thus a dynamic picture of elongation is emerging.

Awareness of the involvement of rRNA in ribosomal functions has grown during the last decades. Seminal in this context was certainly Carl Woese with his speculations on the origin of the protein synthetic machinery [294]. But it was the characterization of catalytic activities of precursor ribosomal RNA initiated by Thomas Cech [295] that turned the 'protein paradigm' of the ribosome, prevalent in the 1960s and 1970s, back into an 'RNA paradigm'. (Indeed, in the early days of ribosomology, rRNA had been closely associated with ribosomal function. That function — of a template — however, did not survive history.) Indications accumulated that 23S RNA is involved in the peptidyltransferase reaction [296, 297], which until then was thought to be a domain of the ribosomal proteins. Efforts to achieve peptidyltransfer activity with ribosomal RNA alone have so far not been successful [298, 299]. The atomic model of the 50S subunit now appears to suggest that ribosomal RNA may indeed be able to do the job without direct involvement of proteins [300]. On this view, the ribosome would finally turn out to be a veritable ribozyme. However, to adduce direct biochemical evidence concerning the catalytic mechanism of the transfer reaction remains a task for the future.

References

1 H.-J. Rheinberger: *Toward a History of Epistemic Things. Synthesizing Proteins in the Test Tube,* Stanford University Press, Stanford 1997.

2 P. C. Zamecnik, An historical account of protein synthesis, with current overtones – a personalized view, *Symp. Quant. Biol.* **1969,** *34,* 1–16.

3 F. Lipmann: *Wanderings of a Biochemist,* Wiley-Interscience, New York 1971.

4 P. C. Zamecnik, Historical aspects of protein synthesis, *Ann. NY Acad. Sci.* **1979,** *325,* 269–301.

5 P. Siekevitz, P. C. Zamecnik, The ribosome and protein synthesis, *J. Cell. Biol.* **1981,** *91 (3),* 53S–65S.

6 F. Jacob: *The Statue Within,* Basic Books, New York 1988.

7 F. H. C. Crick: *What Mad Pursuit,* Basic Books, New York 1988.

8 M. Nomura: History of Ribosome Research: A Personal Account, in *The Ribosome. Structure, Function and Evolution,* eds W. E. Hill, P.B. Moore, A. Dahlberg et al., ASM Press, Washington, DC 1990, 3–55.

9 A. Spirin: Ribosome Preparation and Cell-free Protein Synthesis, in *The Ribosome. Structure, Function and Evolution,* eds W. E. Hill, P.B. Moore, A. Dahlberg et al., ASM Press, Washington, DC 1990, 56–70.

10 M. Hoagland: *Toward the Habit of Truth,* W. W. Norton and Company, New York 1990.

11 F. H. Portugal, J. S. Cohen: *A Century of DNA,* The MIT Press, Cambridge, MA 1977.

12 H. F. Judson: *The Eighth Day of Creation,* Simon and Schuster, New York 1979.

13 H.-J. Rheinberger: Experiment, difference, and writing. I. Tracing protein synthesis, and II. The laboratory production of transfer RNA, *Stud. Hist. Philos. Sci.* **1992,** *23,* 305–331, 389–422.

14 H.-J. Rheinberger, Experiment and orientation: early systems of *in vitro* protein synthesis, *J. Hist. Biol.* **1993,** *26,* 443–471.

15 H.-J. Rheinberger, From microsomes to ribosomes: 'strategies' of 'representation', *J. Hist. Biol.* **1995,** *28,* 49–89.

16 R. M. Burian, Task definition, and the transition from genetics to molecular genetics: aspects of the work on protein synthesis in the laboratories of J. Monod and P. Zamecnik, *J. Hist. Biol.* **1993,** *26,* 387–407.

17 M. Morange: *Histoire de la biologie moléculaire,* Editions La Découverte, Paris 1994.

18 L. E. Kay: *Who Wrote the Book of Life?* A History of the Genetic Code, Stanford University Press, Stanford 2000.

19 Symp Quant Biol **1969,** *34* : The Mechanism of Protein Synthesis.

20 M. Nomura, A. Tissières, P. Lengyel (eds): *Ribosomes,* Cold Spring Harbor, New York 1974.

21 G. Chambliss, G. R. Craven, J. Davies et al. (eds): *Ribosomes. Structure, Function, and Genetics,* University Park Press, Baltimore 1980.

22 B. Hardesty, G. Kramer (eds): *Structure, Function, and Genetics of Ribosomes,* Springer, New York 1985.

23 W. E. Hill, P. B. Moore, A. Dahlberg et al. (eds): *The Ribosome. Structure, Function, and Evolution,* ASM Press, Washington, DC 1990.

24 K. H. Nierhaus, F. Franceschi, A. R. Subramanian et al. (eds): *The Translational Apparatus. Structure, Function, Regulation, Evolution,* Plenum Press, New York 1993.

25 A. T. Matheson, J. E. Davies, P. P. Dennis et al. (eds): *Frontiers in Translation. An International Conference on the Structure and Function of the Ribosome,* National Research Council Canada, Ottawa 1995.

26 R. A. Garrett, S. R. Douthwaite, A. Liljas et al. (eds): *The Ribosome. Structure, Function, Antibiotics, and Cellular Interactions,* ASM Press, Washington, DC 2000.

27 P. C. Zamecnik: Historical and current aspects of the problem of protein synthesis, *Harv. Lect.* **1960,** *54,* 256–281.

28 F. Hofmeister, Über den Bau des Eiweißmolecüls, *Naturwiss. Rundsch.* **1902,** *17,* 529–533, 545–549.

29 E. Fischer: *Untersuchungen über Aminosäuren, Polypeptide und Proteïne (1899–1906)*. Springer, Berlin 1906.

30 H. Borsook, J. W. Dubnoff, The biological synthesis of hippuric acid *in vitro, J. Biol. Chem.* **1940,** *132,* 307–324.

31 F. Lipmann, Metabolic generation and utilization of phosphate bond energy, *Adv. Enzymol.* **1941,** *1,* 99–162.

32 M. Bergmann, A classification of proteolytic enzymes, *Adv. Enzymol.* **1942,** *2,* 49–68.

33 R. Schoenheimer: *The Dynamic State of Body Constituents,* Harvard University Press, Cambridge, MA 1942.

34 D. Rittenberg, The state of the proteins in animals as revealed by the use of isotopes, *Symp. Quant. Biol.* **1941,** *9,* 283–289.

35 T. Caspersson, Studien über den Eiweißumsatz der Zelle, *Nat. Wiss.* **1941,** *29,* 33–43.

36 J. Brachet, La localisation des acides pentosenucléiques dans les tissus animaux et les oeufs d'Amphibiens en voie de développement, *Arch. Biol.* **1942,** *53,* 207–257.

37 F. Sanger, The arrangement of amino acids in proteins, *Adv. Prot. Chem.* **1952,** *7,* 1–67.

38 G. E. Palade, A small particulate component of the cytoplasm, *J. Biophys. Biochem. Cyt.* **1955,** *1,* 59–68.

39 L. E. Kay: *The Molecular Vision of Life,* Oxford University Press, Oxford 1993.

40 O. T. Avery, C. M. MacLeod, M. McCarty, Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus Type III, *J. Exp. Med.* **1944,** *79,* 137–158.

41 M. Delbrück, A theory of autocatalytic synthesis of polypeptides and its application to the problem of chromosome reproduction, *Symp. Quant. Biol.* **1941,** *9,* 122–126.

42 F. Haurowitz, Biological problems and immunochemistry, *Quart. Rev. Biol.* **1949,** *24,* 93–101.

43 M. Bergmann, J. S. Fruton, The specificity of proteinases, *Adv. Enzymol.* **1941,** *1,* 63–98.

44 J. S. Fruton, R. B. Johnston, M. Fried, Elongation of peptide chains in enzyme-catalyzed transamidation reactions, *J. Biol. Chem.* **1951,** *190,* 39–53.

45 H. M. Kalckar, The nature of energetic coupling in biological synthesis, *Chem. Rev.* **1941,** *28,* 71–178.

46 D. Bartels, The multi-enzyme programme of protein synthesis – its neglect in the history of biochemistry and its current role in biotechnology, *Hist. Philos. Life Sci.* **1983,** *5,* 187–219.

47 A. H. Gordon, Electrophoresis and chromatography of amino acids and proteins, *Ann. NY Acad. Sci.* **1979,** *325,* 95–105.

48 S. de Chadarevian: *Designs for Life. Molecular Biology after World War II,* Cambridge University Press, Cambridge 2002.

49 R. R. Bensley, N. L. Hoerr, Studies on cell structure by the freezing-drying method. VI. The preparation and properties of mitochondria, *Anat. Rec.* **1934,** *60,* 449–455.

50 A. Claude, A fraction from normal chick embryo similar to the tumor producing fraction of chicken tumor I, *Proc. Soc. Exp. Biol. Med.* **1938,** *39,* 398–403.

51 A. Claude, The constitution of protoplasm, *Science* **1943,** *97,* 451–456.

52 I. Sapp: *Beyond the Gene: Cytoplasmic Inheritance and the Struggle for Authority in Genetics,* Oxford University Press, Oxford 1987.

53 J. Brachet, R. Jeener, Recherches sur des particules cytoplasmiques de dimensions macromoléculaires riches en acide pentosenucléique, *Enzymology* **1943,** *11,* 196–212.

54 W. C. Schneider, A. Claude, G. H. Hogeboom, The distribution of cytochrome c and succinoxidase activity in rat liver fractions, *J. Biol. Chem.* **1948,** *172,* 451–458.

55 H. Chantrenne, Hétérogénéité des granules cytoplasmiques du foie de souris, *Biochim. Biophys. A* **1947,** *1,* 437–448.

56 C. P. Barnum, R. A. Huseby, Some quantitative analyses of the particulate fractions from mouse liver cell cytoplasm, *Arch. Biochem. Biophys.* **1948,** *19,* 17–23.

57 C. Claude, Studies on cells: morphology, chemical constitution, and distribution of biochemical function, *Harv. Lect.* **1950,** *43,* 121–164.

58 G. H. Hogeboom, W. C. Schneider, G. E. Palade, Cytochemical studies of mammalian tissues. I. Isolation of intact mitochondria from rat liver; some biochemical properties of mitochondria and submicroscopic particulate material, *J. Biol. Chem.* **1948,** *172,* 619–635.

59 J.-P. Gaudillière: *Inventer la biomédicine: La France, L'Amérique et la production des savoirs du vivant (1945–1965),* Éditions de la Découverte, Paris 2002.

60 H.-J. Rheinberger, Putting Isotopes to Work: Liquid Scintillation Counters 1950–1970, in *Instrumentation Between Science, State, and Industry,* eds B. Joerges T. Shinn, Kluwer, Dordrecht 2001, 143–174.

61 A. Creager, The Industrialization of Radioisotopes by the U.S. Atomic Energy Commission, in *Nobel Symposium 123: Science and Industry in the 20th Century,* ed. K. Grandin, in print.

62 J. B. Melchior, H. Tarver, Studies in protein synthesis *in vitro* . I. On the synthesis of labeled cystine (^{35}S) and its attempted use as a tool in the study of protein synthesis; II. On the uptake of labeled sulfur by the proteins of liver slices incubated with labeled methionine (^{35}S), *Arch. Biochem. Biophys.* **1947,** *12,* 301–308, 309–315.

63 T. Winnick, F. Friedberg, D. Greenberg, Incorporation of ^{14}C-labelled glycine into intestinal tissue and its inhibition by azide, *Arch. Biochem. Biophys.* **1947,** *15,* 160–161.

64 F. Friedberg, T. Winnick, D. M. Greenberg, Incorporation of labelled glycine into the protein of tissue homogenates, *J. Biol. Chem.* **1947,** *171,* 441–442.

65 H. Borsook, C. L. Deasy, A. J. Haagen-Smit et al., The incorporation of labelled lysine into the proteins of guinea pig liver homogenate, *J. Biol. Chem.* **1949,** *179,* 689–704.

66 R. B. Loftfield, Preparation of ^{14}C-labelled hydrogen cyanide, alanine, and glycine, *Nucl.* **1947,** *1,* 54–57.

67 W. F. Loomis, F. Lipmann, Reversible inhibition of the coupling between phosphorylation and oxidation, *J. Biol. Chem.* **1948,** *173,* 807–808.

68 I. D. Frantz, Jr., P. C. Zamecnik, J. W. Reese et al., The effect of dinitrophenol on the incorporation of alanine labelled with radioactive carbon into the proteins of slices of normal and malignant rat liver, *J. Biol. Chem.* **1948,** *174,* 773–774.

69 P. C. Zamecnik, The use of labelled amino acids in the study of the protein metabolism of normal and malignant tissues: a review, *Canc. Res.* **1950,** *10,* 659–667.

70 P. Siekevitz, Uptake of radioactive alanine *in vitro* into the proteins of rat liver fractions, *J. Biol. Chem.* **1952,** *195,* 549–565.

71 N. L. R. Bucher, The formation of radioactive cholesterol and fatty acids from ^{14}C-labelled acetate by rat liver homogenates, *J. Am. Chem. Soc.* **1953,** *75,* 498.

72 E. B. Keller, P. C. Zamecnik, Anaerobic incorporation of ^{14}C-amino acids into protein in cell-free liver preparations, *Fed. Proc.* **1954,** *13,* 239–240.

73 E. B. Keller, P. C. Zamecnik, R. B. Loftfield, The role of microsomes in the incorporation of amino acids into proteins, *J. Histochem. Cytochem.* **1954,** *2,* 378–386.

74 V. Allfrey, A. E. Mirsky, S. Osawa, Protein synthesis in isolated cell nuclei, *J. Gen. Physiol.* **1957,** *40,* 451–490.

75 M. B. Hoagland, An enzymic mechanism for amino acid activation in animal tissues, *Biochim. Biophys. A* **1955**, *16*, 288–289.

76 J. A. DeMoss, G. D. Novelli, An amino acid dependent exchange between inorganic pyrophosphate and ATP in microbial extracts, *Biochim. Biophys. A* **1955**, *18*, 592–593.

77 P. Berg, Participation of adenyl-acetate in the acetate-activating system, *J. Am. Chem. Soc.* **1955**, *77*, 3163–3164.

78 P. Berg, Acyl adenylates: the interaction of adenosine triphosphate and L-methionine, *J. Biol. Chem.* **1956**, *222*, 1025–1034.

79 M. B. Hoagland, E. B. Keller, P. C. Zamecnik, Enzymatic carboxyl activation of amino acids, *J. Biol. Chem.* **1956**, *218*, 345–358.

80 E. F. Gale, J. P. Folkes, Effect of nucleic acids on protein synthesis and amino-acid incorporation in disrupted staphylococcal cells, *Nature* **1954**, *173*, 1223–1227.

81 S. Spiegelman, H. O. Halvorson, R. Ben-Ishai, Free Amino Acids and the Enzyme-forming Mechanism, in *A Symposium on Amino Acid Metabolism (14–17 June 1954)*, eds W. D. McElroy and H. B. Glass, The Johns Hopkins Press, Baltimore 1955, 124–170.

82 E. F. Gale: From Amino Acids to Proteins, in *A Symposium on Amino Acid Metabolism (14–17 June 1954)*, eds W. D. McElroy and H. B. Glass, The Johns Hopkins Press, Baltimore 1955, 171–192.

83 M. Grunberg-Manago, S. Ochoa, Enzymatic synthesis and breakdown of polynucleotides; polynucleotide phosphorylase, *J. Am. Chem. Soc.* **1955**, *77*, 3165–3166.

84 M. B. Hoagland, P. C. Zamecnik, M. L. Stephenson, Intermediate reactions in protein biosynthesis, *Biochim. Biophys. A* **1957**, *24*, 215–216.

85 R. W. Holley, An alanine-dependent, ribonuclease-inhibited conversion of AMP to ATP, and its possible relationship to protein synthesis, *J. Am. Chem. Soc.* **1957**, *79*, 658–662.

86 P. Berg, E. J. Ofengand, An enzymatic mechanism for linking amino acids to RNA, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 78–86.

87 T. Hultin, The incorporation *in vitro* of L-^{14}C-glycine into liver proteins visualized as a two-step reaction, *Exp. Cell. Res.* **1956**, *11*, 222–224.

88 K. Ogata, H. Nohara, The possible role of the ribonucleic acid (RNA) of the pH 5 enzyme in amino acid activation, *Biochim. Biophys. A* **1957**, *25*, 659–660.

89 V. V. Koningsberger, Chr. O. Van der Grinten, J. Th. G. Overbeek, Possible intermediates in the biosynthesis of proteins. I. Evidence for the presence of nucleotide-bound carboxyl-activated peptides in baker's yeast, *Biochim. Biophys. A* **1957**, *26*, 483–490.

90 R. S. Schweet, F. C. Bovard, E. Allen et al., The incorporation of amino acids into ribonucleic acid, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 173–177.

91 S. B. Weiss, G. Acs, F. Lipmann, Amino acid incorporation in pigeon pancreas fractions, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 189–196.

92 K. C. Smith, E. Cordes, R. S. Schweet, Fractionation of transfer ribonucleic acid, *Biochim. Biophys. A* **1959**, *33*, 286–287.

93 F. Crick, On degenerate templates and the adaptor hypothesis. A note for the RNA tie club, Unpublished Manuscript, 1955, reference in [94].

94 R. Olby: *The Path to the Double Helix*, Macmillan, London 1974.

95 F. H. C. Crick: *Discussion Note*, in *The Structure of Nucleic Acids and their Role in Protein Synthesis. Biochemical Society Symposium 14* (18 February 1956), *ed.* E. M. Crook, Cambridge University Press, London 1957, 25–26

96 H. Borsook, C. L. Deasy, A. J. Haagen-Smit et al., The uptake *in vitro* of ^{14}C-labelled glycine, L-leucine, and L-lysine by different components of guinea pig liver homogenate, *J. Biol. Chem.* **1950**, *184*, 529–543.

97 T. Hultin, Incorporation *in vivo* of ^{15}N-labeled glycine into liver fractions of newly hatched chicks, *Exp. Cell. Res.* **1950**, *1*, 376–381.

98 E. B. Keller, Turnover of proteins of cell fractions of adult rat liver *in vivo, Fed. Proc.* **1951,** *10,* 206.

99 N. D. Lee, N. M. MacRae, R. H. Williams, Effect of p-dimethylaminoazobenzene on the incorporation of labelled cystine into protein of the subcellular components of rat liver, *Fed. Proc.* **1951,** 10, 363.

100 E. P. Tyner, C. Heidelberger, G. A. LePage, Intracellular distribution of radioactivity in nucleic acid nucleotides and proteins following simultaneous administration of ^{32}P and glycine-2-^{14}C, *Canc. Res.* **1953,** *13,* 186–203.

101 R. M. S. Smellie, W. M. McIndoe, J. N. Davidson, The incorporation of ^{15}N, ^{35}S and ^{14}C into nucleic acids and proteins of rat liver, *Biochim. Biophys. A* **1953,** *11,* 559–565.

102 V. Allfrey, M. M. Daly, A. E. Mirsky, Synthesis of protein in the pancreas. II. The role of ribonucleoprotein in protein synthesis, *J. Gen. Physiol.* **1953,** *37,* 157–175.

103 P. C. Zamecnik, E. B. Keller, Relation between phosphate energy donors and incorporation of labelled amino acids into proteins, *J. Biol. Chem.* **1954,** *209,* 337–354.

104 J. W. Littlefield, E. B. Keller, J. Gross et al., Studies on cytoplasmic ribonucleo-protein particles from the liver of the rat, *J. Biol. Chem.* **1955,** *217,* 111–123.

105 H. K. Schachman, A. B. Pardee, R. Y. Stanier, Studies on the macromolecular organization of microbial cells, *Arch. Biochem. Biophys.* **1952,** *38,* 245–260.

106 M. L. Petermann, M. G. Hamilton, N. A. Mizen, Eletrophoretic analysis of the macromolecular nucleoprotein particles of mammalian cytoplasm, *Canc. Res.* **1954,** *14,* 360–366.

107 G. E. Palade, P. Siekevitz, Liver microsomes. An integrated morphological and biochemical study, *J. Biophys. Biochem. Cyt.* **1956,** *2,* 171–200.

108 M. L. Petermann, M. G. Hamilton, An ultracentrifugal analysis of the macromolecular particles of normal and leukemic mouse spleen, *Canc. Res.* **1952,** *12,* 373–378.

109 M. L. Petermann, N. A. Mizen, M. G. Hamilton, The macromolecular particles of normal and regenerating rat liver, *Canc. Res.* **1953,** *13,* 372–375.

110 M. L. Petermann, M. G. Hamilton, A stabilizing factor for cytoplasmic nucleoproteins, *J. Biophys. Biochem. Cyt.* **1955,** *1,* 469–472.

111 R. B. Roberts, (ed.): Introduction, in *Microsomal Particles and Protein Synthesis,* Pergamon Press, New York 1958, vii–viii.

112 R. B. Roberts (ed.): Ribosomes. A. General Properties of Ribosomes, in *Studies of Macromolecular Biosynthesis,* Carnegie Inst., Washington, DC 1964, 147–168.

113 F. H. C. Crick, On protein synthesis, *Symp. Soc. Exp. Biol.* **1958,** *12,* 138–163.

114 F.-C. Chao, H. K. Schachman, The isolation and characterization of a macromolecular ribonucleoprotein from yeast, *Arch. Biochem. Biophys.* **1956,** *61,* 220–230.

115 M. L. Petermann, M. G. Hamilton, M. E. Balis et al., Physicochemical and Metabolic Studies on Rat Liver Nucleoprotein, in *Microsomal Particles and Protein Synthesis,* ed. R. B. Roberts, Pergamon Press, New York 1958, 70–75.

116 A. Tissières, J. D. Watson, Ribonucleo-protein particles from *Escherichia coli, Nature* 1958, *182,* 778–780.

117 A. Tissières, J. D. Watson, D. Schlessinger et al., Ribonucleoprotein particles from *Escherichia coli, J. Mol. Biol.* **1959,** *1,* 221–233.

118 H. Friedrich-Freksa, Bei der Chromosomenkonjugation wirksame Kräfte und ihre Bedeutung für die identische Verdopplung von Nucleoproteinen, *Nat. Wiss.* **1940,** *28,* 376–379.

119 H. Chantrenne, Un modèle de synthèse peptidique. Propriétés du benzoyl-phosphate de phényle, *Biochim. Biophys. A* **1948,** *2,* 286–293.

120 A. L. Dounce, Duplicating mechanism for peptide chain and nucleic acid synthesis, *Enzymology* **1952,** *15,* 251–258.
121 V. V. Koningsberger, J. Th. G. Overbeek, On the role of the nucleic acids in the biosynthesis of the peptide bond, *Proc. Kon. Ned. Ak. Wet. Ser. B:. Phys.* **1953,** *56,* 248–254.
122 F. Lipmann: On the Mechanism of some ATP-linked Reactions and Certain Aspects of Protein Synthesis, in *A Symposium on the Mechanism of Enzyme Action,* eds W. D. McElroy and B. Glass, The Johns Hopkins Press, Baltimore 1954, 599–607.
123 G. Gamow, Possible relation between deoxyribonucleic acid and protein structures, *Nature* **1954,** *173,* 318.
124 H. Borsook, The Biosynthesis of Peptides and Proteins, in *Proc. III Int. Congress of Biochemistry,* Brussels, 1955, ed. C. Liébecq, Academic Press, New York 1956, 92–104.
125 R. B. Loftfield, The biosynthesis of protein, *Prog. Biophys. Biophys. Chem.* **1957,** *8,* 348–386.
126 J. D. Watson, F. H. C. Crick, Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid, *Nature* **1953,** *171,* 737–738.
127 L. I. Hecht, P. C. Zamecnik, M. L. Stephenson et al., Nucleoside triphosphates as precursors of ribonucleic acid end groups in a mammalian system, *J. Biol. Chem.* **1958,** *233,* 954–963.
128 M. B. Hoagland, Nucleic acids and proteins, *SA* **1959,** *201,* 55–61.
129 E. F. Gale, J. P. Folkes, Promotion of incorporation of amino-acids by specific di- and tri-nucleotides, *Nature* **1955,** *175,* 592–593.
130 E. F. Gale, J. P. Folkes, The assimilation of amino acids by bacteria. 25. The preparation and activities of a factor involved in the incorporation of amino acids in disrupted staphylococcal cells, *Biochem. J.* **1958,** *69,* 611–619.
131 H.-J. Rheinberger, Ernest. F. Gale and Protein synthesis. Difficulties in analysing a complex system, *TIBS* **1998,** *23,* 362–365.
132 R. W. Holley, J. Apgar, G. A. Everett et al., Structure of a ribonucleic acid, *Science* **1965,** *147,* 1462–1465.
133 H. Chantrenne: *The Biosynthesis of Proteins,* Pergamon Press, New York 1961.
134 J.-P. Gaudillière, J. Monod, S. Spiegelman, L'adaptation enzymatique. Programmes de recherche, cultures locales et traditions disciplinaires, *Hist. Philos. Life Sci.* **1992,** *14,* 23–71.
135 F. Jacob, E. Wollman, Genetic and physical determinations of chromosomal segments in *Escherichia coli, Symp. Soc. Exp. Biol.* **1958,** *12,* 75–92.
136 M. Riley, A. B. Pardee, F. Jacob et al., On the expression of a structural gene, *J. Mol. Biol.* **1960,** *2,* 216–225.
137 L. Astrachan, E. Volkin, Properties of ribonucleic acid turnover in T2-infected *Escherichia coli, Biochim. Biophys. A* **1958,** *29,* 536–544.
138 E. F. Gale, J. Folkes, The assimilation of amino acids by bacteria. 21. The effect of nucleic acids on the development of certain enzymic activities in disrupted staphylococcal cells, *Biochem. J.* **1955,** *59,* 675–684.
139 S. Spiegelman: Protein Synthesis in Protoplasts, in *CIBA Foundation Symp. on Ionizing Radiations and Cell Metabolism,* eds G. E. W. Wolstenholme and C. M. O'Connor, Churchill, London 1956, 185–195.
140 S. Brenner, F. Jacob, M. Meselson, An unstable intermediate carrying information from genes to ribosomes for protein synthesis, *Nature* **1961,** *190,* 576–581.
141 F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **1961,** *3,* 316–356.
142 M. Nomura, B. D. Hall, S. Spiegelman, Characterization of RNA synthesized in *Escherichia coli* after bacteriophage T2 infection, *J. Mol. Biol.* **1960,** *2,* 306–326.

143 F. Gros, H. Hiatt, W. Gilbert et al., Unstable ribonucleic acid revealed by pulse labelling of *E. coli, Nature* **1961,** *190,* 581–585.

144 D. Schachtschabel, W. Zillig, Untersuchungen zur Biosynthese der Proteine. I. Über den Einbau ^{14}C-markierter Aminosäuren ins Protein zellfreier Nucleoproteid-Enzyn-Systeme aus *Escherichia coli* B, *Hoppe-Seylers Zs.* **1959,** *314,* 262–275.

145 M. R. Lamborg, P. C. Zamecnik, Amino acid incorporation into protein by extracts of *E. coli, Biochim. Biophys. A* **1960,** *42,* 206–211.

146 A. Tissières, D. Schlessinger, F. Gros, Amino acid incorporation into proteins by *E. coli* ribosomes, *Proc. Natl. Acad. Sci. USA* **1960,** *46,* 1450–1463.

147 T. Kameyama, G. D. Novelli, The cell-free synthesis of ß-galactosidase by *Escherichia coli, Biochem. Biophys. Res. Commun.* **1960,** *2,* 393–396.

148 D. Nathans, F. Lipmann, Amino acid transfer from aminoacyl-ribonucleic acids to protein on ribosomes of *Escherichia coli, Proc. Natl. Acad. Sci. USA* **1961,** *47,* 497–504.

149 K. Matsubara, I. Watanabe, Studies of amino acid incorporation with purified ribosomes and soluble enzymes from *Escherichia coli, Biochem. Biophys. Res. Commun.* **1961,** *5,* 22–26.

150 J. Ofengand, R. Haselkorn, Viral RNA-dependent incorporation of amino acids into protein by cell-free extracts of *E. coli, Biochem. Biophys. Res. Commun.* **1961/62,** *6,* 469–474.

151 M. Nirenberg, The genetic code, *Nobel Lectures / Physiology or Medicine* 1969, 1–21.

152 R. S. Schweet, H. Lamfrom, E. Allen, The synthesis of hemoglobin in a cell-free system, *Proc. Natl. Acad. USA* **1958,** *44,* 1029–1035.

153 M. W. Nirenberg, J. H. Matthaei, The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides, *Proc. Natl. Acad. Sci. USA* **1961,** *47,* 1588–1602.

154 J. D. Watson: *Molecular Biology of the Gene,* W. A. Benjamin, Menlo Park, CA 1965.

155 F. H. C. Crick, L. Barnett, S. Brenner et al., General nature of the genetic code for proteins, *Nature* **1961,** *192,* 1227–1232.

156 H. -G. Wittmann, Ansätze zur Entschlüsselung des genetischen Codes, *Nat. Wiss.* **1961,** *48,* 729–734.

157 H.-G. Wittmann: Studies on the Nucleic acid–Protein Correlation in Tobacco Mosaic Virus, in *Proc. Vth Int. Congress of Biochemistry,* Moscow, 10-16 August 1961, Vol. I, ed. N.M. Sissakian, Pergamon Press, Oxford **1963,** 240–254.

158 H. Fraenkel-Conrat, A. Tsugita, Biological and Protein–Structural Effects of Chemical Mutagenesis of TMV-RNA, in *Proc. Vth Int. Congress of Biochemistry,* Moscow, 10-16 August 1961, Vol. III, ed. N.M. Sissakian, Pergamon Press, Oxford 1963, 242–244.

159 M. W. Nirenberg, P. Leder, RNA codewords and protein synthesis, *Science* **1964,** *145,* 1399–1407.

160 P. Lengyel, J. F. Speyer, S. Ochoa, Synthetic polynucleotides and the amino acid code, *Proc. Natl. Acad. Sci.USA* **1961,** *47,* 1936–1942.

161 H. G. Khorana, H. Büchi, H. Ghosh et al., Polynucleotide synthesis and the genetic code, *Symp. Quant. Biol.* **1966,** *31,* 39–49.

162 B. F. C. Clark, K. A. Marcker, *N* -formyl-methionyl-s-ribonucleic acid and chain initiation in protein biosynthesis, *Nature* **1966,** *211,* 378–380.

163 J. M. Adams, M. R. Capecchi, *N* -formyl methionyl-sRNA as the initiator of protein synthesis, *Proc. Natl. Acad. Sci. USA* **1966,** *55,* 147–155.

164 S. Brenner, A. O. W. Stretton, S. Kaplan, Genetic code: the 'nonsense' triplets for chain termination and their suppression, *Nature* **1965,** *206,* 994–998.

165 M. G. Weigert, A. Garen, Base composition of nonsense codons in *E. coli, Nature* **1965,** *206,* 992–994.

166 M. S. Bretscher, H. M. Goodman, J. R. Menninger et al., Polypeptide chain termination using synthetic polynucleotides, *J. Mol. Biol.* **1965**, *14,* 634–639.
167 M. C. Ganoza, T. Nakamoto, Studies on the mechanism of polypeptide chain termination in cell-free extracts of *E. coli, Proc. Natl. Acad. Sci. USA* **1966**, *55,* 162–169.
168 J. A. Last, W. M. Stanley, Jr., M. Salas et al., Translation of the genetic message. IV. UAA as a chain termination codon, *Proc. Natl. Acad. Sci.* **1967**, *57,* 1062–1067.
169 J. D. Watson, Involvement of RNA in the synthesis of proteins, *Science* **1963**, *140,* 17–26.
170 F. Lipmann, Messenger ribonucleic acid, *Prog. Nucl. Acid. Res.* **1963**, *1,* 135–161.
171 E. B. Keller, P. C. Zamecnik, Effect of guanosine diphosphate on incorporation of labelled amino acids into proteins, *Fed. Proc.* **1955**, *14,* 234.
172 M. B. Hoagland, Some factors influencing protein synthetic activity in a cell-free mammalian system, *Symp. Quant. Biol.* **1961**, *26,* 153–157.
173 D. Nathans, F. Lipmann, Amino acid transfer from sRNA to microsome. II. Isolation of a heat-labile factor from liver supernatant, *Biochim. Biophys. A* **1960**, *43,* 126–128.
174 J. E. Allende, R. Monro, F. Lipmann, Resolution of the *E. coli* amino acyl sRNA transfer factor into two complementary fractions, *Proc. Natl. Acad. Sci. USA* **1964**, *51,* 1211–1216.
175 Y. Nishizuka, F. Lipmann, Comparison of guanosine triphosphate split and polypeptide synthesis with a purified *E. coli* system, *Proc. Natl. Acad. Sci. USA* **1966**, *55,* 212–219.
176 Y. Kaziro, N. Inoue, Y. Kuriki et al., Purification and properties of factor G, *Symp. Quant. Biol.* **1969**, *34,* 385–393.
177 J. Lucas-Lenard, F. Lipmann, Separation of three microbial amino acid polymerization factors, *Proc. Natl. Acad. Sci. USA* **1966**, *55,* 1562–1566.
178 B. Hardesty, R. Arlinghaus, J. Shaeffer et al., Hemoglobin and polyphenylalanine synthesis with reticulocyte ribosomes, *Symp. Quant. Biol.* **1963**, *28,* 215–222.
179 M. Salas, M. B. Hille, J. A. Last et al., Translation of the genetic message. II. Effect of initiation factors on the binding of formyl-methionyl-tRNA to ribosomes, *Proc. Natl. Acad. Sci. USA* **1967**, *57,* 387–394.
180 M. Revel, J. C. Lelong, G. Brawerman et al., Function of three protein factors and ribosomals subunits in the initiation of protein synthesis in *E. coli. Nature* **1968**, *219,* 1016–1021.
181 M. R. Capecchi, Polypeptide chain termination *in vitro* : isolation of a release factor, *Proc. Natl. Acad. Sci. USA* **1967**, *58,* 1144–1151.
182 A. Von der Decken, T. Hultin, A metabolic isotope transfer from soluble polynucleotides to microsomal nucleoprotein in a cell-free rat liver system, *Exp. Cell. Res.* **1958**, *15,* 254–256.
183 T. Hultin, A. Von der Decken, The transfer of soluble polynucleotides to the ribonucleic acid of rat liver microsomes, *Exp. Cell. Res.* **1959**, *16,* 444–447.
184 L. Bosch, H. Bloemendal, M. Sluyser, Metabolic interrelationships between soluble and microsomal RNA in rat-liver cytoplasm, *Biochim. Biophys. A* **1959**, *34,* 272–274.
185 L. Bosch, H. Bloemendal, M. Sluyser, Studies on cytoplasmic ribonucleic acid from rat liver. I. Fractionation and function of soluble ribonucleic acid, and II. Fractionation and function of microsomal ribonucleic acid, *Biochim. Biophys. A* **1960**, *41,* 444–453, 454–461.
186 M. B. Hoagland, L. T. Comly, Interaction of soluble ribonucleic acid and microsomes, *Proc. Natl. Acad. Sci. USA* **1960**, *46,* 1554–1563.
187 W. Gilbert, Polypeptide synthesis in *E. coli*. II. The polypeptide chain and S-RNA. *J. Mol. Biol.* **1963**, *6,* 389–403.

188 M. Takanami, T. Okamoto, Interactions of ribosomes and synthetic polynucleo-tides, *J. Mol. Biol.* **1963,** *7,* 323–333.

189 H. Kaji, A. Kaji, Specific binding of sRNA with the template-ribosome complex, *Proc. Natl. Acad. Sci. USA* **1964,** *52,* 1541–547.

190 J. R. Warner, A. Rich, The number of soluble RNA molecules on reticulocyte polyribosomes, *Proc. Natl. Acad. Sci. USA* **1964,** *51,* 1134–1141.

191 R. R. Traut, R. E. Monro, The puromycin reaction and its relation to protein synthesis, *J. Mol. Biol.* **1964,** *10,* 63–72.

192 J. D. Watson, The synthesis of proteins upon ribosomes, *Bull. Soc. Chim. Biol.* **1964,** *46,* 1399–1425.

193 G. Brawerman, Role of initiation factors in the translation of messenger RNA, *Symp. Quant. Biol.* **1969,** *34,* 307–312.

194 S. Pestka, Translocation, aminoacyl-oligonucleotides, and antibiotic action, *Symp. Quant. Biol.* **1969,** *34,* 395–410.

195 P. Leder, A. Bernardi, D. Livingston et al., Protein biosynthesis: studies using synthetic and viral mRNAs, *Symp. Quant. Biol.* **1969,** *34,* 411–417.

196 H. Weissbach, N. Brot, D. Miller et al., Interaction of guanosine triphosphate with *E. coli* soluble transfer factors, *Symp. Quant. Biol.* **1969,** *34,* 419–431.

197 T. Caskey, E. Scolnick, R. Tompkins et al., Peptide chain termination, codon, protein factor, and ribosomal requirements, *Symp. Quant. Biol.* **1969,** *34,* 479–491.

198 A.-L. Haenni, F. Chapeville, The behaviour of acetylphenylalanyl soluble ribonucleic acid in polyphenylalanine synthesis, *Biochim. Biophys. A* **1966,** *114,* 135–148.

199 M. B. Yarmolinsky, G. L. de la Haba, Inhibition by puromycin of amino acid incroproation into protein, *Proc. Natl. Acad. Sci. USA* **1959,** *45,* 1721–1729.

200 D. W. Allen, P. C. Zamecnik, The effect of puromycin on rabbit reticulocyte ribosomes, *Biochim. Biophys. A* **1962,** *55,* 865–874.

201 M. Rabinovitz, J. M. Fisher, A dissociative effect of puromycin on the patheway of protein synthesis by Ehrlich ascites tumor cells, *J. Biol. Chem.* **1962,** *237,* 477–481.

202 R. E. Monro, D. Vazquez, Ribosome-catalyzed peptidyl transfer: effects of some inhibitors of protein synthesis, *J. Mol. Biol.* **1967,** *28,* 161–165.

203 C. Coutsogeorgopoulos, Inhibitors of the reaction between puromycin and polylysyl-RNA in the presence of ribosomes, *Biochem. Biophys. Res. Commun.* **1967,** *27,* 46–52.

204 E. Cundliffe, K. McQuillen, Bacterial protein synthesis: the effects of antibiotics, *J. Mol. Biol* **1967,** *30,* 137–146.

205 N. Tanaka, T. Kinoshita, H. Masukawa, Mechanism of protein synthesis inhibition by fusidic acid and related antibiotics, *Biochem. Biophys. Res. Commun.* **1968,** *30,* 278–283.

206 A.-L. Haenni, J. Lucas-Lenard, Stepwise synthesis of a tripeptide, *Proc. Natl. Acad. Sci. USA* **1968,** *61,* 1363–1369.

207 J. Davies, W. Gilbert, L. Gorini, Streptomycin, suppression, and the code, *Proc. Natl. Acad. Sci. USA* **1964,** *51,* 883–890.

208 J. Davies, B. D. Davis, Misreading of ribonucleic acid code words induced by aminoglycoside antibiotics, *J. Biol. Chem.* **1968,** *243,* 3312–3316.

209 R. B. Loftfield, The frequency of errors in protein biosynthesis, *Biochem. J.* **1963,** *89,* 82–92.

210 J. R. Warner, A. Rich, C. E. Hall, Electron microscope studies of ribosomal clusters synthesizing hemoglobin, *Science* **1962,** *138,* 1399–1403.

211 W. Gilbert, Polypeptide synthesis in *E. coli*. I. Ribosomes and the active complex, *J. Mol. Biol.* **1963,** *6,* 374–388.

212 F. O. Wettstein, T. Staehelin, H. Noll. Ribosomal aggregate engaged in protein synthesis: characterization of the ergosome, *Nature* **1963,** *197,* 430–435.

213 A. Gierer, Function of aggregated reticulocyte ribosomes in protein synthesis, *J. Mol. Biol.* **1963,** *6,* 148–157.
214 J. R. Warner, P. M. Knopf, A. Rich, A multiple ribosomal structure in protein synthesis, *Proc. Natl. Acad. Sci. USA* **1963,** *49,* 122–129.
215 D. Schlessinger, G. Mangiarotti, D. Apirion, The formation and stabilization of 30S and 50S ribosome couples in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **1967,** *58,* 1782–1789.
216 R. Kaempfer, Ribosomal subunit exchange during protein synthesis, *Proc. Natl. Acad. Sci. USA* **1968,** *61,* 106–113.
217 M. Nomura, C. V. Lowry, Phage f2 RNA-directed binding of formyl-methionyl-tRNA to ribosomes and the role of 30S ribosomal subunits in initiation of protein synthesis, *Proc. Natl. Acad. Sci. USA* **1967,** *58,* 946–953.
218 P. Traub, M. Nomura, Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins, *Proc. Natl. Acad. Sci. USA* **1968,** *59,* 777–784.
219 M. Nomura, V. A. Erdmann, Reconstitution of 50S ribosomal subunits from dissociated molecular components, *Nature* **1970,** *228,* 744–748.
220 K. H. Nierhaus, F. Dohme, Total reconstitution of functionally active 50S ribosomal subunits from *E. coli, Proc. Natl. Acad. Sci. USA* **1974,** *71,* 4713–4717.
221 K. A. Cammack, H. E. Wade, The sedimentation behavior of ribonuclease-active and -inactive ribosomes from bacteria, *Biochem. J.* **1965,** *96,* 671–680.
222 R. F. Gesteland, Unfolding of *E. coli* ribosomes by removal of magnesium, *J. Mol. Biol.* **1966,** *18,* 356–371.
223 K. S. Kirby, A new method for the isolation of ribonucleic acids from mammalian tissues, *Biochem. J.* **1956,** *64,* 405.
224 A. Gierer, G. Schramm, Infectivity of ribonucleic acid from tobacco mosaic virus, *Nature* **1956,** *177,* 702–703.
225 P. O. P Ts'o, R. Squires, Quantitative isolation of intact RNA from microsomal particles of pea seedlings and rabbit reticulocytes, *Fed. Proc.* **1959,** *18,* Abstract 1351, 341.
226 C. G. Kurland, Molecular characterization of ribonucleic acid from *Escherichia coli* ribosomes. I. Isolation and molecular weights, *J. Mol. Biol.* **1960,** *2,* 83–91.
227 A. S. Spirin, The 'temperature effect' and macromolecular structure of high-polymer ribonucleic acids of various origin, *Biochemistry* **1961,** *26,* 454–463.
228 W. Möller, H. Boedtker: The Integrity of High Molecular Weight Ribonucleic Acid, in *Colloques Internationaux du CNRS No 106, Acides ribonucléiques et polyphosphates. Structure, synthèse et fonctions, Strasbourg, 6-12 Juillet 1961,* Editions du CNRS, Paris 1962, 99–121.
229 R. Rosset, R. Monier, A propos de la présence d'acide ribonucléique de faible poids moléculaire dans les ribosomes d'*Escherichia coli, Biochim. Biophys. A* **1963,** *68,* 653–56.
230 G. G. Brownlee, F. Sanger, B. G. Barrell, The sequence of 5S ribosomal ribonucleic acid, *J. Mol. Biol.* **1968,** *34,* 379–412.
231 F. Sanger, G. G. Brownlee, B. G. Barrell, A two-dimensional fractionation procedure for radioactive nucleotides, *J. Mol. Biol.* **1965,** *13,* 373–398.
232 H. G. Zachau, D. Dütting, H. Feldmann et al., Serine specific transfer ribonucleic acids. XIV. Comparison of nucleotide sequences and secondary structure models, *Symp. Quant. Biol.* **1966,** *31,* 417–424.
233 U. L. Raj Bhandary, A. Stuart, R. D. Faulkner et al., Nucleotide sequence studies on yeast phenylalanine sRNA, *Symp. Quant. Biol.* **1966,** *31,* 425–434.
234 S.-H. Kim, A. Rich, Single crystals of transfer RNA: an X-ray diffraction study, *Science* **1968,** *162,* 1381–1384.

235 F. Cramer, F. Von der Haar, W. Saenger et al., Einkristalle von phenylalanin-spezifischer Transfer-Ribonucleinsäure, *Ang. Chem.* **1968**, *80,* 969-970; *Int. Ed. Engl.* **1968,** *7,* 895–895.

236 J.-P. Waller, J. I. Harris, Studies on the composition of the protein from *E. coli* ribosomes, *Proc. Natl. Acad. Sci. USA* **1961,** *47,* 18–23.

237 D. Elson, A ribonucleic acid particle released from ribosomes by salt, *Biochim. Biophys. A* **1961,** *53,* 232–234.

238 P. Spitnik-Elson, Fractionation of the proteins of *Escherichia coli* ribosomes, *Biochim. Biophys. A* **1962,** *61,* 624–625.

239 W. Möller, J. Widdowson, Fractionation studies of the ribosomal proteins from *Escherichia coli, J. Mol. Biol.* **1967,** *24,* 367–378.

240 E. Kaltschmidt, M. Dzionara, D. Donner et al., Ribosomal proteins. I. Isolation, amino acid composition, molecular weights and peptide mapping of proteins from *E. coli* ribosomes, *Mol. Gen.* **1967,** *100,* 364–373.

241 R. R. Traut, P. B. Moore, H. Delius et al., Ribosomal proteins of *Escherichia coli*. I. Demonstration of different primary structures, *Proc. Natl. Acad. Sci.USA* **1967,** *57,* 1294–1301.

242 S. J. S. Hardy, C. G. Kurland, P. Voynow et al., The ribosomal proteins of *Escherichia coli*. I. Purification of the 30S ribosomal proteins, *Biochemistry* **1969,** *8,* 2897–2905.

243 E. Kaltschmidt, H. Günter Wittmann, Ribosomal proteins. XII. Number of proteins in small and large ribosomal subunits of *Escherichia coli* as determined by two-dimensional gel electrophoresis, *Proc. Natl. Acad. Sci. USA* **1970,** *67,* 1276–1282.

244 P. Carbon, C. Ehresmann, B. Ehresmann et al., The sequence of *Escherichia coli* ribosomal 16S RNA determined by new rapid gel methods, *FEBS Lett.* **1978**, *94,* 152–156.

245 J. Brosius, M. L. Palmer, P. J. Kennedy et al., Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli, Proc. Natl. Acad. Sci.USA* **1978,** *75,* 4801–4805.

246 H. F. Noller: Structure and topography of ribosomal RNA, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 3–22.

247 B. Wittmann-Liebold, Primary structure of *Escherichia coli* ribosomal proteins, *Adv. Prot. Chem.* **1984,** *36,* 56–78.

248 A. M. Maxam, W. Gilbert, A new method for sequencing DNA, *Proc. Natl. Acad. Sci. USA* **1977,** *74,* 560–564.

249 F. Sanger, S. Nicklen A. R. Coulson, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA* **1977,** *74,* 5463–5467.

250 S. A. Yanofsky, S. Spiegelman, The identification of the ribosomal RNA cistron by sequence complementarity. I. Specificity of complex formation, *Proc. Natl. Acad. Sci. USA* **1962,** *48,* 1069–1078.

251 P. Sypherd, S. Osawa: Ribosome Genetics Revealed by Hybrid Bacteria, in *Ribosome,* eds M. Nomura, A. Tissières and P. Lengyel, Cold Spring Harbor, New York 1974, 669–678.

252 S. R. Jaskunas, L. Lindahl, M. Nomura, Specialized transducing phages for ribosomal protein genes of *Escherichia coli, Proc. Natl. Acad. Sci. USA* **1975,** *72,* 6–10.

253 K. Isono: Genetics of Ribosomal Proteins and their Modifying and Processing Enzymes in *Escherichia coli,* in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et. al., University Park Press, Baltimore 1980, 641–669.

254 E. R. Dabbs: Mutant Studies on the Prokaryotic Ribosome, in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, New York 1985 733–748.

255 J. Shine, L. Dalgarno, The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites, *Proc. Natl. Acad. Sci. USA* **1974,** *71,* 1342–1346.

256 J. A. Steitz, K. Jakes, How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli, Proc. Natl. Acad. Sci. USA* **1975,** *72,* 4734–4738.

257 C. G. Kurland: On the Accuracy of Elongation, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 597–614.

258 J. J. Hopfield, T. Yamane: The Fidelity of Protein Synthesis, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 585–596.

259 G. S. Stent, S. Brenner, A genetic locus for the regulation of ribonucleic acid synthesis, *Proc. Natl. Acad. Sci. USA* **1961,** *47,* 4734–4738.

260 N. O. Kjeldgaard, K. Gausing: Regulation of biosynthesis of ribosomes, in *Ribosomes,* eds M. Nomura, A. Tissières and P. Lengyel, Cold Spring Harbor, New York 1974, 369–392.

261 M. Cashel, J. Gallant: Cellular Regulation of Guanosine Tetraphosphate and Guanosine Pentaphosphate, in *Ribosomes,* eds M. Nomura, A. Tissières and P. Lengyel, Cold Spring Harbor, New York 1974, 733–745.

262 D. Richter, Stringent factor from *Escherichia coli* directs the ribosomal binding and release of uncharged tRNA, *Proc. Natl. Acad. Sci. USA* **1976,** *73,* 707–711.

263 K. H. Nierhaus: Analysis of the Assembly and Function of the 50S Subunit from *Escherichia coli* Ribosomes by Reconstitution, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 267–294.

264 G. Stöffler, R. Bald, B. Kastner et al.: Structural Organization of the *Escherichia coli* Ribosome and Localization of Functional Domains, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 171–205.

265 J. A. Lake: Ribosome Structure and Functional Sites, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 207–236.

266 P. B. Moore: Scattering Studies of the Three-dimensional Organization of the *E. coli* Ribosome, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 111–133.

267 R. R. Traut, J. M. Lambert, G. Boileau et al.: Protein Topography of *Escherichia coli* Ribosomal Subunits as Inferred from Protein Crosslinking, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 89–110.

268 B. S. Cooperman: Functional Sites on the *E. coli* Ribosome as defined by Affinity Labelling, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 531–554.

269 H. Bielka: *The Eukaryotic Ribosome,* Springer, New York 1982.

270 I. Wool: Studies of the Structure of Eukaryotic (mammalian) Ribosomes, in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, New York 1985, 391–411.

271 R. J. Planta, J. H. Meyerink: Organization of the Ribosomal RNA Genes in Eukaryotes, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 871–887.

272 M. Stöffler-Meilicke, G. Stöffler: Topography of the Ribosomal Proteins from *Escherichia coli* within the Intact Subunits as Determined by Immunoelectron Microscopy and Protein–protein Cross-linking, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore

and A. Dahlberg, ASM Press, Washington, DC 1990, 123–133.
273 R. Brimacombe, B. Greuer, P. Mitchell et al.: Three-dimensional Structure and Function of *Escherichia coli* 16S and 23S rRNA as Studied by Cross-linking Techniques, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, 93–106.
274 B. Ehresmann, C. Ehresmann, P. Romby et al.: Detailed Structures of rRNAs: New Approaches, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, 148–159.
275 H. F. Noller, D. Moazed, S. Stern et al.: Structure of rRNA and its Functional Interactions in Translation, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, 73–92.
276 P. B. Moore, M. Capel, M. Kjeldgaard et al.: A 19 Protein Map of the 30S Ribosomal Subunit of *Escherichia coli,* in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, New York 1985, 87–100.
277 V. Nowotny, R. P. May, K. H. Nierhaus: Neutron-scattering Analysis of Structural and Functional Aspects of the Ribosome: The Strategy of the Glassy Ribosome, in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, New York 1985, 101–111.
278 M. Boublik, G. T. Oostergetel, J. S. Wall et al.: Structure of Ribosomes and their Components by Advanced Techniques of Electron Microscopy and Computer Image Analysis, in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, New York 1985, 68–86.
279 Ada. Yonath, W. Bennett, S. Weinstein et al.: Crystallography and Image Reconstructions of Ribosomes, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, 134–147.
280 H.-J. Rheinberger, H. Sternbach, K. H. Nierhaus, Three tRNA binding sites on *E. coli* ribosomes, *Proc. Natl. Acad. Sci. USA* **1981,** *78,* 5310–5314.
281 H.-J. Rheinberger, K. H. Nierhaus, Testing an alternative model of the ribosomal peptide elongation cycle, *Proc. Natl. Acad. Sci. USA* **1983,** *80,* 4213–4217.
282 H.-J. Rheinberger, U. Geigenmüller, A. Gnirke et al.: Allosteric Three-site Model for the Ribosomal Elongation Cycle, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, 318–330.
283 W. Wintermeyer, R. Lill, J. M. Robertson, Role of the tRNA Exit Site in Ribosomal Translocation, in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, 348–357.
284 P. B. Moore, Comments on the 1989 International Conference on Ribosomes. in *The Ribosome Structure, Function, and Evolution,* eds W. E. Hill, P. B. Moore and A. Dahlberg, ASM Press, Washington, DC 1990, xxi–xxiii.
285 H.-G. Wittmann, J. Müssig, H. S. Gewitz et al.: Crystallization of *E. coli* ribosomes. *FEBS Lett.* **1982,** *146,* 217–220.
286 S. D. Trakhanov, M. M. Yusupov, S. C. Agalarov et al., Crystallization of 70S ribosomes and 30S ribosomal subunits from *Thermus thermophilus, FEBS Lett.* **1987,** *220,* 319–322.
287 N. Ban, P. Nissen, J. Hansen et al.: The complete atomic structure of the large ribosomal subunit at 2.4 Å resoultion, *Science* **2000,** *289,* 905–920.
288 J. Harms, F. Schluenzen, R. Zarivach et al.: High resolution structure of the large ribosomal subunit from a mesophilic eubacterium, *Cell* **2001,** *107,* 679–688.
289 F. Schluenzen, A. Tocilj, R. Zarivach et al., Structure of functionally activated small ribosomal subunit at 3.3 Å resolution, *Cell* **2000,** *102,* 615–623.

290 B. T. Wimberly, D. E. Brodersen, W. M. Clemons et al.: Structure of the 30S ribosomal subunit, *Nature* **2000,** *407,* 327–339.

291 J. H. Cate, M. M. Yusupov, G. Z. Yusupova et al., X-ray crystal structures of 70S ribosome functional complexes, *Science* **1999,** *285,* 2095–2104.

292 R. K. Agrawal, C. M. T. Spahn, P. Penczek et al., Visualization of tRNA movements on the *Escherichia coli* 70S ribosome during the elongation cycle, *J. Cell. Biol.* **2000,** *150,* 447–459.

293 H. Stark, M. Rodnina, H.-J. Wieden et al., Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation, *Cell* **2000,** *100,* 301–309.

294 C. R. Woese: Just so Stories and Rube Goldberg Machines: Speculations on the Origin of the Protein Synthetic Machinery, in *Structure, Function, and Genetics,* eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press, Baltimore 1980, 357–373.

295 K. Kruger, P. J. Grabowski, A. J. Zaug et al., Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena, *Cell* **1982,** *31,* 147–157.

296 A. Barta, G. Steiner, J. Brosius et al., Identification of a site on 23S ribosomal RNA located at the peptidyltransferase center, *Proc. Natl. Acad. Sci. USA* **1984,** *81,* 3607–3611.

297 H. F. Noller, M. Asire, A. Barta et al.: Studies on the Structure and Function of Ribosomal RNA, in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, New York 1985, 143–163.

298 H. F. Noller, V. Hoffarth, L. Zimniak, Unusual resistance of peptidyl transferase to protein extraction procedures, *Science* **1992,** *256,* 1416–1419.

299 P. Khaitovich, A. S. Mankin, R. Green et al., Characterization of functionally active subribosomal particles from *Thermus aquaticus, Proc. Natl. Acad. Sci. USA* **1999,** *96,* 85–90.

300 P. Nissen, J. Hansen, N. Ban et al., The structural basis of ribosome activity in peptide bond synthesis, *Science* **2000,** *289,* 920–930.

301 A. Claude, Particulate components of cytoplasm, *Symp. Quant. Biol.* **1941,** *9,* 263–271.

302 M. B. Hoagland, Commentary on 'Intermediate reactions in protein biosynthesis', *Biochim. Biophys. A* **1989,** *1000,* 103–105.

303 P. C. Zamecnik, E. B. Keller, J. W. Littlefield et al., Mechanism of incorporation of labelled amino acids into protein, *J. Cell. Comp. Physiol.* **1956,** *47* (Suppl. 1), 81–101.

304 M. B. Hoagland, On an enzymatic reaction between amino acids and nucleic acid and its possible role in protein synthesis, *Rec. Trav. Chim.* **1958,** *77,* 623–633.

305 M. Nomura, W. A. Held: Reconstitution of Ribosomes: Studies of Ribosome Structure, Function and Assembly, in *Ribosomes,* eds M. Nomura, A. Tissières and P. Lengyel, Cold Spring Harbor, New York 1974, 193–223.

2
Structure of the Ribosome

Gregor Blaha and Pavel Ivanov

Protein synthesis is a complex process with the ribosome as a central player. Its task is to decode the mRNA into the corresponding sequence of amino acids with the aid of amino-acylated tRNAs. We will follow the history of ribosome structure starting with the low-resolution images of the ribosome to arrive at the recent high-resolution, atomic structures for each of the ribosomal subunits, where we will focus on the structural elements that shape them. We will not include detailed structural discussion of 5.5 Å resolution structure of *Thermus thermophilus* 70S [1], since, as Ramakrishnan and Moore [2] emphasized, it is not possible to build a structure *de novo* from a 5.5 Å electron density map. Thus, for example, the structure of the 30S ribosomal proteins in the 70S structure was built by placing those from the 30S subunit structure [3] as rigid bodies into the electron density maps of 70S ribosome [4]. Also the 50S ribosomal subunit of the 70S ribosome structure seems to include some misinterpreted electron density regions. The trace of protein L1 C alpha atoms is partly overlapping with the trace of 23S rRNA phosphorus atoms (PDB code 1 giy) [5, 6] and the orientation of the two domains of L1 deviates from the one observed in L1 [7, 8] or in its complex with rRNA [6].

2.1
General Features of the Ribosome and Ribosomal Subunits

With an approximate mass of 2.6–2.8 MDa the bacterial ribosome has a diameter of 200–250 Å and a sedimentation coefficient of 70S. The 70S ribosome consists of two unequal subunits: a large 50S subunit and a small 30S subunit. Each subunit is a ribonucleoprotein particle with one-third of the mass consisting of protein and the other two-thirds of RNA: a single 16S rRNA (~1500 nt) in the 30S subunit and a 5S (~120 nt) and 23S rRNA (~2900 nt) in the large subunit. The protein fraction consists of approximately 20 different proteins in the small and 33 proteins in the large subunit.

The general outline of the 70S and its component subunits was characterized by a variety of electron microscopic techniques during the 1980s. The 30S was described anthropomorphically with a head, connected by a neck to a body with a shoulder and a so-called platform (Fig. 2-1A). A more compact structure for the 50S was defined, consisting of a rounded base with three almost cylindrical extensions. The three protuberances seen from the 50S side are called from left to right, the L1 protuberance,

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

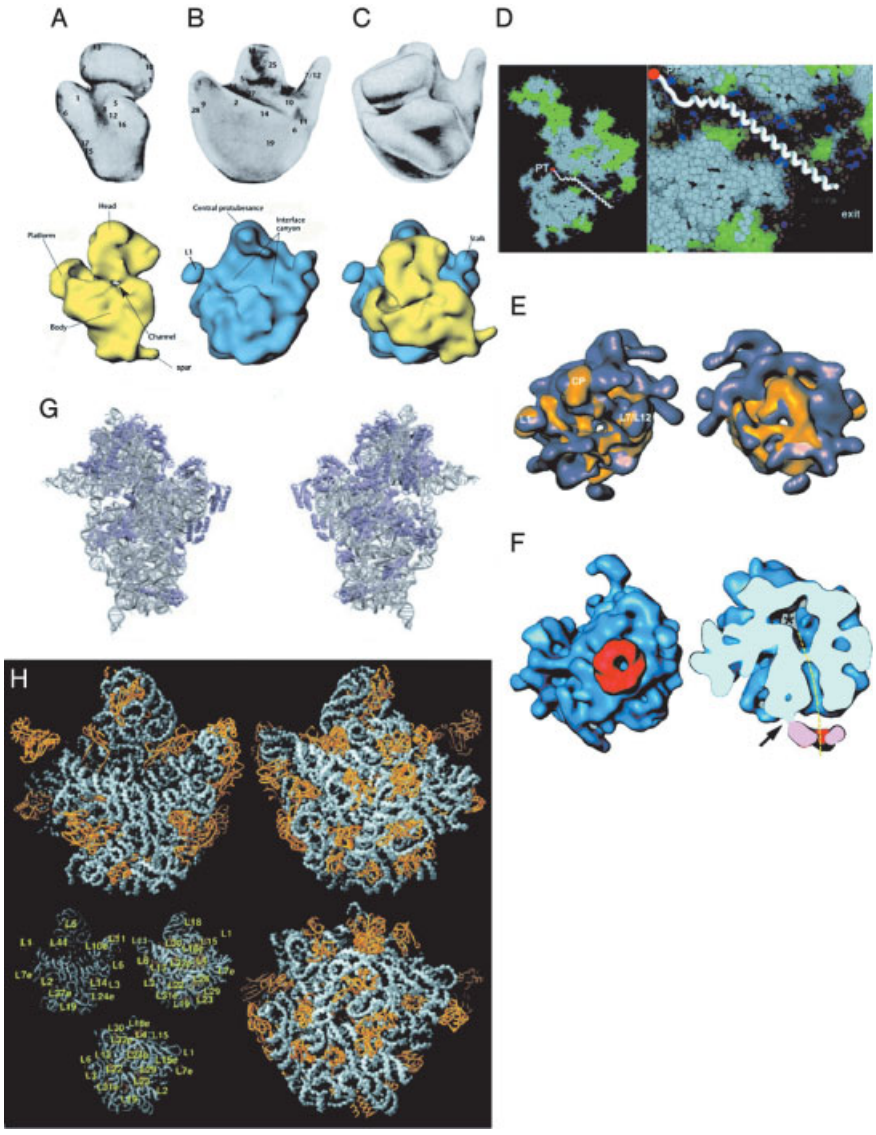
ISBN 3-527-30638-2

the central protuberance, and the L7/L12 stalk (Fig. 2-1B). Both subunits form a 70S ribosome as shown in Fig. 2-1(C). A leap in resolution was achieved by the introduction of single-particle reconstruction of cryo-electron microscopic images [9, 10]. As the resolution improved, the general structural features of the ribosome remained, but more detailed structural features appeared, such as the beak and toe or spur on the 30S (Fig. 2-1A) and a tunnel through the 50S (Fig. 2-1D).

2.2 A Special Feature of the 50S Subunit: The Tunnel

A tunnel transverses the 50S subunit, running from the peptidyl-transferase (PTF) center at the foot of the central protuberance up to the base at the cytoplasmic side of the large subunit with a length of about 100 Å and a width of 10–20 Å (Fig. 2-1D; [11, 10]). The first hint that this tunnel existed was provided by electron microscopy (EM) of two-dimensional (2D) crystals of 80S isolated from chicken embryos [12] and 50S subunits from *Bacillus stearothermophilus* [13]. By that time it had already been shown by immuno-EM that the relative orientation of the exit site of the nascent chain in prokaryotic and eukaryotic ribosomes was identical, located at the lower back (cytoplasmic side) of the large subunit [14]. Moreover, the alignment of cryo-EM structures from rat liver ribosomes with those of *Escherichia coli* proved that not only the central structural features of the ribosome, i.e., L1, L7/L12 and central protuberance, can be superimposed but also the tunnel, suggesting that the tunnel is another universally conserved feature of the ribosome and probably of high functional importance (Fig. 2-1E; [15]). This was subsequently confirmed by the cryo-EM investigations of ribosome–Sec61 complexes from yeast, where the trimeric Sec61 complex, the major component of the endoplasmic pores which conducts the growing nascent peptide chain into the endoplasmic reticulum (ER), was positioned over the exit of the tunnel (Fig. 2-1F; [16]). Recent studies demonstrated that proteins, which are translocated through the ER membrane, indeed exit the ribosome from the ribosomal tunnel [17, 18].

Figure 2-1 Features of the ribosomes. Comparison of ribosomal 30S (A) and 50S subunits (B) and the 70S ribosome (C) from early EM pictures (top row; [104] with the corresponding views obtained by recent cryo-EM reconstructions (bottom row; according to Frank and Agarwal [105]. (D) A cut through the 50S subunit which bisects the central protuberance and the tunnel along the entire length. All ribosome atoms are shown in spacefilling representation, with all RNA atoms that do not contact solvent shown in white and all protein atoms that do not contact solvent shown in green. Surface atoms of both protein and RNA are color-coded: yellow, carbon; red, oxygen; and blue, nitrogen. A possible trajectory for a polypeptide passing through the tunnel is shown as a white ribbon. PT, peptidyl-transferase site [20]. (E) Both the mammalian and the bacterial large subunits have been superimposed. Wherever the bacterial contour lies outside the mammalian one, the resulting surface is gold; wherever the mammalian contour lies outside the bacterial one, the resulting surface is blue. The views are from the 30S subunit side (left) and from the tunnel exit (right). From Ref. [15] with permission. (F) Three-dimensional reconstruction of the ribosome–Sec61 complex with Sec61 oligomer shown in red. Right panel: a cut along a plane that cross sections the pore of the Sec61 oligomer and the ribosome tunnel is shown.



The arrow indicates the stem connecting the ribosome with the Sec61 oligomer; the ribosomal tunnel and its alignment with the Sec61 pore is indicated by a broken yellow line. From Ref. [16], with permission. (G) Distribution of r-proteins in 30S: left seen form the 50S and right the cytosolic side of the 30S. Grey, RNA; blue, proteins. From Ref. [3] with permission. (H) Proteins that appear on the surface of the large ribosomal subunit. The RNA of the subunit is shown in gray and protein backbones are shown in gold. Left panel: the crown view facing the small subunit ; right panel, back side of the subunit (solvent side) in the 180° rotated crown view orientation; bottom right, view from the bottom of the 50S subunit; bottom left, key for the ribosomal proteins. From Ref. [11] with permission.

The 50S crystal structure of *Deinococcus radiodurans* [19] and *Haloarcula marismortui* [11] has now revealed the tunnel at high resolution and confirmed the previously determined dimensions and, furthermore, show that the wall of the tunnel is composed of nucleotides from domains I through V of the 23S rRNA, as well as of the non-globular parts of ribosomal proteins L4 and L22. The narrowest part of the tunnel is formed mainly by ribosomal proteins L4 and L22, where the, β-hairpin of L22 intercalates between rRNA segments of the 23S rRNA. The tunnel surface should minimize unfavorable interactions with growing nascent chains and accordingly no large hydrophobic patches have been observed lining the wall; instead the lining of the tunnel wall is made up of large hydrophilic non-charged groups, thereby facilitating the passage of all kinds of peptide sequences [20]. There seems to exist a system of tunnels, the main tunnel of which represents the shortest route from the entrance to the exterior surface of the ribosome that binds to the bacterial membrane (exit 1), whereas three additional tunnels might communicate with the solvent.

One of these additional routes, which branches off the main tunnel in the segment formed by domains I and III of 23S RNA close to the main exit, was originally discovered at the 25 Å resolution [21]; the branching point is 70 Å away from the tunnel entrance. *H. marismortui* proteins L15 and L29 are closest to the end of this second branch (exit 2; [11]), whose length is 50 Å. Two other branches that start approximately in the same region, which is in fact the widest (20 Å) segment of the main tunnel, partially follow the extensions of proteins L4 (exit 3) and L22 (exit 4). The lengths of these branches from the common branching point to the corresponding exits are approximately 100 and 40 Å, respectively [22, 23]. A similar network of tunnels was also found in *E. coli* ribosomes [22, 23], within the 50S crystal structure of *D. radiodurans* (eubacteria, [19]) and *H. marismortui* (archea, [11]), and in cryo-EM reconstruction of yeast 80S ribosome (eukaryotes, [22, 24]) suggesting that the nascent peptide could in theory emerge into the cytoplasm via one these sub-braches, and thus alternative routes for the nascent peptide chain might be a universal feature of ribosomes.

If the main tunnel is the shortest route and most simple way out of the ribosome, why then does the ribosome need these additional routes branching off the main pathway near the main exit? One possibility is that these openings could maintain the necessary chemical equilibrium in the tunnel system, providing access for water and ion molecules. Another is that they could be used for a more complex regulation of peptide translocation and modification, i.e., the idea being that different polypeptides would utilize different pathways depending on (i) their subcellular destination, (ii) co-factors, e.g., chaperones, which they need for folding or (iii) whether they require post-translational modifications such as methylations, acetylations, or phosphorolations. The implication of this latter view is that the ribosome tunnel would need to play some active role in directing the native peptides in the correct direction. The corollary of this is that there should in fact be specific interactions between the polypeptide and ribosomal components.

This is exactly what a number of recent experiments are clearly indicating: specific peptide sequences have been shown to interact with the interior of the tunnel and

thereby affect protein synthesis on the ribosome. Such sequence-specific interactions between the exit tunnel and nascent peptides suggest that the ribosome, similar to the RNA polymerases [25], can recognize *cis*-acting signals in the synthesized heteropolymeric chain and use them in possibly important intracellular control systems. Nascent peptides in prokaryotes and eukaryotes contain special sequence motifs, and when these effector sequences are situated in the exit tunnel of translating ribosomes, they can significantly affect both protein elongation and peptide termination (Table 2-1) [26, 27]. In all known cases, the peptides with effector motifs act only in *cis* and thus only affect the ribosome on which they are synthesized. Secondly, most effector sequences give rise to ribosomal complexes that are stalled either in the elongation or termination phase of protein synthesis. Thirdly, several of the active peptides have a co-effector and the interplay between an effector motif and a co-effector is key to several intracellular control systems. The co-effector can, for example, be an antibiotic (leading to expression of resistance genes), an amino acid (leading to induction of an amino acid degradation operon), or a polyamine (leading to repression of polyamine synthesis; summarized in Ref. [23]).

An interesting and surprising involvement for the tunnel is in the regulation of a tryptophan catabolite pathway is seen in the expression of the tryptophanase (tna) operon in *E. coli* [28]. Tna is a catabolic enzyme that degrades tryptophan to indole, pyruvate, and ammonia, allowing tryptophan to serve as a carbon or nitrogen source [29]. The tna operon begins with a tnaC gene coding for a 24-amino-acid-long oligopeptide (leader peptide) followed by the structural genes tnaA and tnaB coding for a the tryptophanase and tryptophan permease, respectively [30]. The regulation is orchestrated by the following elements: The 12th and the 24th (last) position of the leader peptide are tryptophan and proline, respectively, followed by the RF2-dependent UGA stop codon. Another set of essential elements include transcriptional pause sites located immediately after the tnaC gene, where Rho-factor-mediated transcription termination can occur and thus prevent synthesis of tryptophanase. First analyses have demonstrated that high tryptophan concentration prevents Rho action [31] and interferes with RF2-dependent hydrolysis [32].

The following mechanism has emerged: The ribosome pursuing the transcriptase during translation of the leader peptide will carry an aa_{23}-Pro-$tRNA^{Pro}$ at its P-site with the 12th Trp residue in the tunnel just where L4/L22 form the kink of the tunnel (see Fig. 2-1D). This constellation with the prolyl residue at the P-site next to the UGA and the tryptophanyl residue at the tunnel kink provokes a stalling of the ribosome and a retardation of the RF2-dependent hydrolysis. At this moment, obviously, the amino acid Trp (not Trp-tRNA!) at high Trp concentration binds to ribosome (possibly to the A-site region of the PTC) thus preventing the hydrolysis of the pp-tRNA and blocking the ribosome on a transcript site of the mRNA required for Rho factor binding. The result is a continuation of transcription into the structural genes tnaA and tnaB fostering the degradation of tryptophan [28].

Another example of a peptide with an effector sequence is the *secM* (secretion monitor) gene of *E. coli* encoding a unique secretory protein that monitors cellular activity for protein export and accordingly regulates translation of the downstream

Table 2-1 Nascent peptides causing ribosome stalling (data taken from Ref. [23])

Organism	Gene	Active sequence	Reference	Co-effector	Notes
Eubacteria	cat, cmlA	**MVKTD** **MSTSKNAD**	26	Chloramphenicol	Present in the middle of a small uORF; low concentration of antibiotics cause stalling of the ribosome and therefore rearrangment of mRNA secondary structure releasing the otherwise trapped translation-initiation region of a downstream ORF coding for an enzyme responsible for antibiotic resistance
Eubacteria	ErmC	**SIFVI**	95	Erythromycin	
Eubacteria	TnaC	**KWFNID**	32, 28	Tryptophan	Trp-induced ribosome stalling at the end of a small uORF is essential for attenuation of transcription; ribosome is inhibited at the stage of termination
Eubacteria	secM	**FXXXXWIXXXXGIRAGP**	35	Membrane translocation (secA)	Sensor of protein translocation across the membranes; inhibits translating ribosome; translocation of the nascent chain relieves inhibition
Yeast	CPA1	**NSQYTCQDYISDHIWKTS**	96	Arginine	Small uORFs in the eucaryotic mRNAs cause stalling of the ribosome. This inhibits scanning of the initiation complexes to the main ORF downstream. The Arg-responsible nascent chain CAP1 and arg can act on either terminating or elongating ribosomes. AdoMetDC and CMV UL4 can inhibit termination
Fungi	arg	**PSXFTSQDYXSDHLWXAX**	97	Arginine	
Mammals	AdoMet DC β_2-AdRec Rar- β_2	**MAGDIS** **MKLPGVRPRPAAPRRRCTR** **MIRGWEKDQQPTCQKRGRV**	98–101	Spermidine, spermine	
Mammalian virus	CMV UL4	**MQPLVLSAKKLSSLLTCKY IPP**	102		

secA gene [33]. SecM is exported to the periplasm, where it is rapidly degraded by a tail-specific protease [34]. The regulation works again via a translational arrest, and, interestingly, with a similar sequence signature as in the previous example: the motif critical for the arrest is FXXXXWIXXXXGIRAGP with a polypeptide Pro-$tRNA^{Pro}$ at the P-site and a Trp residue located 12 aa residues (towards the N-terminal; [35]). This arrest will be only relieved if the ribosomal complex can contact SRP · SecA, thus triggering the export of nascent SecM. Only the stalled ribosome allows the display of the ribosomal binding site for the translation of SecA, whereas relief of the blocked SecM translation allows folding of the secM-secA mRNA, which hides the translational-initiation site of SecA-mRNA region. It follows that a lack of SecA induces synthesis of SecA.

This arrest can be suppressed by each of three amino acid mutations in L22, namely Gly91 to Ser, Ala93 to Thr, and Ala93 to Val. The two residues, 91 and 93, are located on the segment of L22 that protrudes into the exit tunnel at the constricted region. We also see in this example that the regions of L4 and L22 at the tunnel kink (and probably influencing the tunnel shape at this point) might sense the nascent chain in an unknown way, thus influencing essential ribosomal functions occurring not in the adjacent neighborhood such as peptide-bond formation and tRNA translocation.

Modeling of this polypeptide in the tunnel revealed that the conserved Trp-Ile (WI) of the motif would be placed within the most constricted region of the tunnel in close proximity to the tip of the β-hairpin of L22 [36]. Furthermore, the binding of the macrolide troleandomycin with the *D. radiodurans* 50S subunit coincided with this hairpin such that the hairpin was pushed across the tunnel lumen to contact the wall on the other side of the tunnel [36]. This led Yonath and co-workers to suggest that this swung conformation is related to the gating mechanism that is involved in secM-induced translational stalling, i.e., the interaction of the Trp-Ile (both relatively bulky residues) may also induce similar structural rearrangements in L22 such that the tunnel is temporarily closed and therefore translation blocked [36]. Further speculations are that the known ribosomal arrest suppression mutations of L22 (G91S, A93T, and A93V) may stabilize the swung conformation [36]. However, confirmation of this mechanism will require structures of nascent chain-ribosome complexes.

2.3
Features of the Ribosomal Subunits at Atomic Resolution

The first attempts to crystallize the ribosome were undertaken in the 1980s with the first 3D crystals obtained from *B. stearothermophilus* [37]. Owing to continual improvements in the quality of the crystals and in sampling techniques of diffraction patterns (discussed in Ref. [38]), the structure of both subunits at atomic resolution was revealed in the past few years [11, 19, 39, 3].

Although it may not be obvious at first glance, but both large and small subunits have structural features in common at a global as well as at an atomic level. The overall shapes of the atomic resolution structures are in good agreement with those derived from cryo-EM (see Refs. [40, 41] for comparison). The interfaces of both subunits, with the exception of S12 in the small subunit, are essentially protein-free [2],

which had been predicted by neutron scattering [42]. This was later on confirmed by cryo-EM not only for ribosomes from bacteria but also for yeast ribosomes ([43]; see also Fig. 2-1G). In the 50S subunit, the proteins are evenly scattered over the cytosolic surface (Fig. 2-1H), whereas in the 30S they are concentrated mainly in the head and shoulder and platform regions of the body (Fig. 2-1G).

The ribosomal proteins are often bound to junctions between helices, thereby often connecting separate domains, for example S17, which simultaneously contacts helix 7 (h7), h11 of 5′ domain and h21 in the central domain, and L18 which links the helical regions 1 and 2/3 of the 5S rRNA with H87 (note the helices of the rRNA of small ribosomal subunit are denoted with a lower-case letter "h", those of the large subunit with upper-case letter "H" and are counted in the phylogenetically derived secondary structure from the 5′ to 3′ as they occur; see also Sect. 1.4 and Ref. [44]) of 23S rRNA (for helix numbering see below Figs. 2-4A and C; [39, 10, 3]). Many proteins in both subunits have globular domains, generally found on the surface of the subunits, with long extensions that reach far into the RNA core, where they make intimate contacts with the rRNA. These extensions lack tertiary structure and in many regions, even secondary structure, as exemplified by the proteins that neighbor the PTF center (PTC) of the large subunit (Fig. 2-2, [20]). These long extensions from a globular domain represent a new and typical feature of ribosomal proteins and explain the numerous painful and unsuccessful attempts to crystallize many of the ribosomal proteins. A classic example being L2, where only fragments could be crystallized [45], and L4, where crystals were obtained only from a thermophilic bacterium and a halophilic archeon [46]. Almost half of the 30S proteins belong to the category of "globular domain plus long extensions" (such as S2, S6, S9, S11, S12, S14, S16, S17, and protein Thx specifically found in *T. thermophilus*) as well as many of the large subunit proteins (in *H. marismortui*: L2, L3, L4, L15, L18, L19, L22, L24, L37e, L44e, L15e, L37ae and in *D. radiodurans*: L3, L4, L5, L13, L24, L31 (counterpart L15e), L35 (no counterpart in *H. marismortui*)).

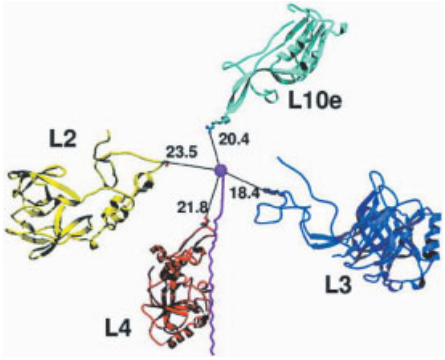


Figure 2-2 A view of the active PTF site with the RNA removed. The proteins with closest extensions to the entrance of the tunnel (pink) through the 50S subunit are shown as ribbons with their closest side chains in all-atom representation. From Ref. [20] with permission.

In the 30S at least, the feature of protein extensions is found only with late-assembly proteins. The large extensions that are rich in basic amino acids (to mask the negative charges of the phosphates in the rRNA backbone) obviously fix the fold of the rRNAs at a late-assembly stage and stabilize the 3D fold of both proteins and rRNAs. In the 50S subunit, the proteins L3, L4, L22, and L25 belong to the proteins that determine a fold of the 23S rRNA essential for the early assembly [47]; in this case they could act initially to connect two distant domains and facilitate their coming together [4].

Even though the general microscopic features, such as the RNA fold or the protein distribution of the different 50S structures, are similar there are some significant differences [19]. For example, the entire L1 stalk in the unbound *D. radiodurans* 50S is tilted by about 30° away from its position in the *T. thermophilus* 70S ribosome yielding a maximum distance of over 30 Å of the outermost points (Fig. 2-3A) [19, 48].

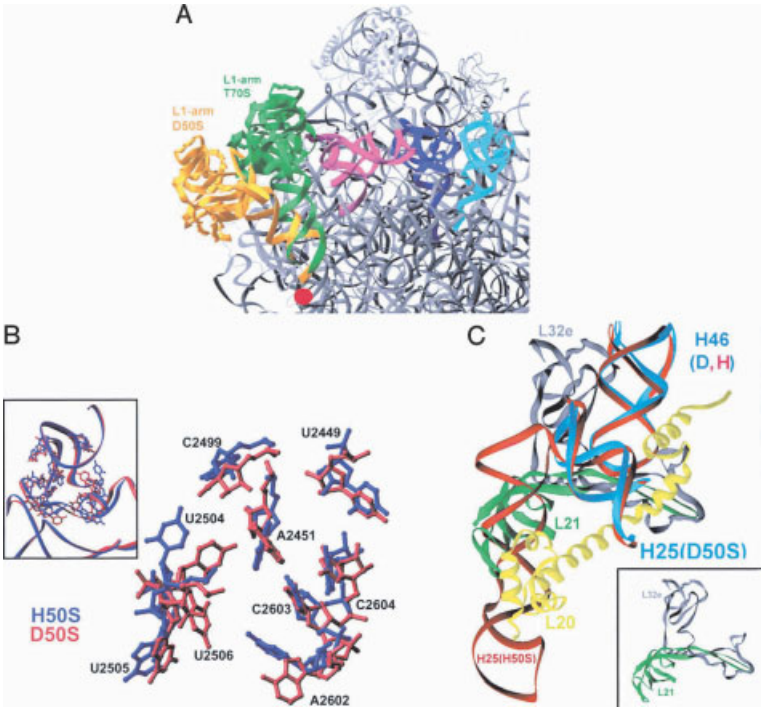


Figure 2-3 Specific features and differences of *D. radiodurans* 50S compared with 50S from *T. thermophilus* 70S and *H. marsimortui* 50S. (A) Movement of L1 stalk. The *D. radiodurans* 50S structure is displayed as gray ribbons with the L1-arm highlighted in gold. The over-laid L1-stalk of T70S is displayed in green. (B) Comparison of the nucleotides within the peptidyl-transferase center of *D. radiodurans* 50S with the corresponding ones from *H. marismortui* 50S. Inset: the overall fold of the peptidyl-transferase center to emphasize the back-bone similarity within *D. radiodurans* and *H. marismortui* 50S. (C) Overlay of H25 in *D. radiodurans* and *H. marismortui* (for details see text). Inset displays proteins L21 and L23e, which are related by an approximate 2-fold. From Ref. [19] with permission.

Also the L7/L12 stalk and the GTPase-associated center consisting of H42–H44 and proteins L7/L12 and L10 are shifted (by 3–4 Å) between *D. radiodurans* 50S and *T. thermophilus* 70S ribosome [1]. The helices H42–H44 show a rotation of about 12° in the case of the *H. marismortui* 50S (PDB 1JJ2) from its position in *D. radiodurans* 50S [19, 48]. These observed flexibilities of the stalks are in line with cryo-EM studies of ribosome complexes in different functional states [49, 50, 24, 51]

H. marismortui 50S crystals derived according to Ban et al. [11] were used in kinetic and crystallographic studies of the PTF reaction. The appearance of peptide product, bound to the PTF ring in the electron density maps of 50S crystal soaked with substrate and the strict dependence of the peptide formation on the presence of 50S crystals clearly demonstrates the catalytic activity of 50S in the disputed crystal form (Refs. [52, 53]; see also Chap. 8.4 on the peptidyl-transferase reaction).

L27, which is located at the base of the central protuberance of *D. radiodurans* and has no homolog in *H. marsimortui*, is proposed to be involved in the proper placement of the 3′-end of the A-site tRNA at the PTC during the PTF reaction [54]. In *H. marismortui,* the non-homologous L21e replaces L27; however, the tail of L21e folds back towards the interior of the subunit and therefore cannot make contact with the P-site tRNA [19].

Other *D. radiodurans*-specific large ribosomal proteins are the L25 analog CTC, which fills the gap between the central protuberance and L7/L12 stalk; the extended α-helical protein L20, which is replaced by 47 n extension in H25 of domain II in *H. marismortui* (see Fig. 2-3C); and the two Zn-finger proteins L32 and L36 [19].

2.4 The Domain Structure of the Ribosomal Subunits

The shear complexity of protein synthesis forces any participating component to maintain their structure and function through evolution. This principle justifies the assumption that not only all tRNAs, but also all 16S (and 16S-like) and 23S (and 23S-like) rRNAs have the same general secondary and tertiary structures [55]. Therefore, the secondary structures of 16S, 23S, and 5S rRNA could be derived by analyzing the pattern of variation within aligned rRNA sequences from different species (see Figs. 2-4A and C; [56]). The resulting secondary-structure diagrams consist of a complex arrangement of A-form helices and non-helical regions (loops or bulges) [55]

In the 16S rRNA the different domains branch from a central pseudo-knot and, beginning from the 5′-end, are termed the 5′, central, 3′-major and 3′-minor domains (see Fig. 2-4A). In striking contrast to the 50S subunit (see following), the domains are not interwoven in the tertiary fold and can be assigned easily to the structural landmarks of the 30S subunit (Fig. 2-4B). The 5′-domain forms the 30S body, starting from the neck of 30S subunit it goes down to the toe and finally turns back to form the shoulder. The central domain constitutes the platform, the 3′-major domain the head. The 3′-minor domain consists of h44 and h45; h44 runs down the 30S along the inter-subunit surface and returns back to the neck, followed by h45 and a single-stranded 3′-end containing the anti-Shine–Dalgarno sequence.

In the 23S rRNA secondary structure the 5′- and 3′-terminal ends are brought together to form a helix (H1 in Fig. 2-4D). Radiating from the loop of this helix are 11 stem-loop structures of differing degrees of complexity. These stem-loop structures

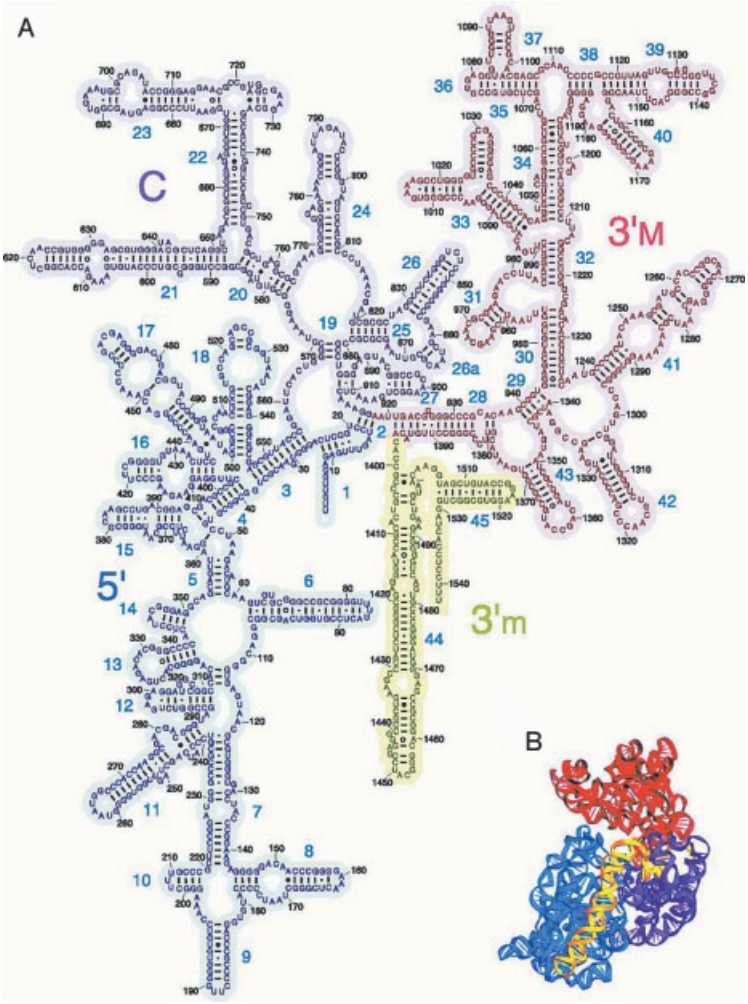


Figure 2-4 Secondary structures of 16S, 23S, and 5S rRNAs. (A) Secondary structure of *T. thermophilus* 16S rRNA, with its 5′, central, 3′-major, and 3′-minor domains shaded in blue, magenta, red, and yellow, respectively. (B) Three-dimensional fold of 16S rRNA in 70S ribosomes, with its domains colored as in (A). (C) Secondary structures of *T. thermophilus* 23S and 5S rRNAs, indicating domains I (blue), II (cyan), III (green), IV (yellow), V (red), and VI (magenta) of 23S rRNA. The rRNAs are numbered according to *E. coli*. (D) Three-dimensional folds of 23S and 5S rRNAs, with their domains colored as in (C).

are bundled into six different domains and, in analogous fashion to the helices, are numbered from the 5′- to 3′-end. The last change in the assignment of the secondary structure to the various domains was contributed by crystallography. Helix 25, which was originally considered as being part of domain I, was reassigned to domain II, because it exhibits stronger interactions with this domain than with the elements of domain I [11]. The six domains of 23S and 5S rRNA all have compact shapes, which are intertwined (Fig. 2-4D). The domains form structural units, as the vast majority of

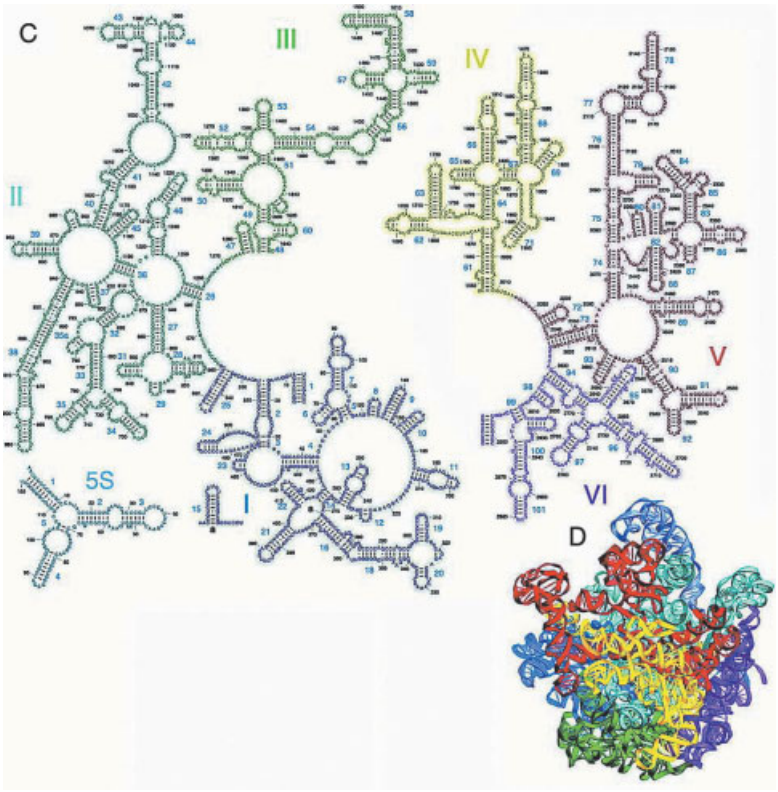


Figure 2-4 From Ref. [1] with permission. (E, F) Comparison of the current comparative structure models for the 16S and 23S rRNAs with the corresponding ribosomal subunit crystal structures. (E) 16S rRNA versus the *T. thermophilus* structure (GenBank accession no. M26923; PDB code 1FJF). (F) 23S rRNA, 5′-half and 3′-half versus the *H. marismortui* structure (GenBank accession no. AF034620; PDB code 1JJ2). Nucleotides are replaced with colored dots that show the sources of the interactions: red, present in both the covariation-based structure model and the crystal structure; green, present in the comparative structure and not present in the crystal structure; blue, not present in the comparative structure and present in the1 crystal structure; purple, positions that are unresolved in the crystal structure.

interactions involving two or more hydrogen bonds occur within domains, rather than between them. This is also the reason for that the interchange of the domain V from *E. coli* against *Staphylococcus aureus* in the 23S rRNA of *E. coli* was possible, even though it introduces 132 changes into the *E. coli* rRNA sequence, only one additional mutation of U1782C for was necessary for viability [57].

Nearly all of the secondary-structure base pairings and a few of the tertiary base pairs observed in the crystal structure had already been predicted by comparative structure models. Specifically, more than 1250 base pairs predicted were indeed present in the 16S and 23S rRNA crystal structures. The ~35 predicted base pairs, which were not found in the crystal structures, could simply not occur at all or possibly only at certain stages of protein synthesis, for example, in the 30S subunit, the A•A base pair between positions 1408 and 1493 is broken upon binding of tRNA and mRNA (cf. 1FJF and 1IBM) [58, 59]. The crystal structures of small and large subunits enriched the secondary-structure diagrams by ~170 base pairs in 16S and ~415 in 23S rRNA, i.e., these were not predicted by comparative methods. Essentially, all the "mis-assigned" base pairs have no significant amount of variation (Ref. [55], see Figs. 2-4E and F).

The fact that the domain secondary structures form well-defined structural domains of quaternary structure in the small subunit, but not in the large subunit, may result from two reasons that need not be mutually exclusive: (i) the small subunit might require larger flexibility for ribosomal functions [2], and the difference in the organization of the secondary structures might indicate that (ii) the 50S subunit is older in evolutionary terms, because more time would be required to evolve such a complicated interwoven structure similar to that of the 50S subunit. As Schimmel and Henderson [60] noted, three elements of protein synthesis, viz. tRNA, synthetases and the ribosome, separate into two domains of different evolutionary ages that have probably co-evolved. The "old" domain of the tRNA is the aminoacyl stem (the short arm of the L-shaped tRNAs or mini-helix), which corresponds to the catalytic domain of synthetases in charging the tRNAs and to the large ribosomal subunit involved in peptide-bond formation. The "young" domains are represented by the long arm of tRNAs bearing the anticodon loop, the recognition domain of the synthetases, and the small ribosomal subunit that interacts with the anticodon loop and some of the stem base-pairs [61, 1].

2.5 Interactions of RNA with RNA or Struts and Bolts in the Three-dimensional Fold of rRNA: Coaxial Stacking and A-minor Motifs

Upon the publication of the structures of the two ribosomal subunits the amount of RNA structure known at atomic resolution increased about 8-fold [62, 2]. However, most of the structure motifs had been seen before, suggesting that the possible number of RNA structure motifs is limited [2]. In this and the following section (2.6), we will restrict our analysis to a specific subsection of structural elements, since tetraloops, tetraloop receptors, adenosine platforms, U-turns, E-loops, sarcin/

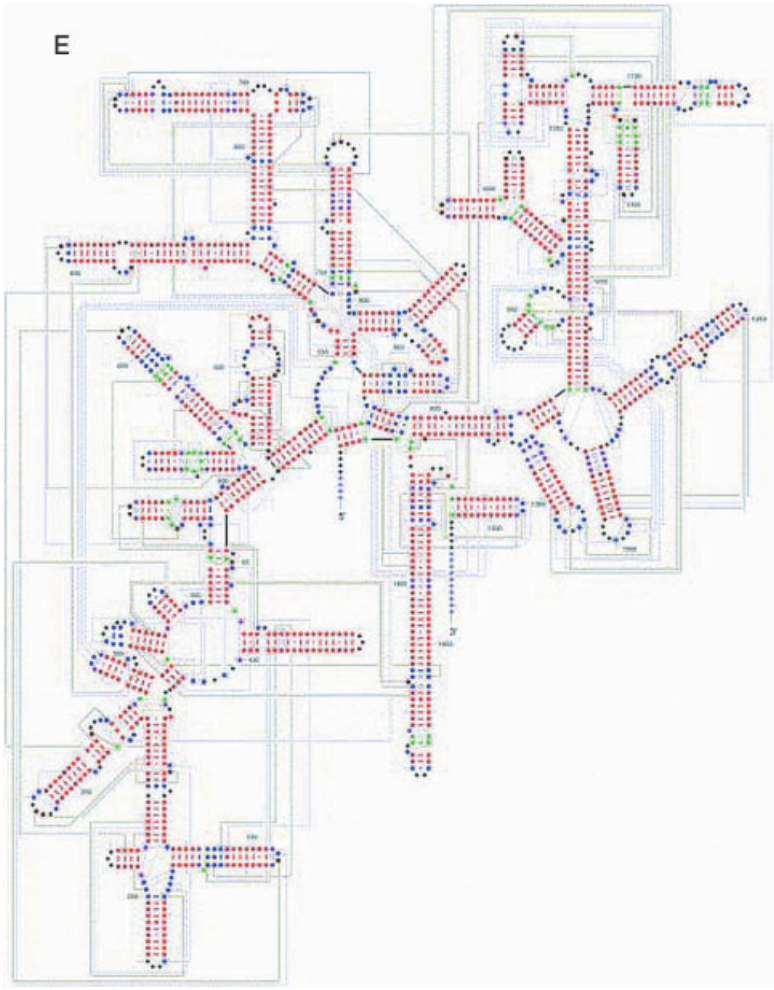


Figure 2-4 (contd.)

ricin motif, and cross-strand purine stacks have been extensively discussed elsewhere [63–66]. In the crystal structure, many of the single-stranded loop regions in the secondary structure turned out to be slightly irregular double-stranded extensions of neighboring regular helices. Thus, most rRNA may be described as helical or approximately helical [3]. These helical elements are organized via vertical co-axial stacking of helices, by A minor interactions and ribose zippers.

2.5.1 Coaxial Stacking

Coaxial stacking is the end-to-end stacking of separate helical RNA parts to form long quasi-continuous helical structures (see as example h16/h17 in 30S or H34/H35 in

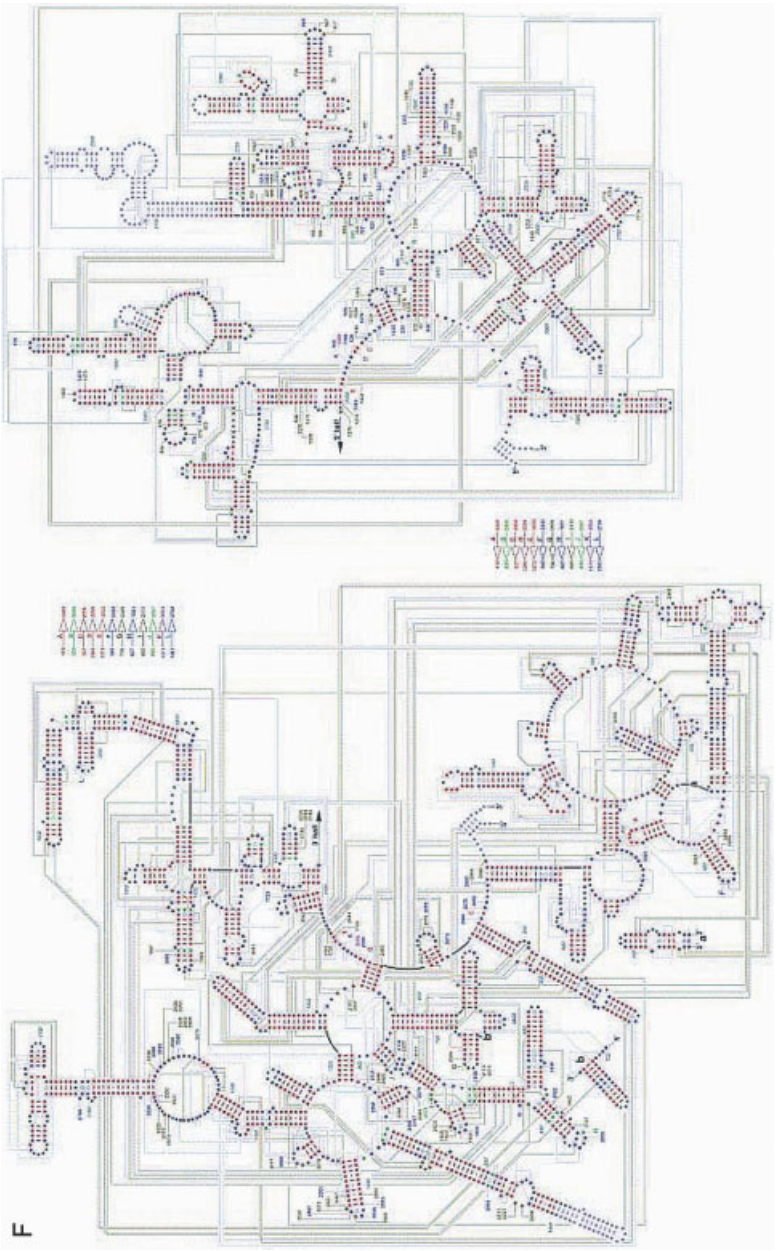


Figure 2-4 (contd.)

domain II of 50S in *H. marismortui* [64, 67]). Stacking of nucleic acids is driven by the highly energetically favored stacking interactions between the π-electron system of the nucleic acid bases [68]. The free energy from coaxial stacking of helices ending with Watson–Crick base pairs yields a $\Delta\Delta G^{\circ}$ of –1.0 to –4.5 kcal mol^{-1}, depending on the context. This is in the range of the contribution of next-neighbor interaction to the free energy of an intact helix (–2.0 and –3.4 kcal mol^{-1}) [69, 70]. Also in the same range of free energy is the contribution from coaxial stacking with a single G•A base pair at the interface of the stacking helices (about –2 kcal mol^{-1}, [71]). Therefore, 92% (11/12) of the potential coaxial stacking in the 16S rRNA and 50% (11/22) in the 23S rRNA are observed in the crystal structure. Potential coaxial helix stacking is defined by two helices with an A•G or A•A base-pair at their interface and no unpaired nucleotides in the strand connecting them [58]. Often the A•A and A•G, with the G 3′ to the helix, at the end of helices are inter-convertible.

A good example of a coaxially stacked helix is seen in a classic pseudoknot. This motif consists of a hairpin loop, which base pairs with a complementary single-stranded sequence adjacent to the hairpin stem, to form a contiguous helical structure. Similar to junctions, the coaxial stacking observed in the pseudoknot requires either Mg^{2+} ions or a high concentration of Na^{+} ions for stabilization [63]. Sequence-independent packing of ordinary helices is very seldom seen. It consists of an insertion of a phosphate ridge of one helix into a minor groove of the other at a fixed inter-helical angle of about 80° (e.g., in the 30S subunit for helices h7 and h21 with an angle of 93°). This kind of interaction is already known from the crystal structure of the 59-GGCGCUUGCGUC-39 RNA duplex. In this structure, the duplex forms quasi-continuous helices that pack against each other, with the backbone of one helix in the minor groove of a perpendicular helix [72].

Owing to the limitation to four principal nucleotides, the primary RNA sequence is by itself not enough to define a motif. Moore [66] suggested classifying only structural elements as motifs if they have a defined sequence length or specific loop sizes. This definition would exclude pseudoknots as motifs, as the sequence length and the loop size are unrestricted. Westhof tried to define an RNA motif as an ordered array of stacked non-Watson–Crick base-pairs and thereby focusing plainly on the 3D aspect of a motif. This is well exemplified by the complex topology around G911 in 23S rRNA. The 3D structure of this G911 topology (consisting of nucleotides 1068–1071 and 1292–1295 in one strand and nucleotides 910–914 in the base-pairing strand, shown in red in Fig. 2-5) superimposes well onto the sarcin/ricin motif, as seen in the internal-loop motif G225 in 23S rRNA (shown in blue in Fig. 2-4). The only parts of the motifs that do not superimpose well belong to the backbone, which in the composite motif connects to other parts of 23S RNA [65]. Although this definition of an RNA motif is quite successful, it does not encompasses the K-turn motif, where the array of stacked base pairs is interrupted by a triple loop to introduce a ~120° kink (see below).

As seen above, both secondary- and tertiary-structural aspects must be considered to provide an accurate definition of an RNA motif. In spite of these complications, evidence is accumulating that the modular structure of RNA motifs, which mediate

RNA–RNA (e.g., sarcin/ricin motif [65]) and RNA–protein interactions (e.g., K-turn, see later), are recurrent and are also found in the high-resolution structures of the ribosomal subunits.

2.5.2 A-minor Motifs

From the crystal structure of the *H. marismortui* 50S, the importance of the so-called A-minor motifs for stabilization of the 3D fold of large RNA became evident [73]. Phylogenetic co-variation analysis had already revealed a very strong bias for adenine (A) in single-stranded regions as compared with helical ones, hinting at the important structural or functional role of these adenines (see Table 2-2). One reason for this bias is the crucial importance of A-minor motifs in helix–helix packing, in helix–loop interactions and at helix junctions [73]. In the type-I A-minor motif, the N1-C2-N3 edge of a purine, preferentially a highly conserved A, interacts with the minor groove face of another helix (see Fig. 2-5). The A-minor motif was originally observed in the P4-P6 domain of group I ribozymes, where it was part of the ribose zipper [74], and was subsequently seen in the L11-protein-binding domain of 23S rRNA [75, 76].

Nissen et al. [73] distinguish between four variations in the A-minor motif (types O, I, II and III); however, we will focus primarily on types I and II, as they seem to be the most essential for ribosome function. Examples include the fixation of the CCA end of the tRNA at the PTC [20, 73], A-site recognition of the correct codon–anticodon interaction during the selection of the correct aminoacyl-tRNAs (Fig. 2-6B [59, 77], see also Chaps. 8.2 and 8.4) and subunit association (A702 of 16S rRNA with the minor groove of H68 of 23S rRNA; [1]).

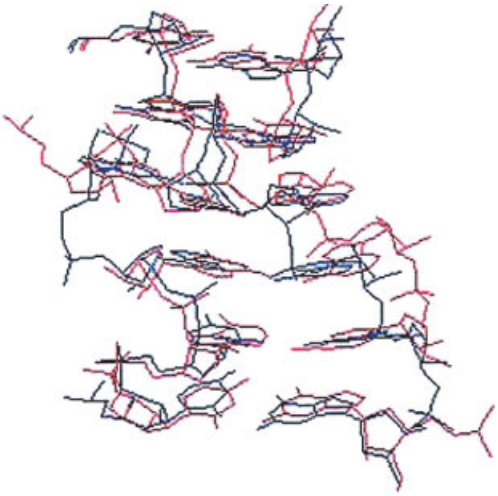


Figure 2-5 Superposition of crystal structures of composite sarcin motif, 23S G911 (red), and internal loop sarcin motif, 23S G225 (blue). From Ref. [65] with permission.

In all A-minor motifs, the ribose-phosphate backbone of the interacting adenosine is closer to one of the strands of the receptor helix. The orientation of the receptor helix strand closest to the adenosine is anti-parallel to the orientation of the adenosine of the donor strand. Type I and II A-minor motifs are different with respect to the position of the 2′-OH and N3 atoms of the adenosine residue relative to the closer strand of the receptor helix (Fig. 2-6; [73]). In type I, both the 2′-OH and the N3 of the adenosine residue are within the minor groove of the receptor helix (Fig. 2-6A), thereby maximizing the number of possible hydrogen bonds to the inserting adenosine. In type II, the N1, C2, N3, and 2′-OH of the adenosine contact approximately half of the minor groove. The 2′-OH of the inserting adenosine is positioned outside of the minor groove with respect to the 2′-OH of the closer strand of the receptor helix, whereas the N3 of the A is inside the minor groove (Fig. 2-6A; [73]). Both types of A-minor motifs are highly specific for adenine bases and show a strong preference for C-G receptor base pairs (Table 2-2; note the preference of adenosine in unpaired regions and the preference of G:C pairs in helices; Fig. 2-6; [73]).

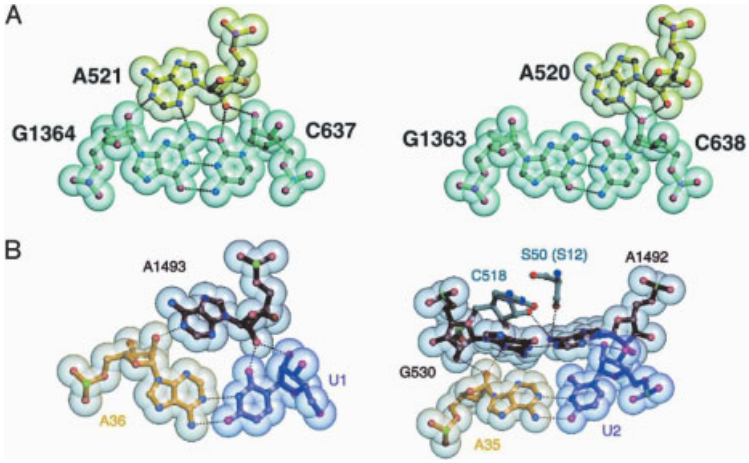


Figure 2-6 A-minor types I and II. (A) Ribosomal examples of A-minor type I and II. Each type is defined by the position of the 2′-OH group of the interacting adenosine relative to the positions of the two 2′-OH groups of the receptor base pair (see text for details). From Ref. [73] with permission. (B) Principles of decoding according to Ref. [59], with permission. For details see text.

Table 2-2 Frequency and distribution % of single nucleotides in bacterial 16S and 23S rRNAs comparative structure models [103]

Nucleotides	G	C	A	U
Overall	31.4	22.4	25.7	20.5
In helices	36.6	28.8	14.8	19.8
In unpaired regions	23.6	12.5	42.6	21.3
Unpaired / total number of positon of nt	30.1	22.3	66.2	41.5

2.5.3 Ribose Zippers and Patches of A-minor Motifs

A ribose zipper is defined as two consecutive hydrogen-bonding interactions between ribose 2′-OH from two different RNA segements. The orientation of the two chains linked by ribose zipper is always antiparallel. A total of 97 ribose zippers are present in the ribosomal subunit crystal structures: 20 in the *T. thermophilus* 16S rRNA, 44 in *H. marismortui* 23S rRNA (plus two ribose zippers bridging 5S rRNA loop E with H38 and the internal loop of H38 and the 5S rRNA helix 4) and 30 in *D. radiodurans* 23S rRNA (plus one bridging 5S and 23S rRNAs) [78]. Out of the 11 possible types of ribose zipper, seven are found in 23S and 16S rRNAs. From these only the canonical and single-base ribose zipper occur more than once in 23S and 16S rRNA [78].

2.5.3.1 Canonical Ribose Zipper

In this type of zipper, the 2′-OH hydrogen bonds are supported by additional hydrogen-bond interactions between the base at the 5′-end and the ribose 2′-OH of the 3′-end on the opposite zipper strand (see Fig. 2-7A). On the base side, the purine N3 or the pyrimidine O2 functions as a hydrogen acceptor. The ribose zipper formed by C376, C377 and A243, A242 in 23S rRNA of *H. marismortui,* has a prototypical topology for the 40 canonical ribose zippers found in the 16S and 23S rRNAs. A243 is inserted into the minor groove of the C376:G274 base-pair forming a type I A-minor motif. A242 interacts via a type II A-minor motif with C377:G273 base pair and is stacked onto A243 (Fig. 2-7B). This tandem A-minor motif seen in 31 instances of the 40 canonical ribose zipper in *T. Thermophilus* 30S and *H. marismortui* 50S had been noticed earlier by Nissen et al. [73], where it was referred to as an "A patch".

The adenosine in the type I A-minor motif, which exhibits stronger conservation than the type II, shows a CG>GC>UA=AU order of preference with regard to the Watson–Crick pairs it interacts with. In contrast, no base-pair preference was detected for the type II A-minor motif [78]. This is in line with experimental and phylogenetic covariation analysis on group I introns [79]. Moreover, with the A-minor patches found in group I introns, the average contribution of the ribose hydrogen bond to the tertiary fold was determined to a $\Delta\Delta G°$ of –0.4 to –0.5 [80] and –6.6 kcal/mol^{-1} for type I A-minor motif [79].

2.5.3.2 Single-base Ribose Zipper

The single-base ribose zipper is a canonical ribose zipper with one obstructed base 2′-OH interaction. Two possible types of single-base ribose zippers can be distinguished: types A and B; the definition depending on which base ribose interaction is interrupted, i.e., in either the type I or type II A-minor position, respectively. Consequently, the 15 single-base ribose zippers identified within the *T. thermophilus* 30S and *H. marismortui* 50S can be separated into 10 type A and 5 type B. Almost all single-base ribose zippers have adenine in the type I A-minor position, which is, in the

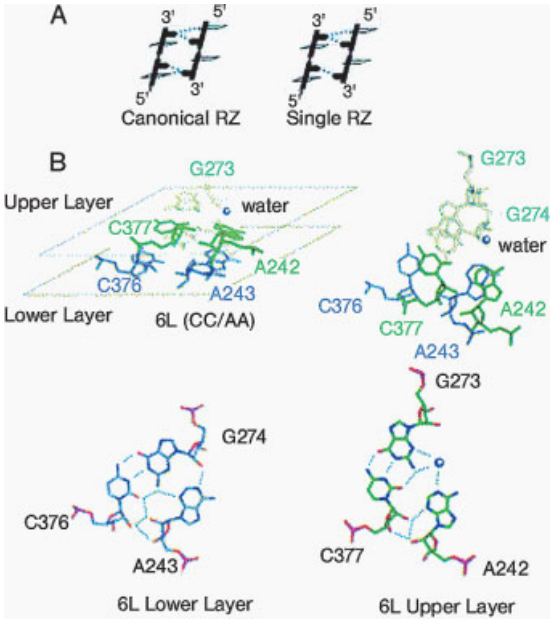


Figure 2-7 Ribose zipper. (A) Schematic representation of canonical and single-base ribose zippers type A. Light blue colored broken lines represent hydrogen bonds. (B) Stick diagrams of the canonical ribose zipper. Top panels: canonical ribose zipper where C377 and A242 belong to the upper layer (blue green) and C376 and A243 belong to the bottom layer (slate) (top left panel) viewed perpendicular to the back-bone, showing the ribose–ribose inter-actions and (top right panel) from above the upper layer. Lower panels: hydrogen-bond network in the lower layer (to the right) and the upper layer (to the left). From Ref. [78] with permission.

case of type A, rotated away from the "acceptor base pair". In the type B single-base ribose zipper the type II A-minor position is a G in all instances [78]. Comparison of the two available 50S structures revealed that 10 of the canonical ribose zippers that are present in *H. marismortui* are also found in *D. radiodurans*, although two canonical ribose zippers in *H. marismortui* are single-base ribose zippers in *D. radiodurans*.

All ribose zippers of *H. marismortui* 50S structure are either conserved or their composing bases are in close proximity to each other in *D. radiodurans* structure. For example, 10 *canonical ribose zippers* found specifically in *H. marismortui* have similar positions in *D. radiodurans*; however, the bases are too far apart from each other to form hydrogen bonds.

2.6 Progress and New Developments in Understanding rRNA Structures

To finalize the overview of motifs recognized in the ribosomal subunit, the K-turn and lonepair triloop will be reviewed as well as the attempts that have been made to categorize non-Watson–Crick base interactions and RNA motifs in databases.

2.6.1 K-turn

After careful analysis of crystal structure of the large subunit, Klein et al. [62] recognized a new motif, consisting of a helix–internal loop–helix. A kink in the phosphodiester backbone of the internal loop bends the RNA helix axis by ~120° and also gives the motif its name 'Kink-turn' or 'K-turn'.

The first helical stem termed 'canonical stem' (C-stem) ends at the internal loop with two Watson–Crick base pairs, typically C:Gs. The second helical stem termed 'non-canonical stem' (NC-stem) follows the internal loop and starts with two non-Watson–Crick base pairs, typically sheared G•A base pairs. The internal loop between the helical stems is always asymmetrical and usually contains three unpaired nucleotides in one strand but none in the other (Fig. 2-8A). The close helical packing between the helical stems is stabilized by a type I A-minor interaction [73] between the last C:G of the C-stem and the A of the G•A base pair in the NC-stem (Fig. 2-8B). The requirement for a type I A-minor interaction may account for the conservation of C:G in the C-stem and A•G in the NC-stem and also the high conservation with the consensus secondary structure for the K-turn, which includes 10 consensus nucleotides out of the possible 15 [62]. Although the eight K-turns identified in the *H. marismortui* 23S and *T. thermophilus* 16S rRNA vary somewhat in sequence, each has essentially the same distinctive 3D structure. The six K-turns in *H. marismortui* 23S rRNA superimpose with an r.m.s.d. of 1.7 Å (r.m.s.d.: root-mean-square deviation). All six of these K-turn motifs appear at or near the surface of the 50S particle, in regions that are less well conserved among the three kingdoms (Fig. 2-8C). The two K-turns identified in the structure of *T. thermophilus* 30S are localized in the intersubunit surface and probably play a structural role in the association of the subunits.

2.6.2 Lonepair Triloop

As implied in the name, the lonepair triloop (LPTL) consists of a lone base pair capped with a loop of three nucleotides. The nucleotide sequence within LPTLs can be described as 5′–FXYZL–3′, where the underlined nucleotides F and L form the lonepair and the three nucleotides, X, Y and Z, the triloop [81]. Twenty-three LPTLs occur in the ribosomal RNAs, of which seven are in 16S rRNA of *T. thermophilus* 30S, 15 in the 23S and one in 5S rRNA of the *H. marismortui* 50S. The LPTLs in *D. radiodurans* 50S are at the corresponding positions of 23S RNA of *H. marismortui* and are structurally equivalent. In addition to the LPTLs recognized in the ribosomal RNA an additional one was found in the T-loop of tRNAs [81]. Nearly all of the LPTL sites in rRNAs are conserved and most of them contribute to rRNA packing via tertiary interactions with RNA segments that are distant in terms of the secondary structure [81].

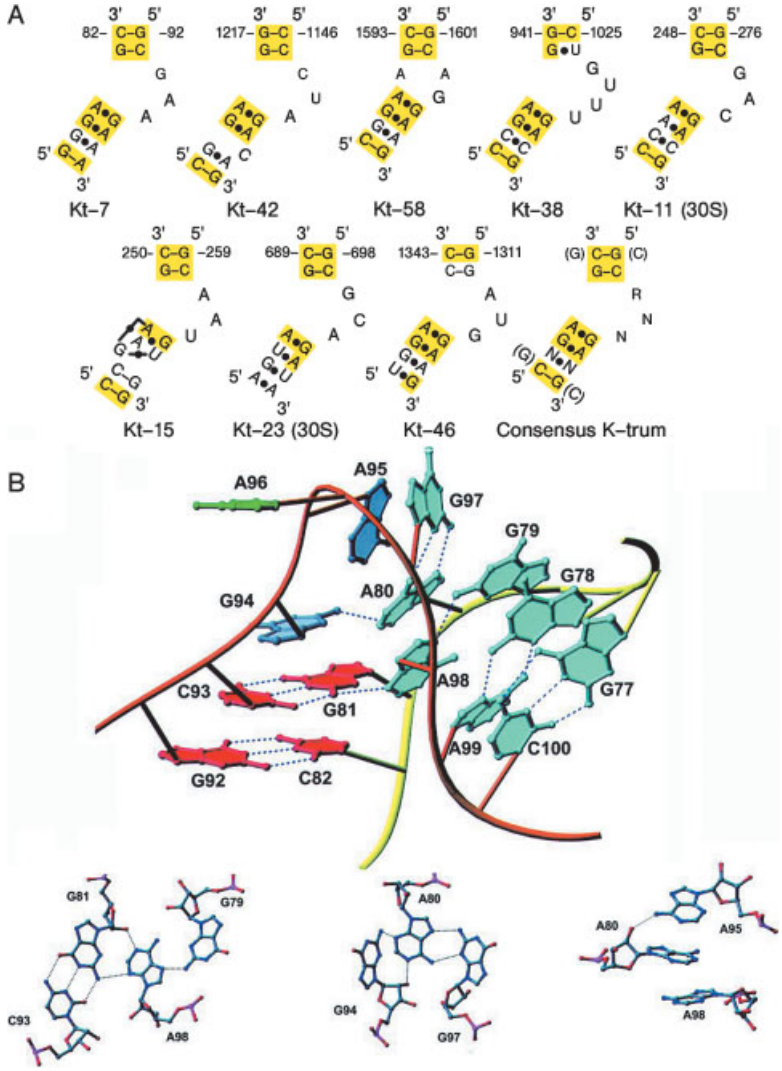


Figure 2-8 K-turn motif. (A) Secondary-structure diagrams of the eight K-turns found in the *H. marismortui* 50S and *T. thermophilus* 30S subunit structures and a derived consensus sequence. (B) Three-dimensional representation of KT-7 with the phosphate backbone of the kinked strand in orange and the unkinked strand in yellow (upper panel). Lower panels displays atomic details of individual base–base and base–backbone hydrogen bonds. (C) Location of the K-turns in the *H. marismortui* 50S structure. K-turns in blue are shown in the 3D structure: crown view from the interface (top left panel), cytosolic face (top right panel) and in the secondary structure (lower panel). From Ref. [62] with permission.

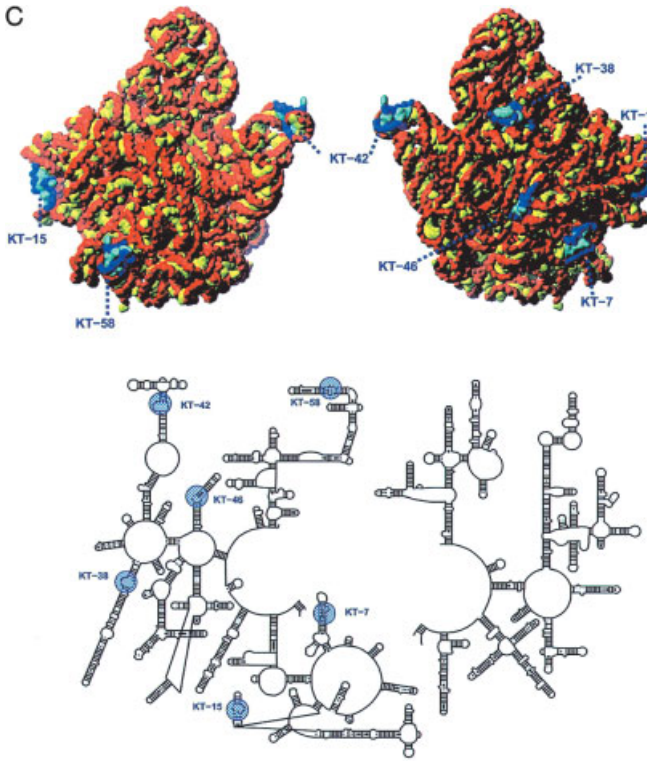


Figure 2-8 cond.

2.6.2.1 **Classification of Lonepair Triloops**

All but one of the LPTLs are coaxially stacked on the nearest 5′-helix. The LPTLs can be classified as directly (class I) or indirectly (mediated by a short helical region (class II)), coaxially stacking LPTLs. In both classes, the 5′-base of the lonepair (F) is stacked onto the first nucleotide of the triloop (X), which is connected to the second nucleotide of the triloop (Y) by a 220–230° U-turn. The second and third nucleotides are facing into the minor groove, with the third base forming hydrogen bonds to the first nucleotide of the triloop (see Figs. 2-9A and B).

In a subpopulation of type I and II LPTLs, a tertiary base is recruited by the triloop by forming a base pair to the first base of the triloop (X). This recruited tertiary base stacks between the third base (Z) and the 3′-base in the lonepair (L) and yields a structural conformation resembling that of the tetraloop motif. In contrast to the above-described type A LPTLs, no tertiary base is recruited by type B LPTLs [81]. A third class of LPTLs includes all LPTLs within a helical region. This category of LPTLs is structurally distinct from the first two classes in having at least two of the triloop bases base-pairing to form part of a regular helical stem and in missing the U-turn

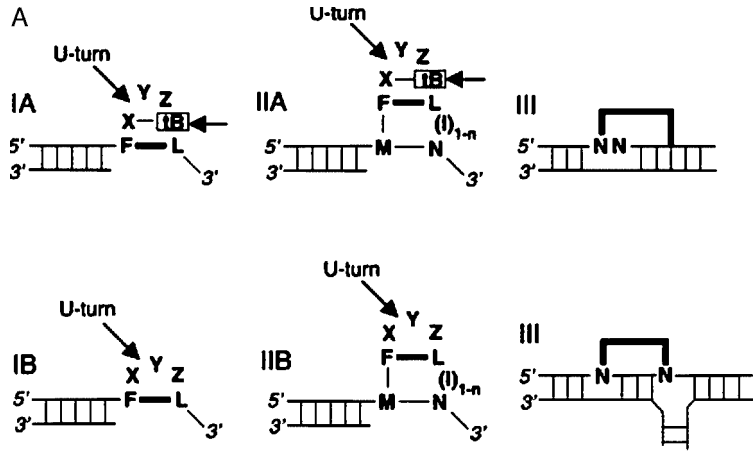


Figure 2-9 Lonepair triloop. (A) Schematic representation of IA, IB, IIA, IIB, and III LPTLs. The lonepair F:L is appended either directly (class I) or indirectly (class II) through the intervening base-pair(s) M:N to its 5′-helix, which is shown with lines (see text for details). (B) Three-dimensional structure of representatives of type IA, IB, and IIA LPTLs. Nucleotides are numbered in black give *T. thermophilus* numbers for 16 S rRNA and *H. marismortui* numbers for 23 S rRNA and in red *E. coli* numbers. From Ref. [81] with permission.

in the triloop [81]. The LPTLs reviewed here include the earlier reported T-loop RNA fold as a type I LPTL [82].

2.6.3 Systemizing Base Pairs

To facilitate the understanding of non-canonical base interactions, which have proven to stabilize secondary and tertiary structures, different groups have put together compilations of non-Watson–Crick interactions. Westhof and colleagues have sorted the base–base interaction by the C1′–C1′ distance and the orientation of the glycosidic bonds. Two base pairs with nearly the same C1′–C1′ distance and same glycosidic bond orientation can replace each other without drastically changing the 3D path and relative geometric orientations of the phosphate-sugar backbones. These base pairs are called isosteric, although they will not always occupy the same volume of space. This collection of isosteric base-pair interactions will be useful in the development of more accurate 3D RNA models.

In the course of this work, Leontis and Westhof [83] systemized the nomenclature of base pairs, based on the three potential hydrogen bond forming edges of a base. These are the Watson–Crick, the Hoogsteen edge for purines and the CH edge for pyrimidines, and the sugar or the shallow groove edge (Fig. 2-10).

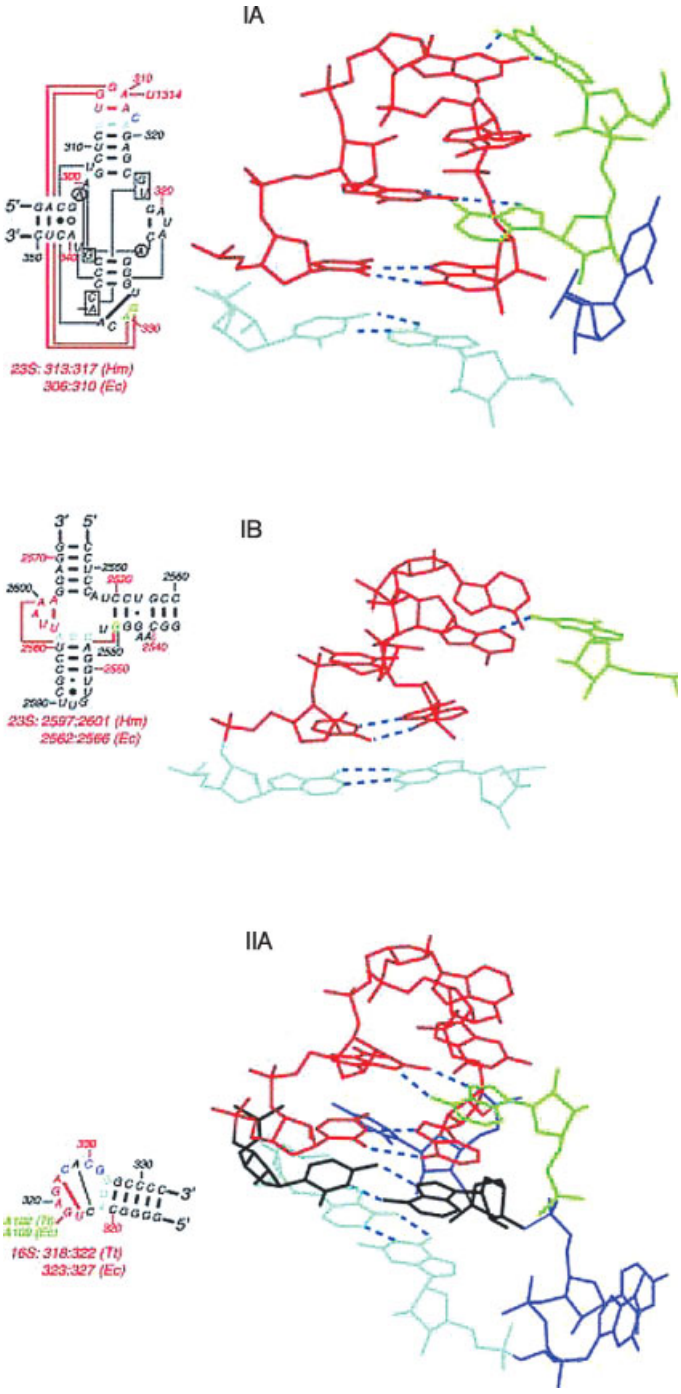


Figure 2-9 (cond.)

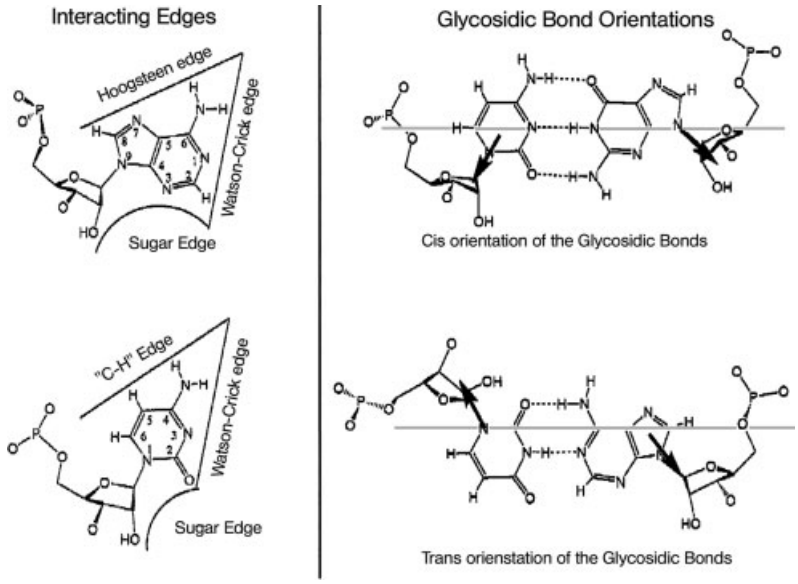


Figure 2-10 Proposed nomenclature for base pairs by Leontis and Westhof. Left panel: purine (A or G, indicated by "R") and pyrimidine (C or U, indicated by "Y") bases provide three edges for interaction, as shown for adenosine and cytosine. Right panel: the *cis* and *trans* orientations are defined relative to a line drawn parallel to and between the *base-to-base* hydrogen bonds in the case of two hydrogen bonds or, in the case of three hydrogen bonds, along the middle hydrogen bonds. From Ref. [83] with permission.

In the NCIR (non-canonical interactions in known RNA structures) database, Fox and colleagues [84] compiled all the RNA structures that contain non-Watson–Crick base–base interactions. In addition, this database provides a summary of the known properties of the base–base interaction (http://prion.bchs.uh.edu/bp_type/).

2.6.4 Systemizing RNA Structural Elements

In their structural classification of RNA (SCOR) database (http://scor.lbl.gov), Fox and colleagues categorized RNA motifs either as an external or internal loop. To define loops, they applied a strict definition for each loop; external loops being a covalently connected series of residues non-Watson–Crick-paired to each other, which are closed on one side by a Watson–Crick base pair, whereas internal loops consist of one (for bulge loops) or two non-Watson–Crick paired residues, closed on both sides by Watson–Crick base pairs. Using their definitions, bulged loops or G•U base pairs within a standard helix are considered internal loops [85]. Two hundred and twenty-three internal and 203 external loops extracted from the 259 NMR and

X-ray structures are complied in the SCOR database. The internal loops are further divided into nine subclasses. The external loops are categorized by size and stacking pattern within the loop. Examples of RNA tertiary interactions are also included in the SCOR database, such as coaxial helical stacking, ribose zippers, A-minor interactions or pseudoknots [85].

2.7 RNA–protein Interactions

The ribosome consists of approximate 53 ribosomal proteins, which are crucial for the smooth functioning of protein synthesis. Therefore, we will turn our attention now to principles governing RNA–protein interaction. We will start by explaining the differences between RNA and DNA and the implication this has for binding protein, which will lead us to the principal modes of interaction between protein and RNA, and ending in the specifics of ribosomal protein interaction with rRNA.

2.7.1 Problem of RNA Recognition

The A-form helix accounts for as much as 50% of the residues in an average non-messenger RNA, which includes the bases of Watson–Crick base pairs as well as G•U wobble base pairs in the runs of Watson–Crick base pairs [66]. In the A-form, the minor groove is shallow and broad (10–11 Å in width and 2.8 Å in depth) whereas the major groove is deep and narrow (4–5 Å in width and 13.5 Å in depth), when compared with B-form helix of DNA (major groove: 12 Å in width and 8.5 Å in depth, minor groove: 5.8 Å in width, 7.5 Å in depth). This allows functional groups of the ribose sugar, in particular the 2′-OH, to participate in interactions. The 2′-OH group can act as a hydrogen bond donor and/or acceptor, making it versatile when it comes to interaction with both RNA and protein. The functional importance of 2′-OH group for protein synthesis was experimentally reinforced in many instances, such as decoding [86], for which the involvement of 2′-OH of the mRNA was structurally rationalized [59], translocation of tRNA(Met) from P-site to the E-site, which is dependent on 2′-OH groups at positions 71 and 76 in the 3′-acceptor arm of the tRNA [87], or tRNA binding to EF-Tu or to aminoacyl-tRNA synthetases [88, 89]

Sequence-specific interactions with regular A-form RNA helices via direct H-bonding and van der Waals interactions *cannot* distinguish, in the minor groove, a G:C from C:G or an A:U from a U:A base pair, but can distinguish G:C from A:U types [90]. This fact degenerates the recognition of base pairs from a quaternary mode (four base pairs) to a binary mode (two kinds of base pairs). This contrasts with the recognition, in the major groove, of B-form DNA where discrimination of all four Watson–Crick base pairs (GC, CG, AT, and TA) is possible (Fig. 2-11; [90]).

This difference in discrimination is reflected in the different mode of sequence-specific recognition of DNA and RNA. DNA recognition is typically accomplished through the recognition of a particular nucleotide sequence in double-stranded DNA (dsDNA). However, the sequence-specific RNA recognition results largely through single-stranded regions, bulges, or internal and terminal loops [91, 92]. The complex

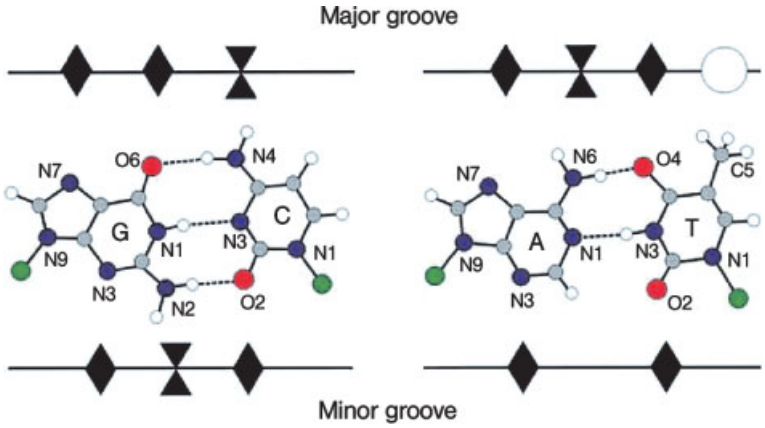


Figure 2-11 A schematic representation of the different patterns of hydrogen-bond donors (two triangles connected on one tip) and acceptors (diamonds) presented by Watson–Crick pairs to the major and the minor grooves. A varied pattern of hydrogen donors and acceptors in the major groove allows easy discrimination of AT, TA, GC, and CG, whereas in the minor groove only the discrimination between AT and GC base pairs is possible due to the symmetric distribution of the donors and acceptors. Adapted From Ref. [90].

structure of the RNA moiety within the protein–RNA complex hampers tight packing at the protein–RNA interface. Therefore, the protein–RNA interface is less well packed in comparison with protein–dsDNA or protein–ssDNA (single-stranded DNA) complexes. Within the protein–RNA complexes, the sequence-specific complexes achieve the best packing with, surprisingly, the least polar protein interface [92].

The difference in sequence-specific recognition of dsDNA and ssRNA also resonates in the hydrogen-bond pattern between amino acids and bases. In dsDNA–protein complexes the functional groups of amino acid side-chains interact with the nucleotide bases in the accessible major groove. In RNA–protein complexes the amide backbone is frequently used for specific base interactions, especially the carbonyl group of the amide group as a proton acceptor. A duplex helix hampers close contact of a peptide backbone with a nucleotide base to a larger extent than ssRNA does. Interestingly, the extent to which an amide group is used for phosphodiester backbone contacts is approximately the same in RNA and DNA complexes [91].

2.7.2 Chemistry of RNA–protein Interactions

In most of the recently published studies on protein–RNA interaction, the structure of *H. marsimortui* 50S is excluded because the constrains of protein–RNA interaction is compounded by the problem of counteracting the high-salt conditions found in *H. marsimortui* cells. A statistical analysis of protein–RNA interactions of the 30S

subunit, synthetase–tRNA complexes and other RNA–protein complexes revealed that 22% of the amino acids in the RNA–protein interface are hydrophobic, 40% charged (positive 32%, negative 8%), 30% polar and 8% glycine. The preferred amino acids at the RNA–protein interface are Arg, Ser, His, Tyr, Lys, with the underrepresented ones being Ala, Val, Glu, Met, Leu, Ile [93]. Statistical analysis has also established that some amino acids have preferences for interacting with either the phosphate group, the ribose moiety or the nucleotide base of the RNA; Arg prefers to interact with the phosphate backbone; Met, Phe and Tyr favor the ribose; Pro and Asn display stronger affinity for bases over ribose or phosphate groups. Additionally, the four bases have individual preferences, such that adenosine favors Ile, Pro, Ser; cytosine, Leu; guanosine, Asp and Gly, and uracil, Asn [93].

2.7.3 rRNA–protein Interaction

Ribosomal proteins make more extensive use of contacts to the sugar-phosphate backbone, which is reflected in the preference for hydrogen bonds to phosphate groups, over ribose, followed by bases. Whereas proteins with sequence-specific binding to ssRNA-like structures make a surprisingly small number of contacts to the RNA backbone and prefer, by far, hydrogen bonds to the base. In-between these two extremes is the tRNA synthetase family, using phosphate, ribose and bases equally for hydrogen-bond contacts [91, 93].

Based on the hypothesis that the ribosome is evolutionary older than tRNA synthetases, which are older than ssRNA–protein complexes, Allers and Shamoo [91] put forward an intriguing idea: They proposed that in the earliest interaction of an "RNA world" RNA-like molecules had only the minimalist chemical interactions with amino acids and therefore would have only take advantage of the abundant peptide backbone amide and carbonyl groups.

The preference of ribosomal proteins to interact with the RNA backbone led to the hypothesis that shape and charge complementarity, rather than specific sequences, are responsible for the specificity observed in ribosomal protein–RNA interactions [75, 76]. This hypothesis is in agreement with co-variation sequence analysis and can also explain the latter's inefficiency in recognizing r-protein binding sites.

As explained above, some of the forces within an RNA–protein complex are understood; however, the understanding of RNA–protein recognition and the mutual interplay between protein and RNA, as seen in the ribosome assembly, is still a long way off [91, 94].

References

1 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001**, *292*, 883–996.
2 V. Ramakrishnan, P. B. Moore, *Curr. Opin. Struct. Biol.* **2001**, *11*, 144–154.
3 B. T. Wimberly, D. E. Brodersen, W. M. Clemons et al., *Nature* **2000**, *407*, 327–339.
4 D. E. Brodersen, W. M. Clemons, Jr, A. P. Carter et al., *J. Mol. Biol.* **2002**, *316*, 725–768.
5 N. Nevskaya, S. Tishchenko, M. Paveliev et al., *Acta Crystallogr.* **2002**, *58*, 1023–1029.
6 A. Nikulin, I. Eliseikina, S. Tishchenko et al., *Nat. Struct. Biol.* **2003**, *10*, 104–108.
7 N. Nevskaya, S. Tischenko, R. Fedorov et al., *Struct. Fold. Des.* **2000**, *8*, 363–371.
8 S. Nikonov, N. Nevskaya, I. Eliseikina et al., *EMBO J.* **1996**, *15*, 1350–1359.
9 J. Frank, *Electron Microsc. Rev.* **1989**, *2*, 53–74.
10 H. Stark, F. Mueller, E. V. Orlova et al., *Structure* **1995**, *3*, 815–821.
11 N. Ban, P. Nissen, J. Hansen et al., *Science* **2000**, *289*, 905–920.
12 R. A. Milligan, P. N. Unwin, *Nature* **1986**, *319*, 693–695.
13 A. Yonath, K. R. Leonard, H. G. Wittmann, *Science* **1987**, *236*, 813–816.
14 C. Bernabeu, E. M. Tobin, A. Fowler et al., *J. Cell. Biol.* **1983**, *96*, 1471–1474.
15 P. Dube, M. Wieske, H. Stark et al., *Structure* **1998**, *6*, 389–399.
16 R. Beckmann, D. Bubeck, R. Grassucci et al., *Science* **1997**, *278*, 2123–2126.
17 R. Beckmann, C. M. Spahn, N. Eswar et al., *Cell* **2001**, *107*, 361–372.
18 J. F. Menetret, A. Neuhof, D. G. Morgan et al., *Mol. Cell* **2000**, *6*, 1219–1232.
19 J. Harms, F. Schluenzen, R. Zarivach et al., *Cell* **2001**, *107*, 679–688.
20 P. Nissen, J. Hansen, N. Ban et al., *Science* **2000**, *289*, 920–930.
21 J. Frank, J. Zhu, P. Penczek et al., *Nature* **1995**, *376*, 441–444.
22 I. S. Gabashvili, S. T. Gregory, M. Valle et al., *Mol. Cell* **2001**, *8*, 181–188.
23 T. Tenson, M. Ehrenberg, *Cell* **2002**, *108*, 591–594.
24 C. M. Spahn, R. Beckmann, N. Eswar et al., *Cell* **2001**, *107*, 373–386.
25 J. Greenblatt, *Curr. Opin. Cell Biol.* **1997**, *9*, 310–319.
26 P. S. Lovett, E. J. Rogers, *Microbiol. Rev.* **1996**, *60*, 366–385.
27 D. R. Morris, A. P. Geballe, *Mol. Cell Biol.* **2001**, *20*, 8635–8642.
28 F. Gong, C, Yanofsky, *Science* **2002**, *297*, 1864–1867.
29 M. N. Kazarinoff, E. E. Snell, *J. Biol. Chem.* **1977**, *252*, 7598–7602.
30 M. C. Deeley, C. Yanofsky, *J. Bacteriol.* **1981**, *147*, 787–796.
31 V. Stewart, C. Yanofsky, *J. Bacteriol.* **1986**, *167*, 383–386.
32 F. Gong, C, Yanofsky, *J. Biol. Chem.* **2001**, *276*, 1974–1983.
33 D. Oliver, J. Norman, S. Sarker, *J. Bacteriol.* **1998**, *180*, 5240–5242.
34 H. Nakatogawa, K. Ito, *Mol. Cell* **2001**, *7*, 185–192.
35 H. Nakatogawa, K. Ito, *Cell* **2002**, *108*, 629–636.
36 R. Berisio, F. Schluenzen, J. Harms et al., *Nat. Struct. Biol.* **2003**, *10*, 366–370.
37 A. E. Yonath, J. Mussig, B. Tesche et al., *Biochem. Int.* **1980**, *1*, 428–435.
38 W. M. Clemons, Jr, D. E. Brodersen, J. P. McCutcheon et al., *J. Mol. Biol.* **2001**, *310*, 827–843.
39 F. Schluenzen, A. Tocilj, R. Zarivach et al., *Cell* **2000**, *102*, 615–623.
40 N. Ban, B. Freeborn, P. Nissen et al., *Cell* **1998**, *93*, 1105–1115.
41 I. S. Gabashvili, R. K. Agrawal, C. M. Spahn et al., *Cell* **2000**, *100*, 537–549.
42 D. I. Svergun, M. H. Koch, J. S. Pedersen et al., *J. Mol.Biol.* **1994**, *240*, 78–86.
43 C. M. Spahn, P. A. Penczek, A. Leith et al. *Struct. Fold Des.* **2000**, *8*, 937–948.

44 C. Glotz, R. Brimacombe, *Nucleic Acids Res.* **1980,** *8,* 2377–2395.
45 A. Nakagawa, T. Nakashima, M. Taniguchi et al., *EMBO J.* **1999,** *18,* 1459–1467.
46 M. Worbs, R. Huber, M. C. Wahl, *EMBO J.* **2000,** *19,* 807–818.
47 S. Spillmann, F. Dohme, K. H. Nierhaus, *J. Mol. Biol.* **1977,** *115,* 513–523.
48 A. Yonath, *Curr. Protein Pept. Sci.* **2002,** *3,* 67–78.
49 R. K. Agrawal, J. Linde, J. Sengupta et al., *J. Mol. Biol.* **2001,** *311,* 777–787.
50 M. G. Gomez-Lorenzo, C. M. Spahn, R. K. Agrawal et al., *EMBO J.* **2000,** *19,* 2710–2718.
51 M. Valle, A. Zavialov, J. Sengupta et al., *Cell* **2003,** *114,* 123–134.
52 J. L. Hansen, T. M. Schmeing, P. B. Moore et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 11670–11675.
53 T. M. Schmeing, A. C. Seila, J. L. Hansen et al., *Nat. Struct. Biol.* **2002,** *9,* 225–230.
54 S. V. Kirillov, J. Wower, S. S. Hixson et al., *FEBS Lett.* **2002,** *514,* 60–66.
55 R. R. Gutell, J. C. Lee, J. J. Cannone, *Curr. Opin. Struct. Biol.* **2002,** *12,* 301–310.
56 R. R. Gutell: in *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis,* eds R. A. Zimmermann and A. E. Dahlberg, CRC Press, Boca Raton, FL 1996, 111–128.
57 J. Thompson, W. E. Tapprich, C. Munger et al., *RNA* **2001,** *7,* 1076–1083.
58 T. Elgavish, J. J. Cannone, J. C. Lee et al., *J. Mol. Biol.* **2001,** *310,* 735–753.
59 J. M. Ogle, D. E. Brodersen, Jr. W. M. Clemons et al., *Science* **2001,** *292,* 897–902.
60 P. Schimmel, B. Henderson, *Proc. Natl. Acad. Sci. USA* **1994,** *91,* 11283–11286.
61 M. A. Schafer, A. O. Tastan, S. Patzke et al., *J. Biol. Chem.* **2002,** *277,* 19095–19105.
62 D. J. Klein, T. M. Schmeing, P. B. Moore et al., *EMBO J.* **2001,** *20,* 4214–4221.
63 R. T. Batey, R. P. Rambo, J. A. Doudna, *Angew. Chem.-Int. Ed.* **1999,** *38,* 2327–2343.
64 T. Hermann, D. J. Patel, *J. Mol. Biol.* **1999,** *294,* 829–849.
65 N. B. Leontis, J. Stombaugh, E. Westhof, *Biochimie* **2002,** *84,* 961–973.
66 P. B. Moore, *Annu. Rev. Biochem.* **1999,** *68,* 287–300.
67 S. R. Holbrook, S. H. Kim, *Biopolymers* **1997,** *44,* 3–21.
68 W. Saenger: *Principles of Nucleic Acid Structure,* ed. C. R. Cantor, Springer, New York 1984.
69 A. E. Walter, D. H. Turner, *Biochemistry* **1994,** *33,* 12715–12719.
70 A. E. Walter, D. H. Turner, J. Kim et al., *Proc. Natl. Acad. Sci. USA* **1994,** *91,* 9218–9222.
71 J. Kim, A. E. Walter, D. H. Turner, *Biochemistry* **1996,** *35,* 13753–13761.
72 S. E. Lietzke, C. L. Barnes, J. A. Berglund et al., *Structure* **1996,** *4,* 917–930.
73 P. Nissen, J. A. Ippolito, N. Ban et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 4899–4903.
74 J. H. Cate, A. R. Gooding, E. Podell et al., *Science* **1996,** *273,* 1678–1685.
75 G. L. Conn, D. E. Draper, E. E. Lattman et al., *Science* **1999,** *284,* 1171–1174.
76 B. T. Wimberly, R. Guymon, J. P. McCutcheon et al., *Cell* **1999,** *97,* 491–502.
77 D. N. Wilson, G. Blaha, S. R. Connell et al., *Curr. Protein Pept. Sci.* **2002,** *3,* 1–53.
78 M. Tamura, S. R. Holbrook, *J. Mol. Biol.* **2002,** *320,* 455–474.
79 E. A. Doherty, R. T. Batey, B. Masquida et al., *Nat. Struct. Biol.* **2001,** *8,* 339–343.
80 S. K. Silverman, T. R. Cech, *Biochemistry* **1999,** *38,* 8691–8702.
81 J. C. Lee, J. J. Cannone, R. R. Gutell, *J. Mol. Biol.* **2003,** *325,* 65–83.
82 U. Nagaswamy, G. E. Fox, *RNA* **2002,** *8,* 1112–1119.
83 N. B. Leontis, E. Westhof, *RNA* **2001,** *7,* 499–512.

84 U. Nagaswamy, M. Larios-Sanz, J. Hury et al., *Nucleic Acids Res.* **2002,** *30,* 395–397.

85 P. S. Klosterman, M. Tamura, S. R. Holbrook et al., *Nucleic Acids Res.* **2002,** *30,* 392–394.

86 A. P. Potapov, F. J. Triana-Alonso, K. H. Nierhaus, *J. Biol. Chem.* **1995,** *270,* 17680–17684.

87 J. S. Feinberg, S. Joseph, *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 11120–11125.

88 J. A. Pleiss, O. C. Uhlenbeck, *J. Mol. Biol.* **2001,** *308,* 895–905.

89 J. A. Pleiss, A. D. Wolfson, O. C. Uhlenbeck, *Biochemistry* **2000,** *39,* 8250–8258.

90 T. A. Steitz: in *The RNA World : the Nature of Modern RNA Suggests a Prebiotic RNA,* eds R. Gesteland, T. R. Cech and J. Atkins, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 1999, 427–450.

91 J. Allers, Y. Shamoo, *J. Mol. Biol.* **2001,** *311,* 75–86.

92 S. Jones, D. T. Daley, N. M. Luscombe et al., *Nucleic Acids Res.* **2001,** *29,* 943–954.

93 M. Treger, E. Westhof, *J. Mol. Recog.* **2001,** *14,* 199–214.

94 J. R. Williamson, *Nat. Struct. Biol.* **2000,** *7,* 834–837.

95 B. Weisblum, *Antimicrob. Agents Chemother.* **1995,** *39,* 797–805.

96 P. Delbecq, O. Calvo, R. K. Filipkowski et al., *Curr. Genet.* **2000,** *38,* 105–112.

97 P. Fang, Z. Wang, M. S. Sachs, *J. Biol. Chem.* **2000,** *275,* 26710–26719.

98 G. L. Law, A. Raney, C. Heusner et al., *J. Biol. Chem.* **2001,** *276,* 38036–38043.

99 A. L. Parola, B. K. Kobilka, *J. Biol. Chem.* **1994,** *269,* 48497–4505.

100 A. Raney, G. L. Law, G. J. Mize et al., *J. Biol. Chem.* **2002,** *277,* 5988–5994.

101 K. Reynolds, A. M. Zimmer, A. Zimmer, *J. Cell Biol.* **1996,** *134,* 827–835.

102 J. P. Alderete, S. Jarrahian, A. P. Geballe, *J. Virol.* **1999,** *73,* 8330–8337.

103 R. R. Gutell, J. J. Cannone, Z. Shang et al., *J. Mol. Biol.* **2000,** *304,* 335–354.

104 M. Stöffler-Meilicke, G. Stöffler: in *The Ribosome. Structure, Function and Evolution,* eds W. E. Hill, A. Dahlberg, R. A. Garrett et al., American Society for Microbiology, Washington, DC 1990 123–133.

105 J. Frank, R. Agrawal: in *Embryonic Encyclopedia of Life Sciences Nature Publishing Group,* London 1999, www.els.net.

3
Ribosome Assembly

3.1
Assembly Of The Prokaryotic Ribosome

Knud H. Nierhaus

3.1.1
Introduction

A ribosome consists of a large number of different components (e.g., *Escherichia coli*: 54 r-proteins and three rRNAs), all of which are present in one copy per ribosome, the only exception being that of the ribosomal protein L7/L12, which is present in four copies. This fact has two important consequences for the ribosomal biogenesis: both the synthesis and the assembly of the ribosomal components must occur in a highly coordinated fashion. The principles of ribosomal assembly in prokaryotes, such as the assembly maps, rate-limiting steps and that the assembly gradient follows the transcription of the ribosomal RNA were uncovered primarily by *in vitro* techniques. The assembly of eukaryotic ribosomes is described in Chap. 3.2.

The requirement for a highly coordinated synthesis is particularly demanding in the cases in which ribosomes contribute significantly to the dry mass of the cell. In bacteria, the ribosomes can amount to more than 50% of the dry mass [1], whereas in eukaryotes they represent not more than 5% [2]. In fact, cells of *E. coli* appear as sacs filled with ribosomes in images of transmission electron microscopy. Correspondingly, more than 50% of the total energy production of bacteria is consumed by ribosomal biogenesis. Therefore, it is understandable that a coordinated synthesis is not only a prerequisite for a successful and effective assembly, but is also a necessity for economic consumption of the energy available to the cell. We thus find an intricate network of regulatory mechanisms for the synthesis of ribosomal components in bacteria, which do not exist to the same degree in eukaryotes (e.g., translational control of mRNAs carrying cistrons for ribosomal proteins and the stringent response – Chap. 11; but see the eIF2 regulation in yeast – Chap. 7.2).

Third-order reactions practically do not exist, since the probability of three substrates reacting simultaneously is negligibly low. By extension, the assembly of more than 20 components (in the case of the small ribosomal subunit) to a defined and relatively compact particle is a series of reactions. This construction is a self-assembly process, i.e. the total information for the pathway as well as for the quaternary

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

structure of the active ribosomes resides completely in the primary sequences of the ribosomal proteins and rRNAs. The fact that fully active ribosomes can be reconstituted from the isolated components with the remarkably high efficiency of 50–100% of the input material substantiates this assumption.

The self-assembly character *in vitro* does not preclude the involvement of additional factors *in vivo* to facilitate and accelerate the whole process, for example, by reducing activation energies of distinct or otherwise rate-limiting reactions. One of these factors is probably the "assembly gradient" that marks the coupling of rRNA synthesis and ribosomal assembly in prokaryotes and eukaryotes, which means that the state of the rRNA synthesis dictates the progress of assembly [3–6]. It is further possible that proteins exist which maintain an unfolded state of the *de novo* synthesized ribosomal proteins thus favoring the integration into ribosomal particles ("chaperonins"). A recent report suggested that the Hsp70 DnaK system is involved in facilitating the ribosomal assembly by reducing the activation energy barrier [7]; however, these assertions could not be confirmed [8]. Circumstantial evidence for a corresponding activity for the rRNA ("helicases"), possibly facilitating the attainment of distinct RNA conformations that favor the assembly process, has been reported previously [9].

3.1.2 Processing of rRNAs

There are two main aspects to the processing of rRNAs: (1) trimming of the rRNAs to yield the mature molecules found in native, active ribosomes and (2) modification (mostly methylations and pseudouridylation) of the rRNAs.

The common order of rRNA genes in the seven rRNA operons in *E. coli* is 16S – internal transcribed spacer I (ITS-1) – (either $tRNA^{Ala}$ and $tRNA^{Ile}$ or $tRNA^{Glu}$) – ITS-2 – 23S – ITS-3 – 5S (see Fig. 3.1-1). The complete, intact transcript, the "30S precursor rRNA", is found at low levels in wild-type cells (1–2% of rRNA). Endonucleases are the primary processing enzymes, among which RNase III plays a major role in the maturation of 23S rRNA. RNase III cleaves within the spacer sequences bordering 16S and 23S rRNA. The spacer sequences can form impressive secondary structures flanking both 16S and 23S rRNA; however, these intramolecular interactions are not a prerequisite for RNase III activity, since RNase III can cleave at the 5′ end before the 3′ end of the same molecule is fully transcribed. A general feature of processing is that it begins before transcription of a ribosomal (rrn) operon is finished. This sequential processing in the 5′ → 3′ direction is compatible with the hypothesis that at least some processing steps are coupled with ribosomal assembly. The final maturation steps of pre-16S, pre-23S and pre-5S are performed by exonucleases (maturases, secondary processing enzymes) that are not yet characterized and are termed M16, M23, and M5, respectively (see Ref. [10] for review). The final processing steps in the 50S subunit occur even after ribosomes are formed probably during early steps of protein biosynthesis, whereas mature 16S rRNA is required to obtain functional competence (see below).

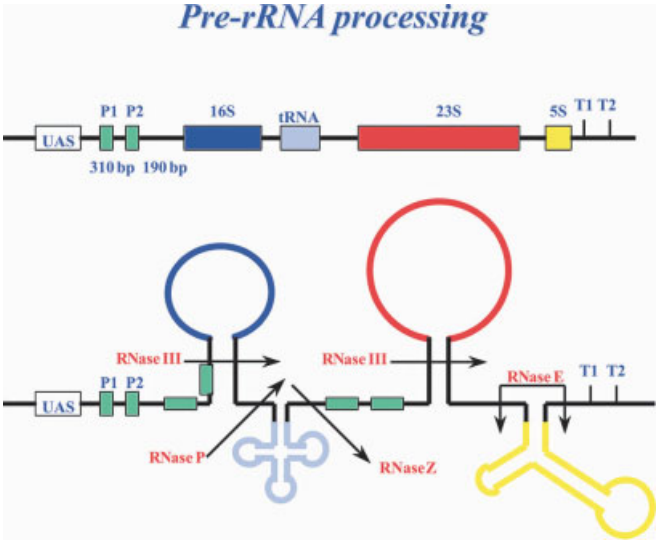


Figure 3.1-1 Operon structure of the rRNA genes in bacteria. UAS, upstream activating sequence, binding region of Fis, a DNA-binding protein that bends the DNA and increases rRNA transcription about 10-fold. P1 and P2, the two promoters, the first of which can efficiently be regulated by the stringent response (see Chap. 11); T1 and T2, terminators of transcription. Below the operon is shown with the symbolized secondary structure of the respective RNAs and the cleavage sites of some of the processing RNases.

RNase III cleavage yields precursor species of rRNA. The pre-16S species retains additional stretches of 115 and 33 nucleotides at the 5′- and 3′- ends, respectively, whereas the pre-23S has stretches of only 7 and 7–9 nucleotides, respectively [10, 11]. RNase III cleavages are not essential processing steps since mutants lacking RNase III are viable. In the absence of RNase III, 50S with pre-23S are found, whereas mature 16S rRNA molecules are formed at the same rate as in the wild-type strain. Interestingly, 30S subunits containing pre-16S, where the sequence flanking the mature 16S rRNA are base-paired and form a long helix, seem to be inactive (i.e., mutants deficient in M16 are not viable) in contrast with 50S subunits containing pre-23S. Note that in mature 30S subunits the 5′- and 3′- ends are far apart from each other, whereas in 50S the 5′- and 3′- ends are base-paired as seen clearly in the crystal structure of bacterial 30S and 50S subunits (Figs. 3.1-2 and B, respectively [12, 13]). Consequently, the maturation from pre-16S to mature 16S rRNA within 30S particles (removing the secondary structure flanking the 16S rRNA) triggers the activation of 30S subunits [14]. When the 16S rRNA processing is coupled with and depends on a correct 30S assembly, this final processing step guarantees that only active 30S subunits can initiate protein synthesis.

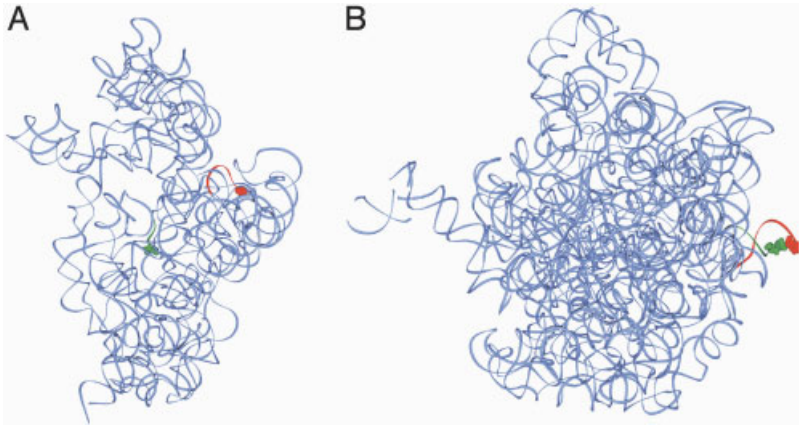


Figure 3.1-2 5′- and 3′-ends of 16S rRNA within the 30S subunit (A) and of 23S rRNA within the 50S subunit (B) viewed from the interface. (A) 30S subunit; *Thermus thermophilus* 30S (pdb 1fka): ribbon representation of rRNA with the 5′- and 3′-ends of the 16S rRNA colored green and red, respectively. The residues U6 (green) and C1512 (red) are in spacefill and are approximately 80 Å apart. (B) 50S subunit; *Deinococcus radiodurans* 50S (1kpj): ribbon representation of rRNA with the 5′- and 3′-ends of the 23S rRNA colored green and red, respectively. The residues G1 (green) and A2877 (red) are shown in spacefill and are in contact with one another.

Processing of 5S rRNA requires RNase E. RNase E-deficient mutants accumulate a 9S species that has not been detected in wild-type cells. RNase E forms pre-5S with three extra nucleotides at both its 5′- and 3′- ends. The final processing of 5S rRNA might also occur during protein synthesis or at least in active 70S ribosomes, since pre-5S was found in polysomes. *In vitro* reconstitution studies have revealed that 5S rRNA can be incorporated into the large subunit at any stage of the 50S assembly [15] reflecting its exposed location in the central protuberance of the 50S subunit.

In *E. coli*, 11 and 23 nucleosides are modified in 16S and 23S rRNA, respectively (Table 3.1-1; [16, 17]). The modifications of the 16S rRNA are late events during *in vivo* assembly and are not essential for assembly *per se*, whereas most of the 23S modifications, some of which are essential, occur early during assembly. Most of the modifications are base-methylations, and nine are pseudo-uridylations (Ψ) in the 23S rRNA. The methylations are not required for the trimming processes described above. In fact, a few methyl groups are found at the 2′-ribose position (e.g., Cm2498 of 23S rRNA, see Table 3.1-1) and might protect sensitive rRNA regions against RNase attack. The presence of a modified adenine at position 2071 (A* in Table 3.1-1) is uncertain. Most of the methyl groups are modifications of bases exposed at the ribosomal surface and are clustered at functionally active sites of the ribosome, e.g., 20 of the 23 modifications occur at or near the peptidyl transferase ring of domain V of the 23S rRNA, none of which are universally conserved.

Table 3.1-1 Modified nucleosides in E. coli rRNA. (see Ref. [44] for review and references)

16S rRNA	Location (nucleotide)	23S rRNA	Location (nucleotide)
Ψ	516	m^1G	745
m^7G	527	Ψ	746
m^2G	966	m^5U	747
m^5C	967	Ψ	955
m^2G	1207	m^6A	1618
m^4Cm	1402	m^2G	1835
m^5C	1407	Ψ	1911
m^3U	1498	$m^3Ψ$	1915
m^2G	1516	Ψ	1917
$m^6{}_2A$	1518	m^5U	1939
$m^6{}_2A$	1519	m^5C	1962
Total	11	m^6A	2030
		m^7G	2069
		A^*	2071
		Gm	2251
		m^2G	2445
		D	2449
		Ψ	2457
		Cm	2498
		m^2A	2503
		Ψ	2504
		Um	2552
		Ψ	2580
		Total	23

Therefore, they are probably more important for fine-tuning and stability of structural motifs rather than being directly involved in ribosome function. The two adjacent m^6_2A residues at positions A1518/A1519 in 16S rRNA are the only universally conserved modifications of rRNA. They improve the formation of initiation complexes, in particular the binding of IF-3, which has an anti-association activity during the initiation of translation (see Chap. 7.1). The absence of these four methyl groups confers resistance to the drug kasugamycin (see Chap. 12).

Recently, evidence was reported that occasionally defined rRNA fragments occur during the synthesis of 16S and 23S rRNA. These fragments are so efficiently degraded by polynucleotide phosphorylase and RNase R that they escaped the attention in former analyses of rRNA synthesis. Cells are not viable when both enzymes are inactivated. The fragments, if accumulated, might bind some ribosomal proteins, thus compromising the assembly of the mature rRNAs and eventually leading to cell death [18].

3.1.3 Precursor Particles and Reconstitution Intermediates

Usually, the assembly process is described from the point of view of the largest component, i.e., the 16S or 23S type of rRNA and the sequence of addition of these components is considered. This process requires a concatenation of reactions that differ in their respective activation energies. The highest activation energies function as energy barriers allowing precursor particles to accumulate.

In fact, precursor particles have been found *in vivo*. The assembly of the small subunit (30S, *E. coli*) passes through at least two different intermediate particles termed $p_1$30S and $p_2$30S (p for precursor). The $p_1$30S particle sediments with 21S. The $p_2$30S particle contains the full complement of S-proteins (S for proteins from the small subunit), but still an immature "17S" rRNA, which is longer at both its 5′- and 3′-ends with respect to the mature 16S rRNA (see the preceding section; [10]).

Only one reconstitution intermediate, RI_{30}, is found during the assembly *in vitro* of the 30S subunit [19]. The total reconstitution ("total" marks a reconstitution from completely separated ribosomal proteins and rRNA) is a one-step procedure according to the formula

$$16S\ rRNA + TP30 \xrightarrow{20\ mM\ Mg^{2+},\ 40\ C,\ 20\ min} 30S, \qquad (1)$$

where TP30 stands for total proteins derived from 30S subunits. An incubation of both 16S rRNA and TP30 at 0°C leads to the RI_{30} particle. This particle has to undergo a conformational change ("activation") according to the equation

$$RI_{30} \xrightarrow{40\ °C,\ 20\ min} RI_{30}^{*}, \qquad (2)$$

where RI_{30}^{*} is a particle with an unchanged composition but a tighter packing.

Only the RI_{30}^{*} particle can bind at 0°C, the lacking S-proteins to form an active 30S particle. Interestingly, the protein content of the RI_{30} particle is very similar to that of the $p_1$30S precursor. Equation (2) describes the rate-limiting step of the 30S assembly, the activation energy of which is 63 kcal mol^{-1} [19].

The assembly *in vivo* of the large subunit (50S, *E. coli*) occurs via three precursor particles $p_1$50S, $p_2$50S, and $p_3$50S sedimenting with 34S, 43S and "near 50S", respectively [20]. The final precursor ($p_3$50S) contains again a full complement of L-proteins (L for proteins derived from the large subunit). The $p_2$50S particles can be converted to active 50S subunits in the presence of TP50 under methylating conditions (S-adenosyl methionine, postribosomal supernatant), whereas the $p_1$50S particles could not, suggesting that this initial process might require additional processing steps other than simple methylations [21].

A two-step procedure is required for the total reconstitution of active 50S subunits [22]:

$$(23S+5S)\ rRNA+TP50 \xrightarrow[20\ mM\ Mg^{2+},\ 50°C,\ 90\ min]{4\ mM\ Mg^{2+},\ 44°C,\ 20\ min} 50S. \qquad (3)$$

The two-step procedure is a consequence of the fact that the rate-limiting steps of early and late assembly involve conformational changes that differ in their ionic optima *in vitro* ([23]; see also below). The two-step procedure is therefore a convenient way to separate early- and late-assembly events, and indeed allowed for a detailed analysis of the possible reconstitution intermediates.

Three reconstitution particles have been identified; protein analysis revealed that the first and the second particles contained the same complement of rRNAs and L-proteins in spite of the drastic difference in their respective *S* values (33S and 41S, respectively, see Table 3.1-2), whereas the third particle contained all the components of the active 50S subunit but was totally inactive. Accordingly, the three reconstitution intermediates were termed $RI_{50}(1)$, $RI^{*}_{50}(1)$ and $RI_{50}(2)$. It appears that the rate-limiting step of the first incubation is the conformational change $RI_{50}(1) \rightarrow RI^{*}_{50}(1)$, and that of the second incubation the conformational change $RI_{50}(2) \rightarrow 50S$ (Table 3.1-2). The corresponding activation energies have been determined as 45 and 55 kcal mol^{-1}, respectively. The precursor particles and the corresponding reconstitution intermediates have similar protein compositions as well as similar *S* values, indicating that assembly *in vivo* proceeds via rate-limiting steps that are very similar, if not identical, to the corresponding ones of the assembly *in vitro* [23].

3.1.4 Assembly-initiator Proteins

Ribosomal proteins (r-proteins) that bind *in vitro* specifically to naked rRNA are members of the "RNA-binding" family of proteins. About two-thirds of all r-proteins are RNA-binding proteins (see Fig. 3.1-3). The intriguing question was whether all these RNA-binding proteins, for example, about 20 L-proteins in the large subunit and 7 S-proteins in the small one, also bind directly to rRNAs *in vivo* without the help of other r-proteins (without cooperativity), i.e., whether these proteins are independent assembly-initiation events.

There were indications from *in vivo* studies that only a small number of r-proteins were able to initiate the assembly process. Under unfavorable growth conditions, when the doubling time for growth was about 10 h, i.e., 30 times longer than the optimal doubling time, the balanced synthesis of rRNA and r-proteins was lost and rRNA was produced in a three molar excess over r-proteins [24]. If, under such conditions, all 20 RNA-binding L-proteins could initiate the assembly process independently then they would be distributed evenly over the excess of rRNA, and therefore the yield of active particles with a full complement of r-proteins would be negligibly small. Since *E. coli* cells produce significant amounts of active ribosomes

Table 3.1-2 Sequential addition of proteins in the course of total reconstitution. Proteins in bold indicate proteins essential and sufficient for the RI* formation. For further details see Refs. [19] (30S subunit) and [23] (50S subunit).

Subunit	Step	rRNA (RI)		Proteins	Temperature (mM Mg^{2+})	Reconstitution intermediate (RI)	Sedimentation coefficient
Small subunit	1	16S	+	**S4**, S5, S6, **S7**, **S8**, S9, S11, S12, S13, S15, **S16**, S17, S18, **S19**, S20.	0°C →	RI_{30}	21-22S
	2	RI_{30}			37°C →	RI^{*}_{30}	25-26S
	3	RI^{*}_{30}		S1, S2, S3, S10, S14, S21.	0°C →	30S	
Large subunit	1	23S+5S	+	L1, L2, **L3**, **L4**, L5, L7/L12, L9, L10, L11, **L13**, L15, L17, L18, **L20**, L21, L22, L23, **L24**, L26, L29, L33, L34.	0°C (4) →	RI_{50}(I)	33S
	2	RI_{50}(I)			44°C (4) →	RI^{*}_{50}(1)	41S
	3	RI^{*}_{50}(I)	+	L6, L14, L16, L19, L25, L27, L28, L30, L31, L32.	44°C (4) / 50°C (20) →	RI_{50}(2)	48S
	4	RI_{50}(2)			50°C (20) →	50S	

even under unfavorable growth conditions, the number of assembly-initiator proteins must be significantly smaller than that of the total number of RNA-binding r-proteins.

An assembly-initiator protein is defined as an r-protein, which binds without cooperativity to an rRNA molecule and is essential for the formation of an active ribosomal subunit. Only those rRNA molecules with a complete set of initiator proteins are able to assemble correctly to form fully active ribosomal particles.

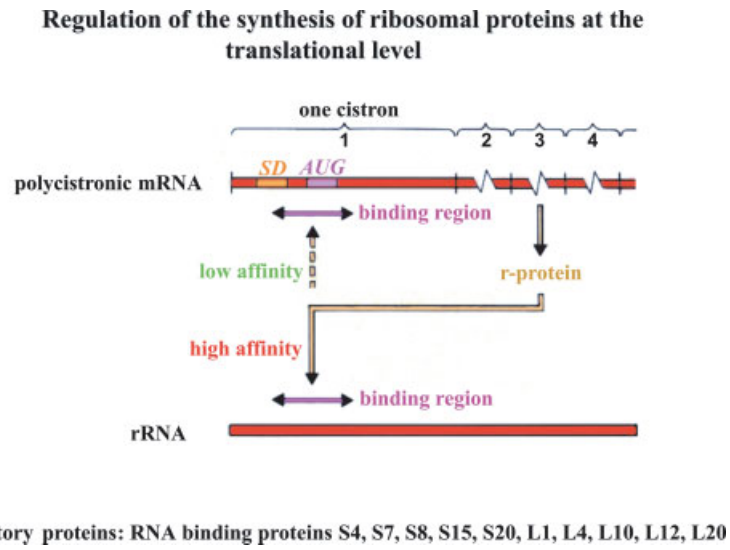


Figure 3.1-3 Translational regulation of the ribosomal proteins. Usually the second or third cistrons of a polycistronic mRNA code for an rRNA-binding protein, which can also bind to the region of the ribosomal binding site of the first cistron of its own mRNA, thus competing with initiating 30S subunits. Therefore, this regulatory protein will inhibit the translation of its own polycistronic mRNA, when a signi-ficant free pool of this protein is in the cell.

The dependence of the amount of active particles on the number of assembly-initiator proteins and the excess of rRNA is governed by the formula [25]

$$A = E^{1-n}, \tag{4}$$

where A is the fraction of total proteins that appears in active particles (A for activity), E the molar ratio of rRNA to r-proteins (E for rRNA excess), and n represents the number of assembly-initiator proteins. Let us assume that only one initiator protein is present. If the molar ratio rRNA : TP is E, then the probability of finding the initiator protein on one distinct rRNA molecule is E^{-1}, whereas for n initiator proteins the probability is E^{-n} (independent events). Since the probability is the same for each rRNA molecule, the overall probability of obtaining a complete initiation complex (i.e., a complex of rRNA and all n initiator proteins) is $E \times (E^{-n}) = E^{1-n}$. This is identical ($A$) with the fraction of TP, which appears in active particles, since only complete initiation complexes will form active particles. Hence, $A = E^{1-n}$ and

$$\ln A = (1-n)\ln E. \tag{5}$$

Equation (5) provides us with direct access to the experimental strategy for the elucidation of the number of initiator proteins, for example, one keeps the level of TP50

constant and increases the input *E* of (23S+5S) rRNA. The reconstitution is then performed and the yield of active particles, *A*, is determined. This kind of analysis revealed that only two L-proteins actually function as assembly-initiator proteins, and in an additional reconstitution analysis with purified proteins, L3 and L24 were identified as the two assembly-initiator proteins of the large subunit by applying a variation of this strategy [25]. Two proteins were also identified as assembly-initiator proteins for the small ribosomal subunit, namely the proteins S4 and S7 [26].

We understand from Eq. (4) that a high yield of active particles, under conditions where rRNAs are synthesized in excess over r-proteins, correlates with a small number of assembly-initiator proteins. Why then are there two initiator proteins rather than one?

As already mentioned, the mechanisms regulating the mutual adaptation of the synthesis of rRNA and that of r-proteins uncouple during extremely unfavorable growth conditions leading to an excess synthesis of rRNA. At this point, feedback mechanisms and autoregulatory circuits become increasingly important, such as the translational regulation of the r-proteins. The principle of the translational regulation is demonstrated in Fig. 3.1-3. Usually, the second or third cistrons of a polycistronic mRNA of r-proteins code for an RNA-binding protein that can also bind to the Shine–Dalgarno region of the first cistron from its own mRNA, thus competing with initiating 30S subunit and reducing the frequency of translation of the mRNA. Ten ribosomal proteins involved in translational regulation have been identified; these are S4, S7, S8, S15, S20, L1, L4, L10, L12 and L20 ([27], see also Chap. 11).

If only one assembly-initiator protein existed, then in the presence of excess rRNA all the r-proteins would flow into the formation of active particles leaving no free pool of r-proteins for regulatory tasks. It is the existence of two initiator proteins that is responsible for assembly-dead ends. These assembly-dead ends are loose protein–rRNA complexes from which r-proteins are provided for the translational control. Therefore, two initiator proteins seem to represent an optimum, the number must be small to allow the formation of significant amounts of active ribosomes in the presence of rRNA excess; on the other hand, the number must be larger than one to enable translational control under unfavorable growth conditions.

Protein L24 is essential only for the early-assembly event, being dispensable during later assembly events and not necessary for ribosomal function. The existence of a mutant lacking L24 is a surprise. The mutant strain produces assembly-defective 50S subunits in agreement with the findings described above. The mutation is conditionally lethal, exhibiting temperature sensitivity, i.e., it does not grow at temperatures above 36°C. Even at permissive temperatures, severe growth defects are observed (the doubling time is 5–7 times longer than that of wild-type *E. coli* strain), and the molar ratio of 30S : 50S ribosomal subunits is only 1 : 0.5. An analyses *in vitro* of the assembly of the mutant ribosomes demonstrated that L24 is absolutely required for the total reconstitution of active 50S subunits when the incubation temperature of the first step (normally 44°C, see Eq. (3)) was above 40°C. Below 40°C, active particles could be formed in the absence of L24; however, the maximal output of active 50S subunits was only 50% at permissive temperatures as compared with optimal conditions, i.e., half that in the presence of functional L24, and the activation

energy of the rate-limiting reaction during the first step incubation was twice as large as that in the presence of L24. Obviously, another protein replaces L24 at "permissive temperatures" (below 40°C) with reduced efficiency, and a systematic study with isolated L-proteins revealed that this protein is L20 [28].

3.1.5
Proteins Essential for the Early Assembly: The Assembly Gradient

The transition of the reconstitution intermediate $RI_{50}(1) \rightarrow RI^{*}_{50}(1)$ is marked by a drastic S-value shift from 33S to 41S (Table 3.1-2). $RI^{*}_{50}(1)$ is the essential product of the early assembly during the first-step incubation that cannot be formed during the second step. Both intermediate particles consist of 23S and 5S rRNA and 20 different proteins. Are all these multiple components needed for the critical transition from $RI_{50}(1) \rightarrow RI^{*}_{50}(1)$?

The results of a systematic analysis with purified ribosomal proteins were surprising. Only 23S rRNA and five proteins (L4, L13, L20, L22, and L24) were necessary and sufficient for establishing the functionally important $RI^{*}_{50}(1)$ conformation, although L3 stimulated the formation, the 5S rRNA and all the other ribosomal proteins were not required [6].

Comparison of the known binding sites of the early assembly proteins revealed that all these essential proteins have a binding site located towards the 5′-end of the 23S rRNA and only the stimulatory L3 binds near the 3′-end. Since all polymerase-dependent synthesis of nucleic acids starts at the 5′-end, this observation has an important consequence. Those proteins that determine the early-assembly reactions bind *in vivo* immediately after the onset of transcription of the 23S rRNA and before the completion of its synthesis. This coupling of rRNA synthesis and ribosomal assembly was termed the "assembly gradient" and states that the progress of rRNA synthesis dictates the progress of assembly [4]. Therefore, the entropic situation of *in vivo* assembly is much simpler than that of the total reconstitution *in vitro*: i*n vivo*, the assembly starts with a relatively short 5′-sequence of the rRNA and the five proteins essential for the early assembly, whereas *in vitro* the mature 23S rRNA is exposed to all 33 L-proteins. The entropic advantage of the initiation phase of assembly *in vivo* is seen in the following: the disorder of assembly initiation *in vivo* is drastically lower than under *in vitro* conditions, since *in vivo* the assembly commences *before* the transcription of the rRNA is finished and thus assembly initiation deals only with a limited number of components. These are the freshly transcribed 5′-region of the rRNA and the five proteins that can bind to the 5′-terminal sequence and determine the early assembly events. In striking contrast, when the mature 23S rRNA is present, all L-proteins compete for the assembly process *in vitro*, thus defining a much larger disorder under reconstitution conditions. This entropic advantage of the *in vivo* assembly over the *in vitro* reconstitution cannot be overestimated. It might at least partially explain why *in vivo* the 50S assembly of *E. coli* ribosomes lasts a couple of minutes at 37°C, whereas *in vitro* 90 min at 50°C is needed. *In vivo* studies have confirmed the existence of an "assembly gradient" [3].

The longer the rRNA, the more importance the assembly gradient assumes for an energetically favorable assembly process. The "assembly gradient" is presently the only explanation for the extremely complicated process of ribosomal assembly in eukaryotes. Ribosomal proteins are imported to the nucleoli, where rRNA transcription occurs (with the exception of the 5S rRNA). The near mature ribosomal subunits are exported from the nucleoli to the cytoplasm. It is assumed that this mechanism is required by the necessity to couple rRNA transcription and ribosomal assembly and thus to retain the entropic advantage of the assembly gradient. See Chap. 3.2 for details of the eukaryotic ribosome assembly.

The early-assembly proteins responsible for the $RI_{30} \rightarrow RI^{*}_{30}$ transition on the pathway to active 30S subunits were identified in the 1960s long before the 50S analysis described above [29]. The proteins essential for the transition were S4, S7, S8, S16 and S19, but the corresponding binding sites do not show such a clear preference for 16S rRNA regions near the 5′-end, as the early-assembly proteins do for the 23S rRNA. However, kinetic studies also revealed a sequential assembly with a 5′- to 3′-polarity along the 16S rRNA chain [5].

3.1.6
Late-assembly Components

One can define "late-assembly components" as those components of the 50S subunit that can be added to the second incubation step of the two-step incubation to yield active particles. In this classification, 5S rRNA and all L-proteins (except those five proteins essential for the formation of $RI^{*}_{50}(1)$ particles) would be included. However, here we define "late-assembly components" in the narrow sense as those components which play a decisive role in the late-assembly process, regardless of whether or not they are in addition important for stabilization of the structure and/or participate directly in ribosomal functions.

Until now, only two L-proteins have been identified, namely L15 and L16, whose main task is the organization of the late assembly step [30]. A mutant that lacks L15 exists (see also the next section), whereas a mutant without L16 has not yet been described. However, ribosomes lacking either L15 or L16 or both proteins can be reconstituted and are active in poly(U)-dependent poly(Phe) synthesis. Both proteins have been shown to accelerate the assembly process independently by a factor of 2–4; the proteins act synergistically since together they increase the assembly rate by a factor of 20. The stimulatory effects of both proteins are also observed when they were added after the two-step reconstitution during a short third incubation under the conditions of the second step. This means that particles could be reconstituted during the standard two-step procedure in the absence of L15 and L16, and the late addition of these proteins in a short third incubation was sufficient to form active particles. The latter feature underlines their involvement in the late assembly.

5S rRNA also fulfill's the criteria of a late-assembly component, viz., it can also be added after the two-step reconstitution to a third incubation. The activation effect of 5S rRNA is heat-dependent, i.e., it induces a conformational change. However, in

contrast with L15 and L16, fully active particles without 5S rRNA cannot be reconstituted. Therefore, a direct or indirect involvement of 5S rRNA in ribosomal functions is probable, in addition to its assembly activities [15].

3.1.7
Proteins Solely Involved in Assembly

A comparison of the secondary structures of rRNAs from organisms of the various kingdoms reveals that about two-thirds of the *E. coli* 16S and 23S rRNAs is universally conserved (Fig. 3.1-4). The regions of the remaining one-third are randomly scattered over the rRNAs and can be shorter, longer or even absent in other organisms. Similarly, one-third of the r-proteins are dispensable in *E. coli*, since mutants have been described lacking one or other of these r-proteins (Table 3.1-3). A recent

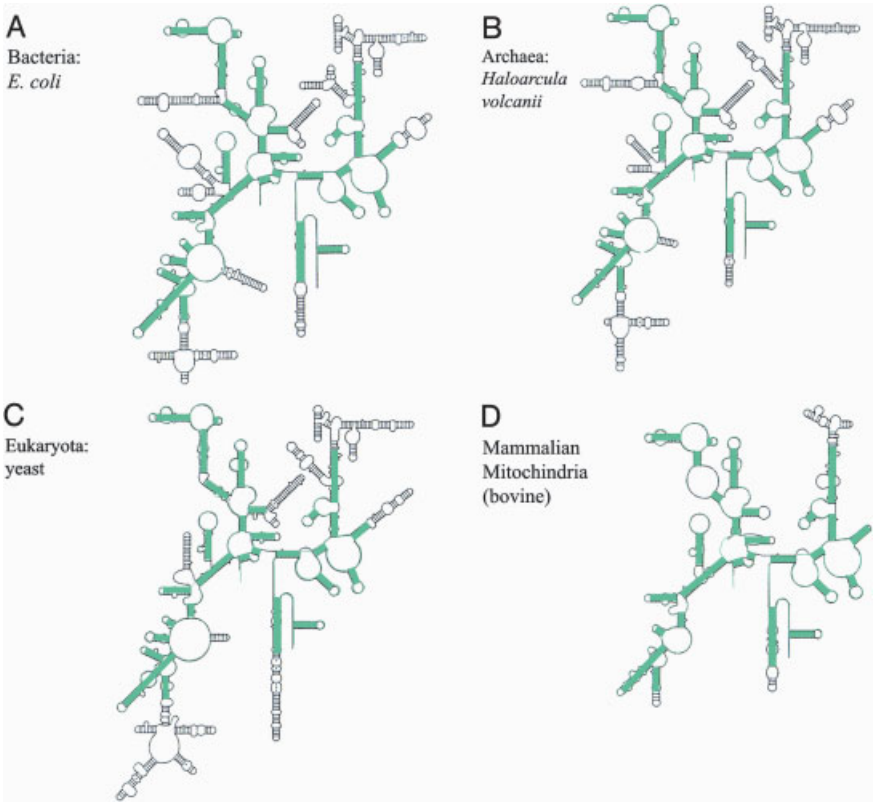


Figure 3.1-4 The core structure (green lines) common to all 16S-type rRNAs from ribosomes of various organisms: (A) *E. coli*; (B) *Halobacterium volcanii*; (C) yeast, cytoplasmic ribosomes; (D) mitochondria of plants (maize). According to Ref. [46], modified.

comparison of sequenced genomes from organisms of all three evolutionary domains, viz., bacteria, archaea and eukarya, found that about one-third of the *E. coli* proteins are universally conserved ([31]; Fig 3.1-5). This suggests that these rRNA regions and r-proteins are dispensable for ribosomal assembly, structure or function, although they might accelerate assembly, stabilize structures or fine-tune functions.

Proteins that accelerate assembly represent one class of proteins solely involved in assembly. S16 and L15 belong to this class. Fully active 30S subunits can be assembled in the absence of S16, but the protein accelerates the assembly [32]. L15 is absolutely required for the formation of active particles under standard reconstitution conditions, but this requirement is relieved after decreasing the NH_4Cl concentration from 400 to 240 mM [30]. Under these conditions, L15 accelerates the assembly process by a factor 2–4, which correlates well with the prolonged generation time (2–3 fold) of the mutant lacking L15. This observation suggests that the production of the large ribosomal subunits is the rate-limiting factor of the generation time of *E. coli* cells under optimal growth conditions.

A second class of "assembly-only proteins" consists of a group of proteins essential for achieving a distinct assembly stage, which is a necessary intermediate in the path towards an active subunit. If mutants lacking such a protein exist at all, they are very sick. L20, L24, and probably L16, belong to this class. L16 is an assembly protein; it accelerates the late assembly, and particles lacking L16 can be reconstituted and show a good activity in poly(Phe) synthesis, although reconstituted 50S subunit with a full complement of proteins are four times more active [30]. Crystal-structure analysis of 50S subunits suggests that L16 might help to position tRNAs at P-site (see Chap. 6). A mutant lacking L16 has not yet been identified, possibly because of

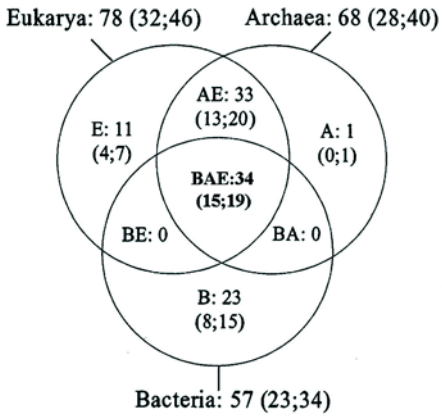


Figure 3.1-5 Conservation of ribosomal proteins from bacteria (B), eukarya (E) and archaea (A). 34 proteins are universally conserved, no proteins are common between bacteria and archaea or between bacteria and eukarya, whereas 33 proteins are present in archaea and eukarya. This is an impressive example that the common ancestor of archaea and eukarya separated from bacteria before the domain separation between eukarya and archaea. From Ref. [31] with permission.

Table 3.1-3 Mutants from *E. coli* lacking r-proteins. (taken from Ref. [45]).

Subunit	Missing protein	Phenotype
30S	S1	
	S6	
	S9	Cold-sensitive
	S13	
	S17	
	S20	Temperature-sensitive
50S	L1	
	L11	
	L15	Cold-sensitive
	L19	
	L24	Temperature-sensitive, Very slow growth
	L27	Cold-sensitive
	L28	Cold-sensitive
	L29	
	L30	
	L33	Cold-sensitive

this role in positioning tRNAs. A mutant lacking L20 is also yet to be described. As mentioned previously, the mutant lacking L24 is severely handicapped as expected.

Both L20 and L24 are essential for the formation of the obligatory, early intermediate $RI^{*}_{50}(1)$, but they are not involved in late assembly nor in ribosomal functions. How can one test that a protein has an essential role in the assembly but no role in function? The intriguing observation was that L20, L24, and other proteins (L4, L14, and L22) are essential for the formation of the $RI^{*}_{50}(1)$ conformation, but once this conformation has been achieved, at least for L20 and L24, they can be again removed by high-salt washes without losing the $RI^{*}_{50}(1)$ conformation. If the resulting core particle is reconstituted with TP50 lacking either L20 or L24, fully active particles are obtained. Therefore, both proteins are essential for the early assembly but play no role in either late assembly or ribosomal functions [33, 34].

3.1.8 Assembly Maps

In addition to the formal assembly pathway described in the preceding sections, the precise sequence of binding reactions starting with the 16S or 23S rRNA has been unravelled. The experimental results of such binding analyses are summarized in "assembly maps". Primary binding proteins can individually form a stable complex

with the rRNA and are connected to the rRNA by a thick arrow in Fig. 3.1-6. Secondary binding proteins require the help of other proteins. The assembly map reflects the interdependencies of the proteins for their incorporation into the ribosomal particle. Consistently, the sequence of stripping-off r-proteins with increasing LiCl concentrations is the exact reverse of the assembly order [35].

An interdependence of binding of two proteins does not necessarily reflect physical proximity. It is conceivable that a protein induces a conformational change of the ribonucleoprotein particle (RNP) upon integration, thus generating new binding sites for other components. Crystal structures of the small [36, 13] and the large subunits [37, 12] at atomic resolution have enabled a quantum leap in our understanding of the assembly process. We can infer the precise protein–rRNA interactions for each protein as demonstrated with the 30S subunit [38], and thus it is now clear that the primary binding proteins (those proteins in Fig. 3.1-6 directly connected with the rRNA) help to nucleate the folding of the rRNA domains. An assembly feature seen for the three major secondary structure domains of the 30S subunit – namely that each of these domains (Figure 3.1-7) represent an independent assembly and folding domain [39–42] is strictly true only for the 3′ major domain forming the head of the 30S subunit. The proteins involved form a separate S7-dependent branch of the 30S assembly map. With regard to the other domains, the above-mentioned study of protein–16S rRNA interactions revealed that most of the proteins contact more than one domain [38]. This is particularly true for the 5′ (30S body) and central domains (platform), which are tightly associated with each other. This observation can probably be extrapolated to the assembly of these domains and thus explains why the 30S subunit has only two assembly-initiator proteins [26], S4 initiating the assembly of the 5′- domain and the central domain (body and platform, respectively), and S7 the 3′ major domain (head).

Since assembly follows transcription of the rRNA from 5′- to the 3′- end, the assembly proceeds in three main steps, the 5′ domain forms the body of the 30S subunit, followed by the formation of the platform (central domain) and eventually by that of the head (3′ major domain). The 3′ minor domain is formed from the long helix (h44) that runs down and back up the interface of the 30S subunit and comprises elements of the decoding center (see Chap. 8.2), as well as the short h45 and the anti-Shine-Dalgarno stretch at the 3′- end of the 16S rRNA (see Chap. 7).

The assembly of the domains of the 30S subunit was the topic of an interesting experiment *in silico* [43]. The 30S crystal structure was simplified by considering each nucleotide of the 16S rRNA as a pseudo-atom P and each aminoacyl residue of the ribosomal proteins as a pseudo-atom C. Interactions between proteins and 16S rRNA were assumed when Ps and Cs were closer than 3 Å. The resulting interaction diagram (Fig. 3.1-8) shows again the exclusive interaction of the proteins from the S7-dependent assembly branch with the head domain. In the next step the secondary structure map of 16S rRNA was taken, i.e., the tertiary folding was removed and the secondary structure alone considered, the proteins were added according to the sequence of the assembly map and the 16S rRNA secondary structure folded following the derived contact points of Fig. 3.1-8. The result was that the domains obtained were

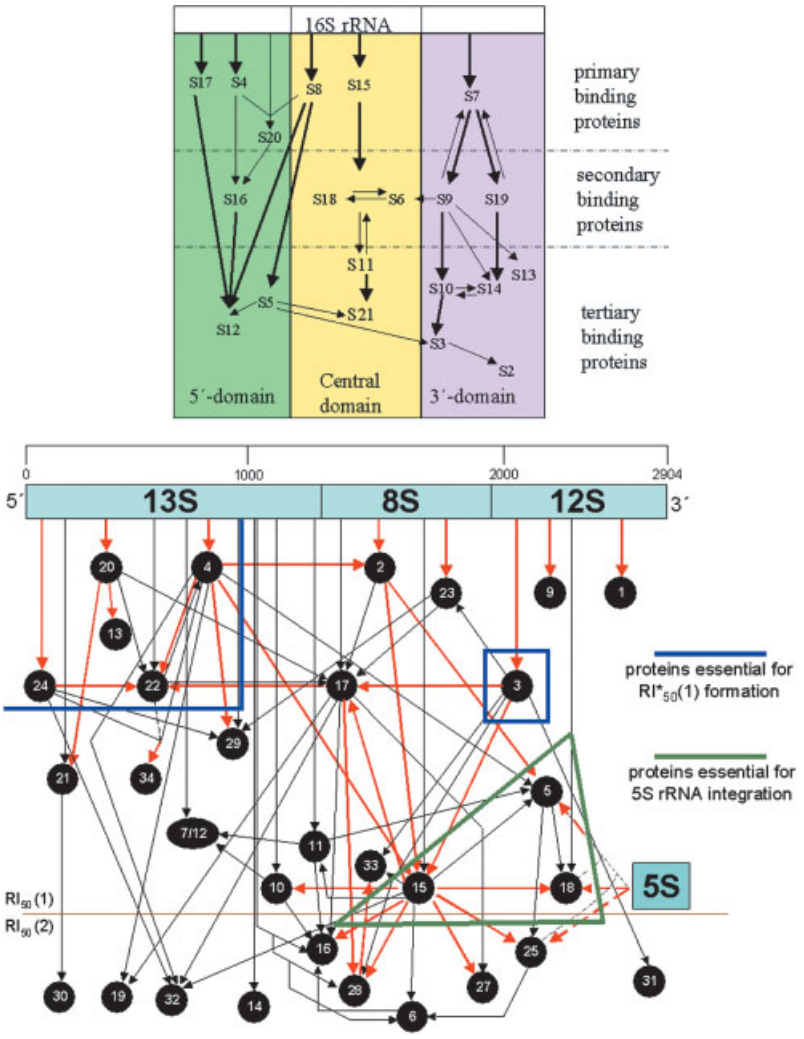


Figure 3.1-6 Assembly maps of the ribosome from *E. coli*. (A) The small 30S subunit ([47]; modified according to Ref. [48]). Assembly proteins in the green, yellow, and purple field drive the assembly of the 5′-domain, the central domain and the 3′-domain. Proteins above the dotted line are those either required for the formation of RI^{*}_{30} particles or found in the isolated 21 RI_{30} particles. (B) The large 50S subunit. The three main fragments of 23S rRNA (13S, 8S, and 12S) are indicated, and proteins are arranged according to their binding regions on 23S rRNA. The proteins boxed with a blue line are required for the transition $RI_{50}(1) \rightarrow RI_{50}(1)^{*}$ of the early assembly. Proteins above the brown line are those found in the $RI^{*}_{50}(1)$ particles. L5, L15 and L18, circled by the green line, are the proteins important for mediating the binding of 5S rRNA to 23S rRNA (according to Ref. [49], modified).

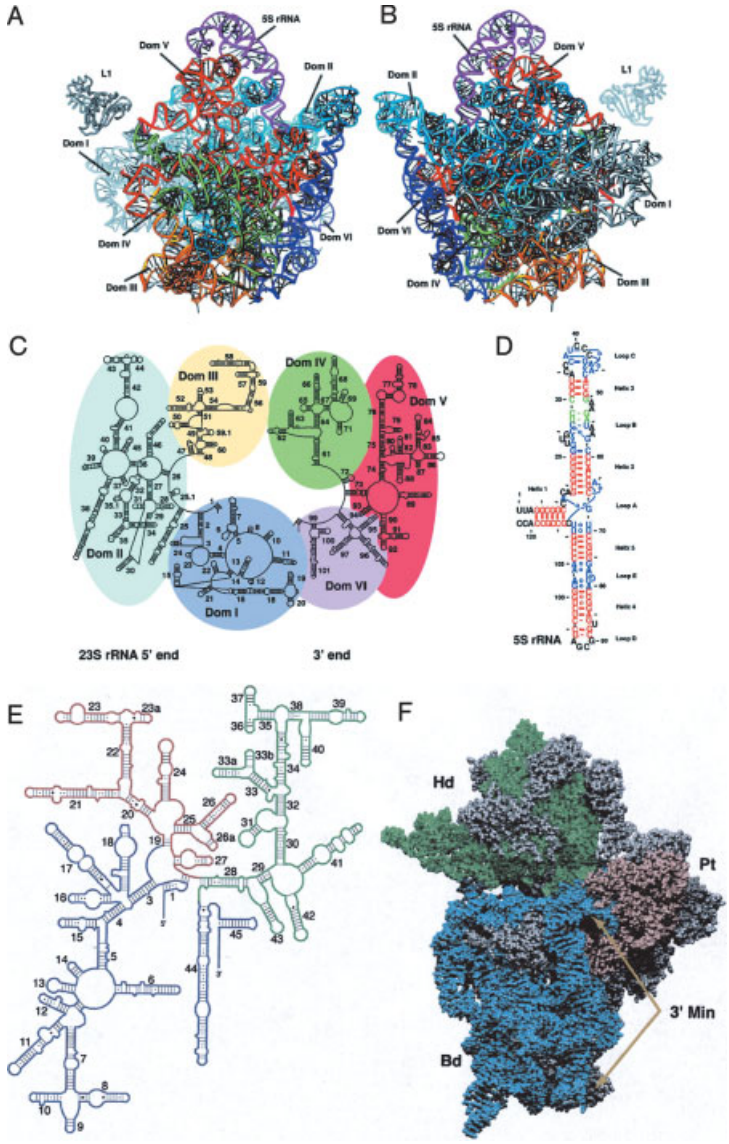


Figure 3.1-7 Ribosomal secondary and tertiary structure within the ribosome; the domains are marked with the same color in the secondary and the tertiary structures. (A) Structures of the 16S rRNA; the 5′, central and 3′ major domain are in blue, red and green; the 3′ minor domain in blue (3′ MIN); Hd, head; Bd, body; Pt, platform; the proteins are in gray. From Ref. [43], with permission.
(B) Structures of the 23S rRNA. Note that the colors of the domains are different in the maps of the secondary structure and in the tertiary structure. From Ref. [50] with permission.

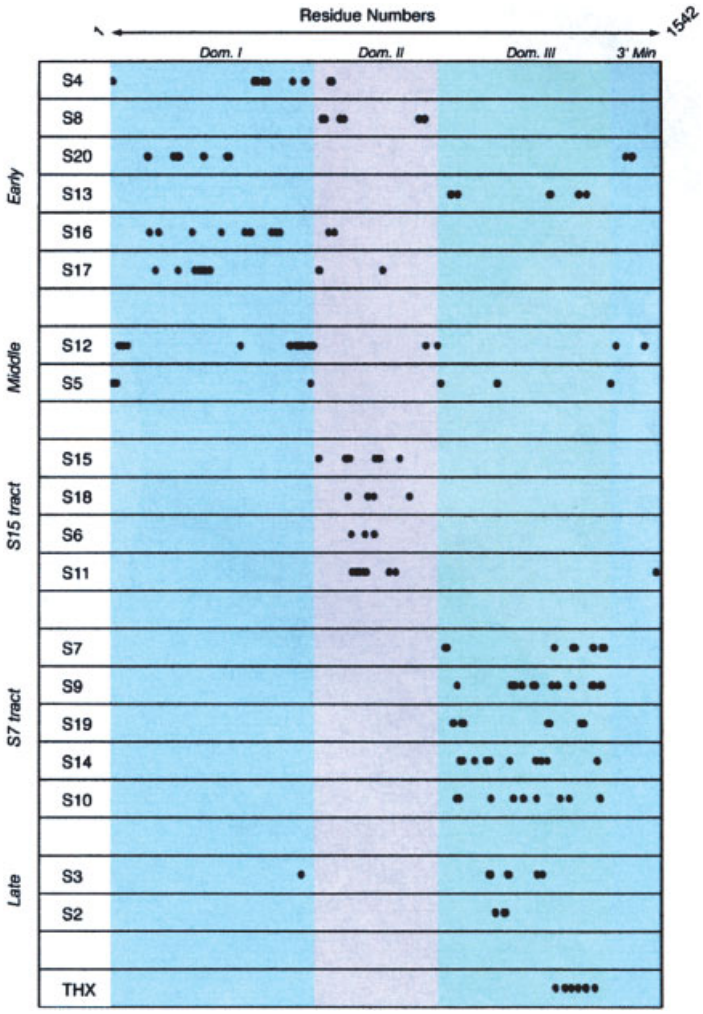


Figure 3.1-8 The proteins of the 30S subunit are listed according to their contacts with the 16S rRNA. From Ref. [43] with permission

strikingly similar to those of the crystal structure, whereas the interdomain arrangements deviated significantly. The conclusion was that the sequential addition of the proteins during the assembly gradient dictates the folding pathway for the 16S rRNA, whereas the mutual arrangement of the domains might be governed by rRNA–rRNA interactions. The implicit assumption of this study, namely that the structure of the ribosomal proteins before and after assembly is practically the same, seems to be justified (see Ref. [38]).

Despite our rapid increase in understanding of the assembly processes through the ribosome subunit crystal structures, we have but scant knowledge of the features responsible for the extremely efficient assembly of ribosomes *in vivo*. This is emphasized, for example, by the requirement of an incubation step at 50°C for 90 min during reconstitution of a 50S subunit *in vitro* [22] with the *in vivo* assembly of the large subunit in 20 min at 37°C. Although the assembly gradient as mentioned contributes to this enhancement, there are undoubtedly multiple helicases and chaperones that facilitate and accelerate the assembly process. The elucidation of the concerted interplay of elements supporting the assembly *in vivo* is a challenge for future assembly studies.

References

1 A. Tissières, J. Watson, D. Schlessinger et al., *J. Mol. Biol.* **1959,** *1,* 221–233.
2 G. Blobel, V. R. Potter, *J. Mol. Biol.* **1967,** *26,* 279–292.
3 B. T. U. Lewicki, T. Margus, J. Remme et al., *J. Mol. Biol.* **1993,** *231,* 581–593.
4 K. H. Nierhaus: in Analysis of the Assembly and Function of the 50S Subunit from Escherichia coli Ribosomes by Reconstitution, eds G. Chambliss, G. R. Craven, J. Davies et al., University Park Press Baltimore, 1980.
5 T. Powers, G. Daubresse, H. F. Noller, *J. Mol. Biol.* **1993,** *232,* 362–374.
6 S. Spillmann, F. Dohme, K. H. Nierhaus, *J. Mol. Biol.* **1977,** *115,* 513–523.
7 J. A. Maki, D. J. Schnobrich, G. M. Culver, *Mol. Cell* **2002,** *10,* 129–138.
8 J.-H. Alix, K. H. Nierhaus, *RNA* **2003,** *9,* 787–793.
9 K. Nishi, J. Schnier, *EMBO J.* **1986,** *5,* 1373–1376.
10 A. K. Srivastava, D. Schlessinger: in *rRNA Processing in Escherichia coli,* eds N. Hill, A. Dahlberg, R. Garrett et al., American Society for Microbiology, Washington, DC 1990.
11 D. Schlessinger, R. I. Bolla, R. Sirdeshmukh et al., *Bioessays* **1985,** *3,* 14–18.
12 J. Harms, F. Schluenzen, R. Zarivach et al., *Cell* **2001,** *107,* 679–688.
13 B. T. Wimberly, D. E. Brodersen, W. M. Clemons et al., *Nature* **2000,** *407,* 327–339.
14 J. W. Wireman, P. S. Sypherd, *Nature* **1974,** *247,* 552–554.
15 F. Dohme, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1976,** *73,* 2221–2225.
16 J. Ofengand, K. E. Rudd: in *Bacterial, Achaeal, and Organellar* rRNA: *Pseudouridines and Methylated Nucleosides and their Enzymes,* eds W. Hill, A. Dahlberg, R. Garrett et al., American Society for Microbiology, Washington, DC 1990.
17 J. Rozenski, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1999,** *27,* 196–197.
18 Z. F. Cheng, M. P. Deutscher, *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 6388–6393.

19 P. Traub, M. Nomura, *J. Mol. Biol.* **1969,** *40,* 391–413.
20 L. Lindahl, *J. Mol. Biol.* **1975,** *92,* 15–37.
21 K. Nierhaus, K. Bordasch, H. E. Homann, *J. Mol. Biol.* **1973,** *74,* 587–597.
22 K. H. Nierhaus, F. Dohme, *Proc. Natl. Acad. Sci. USA* **1974,** *71,* 4713–4717.
23 F. Dohme, K. H. Nierhaus, *J. Mol. Biol.* **1976,** *107,* 585–599.
24 K. Gausing, *J. Mol. Biol.* **1977,** *115,* 335–354.
25 V. Nowotny, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1982,** *79,* 7238–7242.
26 V. Nowotny, K. H. Nierhaus, *Biochemistry* **1988,** *27,* 7051–7055.
27 M. Nomura, R. Gourse, G. Baughman, *Ann. Rev. Biochem.* **1984,** *53,* 75–117.
28 F. J. Franceschi, K. H. Nierhaus, *Biochemistry* **1988,** *27,* 7056–7059.
29 P. Traub, M. Nomura, *J. Mol. Biol.* **1969,** *40,* 391–413.
30 F. Franceschi, K. H. Nierhaus, *J. Biol. Chem.,* **1990,** *265,* 16676-16682.
31 O. Lecompte, R. Ripp, J. C. Thierry et al., *Nucleic Acids Res.* **2002,** *30,* 5382–5390.
32 W. A. Held, M. Nomura, *J. Biol. Chem.* **1975,** *250,* 3179–3184.
33 V. Nowotny, K. H. Nierhaus, *J. Mol. Biol.* **1980,** *137,* 391–399.
34 S. Spillmann, K. H. Nierhaus, *J. Biol. Chem.* **1978,** *253,* 7047–7050.
35 H. E. Homann, K. H. Nierhaus, *Eur. J. Biochem.* **1971,** *20,* 249–257.
36 F. Schluenzen, A. Tocilj, R. Zarivach et al., *Cell* **2000,** *102,* 615–623.
37 N. Ban, B. Freeborn, P. Nissen et al., *Cell* **1998,** *93,* 1105–1115.
38 D. E. Brodersen, W. M. Clemons, Jr., A. P. Carter et al., *J. Mol. Biol.* **2002,** *316,* 725–768.
39 S. C. Agalarov, E. N. Zheleznyakova, O. M. Selivanova et al., *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 999–1003.
40 V. Mandiyan, S. J. Tumminia, J. S. Wall et al., *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 8174–8178.
41 R. R. Samaha, B. Obrien, T. W. Obrien et al., *Proc. Natl. Acad. Sci. USA* **1994,** *91,* 7884–7888.
42 C. J. Weitzmann, P. R. Cunningham, K. Nurse et al., *FASEB J.* **1993,** *7,* 177–180.
43 S. M. Stagg, J. A. Mears, S. C. Harvey, *J. Mol. Biol.* **2003,** *328,* 49–61.
44 G. R. Björk: in *Escherichia coli and Salmonella,* ed. F. C. Neidhardt, American Society for Microbiology, Washington DC 1996, 861–886.
45 E. R. Dabbs: in *Structure, Function and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer-Verlag, New York 1986, 733–748.
46 R. R. Gutell, B. Weiser, C. R. Woese et al., *Prog. Nucleic Acid Res. Mol. Biol.* 1985.
47 W. A. Held, B. Ballou, S. Mizushima et al., *J. Biol. Chem.* **1974,** *249,* 3103–3111.
48 M. I. Recht, J. R. Williamson, *Cold Spring Harb. Symp. Quant. Biol.* **2001,** *66,* 591–598.
49 M. Herold, K. H. Nierhaus, *J. Biol. Chem.* **1987,** *262,* 8826–8833.
50 N. Ban, P. Nissen, J. Hansen et al., *Science* **2000,** *289,* 905–920.

3.2
Eukaryotic Ribosome Synthesis

Denis L. J. Lafontaine

3.2.1
Introduction

Recent proteomic developments are providing the first eukaryotic ribosomal assembly maps. In these, pre-ribosomal assembly appears to be asymmetric, at least biphasic, with the small ribosomal subunit synthesis factors binding first to the pre-rRNAs to be replaced, after the first few pre-rRNA cleavages, by proteins involved in the synthesis of the large ribosomal subunit. Pre-rRNA processing is fairly well characterized with several key-processing enzymes remaining to be identified, including most endoribonucleases. rRNA modification is also well understood and relies extensively on small nucleolar RNAs (snoRNAs) for the recognition of the sites of modification. Nucleolar routing of box C+D snoRNAs (required for sugar 2′–O methylation) involves transit through a specific nuclear locale, the human coiled/cajal body (CB) and yeast nucleolar body (NB); these are conserved sites of small ribonucleoprotein particles (RNP) biogenesis. The first proteins involved in ribosome export are being identified; however, most of these are also required for pre-rRNA processing, and presumably pre-rRNP assembly. Their precise function in RNP transport therefore awaits these effects to be uncoupled. Key factors active in ribosome synthesis are also required for the processing of many other classes of cellular RNAs, suggesting that maturation factors are recruited from a 'common pool' of proteins to specific pathways. Much remains to be done to understand how rRNP processing, modification, assembly and transport are integrated with respect to ribosome synthesis and other cellular biosynthetic pathways.

3.1.1
Prelude

Ribosome synthesis starts in the nucleolus, the site of rDNA transcription. rRNA synthesis occurs at the interface between the fibrillar center(s) (FCs) and the dense fibrillar component (DFC) with the nascent transcripts reaching out in the body of the DFC ([128]; reviewed in Ref. [104]). A dedicated polymerase, RNA Pol I (Pol I), drives the transcription of a large precursor encoding three of the four mature ribosomal RNAs (rRNAs). The fourth rRNA, 5S, is produced from a Pol III promoter. The Pol I primary transcript is modified (specific residues are selected for ribose or base modification and pseudouridines formation), processed (mature sequences are released from the precursors and the non-coding sequences discarded) and assembled with the ribosomal proteins (RPs) in pre-ribosomes (reviewed in Refs. [130, 224, 298, 299, 311, 325]). As these processes occur, the granular component (GC) of the nucleolus emerges. FC, DFC, and GC are morphologically distinct compartments present in most eukaryotes; interestingly, although controversial, recent analysis

indicate that the yeast *Saccharomyces cerevisiae* has no FC (D.L.J. Lafontaine and M. Thiry, unpublished results). The relationship between the subnucleolar structures and the different steps of ribosome synthesis is not clear at present.

The nucleolus is a highly dynamic structure, and RNA and protein components are known to exchange with the surrounding nucleoplasm with high kinetics [40, 211]. The average nucleolar residency time for human *fibrillarin* was estimated to be of only ~40 s, indicating that the remarkably stable organization of the nucleolus may in fact reflect the extremely rapid dynamic equilibrium of its constituents. It is presently unclear whether there are resident, structural, nucleolar proteins or whether the structure simply 'holds together' through multiple, weak, interactions occurring between the nascent pre-rRNAs and the numerous trans-acting factors recruited to the sites of transcription [173]. The recent proteomic characterization of this cellular compartment will probably help to address these issues ([8, 236]; reviewed in Ref. [61]).

Pre-ribosomes are released from the nucleolar structure, reach the nuclear pore complexes (NPC), presumably by diffusion, and are translocated to the cytoplasm. Both the small (40S) and large (60S) ribosomal subunits undergo final cytoplasmic maturation steps. A large number of trans-acting factors follow the pre-ribosomes to the cytoplasm and are recycled to the nucleus. Recent data suggest that the final steps of maturation may be coupled to cytoplasmic translation [240, 286].

RP genes, most often intron-containing, duplicated and expressed at distinct levels (yeast), are transcribed by Pol II. RP pre-mRNAs follow a canonical Pol II synthesis pathway (including capping, splicing, poly-adenylation, etc.; reviewed in Ref. [219]) and are routed to the cytoplasm to be translated. RPs are addressed to the nucleus and the nucleolus. Nuclear targeting operates on the NLS mode (reviewed in Refs. [163, 322, 323]); redundant importins are involved [111, 230]. Nucleolar targeting is less- well defined.

Ribosome synthesis is an extremely demanding process requiring both tremendous amounts of energy and high levels of co-regulation and integration with other cellular pathways (reviewed in Refs. [150, 214, 309]). The production of the resident ribosomal components (4 rRNAs and about 80 RPs), as well as several hundreds of RNAs and protein trans-acting factors (see below) depends on the concerted action of the three RNA polymerases, extensive RNA processing and modification reactions, RNP assembly and transport and the function of several RNPs, including the ribosome itself. With about 2000 ribosomes to be produced per minute in an actively dividing yeast cell, transcription of pre-rRNAs and RP pre-mRNAs alone represent not less than 60 and 40% of the Pol I and Pol II cellular transcription, respectively. With about 150 pores per nucleus, each pore must import close to 1000 RPs and export close to 25 ribosomal subunits per minute.

The nucleolus does not only serve the purpose of 'making of a ribosome'. In fact, it appears that most classes of cellular RNAs, including mRNAs [117, 239], tRNAs [21], snRNAs [81, 87], the SRP [42, 93, 110], RNAse P [113] and the TEL RNP [64, 192, 254] all transit through this organelle on their way to their final destinations, which can either be the nucleoplasm or the cytoplasm. Although the reason for this

trafficking is in most cases unclear at present, this presumably reflects a need to benefit from the pre-ribosomes maturation machinery. In the following, I will try to emphasize instances where common trans-acting factors are used on distinct classes of RNAs. The concept of a 'multifunctional nucleolus' has recently been elegantly reviewed [206].

Most of our current understanding of ribosome synthesis is based on research work in *S. cerevisiae*; this will be reviewed here. Other eukaryotic systems have been used successfully, including trypanosomes, *Xenope*, mouse and humans. Comparison between these various levels of organization is most useful and often highlights a high degree of conservation throughout the eukaryotic kingdom, e.g., most trans-acting factors identified in yeast have human counterparts.

This chapter will present an overview of eukaryotic ribosome synthesis for the non-specialists, with an emphasis on the latest developments and unresolved issues.

3.2.2 Why so many RRPs?

An excess of 200 proteins, here referred to as RRPs (ribosomal RNA processing factors) are required for ribosome synthesis and transiently associate with the pre-ribosomes. RRPs are not found in mature, cytoplasmic, particles but are recruited at various stages in the ribosomal assembly process. Recruitment presumably follows a strict temporal order. A similar number of small, stable RNAs, which localize at steady state in the nucleolus, the snoRNAs, are also involved.

Most RRPs have no known function and, in fact, apart from those few with catalytic activities or well-characterized protein domains, we clearly have no idea of what they do. Best-characterized RRPs include several endo- and exoribonucleases (Table 3.2-1), snoRNA-associated proteins, modification enzymes (ribose and base methyl-transferases, pseudouridine synthase), RNA helicases [47, 262], GTPases [86, 240, 317], AAA-ATPases [14, 77], protein kinases [295, 296], RNA binding or protein–protein interaction domain-containing proteins and proteins with striking

Table 3.2-1 Endo- and exoribonucleolytic activities involved in pre-rRNA processing.

Cleavage site	Cleavage activity	References
B_0	Rnt1p/yeast Rnase III	136
$B_0 \rightarrow B_2$	Rex1p	187
A_0, A_1, A_2	?, ?, ?	
A_3	MRP	159
$A_3 \rightarrow B_{1S}$	Rat1p, Xrn1p	98
B_{1L}	?	
C_2	?	
$C_2 \rightarrow C_{1'}$ and $C_{1'} \rightarrow C_1$	Rat1p, Xrn1p	85
$C_2 \rightarrow E$	Exosome Rex1p, Rex2p ?Ngl2p	176 287 65

homology to RPs [14, 59, 79, 234]. Protein–protein interaction domain include coil–coil domains, WD and HEAT repeats, crooked-neck-like tetratrico peptide repeat, etc.; distinctive RRP motifs include the Brix, GAR, G-patch, and KKD/E-domains [9, 62, 67, 83, 94, 318]. The actual catalytic activity of most RRPs remains to be demonstrated experimentally and the precise substrate of these proteins is, in most cases, not known.

Comprehensive lists of RRPs have recently been compiled by various authors with a short description of protein domains and known or presumed functions; these are freely available on-line (see useful WWW links at the end of this chapter).

3.2.3 (Pre-)ribosome Assembly, the Proteomic Era

In the early 1970s, the joint efforts of the Warner and Planta Labs defined the basics of eukaryotic ribosome assembly; this remained the core of our understanding for the next 30 years [133, 277, 282, 297, 308, 312, 313]. Metabolic labeling experiments and sucrose-gradient analysis revealed that following transcription, an early 90S pre-ribosome is formed and subsequently partitioned into a 43S and a 66S particle, precursors to the 40S and 60S subunits, respectively (see Fig. 3.2-1). The RNA content of these particles was established as 35S, 27S, and 20S pre-rRNAs for the 90S, 66S, and 43S particles, respectively. The conversion of the 43S particles to 40S subunits is closely linked to small subunit export. Few RRPs were known at that time and the protein composition of these RNP complexes was not determined.

In the absence of appropriate tools, most of the research focused on other aspects of ribosome synthesis with most of the progress being made on pre-rRNA processing and modification (see below).

There was no reason to believe *a priori* that there would be a strong bias for the association of RRPs involved in small subunit synthesis with the primary transcript. In fact, since many mutations affecting primarily 25S rRNA synthesis have negative feedback effects on early pre-rRNA cleavages (see Sect. 3.2.4 and Ref. [299]), as part of what we think is a 'quality control' mechanism (see below), the suggestion was made that early and late RRPs interact functionally; such interactions could have occurred in a single, large, 'processome'. Functional interactions between early and late RRPs are most probably prevalent but the simple view of a unique 'processome' has however been recently challenged.

The advent of efficient copurification schemes and mass-spectrometry analysis [162, 228] led several labs to isolate distinct pre-ribosomal species (currently about 12, see Table 3.2-2). Typically, these were purified from one or several epitope-tagged protein components of the rRNPs ([14, 56, 67, 91, 95, 195, 234, 318]; reviewed in Refs. [71, 310]). These preparations have achieved a much better definition in their pre-rRNA content (which parallels our current understanding of pre-rRNA processing, see Figs. 3.2-1 and 3.2-3) and the protein composition of the particles has been established accurately. In combination with high-throughput copurification and two-hybrid schemes ([74, 75, 84, 101, 244] and useful WWW links), these data

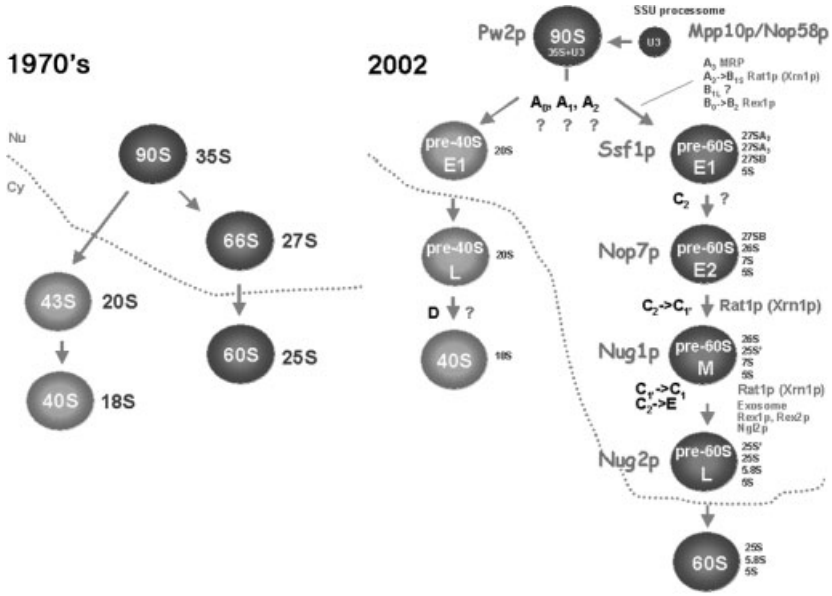


Figure 3.2-1 Ribosomal assembly pathways. See main text for a full description. Cleavage sites, processing activities and RNA content of pre-ribosomal particles are indicated, as well as the TAP-targets used for the purifications (see Table 3.2-2). Pre-ribosomes have tentatively been ordered, based on their protein and RNA content, and assigned and to the early (E), middle (M), or late (L) class. At the time of writing (Christmas 2002), several novel pre-ribosomal species are being isolated (in particular in the 40S subunit branch) and the pathway is expected to be much refined in the next few months. Largely inspired by Fatica and Tollervey [71]). Nu, nucleus; Cy, cytoplasm.

Table 3.2-2 TAP-tagged purified pre-ribosomes, as of Christmas 2002.

Pre-ribosomes	TAP targets	References
90S and U3/SSU processome	Pw2p, Rrp9p, Nop58p, YDR449c, Krr1p, Noc4p, Kre31p, Bud21p, YHR196w, YGR090w, Enp1p, YJL109c, Nop14p	91
U3/SSU processome	Mpp10p and Nop58p	56
Pre-60S E1	Ssf1p	67
Pre-60S E2	Nop7p	95
Pre-60S M	Nug1p	14
Pre-60S L	Nug2p/Nog2p	234
Seven species of early (E), medium (M) and late (L) pre-60S	Nsa3p, Nop7p, Sda1p, Rix1p, Arx1p, Kre35p, Nug1p	195

The TAP technology (Tandem Affinity Purification, 228) has been used to isolate most pre-ribosomes described to date.

provide the basis to draw the first eukaryotic (pre-)ribosomal assembly maps (see Figs. 3.2-1 and 3.2-2).

It transpires that ribosomal assembly is strongly asymmetric and at least biphasic ([56, 91]; reviewed in Ref. [71]). Early RRPs interact with nascent pre-rRNAs, mostly in association with the U3 snoRNP, now also referred to as the small subunit processome ('SSU processome'; [56], see below). Following the first three pre-rRNA cleavages, at sites A_0, A_1, and A_2 (see Figs. 3.2-1–3.2-3 and pre-rRNA processing section), this first set of factors essentially cycles-off the pre-ribosomes and is replaced by the large ribosomal subunit RRPs (Fig. 3.2-2). Pre-40S subunits are then left associated with very few factors, about a dozen of them, referred to as the SSU RRP complex [235, 335]; pre-60S subunits acquire several dozens of novel RRPs. As anticipated, there is a steady decrease in the number of these pre-60S-associated RRPs as the pre-ribosomes undergo the complex 5.8S–25S pre-rRNA processing pathway and reach the NPC. 90S and 66S particles were long known to have a higher ratio of protein to RNA content than the mature 60S subunits, as judged from buoyant densities in CsCl gradients (see, e.g., Ref. [277]). This is in contrast with 43S pre-ribosomes and 40S subunits that have about the same protein content. Late nuclear pre-60S ribosomes show the reassuring presence of known transport factors, such as the well-characterized Nmd3p/Rpl10p couple (see Sect. 3.2.7).

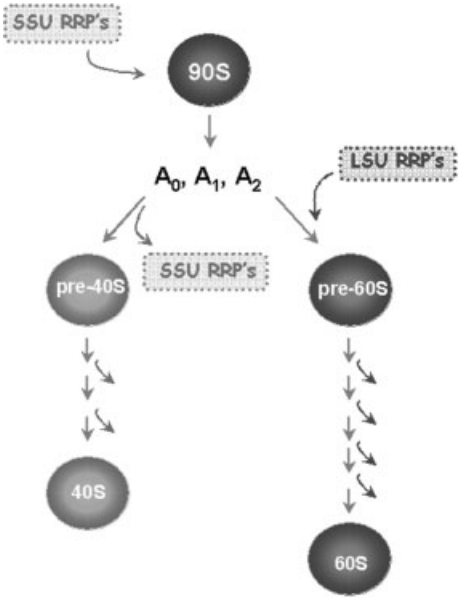


Figure 3.2-2 The 'biphasic model' for ribosomal assembly. The SSU RRPs (including the U3 snoRNP/SSU processome') associate with the primary Pol I transcript, generating the 90S pre-ribosomes. This first set of RRPs is replaced after the first three pre-rRNA processing reactions (A_0–A_2) by the LSU RRPs.

Most striking from the currently described pre-rRNPs is the conspicuous absence of most known cleavage enzymes; this could either reflect the low abundance or transient associations of these activities with the rRNP complexes.

X-ray crystallographic analysis of large ribosomal subunits revealed that while many RPs are located on the exterior of the rRNA core, several RPs show idiosyncratic extensions deeply buried into the body of the subunits in a configuration that is only compatible with concomitant foldings of the RPs and the rRNAs ([13]; reviewed in Refs. [55, 143, 220, 222]). This presumably underlies the need for close to 30 distinct remodeling activities (helicases, GTPases, and AAA-ATPases). It is remarkable that several RRPs are strikingly homologous to RPs (e.g., Imp3p/Rps9p, Rlp7p(Rix9p)/Rpl7p, Rlp24p/Rpl24p, Yh052p/Rpl1p [14, 59, 79, 234]), suggesting that they possibly 'hold in place' pre-rRNP structures during the assembly process, preventing premature, irreversible, folding steps to occur before being swapped for their homologous RPs. This strategy may even couple late pre-rRNA processing reactions to translation as eIF3j/Hcr1p is required for both 20S pre-rRNA processing and translation initiation and the RRP Efl1p is homologous to the ribosomal translocase EF-2 [240, 286].

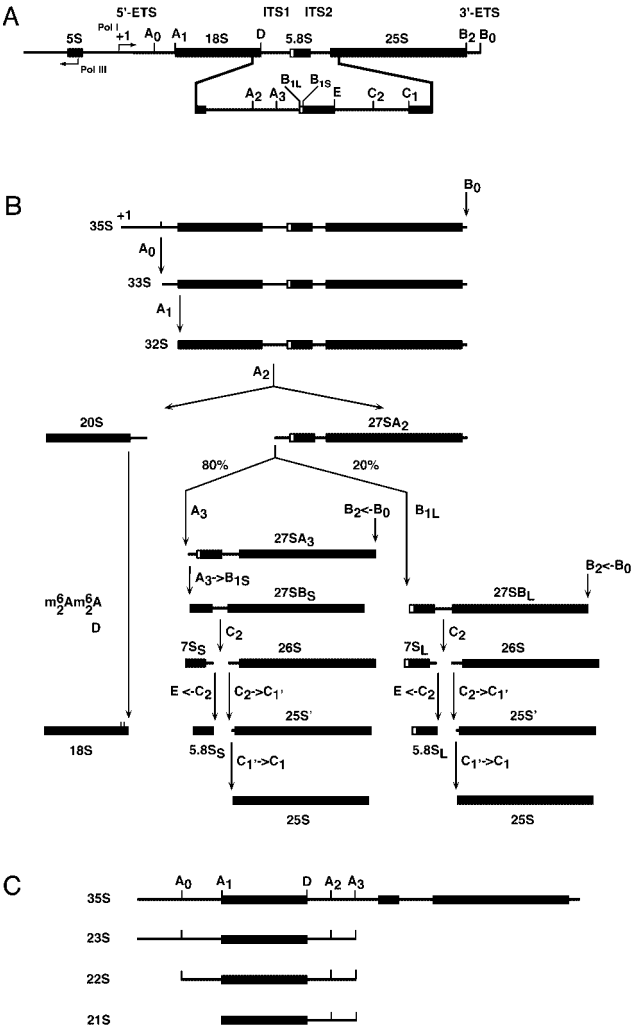
Pre-rRNP particles currently described were isolated from tagged RRPs and although clearly distinct in composition, as expected from the substantial remodeling of the pre-ribosomes that take place along the pathway, represent mixed pre-rRNP populations. It is also important to note that it is in fact mostly pre-ribosomal assembly rather than ribosomal assembly *per se* that has been addressed so far. Indeed, RPs present a particular challenge; there are usually small, highly basic and coprecipitate at high degrees with targets that are not related to ribosome synthesis. Despite these limitations, a major step has however been achieved with the isolation of particles which have a lifetime of presumably less than a minute.

Much remains to be done to understand what the RRPs exactly do, how and when they interact with the pre-ribosomes and how they 'talk' to each other.

3.2.4
Ribosomal RNA Processing, Getting there...

Pol I transcription drives the synthesis of a large pre-rRNA, 35S in yeast, containing the mature sequences for the small subunit rRNA (the 18S rRNA) and two of the large subunit rRNAs (5.8S and 25S rRNAs). Completion of transcription requires about 5 min. Mature sequences are flanked with non-coding spacers (Fig. 3.2-3A).

Figure 3.2-3 rDNA and pre-rRNA processing pathway. (A) Structure of the yeast rDNA. The 18S, 5.8S and 25S rRNAs are encoded in a single, large, Pol I transcript (35S). Mature sequences are separated by non-coding spacers, the 5′- and 3′-external transcribed spacers (ETS) and the internal transcribed spacers 1 and 2 (ITS). Processing sites (A_0 to E) are indicated. 5S is transcribed independently, in the opposite direction, by Pol III as 3′-extended precursors. The production of 5S mature 3′-ends is a multi-step process that requires Rex1p. (B) Pre-rRNA processing pathway in wildtype strains. See main text for a description of our current understanding of the processing pathway. Processing at sites $C_2 \rightarrow E$ is detailed in Fig. 3.2-4.



Note that C_2 (referred to as $C_{2'}$ in Ref. [85]) was recently mapped precisely by primer extension at a position located 94 nucleotides upstream of site C_1. Previous mapping, by RNase protection, located this site slightly upstream (at position +101 with respect to C_1). Although both sites may be used *in vivo*, it is more probable that this difference reflects limitations inherent to the RNase mapping strategy used. In Ref. [85], the $C_{2'}$-B_2 RNA is referred to as 25.5S (C_2-B_2 is 26S here).

Note that a cryptic processing site (A_4) has recently been identified in ITS1 between A_2 and A_3 in rrp5 mutants [63]. (C) Aberrant RNA precursors commonly detected in RRP mutants. Delays in early pre-rRNA processing often results in the accumulation of the 23S, 22S or 21S RNA. These are generally not further matured.

A Pol III precursor, 7S, is processed in 3′ by the Rex1p/Rna82p exoribonuclease into 5S; the third large ribosomal subunit rRNA [213, 287]. In most eukaryotes but *S. cerevisiae*, 5S rDNA is located in extranucleolar loci as individual or repeated copies. In yeast, 35S and 7S are encoded on opposite strands within 150–200 repeated 9 kb rDNA arrays located on chromosome XII (Fig. 3.2-3A).

Mature sequences are generated from the 35S pre-rRNA following a complex multi-step processing pathway requiring both endo- and exoribonucleolytic digestions (Fig. 3.2-3B). It is thought that most cleavage sites are known and occur within about 2 min following a precise temporal order. There is a strong bias for cleavages from the 5′- to the 3′-end of the primary transcript and the synthesis of the small and large ribosomal subunits is relatively independent.

The 35S pre-rRNA is successively cleaved in the 5′ external-transcribed spacer (5′-ETS) at sites A_0 and A_1 and in the internal-transcribed spacer 1 (ITS1) at site A_2 (Fig. 3.2-3B). Endonucleolytic digestions at sites A_0 and A_1 produce the 33S and 32S pre-rRNAs, respectively. Precursors to the small and large subunit rRNAs (the 20S and $27SA_2$ pre-rRNAs, respectively) are generated by endonucleolytic cleavage of the 32S pre-rRNA at site A_2. The precise mechanism of cleavage at sites A_0–A_2 is not known; however, these reactions are tightly coupled and involve the box C+D snoRNA U3/the 'SSU processome' (see below). The 20S pre-rRNA is then exported to the cytoplasm where endonucleolytic digestion, by an unknown RRP, at site D provides the 18S rRNA [276, 282]. A complex of late small subunit RRPs has recently been described in association with the dimethyl-transferase Dim1p ([295] and see below); the endonucleolytic activity may lie in one of these.

The $27SA_2$ pre-rRNA is cleaved at site A_3 to generate the $27SA_3$ RNA. This cleavage is carried out by the endoribonucleolytic RNP complex RNase MRP. RNase MRP is highly reminiscent to another snoRNP, the ubiquitous RNase P that is involved in the 5′-end formation of tRNAs (reviewed in Refs. [183, 329]). The homology extends both to the structure of their respective RNA as well as to their protein composition (eight of the nine protein subunits are shared between the two enzymes). Snm1p is specific to RNAse MRP; Rpr2p is unique to RNase P [35, 238].

In the absence of cleavage at site A_2, pre-rRNA processing can proceed through the next ITS1 cleavage at site A_3. This can be seen as a 'rescue' pathway for such an essential activity as ribosome synthesis [183].

There are two alternative pathways of synthesis of 5.8S–25S rRNAs [98]. In the major pathway, which represents ~80% of the total processing, the $27SA_3$ pre-rRNA is trimmed to site B_{1S} (the 5′-end of the most abundant form of 5.8S, the $5.8S_S$ rRNA) by the combined action of two 5′–3′ exoribonuclases, Rat1p and Xrn1p. Rat1p is encoded by an essential gene and mostly located to the nucleus; Xrn1p is not essential and mostly localizes to the cytoplasm [114]. These two exoribonucleases often show partially overlapping functions (see, e.g., Refs. [65, 85, 98, 210]).

$27SB_S$ is cleaved by an unknown endonuclease, roughly in the middle of ITS2, at site C_2. Cleavage at C_2 provides the $7S_S$ and 26S pre-rRNAs. Processing of the 3′-end of 5.8S and the 5′-end of 25S requires a complex succession of, mostly, exoribonucleolytic digestions. During these, consecutive substrates are literally 'handed over' from one ribonucleolytic activity to the next.

The 7S is digested to site E by the successive action of the exosome complex [4, 176, 177], the Rex1p exoribonuclease and Ngl2p, a putative endonuclease [65, 287].

The exosome is a remarkable complex of 11 3′–5′ exoribonucleolytic activities involved in the synthesis and degradation of most classes of cellular RNAs ([6, 99, 109, 176]; reviewed in Refs. [178, 289]). A nuclear form of the exosome is specialized in the synthesis and turnover of large RNAs, including rRNAs and pre-mRNAs as well as most classes of small stable RNAs (snoRNAs, snRNAs, tRNAs, pre-mRNAs, SRP, RNase P, etc.); a cytoplasmic form is devoted to mRNA degradation. Rrp6p (*E. coli* RNase D), a non-essential subunit of the exosome, is specific to the nuclear form of the complex [6, 29]. Nuclear and cytoplasmic exosomes also differ by their use of specific cofactors (see, e.g., Refs. [260, 290]). The related DExH putative RNA helicase Dob1p/Mtr4p (nuclear) and Ski2p (cytoplasmic) is an example [48, 109].

7S precursors are first trimmed from site C_2, located at position +134 with respect to the 3′-end of 5.8S, to position +30 [4] (Fig. 3.2-4). This requires all the subunits of the exosome and the nuclear cofactor Dob1p/Mtr4p. 5.8S+30 pre-rRNA is then digested to position +8 by Rrp6p. 5.8S+8, also referred to as 6S, is consequently trimmed to 5.8S+5 by the multiple exoribonuclease activities of Rex1p, Rex2p and the exosome complex (notably the Rrp40p and Rrp45p subunits) [4, 287]. 5.8S+5 is finally matured to 5.8S by Ngl2p [65] (Fig. 3.2-4).

While the relationship between the subnucleolar compartments and the various ribosome synthesis steps is far from being clear, it is probable that the DFC is the site of early pre-rRNA processing, modification and assembly reactions with later processing cleavages and assembly steps occurring in the GC. SnoRNP core proteins

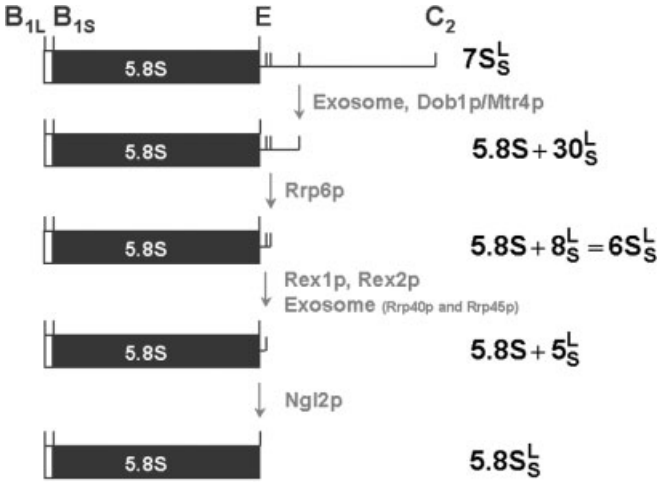


Figure 3.2-4 Multiple steps of ribonucleolytic 'hand-over' are required to synthesize the 5.8S rRNA. Successive pre-rRNA species and trans-acting factors involved are indicated. See main text for a complete description.

involved in 2′–O methylation, pseudouridines formation and early pre-rRNA processing cleavage at sites A_0–A_2 (see below) localize to the DFC [97, 156, 197]. The MRP, involved in cleavage at site A_3, is detected in the GC [225]; this is also where Rlp7p, which is required for cleavage at site C_2, has been localized [79].

Following cleavage at site A_2, the maturation of the small and large rRNAs is relatively independent. However, mutations affecting primarily the synthesis of 5.8S and 25S rRNAs frequently have negative feedback effects on early cleavages at sites A_0–A_2. The mechanism underlying these observations is not known but believed to be part of a 'quality control' mechanism (it would not appear very useful to further initiate the production of pre-ribosomes that will fail to mature properly), which presumably reflects the existence of functional interactions between early and late RRPs.

3′-end formation of other classes of RNAs, such as the snoRNAs and snRNAs, seem to follow a similar strategy of 'exoribonucleolytic hand over' [4, 288]. It is unclear, at present, whether so many distinct nucleolytic activities, with partially overlapping specificity, are required to achieve what would appear to be a fairly straightforward processing. This presumably provides potential for further 'rescue pathways' and quality controls.

The 26S pre-rRNA is trimmed to site C_1 by Rat1p and Xrn1p. This is also probably a multi-step process. Consistently, primer extension through ITS2 from an oligonucleotide specific to the 5′-end of 25S rRNA reveals strong stops at positions +9 and +18 (respective to 25S rRNA 5′-end). The species extending to site +9 (25S′) is lost in some RRP mutants [79]. In the mature subunits, 5.8S and 25S rRNAs are base-paired but the precise timing of this association in the pre-ribosomes is not known.

The major site of Pol I transcription termination (site T_2) is located at position +210 (respective to the 3′-end of 25S rRNA). Precursors extending to this site are however not detected in wild-type cells as primary transcripts are cleaved co-transcriptionally at sites +14/+49 (B_0) on both sides of an AAGN-closed stem-loop structure by the endonuclease Rnt1p [37, 136, 326]. Rnt1p is homologous to bacterial RNase III which similarly cleaves its substrates on both sides of extended stem-loop structures (reviewed in Ref. [121]). Final trimming to site B_2 (the 3′-end of 25S) is carried out by Rex1p/Rna82p [287]. An oligonucleotide specific to sequences located downstream to B_2 detects $27SA_2$ but not 27SB on Northern blots, demonstrating that processing at sites B_1 and B_2 is tightly coupled and presumably concurrent [136].

The minor pathway (used in ~20% of the cases) produce pre-rRNAs and 5.8S rRNA that are extended in 5′ by 7–8 nucleotides. This starts with cleavage of the $27SA_2$ pre-rRNA at site B_{1L} by an unknown enzyme, a presumed endoribonuclease. The resulting $27SB_L$ is then processed into 25S and $5.8S_L$ rRNAs following a pathway that is, as far as we know, essentially identical to the one described above for $27SB_S$.

It is not precisely known when the 5S RNP (5S rRNA associated with RPL5, see [52]) joins pre-60S ribosomes but its recruitment is required for efficient 27SB processing and is therefore presumably concomitant with processing at site C,

thus ensuring that all newly formed 60S subunits contain stoichiometric amounts of the three rRNAs [50, 294].

Alterations in the kinetics of cleavage are seen in many RRP mutants. These usually lead to the accumulation of aberrant precursors that are not faithfully processed to mature rRNAs but rather degraded, notably by the action of the exosome complex [4, 5]. The most often encountered abnormal species, the 23S (extending from sites +1 to A_3), 22S (from sites A_0–A_3) and 21S (A_1–A_3) RNAs, result from alterations in the kinetics of early pre-rRNA processing reactions (Fig. 3.2-3C). Analysis of these species has allowed the description of the processing in the ITS1 and led to the identification of the cleavage site A_3 [98, 154, 155, 184, 248, 268]. Alterations in the order of cleavage at later processing sites are now also known to occur and give rise to the accumulation of a full range of abnormal RNAs; e.g., A_2–C_2, A_2–E, etc. [67, 135].

Over the years, extensive mutagenesis experiments have been performed on rDNA to isolate sequences relevant in cis to pre-rRNA processing reactions. While it is far beyond the scope of this chapter to review this body of data (see Ref. [299]), it should be noted that these experiments have often highlighted how processing reactions distant in the primary rRNA sequence are in fact tightly linked; indeed, mutations in the 5′-ETS, ITS2, or 3′-ETS regions can each inhibit processing in ITS1 (see, e.g., Refs. [7, 20, 292, 293]).

While we now have a fairly complete picture of pre-rRNA processing, much remains to be done to understand the precise kinetics of the processing as well as the extensive connections between early and late cleavage events. Many processing enzymes also remain to be identified, in particular most endoribonucleolytic activities. It is possible that some endoribonucleases have already been assigned to the RRPs and await further attention; the absence of specific motifs in their sequence complicates their identification. The development of *in vitro* reconstitution assays should be most useful in this respect.

It is notable that most known cleavage factors (the exosome, the exoribonucleases Rat1p and Xrn1p, the endoribonuclease Rnt1p) involved in pre-rRNA processing are required for the synthesis and/or degradation of other classes of cellular RNAs (mRNAs, snRNAs, snoRNAs, tRNAs, SRP, RNase P, etc.). All seem to indicate that general maturation factors are recruited from a 'common pool' of proteins to specific cellular pathways. This is also illustrated by the over-increasing sets of proteomic data supporting the existence of extensive integration between ribosome synthesis and other biosynthetic pathways.

3.2.5
Ribosomal RNA Modification: A Solved Issue?

Ribosomal RNAs are extensively modified with a large majority of the modifications clustering at the most functionally relevant and conserved sites of the ribosome (tRNA- and mRNA-binding sites, peptidyl transferase center, intersubunit bridges, entry of the exit tunnel, etc.; see Chapters 6 and 8 for a functional description of the ribosome). This has recently been highlighted on three-dimensional

maps based on crystallographic analysis of archaeal and bacterial ribosomal subunits (see Ref. [49] and useful WWW links). The atomic resolution structure of the ribosome established it as a ribozyme; the peptidyl transferase center is surrounded by an RNA cage leaving little, if any, chance to the RPs to be involved in the peptidyl-transfer reaction *per se* (reviewed in Refs. [55, 143, 181] and Sect. 8.3). rRNA spacers are consistently devoid of modification.

The most frequent RNA modifications are 2′–O methylation of ribose moieties (Nm) and uridine isomerization (pseudouridines, ψ) (~50 of each in yeast; twice this amount in humans) (Figs. 3.2-5b and d). The sites of these modifications are virtually all selected by base pairing with the snoRNAs. Less abundant are the base modifications. These are also essentially modified by methylation (mN) and rely, as far as we know, on protein-specific enzymes rather than the snoRNPs.

3.2.5.1 Ribose Methylation, Pseudouridines formation and the snoRNAs

There are essentially two families of snoRNAs, the box C+D (involved in sugar 2′–O methylation) and the box H+ACA (required for pseudouridines formation) (Fig. 3.2-5). A third class is defined by the related RNAse P/RNAse MRP RNAs. Yeast snoRNAs range in size from about 60 to about 600 nucleotides.

Box C+D snoRNAs consist of a stem-loop structure with boxes C (UGAUGA) and D (CUGA) flanking a terminal helix; duplicated boxes C′ and D′ are also observed

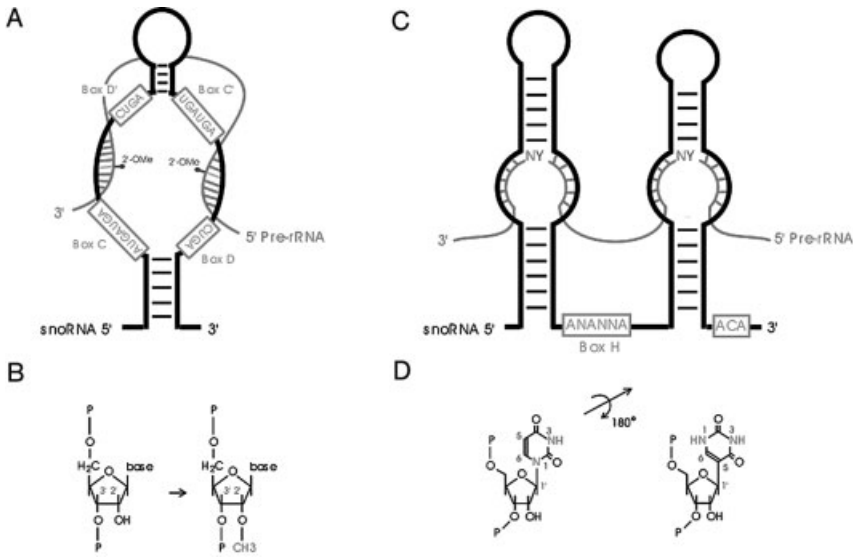


Figure 3.2-5 snoRNA in pre-rRNA modification. SnoRNA/pre-rRNA hybrids at sites of 2′–O methylation (a) and pseudouridine formation (c). Sugar methylation (b) and pseudouridines (d). See main text. Adapted from Ref. [140].

(reviewed in Refs. [72, 140, 320]) (Fig. 3.2-5a). Box H+ACA RNAs show two consecutives hairpin structures, bridged by a conserved H-box (ANANNA, where N is any residue) or hinge motif (hence its name); the triplet ACA is always located 3 nucleotides upstream to the 3′-end of the RNA (Fig. 3.2-5c).

At each site of modification, a duplex is formed by Watson–Crick base-pair interactions between a specific snoRNA and the RNA substrate. This results in the formation of a snoRNA/pre-rRNA hybrid that precisely position the residue to be modified on the substrate with respect to conserved boxes on the snoRNA. For the box C+D snoRNAs, the guide or 'anti-sense' elements are located upstream of boxes D or D′ and provide the potential to form between 10 and 21 consecutive base-pairs with the pre-rRNAs; including the site of 2′–O methylation invariably located five nucleotides upstream of boxes D or D′ (reviewed in Refs. [11, 124]). For the box H+ACA snoRNAs, the 'anti-sense' motifs are within internal bulges (also known as 'pseudouridylation pockets') in the hairpin stems, and target the formation of two short helices of 3–10 base-pairs with the substrate; these are interrupted by the uridine to be altered by rotation into ψ (reviewed in Refs. [12, 125]). This uridine is usually located at about 14 residues from boxes H or ACA. SnoRNAs show a high degree of divergence outside the conserved boxes; including, obviously, the 'anti-sense' elements.

SnoRNAs are associated with a limited set of specific core proteins. Snu13p (15.5K in humans), Nop1p (yeast *Fibrillarin*) and the related Nop56p and Nop58p-KKD/E containing proteins are associated with all box C+D snoRNAs [83, 144, 146, 237, 316]. The human 15.5K is expected to nucleate box C+D snoRNP assembly through direct binding to a conserved RNA-fold (K-turn, see below) generated by interactions between boxes C and D [316]. Remarkably, Snu13p is also a component of the spliceosomal U4/U6•U5 tri snRNP [316]. Cbf5p (NAP57 in rodents, *Dyskerin* in humans), Gar1p, Nhp2p and Nop10p are all associated with the H+ACA snoRNAs [90, 97, 141, 315]. Concurring evidences support that Nop1p and Cbf5p are the methyltransferase and the pseudouridine synthase, respectively [102, 141, 201, 271, 307, 333]. The localization of the snoRNAs and their associated core proteins in the DFC of the nucleolus suggest that this is the site of rRNA modification [97, 156, 197].

Telomerase is an RNP reverse transcriptase that maintains telomere length by adding telomeric DNA repeats onto the ends of eukaryotic chromosomes [23, 36, 165]. In humans, the telomerase RNA (hTERT) has a canonical H+ACA motif at its 3′-end that is bound by the core H+ACA proteins [53, 57, 174, 217]. In yeast, the TEL RNA is bound by the spliceosomal Sm proteins [241]; another interesting evolutionary crosstalk. Human telomerase also interacts with La and SMN (see below and Refs. [10, 73]).

Strikingly, several self-immune and genetic human diseases map to key nucleolar RNA-processing factors and snoRNP proteins, such as the exosome, *fibrillarin*, *dyskerin* and the RNAse MRP [6, 30, 96, 159, 175, 227, 231, 274, 334].

3.2.5.2 The Emergence of the snoRNAs

Eukaryotes have about 10 times (in the range of the hundred) more modifications in their ribosomal RNAs than prokaryotes do. An observation that, a *posteriori,* seems to fully justify the emergence of the snoRNAs. Why would a cell evolve and produce several dozens of protein enzymes with distinct substrate specificities when it can rely on a single snoRNA-associated protein? In addition, the snoRNA-based system of RNA modification is very flexible as the guide sequences are not conserved (they have little, if any, functions in snoRNA synthesis, stability, and nucleolar targeting) and are therefore prone to rapid evolution. The accumulation of point mutations in snoRNAs generates new 'anti-sense' elements that, eventually, will find new RNA targets.

In fact, there is a steady increase in the number of modification across evolution with bacteria and eukaryotes on both ends of the range and the archaea showing intermediate distributions. This raised the possibility that these too may rely on a 'snoRNA-like mechanism' to select their sites of RNA modification (discussed in Ref. [140]). An assumption that turned out to be correct as a large family of archaeal box C+D and H+ACA sRNAs (archaea lacking a clear nucleolar structure) and a full set of core sRNPs proteins has now been described ([82, 137, 200, 201]; reviewed in Refs. [51, 266]). Remarkably, in archaea, the snRNAs not only target the modification of rRNAs but also of tRNAs [45]. A model, based on the assumption that 2′–O methylation confers extra thermostability, has been proposed that correlates the distribution of archaeal rRNA modification with the temperature of their ecological niches [196].

Archaeal box C+D sRNAs, active in methylation, have been reconstituted *in vitro* from individually produced components [201]. In these experiments, assembly appeared to follow a strict order with the aL7a (archaeal Snu13p) binding first to the RNAs, followed by aNOP56 binding (archaeal Nop56/58p) and then finally association with aFib (archaeal Nop1p). These analysis led further support to the predictions that Snu13p may nucleate the step-wise assembly of box C+D snoRNPs and that Nop1p carries the methyltransferase activity (mutations in aFib catalytic motifs were inactive in methylation). In yeast, Nop56p was dependent on Nop1p for binding to the snoRNAs whereas Nop58p was found to bind independently [146].

Recent studies have revealed that archaea assemble symmetric sRNPs with a complete set of core proteins (L7a, the single Nop56/58p homolog and fibrillarin) at both box C+D and C′+D′ motifs [275]. In contrast, eukaryotes snoRNPs appeared asymmetric with a distinct set of core proteins bound to each motifs; 15.5K, Nop58p, and fibrillarin were all detected at the terminal C+D motif, whereas Nop56p, fibrillarin, but no apparent 15.5K, were present at the internal C′+D′ position [31, 321]. A rationale to this key difference in protein composition is provided by the observation that during evolution, the 15.5K seemed to have lost its ability to recognize internal C′+D′ motif [275]. The box C′+D′ motif is degenerated and suboptimal for tight association with the core proteins. Significantly, the recent resolution of the 3D structure of an archaeal Nop58p–fibrillarin complex bound to S-adenosyl methionine

(SAdoMet), the universal methyl donor, strongly suggests that the C-terminal coiled-coil domain of Nop58p may promote its homodimerization and allow the assembly of a core complex at the suboptimal C′+D′ motif ([2]; reviewed in Ref. [70]). In eukaryotic snoRNPs, this interaction would take place between the C-terminal tails of Nop58p and Nop56p at the C+D and C′+D′ motifs, respectively, and possibly compensate for the absence, at this site, of the nucleation activity carried out by 15.5K.

Snu13p belongs to a family of related RNA-binding proteins including several RPs of both subunits: yeast L30 (which binds to its own mRNA for autoregulation, see Refs. [303, 304]) and human L7a and S12, the box H+ACA snoRNP protein Nhp2p [97], SBP2 (which binds to the stem-loop SECIS element in the 3′-UTR of selenocysteine protein-encoding mRNAs, see Refs. [3, 132]) and eRF1 (a subunit of the translation termination release factor). These proteins have been shown, or predicted to, bind to a ubiquitous RNA structural motif, known as 'kink-turn' (K-turn, [126]) or 'GA motif' suggesting that they share a similar strategy for binding to their substrates.

Interestingly, archaeal Snu13p not only binds to the box C+D sRNAs but also to the LSU (23S) rRNA contacting a K-turn and suggesting that ancestors to small stable RNAs may have evolved from rRNA segments; an assumption further supported by the identification of an archaeal box C+D sRNA within a non-coding rRNA spacer region [261].

In the widely accepted concept of the 'prebiotic RNA world', RNAs preexisted proteins and most essential functions were carried out by 'RNA-based machines'. In contrast, the model proposed for the emergence of the snoRNAs is a case where a function initially performed by individual proteins has slowly been taken over by RNPs to achieve greater efficiency (see Ref. [140] for further discussions).

3.2.5.3 Non-ribosomal RNA Substrates for the snoRNAs

Although originally described in pre-rRNA modification, snoRNAs and alike (archaeal sRNAs, human scaRNAs, see below) have now been demonstrated to work on other RNA substrates, including spliceosomal U RNAs U1, U2, U4 and U5 (Pol II transcripts) and U6 (Pol III), tRNAs and possibly mRNAs [32, 45, 46, 81]. An interesting case of putative mRNA guide is a tissue-specific (brain) snoRNA, expressed from an imprinted region of the genome that is linked to the neurodegenerative genetic disease Prader-Willi syndrome [32–34]. Remarkably, this snoRNA is expected to target a site of RNA 2′–O methylation on a serotonin receptor mRNA at a position that is also subjected to A to I editing.

Orphan snoRNAs are waiting for their RNA target to be identified and many more classes of RNAs are expected to use a similar strategy for their modification. Viral RNAs are particularly interesting to consider in this respect, as these would require additional co-evolution with their hosts.

3.2.5.4 Possible function(s) of RNA modifications

Several structural and thermodynamic effects have been proposed for RNA modifications, including altered steric properties, different hydrogen-bonding potential and increased local base stacking (ψ), increased structural rigidity (ψ and Nm) and protection from hydrolysis of inter-nucleotides bonds (Nm) [147, 198]. However, the precise function of these modifications is not known and we have failed to identify a single modification that is essential for ribosome synthesis or function, although the selective loss of the ψ's surrounding the peptidyl transferase center significantly reduce translation efficiency [123]. It is therefore probable that each modification contributes a little benefit and that it is the overall modification pattern that significantly improves ribosome synthesis and/or function. It is quite remarkable that three sites of ψ and three sites of 2′–O methylation are common to bacteria and eukaryotes; these have been selected independently twice during evolution and are made by distinct mechanisms (snoRNPs versus protein-specific enzymes; see Ref. [140]). In addition, most known modification enzymes carry additional, presumably indirect, essential functions in ribosome synthesis, notably in pre-rRNA cleavage (e.g., Nop1p, Cbf5p, Dim1p; reviewed in Ref. [145]).

An attractive hypothesis certainly remains that RNA modifications are simply 'by-products' reflecting the involvements of the snoRNAs in pre-rRNA processing and pre-rRNP assembly. Through extensive base pairing with the rRNA precursors, snoRNAs dictate specific pre-rRNA structures and fold them into conformations that are competent for processing and assembly. Modifications could then be seen as mere triggers to unleash the snoRNAs from the pre-ribosomes following the precise kinetics of ribosome assembly. In yeast, methylation of the rRNA occurs immediately after the completion of transcription [226, 283], implying that the snoRNAs are associated with the growing chain as it is being transcribed and potentially circumvent early unwanted folding.

3.2.5.5 Base methylation

Several putative base methyl-transferases have been described and, as far as we know, do not involve the snoRNAs for their function [103, 131, 212, 249, 327].

A well-characterized example of base methylation is the 18S rRNA dimethylation carried out by Dim1p (KsgAp in *E. coli*). Both the site of modification (the 3′-terminal SSU hairpin located at the subunit interface where interactions important for ribosome function occur) and the modification itself (a twin methylation at position 6 on two adjacent adenosine residues) are highly conserved in evolution [145, 291]. Methylation of the pre-rRNAs by Dim1p is a fairly late event in the SSU assembly pathway, possibly linked to 40S subunit export and occurring in the cytoplasm. However, Dim1p binds to the pre-rRNAs in the nucleolus and is required for early cleavages at sites A_1 and A_2 [138, 139, 142]. This is further evidence for the existence of 'quality control' mechanisms in ribosome synthesis. Processing does not occur on pre-rRNAs that have failed to bind Dim1p and will consequently not be methylated. Consistently, the Dim1p methylation is essential for ribosome function *in vitro* and is

favorable to translation *in vivo* (reduced rates of frame-shifting and misreading; D. Demonté and D.L.J. Lafontaine, unpublished results).

A thermosensitive conditional mutation in Dim1p is suppressed on overexpression of RPL23, a primary binding protein of the large ribosomal subunit (D. Demonté and D.L.J. Lafontaine, unpublished results). This indicates that alteration in the kinetics of LSU assembly (the process is presumably prematurely triggered on RPL23 overexpression) overcomes the need for the 'quality control' exerted by Dim1p in early pre-rRNA processing and small subunit synthesis.

In bacteria, the methylation is conserved but KsgAp is not essential, indicating that eukaryotic Dim1p evolved an additional function in ribosome synthesis.

3.2.5.6 U3 snoRNP, the 'SSU Processome', and the Central Pseudoknot

Several snoRNAs are involved in pre-rRNA cleavage rather than pre-rRNA modification. In yeast, these include the box C+D snoRNAs U3 and U14 and the box H+ACA snoRNAs snR10 and snR30. U3, U8, U14 and U22 are also involved in pre-rRNA cleavage in metazoans (reviewed in Ref. [270]). In yeast, U3, U14, snR10, and snR30 are required for the first three pre-rRNA cleavages at sites A_0–A_2; these are either delayed (snR10) or inhibited (U3, U14, and snR30) [105, 153, 184, 269]. U14 and snR10 are also required for pre-rRNA modification. For snR10, a point mutation in the guide sequence could efficiently uncouple its requirement for pre-rRNA processing and modification [123]. Metazoans have an additional member (U8) involved in ITS2 processing [204, 205, 272, 273]; no equivalent has thus been found in yeast.

U3 is undoubtedly the best-characterized member of this class of snoRNAs both in structure, function, and synthesis (see below). U3 is larger than most box C+D snoRNAs (333 nucleotides in yeast) and carry, in addition to the conserved core motifs, sequences (including a protruding 5′-extension, largely unfolded, and ending with a stem-loop) that are known, or presumed, binding sites for about a dozen of U3-specific proteins: Mpp10p, Imp3p, Imp4p, Sof1p, Dhr1p, Lcp5p, Rrp9p/h55K, Rcl1p, and Bms1p [22, 44, 60, 112, 151, 216, 300, 317, 324].

Recently, U3 has been isolated in association with 28 proteins [56, 319]; 10 of which were known U3-specific RRPs, another was a known RRP involved in early and late pre-rRNA processing (Rrp5p), the remaining 17 (Utp1-17p) were all nucleolar and required for 18S rRNA synthesis. This complex is now referred to as the 'SSU processome' and on the basis of its calculated mass (>2 200 000 kDa) and large size (~80S; roughly the size of a mature ribosome or the spliceosome) has been proposed to correspond to the terminal balls visualized at the 5′-ends of nascent transcripts in chromatin spreads [172, 187]; depletion of several 'SSU processome' components led to the disappearance of these structures [56].

The function of U3 in pre-rRNA processing is mediated through at least two Watson–Crick base-pair interactions between U3-specific motifs and the pre-rRNAs. An interaction between an essentially unstructured region of the 5′-extension of U3 and the 5′-ETS (at site +470) is required for cleavages at sites A_0–A_2 [18–20]. A second interaction between a conserved motif (box A) in the 5′-stem-loop of U3 and the

pre-rRNA at the 5′-end of the mature 18S rRNA is necessary for cleavage at sites A_1 and A_2 [242]. The interaction between box A and the 18S rRNA 5′-end is mutually exclusive with the formation of the central pseudoknot, a conserved long-range interaction, which brings together, in the mature particles, sequences that are located more than a kb apart. The formation of the central pseudoknot is a major structural rearrangement in the SSU rRNA and as such is most probably an irreversible step in ribosomal assembly. Dhr1p, a U3-specific DEAH putative RNA helicase required for pre-rRNA processing at sites A_1 and A_2, has been proposed to be involved in this RNA isomerization [44]. One possibility is that the action of Dhr1p is regulated such as to leave sufficient time for early pre-rRNP assembly to occur prior to the formation of the central pseudoknot. Growing yeast cells have about enough copies of the U3 snoRNP to support ribosome synthesis for only ~1 min in the absence of recycling (considering a production rate of ~2000 ribosomes/min). A function of Dhr1p in recycling the U3 snoRNP and in SSU-processome assembly is therefore also probable. This is currently under investigation.

3.2.6 SnoRNA Synthesis and Intranuclear Trafficking

3.2.6.1 SnoRNAs Synthesis

SnoRNAs have adopted a large range of strategies for their expression. Their synthesis, in the nucleoplasm, can either proceed from individual Pol II (most snoRNAs) or Pol III (U3 in plants) promoters and produce mono- (most yeast snoRNAs) or polycistronic units (many plants snoRNAs; several yeast snoRNAs) or be expressed from introns of house-keeping genes (most vertebrates snoRNAs; several yeast snoRNAs) (reviewed in Refs. [72, 164, 320]). Host genes are often somehow related to ribosomal synthesis or function and, in extreme cases, do not seem to have any additional function than to carry the snoRNAs, i.e., no proteins are expressed from the spliced mRNAs [26, 32, 207, 251, 281].

SnoRNA maturation is complex. Processing of independently encoded or polycistronic units is initiated by endonucleolytic, possibly co-trancriptional, cleavage in the 3′-portion of the primary transcript and requires Nrd1p, the Sen1p helicase and the cleavage factor IA activity of the mRNA polyadenylation machinery [69, 182, 257]. SnoRNAs encoded in polycistronic units are separated by the endonucleolytic activity of Rnt1p/yeast RNase III [38, 39, 221]; precursors transcripts containing a single snoRNA may also be cleaved at their 5′-ends by Rnt1p [38].

Intron-encoded snoRNAs are usually synthesized from the excised intron lariat following splicing and debranching by Dbr1p, and exonucleolytic trimming on both ends [202, 210]. In a minor, splicing-independent pathway, the pre-mRNA is directly cleaved endonucleotically to provide entry sites for exoribonucleases.

In all cases, final pre-snoRNA maturation steps require exonucleolytic digestions to the mature ends. This involves 3′ to 5′ exonucleolytic digestion (exosome) [4, 288] and, at least in the case of intronic or polycistronic snoRNAs, 5′ to 3′ exonuclease digestion (Rat1p, Xrn1p) [210, 221].

The best-characterized pre-snoRNA processing pathway is for the box C+D snoRNA U3 (Fig. 3.2-6). As for many other snoRNAs, U3 is synthesized with 3′-extensions; these require endonucleolytic cleavage (Rnt1p) to provide an entry site for a processing 'hand over' by the exosome subunits [134]. This processing is literally 'timed' by the binding of yeast Lhp1p (human La) to poly(U)-rich tracks located close to the RNA 3′-ends [134]. Displacement of La is concomitant with snoRNP assembly (the core snoRNP proteins bind to the RNA, presumably conferring 3′-ends protection) and allows final trimming by the exosome to produce the mature 3′-ends. The binding of La to the pre-snoRNAs presumably provides sufficient time for snoRNP assembly to occur prior to the final action of the exosome complex. U3 additionally requires the concomitant splicing of an intron.

Individually expressed Pol II snoRNA precursors are produced with a 5′-terminal 7-monomethylguanosine (m^7G) cap that is retained in many snoRNAs and hypermethylated to 2,2,7-trimethylguanosine ($m^{2,2,7}G$ or TMG) by Tgs1p [186]; the timing of this modification is not known. Tgs1p is also active on snRNAs. For U3, cap trimethylation is dependent on boxes C and D and is concomitant with 3′-end formation and snoRNP assembly as 3′-extended forms of U3 are not bound by the core proteins and are not precipitated by anti-TMG antibodies [134, 252, 253, 264]. In plants, U3 is transcribed by Pol III and carries a γ-monomethyl phosphate cap [245].

3.2.6.2 Non-core snoRNP Proteins required for snoRNA Accumulation

Besides the core components, several proteins have been linked physically or functionally to the snoRNPs but are not found in mature snoRNPs. Such proteins are the Rvb2p(p50)/p55 putative NTPases [122], the putative DEAD-box helicase Sen1p [285], the Naf1p/Shq1p complex [54, 68, 330] and Nopp140 [331]. These are required for snoRNA accumulation, through presumed transient interactions, and are potentially involved in snoRNA synthesis, snoRNP assembly, and/or nucle(ol)ar trafficking.

The nucleoplasmic p50/p55 complex is required for the stability of both box C+D and box H+ACA snoRNAs as well as for proper nucleolar localization of the core proteins Nop1p and Gar1p. Mammalian orthologs have DNA unwinding activity *in vitro* and have been linked to chromatin remodeling and transcription (see Ref. [122] and references therein).

Sen1p is required for snoRNA accumulation of both families as well as several other classes of RNAs (including rRNAs, tRNAs, and snRNAs) [223, 285]. Nop1p is mislocalized on Sen1p inactivation [284].

The Naf1p/Shq1p complex is specific to box C+D snoRNAs accumulation. Naf1p is mostly localized to the nucleoplasm and can be co-precipitated at low levels with several snoRNP components [54, 68, 330]. Naf1p interacts directly with the RNA *in vitro*, and most interestingly, is found in association with the phosphorylated form of the C-terminal domain (CTD) of Pol II. This provides a further link to RNA synthesis [68].

Nopp140 (yeast Srp40p) [166, 168], a highly phosphorylated nucleolar- and CB-specific protein, is found in association with both box C+D and box H+ACA

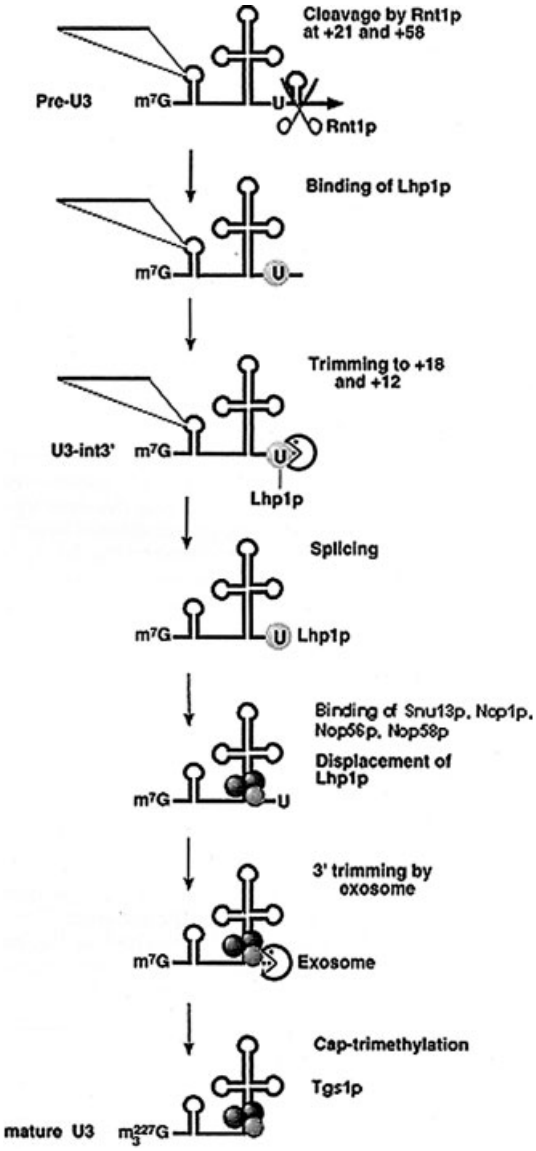


Figure 3.2-6 U3 synthesis pathway. The box C+D snoRNA U3 is synthesized with 3′-extensions; these are cleaved co-transcriptionally by Rnt1p/yeast RNase III. Yeast La (Lhp1p) binds to 3′-terminal poly(U) tracks. Lhp1p-bound precursors are monomethylated and are not assembled with the core proteins. SnoRNP assembly is concomitant with the displacement of La and the production of mature 3′-ends by the exosome; the cap is trimethylated by Tgs1p. The yeast U3 genes are unusual in that they contain an intron; this is spliced out from the 3′-extended precursors. Adapted from Ref. [134].

snoRNPs [108, 331]; association with the box H+ACA is more avid. The interaction with the snoRNPs is dependent on Nopp140 phosphorylation [306]. The expression of a dominant-negative allele of Nopp140 depletes core snoRNP proteins (NAP57, Gar1p, and fibrillarin) from nucleoli and inactivates Pol I transcription [331]. Nopp140 was also coimmunoprecipitated with the largest subunit of Pol I; strengthening a link between snoRNP metabolism and transcription [41]. NAP57/Cbf5p was originally isolated as a Nop140-associated protein [167]; Nopp140 is however not required for *in vitro* pseudouridine formation [306]. Box H+ACA snoRNAs are lost on *srp40* deletion in a yeast synthetic lethal background [331].

3.2.6.3 Interactions between Cleavage Factors and Core snoRNP Proteins

Interaction between Rnt1p and Gar1p is required for optimal Rnt1p activity in pre-rRNA processing, nucleolar localization of the core H+ACA proteins and pseudouridylation. This provides a link between snoRNP synthesis and transport and between RRPs involved in 3′-ETS co-transcriptional cleavage (Rnt1p) and 5′-ETS pre-rRNA processing (Gar1p) [278]. This possibly ensures proper pre-rRNA kinetics and coordinated cleavages on both ends of the primary transcript and prevents processing of incomplete molecules.

In addition, Rnt1p accurately cleaves most of the snoRNA substrates *in vitro* in the absence of other cofactors, with the exception of the U18 intron-encoded snoRNA, which requires the additional presence of the box C+D snoRNP protein Nop1p; Rnt1p and Nop1p interact with each other in pull-down experiments [89].

3.2.6.4 SnoRNAs Trafficking

The synthesis of the snoRNAs in the nucleoplasm but their function, in pre-rRNA processing and/or modification, in the DFC of the nucleolus raise interesting questions as to their localization pathway. Nucleolar targeting and localization of the snoRNAs is probably achieved by diffusion through the nucleoplasm followed by retention through multiple interactions with nucleolar components.

The cis-acting elements involved in this nucleolar targeting have been identified and, unsurprisingly, precisely map to the conserved boxes C and D and H and ACA [148, 149, 192, 193, 232]. These are the only sequences conserved in the snoRNAs and are, known or presumed, protein-binding sites. The ACA element in the telomerase RNA is also required for its nucleolar trafficking [158, 192].

Trans-acting factors involved in this process have only started to be addressed in yeast, with most attention being paid to the box C+D snoRNAs. All core proteins are required as well as several nucleolar proteins of previously ill-defined or unknown functions such as Srp40p (Nopp140 in rodents) and Nsr1p (human nucleolin). The Ran cycle is not involved [191].

Nucleolar routing involves transit through the CB in plants and vertebrates and their recently identified homolog in yeast, the NB [193, 243, 301, 302] (Fig. 3.2-7).

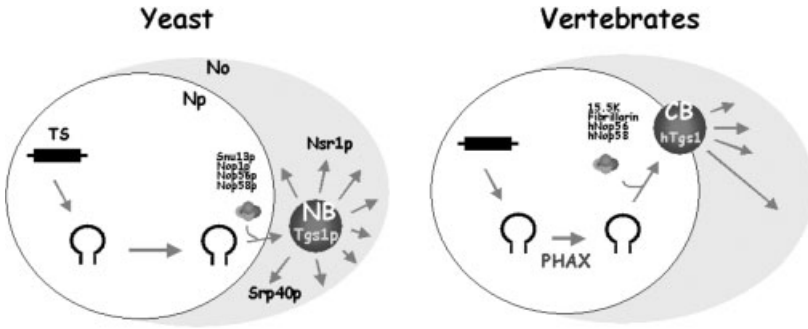


Figure 3.2-7 Intra-nuclear trafficking of box C+D snoRNAs. A comparison between the yeast and vertebrate systems is provided. Box C+D snoRNA nucleolar targeting involves transit through a conserved nuclear locale, the NB and CB, in yeast and vertebrates, respectively. The cap trimethyl-transferase (Tgs1p/hTgs1) is a specific antigen of this cellular compartment. SnoRNA nucleolar routing is a multi-step process. In mammals, PHAX, the phosphorylated adaptor for snRNA export, drives the snoRNAs from their transcription sites (TS) to the CB (E. Bertrand, pers. Comm.). SnoRNP assembly and captrimethylation presumably occur in the NB/CB. Np, nucleoplasm; No, nucleolus.

Overexpression of artificial box C+D snoRNAs in yeast led to their accumulation in a single, roughly spherical structure of ~200 nm in diameter always contiguous to the fibrillar component of the nucleolus [302]. This structure was highly reminiscent to the CBs, also often found in close association with the nucleolus and functionally linked to this nuclear locale ([24, 215, 250]; reviewed in Ref. [80]). Expression of a human GFP-SMN reporter construct (a CB-specific antigen) specifically co-localized with the NB strongly supporting this assumption. In addition, the cap trimethyl-transferase Tgs1p, specifically localized to the CB and NB in yeast and humans, respectively, providing further supporting evidences [301]. Most importantly, NBs were later detected with endogenous snoRNAs, in the absence of snoRNA overexpression, supporting the physiological importance of this novel nuclear compartment [301].

Survival of motor neurons (SMN) is the causative agent for spinal muscular atrophy, a neurodegenerative disease and most frequent genetic cause of infant mortality ([152]; reviewed in Refs. [76, 194]). SMN is present in multiple RNP complexes and notably interacts with core snoRNP proteins of both families (Nop1p and Gar1p), the human telomerase RNP and the human cap trimethyl-transferase, hTgs1 [10, 116, 185, 208]. In the best-described complex, SMN is associated with Gemins 2–6 and is involved in snRNP metabolism and pre-mRNA splicing ([209]; reviewed in Refs. [169, 203, 265]). In a Gemin 3 (a putative DEAD-box helicase), gemin 4 and eiF2C-specific complex, SMN has also recently been linked to the metabolism of the micro RNPs (miRNPs) [188].

The accumulation of snoRNAs in NBs on RNA overexpression suggested that nucleolar targeting is a saturable, multi-step process (Fig. 3.2-7); snoRNAs would first transit from transcription sites (TS) to NB/CB before being redistributed to the entire nucleolus. Both Nsr1p and Srp40p were involved in the emergence of the NB [302].

The first step in the nucleolar routing pathway has recently been successfully uncoupled from the subsequent nucleolar distribution and imaged in time-lapsed microscopy [27]. Transcription sites and CBs were relatively static as to their locations (at least within the time frame used, ~1 h); snoRNPs appeared to transit from TS to the vicinity of CBs within minutes but strikingly lagged for up to 60 min before being incorporated into this compartment [27].

PHAX (phosphorylated adaptor for snRNA export; [199]) is localized in the nucleoplasm and the CBs, binds specifically to box C+D 3′-extended precursors, and is able to target artificial RNA substrates from their transcription sites to CB, supporting a direct role for this protein in the first step of nucleolar routing (Fig. 3.2-7). PHAX interacts with the 15.5 K (human Snu13p) *in vitro* and contact the snoRNPs, at least in part in an hSnu13p-dependent fashion (E Bertand, pers. comm.). 15.5K is also present in the spliceosomal U4 snRNP (see Sect. 3.2.5.4), raising interesting questions as to the discrimination of snRNAs and snoRNAs for their trafficking. Studies on the U3 box B+C motif, which is also bound by 15.5 K, indicate that specific flanking sequences and/or structure, surrounding a conserved 15.5 K-binding site, probably provide the specificity for the recruitment of additional complex-specific proteins [92].

The recent identification of box C+D and/or box H+ACA containing small RNAs localized at steady-state in the CB [46], hence their name scaRNAs (small cajal bodies specific RNAs) and active in snRNAs modifications raise additional questions as to the presence of specific cis- or trans-acting determinants in these RNAs for CB retention.

3.2.6.5 **CB/NB are Conserved Sites of Small RNP Synthesis**

Our current view is that NBs/CBs are conserved sites of small RNPs biogenesis; maturation steps occurring in NBs/CBs include snoRNA cap trimethylation (presence of Tgs1p), snRNA internal modification (identification of the scaRNAs) and snoRNA 3′-end formation and snoRNP assembly (occurrence of unassembled 3′-end-extended snoRNA precursors and core snoRNP proteins).

3.2.7
Ribosome Intranuclear Movements and Ribosome Export

Once released from the nucleolus, pre-ribosomes transit through the nucleoplasm to reach the NPC. The precise mechanisms of ribosomes intranucle(ol)ar movements are unknown. This presumably occurs by diffusion and may involve unleashing the pre-ribosomes from successive nucle(ol)ar retention sites.

Interestingly, three related couples of proteins, originally identified in a large Pol I transcription-related nucleolar complex [66], have recently been involved in this process. In these, Noc1p (Mak21p), Noc2p (Rix3p), and Noc4p share a 45-amino-acid-long domain (Noc domain) [170, 171]. Noc2p organizes two distinct nucleolar complexes, Noc1p/Noc2p and Noc2p/Noc3p (a related nucleolar protein which does not show a Noc motif). The Noc complexes differ both in their intranuclear

localization and association with the pre-ribosomes. Noc2p/Noc3p is mainly nucleoplasmic and interacts with 66S particles; Noc1p/Noc2p is nucleolar-enriched and associates with the 90S and 66S pre-ribosomes [170]. The Noc1p/Noc2p and Noc2p/Noc3p complexes are required for pre-60S export. The Noc1p homolog, Noc4p, is associated with Nop14p (another unrelated nucleolar protein) [157]; the Noc4p/Nop14p complex is nucleolar, associated with 90S and presumably 43S pre-ribosomes and is involved in pre-40S export [171]. The dynamic intranuclear distribution of the Noc proteins (potential to shuttle between the nucleolus and the nucleoplasm) and their association with distinct species of pre-ribosomes supports a role in intranuclear movements.

A problem faced with many RRP mutants defective in ribosome export (also referred to as Rix, for ribosome export) is that they are, in addition, impaired in pre-rRNA processing. Typically, strains defective for pre-60S export show inhibitions in early pre-rRNA processing reactions (sites A_0–A_2). This suggests that efficient pre-rRNA processing is dependent on ongoing ribosome export. Most importantly, in this respect, overexpression of the Noc domain results in a dominant-negative phenotype for growth and nuclear accumulation of the pre-ribosomes in the absence of pre-rRNA processing defects [170]. In this case, pre-rRNA processing and transport defects were efficiently uncoupled, strongly supporting a direct involvement of the Noc proteins in intranuclear movement and nuclear exit of the ribosomes. Another RRPs, the ribosomal-like protein Rlp7p, has also been recently involved in pre-60S subunits release from the nucleolus [79].

Export assays based on microinjections in *Xenopus* oocytes and the use of isolated *Tetrahymena* nuclei concluded that ribosome nuclear exit is a unidirectional, saturable (involvement of trans-acting factors, including components of the NPC), energy- and temperature-dependent process [17, 88, 120, 218, 328]; subunits are believed to transit to the nucleoplasm independently.

In yeast, the intranucle(ol)ar accumulation of pre-ribosomes is either monitored *in vivo* by the use of fluorescent reporter RPs (e.g. Rps2p-eGFP, Rpl11p-GFP, and Rpl25p-eGFP) [78, 106, 171, 56] or on fixed samples by FISH (e.g., a probe specific to the 5′-portion of ITS1 has been used to follow pre-40S export) [189, 190]. Although none of these strategies is entirely satisfactory (the RPs assay relies on proper incorporation of the reporter constructs in strains that are also potentially defective for assembly; the FISH assay largely used a *xrn1Δ* strain that accumulates high levels of cytoplasmic 20S and/or ITS1 D-A_2 fragment but with a *plethora* of associated phenotypes in unrelated processes as diverse as mRNA turnover, microtubule function, DNA replication, telomere length, karyogamy, etc.; see discussion in Ref. [189]), they nevertheless succeeded in identifying a role in ribosome export for a subset of RPs, several nucleoporins, the Ran-system, as well as a, very large number of known or novel RRPs.

A well-characterized set of Rix proteins is the Rpl10p/Nmd3p/Xpo1p complex. Rpl10p binds late to the pre-60S ribosomes and interacts with Nmd3p, a nucleo-cytoplasmic shuttling protein which serves as a transport adaptor providing a leucine-rich nuclear export signal (NES) to the exportin Xpo1p/Crm1p ([78, 100, 255];

reviewed in Refs. [1, 115]); the RRP Rsa1p was involved in facilitating the loading of Rpl10p onto pre-ribosomes [129]. The Nmd3p-mediated pathway of LSU export is conserved in metazoans [267, 279]. It is most probable that additional such NES are provided, either directly or not, by the RPs. Consistently, Yrb2p, a Ran-GTP-binding protein required for the efficient export of NES-containing protein has recently been involved in 40S subunit export [190, 263]. In addition, a specific conditional inactivation of Mtr2p, which is required for mRNA export [233], led to the nuclear accumulation of pre-60S ribosomes and was synthetic lethal with Nmd3p [14]; the mechanism underlying these observation is not known at present.

Proteomic analyses of late nuclear pre-60S complexes revealed the presence of the Rpl10p complex as well as several RRPs that were also isolated in NPC purifications [14, 229].

Since the size of the NPC is just about enough (~20–25 nm in diameter) to accommodate that of individual ribosomal subunits (25–30 nm), it is anticipated that extensive remodeling is needed prior to, during passage through the pore, and following nuclear exit of such large RNPs. The recently identified AAA-ATPase Rix7p is a good candidate to be involved in such structural rearrangements [77].

How late pre-rRNP cleavage, modification and assembly are coupled to intranuclear movements and translocation of pre-ribosomes through the NPC is the subject of ongoing research.

3.2.8 The Cytoplasmic Phase of Ribosome Maturation

Following nuclear exit, both the small and large ribosomal subunits undergo final cytoplasmic maturation steps; these include structural rearrangements, the addition of late RPs, and possibly, late pre-rRNA processing and modification reactions. These steps underlie the long-standing observation that ribosomal subunits undergo a significant cytoplasmic lag before their incorporation into polysomes [133, 276, 282, 314]. The recent identification of the first trans-acting factors involved in these reactions led to an important novel concept in the field, several RRPs follow the pre-ribosomes to the cytoplasm and, at least for some of them, are recycled to the nucle(ol)ar pre-rRNA processing machinery [78, 100, 195, 240, 332]. Such nucleo-cytoplasmic shuttling was observed more than a decade ago for nucleolin/C23 and No38/B2, two important vertebrate nucleolar antigens [25]. However, the interpretation of these data was not clear at that time.

The 20S pre-rRNA is exported to the cytoplasm where cleavage at site D, by an unknown RRP, generates the 18S rRNA. A conclusion largely based on cell fractionation experiments [276, 282] and indirectly supported by the following concurring evidences: (i) strains deleted for the major cytoplasmic 5′–3′ exoribonucleolytic activity (Xrn1p) accumulates high levels of the D-A_2 fragment in the cytoplasm [189, 258]; (ii) strains genetically depleted for Rio1p or Rio2p, two putative protein kinases, accumulate increased amounts of cytoplasmic 20S pre-rRNAs [295, 296]; and (iii) deletion of the translation initiation factor eIF3j (Hcr1p) slightly impairs 20S pre-rRNA processing [286].

Although cleavage at site D is certainly very closely linked to small subunit export, it should be kept in mind that (i) Xrn1p works cooperatively with Rat1p in multiple nuclear reactions (see above) and, consistently, Xrn1p has recently been copurified with nucle(ol)ar pre-60S subunits [195], and that (ii) the D-A_2 fragment, or the 20S pre-rRNA (none of which has ever been directly detected in the cytoplasm in a wild-type strain, see Refs. [189, 190]) could be leaking through the NPC in *xrn1Δ* backgrounds or *rio* mutants. Furthermore, although mostly located in the cytoplasm, eIF3j is also detected within the nucleus. The formal possibility that cleavage at site D occurs shortly prior to nuclear exit or during passage through the NPC prevails. The Dim1p dimethylation was also reported to be a late, cytoplasmic event based on crude cell fractionation and fingerprint analysis [28, 127, 160, 161]. Nucle(ol)ar pre-rRNA precursors are dimethylated when pre-rRNA processing kinetics is altered [98, 139] and dimethylation too could still formally be a late nucleoplasmic event closely linked to export in wild-type strains.

Elongation factor-like 1 (Efl1p), a cytoplasmic GTPase homologous to the ribosomal translocases EF-G and EF-2, has recently been involved in nucleolar pre-rRNA processing at sites A_0–A_2 [240]. It turned out that in strains deficient for Efl1p, Tif6p (a nuclear protein involved in early pre-rRNA processing [16]) is mislocalized to the cytoplasm. We proposed that the pre-ribosomes exit the nucleus in association with Tif6p and that the latter is unleashed from the particles and allowed to recycle to the nucle(ol)us following a structural rearrangement mediated by the GTPase activity of Efl1p [240] (Fig. 3.2-8). The homology between Efl1p and ribosomal translocases further suggests that Efl1p may check on the pre-ribosomes that the binding sites for the

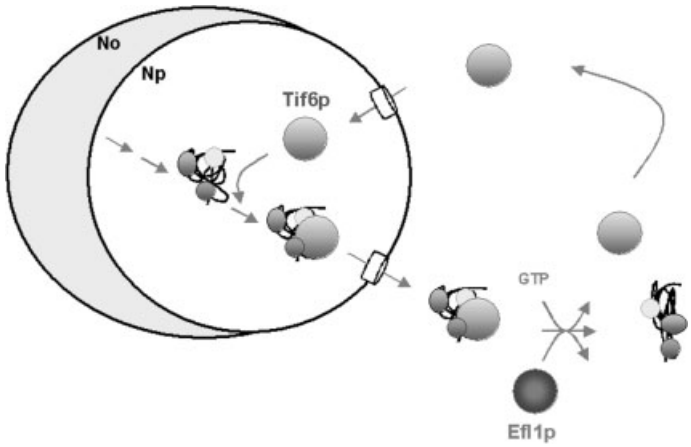


Figure 3.2-8 The cytoplasmic phase of ribosome maturation. Several RRPs follow the pre-ribosomes to the cytoplasm during the assembly process. A case is provided here for Tif6p. A structural rearrangement in cytoplasmic pre-ribosomes, mediated by the GTPase activity of Efl1p, is proposed to facilitate the release of Tif6p and its recycling to the nucle(ol)ar pre-rRNA processing machinery.

elongation factors have the proper configuration for interaction before ribosomes engage in translation. Furthermore, the nuclear exit of Tif6p has recently been shown to be dependent on phosphorylation [15]. Finally, Lsg1p/Kre35p, another cytoplasmic GTPase, may also be involved in recycling RRPs to the nucle(ol)us [118].

3.2.9
Regulatory Mechanisms, all along

Many examples of what we currently interpret as 'quality control' mechanisms have been provided here (coupling between early and late cleavages, the Dim1p dimethylation, involvement of Dhr1p in pseudoknot formation, Efl1p and late LSU structural rearrangement, Rlp's versus RPs binding, Noc's in intranuclear movements, Rix's in nuclear exit, etc.).

In most cases, 'quality control' steps potentially circumvent premature, irreversible events to occur such as to drive properly the pre-rRNPs from one assembly step to the next. To put it simply, cells have evolved complex strategies to keep the 'assembly line' in good order. In other instances, checkpoints possibly signal upstream processing events to abort the production of what would be unfaithful and non-productive synthesis. In wild-type cells, synthesis is presumably only delayed until the proper event occurs (i.e., RRP or RP binding, a specific structural rearrangement, a cleavage, modification, or transport reaction).

3.2.10
And Now ... What's Next?

The next few years will undoubtedly refine the ribosomal assembly pathway. Much attention needs to be paid to the RPs; and as mass-spectrometry techniques develop, to quantitation of the various components in distinct pre-rRNP particles.

It is probable that several dozens of novel RRPs will be identified adding to the over increasing list of such factors and that, eventually, the endoribonucleases will uncover. Their identification may, however, await the availability of *in vitro* reconstitution assays.

It is quite surprising, considering the amount of work put into the functional characterization of the snoRNAs, that we still barely have a clue to what they do in pre-rRNA processing and ribosome assembly.

A major challenge will be to try to understand what the known RRPs are doing and, as further connections between ribosome synthesis and other biosynthetic pathways unfold, it will become essential to distinguish properly the primary versus secondary effects of these trans-acting factors.

It will also become necessary to better define the relationships between the various morphological subnucle(ol)ar compartments and the biochemical reactions that occur during ribosome synthesis.

Ribosome turnover has not been properly addressed yet. (Pre)-ribosomal assembly studies indicate that RRPs probably recycled but it is presently unclear whether this also applies to some mature ribosomal components.

Recent observations are suggesting the existence, in higher eukaryotes, of a nuclear translation-like mechanism [107]. Are the particles involved fully matured, considering the essential cytoplasmic synthesis steps described in yeast – these steps are not known to occur in humans? What is the relationship, if any, between this currently ill-defined process and ribosome synthesis? Does this add a further level of complexity in the assembly process through a connection with pre-mRNA metabolism?

3.2.11 Epilogue

It is becoming more and more evident that ribosome synthesis is fully integrated with respect to most other essential cellular pathways. The importance of these connections is only starting to emerge and so far evidences have been provided for a link to transcription, pre-mRNA splicing, mRNA turnover, translation and telomere function (see above), as well as to the secretory pathway [179, 180, 280] and the cell cycle (see, e.g., Refs. [58, 246, 247, 259, 305]). This is a promising and exciting area of research for the future.

Remarkably, two proteins encoded within rDNA or rDNA-like sequences (Tar1p and *Ribin*, respectively) have recently been identified; these are transcribed in the antisense direction with respect to 25S or 25S-like sequences [43, 119]. Yeast Tar1p is a mitochondrial protein that is capable of rescuing respiration-deficient strains. Mouse *Ribin* is linked to rDNA transcription; its expression is regulated by physiological changes. These are fascinating observations suggesting stringent coevolution between these short proteins (14 and 32 kDa for Tar1p and *Ribin*, respectively) and rDNA sequences and providing compelling evidences for a high level of integration between ribosome synthesis and other biosynthetic pathways.

3.3.12 Useful WWW links

>http://www.expasy.org/linder/proteins.html

- A comprehensive list of the yeast RRPs with a short description of their known or putative functions.

>http://www.pre-ribosome.de/; http://yeast.cellzome.com/; http://genome-www.stanford.edu; http://biodata.mshri.on.ca/yeast_grid/Servlet/Search Page

- A list of physical and functional interactions between RRPs and between RRPs and proteins involved in unrelated biosynthetic pathways. These mostly rely on data sets from extensive co-immunoprecipitation and two-hybrid schemes.

>http://www.umass.edu/molvis/pipe/ribosome/opinion/index.htm

- 3D maps of rRNA modifications.

>http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html

- A most useful database of the yeast snoRNAs.

Acknowledgements

I am indebted to David Tollervey (Wellcome Trust Center for Cell Biology, Edinburgh) for expert training, constant support and countless advices. I thank R. Bordonné and E. Bertrand (CNRS, Montpellier) for allowing to quote unpublished material. The research carried out at my Lab is currently supported by Banque Nationale de Belgique, EMBO, Fonds National de la Recherche Scientifique (FNRS), Fonds pour la Recherche dans l'Industrie et l'Agriculture (FRIA), Fonds Emile Defay, Fonds Brachet, Fonds Alice and David van Buuren and Université Libre de Bruxelles.

References

1 J. D. Aitchison, M. P. Rout, *J. Cell. Biol.* **2000**, *151*, F23–6.
2 M. Aittaleb, R. Rashid, Q. Chen et al., **2003**, *10*, 256–263.
3 C. Allmang, P. Carbon, A. Krol, *RNA* **2002**, *8*, 1308–1318.
4 C. Allmang, J. Kufel, G. Chanfreau et al., *EMBO J.* **1999**, *18*, 5399–5410.
5 C. Allmang, P. Mitchell, E. Petfalski et al., *Nucleic Acids Res.* **2000**, *28*, 1684–1691.
6 C. Allmang, E. Petfalski, A. Podtelejnikov et al., *Genes Dev.* **1999**, *13*, 2148–2158.
7 C. Allmang, D. Tollervey, *J. Mol. Biol.* **1998**, *278*, 67–78.
8 J. S. Andersen, C. E. Lyon, A. H. Fox et al., *Curr. Biol.* **2002**, *12*, 1–11.
9 L. Aravind, E. V. Koonin, *Trends Biochem. Sci.* **1999**, *24*, 342–344.
10 F. Bachand, F. M. Boisvert, J. Cote et al., *Mol. Biol. Cell* **2002**, *13*, 3192–3202.
11 J. P. Bachellerie, J. Cavaille, A. Huttenhofer, *Biochimie* **2002**, *84*, 775–790.
12 J. P. Bachellerie, J. Cavaillé, L.-H. Qu: Nucleotide Modifications of Eukaryotic rRNAs: the World of Small Nucleolar RNA Guides Revisited, in eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions*, ASM, Washington, DC 2000.
13 N. Ban, P. Nissen, J. Hansen et al., *Science* **2000**, *289*, 905–920.
14 J. Bassler, P. Grandi, O. Gadal et al., *Mol. Cell* **2001**, *8*, 517–529.
15 U. Basu, K. Si, D. H., U. Maitra et al., *Mol. Cell. Biol.* **2003**, *23*, 6187–6199.
16 U. Basu, K. Si, J. R. Warner et al., *Mol. Cell. Biol.* **2001**, *21*, 1453–1462.
17 N. Bataille, T. Helser, H. M. Fried, *J. Cell. Biol.* **1990**, *111*, 1571–1582.
18 M. Beltrame, Y. Henry, D. Tollervey, *Nucleic Acids Res.* **1994**, *22*, 5139–5147.
19 M. Beltrame, D. Tollervey, *EMBO J.* **1995**, *14*, 4350–4356.
20 M. Beltrame, D. Tollervey, *EMBO J.* **1992**, *11*, 1531–1542.
21 E. Bertrand, F. Houser-Scott, A. Kendall et al., *Genes Dev.* **1998**, *12*, 2463–2468.
22 E. Billy, T. Wegierski, F. Nasr et al., *EMBO J.* **2000**, *19*, 2115–2126.
23 E. H. Blackburn, *Cell* **2001**, *106*, 661–673.
24 K. Bohmann, J. A. Ferreira, A. I. Lamond, *J. Cell. Biol.* **1995**, *131*, 817–831.
25 R. A. Borer, C. F. Lehner, H. M. Eppenberger et al., *Cell*, **1989**, *56*, 379–390.
26 M. L. Bortolin, T. Kiss, *RNA* **1998**, *4*, 445–454.
27 S. Boulon, E. Basyuk, J. M. Blanchard et al., *Biochimie* **2002**, *84*, 805–813.
28 R. C. Brand, J. Klootwijk, T. J. Van Steenbergen et al., *Eur. J. Biochem.* **1977**, *75*, 311–318.
29 M. W. Briggs, K. T. Burkard, J. S. Butler, *J. Biol. Chem.* **1998**, *273*, 13255–13263.
30 R. Brouwer, G. J. Pruijn, W. J. van Venrooij, *Arth. Res.* **2001**, *3*, 102–106.

31 N. M. Cahill, K. Friend, W. Speckmann et al., *EMBO J.* **2002,** *21,* 3816–3828.
32 J. Cavaille, K. Buiting, M. Kiefmann et al., *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 14311–14316.
33 J. Cavaille, H. Seitz, M. Paulsen et al., *Hum. Mol. Genet.* **2002,** *11,* 1527–1538.
34 J. Cavaille, P. Vitali, E. Basyuk et al., *J. Biol. Chem.* **2001,** *276,* 26374–26383.
35 J. R. Chamberlain, Y. Lee, W. S. Lane et al., *Genes Dev.* **1998,** *12,* 1678–1690.
36 S. W. Chan, E. H. Blackburn, *Oncogene* **2002,** *21,* 553–563.
37 G. Chanfreau, M. Buckle, A. Jacquier, *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 3142–3147.
38 G. Chanfreau, P. Legrain, A. Jacquier, *J. Mol. Biol.* **1998,** *284,* 975–988.
39 G. Chanfreau, G. Rotondo, P. Legrain et al., *EMBO J.* **1998,** *17,* 3726–3737.
40 D. Chen, S. Huang, *J. Cell. Biol.* **2001,** *153,* 169–176.
41 H. K. Chen, C. Y. Pai, J. Y. Huang et al., *Mol. Cell. Biol.* **1999,** *19,* 8536–8546.
42 L. F. Ciufo, J. D. Brown, *Curr. Biol.* **2000,** *10,* 1256–1264.
43 P. S. Coelho, A. C. Bryan, A. Kumar et al., *Genes Dev.* **2002,** *16,* 2755–2760.
44 A. Colley, J. Beggs, D. Tollervey et al., *Mol. Cell. Biol.* **2000,** *20,* 7238–7246.
45 B. C. d'Orval, M. L. Bortolin, C. Gaspin et al., *Nucleic Acids Res.* **2001,** *29,* 4518–4529.
46 X. Darzacq, B. E. Jady, C. Verheggen et al., *EMBO J.* **2002,** *21,* 2746–2756.
47 J. de la Cruz, D. Kressler, P. Linder, *Trends Biochem. Sci.* **1999,** *24,* 192–198.
48 J. de la Cruz, D. Kressler, D. Tollervey et al., *EMBO J.* **1998,** *17,* 1128–1140.
49 W. A. Decatur, M. J. Fournier, *Trends Biochem. Sci.* **2002,** *27,* 344–351.
50 A. M. Dechampesme, O. Koroleva, I. Leger-Silvestre et al., *J. Cell. Biol.* **1999,** *145,* 1369–1380.
51 P. P. Dennis, A. Omer, T. Lowe, *Mol. Microbiol.* **2001,** *40,* 509–519.
52 M. Deshmukh, Y. F. Tsay, A. G. Paulovich et al., *Mol. Cell. Biol.* **1993,** *13,* 2835–2845.
53 C. Dez, A. Henras, B. Faucon et al., *Nucleic Acids Res.* **2001,** *29,* 598–603.
54 C. Dez, J. Noaillac-Depeyre, M. Caizergues-Ferrer et al., *Mol. Cell. Biol.* **2002,** *22,* 7053–7065.
55 J. A. Doudna, V. L. Rath, *Cell* **2002,** *109,* 153–156.
56 F. Dragon, J. E. Gallagher, P. A. Compagnone-Post et al., *Nature* **2002,** *417,* 967–970.
57 F. Dragon, V. Pogacic, W. Filipowicz, *Mol. Cell. Biol.* **2000,** *20,* 3037–3048.
58 Y. C. Du, B. Stillman, *Cell* **2002,** *109,* 835–848.
59 D. A. Dunbar, F. Dragon, S. J. Lee et al., *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 13027–13032.
60 D. A. Dunbar, S. Wormsley, T. M. Agentis et al., *Mol. Cell. Biol.* **1997,** *17,* 5803–5812.
61 M. Dundr, T. Misteli, *Mol. Cell.* **2002,** *9,* 5–7.
62 F. Eisenhaber, C. Wechselberger, G. Kreil, *Trends Biochem. Sci.* **2001,** *26,* 345–347.
63 N. A. Eppens, A. W. Faber, M. Rondaij et al., *Nucleic Acids Res.* **2002,** *30,* 4222–4231.
64 K. T. Etheridge, S. S. Banik, B. N. Armbruster et al., *J. Biol. Chem.* **2002,** *277,* 24764–24770.
65 A. W. Faber, M. Van Dijk, H. A. Raue et al., *RNA* **2002,** *8,* 1095–1101.
66 S. Fath, P. Milkereit, A. V. Podtelejnikov et al., *J. Cell. Biol.* **2000,** *149,* 575–590.
67 A. Fatica, A. D. Cronshaw, M. Dlakic et al., *Mol. Cell.* **2002,** *9,* 341–351.
68 A. Fatica, M. Dlakic, D. Tollervey, *RNA* **2002,** *8,* 1502–1514.
69 A. Fatica, M. Morlando, I. Bozzoni, *EMBO J.* **2000,** *19,* 6218–6229.
70 A. Fatica, D. Tollervey, *Nat. Struct. Biol.* **2003,** *10,* 237–239.
71 A. Fatica, D. Tollervey, *Curr. Opin. Cell. Biol.* **2002,** *14,* 313–318.
72 W. Filipowicz, V. Pogacic, *Curr. Opin. Cell. Biol.* **2002,** *14,* 319–327.
73 L. P. Ford, J. W. Shay, W. E. Wright, *RNA* **2001,** *7,* 1068–1075.

74 M. Fromont-Racine, J. C. Rain, P. Legrain, *Meth. Enzymol.* **2002,** *350,* 513–524.
75 M. Fromont-Racine, J. C. Rain, P. Legrain, *Nat. Genet.* **1997,** *16,* 277–282.
76 T. Frugier, S. Nicole, C. Cifuentes-Diaz et al., *Curr. Opin. Genet. Dev.* **2002,** *12,* 294–298.
77 O. Gadal, D. Strauss, J. Braspenning et al., *EMBO J.* **2001,** *20,* 3695–3704.
78 O. Gadal, D. Strauss, J. Kessl et al., *Mol. Cell. Biol.* **2001,** *21,* 3405–3415.
79 O. Gadal, D. Strauss, E. Petfalski et al., *J. Cell. Biol.* **2002,** *157,* 941–951.
80 J. G. Gall, *Annu. Rev. Cell. Dev. Biol.* **2000,** *16,* 273–300.
81 P. Ganot, B. E. Jady, M. L. Bortolin et al., *Mol. Cell. Biol.* **1999,** *19,* 6906–6917.
82 C. Gaspin, J. Cavaille, G. Erauso et al., *J. Mol. Biol.* **2000,** *297,* 895–906.
83 T. Gautier, T. Berges, D. Tollervey et al., *Mol. Cell. Biol.* **1997,** *17,* 7088–7098.
84 A. C. Gavin, M. Bosche, R. Krause et al., *Nature* **2002,** *415,* 141–147.
85 T. H. Geerlings, J. C. Vos, H. A. Raue, *RNA* **2000,** *6,* 1698–1703.
86 D. Gelperin, L. Horton, J. Beckman et al., *RNA* **2001,** *7,* 1268–1283.
87 S. A. Gerbi, T. S. Lange, *Mol. Biol. Cell.* **2002,** *13,* 3123–3137.
88 G. Giese, F. Wunderlich, *J. Biol. Chem.* **1983,** *258,* 131–135.
89 C. Giorgi, A. Fatica, R. Nagel et al., *EMBO J.* **2001,** *20,* 6856–6865.
90 J. P. Girard, H. Lehtonen, M. Caizergues-Ferrer et al., *EMBO J.* **1992,** *11,* 673–682.
91 P. Grandi, V. Rybin, J. Bassler et al., *Mol. Cell.* **2002,** *10,* 105–115.
92 S. Granneman, G. J. Pruijn, W. Horstman et al., *J. Biol. Chem.* **2002,** *277,* 48490–48500.
93 H. Grosshans, K. Deinert, E. Hurt et al., *J. Cell. Biol.* **2001,** *153,* 745–762.
94 B. Guglielmi, M. Werner, *J. Biol. Chem.* **2002,** *277,* 35712–35719.
95 P. Harnpicharnchai, J. Jakovljevic, E. Horsey et al., *Mol. Cell.* **2001,** *8,* 505–515.
96 N. S. Heiss, S. W. Knight, T. J. Vulliamy et al., *Nat. Genet.* **1998,** *19,* 32–38.
97 A. Henras, Y. Henry, C. Bousquet-Antonelli et al., *EMBO J.* **1998,** *17,* 7078–7090.
98 Y. Henry, H. Wood, J. P. Morrissey et al., *EMBO J.* **1994,** *13,* 2452–2463.
99 P. Hilleren, T. McCarthy, M. Rosbash et al., *Nature* **2001,** *413,* 538–542.
100 J. H. Ho, G. Kallstrom, A. W. Johnson, *J. Cell. Biol.* **2000,** *151,* 1057–1066.
101 Y. Ho, A. Gruhler, A. Heilbut et al., *Nature* **2002,** *415,* 180–183.
102 C. Hoang, A. R. Ferre-D'Amare, *Cell* **2001,** *107,* 929–939.
103 B. Hong, J. S. Brockenbrough, P. Wu et al., *Mol. Cell. Biol.* **1997,** *17,* 378–388.
104 S. Huang, *J. Cell. Biol.* **2002,** *157,* 739–741.
105 J. M. Hughes Jr., M. Ares, *EMBO J.* **1991,** *10,* 4231–4239.
106 E. Hurt, S. Hannus, B. Schmelzl et al., *J. Cell. Biol.* **1999,** *144,* 389–401.
107 F. J. Iborra, D. A. Jackson, P. R. Cook, *Science* **2001,** *293,* 1139–1142.
108 C. Isaac, Y. Yang, U. T. Meier, *J. Cell. Biol.* **1998,** *142,* 319–329.
109 J. S. Jacobs, A. R. Anderson, R. P. Parker, *EMBO J.* **1998,** *17,* 1497–1506.
110 M. R. Jacobson, T. Pederson, *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 7981–7986.
111 S. Jakel, D. Gorlich, *EMBO J.* **1998,** *17,* 4491–4502.
112 R. Jansen, D. Tollervey, E. C. Hurt, *EMBO J.* **1993,** *12,* 2549–2558.
113 N. Jarrous, J. S. Wolenski, D. Wesolowski et al., *J. Cell. Biol.* **1999,** *146,* 559–572.
114 A. W. Johnson, *Mol. Cell. Biol.* **1997,** *17,* 6122–6130.
115 A. W. Johnson, E. Lund, J. Dahlberg, *Trends Biochem. Sci.* **2002,** *27,* 580–585.
116 K. W. Jones, K. Gorzynski, C. M. Hales et al., *J. Biol. Chem.* **2001,** *276,* 38645–38651.
117 T. Kadowaki, R. Schneiter, M. Hitomi et al., *Mol. Biol. Cell.* **1995,** *6,* 1103–1110.
118 G. Kallstrom, J. Hedges, A. Johnson, *Mol. Cell. Biol.* **2003,** *23,* 4344–4355.

119 M. Kermekchiev, L. Ivanova, *Mol. Cell. Biol.* **2001,** *21,* 8255–8263.
120 A. Khanna-Gupta, V. C. Ware, *Proc. Natl. Acad. Sci. USA* **1989,** *86,* 1791–1795.
121 T. C. King, R. Sirdeskmukh, D. Schlessinger, *Microbiol. Rev.* **1986,** *50,* 428–451.
122 T. H. King, W. A. Decatur, E. Bertrand et al., *Mol. Cell. Biol.* **2001,** *21,* 7731–7746.
123 T. H. King, B. Liu, R. R. McCully et al., *Mol. Cell.* **2003,** *11,* 425–435.
124 T. Kiss, *EMBO J.* **2001,** *20,* 3617–3622.
125 T. Kiss, *Cell* **2002,** *109,* 145–148.
126 D. J. Klein, T. M. Schmeing, P. B. Moore et al., *EMBO J.* **2001,** *20,* 4214–4221.
127 J. Klootwijk, R. C. van den Bos, R. J. Planta, *FEBS Lett.* **1972,** *27,* 102–106.
128 K. Koberna, J. Malinsky, A. Pliss et al., *J. Cell. Biol.* **2002,** *157,* 743–748.
129 D. Kressler, M. Doere, M. Rojo et al., *Mol. Cell. Biol.* **1999,** *19,* 8633–8645.
130 D. Kressler, P. Linder, J. de la Cruz, *Mol. Cell. Biol.* **1999,** *19,* 7897–7912.
131 D. Kressler, M. Rojo, P. Linder et al., *Nucleic Acids Res.* **1999,** *27,* 4598–4608.
132 A. Krol, *Biochimie,* **2002,** *84,* 765–774.
133 T. Kruiswijk, R. J. Planta, J. M. Krop, *Biochim. Biophys. Acta* **1978,** *517,* 378–389.
134 J. Kufel, C. Allmang, G. Chanfreau et al., *Mol. Cell. Biol.* **2000,** *20,* 5415–5424.
135 J. Kufel, C. Allmang, E. Petfalski et al., *J. Biol. Chem.* **2002,** *278,* 2147–2156.
136 J. Kufel, B. Dichtl, D. Tollervey, *RNA* **1999,** *5,* 909–917.
137 J. F. Kuhn, E. J. Tran, E. S. Maxwell, *Nucleic Acids Res.* **2002,** *30,* 931–941.
138 D. Lafontaine, J. Delcour, A. L. Glasser et al., *J. Mol. Biol.* **1994,** *241,* 492–497.
139 D. Lafontaine, J. Vandenhaute, D. Tollervey, *Genes Dev.* **1995,** *9,* 2470–2481.
140 D. L. Lafontaine, D. Tollervey, *Trends Biochem. Sci.* **1998,** *23,* 383–388.
141 D. L. J. Lafontaine, C. Bousquet-Antonelli, Y. Henry et al., *Genes Dev.* **1998,** *12,* 527–537.
142 D. L. J. Lafontaine, T. Preiss, D. Tollervey, *Mol. Cell. Biol.* **1998,** *18,* 2360–2370.
143 D. L. J. Lafontaine, D. Tollervey, *Nat. Rev. Mol. Cell. Biol.* **2001,** *2,* 514–520.
144 D. L. J. Lafontaine, D. Tollervey, *RNA* **1999,** *5,* 455–467.
145 D. L. J. Lafontaine, D. Tollervey: Regulatory Aspects of rRNA Modifications and Pre-rRNA Processing, in *Modification and Editing of RNA,* eds H. Grosjean and R. Benne, ASM Press, Washington, DC 1998, 281–288.
146 D. L. J. Lafontaine, D. Tollervey, *Mol. Cell. Biol.* **2000,** *20,* 2650–2659.
147 B. G. Lane, J. Ofengand, M. W. Gray, *Biochimie* **1995,** *77,* 7–15.
148 T. S. Lange, A. Borovjagin, E. S. Maxwell et al., *EMBO J.* **1998,** *17,* 3176–3187.
149 T. S. Lange, M. Ezrokhi, F. Amaldi et al., *Mol. Biol. Cell* **1999,** *10,* 3877–3890.
150 D. J. Leary, S. Huang, *FEBS Lett.* **2001,** *509,* 145–150.
151 S. J. Lee, S. J. Baserga, *Mol. Cell. Biol.* **1999,** *19,* 5441–5452.
152 S. Lefebvre, P. Burlet, Q. Liu et al., *Nat. Genet.* **1997,** *16,* 265–269.
153 H. D. Li, J. Zagorski, M. J. Fournier, *Mol. Cell. Biol.* **1990,** *10,* 1145–1152.
154 L. Lindahl, R. H. Archer, J. M. Zengel, *Nucleic Acids Res.* **1994,** *22,* 5399–5407.
155 L. Lindahl, R. H. Archer, J. M. Zengel, *Nucleic Acids Res.* **1992,** *20,* 295–301.
156 M. A. Lischwe, R. L. Ochs, R. Reddy et al., *J. Biol. Chem.* **1985,** *260,* 14304–14310.
157 P. C. Liu, D. J. Thiele, *Mol. Biol. Cell* **2001,** *12,* 3644–3657.
158 A. A. Lukowiak, A. Narayanan, Z. H. Li et al., *RNA* **2001,** *7,* 1833–1844.
159 Z. Lygerou, H. Pluk, W. J. van Venrooij et al., *EMBO J.* **1996,** *15,* 5936–5948.
160 B. E. Maden, M. Salim, *J. Mol. Biol.* **1974,** *88,* 133–152.
161 B. E. Maden, M. Salim, D. F. Summers, *Nat. New Biol.* **1972,** *237,* 5–9.
162 M. Mann, R. C. Hendrickson, A. Pandey, *Annu. Rev. Biochem.* **2001,** *70,* 437–473.

163 I. W. Mattaj, L. Englmeier, *Annu. Rev. Biochem.* **1998**, *67*, 265–306.
164 E. S. Maxwell, M. J. Fournier, *Annu. Rev. Biochem.* **1995**, *64*, 897–934.
165 M. J. McEachern, A. Krauskopf, E. H. Blackburn, *Annu. Rev. Genet.* **2000**, *34*, 331–358.
166 U. T. Meier, *J. Biol. Chem.* **1996**, *271*, 19376–19384.
167 U. T. Meier, G. Blobel, *J. Cell. Biol.* **1994**, *127*, 1505–1514.
168 U. T. Meier, G. Blobel, *Cell* **1992**, *70*, 127–138.
169 G. Meister, C. Eggert, U. Fischer, *Trends Cell. Biol.* **2002**, *12*, 472–478.
170 P. Milkereit, O. Gadal, A. Podtelejnikov et al., *Cell* **2001**, *105*, 499–509.
171 P. Milkereit, D. Strauss, J. Bassler et al., *J. Biol. Chem.* **2002**, *278*, 4072–4081.
172 Jr. O. L. Miller, B. R. Beatty, *Science* **1969**, *164*, 955–957.
173 T. Misteli, *J. Cell. Biol.* **2001**, *155*, 181–185.
174 J. R. Mitchell, J. Cheng, K. Collins, *Mol. Cell. Biol.* **1999**, *19*, 567–576.
175 J. R. Mitchell, E. Wood, K. Collins, *Nature* **1999**, *402*, 551–555.
176 P. Mitchell, E. Petfalski, A. Shevchenko et al., *Cell* **1997**, *91*, 457–466.
177 P. Mitchell, E. Petfalski, D. Tollervey, *Genes Dev.* **1996**, *10*, 502–513.
178 P. Mitchell, D. Tollervey, *Nat. Struct. Biol.* **2000**, *7*, 843–846.
179 K. Miyoshi, R. Tsujii, H. Yoshida et al., *J. Biol. Chem.* **2002**, *277*, 18334–18339.
180 K. Mizuta, J. R. Warner, *Mol. Cell. Biol.* **1994**, *14*, 2493–2502.
181 P. B. Moore, T. A. Steitz, *Nature* **2002**, *418*, 229–235.
182 M. Morlando, P. Greco, B. Dichtl et al., *Mol. Cell. Biol.* **2002**, *22*, 1379–1389.
183 J. P. Morrissey, D. Tollervey, *Trends Biochem. Sci.* **1995**, *20*, 78–82.
184 J. P. Morrissey, D. Tollervey, *Mol. Cell. Biol.* **1993**, *13*, 2469–2477.
185 J. Mouaikel, U. Narayanan, C. Verheggen et al., *EMBO Rep.* **2003**, *4*, 616–622.
186 J. Mouaikel, C. Verheggen, E. Bertrand et al., *Mol. Cell.* **2002**, *9*, 891–901.
187 E. B. Mougey, M. O'Reilly, Y. Osheim, et al., *Genes Dev.* **1993**, *7*, 1609–1619.
188 Z. Mourelatos, J. Dostie, S. Paushkin et al., *Genes Dev.* **2002**, *16*, 720–728.
189 T. I. Moy, P. A. Silver, *Genes Dev.* **1999**, *13*, 2118–2133.
190 T. I. Moy, P. A. Silver, *J. Cell. Sci.* **2002**, *115*, 2985–2995.
191 A. Narayanan, J. Eifert, K. A. Marfatia et al., *J. Cell. Sci.* **2003**, *116*, 177–186.
192 A. Narayanan, A. Lukowiak, B. E. Jady et al., *EMBO J.* **1999**, *18*, 5120–5130.
193 A. Narayanan, W. Speckmann, R. Terns et al., *Mol. Biol. Cell.* **1999**, *10*, 2131–2147.
194 S. Nicole, C. C. Diaz, T. Frugier et al., *Muscle Nerve* **2002**, *26*, 4–13.
195 T. A. Nissan, J. Bassler, E. Petfalski et al., *EMBO J.* **2002**, *21*, 5539–5547.
196 K. R. Noon, E. Bruenger, J. A. McCloskey, *J. Bacteriol.* **1998**, *180*, 2883–2888.
197 R. L. Ochs, M. A. Lischwe, W. H. Spohn et al., *Biol. Cell.* **1985**, *54*, 123–133.
198 J. Ofengand, M. J. Fournier: The Pseudouridine Residues of rRNA: Number, Location, Biosynthesis, and Function. in *Modification and Editing of RNA*, eds H. Grosjean and R. Benne, ASM Press, Washington, DC 1998, 229–253.
199 M. Ohno, A. Segref, A. Bachi et al., *Cell* **2000**, *101*, 187–198.
200 A. D. Omer, T. M. Lowe, A. G. Russell et al., *Science* **2000**, *288*, 517–522.
201 A. D. Omer, S. Ziesche, H. Ebhardt et al., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5289–5294.
202 S. L. Ooi, D. A. Samarsky, M. J. Fournier et al., *RNA* **1998**, *4*, 1096–1110.
203 S. Paushkin, A. K. Gubitz, S. Massenet et al., *Curr. Opin. Cell. Biol.* **2002**, *14*, 305–312.
204 B. A. Peculis, J. A. Steitz, *Cell* **1993**, *73*, 1233–1245.
205 B. A. Peculis, J. A. Steitz, *Genes Dev.* **1994**, *8*, 2241–2255.
206 T. Pederson, *Nucleic Acids Res.* **1998**, *26*, 3871–3876.
207 P. Pelczar, W. Filipowicz, *Mol. Cell. Biol.* **1998**, *18*, 4509–4518.

208 L. Pellizzoni, J. Baccon, B. Charroux et al., *Curr. Biol.* **2001,** *11,* 1079–1088.
209 L. Pellizzoni, J. Yong, G. Dreyfuss, *Science* **2002,** *298,* 1775–1779.
210 E. Petfalski, T. Dandekar, Y. Henry et al., *Mol. Cell. Biol.* **1998,** *18,* 1181–1189.
211 R. D. Phair, T. Misteli, *Nature* **2000,** *404,* 604–609.
212 L. Pintard, D. Kressler, B. Lapeyre, *Mol. Cell. Biol.* **2000,** *20,* 1370–1381.
213 P. W. Piper, J. A. Bellatin, A. Lockheart, *EMBO J.* **1983,** *2,* 353–359.
214 R. J. Planta, *Yeast* **1997,** *13,* 1505–1518.
215 M. Platani, I. Goldberg, J. R. Swedlow et al., *J. Cell. Biol.* **2000,** *151,* 1561–1574.
216 H. Pluk, J. Soffner, R. Luhrmann et al., *Mol. Cell. Biol.* **1998,** *18,* 488–498.
217 V. Pogacic, F. Dragon, W. Filipowicz, *Mol. Cell. Biol.* **2000,** *20,* 9028–9040.
218 N. J. Pokrywka, D. S. Goldfarb, *J. Biol. Chem.* **1995,** *270,* 3619–3624.
219 N. J. Proudfoot, A. Furger, M. J. Dye, *Cell* **2002,** *108,* 501–512.
220 J. D. Puglisi, S. C. Blanchard, R. Green, *Nat. Struct. Biol.* **2000,** *7,* 855–861.
221 L. H. Qu, A. Henras, Y. J. Lu et al., *Mol. Cell. Biol.* **1999,** *19,* 1144–1158.
222 V. Ramakrishnan, P. B. Moore, *Curr. Opin. Struct. Biol.* **2001,** *11,* 144–154.
223 T. P. Rasmussen, M. R. Culbertson, *Mol. Cell. Biol.* **1998,** *18,* 6885–6896.
224 H. A. Raue, R. J. Planta, *Prog. Nucleic Acid Res. Mol. Biol.* **1991,** *41,* 89–129.
225 G. Reimer, I. Raska, U. Scheer et al., *Exp. Cell. Res.* **1988,** *176,* 117–128.
226 J. Retel, R. C. van den Bos, R. J. Planta, *Biochim. Biophys. Acta.* **1969,** *195,* 370–380.
227 M. Ridanpaa, H. van Eenennaam, K. Pelin et al., *Cell* **2001,** *104,* 195–203.
228 G. Rigaut, A. Shevchenko, B. Rutz et al., *Nat. Biotechnol.* **1999,** *17,* 1030–1032.
229 M. P. Rout, J. D. Aitchison, A. Suprapto et al., *J. Cell. Biol.* **2000,** *148,* 635–651.
230 M. P. Rout, G. Blobel, J. D. Aitchison, *Cell* **1997,** *89,* 715–725.
231 D. Ruggero, S. Grisendi, F. Piazza et al., *Science* **2003,** *299,* 259–262.
232 D. A. Samarsky, M. J. Fournier, R. H. Singer et al., *EMBO J.* **1998,** *17,* 3747–3757.
233 H. Santos-Rosa, H. Moreno, G. Simos et al., *Mol. Cell. Biol.* **1998,** *18,* 6826–6838.
234 C. Saveanu, D. Bienvenu, A. Namane et al., *EMBO J.* **2001,** *20,* 6475–6484.
235 T. Schafer, D. Strauss, E. Petfalski et al., *EMBO J.* **2003,** *22,* 1370–1380.
236 A. Scherl, Y. Coute, C. Deon et al., *Mol. Biol. Cell.* **2002,** *13,* 4100–4109.
237 T. Schimmang, D. Tollervey, H. Kern, et al., *EMBO J.* **1989,** *8,* 4015–4024.
238 M. E. Schmitt, D. A. Clayton, *Genes Dev.* **1994,** *8,* 2617–2628.
239 R. Schneiter, T. Kadowaki, A. M. Tartakoff, *Mol. Biol. Cell.* **1995,** *6,* 357–370.
240 B. Senger, D. L. J. Lafontaine, J. S. Graindorge et al., *Mol. Cell.* **2001,** *8,* 1363–1373.
241 A. G. Seto, A. J. Zaug, S. G. Sobel et al., *Nature* **1999,** *401,* 177–180.
242 K. Sharma, D. Tollervey, *Mol. Cell. Biol.* **1999,** *19,* 6012–6019.
243 P. J. Shaw, A. F. Beven, D. J. Leader et al., *J. Cell. Sci.* **1998,** *111,* 2121–2128.
244 A. Shevchenko, D. Schaft, A. Roguev et al., *Mol. Cell. Prot.* **2002,** *1,* 204–212.
245 S. Shimba, B. Buckley, R. Reddy et al., *J. Biol. Chem.* **1992,** *267,* 13772–13777.
246 W. Shou, K. M. Sakamoto, J. Keener et al., *Mol. Cell.* **2001,** *8,* 45–55.
247 W. Shou, J. H. Seol, A. Shevchenko et al., *Cell* **1999,** *97,* 233–244.
248 K. Shuai, J. R. Warner, *Nucleic Acids Res.* **1991,** *19,* 5059–5064.
249 K. Sirum-Connolly, T. L. Mason, *Science* **1993,** *262,* 1886–1889.
250 J. Sleeman, C. E. Lyon, M. Platani et al., *Exp. Cell. Res.* **1998,** *243,* 290–304.
251 C. M. Smith, J. A. Steitz, *Mol. Cell. Biol.* **1998,** *18,* 6897–6909.
252 W. Speckmann, A. Narayanan, R. Terns et al., *Mol. Cell. Biol.* **1999,** *19,* 8412–8421.
253 W. A. Speckmann, R. M. Terns, M. P. Terns, *Nucleic Acids Res.* **2000,** *28,* 4467–4473.
254 C. Srisawat, F. Houser-Scott, E. Bertrand et al., *RNA* **2002,** *8,* 1348–1360.
255 K. Stade, C. S. Ford, C. Guthrie et al., *Cell* **1997,** *90,* 1041–1050.

256 T. Stage-Zimmermann, U. Schmidt, P. A. Silver, *Mol. Biol. Cell.* **2000,** *11,* 3777–3789.

257 E. J. Steinmetz, N. K. Conrad, D. A. Brow et al., *Nature* **2001,** *413,* 327–331.

258 A. Stevens, C. L. Hsu, K. R. Isham et al., *J. Bacteriol.* **1991,** *173,* 7024–7028.

259 A. F. Straight, W. Shou, G. J. Dowd et al., *Cell* **1999,** *97,* 245–256.

260 N. Suzuki, E. Noguchi, N. Nakashima et al., *Genetics,* **2001,** *158,* 613–625.

261 T. H. Tang, T. S. Rozhdestvensky, B. C. d'Orval et al., *Nucleic Acids Res.* **2002,** *30,* 921–930.

262 N. K. Tanner, P. Linder, *Mol. Cell.* **2001,** *8,* 251–262.

263 T. Taura, G. Schlenstedt, P. A. Silver, *J. Biol. Chem.* **1997,** *272,* 31877–31884.

264 M. P. Terns, C. Grimm, E. Lund et al., *EMBO J.* **1995,** *14,* 4860–4871.

265 M. P. Terns, R. M. Terns, *Curr. Biol.* **2001,** *11,* R862–R864.

266 M. P. Terns, R. M. Terns, *Gene Expr.* **2002,** *10,* 17–39.

267 F. Thomas, U. Kutay, *J. Cell. Sci.* **2003,** *116,* 2409–2419.

268 D. Tollervey, *EMBO J.* **1987,** *6,* 4169–4175.

269 D. Tollervey, C. Guthrie, *EMBO J.* **1985,** *4,* 3873–3878.

270 D. Tollervey, T. Kiss, *Curr. Opin. Cell. Biol.* **1997,** *9,* 337–342.

271 D. Tollervey, H. Lehtonen, R. Jansen et al., *Cell* **1993,** *72,* 443–457.

272 N. Tomasevic, B. Peculis, *J. Biol. Chem.* **1999,** *274,* 35914–35920.

273 N. Tomasevic, B. A. Peculis, *Mol. Cell. Biol.* **2002,** *22,* 4101–4112.

274 V. J. Tormey, C. C. Bunn, C. P. Denton et al., *Rheum. (Oxford)* **2001,** *40,* 1157–1162.

275 E. J. Tran, X. Zhang, E. S. Maxwell, *EMBO J.* **2003,** *22,* 3930–3940.

276 J. Trapman, R. J. Planta, *Biochim. Biophys. Acta.* **1976,** *442,* 265–274.

277 J. Trapman, J. Retel, R. J. Planta, *Exp. Cell. Res.* **1975,** *90,* 95–104.

278 A. Tremblay, B. Lamontagne, M. Catala et al., *Mol. Cell. Biol.* **2002,** *22,* 4792–4802.

279 C. R. Trotta, E. Lund, L. Kahan et al., *EMBO J.* **2003,** *22,* 2841–2851.

280 A. Tsuno, K. Miyoshi, R. Tsujii et al., *Mol. Cell. Biol.* **2000,** *20,* 2066–2074.

281 K. T. Tycowski, M. D. Shu, J. A. Steitz, *Nature* **1996,** *379,* 464–466.

282 S. A. Udem, J. R. Warner, *J. Biol. Chem.* **1973,** *248,* 1412–1416.

283 S. A. Udem, J. R. Warner, *J. Mol. Biol.* **1972,** *65,* 227–242.

284 D. Ursic, D. J. DeMarini, M. R. Culbertson, *Mol. Gen. Genet.* **1995,** *249,* 571–584.

285 D. Ursic, K. L. Himmel, K. A. Gurley et al., *Nucleic Acids Res.* **1997,** *25,* 4778–4785.

286 L. Valasek, J. Hasek, K. H. Nielsen et al., *J. Biol. Chem.* **2001,** *276,* 43351–43360.

287 A. van Hoof, P. Lennertz, R. Parker, *EMBO J.* **2000,** *19,* 1357–1365.

288 A. van Hoof, P. Lennertz, R. Parker, *Mol. Cell. Biol.* **2000,** *20,* 441–452.

289 A. van Hoof, R. Parker, *Cell* **1999,** *99,* 347–350.

290 A. van Hoof, R. R. Staples, R. E. Baker et al., *Mol. Cell. Biol.* **2000,** *20,* 8230–8243.

291 P. H. van Knippenberg: Structural and Functional Aspects of the N6, N6 Dimethyladenosines in 16S Ribosomal RNA, in *Structure, Function, and Genetics of Ribosomes,* eds B. Hardesty and G. Kramer, Springer, Berlin 1986, 412–424.

292 R. W. van Nues, J. M. Rientjes, S. A. Morre et al., *J. Mol. Biol.* **1995,** *250,* 24–36.

293 R. W. van Nues, J. Venema, J. M. Rientjes et al., *Biochem. Cell. Biol.* **1995,** *73,* 789–801.

294 D. I. Van Ryk, Y. Lee, R. N. Nazar, *J. Biol. Chem.* **1992,** *267,* 16177–16181.

295 E. Vanrobays, J. P. Gelugne, P. E. Gleizes et al., *Mol. Cell. Biol.* **2003,** *23,* 2083–2095.

296 E. Vanrobays, P. E. Gleizes, C. Bousquet-Antonelli et al., *EMBO J.* **2001,** *20,* 4204–4213.

297 M. H. Vaughan, J. R. Warner, J. E. Darnell, *J. Mol. Biol.* **1967,** *25,* 235–251.

298 J. Venema, D. Tollervey, *Yeast,* **1995,** *11,* 1629–1650.
299 J. Venema, D. Tollervey, *Annu. Rev. Gen.* **1999,** *33,* 261–311.
300 J. Venema, H. R. Vos, A. W. Faber et al., *RNA* **2000,** *6,* 1660–1671.
301 C. Verheggen, D. L. J. Lafontaine, D. Samarsky et al., *EMBO J.* **2002,** *21,* 2736–2745.
302 C. Verheggen, J. Mouaikel, M. Thiry et al., *EMBO J.* **2001,** *20,* 5480–5490.
303 J. Vilardell, J. R. Warner, *Mol. Cell. Biol.* **1997,** *17,* 1959–1965.
304 J. Vilardell, S. J. Yu, J. R. Warner, *Mol. Cell.* **2000,** *5,* 761–766.
305 R. Visintin, E. S. Hwang, A. Amon, *Nature* **1999,** *398,* 818–823.
306 C. Wang, C. C. Query, U. T. Meier, *Mol. Cell. Biol.* **2002,** *22,* 8457–8466.
307 H. Wang, D. Boisvert, K. K. Kim et al., *EMBO J.* **2000,** *19,* 317–323.
308 J. R. Warner, *J. Mol. Biol.* **1966,** *19,* 383–398.
309 J. R. Warner, *Trends Biochem. Sci.* **1999,** *24,* 437–440.
310 J. R. Warner, *Cell* **2001,** *107,* 133–136.
311 J. R. Warner, *Microbiol. Rev.* **1989,** *53,* 256–271.
312 J. R. Warner, A. Kumar, S. A. Udem et al., *Biochem. Soc. Symp.* **1973,** *37,* 3–22.
313 J. R. Warner, A. Kumar, S. A. Udem et al., *Biochem. J.* **1972,** *129,* 29P–30P.
314 J. R. Warner, S. A. Udem, *J. Mol. Biol.* **1972,** *65,* 243–257.
315 N. J. Watkins, A. Gottschalk, G. Neubauer et al., *RNA* **1998,** *4,* 1549–1568.
316 N. J. Watkins, V. Segault, B. Charpentier et al., *Cell* **2000,** *103,* 457–466.
317 T. Wegierski, E. Billy, F. Nasr et al., *RNA* **2001,** *7,* 1254–1267.
318 K. A. Wehner, S. J. Baserga, *Mol. Cell.* **2002,** *9,* 329–339.
319 K. A. Wehner, J. E. Gallagher, S. J. Baserga, *Mol. Cell. Biol.* **2002,** *22,* 7258–7267.
320 L. B. Weinstein, J. A. Steitz, *Curr. Opin. Cell. Biol.* **1999,** *11,* 378–384.
321 L. B. Weinstein Szewczak , S. J. DeGregorio, S. A. Strobel et al., *Chem. Biol.* **2002,** *9,* 1095–1107.
322 K. Weis, *Trends Biochem. Sci.* **1998,** *23,* 185–189.
323 K. Weis, *Curr. Opin. Cell. Biol.* **2002,** *14,* 328–335.
324 T. Wiederkehr, R. F. Pretot, L. Minvielle-Sebastia, *RNA* **1998,** *4,* 1357–1372.
325 J. L. J. Woolford, J. R. Warner, The Ribosome and its Synthesis, in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics,* vol. I., eds J. R. Broach, J. R. Pringle, E. W. Jones, Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1991, 587–626.
326 H. Wu, P. K. Yang, S. E. Butcher et al., *EMBO J.* **2001,** *20,* 7240–7249.
327 P. Wu, J. S. Brockenbrough, M. R. Paddy et al., *Gene* **1998,** *220,* 109–117.
328 F. Wunderlich, G. Giese, H. Falk, *Mol. Cell. Biol.* **1983,** *3,* 693–698.
329 S. Xiao, F. Scott, C. A. Fierke et al., *Annu. Rev. Biochem.* **2002,** *71,* 165–189.
330 P. K. Yang, G. Rotondo, T. Porras et al., *J. Biol. Chem.* **2002,** *277,* 45235–45242.
331 Y. Yang, C. Isaac, C. Wang et al., *Mol. Biol. Cell.* **2000,** *11,* 567–577.
332 N. I. Zanchin, P. Roberts, A. DeSilva et al., *Mol. Cell. Biol.* **1997,** *17,* 5001–5015.
333 Y. Zebarjadian, T. King, M. J. Fournier, et al., *Mol. Cell. Biol.* **1999,** *19,* 7461–7472.
334 X. Zhou, F. K. Tan, M. Xiong et al., *J. Immunol.* **2001,** *167,* 7126–7133.
335 E. Vanrobays, J. P. Gelugne, M. Caizergnes-Ferrer et al., *RNA* **2004,** *10,* 645–656.

4
tRNA and Synthetases

4.1
tRNA: Structure and Function

Viter Marquéz and Knud H. Nierhaus

4.1.1
Introduction

Even before deciphering the genetic code during the 1960s, Francis Crick had postulated, in 1956, that protein synthesis is mediated by "adaptor" RNA molecules [1]. Two years later, Hoagland et al. [2] discovered a nucleic acid fraction of low molecular weight that served as a carrier for amino acids, transporting them to the place where polypeptides are synthesized. This fraction was termed soluble RNA (sRNA) and was described as mixture of components, each with an adaptor ability for a particular amino acid [3]. Nowadays, we know that the sRNA or adaptor molecules are the transfer RNAs (tRNAs) and they are the linking factors between the RNA world, deciphering the triplet code or codon encoded in the messenger RNA (mRNA), and the protein world, because they carry amino acids to the ribosomes where they are linked together to form proteins. The specificity of deciphering results from the fact that tRNAs contain at one tip of their L-shaped tertiary structure an anticodon complementary to a specific codon and at the other tip the corresponding aminoacyl residue linked by an energy-rich ester bond ($\Delta G^{\circ\prime} = \sim$ -6 kcal mol^{-1}). It follows that the charging of a tRNA with its amino acid is the true translation step, whereas it is the astounding task of the ribosome to translate the sequence of codons of an mRNA into the corresponding protein sequence in a fast and accurate fashion (see Chaps. 8.2 and 8.3). Charging of tRNAs is performed by synthetases (aaRS), and all tRNAs that can carry the same amino acid (isoacceptor tRNAs) are usually recognized by one and the same enzyme (see the next chapter).

Many milestones in molecular biology were achieved as a product of tRNA research:

(1) binding assays of tRNA to ribosomes were essential for deciphering the genetic code [4];
(2) $tRNA^{Ala}$ from yeast was the first nucleic acid molecule, whose complete sequence was determined [5];

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

(3) $tRNA^{Phe}$ from yeast was the first RNA molecule, for which the crystal structure could be resolved [6, 7];
(4) a tRNA gene ($tRNA^{Ala}$ from yeast) was the first chemically synthesized gene that showed activity *in vivo* [8];
(5) the structural motif called a pseudoknot was first described in studies of tRNA-like structures.

In fact, for a long time tRNAs were the only RNA molecules that could be produced in large amounts and be obtained in homogenous form; thus many of the techniques used nowadays in the study of RNA and RNA–protein interactions were developed using tRNAs [9]. Additionally, the tRNA primary sequence carries information about the age of the genetic code. By comparison of sequence alignments of tRNAs, applying a method called "statistical geometry in sequence space", it was possible to draw conclusions about phylogeny (common progenitor of the kingdoms) and the origin of the genetic code. With the assumption that the kingdom separation occurred around 2.5±0.5 billion years ago, the age of the genetic code was determined to 3.8 (±0.6) billion years [10]. Now the rRNAs have replaced the tRNAs as the preferential tool to study evolutionary relationships.

The central role of tRNAs is to carry amino acids to the ribosome, but even beyond this adaptor function, tRNAs also dictate the functional states of the ribosome. The tRNA locations on the ribosome define the pre- and post-translocational states, namely the two main conformations between which ribosomes oscillate during the elongation cycle (see Chap. 8).

4.1.2 Secondary Structure

Three main species of RNA molecules exist in all living cells: rRNA, mRNA and tRNA. tRNA constitutes only 10–15% of the total RNA in *Escherichia coli*, each tRNA has a molecular weight of about 25 kDa and a relative sedimentation coefficient of 4S giving rise to the original name "4S RNA". The size of tRNAs is variable, but on average they have a length of 76 nucleotides (nt), the longest tRNA identified so far is $tRNA^{Ser}$ from *E. coli* having 93 nt whereas in nematode mitochondria very short cripple tRNAs are found lacking either the D or the TΨC stem loop (about 56 nt, see Ref. [11]).

Holley et al. [5] proposed several possible secondary structures, when the first tRNA sequence of $tRNA^{Ala}$ from yeast was obtained. However, when the yeast tRNA sequences of $tRNA^{Ser}$ [12], $tRNA^{Tyr}$ [13], and $tRNA^{Phe}$ [14] also became available, a planar cloverleaf secondary structure was the only one that best satisfied all sequences (Fig. 4.1-1). Sequence comparison of other tRNAs revealed that all tRNAs adopt the cloverleaf secondary structure. Figure 4.1-2(a) shows a graphic representation of the distribution of conserved bases in elongator tRNAs [15] on the basis of 932 sequences. Figure 4.1-2(b) shows the generally accepted numbering of the tRNA nucleotides [16].

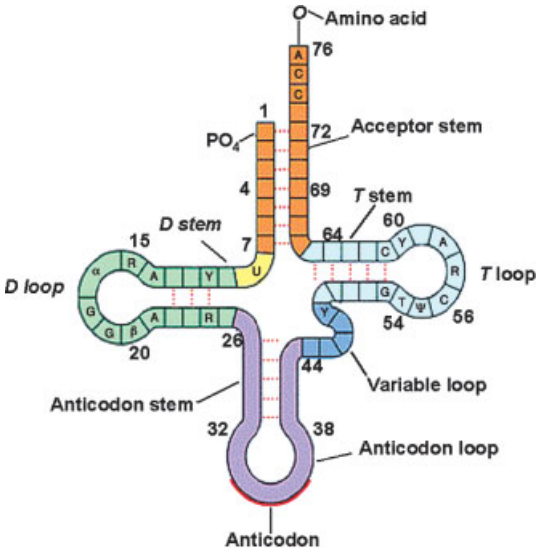


Figure 4.1-1 Cloverleaf secondary structure of tRNA. The various secondary structure motifs are shown in different colors. Abbreviations: R, purine base (A or G); Y, pyrimidine base (C or U); T, ribothymidine; Y, pseudouridine. Taken from Ref. [52] with permission.

The cloverleaf structure is characterized by three stem loops and four helices. The acceptor helix (termed acceptor stem) is formed by seven base pairs combining the 5′ and 3′-end of the molecule. In all elongator tRNAs, the first base pair of the acceptor stem is a G1-C72 Watson–Crick interaction. However, in eubacterial initiator $tRNA^{fMet}$, a mismatch between residue C1 and A72 is observed. This stem contains, at the 3′-end, a single-strand sequence N^{73}CCA-3, where N73 could be any nucleotide representing a "discriminator" base important for the specificity and efficiency of tRNA synthetase activity [17]. The universally conserved CCA sequence carries at its 3′-end the aminoacyl residue via an ester link between the ribose 2′- or 3′-OH group of the carboxyl group of the amino acid. Ten synthetases are linking "their" amino acids to the 2′-end of the ultimate A, the other 10 at the 3′-end (see Chap. 4.2). This has no direct consequence to protein synthesis, since the aminoacyl residue can trans-esterify between 2′- and 3′-OH groups with a rate of 1–10 s^{-1} [18], which might be increased and fixed at the 3′ position by elongation factor EF-Tu·GTP (M. Sprinzl, pers. comm.). This factor carries an aminoacyl-tRNA to the A-site of a ribosome (see Chap. 8). On the ribosome, the aminoacyl and peptidyl residues are linked to the 3′-OH group via an ester bond; the 2′-OH group is essential at least for the translocation of tRNAs on the ribosome from A- to the P-sites as well as from the P- to the E-sites [19, 20].

The CCA ends are the docking sites of tRNAs at the A- and P-site regions of the ribosomal peptidyl-transferase center on the 50S subunit (see Chap. 8.3). Eubacteria

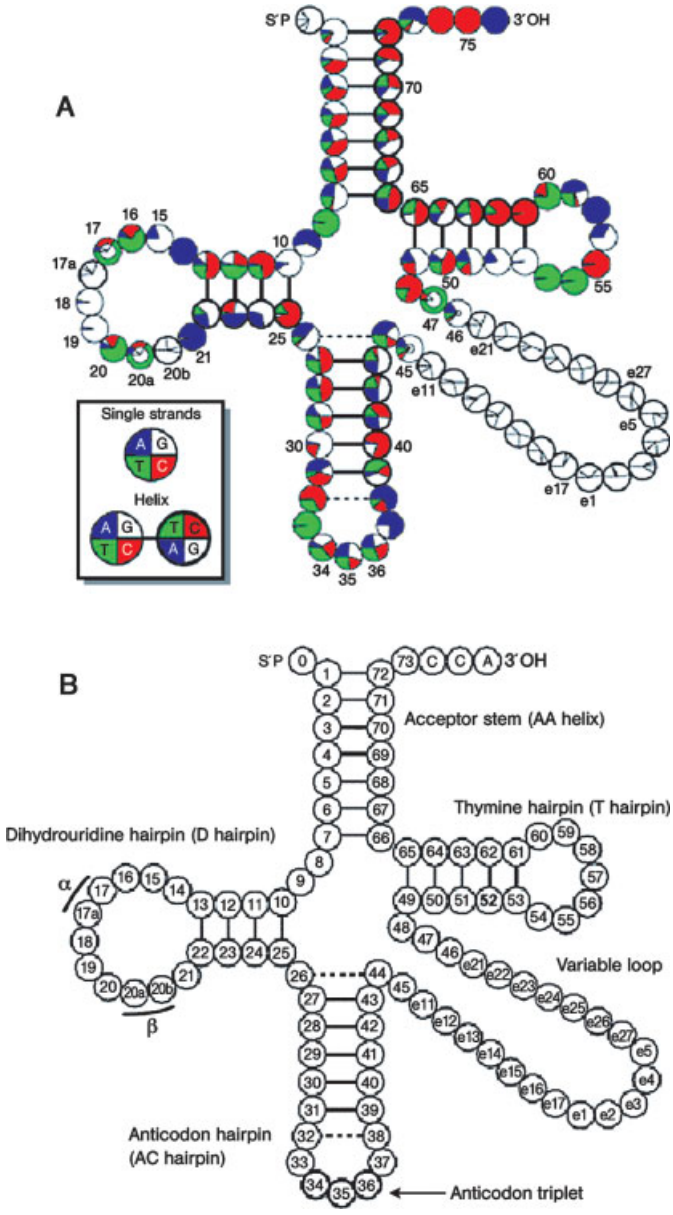


Figure 4.1-2 Features of the tRNA secondary structure. (A) Cloverleaf secondary structure of tRNA showing the distribution and conserved position of nucleotides found in all tRNA sequences known. The nucleotides are colored (A, blue; G, white; T(U), green; C, red) and their occurrence are indicated by the fractions of the circle area. According to Ref. [53] with permission, modified. (B) Conventional numbering of nucleotides according to Ref. [16], with permission.

such as *E. coli* often encode the 3′-terminal CCA in the tRNA genes, whereas in most organisms the CCA end is added post-transcriptional with a ATP (CTP): tRNA nucleotidyl-transferase or CCase [21], thus representing the most common editing mechanism. In all cells, the ATP (CTP):tRNA nucleotidyl-transferase is an essential enzyme, since it functions also in the repair of damaged CCA ends.

The second stem-loop structure is the D stem loop, where the helical region consists of 3–4 base pairs and the loop, between 8 and 11 nucleotides. The loop contains two hydrouridine bases, hence the name dihydrouridine-stem-loop for this substructure.

The anticodon stem loop at the opposite end of the molecule to the acceptor stem contains the anticodon in the middle of its loop. The loop has a universal length of 7 nt with a consensus sequence Py_{32}-U_{33}-XYZ-Pu (modified)-N_{38}, where Py represents a pyrimidine, XYZ is the anticodon, Pu a purine base and N any nucleotide. The stem always contains 5 bps and the nucleotide at position 33 of the anticodon loop is a universally conserved U in all tRNAs.

Like the anticodon stem loop, the fourth helix also comprises 5 bps with a 7 nt loop. The loop contains the sequence TΨC that gives rise to the name "TΨC-loop", where T stands for ribose-thymidine and Ψ for pseudouridine.

In addition to the defined substructures described above, a variable region exists between the T-loop and the anticodon loop, which can be anywhere between 4 and 24 nts. According to the length of this variable loop, the tRNAs have been classified (not extremely useful) as class I (4–5 nts, the vast majority) and class II (10–24 nts, $tRNA^{Leu}$, $tRNA^{Ser}$, $tRNA^{Tyr}$ in eubacteria and some organelles).

4.1.3 Tertiary Structure

The crystal structures at 3 Å resolution of yeast $tRNA^{Phe}$ [6, 7] and later $tRNA^{Asp}$ [22] confirmed that the tRNA molecule adopts an L shape. It is product of a double coaxial stacking between the acceptor stem with its CCA end and the T stem loop forming the short arm of the L arm, and the anticodon stem loop and the D stem-loop forming the long arm (Fig. 4.1-3A). In this way, the cloverleaf structure of the tRNA is transformed into two main domains: the acceptor and anticodon arms, respectively (Fig. 4.1-3B), enclosing an angle of about 90°. The extremities of both domains represent the functional "hot spots" of tRNAs: at the tip of the acceptor arm the amino acid is covalently attached, whereas at the tip of the anticodon arm, the anticodon is located, and are thus separated from each other by a distance of 75–80 Å. This is precisely the distance from the decoding center on the 30S ribosomal subunit to the peptidyl-transferase center on the 50S ribosomal subunit. Since a rigid and straight rod-shaped molecule could also fulfill the distance requirement, it raises the question as to why tRNAs have a universally conserved L shape?

The answer becomes clear when we consider other functions of the tRNAs on the ribosome. We will see in the translocation chapter (Sect. 8.4), where there is good evidence that during translocation the tRNAs are the handle to move the $tRNA_2$•mRNA complex, and thus there is a need to link the mRNAs tightly to the tRNAs via two

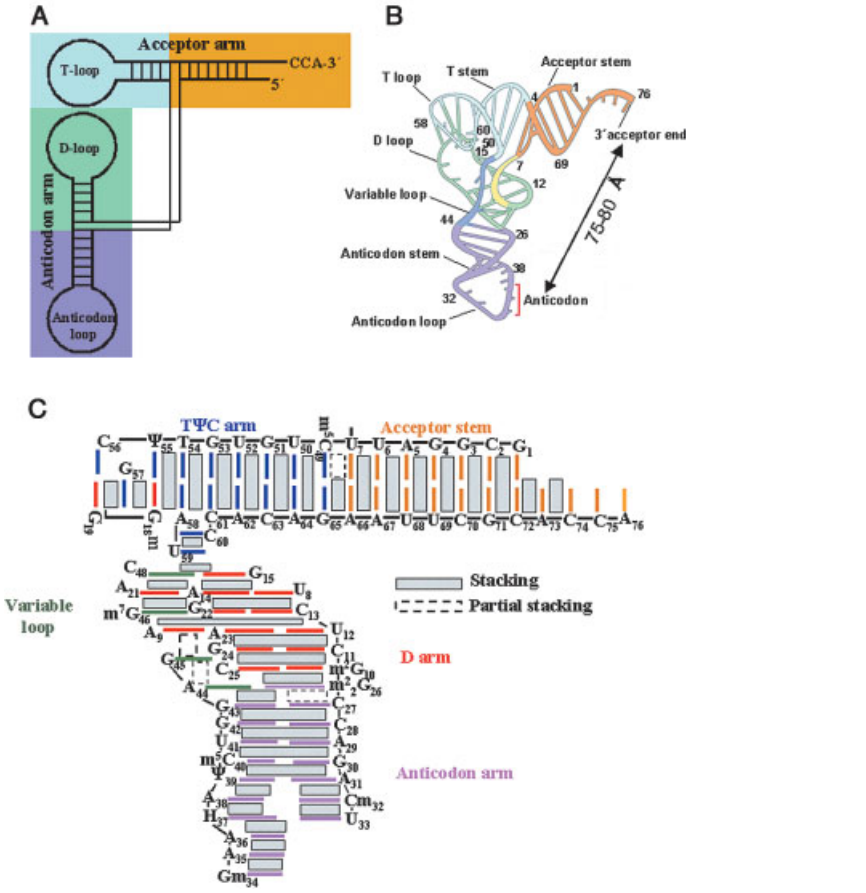


Figure 4.1-3 Features of the tRNA tertiary structure. (A) Double coaxial stackings between the anticodon stem loop and D-stem loop (anticodon arm) and between the acceptor stem and the T-stem loop (acceptor arm) transform the cloverleaf secondary structure of a tRNA into a two domain structure that include an angle of about 90°. According to Ref. [30] with permission, modified.
(B) The functional hot spots of a tRNA are the anti-codon and the CCA-3′-end that are separated by 75–80 Å in almost all canonical L-shape tRNAs. The aminoacyl or peptidyl residue is linked to the CCA-3′-end via an ester bond. This precise architecture allows codon–anticodon interaction at the decoding center of the small ribosomal subunit and contacts of the aminoacyl-(peptidyl) residue with the peptidyl transferase center on the large ribosomal subunit. From Ref. [47] with permission, modified. The colors correspond to those in (A).
(C) Base stacking (gray blocks) is an important element for the stability of the tertiary L-shaped tRNA. Colored lines indicate base interactions (mostly base pairs), the colors indicate secondary-structure motifs such as acceptor stem or D arm. Four bases (16, 17, 20, and 47) that are not involved in base stacking are not included in the figure for the sake of clarity. According to Ref. [54] with permission, modified.

adjacent codon–anticodon interactions. However, since a tRNA has a diameter of 20 Å (that of a double helix) but a codon length of only ~10 Å, it is immediately apparent that it is the L shape that prevents a steric clash of the tRNA bodies and allows simultaneous codon–anticodon interaction of both tRNAs at one end and at the same time a neighborhood of the CCA ends at the A- site and P-site regions of the peptidyl-transferase center at the other end. In fact, an angle of about 40° has been detected between the planes defined by the two L-shaped tRNAs at A- and P-sites [23–25] and about 140° between the A- and P-site codons [26].

A tRNA in solution is rigid and stable – features that are more typical for a protein than for an RNA of 76 nt. Aside from the Watson–Crick interactions seen in the cloverleaf structure and the base-stacking effects of the helices, a multitude of unusual interactions are involved in establishing the tertiary structure of a tRNA. One important feature at the tRNA elbow is the interspersed base stacking between the two main domains of the tertiary structure of tRNA: the D-arm is participating with two bases G18 and G19 in the coaxial stacking of the acceptor and the TΨC arms, and the reverse is also true: the TΨC arm participates with C60 and U59 ($tRNA^{Phe}$) in the coaxial stacking of the anticodon and D-arms (Fig. 4.1-3C). Other unusual features are listed below:

(1) At the junction of T- and D-loops, the elbow of the tRNA, a cross-strand stacking of four purines form a "base zipper structure" [27], which is important for stabilizing the "elbow architecture" of a tRNA (see Fig. 4.1-4A).

(2) Non-canonical Watson–Crick base pairs such as G:U are quite common in RNA helices (see Fig. 4.1-4B), but also more rare interactions such as the iminoG26:A44 base pair are observed (see Fig. 4.1-4C). Hoogsteen base pairs are formed when the nitrogen N7 of the imidazole ring of the adenine base is involved in the hydrogen-bonding interaction instead of the pyrimidine edge as in the normal Watson–Crick base pairing. These non-canonical base pairs are observed between the conserved residues U8 and A14 organizing the sharp turn of the D-loop (see Fig. 4.1-4D). Another example is the base pairing between the universally conserved ribothymidine residue T54 and A58 (see Fig. 4.1-4E).

(3) Non-Watson–Crick base pairs from residues located in a single-stranded region with a base pair within a helix, form base triplet interactions. For example, residue G45 of the variable loop interacts with the base 10 of base pair m^2 G10-C25 of $tRNA^{Phe}$ from yeast (Fig. 4.1-4F).

(4) The sugar-phosphate backbone participates in complex interactions to hold the tRNA compact. Single-stranded regions in the tRNA can adopt the C2′-endo configuration, although the C3′-endo configuration is found in the helical regions (A-form) [28]. Furthermore, the 2′-OH of the ribose and the oxygen of the anionic phosphates groups are involved in a series of hydrogen bonds, which contribute to the stability of the tRNA tertiary structure. For instance, the N1 of the conserved base A21 interacts with the 2′-OH of the highly conserved U8 and the 2′-OH of the residue A58 donates a proton to an anionic oxygen of the phosphate at position 60 (Fig. 4.1-4G).

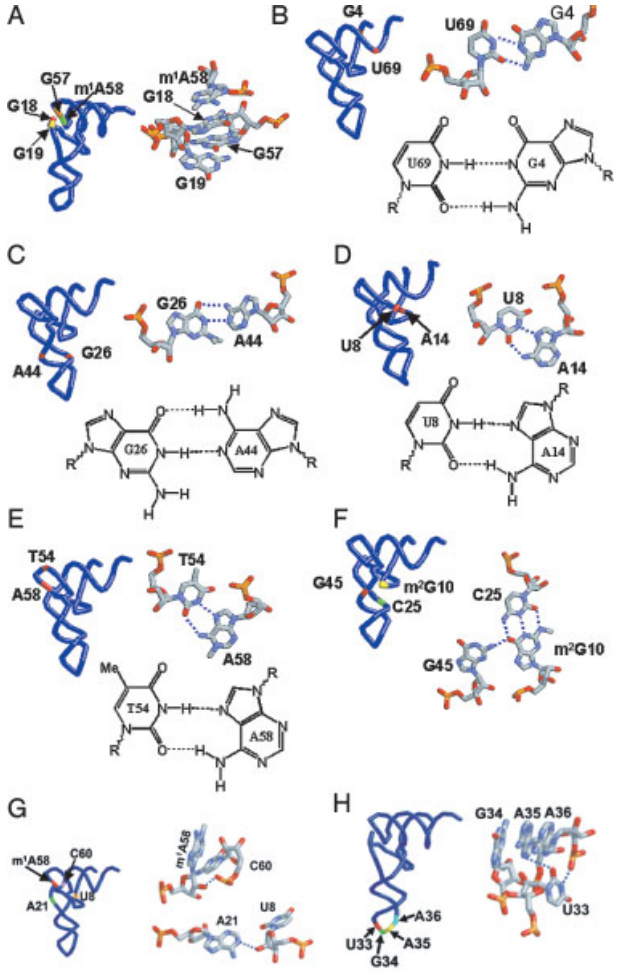


Figure 4.1-4 Unusual interactions involved in establishing the tertiary structure of tRNA. (A) "Base-zipper structure" characterized by the stacking of four purines coming from different secondary-structure motifs, viz. the D-loop (G18 and G19) and the T-loop (G57 and m^1A58); (B) cis Watson–Crick G4–U69 pair; (C) cis Watson–Crick G26–A44 inter-action (imino G-A pair); (D) trans Hoogsteen U8–A14 pair; (E) trans Hoogsteen T54-A58 pair; (F) base triplet interaction m^2G10-C25-G45; (G) hydrogen-bonding formation between sugar phosphate backbone and a nitrogen base or an anionic oxygen of a phosphate. Two examples are shown, the U8(2′OH)-A21(N1) and C60(PO)-m^1A58(2′OH) hydrogen bonds. H, the U turn around U33 is essential for an optimal presentation of the bases of the anticodon loop, the 2′OH is important for an effective translocation of the tRNA from the P- to the E-Site (see text for details). All figures were prepared using 6tna pdb files processed by RasMol software.

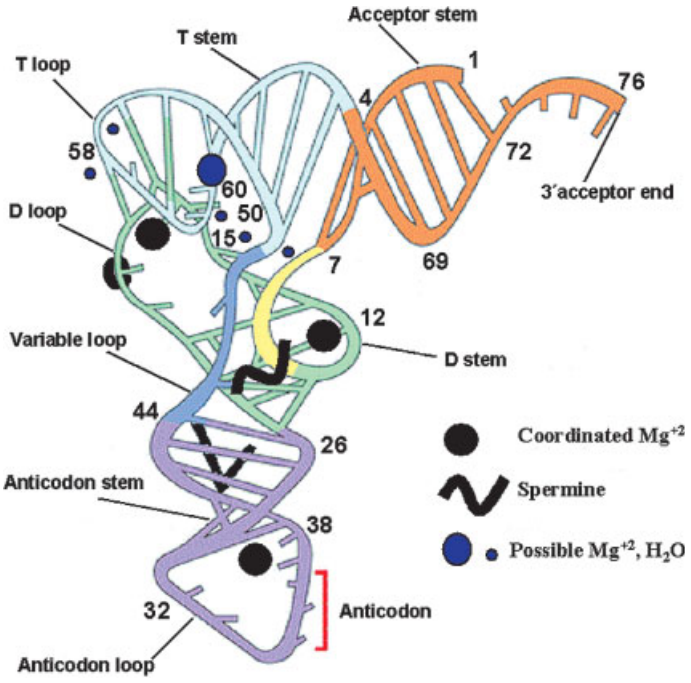


Figure 4.1-5 Coordinated Mg^{2+} ions and polyamines interact with the tRNA sugar-phosphate backbone where unusual folding of tRNA might be stabilized by these interactions. According to Ref. [55], modified.

(5) Finally, Mg^{2+} and polyamines interact with the phosphate backbone of RNA through electrostatic interactions to stabilize the folded backbone structure of the tRNA. The binding of the Mg^{2+} ion is coordinated by six molecules of water, some of which form hydrogen bonds with phosphate oxygens. The magnesium ions can also be directly coordinated by one or two phosphate oxygens and the rest of the sites can be occupied by water molecules which participate in hydrogen bonding with nitrogens or oxygens of the bases (Fig. 4.1-5).

As mentioned already, U33 in the anticodon loop is universally conserved, and this residue is of prominent importance for protein synthesis, including the reactions of tRNA binding to the P and A sites and translocation. U33 is instrumental for the "U-turn", a sharp 180° turn within a short stretch of three nucleotides in the anticodon loop that was in fact the first structural motif recognized in this loop (Figure 4.1-4H). The U-turn is characterized by two hydrogen bonds and a stacking interaction of U33 with O-P of residue 35 leading to a sharp turn in the phosphodiester bond following U33. The first hydrogen bond goes from U33 N3-H to O-P of residue 36 and the second from O2′ of U33 to N7 of A35 (or O4 of uracil or N4 of cytosine, if residue 35 is a U or C, respectively; [29]). The 2′-OH group of U33 is of outstanding

importance for P site binding and translocation, and less important for A site binding. Replacing the 2′-OH with a hydrogen or a 2′-O-methyl group reduces the P site binding by 50- and 100-fold, respectively [30, 31]. Similarly, a 2′-deoxy substitution impairs the translocation reaction by 50-fold, whereas the A site binding is reduced only by a factor of 7 [32]. U33 is also instrumental for the 3′-stack conformation of the anticodon loop, where the anticodon is stacked on the 3′-side of the loop. This seems to be the anticodon conformation being present under all binding conditions of a tRNA (see Ref. [34] for review).

4.1.4
tRNA Modifications

tRNAs are the most modified RNA molecules; almost 25% of the nucleotides of a tRNA are modified. Eighty nucleoside modifications out of a total of 95 reported modifications in all RNA molecules have been observed in tRNAs. Specific enzymes modify tRNAs during their maturation. Table 4.1-1 lists all types of known RNA modifications, and those that are found in tRNAs are indicated [34].

Usually the modification reaction is an alteration of, or addition to, existing bases in the tRNA, an exception being the base queuosine. This base is found 5′ to the anticodon at position 34 of tRNAs that read NAU or NAC codons (where N is any nucleotide), and the modification requires an enzyme that exchanges free queuosine with guanosine. Many examples of tRNA modifications include ribose/base methylations (Gm, Cm / m^5C), base isomerization (U to pseudouridine Ψ), base reduction (U to D; dihydro-uridine), base thiolation (s^2C, s^2U, s^4U) and base deamination (inosine). Some modifications are conserved features of all tRNA molecules (D residues that give rise to the name of the D-arm, Ψ found in the TΨC sequence).

Several functions have been attributed to tRNA modifications such as tertiary structure stabilization, increase in the specificity of the tRNA for its cognate tRNA synthetase, increase in the surface exposure of the tRNA, increase of interaction with initiation factors and elongation factors and involvement in the decoding of the RNA on the ribosome. The most direct effect of modification is seen in the anti-codon. Inosine, which is generated by deamination of adenine (I, instead of A), is often present at the first position of the anticodon (pairs with third position of the codon), where it is capable of pairing with any one of three bases: U, C, and A (wobble position). Curiously, although the inosine base is derived from adenine, its behavior is most similar to that of guanine in terms of potential base-pairing formation.

4.1.5
Recognition of tRNA by tRNA synthetase: Identity Elements

An important feature warranting the transfer of the genetic information is seen in the bridge function that a tRNA has, i.e., serving as linker between the RNA world and the protein world. Although codon:anticodon is the key interaction that directs the translational process, the correct recognition and aminoacylation of the tRNA by its cognate tRNA synthetase is of comparable importance. Mis-charging of tRNAs

can lead to the incorporation of the wrong amino acid into the polypeptide, thus impairing the fidelity of the translational process, which can lead to production of inactive or worse, toxic protein products.

Table 4.1-1 Abbreviations and chemical names of modified nucleosides found in RNA
Data taken from Ref. [34].

	Symbol	Name	Found in tRNA
1	Am	2′-*O*-methyladenosine	+
2	m^2A	2-Methyladenosine	+
3	m^6A	N^6-methyladenosine	+
4	$m^6{}_2A$	N^6,N^6-dimethyladenosine	
5	m^6Am	N^6,2′-O-dimethyladenosine	
6	$m^6{}_2Am$	N^6, N^6,2′-O-trimethyladenosine	
7	ms^2m^6A	2-Methylthio-N^6-methyladenosine	+
8	i^6A	N^6-isopentenyladenosine	+
9	ms^2i^6A	2-Methylthio-N^6-isopentenyladenosine	+
10	io^6A	N^6-(*cis*-hydroxyisopentenyl)-adenosine	+
11	ms^2io^6A	2-Methylthio-N^6-(*cis*-hydroxyisopentenyl)-adenosine	+
12	g^6A	N^6-glycinylcarbamoyladenosine	+
13	t^6A	N^6-threonylcarbamoyladenosine	+
14	m^6t^6A	N^6-methyl-N^6-threonylcarbamoyladenosine	+
15	ms^2t^6A	2-Methylthio-N^6-threonylcarbamoyladenosine	+
16	hn^6A	N^6-hydroxynorvalylcarbamoyladenosine	+
17	ms^2hn^6A	2-Methylthio-N^6-hydroxynorvaylcarbamoyladenosine	+
18	Ar(p)	2′-*O*-ribosyladenosine (phosphate)	+
19	m^1A	1-Methyladenosine	+
20	I	Inosine	+
21	Im	2′-*O*-methylinosine	
22	m^1I	2′-*O*-methylinosine	+
23	m^1Im	1,2′-*O*-dimethylinosine	+
24	Um	2′-*O*-methyluridine	+
25	s^2U	2-Thiouridine	+
26	s^2Um	2-Thio-2′-*O*-methyluridine	+
27	m^3U	3-Methyluridine	
28	m^3Um	3,2′-*O*-dimethyluridine	
29	acp^3U	3-(3-Amino-3-carboxypropyl)uridine	+
30	s^4U	4-Thiouridine	+
31	m^5U	Ribosylthymine	+
32	m^5Um	5,2′-*O*-dimethyluridine	+
33	m^5s^2U	5-Methyl-2-thiouridine	+

	Symbol	Name	Found in tRNA
34	ho^5U	5-Hydroxyuridine	+
35	mo^5U	5-Methoxyuridine	+
36	cmo^5U	Uridine 5-oxyacetic acid	+
37	$mcmo^5U$	Uridine 5-oxyacetic acid methyl ester	+
38	cm^5U	5-Caboxymethyluridine	
39	mcm^5U	5-Methoxycarbonylmethyluridine	+
40	mcm^5Um	5-Methoxycarbonylmethyl-2′-*O*-methyluridine	+
41	mcm^5s^2U	5-Methoxycarbonylmethyl-2-thiouridine	+
42	ncm^5U	5-Carbamoylmethyluridine	+
43	ncm^5Um	5-Carbamoylmethyl-2′-*O*-methyluridine	+
44	chm^5U	5-(Carboxyhydroxymethyl)uridine	+
45	$mchm^5U$	5-(Carboxyhydroxymethyl)uridinemethyl ester	+
46	nm^5s^2U	5-Aminomethyl-2-thiouridine	+
47	mnm^5U	5-Methylaminomethyluridine	+
48	mnm^5s^2U	5-Methylaminomethyl-2-thiouridine	+
49	mnm^5se^2U	5-Methylaminomethyl-2-selenouridine	+
50	$cmnm^5U$	5-Carboxymethylaminomethyluridine	+
51	$cmnm^5Um$	5-Carboxymethylaminomethyl-2′-*O*-methyluridine	+
52	$cmnm^5s^2U$	5-Carboxymethylaminomethyl-2-thiouridine	+
53	D	Dihydrouridine	+
54	m^5D	Dihydroribosylthymine	
55	Ψ	Pseudouridine	+
56	Ψm	2′-*O*-methylpseudouridine	+
57	$m^1\Psi$	1-Methylpseudouridine	+
58	$m^3\Psi$	3-Methylpseudouridine	
59	m^1acp Ψ	1-Methyl-3-(3-amino-3-carboxypropyl)pseudouridine	
60	Gm	2′-*O*-methylguanosine	+
61	m^1G	1-Methylguanosine	+
62	m^2G	N^2-methylguanosine	+
63	$m^2{}_2G$	N^2,N^2-dimethylguanosine	+
64	m^2Gm	N^2,2′-*O*-dimethylguanosine	+
65	$m^2{}_2Gm$	N^2,N^2,2′-*O*-trimethylguanosine	+
66	Gr(p)	2′-*O*-ribosylguanosine (phosphate)	+
67	m^7G	7-Methylguanosine	+
68	$m^{2,7}G$	N^2,7-dimethylguanosine	
69	$m^{2,2,7}G$	N^2,N^2,7-trimethylguanosine	
70	imG	Wyosine	+
71	mimG	Methylwyosine	+
72	OHyW*	Undermodified hydroxywybutosine	+

	Symbol	Name	Found in tRNA
73	yW	Wybutosine	+
74	OHyW	Hydroxywybutosine	+
75	o_2yW	Peroxywybutosine	+
76	Q	Queuosine	+
77	oQ	Epoxyqueuosine	+
78	GalQ	Galactosyl-queuosine	+
79	manQ	Mannosyl-queuosine	+
80	$PreQ_0$	7-Cyano-7-deazaguanosine	+
81	gQ(G+)	Archaeosine (alternate name 7-formamidino-7-deazaguanosine)	+
82	$PreQ_1$	7-Aminomethyl-7-deazaguanosine	+
83	Cm	2′-*O*-methylcytidine	+
84	m^4C	*N4*-methylcytidine	
85	m^4Cm	*N4*,2′-*O*-dimethylcitidine	
86	ac^4C	*N4*-acetylcytidine	+
87	ac^4Cm	*N4*-acetyl-2′-*O*-methylcytidine	+
88	m^5C	5-Methylcytidine	+
89	m^5Cm	5,2′-*O*-dimethylcytidine	+
90	hm^5C	5-Hydroxymethylcytidine	
91	f^5C	5-Formylcytidine	+
92	f^5Cm	2′-O-methyl-5-formylcytidine	+
93	m^3C	3-Methylcytidine	+
94	s^2C	2-Thiocytidine	+
95	k^2C	Lysidine	+

There is a distinct synthetase that recognizes every tRNA that participates in the decoding of the same amino acid, and this group of tRNAs are termed “isoaccepting” tRNAs. For such a situation to exist, the isoaccepting tRNAs must carry identical signals for the recognition of their synthetase. These common signals define the recognition identity of the isoacceptor tRNAs and accordingly represent the “identity elements”, which have been discovered in the past 15 years by many groups (see Table 4.1-2). An identity element or positive element is defined as a recognition site on the tRNA that allows the unique aminoacylation by its cognate aaRSs. Table 4.1-2 gives a survey of the currently known identity elements described in Ref. [35].

An interesting case is the so-called “negative determinants”: these modifications do not improve the recognition by the cognate synthetase (i.e., the corresponding, correct synthetase), but prevent or impair recognition by a non-cognate synthetase and thus mis-charging by a non-cognate synthetase [35].

Table 4.1-2 Identity elements in tRNAs aminoacylated by class I (A) and class II (B) synthetases
Data taken from Ref. [35]

	E. coli[a]	***S. cerevisiae***[b]	***T. thermophilus***[c]	**Others**[d]
(A) *Aminoacylated by class I synthetases*				
Val	(a) **A73**	***A73***		
	G3:C70, **U4:A69**	–		
	(b) A35, C36	***A35***		
Ile	(a) **A73**			
	C4:G69			
	(b)**L**/G34, A35, U36	**I**34,**A35**,**U36**		
	t^6**A37**, **A38**			
	(c) **U12:A23**, **C29:G41**			
Leu	(a) **A73**	***A73***		***A73***
				C3:G70, **A4:U69**
				G5:C68
	(b) –	***A35***		
		G37	–	
	(c) **U8•A14**			***C20a***
Met (fMet)	(a) **A73**	A73		
	(**G2:C71**,**C3:G70**)			
	U4:A69,A5:U68			
	(b) **C34**,**A35**,**U36**	**C34**,**A35**,**U36**		
	(**C32**,**U33**,**A37**)	**& the 4 other AC loop nts**		
	(c)	**D-arm**		
Cys	(a) U73	***U73***		
	G2:C71,C3:G70			
	(b) G34,*C35,A36*			
	(c) **G15•G48**,**A13•A22**			
Tyr	(a) A73	**A73**		
		C1:G72		**C1:G72**
	(b) U35	*G34,Ψ35*		
Trp	(a) G73			G73
	A1:U72,*G2:C71*			A1:U72
	G3:C70			G5:C68,A9
	(b) C34,C35,A36	C24,C35		C34,C35,C36

	E. coli[a]	*S. cerevisiae*[b]	*T. thermophilus*[c]	Others[d]
Glu	(a) **G1:C72,U2:A71** (b) s[4]**U34**,U35 ***A37*** (c) **U11:A24** **U13:G22-A46**, Δ**47**			
Gln	(a) **G73** **U1:A72**, G2:C71 G3:C70 (b) **Y34,U35,G36** **A37,U38** (c) **G10**			
Arg	(a) A/G73 (b) **C35,U/G36** (c) **A20**	**C35,U/G36**		
(B) *Aminoacylated by class II synthetases*				
Ser	(a) *G73* *C72,*G2:C71, A3:U70 *C11:G24,***R4:Y69** (c) *C11:G24* Variable loop	Variable loop		G73 Variable loop
Thr	(a) **G1:C72,C2:G71** (b) **G34,G35,U36**	**G1:C72** G35,U36	U73 **G1:C72,U3:A70** G35,U36	
Pro	(a) **A73** G72 (b) **G35,G36** (c) **G15•C48**			**G35,G36**
Gly	(a) U73 G1:C72,C2:G71 G3:C70 (b) C35,**C36** (c)	**A73** **C2:G71, G3:C7O** **C35,C36**	**U73** **G1:C72, C2:G71** **(G3:C70)** C35,C36 **(G10:C25)**	**A73** **C2:G71**
His	(a) C73 **G-1** (b) anticodon	**A73** **G-1** **G34,U35**		

	E. coli[a]	*S. cerevisiae*[b]	*T. thermophilus*[c]	Others[d]
	(a) **G73**	**G73**	**G73**	
	G2:C71			
Asp	(b) **G34,U35,C36**	**G34,U35,C36**	**G34,U35,C36**	
	C38	**C38**	**C38**	
	(c) **G10**	**G10•U25**	**G10**	
	(a) A73			
Lys	(b) **U34**,U35,**U36**			
	(mnm^{5}s^{2}U)34			
Asn	(a) *G73*			
	(b) *G34,U35,U36*			
	(a) *A73*	**A73**	**A73**	**A73**
	(b) *G34,A35,A36*	**G34,A35,A36**	**G34,A35,A36**	**G34,A35,A36**
Phe	G27:C43,G28:C42	i^{6}A37		**G30:C40**
	(c) *U20*	**G20**		**A31:U39,G20**
	G44,U45,U59,U60			
	(a) **A73**			
Ala	G2:C71,G3•U70	**G3•U70**		**G3•U70**
	G4:C69			
	(c) **G20**			

The tRNAs are listed according to the synthetase classification in two classes with subclasses. Identity elements are classified according to their location in the amino acid accepting stem (a), anticodon region (b), and other tRNA domains (c). Identity nucleotides in bold were identified by the *in vitro* approach, those in italics by the *in vivo* approach, and those in normal scripts by both approaches; when underlined, the identity element is the modified nucleotide. Numbering of residues is according to Ref. [16] and nomenclature of modified nucleotides according to Ref. [56]. In the case of base pair, (:) denotes WC pair, (:) non-WC pairs, and (-) tertiary pairs; (/) indicates that two residues can be identity elements at the same position. R, purine; Y, pyrimidine.

4.1.6
Is the tRNA Cloverleaf Structure a Pre-requisite for the L-shape?

The answer to this question is "no", since there are a number of structures that mimic the tertiary structure of an L-shaped tRNA but do not contain the canonical secondary cloverleaf structure. These variant tRNA structures can be recognized and aminoacylated by the cognates aaRSs.

Examples are tRNA-like molecules such as bacterial tmRNA (= 10Sa) that contain about 350 nt (cf. 75 nt for a tRNA –>4 × larger!), but 5′- and 3′-regions form a tRNA-like structure without T stem loop and the anticodon loop. The tmRNA contains an mRNA "module" that codes for about 9–30 amino acids; it is charged by alanine-tRNA synthetase (AlaRS) [36]. This RNA displays both tRNA and mRNA functions

(hence the name tmRNA) and plays an important role in recycling 70S ribosomes that are stuck at the 3′-end of fragmented mRNAs lacking a stop codon. The mRNA part of the tmRNA encodes an oligopeptide sequence that is tagged onto the incomplete polypeptide, targeting it for rapid degradation (see Sect. 8.2.5)

Another example is the tRNA-like structure in 5′-untranslated region of *thrS* mRNA (regulatory domain of threonyl-tRNA synthetase gene), which is recognized by ThrRS [37]. The tRNA-like structure at the 3′-end of the RNA from the turnip yellow mosaic virus (TYMV) can be charged by ValRS [38] and seems to play a role during replication of the virus [39]

Here we should also mention the "crippled" tRNAs that are found in the mitochondria of nematodes [40, 11], where the T-loop has been reduced to few base pair or deleted completely, whereas the $tRNA^{Ser}$ in this organelle is lacking the D stem loop instead of the T-loop, and also possesses its own EF-Tu factor. Human $tRNA^{Ser}$ with the anticodon GCU is another example where T-stem is not present at all, similar to nematode $tRNA^{Ser}$ [41].

4.1.7
Other Functions of tRNA outside the Ribosomal Elongation Cycle

In addition to the main role of the tRNAs in protein synthesis during the ribosomal elongation cycle, tRNAs are also involved in a series of other reactions beyond protein synthesis.

1. Viral reverse transcriptase of the human immunodeficiency virus (HIV) uses $tRNA_3^{Lys}$ as a primer for the synthesis of DNA [42].
2. Some tRNAs induce the formation of anti-termination structures of the nontranslated region upstream (UTR) of the structural genes of some amino acid operons (*ilv-leu*, *his, trp*) and of some tRNA synthetase genes (*thrS*, *tyrS, lueR, pheS*). Under starvation conditions, deacylated tRNA seems to base-pair via its anticodon and the NCCA-3′-end with complementary sequences in the leader UTR promoting the formation of anti-termination structures in these systems (reviewed in Ref. [43]).
3. Under nutrient deprivation conditions, bacterial cells down-regulate the transcription of genes that belong to the fields of molecular genetics such as replication, transcription, and translation. This most important regulation circuit in bacteria is called the "stringent response" and is mediated by the synthesis of (p)ppGpp. Binding of a deacylated tRNA to the ribosomal A-site activates via the ribosomal protein L11 the ribosome-bound enzyme, RelA, which synthesizes the signaling molecule (p)ppGpp [44] (see Chapter 11.2.3).
4. Glu-$tRNA^{Glu}$ is an activated intermediate in the biosynthetic pathway of δ-aminolevulinate (ALA), a tetrapyrrole precursor of porphyrins in plants and bacteria. The ALA biosynthesis starts with the aminoacylation of $tRNA^{Glu}$ by GluRS, then a NADPH-dependent reduction reaction catalyzed by glutamyl-tRNA reductase occurs on Glu-$tRNA^{Glu}$ to yield glutamate 1-semialdehyde. Finally, the amino group of glutamate 1-semialdehyde is transferred to its terminal carbon by an intramolecular reaction catalyzed by a specific aminotransferase forming ALA [45].

5. Amino acid residues from aminoacyl-tRNAs are used in a cross-linking reaction that occurs during peptidoglycan synthesis of the bacterial cell wall. By this reaction, the pentapeptide moieties attached to the *N*-acetyl muramic acid residue of both the disaccharides, *N*-acetyl muramic acid-*N*-acetyl glucosamine units, become covalently bound [46, 47].

6. RNA polymerase III activity in silkworm depends on several transcription factors. Among these transcription factors, TFIIIR stands out since it contains a $tRNA^{Ile}$ with the anticodon IAU, where I stands for inosine [48].

7. Aminoacyl-tRNAs are involved in a proteolytic pathway, the so-called the "N-end rule" pathway. The N-end pathway governs the half-life of a protein in a cell with respect to the identity of its N-terminal amino acid residue. For instance, arginine-tRNA-protein transferase (R-transferase) is an enzyme that uses Arg-$tRNA^{Arg}$ to "arginylate" polypeptides whose N-terminal residue is Asp or Glu in bacteria and Cys in mammals. This arginylation is the signal for the proteolytic machinery for the protein degradation [49].

8. Many tRNA-like structures are specifically aminoacylated and participate actively in protein synthesis. An example of such structures is found at the 3′-end of the genome of several plant viral RNAs. In the case of 3′-untranslated region of turnip yellow mosaic virus (TYMV), a tRNA-like structure aminoacylated with valine is involved in virus replication and indispensable for virus viability [34]. More recently, a tRNA-like molecule (sRNA85) was identified in the trypanosomatid signal recognition particle in addition to the canonical 7SL RNA homolog. The complex has an *S*-value of ~14S and binds to the ribosomes [50].

4.1.8
Human Neurodegenerative Disorders Associated with Mitochondrial tRNAs

In mammals, many diseases are known that are caused by tRNA defects in the mitochondria. Often they are related to human neurodegenerative disorders. Table 4.1-3 summarizes mutations in human mitochondrial tRNA genes associated with the corresponding disease or phenotype.

Table 4.1-3 Disease-related mutations in human mitochondrial tRNA genes
Data taken from Ref. [51]

	tRNA mutation			
Amino acid specificity	**Gene mutation**	**Domain**	**Position**	**Related pathologies**
Ala	A5628G	AC stem	31–39	CPEO
Asn	A5692G	AC loop	38	CPEO
	C5698T	AC loop	32	PEO
	C5703T	AC stem	27–43	CPEO, MM
Asp	A7543G	AC stem	29–41	MS

tRNA mutation

Amino acid specificity	Gene mutation	Domain	Position	Related pathologies
Cys	A5814G	D stem	13–22	EM, MELAS, PEO
Gln	C4332T	acc. stem	3–70	EM, D
	A4336G	acc. stem	7–66	ADPD
	instT4370	AC loop	After 31	MM, CD
Glu	A14709G	AC loop	37	MM, EM, D
Gly	T9997C	acc. stem	7–66	MHCM
	A10006G	D loop	18	CIPO
	T10010C	D stem	12–23	EM
	A10044G	T loop	59	EM
Ile	A4269G	acc. stem	7–66	FICP, EM
	T4274C	D stem	13–22	CPEO
	T4285C	AC stem	27–43	PEO
	G4298A	AC stem	30–40	CPEO
	G4309A	T stem	51–63	CPEO
	A4317G	T loop	59	FICP
	C4320T	T stem	52–62	ECM
Leu (CUN)	T12297C	AC loop	33	DCM
	G12301A	AC loop	37	AISA
	G12415A	T stem	52–62	CPEO
	A12320G	T loop	57	MM
Leu (UUR)	A3243G	D loop	14	MELAS, DMDF
	A3243T	D loop	14	PEM, MM
	G3249A	D loop	19	KS
	T3250C	D loop	20	MM
	A3251G	D loop	20:01	MM
	A3252G	D loop	21	MELAS
	C3254G	D stem	12–23	MM
	C3256T	D stem	10–25	MERRF-like, MELAS
	T3258C	AC stem	27–43	LA, E1
	A3260G	AC stem	29–41	MMC
	T3264C	AC loop	33	DM
	T3271C	AC stem	30–40	MELAS, DM
	delT3272	AC stem	29–41	PEM
	T3273C	AC stem	28–42	O: EI
	C3275A	Var. region	44	LHON
	A3280G	T stem	49–65	MM
	A3288G	T loop	57	MM

tRNA mutation				
Amino acid specificity	**Gene mutation**	**Domain**	**Position**	**Related pathologies**
	T3291C	T loop	60	MELAS
	A3302G	acc. stem	2–71	MM
	C3303T	acc. stem	1–72	MMC
Lys	A8296G	acc. stem	2–71	DMDF, MERRF
	G8313A	D stem	12–24	MNGIE
	T8316C	AC stem	27–43	MELAS
	G8328A	AC stem	31–39	EM
	G8342A	T stem	53–61	PEO, MS
	A8344G	T loop	55	MERRF
	T8355C	T stem	50–64	PEO, SM
	T8356C	T stem	49–65	MERRF
	T8362G	acc. stem	2–71	SM
	G8363A	acc. stem	1–72	MICM, D, MERRF, LS
Met	T4409C	acc./D stem	8	MM
	G4450A	T stem	53–61	MM
Phe	G583A	acc. Stem	7–66	MELAS
	A606G	AC stem	29–41	M
	T618C	AC stem	29–41	MM
Pro	T15965C	T stem	50–64	ADPD
	G15990A	AC loop	36	MM, O
Ser (AGY)	C12246A	T loop	55	CIPO
	C12258A	acc. stem	7–66	DMDF
Ser (UCN)	insG7472	Var. region	46	PEM
	C7497T	D stem	13–22	MM, PEM, RRF, LA
	A7511G	acc. stem	4–69	DEAF, SNHL
	A7512G	acc. stem	3–70	PEM
Thr	G15915A	AC stem	30–40	MM
	A15923G	AC loop	38	LIMM
	delT15940	T loop	60	MM
	G15950A	acc. stem	3–70	ADPD
Trp	G5521A	D stem	10–25	MM
	insT5537	AC stem	After 27	MILS
	G5540A	AC stem	30–40	PEM, CD
	G5549A	AC stem	31–39	DEMCHO, D, A
Tyr	A5874G	D stem	13–22	EI, LW, CD
Val	G1606A	acc. stem	5–68	AMDF

tRNA mutation

Amino acid specificity	Gene mutation	Domain	Position	Related pathologies
	G1642A	AC stem	27–43	MELAS
	G1644T	Var. region	45	LS

tRNA genes are listed by amino acid specificity in alphabetical order. "Gene mutation" refers to the nucleotide substitution and position of the mutation in human mt genome. "tRNA mutation" refers to the location of the mutation in the gene product. Structural domains affected by the mutations refer to loops and stems, with AC for anticodon; acc., acceptor; Var., variable. Nucleotide numbering is according to classical tRNA numbering [16]. Pathologies are abbreviated as follows:
A, ataxia;
ADPD, Alzeimer's disease and Parkinsons disease;
AISA, acquired idiopathic sideroblastic anemia;
AMDF, ataxia, mental deterioration, deafness;
CD, Cox deficiency;
CIPO, chronic instestinal pseudoobstruction with myopathy;
CPEO, chronic progressive external ophtalmoplegia;
D, diabetes;
DCM, dilated cardiomyopathy;
DEAF, maternally inherited deafness or aminoglycoside-induced deafness;
DEMCHO, DEMentia;
Chorea;
DM, diabetes mellitus;
DMDF, diabetes mellitus, Deafness;
ECM, encephalocardiomyopathy;
EI, exercise intolerance;
EM, encephalomyopathy;
FICP, fatal infantile cardiomyopathy plus a melas-associated cardiomyopathy;
HCM, hypertrophic cardiomyopathy;
KS, Kearns–Sayre syndrome;
LA, lactic acidose,
LHON, leber hereditary
optic neuropathy;
LIMM, lethal infantile mitochondrial myopathy;
LS, leigh syndrome;
LW, limb weakness;
M, myoglobinuria;
MELAS, mitochondrial encephalomyopathy, lactis acidose, Stroke-like episodes;
MERRF, myoclonic epilepsy and ragged red muscle fibers;
MHCM, maternally inherited hypertrophic cardiomyopathy;
MICM, maternal, inherited cardiomyopathy;
MILS, maternal inherited leigh syndrome;
MM, mitochondrial myopathy;
MMC, maternal myopathy and cardiomyopathy;
MNGIE, mitochondrial neurogastrointestinal encephalomyopathy;
MS, myoclonic seizures;
O, ophtalmoplegia;
PEM, progressive encephalomyopathy;
PEO, progressive external ophtalmoplegia;
SM, skeletal myopathy;
SNHL, sensorineural hearing loss.

References

1 F. H. C. Crick, in *The Structure of Nucleic Acids and their Role in Protein Synthesis,* ed. E. M. Crook, Cambridge University Press London 1957.
2 M. B. Hoagland, P. C. Zamecnik, M. L. Stephenson, *BBA* **1957,** *24,* 215.
3 G. Zubay, *J. Mol. Biol.* **1962,** *4.*
4 M. Nirenberg, P. Leder, *Science* **1964,** *145,* 1399–1407.
5 R. Holley, J. Apgar, G. Everett et al., *Science* **1965,** *147,* 1462–1465.
6 S. H. Kim, F. L. Suddath, G. J. Quigley et al., *Science* **1974,** *185,* 435–440.
7 J. D. Robertus, J. E. Ladner, J. T. Finch et al., *Nature* **1974,** *250,* 546–551.
8 H. G. Khorana, *Science* **1979,** *203,* 614–625.
9 U. L. RajBhandary, D. Söll, In: *tRNA: Structure, Biosynthesis, and Functions,* eds D. Söll and U. L. RajBhandary, ASM Washington, DC 1995, 1–4.
10 M. Eigen, B. F. Lindemann, M. Tietze et al., *Science* **1989,** *244,* 673–678.
11 Y. I. Watanabe, H. Tsurui, T. Ueda et al., *J. Biol. Chem.* **1994,** *269,* 22902–22906.
12 H. G. Zachau, H. Dutting, H. Feldmann et al., *Cold Spring Harbor Symp. Quant. Biol.* **1966,** *31,* 417–424.
13 J. T. Madison, G. A. Everett, H. K. Kung, *Cold Spring Harbor Symp. Quant. Biol.* **1966,** *31,* 409–416.
14 U. L. RajBhandary, A. Stuart, R. D. Faulkner et al., *Cold Spring Harbor Symp. Quant. Biol.* **1966,** *31,* 425–434.
15 P. Auffinger, E. Westhof In: *Modification and Editing of RNA,* eds H. Grosjean and R. Benne, ASM, Washington, DC 1998.
16 M. Sprinzl, C. Horn, M. Brown et al., *Nucleic Acids Res.* **1998,** *26,* 148–153.
17 P. Schimmel, L. R. de Pouplana, *Cell* **1995,** *81,* 983–986.
18 M. Taiji, S. Yokoyama, T. Miyazawa, *Biochemistry* **1985,** *24,* 5776–5780.
19 E. G. Wagner, P. C. Jelenc, M. Ehrenberg et al., *Eur. J. Biochem.* **1982,** *122,* 193–197.
20 T. Wagner, M. Sprinzl, *Biochemistry* **1983,** *22,* 94–98.
21 M. A. Augustin, A. S. Reichert, H. Betat et al., *J. Mol. Biol.* **2003,** *328,* 985–994.
22 D. Moras, M. B. Comarmond, J. Fischer et al., *Nature* **1980,** *288,* 669–674.
23 R. K. Agrawal, C. M. T. Spahn, P. Penczek et al., *J. Cell. Biol.* **2000,** *150,* 447–459.
24 K. H. Nierhaus, J. Wadzack, N. Burkhardt et al., *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 945–950.
25 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001,** *292,* 883–896.
26 G. Z. Yusupova, M. M. Yusupov, J. H. Cate et al., *Cell* **2001,** *106,* 233–241.
27 T. Hermann, D. J. Patel, *J. Mol. Biol.* **1999,** *294,* 829–849.
28 Z. Sharbarova, A. Bogdanov, In: *Advance Organic Chemistry of Nucleic Acids,* ed. T. Kellersohn VCH Publishers, Weinheim New York, NY 1994, 335–388.
29 G. J. Quigley, A. Rich, *Science* **1976,** *194,* 796–806.
30 S. S. Ashraf, G. Ansari, R. Guenther et al., *RNA* **1999,** *5,* 503–511.
31 U. von Ahsen, R. Green, R. Schroeder et al., *RNA* **1997,** *3,* 49–56.
32 S. S. Phelps, O. Jerinic, S. Joseph, *Mol. Cell* **2002,** *10,* 799–807.
33 S. M. Stagg, M. Valle, R. K. Agrawal, J. Frank et al., *RNA* **2002,** *8,* 1093–1094.
34 Y. Motorin, H. Grosjean, *tRNA Modification* ed. N. P. Group, Nature Publishing Group, London 1999.
35 R. Giegé, M. Sissler, C. Florentz, *Nucleic Acids Res.* **1998,** *26,* 5017–5035.
36 Y. Hou, P. Schimmel, *Nature* **1988,** *333,* 140–145.
37 M. Graffe, J. Dondon, J. Caillet et al., *Science* **1992,** *255,* 994–996.
38 M. Pinck, P. Yot, F. Chapeville et al., *Nature* **1970,** *226,* 954–956.

39 R. Giegé, C. Florentz, T. W. Dreher, *Biochemie* **1993,** *75,* 569–582.
40 T. Ohtsuki, G. Kawai, K. Watanabe *FEBS Lett.* **2002,** *514,* 37–43.
41 G. Dirheimer, G. Keith, P. Dumas et al., In: *tRNA: Structure, Biosynthesis and Function* eds D. Söll and U. L. RajBhandary, ASM, Washington, DC 1995.
42 A. T. Das, B. Klaver, B. Berkhout *J Virol* **1995,** *69,* 3090–3097.
43 P. Gollnick, P. Babitzke, *BBA* **2002,** *1577,* 240–250.
44 T. M. Wendrich, G. Blaha, D. N. Wilson et al., *Mol. Cell* **2002,** *10,* 779–788.
45 C. G. Kannangara, S. P. Gough, P. Bruyant et al., *Trends Biochem. Sci.* **1988,** *13,* 139–143.
46 R. Plapp, J. L. Strominger, *J. Biol. Chem.* **1970,** *245,* 3667–3674.
47 S. S. Hegde, T. E. Shrader, *J. Biol. Chem.* **2001,** *276,* 6998–7003.
48 H. M. Dunstan, L. S. Young, K. U. Sprague, *Mol. Cell. Biol.* **1994,** *14,* 3588–3595.
49 A. Varshavsky, *Proc. Natl Acad. Sci. USA* **1996,** *93,* 12142–12149.
50 L. Liu, H. Ben-Shlomo, Y. Xu et al., *J. Biol. Chem.* **2003,** *278,* 18271–18280.
51 C. Florentz, *Biosci. Rep.* **2002,** *22,* 81–98.
52 E. Goldman, *Transfer RNA,* ed. N. P. Group, Nature Publisher Group London 1999.
53 E. Westhof, P. Auffinger, *Transfer RNA Structure,* ed. N. P. Group, Nature Publishing Group London 1999.
54 S. H. Kim, *Prog. Nucleic Acid Res. Mol. Biol.* **1976,** *17,* 181–216.
55 G. J. Quigley, M. M. Teeter, A. Rich, *Proc. Natl. Acad. Sci. USA,* **1978,** *75,* 64–68.
56 P. A. Limbach, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1994,** *22,* 2183–2196.

4.2
Aminoacylations of tRNAs: Record-keepers for the Genetic Code

Lluís Ribas de Pouplana and Paul Schimmel

4.2.1
Introduction

Aminoacyl-tRNA synthetases (ARS) catalyze aminoacylation reactions and therefore are essential components of the genetic code [1, 2]. These enzymes aminoacylate each transfer RNA with its cognate amino acid, thus establishing the amino acid–trinucle-otide relationships of the code. Each synthetase recognizes its specific amino acid and all its isoacceptor tRNAs. The reaction takes place in two steps:

$$E + AA + ATP \rightarrow E(AA\text{-}AMP) + PPi$$
$$E(AA\text{-}AMP) + tRNA \rightarrow E + AA\text{-}tRNA + AMP$$

First, the enzyme (E) activates the cognate amino acid by condensing it with ATP to form a transient aminoacyl adenylate (AA-AMP) that remains bound to the enzyme's active site. Secondly, the enzyme catalyzes the formation of an ester linkage between the carboxyl group of the amino acid and a hydroxyl of the ribose of the terminal 3′ adenosine of the tRNA. The aminoacylated tRNAs (AA-tRNAs) are then recognized by translation factors that place them in the ribosome's active site, where protein synthesis takes place.

Every cell requires a synthetase for each of the 20 amino acids of the genetic code. Thus, all cells contain at least 20 synthetases (eukaryotic cellular organelles use an additional set of synthetases) [3]. With minor exceptions, all aminoacyl-tRNA synthetases with the same amino acid specificity are orthologs. For example, all extant aspartyl-tRNA synthetases (the enzymes responsible for aminoacylating $tRNA^{Asp}$ with aspartate) are related to a single ancestor, which has been conserved throughout all speciation events since the last universal common ancestor (LUCA) of all organisms [4].

The concept of LUCA refers to the biological entity that constituted the genetic basis of all extant forms of life. That such a common ancestor existed follows from the universal distribution and composition of the genetic code and its components. Any gene that is found in all extant living species plausibly has an origin that precedes LUCA (this assumption would be false in the case of later-appearing genes transferred to all living species through lateral gene transfer). Most individual ARS are universally distributed. In many cases, the phylogenetic tree derived from their sequences coincides with the evolutionary tree derived from 16S RNA sequences [5, 6]. Hence, most individual synthetases probably predate the separation of the three kingdoms of life – archaea, bacteria, and eukarya.

In addition to the individual evolutionary history of each synthetase, the 20 known ARSs are divided into two classes of homologs, each containing 10 enzymes [7–9]. Each class is identified by a common active site fold, and by certain sequence motifs,

shared by all its members [2, 10]. All enzymes of each class evolved from a common ancestor, which gave rise to the extant 10 types through gene duplication events. Since most ARS may be older than LUCA, most of these duplications took place before LUCA.

Interestingly, a few synthetases had not completely evolved at the time of LUCA [11]. These enzymes are the outcome of branches of the synthetase evolutionary tree that were only fixed in evolution after the separation of the three main branches of life. A particularly interesting subset of this late evolution is that related to the endosymbiotic events that generated mitochondria and chloroplasts. The cohabitation of organelle genomes within eukaryotic cells resulted in selection of new recognition mechanisms between tRNAs and ARS [11]. Among the forces behind these selections is the requirement to preserve faithful recognition of two independent sets of tRNAs. Simultaneously, the recognition mechanisms between tRNAs and ARS might have been influenced by the significant reduction in genome size observed in animal mitochondria. The analysis of these exceptional ARSs provides information about the origin of extant cells after the separation of the three branches of life, as well as about the events that determined the selection of their extant phenotypes [12, 13].

Thus, the origin of aminoacyl-tRNA synthetases is ancient, predating the appearance of the last common ancestor to all living species. The intimate functional link between synthetases and the genetic code suggests a common evolutionary pathway. The evolutionary history of these enzymes can be inferred from structural, sequence, and phylogenetic comparisons. In turn, every aspect of ancient ARS evolution that is solved invariably provides information about general aspects of the origin of life. In this regard, aminoacyl-tRNA synthetases can be used as markers of essential transitions in evolution. Current understanding of the evolution of aminoacyl-tRNA synthetases and the relationship to the development of the code is summarized here.

4.2.2 The Operational RNA Code

The aminoacyl-tRNA synthetases are the actual translators of the genetic code. Their faithful recognition of cognate tRNAs ensures the correct coupling of triplet sequences and amino acids. The recognition of the tRNA molecules by these enzymes depends on the specific interactions between the proteins and identity elements present in the tRNA sequences and structures (see Refs. [23, 25] for a review of identity elements of tRNAs). In some instances, however, the identity elements of tRNAs recognized by ARS do not include the anticodon bases of the tRNA, and are located in the tRNA acceptor stem (Fig. 4.2-1).

For instance, a major identity element of $tRNA^{Ala}$ is a single G:U base pair at the 3:70 position in the acceptor stem of that tRNA [14, 15]. This single G3:U70 base pair is necessary and sufficient to convey alanine acceptance upon many tRNA sequences. Alanine-tRNA synthetase (AlaRS) does not recognize the anticodon region of its cognate tRNA [16].

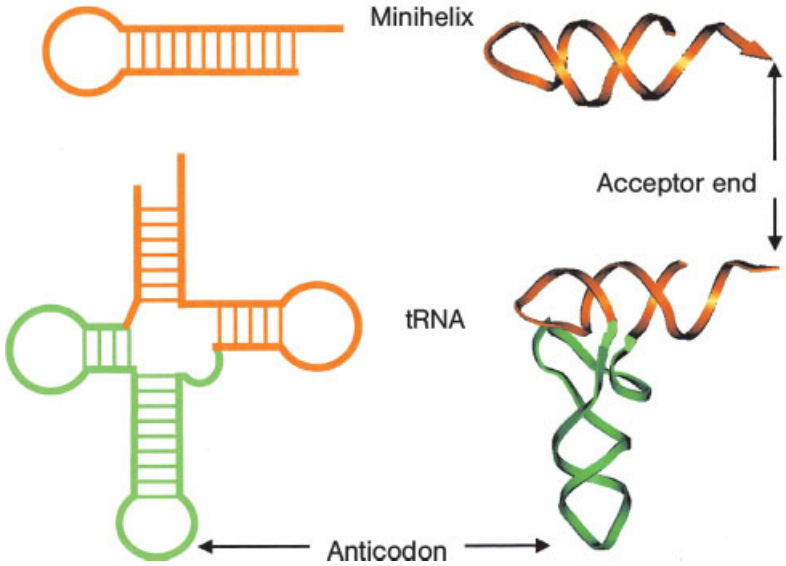


Figure 4.2-1 Secondary and tertiary structure models of an RNA minihelix and a tRNA. The minihelix dumbbell and the corresponding region of the tRNA structure are colored in orange. The rest of the tRNA structure is colored green.

A second example of an ARS that does not recognize the anticodon region is seryl-tRNA synthetase (SerRS). The crystal structure of *Thermus thermophilus* SerRS, complexed with $tRNA^{Ser}$, revealed that this enzyme does not interact with the anticodon triplet [8, 17]. The main identity element in the acceptor stem of $tRNA^{Ser}$ is the discriminator base G73 [18, 19]. The elbow region of the tRNA is also recognized by the enzyme through interactions with an idiosyncratic coiled-coil domain at the N-terminus of SerRS [17].

In these cases, the relationship between an anticodon triplet and an amino acid is indirect, because tRNA recognition is achieved via identity elements embedded in the acceptor stems and not in the anticodons. This set of interactions is known as the operational RNA code [20]. It relates specific RNA sequences/structures in acceptor stems to specific amino acids [20–22]. Indeed, small RNA helices that recapitulate acceptor stems are charged with specific amino acids.

At least 10 synthetases have been shown to aminoacylate specifically RNA minihelices that are based on the acceptor stems of their cognate tRNAs (Fig. 4.2-1) [19, 22–32]. Although the efficiency of aminoacylation of these minimalist structures can be significantly decreased with respect to the full-length tRNA, these minihelices are specifically recognized and aminoacylated by their cognate synthetases. Thus, the set of interactions that constitute the operational RNA code today are possibly the molecular remnants of the set of identity elements that ruled the recognition of the molecular ancestors of tRNA (minihelices).

The operational RNA code might be a testimony of earlier times in genetic code evolution, when RNA stem-loop structures (precursors of modern tRNAs) were aminoacylated by ribozymes in primitive peptide synthesis mechanisms [20, 33, 34]. The modern tRNA shape has been proposed to arise from the fusion of such RNA minihelices, causing in the process the translocation of the early identity signals to the anticodons and acceptor stems of modern tRNAs [34–39]. Thus, identity elements in the acceptor stem and the genetic code are directly linked through the evolutionary history of tRNA [36].

4.2.3 Extant Aminoacyl-tRNA Synthetases

Aminoacyl-tRNA synthetases (ARS) are classified into two distinct structural families: class I and class II [7–9, 40, 41]. Of the twenty aminoacyl-tRNA synthetases, 10 are found in each family (Table 4.2-1) [9]. All the enzymes in each class evolved from a unique single-domain protein that evolved into the active-site characteristic of each class [20, 42]. The only known exception is lysyl-tRNA synthetase (LysRS), which exists as a class I or as a class II enzyme in different organisms [43].

Genetic dissections showed that aminoacyl-tRNA synthetases developed from their ancestral catalytic cores through the addition of domains and insertions [44, 45]. Crystallographic studies confirmed this scheme [2, 10, 46]. The active-site domain recognizes the acceptor stem end of the tRNA, where the amino acid is attached. Most of the class I and II enzymes also recognize the anticodon stem-loop structure of their cognate tRNAs using additional domains that are idiosyncratic to each enzyme [47].

Table 4.2-1 Classes and subclasses of aminoacyl-tRNA synthetases

Subclass	Class I	Class II	Subclass
	LeuRS	AlaRS	
	IleRS	GlyRS	
	ValRS	ThrRS	
Ia	MetRS	SerRS	IIa
	CysRS	ProRS	
	ArgRS	HisRS	
	GluRS	AspRS	
	GlnRS	AsnRS	
Ib	LysRS	LysRS	IIb
	TyrRS	PheRS	
Ic	TrpRS		IIc

a The classification of ARS is based on sequence and structural information.

All members of class I ARS share an active-site domain that forms a Rossmann nucleotide-binding fold. Members of class II have an active-site domain that contains an unusual anti-parallel β-sheet flanked by two long α-helices [2, 10, 46]. The two folds are fundamentally different. Thus, the two classes evolved from two distinct ancestors. Nevertheless, each class is unlikely to have evolved independently, because the composition and tRNA-binding mechanisms of each class are related and complementary (see below).

Within each class, structural classifications further divide the members into three distinct subclasses (Table 4.2-1) [2]. The amino acids recognized by the enzymes in each subclass are chemically related. Each subclass within class I has a matching subclass in class II that recognizes similar amino acids and contains a similar number of enzymes. Class I ARS are subdivided into three subclasses: Ia, Ib, and Ic. Subclass Ia contains enzymes that are specific for the amino acids leucine, isoleucine, valine, methionine, cysteine, and arginine. Subclass Ib enzymes recognize glutamate, glutamine and lysine. Subclass Ic ARS are specific for tyrosine and tryptophan [2].

Similarly, class II enzymes are subdivided into subclasses IIa, IIb, and IIc. Subclass IIa enzymes are specific for serine, threonine, glycine, alanine, proline, and histidine. Subclass IIb enzymes recognize aspartate, asparagine, and lysine. Subclass IIc contains the enzyme specific for phenylalanine [47]. The active-site domains of class I ARS bind the tRNA from the minor groove side of the acceptor stem. In contrast, the active-site domains of class II ARS bind to the tRNA acceptor stem but, in this case, class II enzymes approach the tRNA molecule from its major groove side. These interactions relate acceptor stem sequences/structures to specific amino acids.

These structural observations are consistent with early work showing that most class I ARS (which bind the minor groove of the tRNA acceptor stem) attach their respective amino acid to the 2′–OH group of the terminal ribose of tRNA, whereas most class II ARS (which bind the major groove of the tRNA acceptor stem) attach the amino acid to the 3′–OH [48, 49]. Important exceptions are synthetases that bind aromatic residues. Tyrosyl- and tryptophanyl-tRNA synthetases (TyrRS and TrpRS) are class Ic enzymes, but they bind the tRNA from the major groove side and indistinctly catalyze the attachment of the amino acid to the 2′– or 3′–OH of the tRNA [48, 49]. On the other hand, phenylalanyl-tRNA synthetase (PheRS), a class IIc enzyme, binds the tRNA on the minor groove side and catalyzes the attachment of the amino acid to the 2′–OH [48, 49]. These exceptions give strong support to the 'symmetrical model' for the origin of the two ARS classes (see below).

In addition to their aminoacylation activity, several subclass Ia enzymes possess an editing activity to prevent misacylation of their cognate tRNAs. Valyl-, leucyl-, and isoleucyl-tRNA synthetases activate cognate amino acids that are difficult to discriminate from stereochemically similar ones [50]. In these enzymes, the hydrolysis of noncognate aminoacyl adenylates or misacylated tRNAs is catalyzed by an independent domain [51]. This editing domain is inserted into the catalytic domain for aminoacylation, thereby creating a separate active site [51, 52]. Class II

enzymes glycyl-, alanyl-, prolyl-, and threonyl-tRNA synthetases (GlyRS, AlaRS, ProRS, and ThrRS) also contain editing activities [53–58]. In the cases of AlaRS, ThrRS, and ProRS, these activities are localized to domains that are appended to the catalytic unit of the enzyme, rather than inserted into the active site [56–58]. Those editing domains are completely different in structure than those found in class I enzymes [54].

4.2.4 The Origin of Aminoacyl-tRNA Synthetase Classes: Two Proteins bound to one tRNA

The homology between extant groups of universal ARS implies that several rounds of gene duplication and divergence took place before LUCA to give rise to most of the enzymes that constitute each class. Given that the role of ARS is intrinsically linked to the development of the genetic code, this observation is consistent with the genetic code being completely defined by the time of LUCA.

Because class I and class II synthetases evolved symmetrically to generate two families that display striking similarities and complementarities, their early evolution was probably driven by common constraints. These evolutionary forces shaped the symmetrical nature of the two classes. Based on the analysis of the crystal structures of complexes between ARS and tRNAs, and the structure of the genetic code, a proposal has been put forward that can explain this feature of the ARS classes [59, 60].

The 'symmetry theory' for the origin of the two classes of ARS proposes that the two classes evolved from an ancestral complex where a single tRNA molecule was recognized simultaneously by a class I and a class II ancestor [59]. The extant subclasses would have originated from duplications of the genes coding for these two proteins. This scenario can explain several features displayed by extant ARS. For example, the equivalence in sizes of the two classes, and their subclasses, would result from coupled evolution. Thus, each event of gene duplication and divergence that generated a new tRNA species was followed by the duplication and divergence of the genes coding for the class I- and class II-type active-site domains (Fig. 4.2-4). This process would result in an equivalent numbers of class I and class II ARS. Similarly, the association of a class I and a class II ARS active site with a given tRNA can explain why the synthetases resulting from the evolution of this initial complex recognize sterically similar residues (see below).

The 'symmetry theory' requires that formation of a complex between a single tRNA and two ARS be sterically possible. The association of two extant ARS on a single tRNA would be prevented by steric clashes caused by domains that surround the enzyme's active sites. However, the ARS ancestors were small proteins that contained only the active-site domain [42]. To investigate the possibility that two ancestral ARS active-site domains formed a complex with a single tRNA molecule, the two-synthetase–one-tRNA interactions were modeled using available crystallographic data [59].

The structures of ARS–tRNA complexes were edited to obtain the coordinates of each tRNA bound only to the respective active site domain [52, 54, 61–67]. The

available structures cover at least one representative from each subclass. (Owing of close similarities between enzymes of the same subclass, the mode of binding to the acceptor stem is thought to be the same for each subclass member.) The structures for all possible subclass Ia–c subclass IIa–c pairs bound to tRNA were individually generated. The resulting structures were inspected for steric compatibility of the bound active-site domains. Not all superimpositions generated sterically compatible models. Several pairs, Similar to that of AspRS (subclass IIb) and IleRS (subclass Ia), generated severe steric clashes between large parts of the respective active sites [59].

Several superimpositions generated compatible pairs where two synthetases cover the tRNA acceptor stem without major steric clashes. Remarkably, these pairs link together specific ARS subclasses. In particular, the only combinations that accommodated all enzymes followed exactly a pairing of subclasses. Thus, subclass Ia enzymes (IleRS or ValRS) pair best with subclass IIa enzymes (SerRS or ThrRS). A subclass Ib enzyme (GlnRS) forms a compatible pair with a subclass IIb enzyme (AspRS). Finally, TyrRS (subclass Ic) can only form a compatible pair with PheRS (subclass IIc) [59] (Fig. 4.2-2).

Large translational and rotational differences between the different pairs (with respect to the axis of the tRNA acceptor stem) are an important feature of these complexes. The differences are particularly evident in the Ic–IIc pair (TyrRS and PheRS), which binds the tRNA acceptor stem at a 90° angle of rotation with respect to the other pairs (Fig. 4.2-2). Thus, ancestral ARS pairs have large variations in their orientations around the tRNA acceptor stem [59].

This analysis supported the idea that the two extant classes of synthetases can be interpreted as a consequence of an early interaction of specific synthetase pairs in complex with tRNA. Among the correct predictions derived from the 'symmetry theory' was the assignment of class I LysRS to the subclass Ib prior to the crystallographic evidence. LysRS is an exception among ARS in that it can be found as a class I or class II enzyme (see below). Class I LysRS were discovered recently, in certain archaebacteria and bacteria. Based on the symmetrical pairings of synthetases that recognize similar residues, we predicted that this new enzyme would be a member of the subclass Ib [59]. The crystal structure of the complex between a class I LysRS and $tRNA^{Lys}$ confirmed our prediction, and showed that the complete structures of a class I and class II LysRSs can form a complex around a single $tRNA^{Lys}$ molecule, with almost no steric hindrance [68] (Fig. 4.2-3).

Moreover, the 'symmetry theory' can also explain the uncanny sequence similarities that have been observed between tRNAs that are charged by ARS of opposite classes (i.e., $tRNA^{Tyr}$ and $tRNA^{Phe}$, or $tRNA^{Asp}$ and $tRNA^{Glu}$). The symmetrical pairs may have formed initially to cover and protect the acceptor stem, in a hostile environment where the structure of RNA was susceptible to chemical degradation or denaturation, or where the ester link between the tRNA molecule and its attached amino acid was particularly labile.

Interestingly, the 'symmetry theory' links the evolution of the two ARS families to the development of the genetic code. If the distribution of the two ARS classes

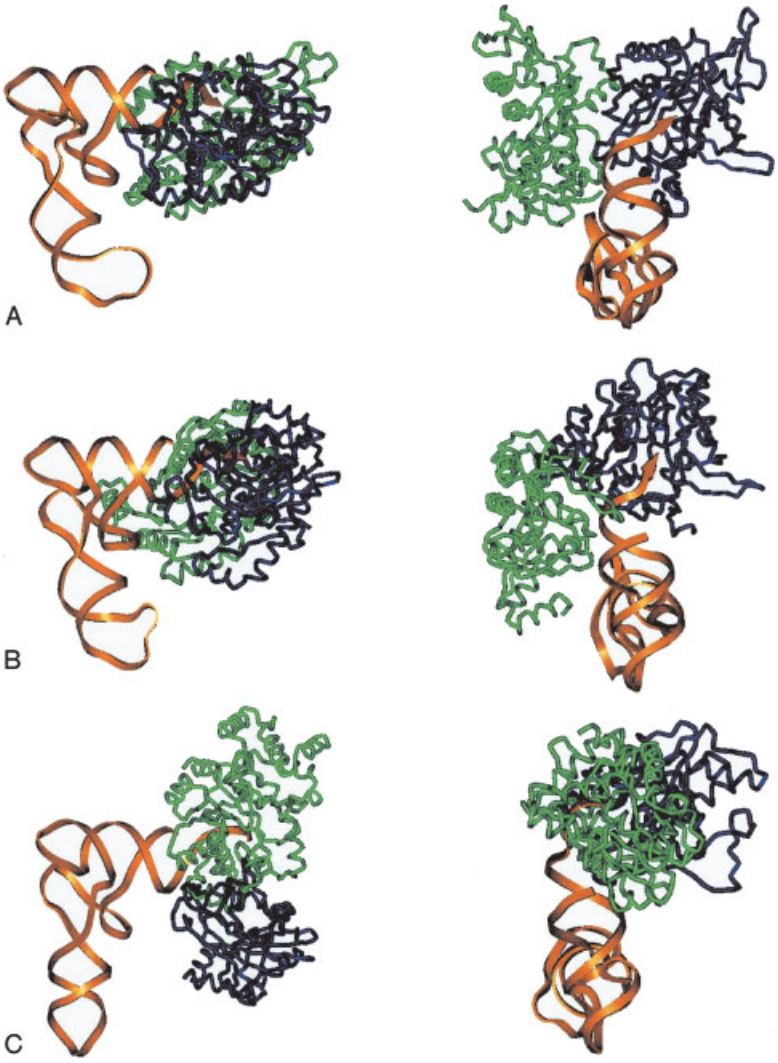


Figure 4.2-2 Graphical representation of complexes formed, respectively, between subclasses Ia–c (green) and IIa–c (blue) synthetase active sites bound simultaneously to a tRNA acceptor stem [59]. The tRNA is depicted in yellow, and each complex is shown in two different orientations. The left complexes are oriented with the plane of the page defined by the axes of the tRNA acceptor stem and anticodon stem helices. The views to the right of the figure show the same molecules along the axis of the anticodon stem-loop, as seen from the acceptor stem side.

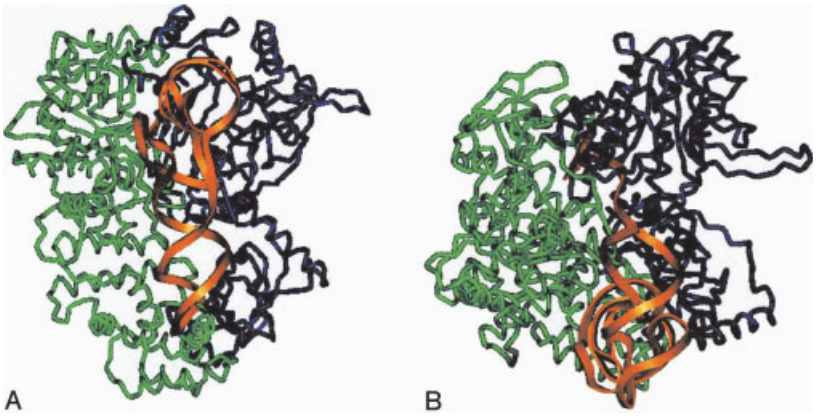


Figure 4.2-3 Graphical representation of the modeled complex between class I and II LysRSs, and $tRNA^{Lys}$ (modified from Ref. [68]). The two proteins were found to be complementary in their mode of binding $tRNA^{Lys}$ [68]. The tRNA molecule is depicted in orange, and the complex is shown in two different orientations as in Fig. 4.2-2.

directly followed the duplications of anticodons, then the ARS paired by the theory are predicted to recognize tRNAs that have related codons. This prediction is largely fulfilled [60], and it produces a general framework within which the growth in complexity of tRNAs and codon families can be examined [60]. The 'symmetrical' theory, however, does not consider the question of the origination of the two class ancestors.

4.2.5
A Common Genetic Origin for all Aminoacyl-tRNA Synthetases ?

Rodin and Ohno [69] first noticed that the coding sequences of class I ARS active sites could be aligned with complementary DNA sequences coding for class II ARS active sites (Fig. 4.2-4). Based on this observation, the genes for the ancestors of the two ARS classes were proposed to be encoded by complementary sequences of RNA or DNA. This proposal offered a mechanistic explanation for the linked duplication of class I and class II ancestral ARS, as stated by the 'symmetry theory'.

More recently, Carter and Duax [70] reported that, in the freshwater mold *Achlya klebsiana,* complementary genes code for proteins having the same folds as class I and II ARS. These genes are complementary to each other in a double-stranded DNA region, with each being transcribed and translated independently. Thus, DNA or RNA complementary strands could have originally coded for the two ARS folds.

The combination of the Rodin–Ohno model and the 'symmetry theory' offers a more detailed explanation for the emergence and evolution of ARS (Fig. 4.2-4). A region of double-stranded RNA or DNA, with complementary coding sequences, would be the starting point. A primitive translation mechanism (possibly RNA-

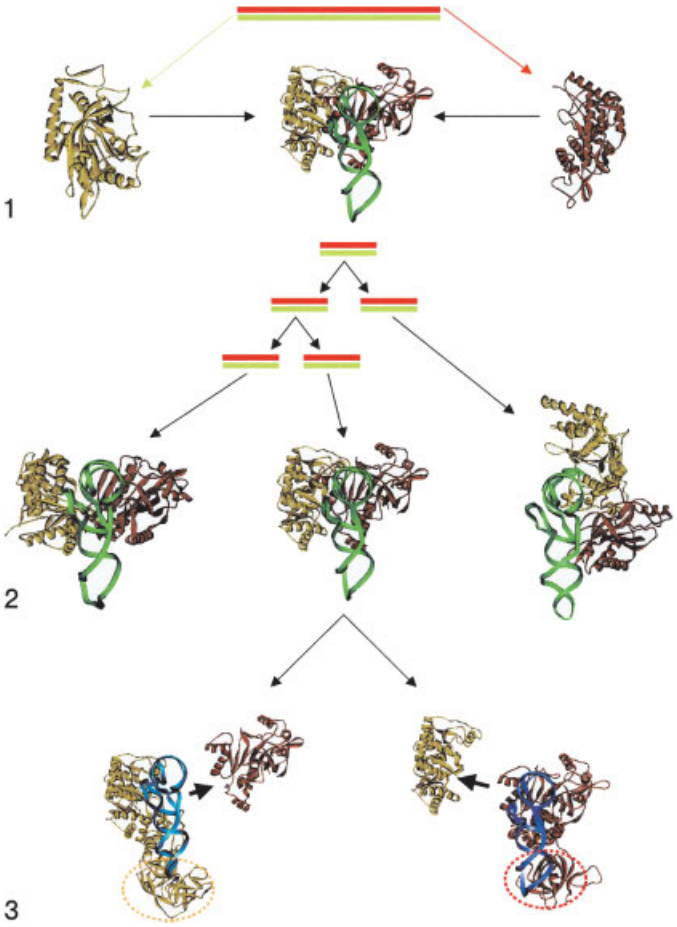


Figure 4.2-4 Evolutionary scheme of the origin and separation of synthetase genes and synthetase pairs, according to the Rodin–Ohno model and the 'symmetry theory' [59, 69]. Panel (1) depicts complementary genes that code for proteins having the same folds as class I and II ARS [70]. These complementary genes code for a pair of ancestral synthetase active sites that bind to a single tRNA (in green) [59]. In this depiction, the class I ancestral domain is shown in yellow on the left side of the complex, and the class II ancestral domain is on the right side in red. During growth of the genetic code, duplication of the complementary genes generates new synthetases that evolve to recognize emerging tRNAs with new identities [59]. Panel (2) represents the idiosyncratic evolution of the new complexes to achieve tRNA and amino acid specificities. Finally, in Panel (3), the ancestral synthetase domain involved in amino acid recognition incorporates other domains (shown as encircled) that allow it to recognize other regions of the tRNA better, giving rise to the modern synthetase structure. The second component of the ancestral pair is lost [60].

based) synthesized two peptides with different folds. These emerging peptides were initially selected for their ability to bind RNA stems simultaneously. Eventually, perhaps through their ability to bring ATP molecules to a pre-existing reaction, these two peptides became functionally involved in the aminoacylation of ancestral tRNA molecules (perhaps minihelix-like structures). The duplication of these tRNA molecules, and divergence of their anticodon sequences, drove the emergence and expansion of the genetic code. Simultaneously, duplication and divergence of the double-stranded region coding for the two ARS ancestors would double the number of available synthetases-like folds, and allow for the evolution of tRNA-specific binding.

It is conceivable that the physical association of the two ARS complementary genes extended to the tRNA genes themselves. Indeed, a comparative sequence analysis of tRNAs suggests that ancestral tRNA genes were coded in pairs by complementary strands of DNA or RNA [36, 69, 71]. This same analysis supported the concept that anticodon sequences arose from duplications of portions of the acceptor stems, thus supporting the idea that the operational RNA code was a precursor to the genetic code.

The proposal that ancestral ARS genes formed complementary DNA or RNA strands has implications for the study of the origin of life. In a primitive RNA world, where metabolic complexity is expected to be lower than in extant organisms, 'double-coding' RNA genomes might have been the norm rather than the exception. If RNA genes evolved to code simultaneously for functional RNA molecules and peptides, a direct physical link could be established between the ancestral machinery for RNA translation (based on ribozymes) and the emerging protein synthesis machinery. More generally, the concept could be extended to other ancient protein families to determine if protein-coding genes can simultaneously code for ribozymes that have the same activities as the proteins they encode. For example, in this scenario, the RNA sequences that code for primitive synthetases might themselves be ribozymes that catalyze aminoacylation.

4.2.5.1 Evolution of Extant Enzymes prior to LUCA

As discussed above, most of the gene duplications that gave rise to extant ARS had been accomplished by the time of appearance of the last universal common ancestor [72–74]. In contrast with phylogenies of whole species, which are not informative about pre-LUCA events, details of pre-LUCA biology might be obtained by analyzing the internal relationships among ARS. Additionally, establishing the order of the duplications that gave rise to the different members of each ARS family would allow us to link this process to the evolution and establishment of the genetic code.

A good example of this type of analysis was provided by studies of sequences of lysyl-tRNA synthetase (LysRS) and $tRNA^{Lys}$. The analysis of the complete genomic sequence of *Methanococcus jannaschii* made apparent that this organism does not contain a gene coding for a canonical class II LysRS [75]. Because all ARSs are essential enzymes, a search was initiated to find the missing activity in related species [43]. The enzyme catalyzing the aminoacylation of $tRNA^{Lys}$ with lysine in

M. maripaludis turned out to be a member of the class I family of ARSs. So far, LysRS represents the only instance of an aminoacylation activity that is catalyzed by a class I, or class II ARS, depending on the species.

The realization that LUCA might have possessed two genes coding for two distinct LysRSs offered, for the first time, the possibility of determining the time of establishment of a tRNA identity relative to the appearance of its cognate ARS. A phylogenetic analysis of the relationships among sequences of $tRNA^{Lys}$ from species bearing class I or class II LysRSs supports the idea that at least one of the extant forms of these enzymes was established in the context of a pre-existing $tRNA^{Lys}$, which remained universally distributed throughout the phylogenetic tree [76]. This prediction subsequently received support from the biochemical analysis of the aminoacylation properties of class I lysyl-tRNA synthetase [77, 78].

Class I LysRSs are mostly limited to archaebacterial species and a small number of eubacteria. On the other hand, class II LysRS are present in all kingdoms of life, including some archaeal species [43, 76]. Explanations for this gene distribution based on a hypothetical late lateral gene transfer of the class I *lysS* genes from bacteria to archaea (or *vice versa*) are not consistent with the phylogenetic analysis of LysRS sequences [43, 76]. More likely, the extant distributions of class I and II LysRSs arose from a situation where an ancestral organism possessed both genes. This redundancy was resolved through the elimination of one of the two genes, either through genetic drift, or by the appearance of selective pressures in favor of one of the two molecules [76].

This situation is, once again, clearly compatible with the 'symmetry' theory. Initially, $tRNA^{Lys}$ was bound by two synthetases, each with the capacity of charging this tRNA with lysine or, at least, of evolving this catalytic activity. As mentioned above, crystallographic studies support the possibility of a complex between $tRNA^{Lys}$ and two LysRSs of opposite classes. The separation of this complex into extant tRNA–ARS interactions took place after LUCA. Most organisms retained the class II fold as LysRS, but some selected the class I LysRS.

4.2.5.2 Changes in Acceptor Stem Identity Elements Correlate with Changes in the Code

The genetic code was first defined as a 'frozen accident' by Francis Crick, who argued that its current structure was due to the fact that its evolution had reached an evolutionary dead-end [79]. Emerging from this cul-de-sac was not possible because the system was incapable of assimilating new changes. This notion of a 'frozen' code has been challenged by the discovery of variations in the code of certain organisms and, more notably, in eukaryotic organelles [80–82]. Nevertheless, the genetic code has remained mostly invariable across the phylogenetic tree. This supports the notion that, for the most part, the code has reached a degree of complexity that does not accept new variations with ease.

As stated above, the 'operational RNA code' for amino acids is the relationship between sequences and structures of acceptor stems and specific amino acids [20, 22, 83]. Through variations of the 'operational RNA code' the genetic code can

change from organism to organism, because any change in the codon–amino acid equivalence has to be adopted by the 'operational RNA code'. Thus, it is conceivable that the frozen state of the genetic code is a consequence of the limitations of the 'operational RNA code'.

Misacylation errors are lethal to cells, and they are prevented through two different mechanisms. On the one hand, potential errors of amino acid recognition (caused by misrecognition of similar residues like isoleucine and valine) are corrected via editing domains contained in the error-prone synthetases [2, 58, 84–88]. On the other hand, potential errors in tRNA recognition are prevented by positive and negative identity elements in each tRNA [31, 32]. But the repertoire of identity elements might have limits. If the capacity of the 'operational RNA code' is limited then new variations in tRNA recognition mechanisms are not possible, because they would result in unacceptable levels of tRNA mischarging by the existing synthetases.

We propose that the fixed state of the genetic code is due to intrinsic limitations of the recognition of tRNAs by aminoacyl-tRNA synthetases. Expansion of the set of tRNAs is restricted because it runs the risk of causing acylation errors. However, incorporation of modifications to the genetic code requires changes in the cellular tRNA set. If the total set of tRNAs in a given organism is reduced, the discrimination problems faced by their cognate synthetases are decreased. This process would facilitate the evolution of tRNA sequences, because the available evolutionary space would increase. In turn, the divergence of tRNA sequences would open the possibility of changes in the genetic code.

Many of the genetic codes found to contain exceptions to the universal codon–amino acid assignments are in animal mitochondria. The first exception to the universal code was detected in the genomes of vertebrate mitochondria, where AUA codes for methionine instead of isoleucine, and UGA codes for tryptophan instead of being a stop triplet. Since that discovery, exceptions to the code have been detected in a large variety of organisms and organelles (reviewed in Ref. [82]). Most of the exceptions, however, are concentrated in metazoans (animals), involving changes of 11 different codons [82].

Additionally, animal mitochondria have experienced a dramatic reduction in their genome size and, in particular, in the number of tRNA genes [89]. If, as we propose, an initial requirement for changes in the code is the relaxation of the recognition constraints between ARS and tRNAs, then the large amount of variations in the genetic code of animal mitochondria should correlate with a large amount of changes in the 'operational RNA code' imbedded in their acceptor stem sequences.

As it can be seen in Table 4.2-2, the percentage of tRNA sequences in mitochondria that contain the recognition elements that are operational in bacteria or eukaryotes is significantly decreased for 16 amino acids. Remarkably, all tRNAs whose identity has been reported to change in mitochondria ($tRNA^{Ile}$, $tRNA^{Arg}$, $tRNA^{Met}$, $tRNA^{Lys}$, and $tRNA^{Ser}$) show important decreases in the conservation of identity elements (Table 4.2-2). Thus, in animal mitochondria, a reduction of tRNA genes is correlated with changes in the mechanisms of recognition between tRNAs and ARS and, simultaneously, with the largest concentration of changes in the genetic code.

Table 4.2-2 Drift of acceptor stem identity elements in animal mitochondrial tRNAs

Amino acid	Identity elements in acceptor stem[a]	Conservation in non-animal mitochondria (%)[b]	Conservation in animal mitochondria (%)[b]
Ala	G3:U70	85	21
Arg	A20	42	5
Asp	G73, G10	92, 80	13, 8
Gln	G2:C71	72	4
Gly	C2:G71, G3:C70	66, 79	46, 0
His	G1	62	0
Ile	C4:C69, C29:G41	47, 52	8, 2
Leu	A73, U8:A14	97, 100	68, 59
Lys	A73, G2:U71	86, 33	70, 10
Met	U4:A69	41	2
Phe	G15:C48	84	2
Pro	G15:C48, A73	37, 88	6, 8
Ser	G73, G2:C71	86, 100	24, 11
Thr	C2:G71	100	38
Tyr	A73	96	86
Val	G3:C70, U4:A69	95, 30	17, 9

a Identity elements in bacterial or eukaryotic tRNAs for review (see Refs. [31, 32]).
b Percentages calculated using all sequences in the Bayreuth tRNA database [90].

This situation was possibly initiated by the reduction of total tRNA genes. This reduction simplified the recognition problem for the mitochondrial synthetases, and allowed the recognition elements to drift into new sequence spaces, thus changing the types of identity elements and 'melting' the genetic code. This process promoted the appearance of new tRNA sequences, because new tRNA variations did not necessarily result in gross aminoacylation errors. These new tRNAs were then capable of acquiring new codon meanings, allowing the genetic code of these organelles to start evolving at a faster pace.

Acknowledgements

This work was supported by grant GM15539 and GM23562 from the National Institutes of Health and by a Fellowship from the National Foundation for Cancer Research.

References

1 P.R. Schimmel, D. Söll, *Ann. Rev. Biochem.* **1979**, *48*, 601–648.
2 S. Cusack, *Curr. Opin. Struct. Biol.* **1997**, *7*, 881–889.
3 A. Tzagoloff, D. Gatti, A. Gampel, *Prog. Nucl. Acid. Res. Mol. Biol.* **1990**, *39*, 129–158.
4 C.R. Woese, G.J. Olsen, M. Ibba et al., *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 202–236.
5 C. Woese, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6854–6859.
6 C.R. Woese, *Microbiol. Rev.* **1987**, *51*, 221–271.
7 T. Webster, H. Tsai, M. Kula et al., *Science* **1984**, *226*, 1315–1317.
8 S. Cusack, C. Berthet-Colominas, M. Hartlein et al., *Nature* **1990**, *347*, 249–255.
9 G. Eriani, M. Delarue, O. Poch et al., *Nature* **1990**, *347*, 203–206.
10 D. Moras, *Trends Biochem. Sci.* **1992**, *17*, 159–164.
11 L. Ribas de Pouplana, P. Schimmel, *J. Biol. Chem.* **2001**, *276*, 6881–6884.
12 T. Hashimoto, L.B. Sanchez, T. Shirakura et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6860–6865.
13 J.W. Chihade, J.R. Brown, P.R. Schimmel et al., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12153–12157.
14 Y.M. Hou, P. Schimmel, *Nature* **1988**, *333*, 140–145.
15 W.H. McClain, K. Foss, *Science* **1988**, *240*, 793–796.
16 S.J. Park, P. Schimmel, *J. Biol. Chem.* **1988**, *263*, 16527–16530.
17 A.D. Yaremchuk, M.A. Tukalo, I.A. Krikliviy et al., *J. Mol. Biol.* **1992**, *224*, 519–522.
18 H. Himeno, T. Hasegawa, T. Ueda et al., *Nucl. Acid. Res.* **1990**, *18*, 6815–6819.
19 M.E. Saks, J.R. Sampson, *EMBO J.* **1996**, *15*, 2843–2849.
20 P. Schimmel, R. Giegé, D. Moras et al., *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8763–8768.
21 C. de Duve, *Nature* **1988**, *333*, 117–118.
22 K. Musier-Forsyth, P. Schimmel, *Acc. Chem. Res.* **1999**, *32*, 368–375.
23 C. Francklyn, P. Schimmel, *Nature* **1989**, *337*, 478–481.
24 M. Frugier, C. Florentz, R. Giegé, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3990–3994.
25 M. Frugier, C. Florentz, R. Giegé, *EMBO J.* **1994**, *13*, 2219–2226.
26 S.A. Martinis, P. Schimmel: in *tRNA, Structure, Biosynthesis, and Function*, eds D. Soll, U.L. RajBhandary, ASM Press, Washington, DC 1994, 349–370.
27 M.E. Saks, J.R. Sampson, J.N. Abelson, *Science* **1994**, *263*, 191–197.
28 O. Nureki et al.: in *The Translational apparatus*, eds K.H. Nierhaus, F. Franceschi, A.R. Subramanian et al., Plenum, New York 1993, 59–66.
29 C.S. Hamann, Y.M. Hou, *Biochemistry* **1995**, *34*, 6527–6532.
30 B. Felden, R. Giegé, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10431–10436.
31 R. Giegé, M. Sissler, C. Florentz, *Nucleic Acids Res* **1998**, *26*, 5017–5035.
32 P.J. Beuning, K. Musier-Forsyth, *Biopolymers* **1999** *52*, 1–28.
33 P. Schimmel, B. Henderson, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11283– 11286.
34 B.S. Henderson, P. Schimmel, *Bioorg. Med. Chem.* **1997**, *5*, 1071–1079.
35 W. Moller, G.M. Janssen, *J. Mol. Evol.* **1992**, *34*, 471–477.
36 S. Rodin, S. Ohno, A. Rodin, *Orig. Life Evol. Biospys.* **1993**, *23*, 393–418.
37 M. Di Giulio, *Orig. Life Evol. Biosph.* **1994**, *24*, 425–434.
38 T.P. Dick, W.A. Schamel, *J. Mol. Evol.* **1995**, *41*, 1–9.
39 S. Rodin, A. Rodin, S. Ohno, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 4537–4542.
40 C. Hountondji, P. Dessen, S. Blanquet, *Biochimie* **1986**, *68*, 1071–1078.
41 S.W. Ludmerer, P. Schimmel, *J. Biol. Chem.* **1987**, *262*, 10801–10806.
42 P. Schimmel, L. Ribas de Pouplana, *Cell* **1995**, *81*, 983–986.

43 M. Ibba et al., *Science* **1997**, *278*, 1119–1122.
44 M. Jasin, L. Regan, P. Schimmel, *Nature* **1983**, *306*, 441–447.
45 K. Shiba, P. Schimmel, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1880–1884.
46 C.W. Carter, Jr., *Ann. Rev. Biochem.* **1993**, *62*, 715–748.
47 S. Cusack, *Biochimie* **1993**, *75*, 1077–1081.
48 M. Sprinzl, F. Cramer, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3049–3053.
49 T.H. Fraser, A. Rich, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3044–3048.
50 P. Schimmel, E. Schmidt, *Trends Biochem. Sci.* **1995**, *20*, 1–2.
51 L. Lin, S.P. Hale, P. Schimmel, *Nature* **1996**, *384*, 33–34.
52 O. Nureki et al., *Science* **1998**, *280*, 578–582.
53 W.C. Tsui, A.R. Fersht, *Nucleric. Acid. Res.* **1981**, *9*, 4627–4637.
54 R. Sankaranarayanan et al., *Cell* **1999**, *97*, 371–381.
55 P.J. Beuning, K. Musier-Forsyth, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8916–8920.
56 A. Dock-Bregeon et al., *Cell* **2000**, *103*, 877–884.
57 P.J. Beuning, K. Musier-Forsyth, *J. Biol. Chem.* **2001**, *276*, 30779–30785.
58 K. Beebe, L. Ribas de Pouplana, P. Schimmel, *EMBO J.* **2003**, *22*, 668–675.
59 L. Ribas de Pouplana, P. Schimmel, *Cell* **2001**, *104*.
60 L. Ribas de Pouplana, P. Schimmel, *Trends Biochem. Sci.* **2001**, *26*, 591–596.
61 V. Biou, A. Yaremchuk, M. Tukalo et al., *Science* **1994**, *263*, 1404–1410.
62 S. Fukai, O. Nureki, S. Sekine et al., *Cell* **2000**, *103*, 793–803.
63 H. Bedouelle, *Biochimie* **1990**, *72*, 589–598.
64 L.F. Silvian, J. Wang, T.A. Steitz, *Science* **1999**, *285*, 1074–1077.
65 M.A. Rould, J.J. Perona, D. Söll et al., *Science* **1989**, *246*, 1135–1142.
66 M. Ruff et al., *Science* **1991**, *252*, 1682–1689.
67 Y. Goldgur, L. Mosyak, L. Reshetnikova et al., *Structure* **1997**, *5*, 59–68.
68 T. Terada, O. Nureki, R. Ishitani et al., *Nat. Struct. Biol.* **2002**, *9*, 257–262.
69 S.N. Rodin, S. Ohno, *Orig. Life Evol. Biospys.* **1995**, *25*, 565–589.
70 C.W. Carter, W.L. Duax, *Mol. Cell* **2002**, *10*, 705–708.
71 S. Rodin, S. Ohno, A. Rodin, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4723–4727.
72 G.M. Nagel, R.F. Doolittle, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 8121–8125.
73 G.M. Nagel, R.F. Doolittle, *J. Mol. Evol.* **1995**, *40*, 487–498.
74 J.R. Brown, W.F. Doolittle, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2441–2445.
75 C.J. Bult et al., *Science* **1996**, *273*, 1058–1073.
76 L. Ribas de Pouplana, R.J. Turner, B.A. Steer et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11295–11300.
77 M. Ibba, H.C. Losey, Y. Kawarabayasi et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 418– 423.
78 P. Schimmel, L. Ribas de Pouplana, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 327–328.
79 F.H. Crick, *J. Mol. Biol.* **1968**, *38*, 367–379.
80 S. Osawa, D. Collins, T. Ohama et al., *J. Mol. Evol.* **1990**, *30*, 322–328.
81 M.A. Santos, T. Ueda, K. Watanabe et al., *Mol. Microbiol.* **1997** *26*, 423–431.
82 M.A.S. Santos, M.F. Tuite: in *The Genetic Code and the Origin of Life*, eds L. Ribas de Pouplana, Landes Biosciences, Houston 2003, in press.
83 P. Schimmel, *J. Mol. Evol.* **1995**, *40*, 531–536.
84 E.W. Eldred, P.R. Schimmel, *Biochemistry* **1972**, *11*, 17–23.
85 A.A. Schreier, P.R. Schimmel, *Biochemistry* **1972**, *11*, 1582–1589.
86 A.R. Fersht, C. Dingwall, *Biochemistry* **1979**, *18*, 2627–2631.
87 E. Schmidt, P. Schimmel, *Science* **1994**, *264*, 265–267.
88 L. Lin, S.P. Hale, P. Schimmel, *Nature* **1996**, *384*, 33–34.
89 C.G. Kurland, *Bioessays* **1992**, *14*, 709–714.
90 M. Sprinzl, K.S. Vassilenko, J. Emmerich et al., http://www.uni-bayreuth.de/departments/biochemie/trna/ 1999.

5
mRNA Decay and RNA-degrading Machines in Prokaryotes and Eukaryotes

Agamemnon J. Carpousis and Marc Dreyfus

5.1
Summary

Research over the past two decades has elucidated the pathways for mRNA decay in *Escherichia coli* and *Saccharomyces cerevisiae*. The study of these model organisms has given us a general overview of mRNA decay in both prokaryotes and eukaryotes. Although the two pathways are largely divergent, some common features are also apparent. Amongst the novel discoveries made in the course of this work was the identification and characterization of the *E. coli* RNA degradosome and the *S. cerevisiae* exosome, which are multienzyme RNA-degrading machines involved in the maturation of stable RNA and the degradation of mRNA. In this chapter, we describe and compare the *E. coli* and *S. cerevisiae* mRNA decay pathways, we discuss the role of RNase E and the RNA degradosome in procaryotic RNA degradation, and then we compare the degradosome to RNase E-based complexes found in other bacteria, and to the eukaryotic exosome.

5.2
Introduction

The ribosome, which decodes genetic information and synthesizes protein, is not the only multicomponent machine that uses RNA as a substrate. Other examples include eukaryotic systems involved in mRNA splicing, 3′-polyadenylation and export from the nucleus as well as eukaryotic and procaryotic complexes involved in RNA maturation and degradation. The *E. coli* RNA degradosome and the *S. cerevisiae* exosome are multienzyme RNA-degrading machines involved in the maturation of stable RNA and the degradation of mRNA. The best-known stable RNAs are the transfer and ribosomal RNAs, which are processed from precursor transcripts to their mature forms. Messenger RNAs are unstable with half-lives in *E. coli* ranging from 30 s to 20 min at 37° C. In eukaryotic cells, mRNA turnover is slower, but the half-lives are usually shorter than the generation time. The instability of mRNA is an important property permitting timely adjustments to changes in growth conditions or to genetically controlled programs of expression. Until recently, transfer and ribosomal RNAs

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

were believed to be protected from degradation by their rapid folding and assembly into compact structures. This simplistic view seems unlikely since the RNA-degrading machinery is more robust than imagined previously. Another widely held preconception was that the enzymes involved in the processing of stable RNA are distinct from those in the degradation of mRNA. With the discovery, in *E. coli* and *S. cerevisiae,* that ribonucleases involved in the maturation of ribosomal RNA are also important in the degradation of mRNA, it is now evident that these processes are closely connected. Several articles at the beginning of the references are recommended for reviews on the degradation of mRNA in bacteria [1–4] and eukaryotes [5–7].

5.3 mRNA Decay in *E. coli*

In *E. coli,* the degradation of mRNA is mediated by the combined action of endo- and exoribonucleases (Fig. 5-1A). The endonucleases initiate mRNA decay by creating fragments that are then degraded by the exonucleases. It is now generally believed that the principal endonuclease involved in mRNA decay is RNase E. Arguments supporting this contention have been marshalled in a recent review [8]. In a subsequent step, two enzymes, RNase II and PNPase, degrade the RNA fragments in a 3′→5′ pathway. Enzymes related to RNase II and PNPase are widespread in bacteria and eukaryotes [9, 10]. RNase II is a hydrolytic enzyme producing nucleotide monophosphates (AMP, etc.). PNPase, which is a phosphorylase, uses inorganic phosphate yielding nucleotide diphosphates (ADP, etc.). Although we sometimes speak of PNPase as a phosphate-dependent ribonuclease, this is not strictly correct since nucleases are hydrolytic by definition. A strain of *E. coli* with mutations in the genes encoding RNase II and PNPase, which is conditionally lethal, accumulates mRNA fragments under conditions that are non-permissive for growth [11]. This result is the principal experimental evidence for mRNA fragments as intermediates in decay. The 3′-ends of many bacterial mRNAs, such as those formed by rho-independent termination, are sequestered in stem-loop structures that protect them from degradation. Intercistronic regions in polycistronic transcripts can also harbor protective RNA structures and nascent transcripts have 3′-ends protected by the RNA polymerase. Thus, RNase II and PNPase are believed to be generally incapable of initiating the decay of an intact mRNA. The endonucleolytic cleavage of an mRNA can remove protective RNA structures or sever the nascent transcript from the RNA polymerase, thus producing a single-stranded 3′-end upstream of the cleavage that is a binding site for the exonucleases. Furthermore, the possibility that a cleavage by RNase E might 'trigger' exonuclease-mediated decay of the downstream mRNA fragment is discussed below. The idea that the initial attack by an endonuclease is followed by exonucleolytic decay of mRNA fragments was an important advance in our concept of the degradation of mRNA in *E. coli* [12, 13].

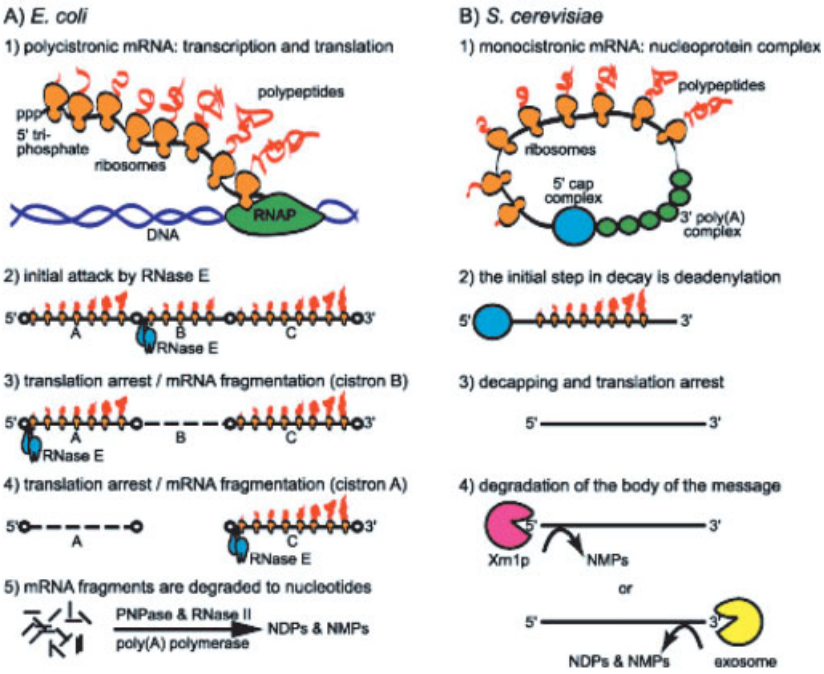


Figure 5-1 Messenger RNA decay in *E. coli* (A) and *S. cerevisiae* (B). The degradation of a hypothetical polycistronic transcript encoding genes A, B, and C is shown in (A). The order of decay depicted here is B, A and then C, but this is an arbitrary choice. The pattern of decay of a real polycistronic transcript in *E. coli* depends on the transcription unit. RNAP (step 1) is the DNA-dependent RNA polymerase. The open circles (steps 2–4) indicate 5′- and 3′-UTRs, and intergenic regions that contain elements controlling translation and mRNA decay. Dashed lines (steps 3–4) indicate cistrons where translation has been arrested and the mRNA has been fragmented by RNase E. The degradation of a hypothetical eukaryotic mRNA is shown in (B). The 5′-cap and 3′-poly(A) complexes are important for translation and mRNA stability. Their removal arrests translation initiation and triggers the degradation of the body of the message. RNase II (A, step 5) and Xrn1p (B, step 4) are hydrolytic enzymes that use water to produce nucleotide monophosphates (NMPs). PNPase (A, step 5) is a phosphorylase that uses inorganic phosphate to produce nucleotide diphosphates (NDPs). The exosome (B, step 4) has both phosphorylytic and hydrolytic activity.

In Fig. 5-1(A) (step 5), poly(A) polymerase assists in the degradation of the mRNA fragments by RNase II and PNPase. RNA 3′-polyadenylation has been implicated in *E. coli* mRNA decay [14–17]. In the PNPase/RNase II double mutant, under non-permissive conditions, the polyadenylation of mRNA decay intermediates is easily detected. *In vivo* and *in vitro* work has shown that 3′-poly(A) addition promotes the exonucleolytic degradation of RNAs whose 3′-ends are sequestered in secondary structure [18–20]. Since the exonucleases are single-strand-specific, the addition of a 3′-poly(A) tail creates a binding site for the exonucleases (see Refs. [21–23] for reviews). Furthermore, *in vitro* experiments using purified PAP I, PNPase and ATP

have demonstrated the degradation of a structured RNA in a reaction involving multiple cycles of polyadenylation [20]. Based on these results, it has been suggested that repeated rounds of polyadenylation and exonucleolytic attack might be necessary for certain hard-to-degrade intermediates. Taken together, this work has shown that RNA 3′-polyadenylation facilitates the decay of certain highly structured mRNA fragments and it suggests that polyadenylation could have a general role in accelerating degradation by the exonucleases.

5.4 mRNA Decay in *S. cerevisiae*

In eukaryotes, the endonucleolytic decay of mRNA appears to be less important although it could be involved in regulating the stability of certain messages [24–26]. Two exonucleolytic pathways have been described in *S. cerevisiae*: 3′→5′ and 5′→3′ pathways [27–31]. Eukaryotic mRNAs are protected by 5′-cap structures and 3′-poly(A) tails. Both elements bind specific proteins that can interact with each other, so that the mRNA is effectively circular (Fig. 5-1B). The cap and poly(A) complexes are important for translation and mRNA stability. Deadenylation by a poly(A)-specific nuclease, followed by decapping, are prerequisites for degradation of the body of the message. Decapping is deadenylation-dependent. In the 5′→3′ degradation pathway, the mRNA is degraded by the exonuclease encoded by *XRN1*, a hydrolytic enzyme producing nucleotide monophosphates. The 3′→5′ degradation pathway involves a multiprotein complex, the exosome (see below), which contains both hydrolytic and phosphorylytic enzymes. Note that since deadenylation promotes decapping and thus the arrest of translation initiation, the 5′→3′ and 3′→5′ pathways are ordered processes in which no new translation can occur during the degradation of the body of the mRNA. In *S. cerevisiae*, the 5′→3′ pathway is apparently the predominant mode of degradation. Whether this is the case in other eukaryotes remains to be clarified. A recent *in vitro* study suggests that the 3′→5′ pathway mediated by the human exosome has an important role in the degradation of short-lived mRNAs encoding certain cytokines and oncogenes [32].

5.5 A Comparison of mRNA Decay in *E. coli* and *S. cerevisiae*

A comparison of the *E. coli* and *S. cerevisiae* pathways in Fig. 5-1 shows that bacterial and eukaryotic mRNA decays are considerably divergent. This probably reflects fundamental differences in the organization of transcription units and the mechanism of translation initiation. *E. coli* does not have a 5′→3′ degradation pathway. The 5′-ends of its mRNA are neither capped nor are there known 5′→3′ ribo-exonucleases. Homologs of the yeast capping enzyme and Xrn1p are found only in other eukaryotes. Thus, capping and the 5′→3′ degradation pathway are apparently specific features of the eukaryotic cell. Eukaryotic messages are generally monocistronic and translation initiation usually involves scanning from the 5′-cap complex to the AUG. Bacterial messages are often polycistronic and translation initiation involves

sequence-specific binding of the ribosome just upstream of the initiating AUG. Owing to this internal mode of entry, the frequency of translation initiation can be independent for each cistron of a polycistronic mRNA. In addition, the inactivation by endonucleolytic cleavage can trigger the decay of a cistron without disrupting neighboring cistrons. Thus, in a bacterial transcription unit with a single promoter, the yield of protein as well as the level of steady-state mRNA can vary considerably from cistron to cistron even though the transcription rate is equivalent. It is difficult to imagine how the independent decay of individual cistrons within a polycistronic transcript could be achieved without the action of an endonuclease. In contrast, there is no obvious advantage in initiating the decay of a monocistronic eukaryotic transcript with an endonuclease although an endonucleolytic cleavage in a 3′-UTR could serve as an alternative pathway for deadenylation.

It has been argued that the *E. coli* mRNA decay pathway outlined in Fig. 5-1(A) is inherently flawed since, in principle, a translated message might be cleaved internally leading to a truncated mRNA without a stop codon and thus a stalled ribosome with an incomplete nascent polypeptide. In the next section, we discuss possible mechanisms whereby many *E. coli* mRNAs may in fact decay via an orderly process in which initiation of decay is coordinated with the arrest of translation, as is the case for eukaryotic mRNAs. Nevertheless, *E. coli* and other bacteria have a mechanism for rescuing ribosomes stalled on an mRNA fragment. The tmRNA, with both tRNA and mRNA function, permits a *trans*-translation step in which the tmRNA initially binds to the stalled ribosome as a tRNA, then it serves as a short mRNA template providing a stop codon [33–35] (see Chap. 11 for more details). Briefly, in this process, the mRNA fragment is released and the truncated polypeptide receives a short C-terminal addition encoded by the tmRNA. The tmRNA-encoded C-terminal tag contains a signal that directs the proteolysis of the incomplete polypeptide. Factors associated with the tmRNA include RNase R, a ribonuclease related to RNase II [36]. Further work is required to establish if RNase R is involved in the decay of the mRNA fragment. It should be noted that a related process, involving the exosome, has been described in yeast [37]. In eukaryotic 'non-stop decay', the exosome degrades the mRNA fragment, although the fate of the stalled ribosome and the associated polypeptide remains to be elucidated. Whereas the tmRNA in *E. coli* can rescue a ribosome stalled at the 3′-end of a non-stop mRNA, it can also release ribosomes that are stalled internally on an intact mRNA. Thus, the emerging view is that the tmRNA has a general role in the rescue of stalled ribosomes [33].

5.6 RNase E Specificity: A Role in Translation Arrest?

The RNase E of *E. coli* is a single-strand specific endo-ribonuclease with a preference for AU-rich sequences [38–41]. *In vitro* work has shown that although RNase E is an endonuclease, its activity is influenced by the 5′-end of the RNA [42]. The hybridization of an oligonucleotide to the 5′-end of a small RNA substrate inhibits endonucleolytic cleavage by RNase E at a downstream site suggesting that single-stranded 5′-ends facilitate substrate binding. These results could explain previous *in vivo* work

which showed that RNA stem loops, sequestering the 5′-end into double-stranded structure, impede RNase E-mediated degradation [43]. Furthermore, the *in vitro* work showed that the initial rate of cleavage is faster with a 5′-monophosphate end than with a 5′-triphosphate, and that covalently closed circular RNAs are resistant to cleavage by RNase E. Thus, RNase E 'senses' RNA topology presumably by recognition of the 5′-end although other results suggests that the 3′-end could be involved as well [44]. The remarkable stability of covalently closed circular mRNA *in vivo* is further evidence for the importance of the RNA ends in controlling degradation [45]. Nevertheless, mRNA that is fully protected at the 5′-end by a stem-loop structure can still be degraded in a slow RNase E-dependent pathway that apparently involves 'internal entry', i.e., via an end-independent mechanism [46].

The interaction of RNase E with the ends of its substrate could help to give an overall direction to mRNA decay. The model in Fig. 5-2 shows how RNase E could help to reinforce a 5′→3′ directionality if the initial cleavage is in the 5′-end. RNase E, which is known to be oligomeric, is generally presumed to be a dimer. After the initial cleavage, RNase E is envisioned to remain tethered to the 5′-end of the decay product (Fig. 5-2). Since the enzyme is dimeric, it could remain bound to the 5′-end and simultaneously interact with a downstream site by 'looping out' the intervening RNA [1]. Repeated cycles of tethering and looping could produce a rapid, processive reaction in which RNase E fragments the message. If the initial cleavage inactivates translation, this could help to resolve the translation-decay conflict by facilitating RNase E cleavage as the ribosomes clear from the message. A 5′→3′ directionality of mRNA decay was first proposed by Apirion as a means of avoiding conflicts between translation and decay [12]. This directionality is indeed evident in long cistrons such as *lacZ*, in which the 5′ region starts decaying before the 3′-end is made [47]. The localization of RNase E to the inner periphery of the *E. coli* cell [48] could also contribute to this directionality, since the most recently synthesized RNA, being associated with the nucleoid in the interior of the cell, might not be accessible to RNase E: it is the 5′-end of the growing transcript that would first reach RNase E.

The model in Fig. 5-2 is probably an oversimplification. Recent work with short synthetic RNAs suggests that RNase E has a preference for distal cleavage sites giving an overall 3′→5′ directionality *in vitro* [49]. The basis for this preference is not known but it could involve recognition of the 3′-end by RNase E. Furthermore, our understanding of the decay of two well-studied mRNAs is at odds with the model in Fig. 5-2. In the *rpsO* message, the initial cleavage by RNase E is 10 nucleotides downstream of the translation termination codon and in the *rpsT* message, in the 3′ half of the coding sequence [50, 51]. Thus, there is no evidence that RNase E cleaves the 5′-UTRs of these messages. Whether our understanding of the *rpsO* and *rpsT* messages can be generalized to other messages is not clear since both are short transcripts whose translation is self-regulated by the binding of their cognate proteins to translational operators in their 5′-UTRs. Regardless of these considerations, other mechanisms for avoiding the translation-degradation conflict are possible. For instance, a relatively minor modification of the scheme depicted in Fig. 5-2 would be that RNase E binds to, but does not cleave, the translation-initiation region in an initial step that arrests translation. Indeed, it has been proposed that RNase E (or other RNA-binding

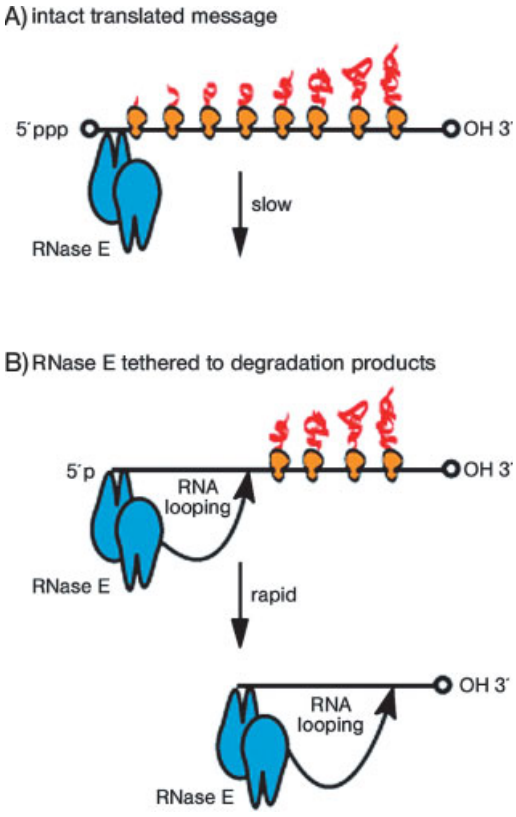


Figure 5-2 A hypothetical scheme in which a slow initial cleavage by dimeric RNase E leads to a rapid processive fragmentation of the mRNA (see text). In this model, the initial cleavage in the 5′-UTR inactivates translation initiation. The subsequent cleavages by RNase E tethered to the monophosphate 5′-end of the decay products occur as the ribosomes clear from the transcript. In this process, RNase E remains bound to the 5′-monophosphate end in a reaction that involves the 'looping out' of the RNA substrate as RNase E searches for cleavage sites. For simplicity, the mRNA is represented as a monocistronic transcript with protective structures in the 5′- and 3′-UTRs (open circles), but this model could also apply to a cistron within a polycistronic message.

proteins) could arrest translation and sequester the message in an inactive form before initiating nucleolytic degradation [52]. Translation initiation sites are unstructured and by definition contain binding sites for the ribosomal protein S1. RNase E is single-strand-specific and contains an S1 RNA-binding domain. Thus, 5′-UTRs and intergenic regions containing translation-initiation regions could be targets for RNase E binding. Other non-nucleolytic models for translation arrest have been proposed. For instance, it has been suggested that decay might be initiated by a 'collapse'

of the translation-initiation region into RNA secondary structure that inhibits ribosome binding [53, 54]. In research on *E. coli* mRNA decay, sorting out the connection between mRNA degradation and translation arrest is an important challenge.

The ideas presented in this section are based principally on the study of short RNA substrates *in vitro* or short monocistronic messages *in vivo*. How these ideas apply to more complicated polycistronic mRNAs remains to be elaborated; however, it seems probable that the intergenic regions of polycistronic transcripts will contain elements that act as initiators of the decay as well as elements that serve as barriers preventing the spread of decay from a cistron to its neighbors.

5.7 The *E. coli* RNA degradosome

Two temperature-sensitive mutations, now known as *rne3071* and *rne1*, were identified because of their effect on the maturation of 5S ribosomal RNA [55] and the degradation of mRNA [56]. Subsequent studies showed that both mutations are in the structural gene for RNase E [57–59]. It is now generally accepted that RNase E has a role in both the maturation of ribosomal RNA and the degradation of mRNA. Recent work has shown that RNase E is also essential for tRNA maturation and evidence from these studies suggests that tRNA deficiency is the ultimate cause of lethality in RNase E mutant strains [60, 61]. It is striking that RNase E appears to have a role in the processing and degradation of nearly every transcript in *E. coli*. RNase E is a large, multidomain enzyme that is part of a complex called the RNA degradosome. Figure 5-3 shows a schematic representation of the primary structure of RNase E. Its

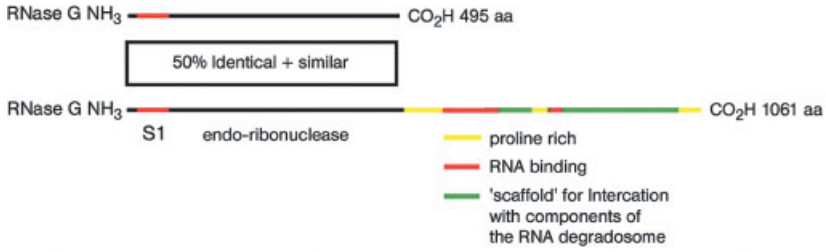


Figure 5-3 The RNase E/G family of enzymes. The RNase E and RNase G of *E. coli* are paralogues in which the N-terminal S1 RNA-binding domain and the endo-ribonuclease catalytic site are conserved [59]. RNase E differs from RNase G by its long C-terminal non-catalytic region containing proline 'hinge' regions, sites that bind RNA, and a 'scaffold' involved in protein–protein interactions with other components of the RNA degradosome (see text). Homologs of RNase G (catalytic domain only) and RNase E (non-catalytic extensions) are found throughout the eubacterial kingdom and in the plant chloroplast [71, 72, 110].

nucleolytic activity resides in the N-terminal half of the protein, which also contains an S1 RNA-binding domain [62–66]. The C-terminal half (CTH) of the protein contains several proline-rich regions, two arginine-rich RNA-binding regions and sites for protein–protein interactions with the other components of the RNA degradosome. *E. coli* encodes a paralogue, now called RNase G, that is about half the size of RNase E [67–69]. Although their catalytic domains are related, RNase G lacks the region corresponding to the CTH of RNase E (Fig. 5-3). It is noteworthy that RNase G is also 5′-end-dependent. This thus appears to be a general property of the RNase E/G family and the determinants involved in the 5′-end preference are apparently part of the conserved N-terminal catalytic domain. Proteins related to RNases E and G are found throughout the eubacterial kingdom and in certain plants [70–72]. The plant homologs are presumably in the chloroplast, which is an organelle of eubacterial origin. The 'RNase E/G' family can be divided into two groups: the large RNase E-like enzymes that can form degradosome complexes and the small RNase G-like proteins that presumably act alone. Although related, these enzymes are not functionally equivalent since in *E. coli*, RNase E is essential for viability whereas RNase G is dispensable. Nevertheless, it has been shown recently that if RNase G is overexpressed, it can complement a knockout of the gene encoding RNase E [73].

A multienzyme complex, now called the RNA degradosome, was discovered during the purification and characterization of *E. coli* RNase E [39, 74–76]. The major components of the RNA degradosome include RNase E, PNPase, and the DEAD-box RNA helicase, RhlB [77–79]. The RNA degradosome also contains enolase, a glycolytic enzyme, as an integral component. Associated proteins, present in substoichiometric amounts, include polyphosphate kinase (PPK), DnaK, and GroEL. Inter-actions with other enzymes such as *E. coli* poly(A) polymerase and the ribosomal protein S1 have also been reported [80, 81]. The role of enolase, PPK and other associated proteins in the degradation of mRNA remains to be clarified. The non-catalytic CTH of RNase E has been shown to contain the protein 'scaffold' upon which the other components of the RNA degradosome assemble [70, 82]. A functional 'minimal' degradosome containing RNase E, RhlB, and PNPase can be reconstituted from purified components [83, 84].

The association of RNase E and PNPase in a complex could provide a direct physical link for their cooperation in degradation. Indeed, there is *in vivo* evidence for coordination between endonucleolytic cleavage at the 5′-end by RNase E and attack at the 3′-end by PNPase [19, 21, 85]. RNA I, a small 108 nucleotide molecule, is a repressor of ColE1 plasmid replication with a short lifetime similar to that of an mRNA. Although the primary transcript with a 5′-triphosphate end and 3′-RNA stem-loop structure is resistant to PNPase attack, removal of five nucleotides from the 5′-end by RNase E triggers the decay of the RNA I-$_5$ intermediate in a pathway that involves 3′-polyadenylation and degradation by PNPase. It has been suggested that poly(A) polymerase and PNPase might be recruited to RNA I-$_5$ by their interaction with RNase E in the degradosome. Indeed, with the *rne131* mutant (see below), which disrupts the RNA degradosome, there is significant stabilization of the RNA I-$_5$ intermediate (M. Dreyfus, unpublished results). The decay of other small regulatory RNAs, such as

Sok, controlling R1 plasmid partition, and CopA, controlling R1 plasmid replication, are also controlled by an initial cleavage at the 5′-end by RNase E followed by a 3′ attack by PNPase [86, 87]. Although it is tempting to believe that the decay of these small regulatory RNAs could serve as a model for mRNA degradation, as discussed above, there is little evidence supporting the idea that RNase E initiates decay by cleavage in the 5′-end of an mRNA. However, a coordination between RNase E cleavage and subsequent steps involving polyadenylation and PNPase attack could be part of a pathway mediating the degradation of structured mRNA decay intermediates or mRNA 3′-end-fragments blocked by stem-loop structures.

The nuclease activity of RNase E is essential, but strains expressing protein with C-terminal truncations are viable. An allele, now known as *rne131*, directing the synthesis of a protein lacking the non-catalytic part of RNase E, was isolated in a screen for extragenic suppressors of a temperature-sensitive *mukB* allele [88]. The suppression resulted from overexpression of the mutant MukB protein. Several other *rne* mutants were obtained, all of them resulting in a truncated protein. The *rne131* mutation was extensively characterized in a subsequent study which showed that the maturation of 5S ribosomal RNA was normal, whereas there was a small but detectable slowdown in the decay of bulk mRNA [89]. It was also demonstrated that certain messages such as the endogenous *thrS* mRNA or messages synthesized by bacteriophage T7 RNA polymerase were preferentially stabilized compared with bulk mRNA. These messages, in which the coupling of transcription and translation is disrupted, might be degraded by RNase E in an alternative pathway distinct from that involving normally translated mRNA (see the next paragraph). Recent work with mutants disrupting various regions in the non-catalytic part of RNase E has demonstrated that it contains both positive and negative elements affecting mRNA-degrading activity [90]. This work also showed that (i) the autoregulation of RNase E synthesis (see the next section) compensates, at least partly, for the defective activity of the mutant enzyme *in vivo*; (ii) the *rne131* mutant has a significant growth defect in the absence of autoregulation; and (iii) even with autoregulation, the mutant strains are less fit than an isogenic wild-type strain in growth competition experiments. Thus, it seems reasonable to suggest that the non-catalytic part of RNase E is involved in fine-tuning RNA-degrading activity although the specific role of each element remains to be elucidated.

It is noteworthy that bacteriophage T7 expresses a protein kinase that phosphorylates a number of *E. coli* proteins including two components of the RNA degradosome, RNase E and RhlB [91]. The target in RNase E, which is heavily phosphorylated, is the non-catalytic region containing the RNA-binding domains and protein scaffold. Bacteriophage T7 encodes its own RNA polymerase. In uninfected cells, mRNA synthesized by this polymerase is exceptionally sensitive to inactivation by RNase E [92]. Since the transcription-elongation rate of the T7 RNA polymerase is 5- to 10-fold faster than its *E. coli* counterpart, it largely exceeds the rate of translation elongation. This leads to long stretches of ribosome-free mRNA proximal to the RNA polymerase. The inactivation of these messages by RNase E appears to involve 'internal entry', i.e., via a 5′-end-independent mechanism [46, 91].

Expression of the T7 protein kinase in uninfected cells stabilizes mRNA transcribed by the T7 RNA polymerase [91]. Thus, during a T7 infection, phosphorylation could help to stabilize the bacteriophage messages although this has not been demonstrated directly. It should be interesting to elucidate the mechanism by which the phosphorylation of RNase E controls 5′-end-independent mRNA decay and to ask if there are cellular protein kinases, perhaps regulated as part of a signal transduction pathway, which modulate RNase E activity depending on conditions of growth or stress.

5.8
The Autoregulation of RNase E and PNPase Synthesis: A Link between Bulk Translation and mRNA Stability

The expressions of RNase E and PNPase are both autoregulated in posttranscriptional pathways that involve the control of mRNA stability via elements in the 5′-untranslated region (UTR) of their messages (Fig. 5-4). The mRNA encoding RNase E contains a 361 nt 5′-UTR region that controls the stability of the *rne* message in response to RNase E levels [93, 94]. Experiments in which fusions were constructed between the 5′-UTR and a *lacZ* reporter gene demonstrated that the *rne* leader regulates functional stability by a mechanism that 'senses' RNase E activity in the cell. More recent work has identified a stem-loop structure in the *rne* leader that is essential for autoregulation and it has been proposed that binding to this site tethers RNase E to the mRNA and promotes its degradation [95, 96]. In the presence of high levels of RNase E, its mRNA would be destabilized thus decreasing expression, whereas low levels would provoke mRNA stabilization and increased expression. The control of RNase E expression can be viewed as a homeostasis that assures adequate RNase E activity. Indeed, recent *in vivo* results have confirmed that this autoregulatory system responds to changes in the demand for RNase E activity [97, 98]. This work shows that the RNase E message is particularly sensitive to changes in RNase E concentration, i.e., it is only partially inactivated at concentrations where other cellular targets are already saturated. How this is achieved remains to be elucidated, but it suggests that the activity of RNase E on its own message is weaker than its activity on other messages or precursors of stable RNA.

PNPase is expressed as part of a polycistronic transcript that begins with the *rpsO* gene encoding the small ribosomal protein, S15 [99]. Maturation of the *pnp* message involves RNase III processing of a double-stranded RNA structure encoded in the intergenic space between *rpsO* and *pnp*. The processing by RNase III is essential for autoregulation [100] and recent work has revealed a novel mechanism for the control of PNPase expression [101]. RNase III processing yields a mature *pnp* mRNA with its 5′-UTR hybridized to an oligoribonucleotide that stabilizes the mRNA (Fig. 5-4B). The degradation of this oligoribonucleotide by PNPase destroys the duplex thus exposing the 5′-monophosphate end and destabilizing the *pnp* message, which is presumably degraded by RNase E although this has not been established. Thus, the oligoribonucleotide acts as a sensor of the level of exonucleolytic activity in the cell.

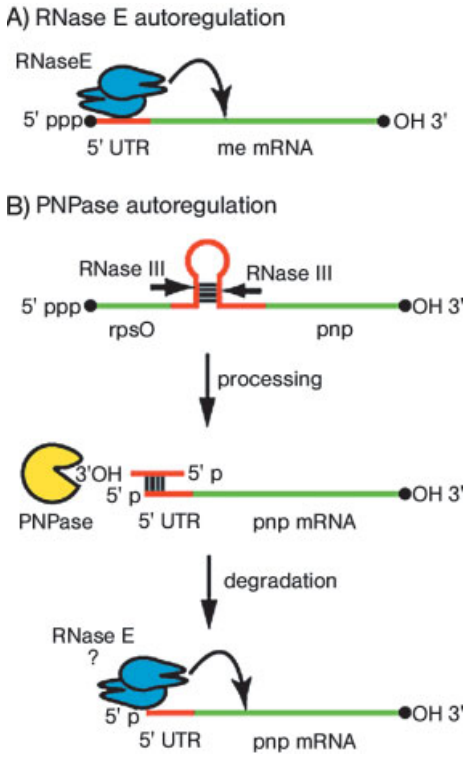


Figure 5-4 The autoregulation of RNase E (A) and PNPase (B) synthesis. The green regions represent coding sequences whereas the red represent untranslated regions (UTRs). In the mechanisms described in both panels, autoregulation involves a process that 'senses' nuclease activity in the cell and controls expression by modulating mRNA stability. In (A), the *rne* mRNA has a long, 361 nt, 5′-UTR that is specifically targeted by RNase E in a process involving 'tethering' to an element in the 5′-UTR (see text). In (B), PNPase expression is regulated in a two-step process. The primary polycistronic transcript encoding *rpsO* and *pnp* is processed to a mature *pnp* message in a pathway that involves degradation of the *rpsO* message (not shown) and the processing of an intergenic RNA stem loop by RNase III. The product of RNase III cleavage is drawn to emphasize that the processed message has a 5′-end that is protected in a double-stranded RNA structure. The oligoribonucleotide hybridized to the 5′-UTR of the *pnp* mRNA has a short protruding 3′-end that is sensitive to attack by PNPase. Degradation of the protective oligoribonucleotide by PNPase promotes the decay of the *pnp* mRNA, which could be mediated by RNase E although this has not been demonstrated.

It should be mentioned that the other major exo-ribonuclease in *E. coli* mRNA decay, RNase II, is also autoregulated and that there is cross-regulation between RNase II and PNPase [102]. The effect of RNase II on PNPase expression could involve degradation of the oligoribonucleotide that stabilizes the *pnp* message although this remains to be tested. The mechanism by which RNase II is autoregulated has not yet been elucidated.

The fact that the stability of the *rne* and *pnp* mRNAs varies with the concentration of their cognate proteins suggests that PNPase and RNase E are never present in excess in the cell. Rather, these proteins must be able to adjust continuously their concentration through autoregulation. Consistent with this view, a burst in the synthesis of an RNase E substrate causes a transient stabilization of the *rne* mRNA until the RNase E pool has expanded to meet the new demand [97]. Similarly, even though poly(A) tails usually destabilize mRNA fragments, overexpression of poly(A) polymerase leads, paradoxically, to the stabilization of the *pnp* and *rne* mRNAs, presumably because the need to degrade the extra poly(A) tails increases the demand for PNPase and RNase E [103]. Interestingly, the homeostatic regulation of RNase E and PNPase expression may be responsible for a seemingly unrelated phenomenon, i.e., the well-known stabilization of bulk mRNA that follows a block in translation. This phenomenon is generally attributed to a protection of mRNAs by stalled ribosomes. However, even untranslated mRNAs are protected from degradation under these circumstances, showing that the stabilization must somehow reflect the reduced activity of the degradation machinery itself. In particular, RNA I and its RNase E cleavage product, RNA I-5, are stabilized, suggesting that both RNase E and PNPase are inhibited under these circumstances [85]. The homeostasy of RNase E and PNPase can provide a straightforward explanation for these effects. Following a translation block, the synthesis of ribosomal RNA is known to be boosted. Moreover, the newly synthesized ribosomal RNA is unstable since it cannot assemble into ribosomes due to the lack of new ribosomal proteins [104]. This results in an increased demand for RNase E and PNPase under conditions where the pools of these enzymes cannot expand. Their titration by the ribosomal RNA thus explains the stabilization of bulk mRNA [85, 97]. Interestingly, a block in translation also causes stabilization of many and perhaps most mRNAs in yeast and higher eukaryotic cells [105]. It will be interesting to learn if the expression of components of the eukaryotic mRNA degradation machinery is also autoregulated.

5.9
RNA-degrading Machines in other Organisms

Several other degradosome-like complexes have been identified and characterized over the past decade (Table 5-1). All act in a 3′→5′ degradation pathway. An RNase E-based complex has been characterized in *Rhodobacter capsulatus*, which is a photosynthetic Gram-negative bacteria that is only distantly related to *E. coli* [106]. Although a PNPase-like activity co-purified with this complex, none of the major polypeptides identified by protein sequencing corresponded to a PNPase homolog.

Table 5-1 RNA-degrading multienzyme complexes

	Degradosome ***E. coli***	**Degradosome** ***R. captulatus***	**?** **chloroplast**	***S. cerevisiae*** **mitochondria**	**Exosome** ***S. cerevisiae*** **nucleus/cytoplasm**
Integral proteins	PNPase 3′→5′phosphorylase (PH1, PH2, S1, KH)		100RNP 3′→5′ phosphorylase (PH1, PH2, S1, KH)	Dss1p 3′→5′ ribonuclease (RNase II homolog)	Rrp41p, 42p, 43p, 45p, 46p, Mtr3p 3′→5′ phosphorylase
	RNase E endonuclease, scaffold	RNase E endonuclease, scaffold	?	p75 ?	Rrp4p, 6p, 40p, 44p, Clsp4p 3′→5′ ribonuclease S1, KH, RNA binding
	RhlB RNA helicase	Rhl1, Rhl2 RNA helicase		Suv3p RNA helicase	
	enolase glycolytic enzyme	Rho transcription termination			
Associated proteins	DnaK, GroEL chaperons	PNPase?	?	?	Mtr4p, Ski2p RNA helicases
	PPK, PAP, S1 specificity?				Ski3p, Ski8p specificity?

Thus, it has been suggested that PNPase might only be loosely associated with this complex. Curiously, two DEAD-box RNA helicases and the transcription termination factor Rho, which is also an RNA helicase, were shown to be associated with RNase E. The role of multiple helicases in the complex is unclear although these proteins might act as adaptors that target the degradosome to specific substrates. The link with Rho is intriguing. In *E. coli*, Rho is an essential factor that is responsible for rho-dependent transcription termination [107]. One manifestation of Rho activity is 'polarity', a phenomenon in which a mutation terminating translation within a cistron provokes transcription termination. The association of Rho with RNase E suggests a link between rho-dependent transcription termination and mRNA degradation, which could involve targeting the degradosome to rho-terminated mRNA. Message decay in the plant chloroplast, an organelle of eubacterial origin, has also been suggested to involve a degradosome. However, despite an earlier report of a degradosome-like association, 100RNP, which is a PNPase homolog, appears to be a hexamer of identical subunits forming a large 600 kDa enzyme [108]. Whether other enzymes associate with the chloroplast PNPase is an open question. In *Streptomyces coelicolor*, a Gram-positive bacteria, an RNase E-like activity was described several years ago [109] and an authentic homolog, RNase ES, has recently been identified [110]. Intriguingly, RNase ES has been shown to associate physically with the PNPase from *S. coelicolor*, suggesting the existence of a degradosome-like complex. Further work will be required to characterize the putative *Streptomyces* degradosome including the identification of other proteins that associate with RNase ES. Considering the very large evolutionary distance between *E. coli* and *S. coelicolor*, these results suggest that the physical association of RNase E and PNPase to form degradosome-like complexes might be widespread in bacteria.

In the yeast *S. cerevisiae*, two complexes have been described, the mtEXO complex and the exosome. The mtEXO complex (Table 5-1), located in the mitochondria, is required for the degradation of introns [111, 112]. Dss1p, in the mtEXO complex, is an exoribonuclease related to RNase II. The RNA helicase Suv3p, an integral component, is required for mtEXO activity both *in vitro* and *in vivo*. The yeast exosome (Table 5-1), with both hydrolytic and phosphorylytic activity, exists in a cytoplasmic form that degrades mRNA and a nuclear form that processes ribosomal RNA and small nuclear RNAs [6]. The nuclear form is also involved in the degradation of pre-mRNA and the rescue of read through transcripts that fail to be cleaved and polyadenylated at the normal processing site [113, 114]. It is from their function in ribosomal RNA processing that many of the components of the exosome derive their "Rrp" nomenclature. A number of co-factors are named "Ski" for the observed super killer phenotype due to overexpression of a toxin from an endogenous RNA. Mtr4p and Ski2p are DEvH-box RNA helicases. Exosome-like complexes have been found in a broad spectrum of eukaryotes ranging from humans to trypanosomes, thus suggesting that they are a highly conserved feature of eukaryotic stable RNA maturation and mRNA decay [115, 116].

A key similarity between eubacterial PNPase and the yeast exosome is that they both have phosphorylytic activity. In addition to PNPase, *E. coli* has a second phosphorylytic RNA-degrading enzyme, RNase PH, which is implicated in the maturation

of tRNA [117, 118]. This phosphorylase is the founder of a superfamily of RNase PH-like enzymes, which include PNPase as well as components of the catalytic core of the eukaryotic exosome [10, 119]. The PNPase of *E. coli* is a trimer of identical subunits. Sequence analysis has suggested that each subunit contains two RNase PH-like domains acquired as the result of an ancient gene-duplication-fusion event [119]. In addition to the two RNase PH domains, PNPase also contains S1 and KH RNA-binding domains in the C-terminal region of the protein (Fig. 5-5A). The recent crystal structure of the PNPase from *Streptomyces antibioticus* has revealed for the first time the architecture of the catalytic site of a member of the RNase PH superfamily [120]. This work shows that the RNase PH domains in the PNPase monomer fold independently and pack closely together to form an intramolecular dimer. These dimers then assemble into a ring structure. Thus, the catalytic core of the PNPase trimer can be viewed as a hexameric ring assembled from the RNase PH domains (Fig. 5-5B). In the X-ray diffraction pattern, the S1 and KH domains were not

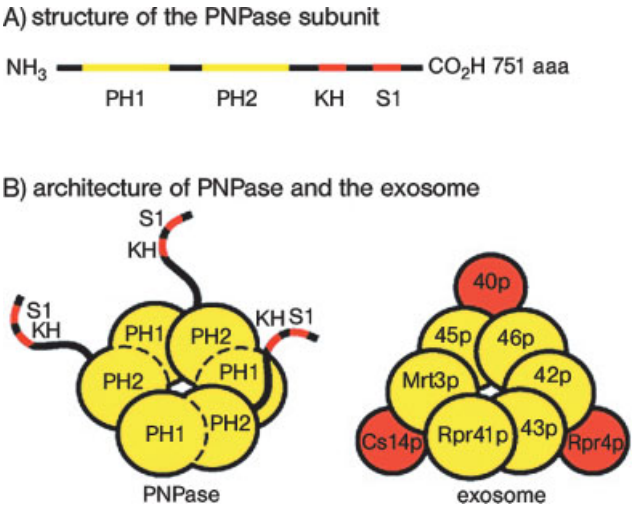


Figure 5-5 The eubacterial PNPase and the yeast exosome. (A) The *E. coli* PNPase subunit contains two RNase PH domains (yellow: PH1 and PH2) and C-terminal KH and S1 RNA-binding domains (red). It has been suggested, based on protein sequence comparisons, that the *pnp* gene arose from a duplication fusion of an ancient gene encoding an RNase PH-like enzyme (see text). (B) X-ray analysis [120, 121] has revealed that the catalytic core of PNPase can be viewed as a hexameric ring of RNase PH domains. In the model for PNPase presented here, the PH domains are in yellow; dashed lines indicate the domain boundaries; bold lines, the subunit boundaries. The 'tails' represent the C-terminal extensions containing the KH and S1 RNA-binding domains (red). The hexameric domain organization of the eubacterial PNPase can serve as a model for the phosphorylytic core of the yeast exosome, which is composed of six RNase PH-like subunits (yellow). The order of the exosome subunits in the hexameric ring (Rrp41p, Rrp43p, Rrp42p, Rrp46p, Rrp45p, and Mrt3p) is taken from a recent prediction [122]. Proteins associated with the exosome core, such as Rpr4p, Rpr40p, and Cs14p (red), contain S1 and KH RNA-binding domains that could serve the same function as the KH and S1 domains that are an integral part of the PNPase subunit.

detected suggesting that they could be part of a flexible structure. Nevertheless, molecular modelling indicates that the S1 and KH domains can form a 'crown', capping the catalytic core, which might serve to 'feed' RNA into the active site. It has been suggested that the ring structure of the catalytic core of PNPase could serve as a model for the organization of the RNase PH-like enzymes in the exosome [121]. A low-resolution structure of the yeast exosome, based on electron microscopy, and the mapping of protein–protein interactions by two-hybrid analysis supports this contention [122,123]. To transform PNPase into the exosome, the RNase PH domains forming the catalytic core of PNPase are replaced by the Rpr41p, 42p, etc., subunits (Fig. 5-5B). It is interesting to note that the non-phosphorylytic subunits of the exosome, Rpr4p, Rpr40p and Cs14p, contain RNA-binding motifs, including the S1 and KH domains found in eubacterial PNPase. Thus, whereas in PNPase, the PH1, PH2, S1 and KH domains are fused into a single polypeptide, in the exosome these domains exist in separate polypeptides. These considerations suggest that PNPase and the exosome might have evolved from an ancient phosphorylytic enzyme with a hexameric ring structure. Considering protein sequence alignments, an archeal 'exosome' has been predicted [124]. Two RNase PH-like proteins as well as a protein related to Rpr4p have been identified as part of an 'operon' in several different archeabacteria. It will be interesting to learn if these proteins actually assemble into an exosome-like complex.

5.10 DEAD-box ATPases

The identification of the DEAD-box ATPase, RhlB, in the *E. coli* degradosome was one of the first indications that RNA helicases could have an active role in the degradation of mRNA. The DEAD-box proteins are a family of putative ATP-dependent RNA helicases that have a conserved core sequence containing eight motifs including the amino acids D-E-A-D [125, 126]. Members of this family have been implicated in a variety of processes involving RNA including ribosome assembly, translation initiation and RNA splicing. The advantage of having an RNA helicase in an RNA-degrading complex was demonstrated *in vitro* with the RNA degradosome [77, 83]. RNAs with internally structured regions often impede the progress of enzymes such as PNPase, forcing the enzyme to pause. RhlB in the degradosome facilitates PNPase-mediated degradation of structured substrates in an ATP-hydrolysis-dependent reaction that is believed to involve the unwinding of RNA double strands. The ATPase activity of RhlB is strongly activated by its interaction with the CTH of RNase E and polypeptides derived from this region can form a complex with RhlB that is capable of unwinding short RNA helices *in vitro* [82]. Thus, the interaction between RNase E and RhlB controls ATPase activity and it could serve to give RhlB specificity via its physical association with an RNA-degrading complex. Since the degradosome, exosome and mtEXO complex all contain putative RNA helicases (Table 5-1), a role for these enzymes appears to be a common feature of complexes involved in 3′→5′ RNA degradation.

In the mtEXO complex, Suv3p, which is a DEAD-box ATPase, is required for RNA-degrading activity both *in vitro* and *in vivo*, and a requirement for ATP hydrolysis has clearly been demonstrated *in vitro* [112, 127]. Thus, Suv3p appears to be providing more than a simple RNA helicase function. It is conceivable that Suv3p could serve as a 'motor' that translocates the RNA substrate. The PNPase of *E. coli* has RNA-degrading activity by itself. However, the enzyme works close to the chemical equilibrium for the reaction. In the presence of low phosphate and high nucleotide concentrations, it can catalyse the reverse reaction, i.e., synthesis of polynucleotides from nucleotide diphosphates. Indeed, in mutant strains of *E. coli* deficient in poly(A) polymerase activity, an RNA synthetic activity attributed to PNPase has been described [128]. It was suggested that under certain 'micro conditions' within the cell, PNPase could work synthetically for brief periods in which short 3′ extensions are added to mRNA decay intermediates. This raises the issue of whether RhlB might have a role in regulating the degradative versus synthetic activity of PNPase. Although RhlB cannot alter the chemical equilibrium of the phosphorolysis reaction, it is conceivable that the energy of ATP hydrolysis could be coupled with the activity of PNPase in a kinetic control that promotes degradation.

5.11 Perspective

In *E. coli*, mRNAs are often polycistronic, and transcription and translation are coupled. Decay is initiated by endonucleases that fragment the mRNA. The principal endonuclease in *E. coli* mRNA decay is RNase E, which is 5′-end-dependent. Nevertheless, recent work suggests that RNase E can also initiate the decay of untranslated or poorly translated mRNA via an 'internal entry' pathway that is 5′-end-independent. RNase II and PNPase, the principal exonucleases in *E. coli* mRNA decay, degrade mRNA fragments to nucleotides in a 3′→5′ pathway. Poly(A) polymerase, which can add 3′ single-strand extensions to mRNA fragments, facilitates attack by the exonucleases. Internal regions of RNA structure that impede exonuclease activity can be unwound by RNA helicases. Thus, the main points for the control of degradation of *E. coli* mRNA are RNA structures in 5′- and 3′-UTRs, and in the intergenic regions of polycistronic messages, which modulate the activity of RNase E and the exonucleases. Whether *E. coli* is a model organism for the entire eubacterial kingdom is debatable. Notably, *B. subtilis* and related Gram-positive bacteria do not have identifiable RNase E homologs [71, 72]. RNase II, which is hydrolytic, and PNPase, which is phosphorylytic, seem to have redundant functions. Nevertheless, related proteins are widespread in the eubacteria and the eukaryotes, suggesting that there must be some advantage in having both types of 3′→5′ exonucleases.

The pathway of mRNA decay in *S. cerevisiae* differs considerably from that of *E. coli*. In yeast and other eukaryotes, the messages, which are monocistronic, are part of ribonucleoprotein complexes containing a wide diversity of RNA-binding proteins [129]. The 5′-cap and 3′-poly(A) structures are important for translation and mRNA stability. Messenger RNA decay is an orderly process in which 3′ deadenylation promotes 5′ decapping; this in turn leads to the arrest of translation initiation

and the degradation of the body of the mRNA. Thus, translation and mRNA stability are intimately linked. Nonsense-mediated mRNA decay, in which messages with premature stop codons are targeted for degradation, and non-stop mRNA decay, in which messages lacking stop codons are targeted for degradation, are examples of the importance of 'translatability' in mRNA stability [130–132]. The body of the mRNA is degraded to nucleotides by two distinct exonucleolytic pathways: 3′→5′ degradation involving the exosome and 5′→3′ degradation involving Xrn1p. The mRNA 5′-cap and the 5′→3′ exoribonuclease Xrn1p are specific features of the eukaryotic mRNA decay pathway. Many of the components involved in mRNA decay in yeast appear to be conserved in higher eukaryotes. It is thus generally believed that the yeast system will serve as a general model for mRNA decay. In higher eukaryotes, the lifetime of specific mRNAs can vary from minutes to days. The challenge now confronting researchers interested in eukaryotic mRNA decay is, in the framework of the established pathways, to elucidate how lifetimes are controlled. The eukaryotic mRNA-binding proteins probably to have a critical role in controlling the stability of specific messages.

In *E. coli*, RNase E and PNPase associate into a complex known as the RNA degradosome, which also contains the DEAD-box ATPase, RhlB. Related complexes, which are RNase E-based, have been described in other eubacteria suggesting that the degradosome-like machinery might be widespread. PNPase is a member of the RNase PH superfamily of phosphorylytic RNA-degrading enzymes, which includes six subunits of the yeast exosome. The crystal structure of a eubacterial PNPase and a recent low-resolution structure of the yeast exosome suggest that the RNase PH-like domains in these complexes have a conserved structure and that they assemble into a conserved hexameric ring architecture. Thus, despite considerable differences between mRNA decay in *E. coli* and *S. cerevisiae*, the core of the phosphorylytic RNA-degrading machinery appears to be conserved in eubacteria and eukaryotes.

Acknowledgements

Research in our laboratories has been supported by the Centre National de la Recherche Scientifique (CNRS) with additional funding from the Cancer Research Association (ARC) and the Fundamental Microbiology Program of the Ministry of Education and Research (MENRT). We thank L. Poljak for critical reading of this manuscript.

References

1 G. A. Coburn, G. A. Mackie, *Prog. Nucleic Acid Res. Mol. Biol.* **1999**, *62*, 55–108.
2 M. Grunberg-Manago, *Annu. Rev. Genet.* **1999**, *33*, 193–227.
3 P. Regnier, C. M. Arraiano, *Bioessays* **2000**, *22*, 235–244.
4 D. A. Steege, *RNA* **2000**, *6*, 1079–1090.
5 M. Tucker, R. Parker, *Annu. Rev. Biochem.* **2000**, *69*, 571–595.
6 P. Mitchell, D. Tollervey, *Nat. Struct. Biol.* **2000**, *7*, 843–846.
7 C. J. Wilusz, M. Wormington, S. W. Peltz, *Nat. Rev. Mol. Cell. Biol.* **2001**, *2*, 237–246.
8 S. R. Kushner, *J. Bacteriol.* **2002**, *184*, 4658–4665.
9 I. S. Mian, *Nucleic Acids Res.* **1997**, *25*, 3187–3195.
10 Y. Zuo, M. P. Deutscher, *Nucleic Acids Res.* **2001**, *29*, 1017–1026.
11 W. P. Donovan, S. R. Kushner, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 120–124.
12 D. Apirion, *Mol. Gen. Genet.* **1973**, *122*, 313–322.
13 J. G. Belasco, C. F. Higgins, *Gene* **1988**, *72*, 15–23.
14 E. B. O'Hara, J. A. Chekanova, C. A. Ingle et al., *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1807–1811.
15 C. A. Ingle, S. R. Kushner, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12926–12931.
16 E. Hajnsdorf, F. Braun, J. Haugel-Nielsen et al., *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3973–3977.
17 J. Haugel-Nielsen, E. Hajnsdorf, P. Regnier, *EMBO. J* **1996**, *15*, 3144–3152.
18 E. Blum, A. J. Carpousis, C. F. Higgins, *J. Biol. Chem.* **1999**, *274*, 4009–4016.
19 F. Xu, S. N. Cohen, *Nature* **1995**, *374*, 180–183.
20 G. A. Coburn, G. A. Mackie, *J. Mol. Biol.* **1998**, *279*, 1061–1074.
21 S. N. Cohen, *Cell* **1995**, *80*, 829–832.
22 A. J. Carpousis, N. F. Vanzo, L. C. Raynal, *Trends Genet.* **1999**, *15*, 24–28.
23 M. Dreyfus, P. Regnier, *Cell* **2002**, *111*, 611–613.
24 J. Ross, *Microbiol. Rev.* **1995**, *59*, 423–450.
25 J. Ross, *Trends. Genet.* **1996**, *12*, 171–175.
26 R. E. Dodson, D. J. Shapiro, *Prog. Nucleic Acid Res. Mol. Biol.* **2002**, *72*, 129–164.
27 A. Jacobson, S. W. Peltz, *Annu. Rev. Biochem.* **1996**, *65*, 693–739.
28 C. A. Beelman, R. Parker, *Cell* **1995**, *81*, 179–183.
29 C. A. Beelman, A. Stevens, G. Caponigro et al., *Nature* **1996**, *382*, 642–646.
30 T. E. LaGrandeur, R. Parker, *EMBO J* **1998**, *17*, 1487–1496.
31 J. S. Jacobs, A. R. Anderson, R. P. Parker, *EMBO J.* **1998**, *17*, 1497–1506.
32 C. Y. Chen, R. Gherzi, S. E. Ong et al., *Cell* **2001**, *107*, 451–464.
33 J. H. Withey, D. I. Friedman, *Curr. Opin. Microbiol.* **2002**, *5*, 154–159.
34 R. Gillet, B. Felden, *EMBO J.* **2001**, *20*, 2966–2976.
35 A. W. Karzai, E. D. Roche, R. T. Sauer, *Nat. Struct. Biol.* **2000**, *7*, 449–455.
36 A. W. Karzai, R. T. Sauer, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3040–3044.
37 A. van Hoof, P. A. Frischmeyer, H. C. Dietz et al., *Science* **2002**, *295*, 2262–2264.
38 R. S. Cormack, G. A. Mackie, *J. Mol. Biol.* **1992**, *228*, 1078–1090.
39 C. P. Ehretsmann, A. J. Carpousis, H. M. Krisch, *Genes Dev.* **1992**, *6*, 149–159.
40 S. Lin-Chao, T. T. Wong, K. J. McDowall et al., *J. Biol. Chem.* **1994**, *269*, 10797–10803.
41 K. J. McDowall, S. Lin-Chao, S. N. Cohen, *J. Biol. Chem.* **1994**, *269*, 10790–10796.
42 G. A. Mackie, *Nature* **1998**, *395*, 720–723.
43 P. Bouvet, J. G. Belasco, *Nature* **1992**, *360*, 488–491.
44 H. Huang, J. Liao, S. N. Cohen, *Nature* **1998**, *391*, 99–102.

45 G. A. Mackie, *J. Biol. Chem.* **2000,** *275,* 25069–25072.
46 K. E. Baker, G. A. Mackie, *Mol. Microbiol.* **2003,** *47,* 75–88.
47 V. J. Cannistraro, D. Kennell, *J. Bacteriol.* **1985,** *161,* 820–822.
48 G. G. Liou, W. N. Jane, S. N. Cohen et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 63–68.
49 Y. Feng, T. A. Vickers, S. N. Cohen, *Proc. Natl. Acad. Sci. U S A* **2002,** *99,* 14746–14751.
50 E. Hajnsdorf, O. Steier, L. Coscoy et al., *EMBO J.* **1994,** *13,* 3368–3377.
51 G. A. Mackie, *J. Bacteriol.* **1991,** *173,* 2488–2497.
52 D. P. Nierlich, G. J. Murakawa, *Prog. Nucleic Acid Res. Mol. Biol.* **1996,** *52,* 153–216.
53 C. Petersen, *Mol. Microbiol.* **1992,** *6,* 277–282.
54 M. Dreyfus, S. Joyce, *The interplay between translation and mRNA decay in procaryotes: a discussion on current paradigms,* Landes Bioscience, Georgetown 2002.
55 B. K. Ghora, D. Apirion, *Cell* **1978,** *15,* 1055–1066.
56 M. Ono, M. Kuwano, *J. Mol. Biol.* **1979,** *129,* 343–357.
57 E. A. Mudd, H. M. Krisch, C. F. Higgins, *Mol. Microbiol.* **1990,** *4,* 2127–2135.
58 P. Babitzke, S. R. Kushner, *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 1–5.
59 K. J. McDowall, R. G. Hernandez, S. Lin-Chao et al., *J. Bacteriol.* **1993,** *175,* 4245–4249.
60 Z. Li, M. P. Deutscher, *RNA* **2002,** *8,* 97–109.
61 M. C. Ow, S. R. Kushner, *Genes Dev.* **2002,** *16,* 1102–1115.
62 S. Casaregola, A. Jacq, D. Laoudj et al., *J. Mol. Biol.* **1992,** *228,* 30–40.
63 R. S. Cormack, J. L. Genereaux, G. A. Mackie, *Proc. Natl. Acad. Sci. USA* **1993,** *90,* 9006–9010.
64 L. Taraseviciene, G. R. Bjork, B. E. Uhlin, *J. Biol. Chem.* **1995,** *270,* 26391–26398.
65 K. J. McDowall, S. N. Cohen, *J. Mol. Biol.* **1996,** *255,* 349–355.
66 M. Bycroft, T. J. Hubbard, M. Proctor et al., *Cell* **1997,** *88,* 235–242.
67 Z. Li, S. Pandit, M. P. Deutscher, *EMBO J.* **1999,** *18,* 2878–2885.
68 M. R. Tock, A. P. Walsh, G. Carroll et al., *J. Biol. Chem.* **2000,** *275,* 8726–8732.
69 X. Jiang, A. Diwa, J. G. Belasco, *J. Bacteriol.* **2000,** *182,* 2468–2475.
70 V. R. Kaberdin, A. Miczak, J. S. Jakobsen et al., *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 11637–11642.
71 C. Condon, D. Brechemier-Baey, B. Beltchev et al., *RNA* **2001,** *7,* 242–253.
72 C. Condon, H. Putzer, *Nucleic Acids Res.* **2002,** *30,* 5339–5346.
73 K. Lee, J. A. Bernstein, S. N. Cohen, *Mol. Microbiol.* **2002,** *43,* 1445–1456.
74 A. J. Carpousis, G. Van Houwe, C. Ehretsmann et al., *Cell* **1994,** *76,* 889–900.
75 B. Py, H. Causton, E. A. Mudd et al., *Mol. Microbiol.* **1994,** *14,* 717–729.
76 A. J. Carpousis, A. Leroy, N. Vanzo et al., *Meth. Enzymol.* **2001,** *342,* 333–345.
77 B. Py, C. F. Higgins, H. M. Krisch et al., *Nature* **1996,** *381,* 169–172.
78 A. Miczak, V. R. Kaberdin, C. L. Wei et al., *Proc. Natl. Acad. Sci. USA* **1996,** *93,* 3865–3869.
79 E. Blum, B. Py, A. J. Carpousis et al., *Mol. Microbiol.* **1997,** *26,* 387–398.
80 L. C. Raynal, A. J. Carpousis, *Mol. Microbiol.* **1999,** *32,* 765–775.
81 Y. Feng, H. Huang, J. Liao et al., *J. Biol. Chem.* **2001,** *276,* 31651–31656.
82 N. F. Vanzo, Y. S. Li, B. Py et al., *Genes Dev.* **1998,** *12,* 2770–2781.
83 G. A. Coburn, X. Miao, D. J. Briant et al., *Genes Dev.* **1999,** *13,* 2594–2603.
84 G. A. Mackie, G. A. Coburn, X. Miao et al., *Meth. Enzymol.* **2001,** *342,* 346–356.
85 P. J. Lopez, I. Marchand, O. Yarchuk et al., *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 6067–6072.
86 F. Soderbom, U. Binnie, M. Masters et al., *Mol. Microbiol.* **1997,** *26,* 493–504.
87 N. Dam Mikkelsen, K. Gerdes, *Mol. Microbiol.* **1997,** *26,* 311–320.

88 M. Kido, K. Yamanaka, T. Mitani et al., *J. Bacteriol.* **1996**, *178*, 3917–3925.
89 P. J. Lopez, I. Marchand, S. A. Joyce et al., *Mol. Microbiol.* **1999**, *33*, 188–199.
90 A. Leroy, N. F. Vanzo, S. Sousa et al., *Mol. Microbiol.* **2002**, *45*, 1231–1243.
91 I. Marchand, A. W. Nicholson, M. Dreyfus, *Mol. Microbiol.* **2001**, *42*, 767–776.
92 I. Iost, M. Dreyfus, *EMBO J.* **1995**, *14*, 3252–3261.
93 E. A. Mudd, C. F. Higgins, *Mol. Microbiol.* **1993**, *9*, 557–568.
94 C. Jain, J. G. Belasco, *Genes Dev.* **1995**, *9*, 84–96.
95 A. Diwa, A. L. Bricker, C. Jain et al., *Genes Dev.* **2000**, *14*, 1249–1260.
96 A. A. Diwa, J. G. Belasco, *J. Biol. Chem.* **2002**, *277*, 20415–20422.
97 S. Sousa, I. Marchand, M. Dreyfus, *Mol. Microbiol.* **2001**, *42*, 867–878.
98 C. Jain, A. Deana, J. G. Belasco, *Mol. Microbiol.* **2002**, *43*, 1053–1064.
99 P. Regnier, M. Grunberg-Manago, C. Portier, *J. Biol. Chem.* **1987**, *262*, 63–68.
100 M. Robert-Le Meur, C. Portier, *EMBO J.* **1992**, *11*, 2633–2641.
101 A. C. Jarrige, N. Mathy, C. Portier, *EMBO J.* **2001**, *20*, 6845–6855.
102 R. Zilhao, F. Cairrao, P. Regnier et al., *Mol. Microbiol.* **1996**, *20*, 1033–1042.
103 B. K. Mohanty, S. R. Kushner, *Mol. Microbiol.* **2002**, *45*, 1315–1324.
104 V. Shen, H. Bremer, *J. Bacteriol.* **1977**, *130*, 1098–1108.
105 G. Brawerman: *mRNA Degradation in Eukaryotic Cells: An Overview,* Academic Press, San Diego 1993.
106 S. Jager, O. Fuhrmann, C. Heck et al., *Nucleic Acids Res.* **2001**, *29*, 4581–4588.
107 T. Platt, *Mol. Microbiol.* **1994**, *11*, 983–990.
108 S. Baginsky, A. Shteiman-Kotler, V. Liveanu et al., *RNA* **2001**, *7*, 1464–1475.
109 J. M. Hagege, S. N. Cohen, *Mol. Microbiol.* **1997**, *25*, 1077–1090.
110 K. Lee, S. N. Cohen, *Mol. Microbiol.* **2003**, *48*, 349–360
111 A. Dziembowski, P. P. Stepien, *Meth. Enzymol.* **2001**, *342*, 367–378.
112 S. P. Margossian, H. Li, H. P. Zassenhaus et al., *Cell* **1996**, *84*, 199–209.
113 C. Torchet, C. Bousquet-Antonelli, L. Milligan et al., *Mol. Cell* **2002**, *9*, 1285–1296.
114 C. Bousquet-Antonelli, C. Presutti, D. Tollervey, *Cell* **2000**, *102*, 765–775.
115 C. Allmang, E. Petfalski, A. Podtelejnikov et al., *Genes Dev.* **1999**, *13*, 2148–2158.
116 A. M. Estevez, T. Kempf, C. Clayton, *EMBO J.* **2001**, *20*, 3831–3839.
117 K. O. Kelly, N. B. Reuven, Z. Li et al., *J. Biol. Chem.* **1992**, *267*, 16015–16018.
118 Z. Li, M. P. Deutscher, *J. Biol. Chem.* **1994**, *269*, 6064–6071.
119 A. Bateman, E. Birney, R. Durbin et al., *Nucleic Acids Res.* **2000**, *28*, 263–266.
120 M. F. Symmons, G. H. Jones, B. F. Luisi, *Struct. Fold Des.* **2000**, *8*, 1215–1226.
121 M. F. Symmons, M. G. Williams, B. F. Luisi et al., *Trends Biochem. Sci.* **2002**, *27*, 11–18.
122 P. Aloy, F. D. Ciccarelli, C. Leutwein et al., *EMBO Rep.* **2002**, *3*, 628–635.
123 R. Raijmakers, W. V. Egberts, W. J. van Venrooij et al., *J. Mol. Biol.* **2002**, *323*, 653–663.
124 E. V. Koonin, Y. I. Wolf, L. Aravind, *Genome Res.* **2001**, *11*, 240–252.
125 S. R. Schmid, P. Linder, *Mol. Microbiol.* **1992**, *6*, 283–291.
126 N. K. Tanner, P. Linder, *Mol. Cell* **2001**, *8*, 251–262.
127 J. Min, R. M. Heuertz, H. P. Zassenhaus, *J. Biol. Chem.* **1993**, *268*, 7350–7357.
128 B. K. Mohanty, S. R. Kushner, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11966–11971.
129 G. Dreyfuss, V. N. Kim, N. Kataoka, *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 195–205.
130 L. E. Maquat, *Curr. Biol.* **2002**, *12*, R196–197.
131 C. J. Wilusz, W. Wang, S. W. Peltz, *Genes Dev.* **2001**, *15*, 2781–2785.
132 S. Vasudevan, S. W. Peltz, C. J. Wilusz, *Bioessays* **2002**, *24*, 785–788.

6
tRNA Locations on the Ribosome

Knud H. Nierhaus

6.1
tRNAs Move through Functional Sites on the Ribosome

The ribosome harbors three well-defined binding sites for tRNAs: the A- and P-sites, where the aminoacyl- and peptidyl-tRNAs reside before peptide-bond formation, respectively, and the E-site, a site specific for deacylated tRNA from which the tRNA exits the ribosome. Localization of tRNA-related functional centers such as the PTF center or the decoding center on the ribosome has always been an important issue in the translational field. Many techniques have been used, developed and even invented to probe the interaction of tRNAs with the ribosome long before high-resolution structures became available.

Site-directed crosslinking (reviewed in Ref. [1]) of tRNAs identified the decoding site on the 30S and the PTF ring in the 50S subunit as functional centers. The anticodon loop of P-site bound tRNA crosslinks to C1400 (h44) of the 16S rRNA [2] and benzophenone attached to the amino acid of the P-site peptidyl-tRNA crosslinks to A2451 and C2452 and from the A-site peptidyl-tRNA to U2584 and U2585 with high yields [3]. In addition, various groups have shown that G2553 is located near the CCA end of an A-site substrate in the PTF center. This topological feature was convincingly confirmed by crosslinking the antibiotic puromycin, which functions as an analog of the tRNA acceptor end, to G2553. After crosslinking, the attached puromycin could still undergo peptide-bond formation [4]. A compilation of tRNA crosslinks can be found in the ribosomal crosslinking database (RDB [5]).

Distinct sets of rRNA bases have been assigned to contact tRNAs in A-, P- or E-sites by applying various techniques (Table 6-1). From the crystal structures of the 70S ribosome and the 30S subunit in the presence of tRNAs or tRNA fragments it is clear that most protections can be explained either by direct contacts with bases or by local conformational changes within the binding regions (discussed in depth in Ref. [6]). On the other hand, some of the P-site protections on the 30S subunit are actually E-site contacts. Protections of bases 1339, 1340 and 1381 are most probably caused by the backbone of the E-site tRNA and the protections in the 690 loop (h23) are caused by the anticodon loop of E-site bound tRNA (34–36). The 790 loop is a contact site for both E- and P-site tRNAs and the protection might result from either of these tRNAs. This mis-assignment has implications for the role of

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

Table 6-1 tRNA contacts with rRNA bases in the A-, P-, and E-sites

tRNA location / method	16S or 23S RNA / residues	Reference
A-site / protection	16S/530 loop: G529, G530, U531; helix 44: A1408, A1492, A1493, G1494; enhanced reactivity, helix 27: A892, G1405 23S/ A1439, C2254, A2439, A2451, G2553, pseudoU2555, A2602, U2609	37, 38, 7
A-site / site-directed muta-genesis	C74 of A-site tRNA base-pairs with G2553	39
P-site / protection	16S/ A532, G693, A794, C795, G926, G966, G1338, A1339, U1381 and in helix 44: C1399, C1400, G140123S/ A1916, A1918, U1926, G2251, G2253, A2439, A2451, G2505, U2506, U2584, U2585, A2602 (enhanced), and G2252 in the loop of H80, the P-loop	37, 38, 7
P-site / site-directed mutagenesis	C74 of P-site tRNA base-pairs with G2252 of H80 (P-loop)	40
P-site / interference	23S/modification of G2252, A2451, U2506, and U2585 prevents tRNA binding to the P-site	37
E-site / protection	23S/ G2112, G2116, A2169 at the L1-binding site; modification of C2394 interferes with E-site binding	7, 41, 42, 47

the E-site in ribosome function and consequences for the hybrid states model of elongation, a model that is interpreted on the basis of these protection experiments (see Chap. 8.1.1).

A ribosome discriminates tRNAs according to the coding sequence of the mRNA; however, during the translation process, they must be capable of binding between 33 and more than 50 (45 in *Escherichia coli*) different tRNA species; note that a tRNA species is defined solely by its anticodon. Thus the ribosome has to utilize conserved features of a tRNA to bind it. One such feature is the universally conserved CCA 3′-end of the tRNAs, which plays an important role in ribosome binding. Seventeen out of the 20 protections observed in the 23S rRNA with complete tRNAs are also seen with CCA fragments alone [7]. Furthermore, the binding of deacylated tRNAs to the E-site is dependent on an intact CCA end [8].

The first evidence that tRNAs do not interact exclusively using the anticodon loop and the CCA acceptor end, but instead are embedded in a ribosomal matrix, derives from phosphorothioate cleavage experiments. Iodine cleavage of phosphorothioated tRNAs bound to the ribosome yielded characteristic protection patterns [9, 10]. Since all tRNAs are conserved in terms of tertiary structure, at least some of the phosphate groups in the backbone might provide important binding determinants. In contrast with the protection pattern of the phosphorothioated mRNA, which locates only within the codon region [11], the cleavage patterns of

tRNAs bound to the ribosome are characteristic for their binding position and the functional state of the ribosome and cover the whole structure of a tRNA [9, 10]. In other words, the mRNA is hardly contacting the ribosome although about a sequence of 40 nt is covered by the ribosome [12, 13]. These rare contacts strikingly contrast with the extensive contacts of a tRNA, leading to the important conclusion that the tRNAs are actively transported during the translocation reaction, whereas the mRNAs are coupled with the movement of the tRNAs by the two adjacent codon–anticodon interactions (see Chap. 8.1).

The tRNA patterns, which differ significantly from the pattern of tRNAs in solution, have been interpreted to reflect the microtopography of the binding site, emphasizing intimate contacts between the ribosome and the entire tRNA surface. Contact patterns between tRNA nucleotides 29 and 43 (comprising anticodon loop and two adjacent stem base-pairs) are due to components of the 30S subunit, whereas the remaining 85% of the tRNA, viz. the acceptor stem, the T and D loops, is in contact with the 50S subunit. The 30S and 50S cleavage patterns are additive to yield the 70S pattern ([14, 15]; see Fig. 6-1A). Crystallographic data obtained with 30S subunits and 70S ribosomes [16, 6] are in perfect agreement with the 30S–50S contacts at the P-site of 70S ribosomes (Fig. 6-1B).

Phosphorothioated tRNAs bound to the ribosome yield two characteristic cleavage patterns: one observed in the P- or E-site (termed ε for its specific appearance at the E-site) and the other in the A- or the P-site (termed α for A-site; see also Chap. 8.1.2). The α- pattern shows few protection sites but several sites of enhancement, whereas, in contrast, the ε- pattern exhibits extensive protection sites and only few positions with enhanced iodine cleavage reactivity. This might reflect that the tRNAs bound at the E- and P-sites are buried in the ribosomal matrix to a higher extent than the A-site tRNA, an observation which is in agreement with the crystal structure of a programed 70S carrying tRNAs [6].

6.2
Visualization of tRNAs on the Ribosome

Biochemical studies have established that the ribosome has three tRNA-binding sites [17–20]. Contrary with this, three-dimensional (3D) cryo-EM has revealed five different tRNA positions on the ribosome, the classic A-, P-, and, E-sites and additional two sites termed P/E and E2 (see Table 6-4 and Refs. [21, 22] for a compilation of identified tRNA sites).

Two early cryo-EM studies identified three tRNA positions on the ribosome [23, 24]. The A-site was localized close to the L7/L12 stalk of the ribosome, the P-site tRNA spanning the inter-subunit space from the neck of the small subunit to the 50S subunit and the E-site tRNA was observed close to the mushroom-shaped L1 protuberance. Although the studies agreed on the position of the P-site tRNA, the locations of the A- and E-site tRNAs were remarkably different. The E-site puzzle was resolved by subsequent studies which showed that the E-site tRNA position was strongly dependent on buffer conditions, and the position of a single tRNA on the ribosome at the P-site on both the ionic conditions and the charging state of the tRNA [25, 26].

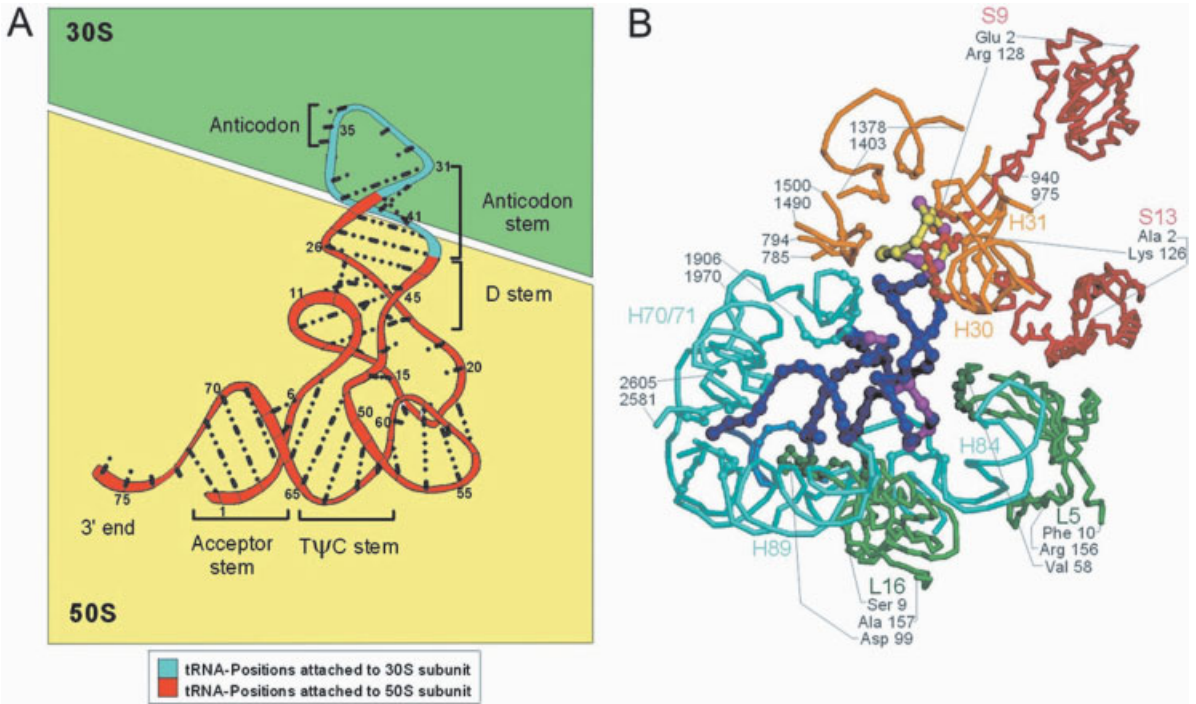


Figure 6-1 tRNA contacts with the ribosomal subunits at the P-site of 70S ribosomes. (A) Contact pattern obtained with protection experiments using phosphorothioated tRNAs (according to Ref. [15]). (B) Ribosomal components that come nearer than 10 Å to the tRNA at the P-site of the 70S ribosome (view from the E-site) according to Ref. [6]. The tRNA is colored: yellow, contacts with the small subunit; blue, contacts with the large subunit; pink, sites near to both tRNA and $tRNA^{Met}$ at the P-site according to Ref. [15]. Regions of the ribosomal components that are within a 10 Å proximity of the P-site tRNA are indicated with spheres. 30S components: gold, 16rRNA; red, S-proteins. 50S components: cyan, 23S rRNA; green, L-proteins.

A tRNA at the E-site was only observed under physiological buffer conditions, whereas under non-physiological conditions a tRNA was instead present at the E2 position that possibly represents an unstable intermediary state following release from the E-site and before dissociation from the 70S ribosome. A recent re-evaluation suggests that the E2 position might be even a misinterpretation caused by a conformational change of the L1 protuberance (C. M. T. Spahn, pers. comm.). Similarly, a tRNA at the hybrid site P/E was exclusively found under non-physiological conditions and thus probably represents a buffer artifact, but certainly not a ribosomal state with a significant population during the elongation cycle [25]; see Chap. 8.1.1 for discussion). We see that a critical discussion melts down the number of tRNA-binding sites on the ribosome again to the classical three sites, A, P, and E. The only exception of this view is the binding of the incoming ternary complex aa–tRNA•EF–Tu•GTP that has been termed A/T site (see, e.g., Fig 8-1; see also Ref. [27]). In fact, in this configuration, codon–anticodon interaction is checked at the decoding center of the A-site as a first step of the A-site occupation corresponding to the "low-affinity state" of the A-site in the allosteric three-site model (see Chap. 8.1.2).

An fMet-tRNA^{fMet} bound to the ribosomal P-site was visualized by cryo-EM at 15 Å and further refined to 11.5 Å [28, 29]. At 11.5 Å resolution, the tRNA X-ray structure could be fitted directly into an L-shaped P-site mass, oriented such that the CCA arm faces toward the entrance of the tunnel on the 50S subunit and the longer anticodon arm faces towards the cleft on small subunit [21]. The tRNA density makes four contacts with the ribosome. The backbone of G57, on the tip of the tRNA elbow, extends towards the 50S central protuberance, the C12–C23 base-pair on the D stem contacts both the 50S body and the 30S platform, and U33 and A37 in the anticodon loop extend into the 30S body and head [29]. The contact sites of the P-site tRNA seen in these cryo-EM studies [28] agreed well with the phosphate contact pattern derived from phosphorothioate studies mentioned in the preceding section [9].

On the 30S subunit, the anticodon ends of P- and A-site tRNAs before translocation (PRE state), as well as P- and E-sites after translocation (POST state), are in close proximity to one another (Figs. 6-2A and B), such that nucleotide 37 in the anticodon loops are 20 and 16 Å away from each other in the PRE and POST states, respectively. Following peptide-bond formation the CCA ends of P- and A-site tRNAs are 17 Å apart (see Chap. 8.4 for more details), whereas after translocation the CCA end of E-site tRNA is turned towards the L1 stalk and measures 60 Å from the P-site CCA end. The position of the CCA end at the E-site has been determined in crystals of 50S subunits of the archeon *Haloarcula marismortui* [30]. The fixation differs from those observed in A- and P-sites, where the CCA ends are held via Watson–Crick base pairs with the rRNA. Instead, at the E-site, the A76 is fixed by an intricate network of hydrogen bonds to nucleotides conserved in all three kingdoms of life (Fig. 6-3): A76 is hydrogen-bonded to C2394 (*E. coli* nomenclature) and to the phospho-oxygen of A2422, the sugar-phosphate backbone of 76 to C2394. Furthermore, A76 is stacked between G2421 and A2422 (Fig. 6-3). The tight packing of A76 at the E-site leaves no room for an amino acid linked to the A76 via an ester bond and thus explains the earlier finding that this site is specific for deacylated tRNA [20], and that

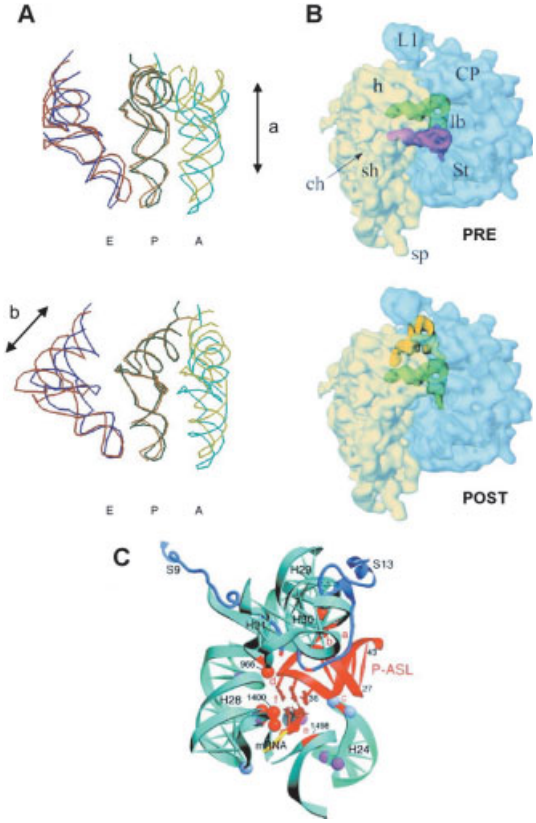


Figure 6-2 tRNA positions on the ribosome. (A) Relative positions of the three tRNAs bound at the A-, P- and E-sites. The P and C1′ atoms of the P-site tRNAs are used to align the molecules of the cryo-EM (brown) with that from the X-ray work (dark green; Refs. [21, 6]). The two studies agree on the position of the tRNAs on the ribosome. From this comparison it can be seen that the A-site tRNAs (cryo-EM in olive green, X-ray in cyan) are shifted relative to each other along the anticodon stem axis (arrow a), and the E-site-bound tRNAs along the acceptor stem axis (arrow b, cryo-EM in red and X-ray in blue). Note that the anticodon regions of all three tRNAs are in close proximity to each other. (B) tRNAs in the PRE and the POST states of the *E. coli* ribosome. Cryo-EM reconstructions of tRNAs bound to the 70S ribosome. PRE state: tRNAs bound to the A-site (pink) and P-site (green). POST state: tRNAs bound to the P-site (green) and E-site (yellow). The small 30S subunit is shown in yellow, the large 50S subunit in blue. To demonstrate the tRNA positions, the 70S ribosome is presented as a semitransparent surface. (C) Fixation of the codon–anticodon duplex at the ribosomal P-site according to Ref. [6]. The 16S rRNA is shown in cyan, ribosomal protein S13 in blue and the anti-codon stem-loop of the P-site tRNA in red. 16S rRNA contacts with the P-site tRNA are indicated and labeled in red (a–f).

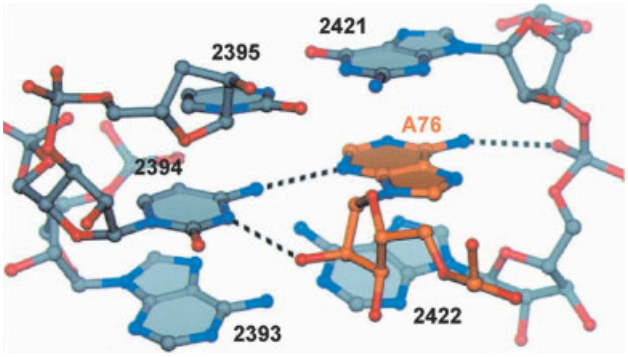


Figure 6-3 A76 of the CCA end at the E-site is held by network of hydrogen bonds. For details see text (taken from Ref. [30]).

a tRNA at the E-site requires an intact CCA end [8]. Only CCA-Gly, the smallest aminoacyl residue, could possibly fit into the E-site. However, this possibility is not relevant for protein synthesis, since there is always a deacylated tRNA at the P-site before translocation and after peptide-bond formation (see, e.g., models of the elongation cycle; Figs. 8.2A and B). The only situation where an aminoacyl-tRNA binds directly to the ribosome is during initiation and here, as mentioned, the fMet moiety itself would prevent the tRNA from binding at the E-site and probably the involvement of the initiation factors for directing this binding would provide further protection against this.

The final two nucleotides of the CCA end of an E-tRNA pass through a loop of protein L44e. Although L44e does not exist in bacteria, the bacterial protein L33 mimics the shape of the globular part of L44e and L31, the extended part of L44e. In fact, it may be the case that binding of a CCA at the E-site of bacterial ribosomes would entail the insertion through the loop extension of L31, although whether this similarity in the involvement of the loop region of the ribosomal proteins is functionally significant is unclear. Since the fold of L31 and L33 is significantly different from that of L44e, whereas the critical C2395 is conserved in over 99% of all species, the authors suspect that the last universal common ancestor (LUCA) fixed the CCA end of the E-tRNA via rRNA, and that protein components were added after separation into kingdoms [30].

Notably, a P-site tRNA occupies virtually the same position in the ribosome before and after translocation (PRE and the POST states, respectively). The angles between the tRNAs in the PRE at A- and P-sites and in the POST state at P- and E-sites are 39° and 35°, respectively [21].

Combined crystal structures of three different tRNA 70S complexes at 7.8 and 5.5 Å resolution yields a wealth of information regarding tRNA–ribosome interactions for all three sites [31, 6] (see Table 6-2 [15]). The positions of the tRNAs on the ribosome are in good agreement with those from the cryo-EM work of Frank and colleagues [21] concerning authentic PRE and POST states of ribosomes,

Table 6-2 Contact sites with tRNA phosphates at the P-site that were strongly protected against iodine (I_2) access in two different elongator tRNAs, viz. $tRNA^{Phe}$ and elongator $tRNA^{Met}$ (adapted from Ref. [15])

5′-phosphate of tRNA base	Residue of rRNA or r-protein nearer than 10 Å (rRNA/nt or r-prot/aa; see Ref. [6])	Evolutionary conservation (in percent of species)	
		Eubact.	Three domains
Y11	23S/1909, 1910, 1923, 1924	80–90, ≥95, ≥95, ≥95	<80, 90–95, ≥95, <80
G30	16S/1230	≥95	80-90
	S13/Lys121	Lys or Arg at position 120	
Y32	16S/1341	≥95	≥95
	S9/Ser126, Lys127, Arg128	126 and 127: ~50% cons. 128: ~90% conserved	
G34	(anticodon)		
C41	16S/1339, 1340	≥95, <80	≥95, <80
T54	23S/2280, 2327	≥95, ≥95	80-90, <80
C56	L5/Arg56 and Glu65	Arg or Lys at positions 56 and 64	
A58	Protected via tertiary folding of tRNA		
U59	Protected via tertiary folding of tRNA		
Y60	Protected via tertiary folding of tRNA		

Note: nt, nucleotidyl residue; aa, aminoacyl residue, eubact., eubacterial domain; three domains, the eubacterial, archeal and eukaryotic domains. The conservation data concerning rRNA were obtained from The Gutell Lab Pages (www.rna.icmb.utexas.edu/csi), the sequences of the ribosomal proteins (r-prot.) were obtained from the Sequence Retrieval System (www.expasy.ch/srs5/), the alignment followed Ref. [43] according to www.expasy.ch/srs5/.

although closer inspection of the relative orientations reveal some minor deviations. In Fig. 6-2(A), the P-site tRNA molecules of the cryo-EM [21] and the X-ray studies [6] are aligned for comparison. In the crystallographic study, the P-site-bound $tRNA^{fMet}$ is slightly kinked at the D stem–anticodon junction in comparison with the X-ray structure of the free tRNA that was used to fit the cryo-EM map. In both the cryo-EM with 12–17 Å resolution and the X-ray analysis with 5–7 Å resolution single-stranded RNA cannot be unequivocally identified, and thus the 3′ single-stranded end of $tRNA^{fMet}$ was deduced from the highly resolved tRNA crystal structure (see, e.g., Ref. [32]). Alignment of the crystal and cryo-EM maps based on the P-site tRNA positions shows the A- and E-site tRNAs in slightly different positions. The A-site tRNAs are shifted with respect to each other along the anticodon stem axes and the E-site tRNAs along the acceptor stem axes. The E-site tRNA in the X-ray study was reported to be substantially distorted [6] relative to the X-ray structure of the free tRNA that was used to fit the cryo-EM data. The distortion might be

due to the fact that the E-site-bound tRNA is non-cognate and thus cannot undergo base-pairing with the E-site codon of the mRNA. Note that during an elongation cycle a deacylated tRNA at the E-site is always a cognate one. The juxtaposition of codon and anticodon at the E-site of the 70S crystal structure makes it probable that under physiological conditions codon–anticodon interaction occurs at this site in agreement with biochemical data [33–35].

6.3 tRNA–ribosome Contacts

In this section, we will consider in detail the contacts of a tRNA at each of the three tRNA-binding sites A, P and E.

Ramakrishnan and colleagues [16] presented the first high-resolution view at 3.1 Å of the P- and A-site tRNA interactions with the 30S subunit. Fortuitously, crystal packing of the *T. thermophilus* 30S subunits placed the spur (h6) of one subunit in the P-site of another, thereby mimicking the anticodon stem-loop of a P-site-bound tRNA. Remarkably, the mRNA base-pairing partner was provided by the 3′-end of 16S rRNA which, folding back upon itself, extended into the decoding center. These crystals were then soaked with an anticodon stem-loop fragment of a tRNA (ASL-tRNA) and a six base poly(U) mRNA fragment to include A-site interactions within the scope of these studies [36]. At the A-site, the ribosome scans the mRNA–tRNA codon–anticodon base-pairing to ensure high-fidelity decoding of aa-tRNAs and to maintain the reading frame (refer Chap. 8.2; [36]).

On the small subunit, the P-site-bound tRNA is fixed very tightly via six interactions with the 16S RNA. RNA elements 1338–1341 (hydrogen bonding to bases) and 1229–1230 (sugar-phosphate backbone) of the 16S rRNA interact in the minor groove of the acceptor stem. Only one hydrogen bond appears to be base-specific. The interaction is supported by the C-terminal tails of proteins S13 and S9. The base corresponding to tRNA position 34 is stacked on C1400, whereas A790 packs against tRNA positions 40 and 41. The P-site codon–anticodon helix is positioned in the major groove of the penultimate helix (h44) and is fixed with a number of "ribosomal fingers" mainly to the sugar-phosphate backbone (Fig. 6-2C; see Table 6-2 for P-site contacts of the ribosome with two different elongator tRNAs; for involvement of h44, bases A1492 and A1493, in the decoding mechanism see Chap. 8.2).

Comprehensive analyses of tRNA:ribosome interactions have been described by Noller and colleagues on the basis of *T. Thermophilus* 70S tRNA co-crystals [31, 6]. A- and P-site tRNAs exhibit similar modes of interaction with the large ribosomal subunit. The 23S rRNA helices, H80-81 in the P-site and H89 in the A-site run parallel to the acceptor stem of the tRNAs making minor groove–minor groove contacts. Proteins, L5 and L16 in A- and P-sites respectively, contact the T-loop at the elbow of the tRNA. Additionally, the A-site finger (H38) contacts the elbow (D and T loops) of the A-site tRNA. H69 and H93 fix both tRNAs simultaneously: helix 69 is "sandwiched" between the top of the D-stem of the P-site tRNA (from the minor groove side) and the D-stem of the A-site tRNA from the major groove side, whereas H93 squeezes between the respective CCA ends. In accordance with bio-

chemical data, G2553 in H89 is positioned to base-pair with C75 of the A-site tRNA, whereas G2252 in H80 base-pairs with C74 of P-site-bound tRNA.

The E-site on the 30S subunit contains proteins S7 and S11, a highly conserved β–hairpin of S11 contacts the backbone of the anticodon stem, whereas α-helix 6 of S7 faces the anticodon side of the anticodon loop. 16S rRNA contacts include h29 (1339, 1340), h28 (1382), the 690 loop, and 790 loop [6].

In the large subunit, the E-site tRNA forms protein contacts at the elbow in similar fashion to the A- and P-site tRNAs. The elbow neighbors protein L1 and H77 of 23S RNA, i.e., both elements that constitute the characteristic L1 protuberance. Other E-site tRNA:23S rRNA contacts are seen with nucleotides 1–5 and 71–76 at the end of the acceptor stem that is buried in a deep pocket made of RNA and protein L33. Here minor groove–minor groove interactions with H68 are evident as well as several interactions with H11, H74, H75, and protein L33.

References

1 R. Brimacombe, *Eur. J. Biochem.* **1995,** *230,* 365–383.
2 J. B. Prince, B. H. Taylor, D. L. Thurlow et al., *Proc. Natl. Acad. Sci. USA* **1982,** *79,* 5450–5454.
3 A. Barta, E. Kuechler, G. Steiner: in *The Ribosome: Structure, Function and Evolution,* eds W. E. Hill, A. Dahlberg, R. A. Garrett et al., American Society for Microbiology, Washington, DC 1990, 359–365.
4 R. Green, C. Switzer, H. Noller, *Science* **1998,** *280,* 286–289.
5 P. V. Baranov, P. V. Sergiev, O. A. Dontsova et al., *Nucleic Acids Res.* **1998,** *26,* 187–189.
6 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001,** *292,* 883–896.
7 D. Moazed, H. F. Noller, *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 3725–3728.
8 R. Lill, A. Lepier, F. Schwägele et al., *J. Mol. Biol.* **1988,** *203,* 699–705.
9 M. Dabrowski, C. M. T. Spahn, K. H. Nierhaus, *EMBO J.* **1995,** *14,* 4872–4882.
10 M. Dabrowski, C. M. T. Spahn, M. A. Schäfer et al., *J. Biol. Chem.* **1998,** *273,* 32793–32800.
11 E. V. Alexeeva, O. V. Shpanchenko, O. A. Dontsova et al., *Nucleic Acids Res.* **1996,** *24,* 2228–2235.
12 D. Beyer, E. Skripkin, J. Wadzack et al., *J. Biol. Chem.* **1994,** *269,* 30713–30717.
13 G. Z. Yusupova, M. M. Yusupov, J. H. Cate et al., *Cell* **2001,** *106,* 233–241.
14 K. H. Nierhaus, C. M. T. Spahn, N. Burkhardt et al.: in *The Ribosome. Structure, Function, Antibiotics, and Cellular Interactions,* eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, 319–335.
15 M. A. Schäfer, A. O. Tastan, S. Patzke et al., *J. Biol. Chem.* **2002,** *277,* 19095–19105.
16 A. P. Carter, W. M. Clemons, D. E. Brodersen et al., *Nature* **2000,** *407,* 340–348.
17 R. A. Grajevskaja, Y. V. Ivanov, E. M. Saminsky, *Eur. J. Biochem.* **1982,** *128,* 47–52.
18 R. Lill, J. M. Robertson, W. Wintermeyer, *Biochemistry* **1984,** *23,* 6710–6717.
19 K. H. Nierhaus, *Mol. Microbiol.* **1993,** *9,* 661–669.
20 H.-J. Rheinberger, H. Sternbach, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1981,** *78,* 5310–5314.
21 R. K. Agrawal, C. M. T. Spahn, P. Penczek et al., *J. Cell. Biol.* **2000,** *150,* 447–459.

22 C. M. T. Spahn, K. H. Nierhaus, *Biol. Chem.* **1998,** *379,* 753–772.
23 R. K. Agrawal, P. Penczek, R. A. Grassucci et al., *Science* **1996,** *271,* 1000–1002.
24 H. Stark, E. V. Orlova, J. Rinke-Appel et al., *Cell* **1997,** *88,* 19–28.
25 R. K. Agrawal, A. B. Heagle, P. Penczek et al., *Nat. Struct. Biol.* **1999,** *6,* 643–647.
26 G. Blaha, U. Stelzl, C. M. T. Spahn et al., *Meth. Enzymol.* **2000,** *317,* 292–309.
27 M. Valle, J. Sengupta, N. K. Swami et al., *EMBO J.* **2002,** *21,* 3557–3567.
28 I. S. Gabashvili, R. K. Agrawal, C. M. T. Spahn et al., *Cell* **2000,** *100,* 537–549.
29 A. Malhotra, P. Penczek, R. K. Agrawal et al., *J. Mol. Biol.* **1998,** *280,* 103–116.
30 T. M. Schmeing, P. B. Moore, T. A. Steitz, *RNA* **2003,** *9,* 1345–1352.
31 J. H. Cate, M. M. Yusupov, G. Z. Yusupova et al., *Science* **1999,** *285,* 2095–2104.
32 S. H. Kim, F. L. Suddath, G. J. Quigley et al., *Science* **1974,** *185,* 435–440.
33 H.-J. Rheinberger, K. H. Nierhaus, *FEBS Lett.* **1986,** *204,* 97–99.
34 H.-J. Rheinberger, H. Sternbach, K. H. Nierhaus, *J. Biol. Chem.* **1986,** *261,* 9140–9143.
35 F. J. Triana–Alonso, K. Chakraburtty, K. H. Nierhaus, *J. Biol. Chem.* **1995,** *270,* 20473–20478.
36 J. M. Ogle, D. E. Brodersen, W. M. Clemons Jr et al., *Science* **2001,** *292,* 897–902.
37 M. Bocchetta, L. Xiong, A. S. Mankin, *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 3525–2530.
38 D. Moazed, H. F. Noller, *Nature* **1989,** *342,* 142–148.
39 D. Kim, R. Green, *Mol. Cell.* **1999,** *4,* 859–864.
40 R. R. Samaha, R. Green, H. F. Noller, *Nature* **1995,** *377,* 309–314.
41 M. Bocchetta, S. Gribaldo, A. Sanangelantoni et al., *J. Mol. Evol.* **2000,** *50,* 366–380.
42 D. Moazed, H. F. Noller, *Cell* **1989,** *57,* 585–597.
43 F. Corpet, *Nucl. Acids Res.* **1988,** *16,* 10881–10890.

7
Initiation of Protein Synthesis

7.1
Initiation of Protein Synthesis in Eubacteria

Daniel N. Wilson

7.1.1
Overview of Initiation in Eubacteria

The initiation phase of protein synthesis is one of the rate-limiting steps of translation and as such is also the principal target of translation regulation (see Chap. 11). There are significant differences between translation-initiation events in eukaryotes (see Chap. 7.2), archea and eubacteria; however, the final state of the ribosome following initiation is principally the same, namely, a ribosome programed with an initiator tRNA and mRNA, such that the start codon and tRNA are both positioned at the P-site. Indeed, the production of functionally active proteins necessitates that translation initiates at the start codon within the mRNA. As well as the use of the correct codon as the start codon, the placement at the P-site of the ribosome must also be precise; since codons are composed of three bases, incorrect placement by one or two bases will result in a complete loss of the correct reading frame. There are two major contributors to ensure the fidelity of this process: (i) the mRNA itself and (ii) a subset of translation factors termed the initiation factors (IFs).

In eubacteria, the majority of mRNAs contain, upstream of the initiation codon in an untranslated region (UTR), a purine-rich sequence called the Shine–Dalgarno (SD) sequence [1, 2], which has sequence complementary to the 3′-end of the 16S rRNA (termed the anti-SD sequence). Base-pairing between these complementary sequences has been conclusively demonstrated using the specialized ribosome system, where expression of mRNAs was shown to be abolished by mutations within the SD sequence and then restored by compensatory mutations (that restore the complementarity with the SD sequence of the mRNA) within an exogenously expressed copy of the 16S rRNA gene ([3]; reviewed in Ref. [4]). This complementarity is thought to enhance the translation of the downstream mRNA, by helping to position the AUG start codon in P-site. Recently, the SD–anti-SD complex was visualized directly in the 5.5 Å structure of *Thermus thermophilus* 70S bound with three tRNAs and SD containing mRNA [5]. This study showed that the SD–anti-SD complex formed a helical structure located above the platform and behind the head of

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

the 30S structure, whereas the AUG codon of the mRNA was in a single-stranded conformation, positioned at the P-site and exposed for interaction with the anti-codon of the P-site tRNA, i.e., if the spacing is too long then the AUG codon will not be positionable in the P-site. Certainly, when the spacing between the SD and the coding becomes to small, the strong interaction between the SD and anti-SD sequence can lead to destabilization of tRNA binding. The RF2 recoding site seems to have taken advantage of this effect since this recoding site contains a SD-like sequence that requires a spacing of two nucleotides and increasing or decreasing the spacing by even a single nucleotide dramatically reduced the frameshifting efficiency [6, 7]. The short spacing and sequence complementarity of the SD–anti-SD interaction encroaches directly on the E-site and causes loss of the tRNA from this site, which in turn destabilizes the translating ribosome and induces +1 frameshifting [8]; see also Chap. 8.2.5). However, it should be noted that there are a special subset of mRNAs, particularly predominant in Gram-positive bacteria and archaea, termed leaderless mRNAs, because the start codon is preceded by only a few nucleotides or simply starts with the 5′-terminal AUG codon (reviewed in Ref. [9]). Translation initation of leaderless mRNAs can follow a different pathway (Fig. 7.1-1B), than that described for canonical mRNAs (Fig. 7.1-1A; Sect. 7.1.2 for more details).

Unlike the multitude of initiation factors present in archea and eukaryotes, only three initiation factors, IF1, IF2 and IF3, are present in eubacteria. IF3 has been proposed as the first initiation factor to associate with the ribosome since this factor has been shown to be involved with dissociation of bacterial 70S ribosomes into their component 30S and 50S subunits (Fig. 7.1-1a). The presence of IF3 may have a role in positioning of the mRNA in conjunction with the SD sequence to move the 30S–mRNA complex from a standby state to one where the mRNA is positioned such that the start AUG codon is at the P-site (as seen in Fig. 7.1-1b). The binding of the initiator fMet-tRNA^{fmet} can occur non-enzymatically by direct binding to the programed 30S subunit or enzymatically in the form of a ternary complex with IF2 and GTP (Fig. 7.1-1c). This second pathway is stimulated by the presence of IF1, although the exact order of binding of IF1 and IF2 is unclear (Fig. 7.1-1c). The presence of all three IFs, the initiator tRNA and mRNA positioned with AUG at the P-site of the 30S subunit is termed the 30S (or pre)-initiation complex. The association of this complex with the 50S subunit results in the release of the initiation factors (Fig. 7.1-1d), presumably release of IF3 is immediate, since the anti-association action of IF3 would prohibit 70S formation. IF1 has also been proposed to be released concomitantly with subunit association (Fig. 7.1-1d). The 50S subunit acts as the GTPase-activator protein (GAP) for IF2, thus stimulating the GTPase activity of IF2, ultimately leading to the release of IF2 from the ribosome (Fig. 7.1-1e). Only after release of IF2 can full accommodation of the initiator-tRNA into the P-site on the 50S subunit occur, resulting in the P_i state. Furthermore, since IF2 binds within the A-site region, overlapping the binding sites of both EF-Tu and the A-site tRNA, release of IF2 is a prerequisite for the binding of the next aminoacyl-tRNA to the A-site, i.e., the first step into the elongation pathway (see Chap. 8).

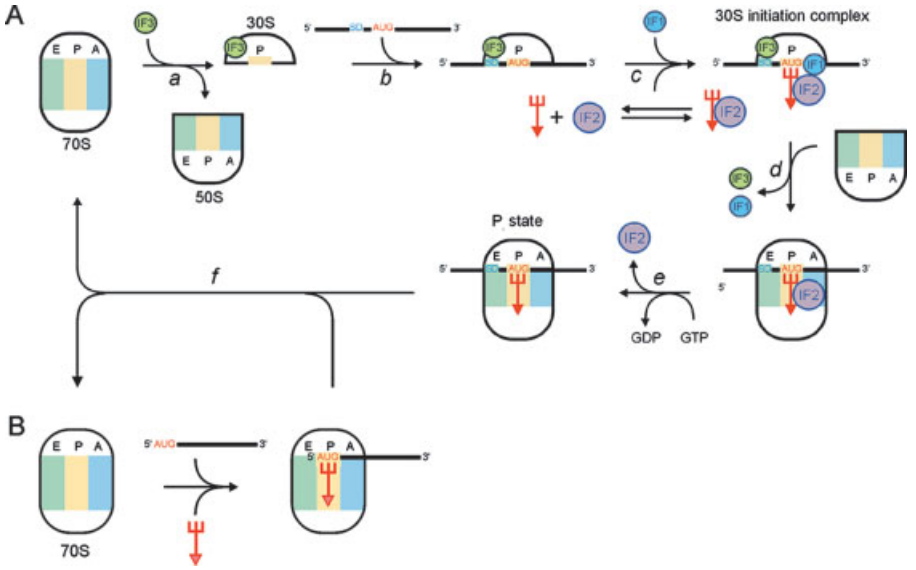


Figure 7.1-1 A schematic representation of the initiation of protein synthesis. A. Canonical mRNAs; (a) Binding of IF3 to the 30S subunit dissociates empty 70S ribosomes into the component 30S and 50S subunits. (b) IF3 aids in positioning of the mRNA such that the AUG start codon is located at the P-site of the 30S subunit. The Shine–Dalgarno (SD) sequence of the mRNA is located in the vicinity of the binding position of IF3, where it makes interaction with the anti-SD sequence of the 16S rRNA in the 30S subunit. (c) Binding of the initiator fMet-tRNA (red) can occur directly or in the form of a ternary complex with IF2 (purple) and GTP and is stimulated by the presence of IF1 (blue) and results in the formation of the 30S initiation complex. (d) Association of the 50S subunit with 30S initiation complex causes the release of IF3 and IF1, but IF2 remains bound at the A-site. (e) The GTPase activity of IF2 is stimulated by the presence of the 50S subunit and ultimately leads to the release of IF2.GDP from the ribosomes, allowing full accommodation of the initiator tRNA at the P-site on the 50S subunit. This complex is termed the Pi state, i.e., P-site is occupied and the A- and E-sites are free. (f) Following translation elongation, termination and ribosome recycling, the empty 70S ribosomes are ready to re-enter the translation-initiation phase. B. Leaderless mRNAs. Binding of leaderless mRNAs may utilize an alternative pathway without requiring initiation factors and using 70S monosomes directly.

7.1.2 Specialized initiation events: translational coupling, 70S initiation and leaderless mRNAs

In the bacteria, most mRNAs are transcribed from transcriptional units that usually contain several, often functionally related, genes. The product is a polycistronic mRNA, where each cistron carries the information of a single protein. In *E. coli*, polycistronic mRNAs usually contain four cistrons. Such mRNAs contain multiple translation initiation sites, one for each cistron, with a Shine-Dalgarno (SD) sequence and an AUG initiation codon. Recognition of the translation start sites within the mRNAs is performed by an initiation complex comprising the small ribosomal subunit (30S), the initiator transfer RNA carrying the amino acid formylmethionine (fMet-tRNA$_f^{Met}$) and three proteins called initiation factors (IF1, IF2, and IF3; Fig. 7.1-1). In principle, the various initiation codons of a bacterial polycistronic mRNA can be recognized independent of one another. Aided by the SD sequences, the 30S initiation complexes can land on any of the available translation initiation sites (30S *de novo* initiation). It has been observed that to be accessible for the 30S subunit, an initiation region (including the start codon and the SD motif) must be in a single-stranded, non-hydrogen-bonded state, i.e. not buried within a secondary structure. This is the rule, but an important exception is seen for the polycistronic mRNAs encoding ribosomal proteins. Here the first cistron is usually accessible for the 30S *de novo* initiation, whereas the second and following initiation sites are sequestered within secondary structure. However, once the first initiation site has been recognized, translation commences and the translating ribosome can unfold the secondary structure to reveal the second initiation site. In this way, the second and all downstream cistrons are translationally coupled, i.e. if one cistron is translated, all the downstream ones are translated. On the other hand, if the first cistron is not translated, then the whole polycistronic mRNA cannot be translated. This phenomenon is termed *translational coupling*.

Translational coupling is exploited for what is called *autogenous translational regulation* (see Chap. 11 for details). Briefly, a repressor protein (usually a translation product of the second or third cistron of the same polycistronic mRNA) will bind to the first initiation site on the mRNA, thereby inhibiting the translation of the first cistron, as well as translation of the downstream cistrons. Repression is relieved and, therefore, translation resumes, only when the regulatory ribosomal protein dissociates from the low-affinity binding site on the mRNA and is recruited by the high-affinity binding site on the rRNA during assembly of ribosomes.

Often, the downstream cistron is translated by re-initiation, meaning that the ribosome that terminates translation of the upstream cistron does not dissociate from the mRNA but proceeds directly to the next cistron, occasionally shifting the reading frame if this is required (70S-type initiation; [104, 105]). This is seen not only for ribosomal proteins but also for translation factors, for example, the *prfB* gene, encoding the translation termination factor RF2, where the final *UGA* stop codon overlaps with the start AUG codon (A*UGA*) of the *hemK* gene, which encodes a methylase

that modifies the termination factor (see Chap. 9 for more details). Therefore, it is easy to envisage that a 70S ribosome, after undergoing termination and peptide release at the UGA stop codon of the first cistron (for example, of the RF2-mRNA), does not dissociate from the polycistronic mRNA, but instead translates the downstream cistron (in this case the hemK mRNA). An empty 70S (following termination and peptide release) is capable of scanning up- and downstream along the mRNA, until it is "caught" by a nearby SD sequence (which occurs through base-pairing with the 3′-end of 16S rRNA). This promotes the correct position of any following AUG-start codons at the P-site. Whether or not the "scanning" 70S ribosomes actually carry an fMet-tRNA is not clear. The 70S type of initiation is the only reason for the formylation of the initiator Met-tRNA, since a 30S subunit can easily form an initiation complex with both Met-tRNA$_f^{Met}$ and fMet-tRNA$_f^{Met}$, whereas the presence of fMet-tRNA$_f^{Met}$ facilitates the formation of the 70S initiation complex [105].

Translation of leaderless mRNAs has been proposed to occur on 70S ribosomes [106], as well as being able to proceed through the 30S pre-initiation pathway (cf. Fig. 1A and B), and there is growing evidence to support this view (reviewed in Ref. [9]). Recently, Ueda and coworkers demonstrated, using an *in vitro* translation system comprising only purified components, that translation of leaderless mRNAs could occur in the absence of initiation factors [107]. Furthermore, the stability of leaderless mRNAs with 70S ribosomes in the presence of initiator-tRNA has been shown to be up to 10-fold higher than with 30S subunits [107, 108]. Since the increased stability of binding of canonical mRNAs with the 30S subunit probably derives from elements within the 5′ untranslated region (UTR), such as the SD sequence, which interact with the 16S rRNA, these sorts of interactions are unavailable to leaderless mRNAs (the 5′ UTR being absent). By forming initiation complexes directly with 70S ribosomes (Fig. 7-1B), rather than through the 30S pre-initiation complex pathway (Fig. 7-1A), the stability of the mRNA–tRNA complex is increased because 85% of the contacts of a P-tRNA are with the 50S subunit in the 70S ribosome [109, 110].

Although downstream stabilization elements have been proposed to exist in leaderless mRNAs, these could not be confirmed. Thus, it seems that the 5′ AUG codon is the major, if not the only, element within the leaderless mRNA required for their efficient translation. Indeed, mutations at this position have been shown to reduce significantly the efficiency of translation, even when the AUG is replaced by other canonical initiation codons, such as CUG, GUG or UUG (see Ref. [9]). In addition, for 30S initiation of leaderless mRNAs, the concentration of the initiation factors is an important factor, such that high concentrations of IF2 stimulate translation, whereas IF3 has an inhibitory effect. Thus, the ratio of IF2 and IF3 seems to influence significantly the expression level of leaderless mRNAs [111].

The additional stability of 70S initiation complexes may explain why translation of leaderless mRNAs, but not canonical mRNAs, continues in the presence of the antibiotic kasugamycin (see Ref. [112]). Kasugamycin has also been shown to affect assembly of the 30S subunit, producing a particle that is deficient in a number of small ribosomal proteins. While this particle cannot translate canonical mRNAs, translation of leaderless mRNAs remains unaffected (U. Blaesi, pers. Comm). Of the

proteins missing, S1 has been shown to be dispensable for translation of leaderless mRNAs [113]. The S1 protein has two N-terminal RNA-binding motifs necessary and sufficient for ribosome binding and four C-terminal RNA-binding motifs associated with mRNA binding. Although S1 is absent in the crystal structures of the 30S subunits [32, 92], the binding position has been located, using cryo-EM, to the platform side of the subunit, in close proximity to the anti-SD sequence [114]. It is therefore easy to envisage that the absence of a 5′ UTR in leaderless mRNAs circumvents the necessity of S1.

As yet, the role of leaderless mRNAs is not clear; certainly there is little correlation (or homology) between the genes encoding leaderless mRNAs in different organisms, let alone across the kingdoms [9]. Despite this, under certain physiological conditions, for example, low temperature or in the presence of antibiotics, the 70S initiation pathway open to leaderless mRNAs might be competitively favorable over that of canonical mRNAs [9, 106]. In this respect, it is interesting to note that in many *Streptomyces* species, a number of antibiotic resistance genes are leaderless mRNAs [9]. Blasi and coworkers [9] have also suggested that leaderless mRNAs may represent remnants of ancestral mRNAs that have acquired canonical start codons at the 5′-end, i.e. the earliest mRNA templates were simply single-stranded polynucleotides.

7.1.3
Initiation Factor 1 Binds to the Ribosomal A-site

The exact role of IF1 within the initiation complex is perhaps the least understood of the IFs. A number of roles have been described for IF1 including (i) subunit association during 70S initiation complex formation, (ii) modulating the binding and release of IF2 and (iii) blocking the binding of tRNAs to the A-site (reviewed in Refs. [10, 11]). Irrespective the role of IF1, the gene encoding IF1, *infA*, is essential for cell viability in *Escherichia coli* [12] indicating its importance in the initiation process.

The structure for IF1 has been determined by NMR spectroscopy and revealed to contain a secondary structure characteristic of the oligomer binding (OB) fold family of proteins, termed because of their ability to bind oligonucleotides and oligosaccharides [13]. The architecture of a classic OB-fold motif includes a five-stranded β-sheet coiled to form a closed β-barrel, and capped by a α-helix as exemplified by IF1 (Fig. 7.1-2A). A number of other translational proteins are members of this family including ribosomal proteins S1, S17, and L2 (reviewed in Refs. [14, 15]), tRNA synthetases, IF5A and eIF2a, as well as the central region of eIF1A (Fig. 7.1-2A; [16]). The presence of domains additional to the common OB component in higher organisms, namely archeal and eukaryotic eIF1A, at least partly correlates with the ability to form binary complexes with eIF5B (see following).

Interestingly, a number of the cold shock protein (Csp) family have high structural homology to IF1 (although little sequence homology). In fact, some strains, predominantly Gram-positive, for example *Bacillus stearothermophilus*, have no obvious IF1 homolog [17]. It has been postulated based on the structural similarity between the Csp and IF1 families that in these strains one of the often many Csps may have assumed this role. In this regard, it is interesting that a double deletion CspB–CspC

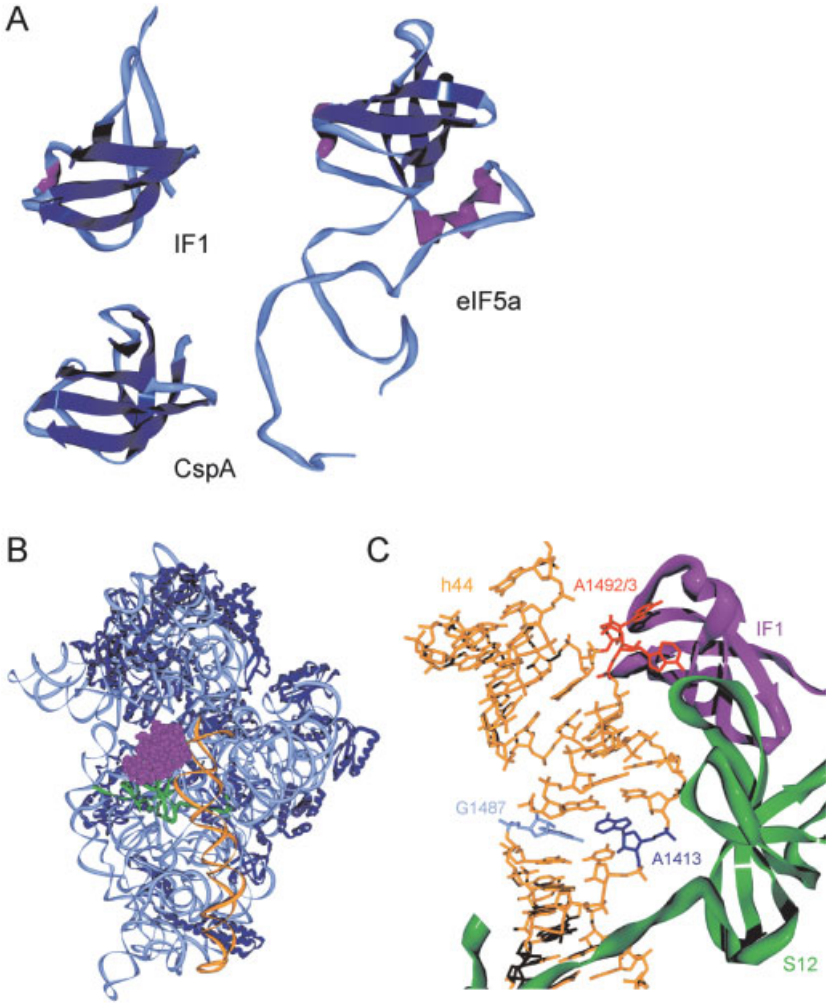


Figure 7.1-2 The binding site of IF1 on the 30S subunit and homology with other factors. (A) The solution structures of IF1 (pdb1ah9; [99]), CspA (pdb1mjc; [100]) and eIF1a (pdb 1d7q; [16]), all shown in ribbon representation with strands (dark blue), helices (purple) and random coil (light blue). (B) Overview of the IF1 binding site on the 30S subunit (pdb1hr0; [19]). Ribbon representation of the 16S rRNA (pale blue) including ribosomal proteins (dark blue) with h44 (yellow) and ribosomal protein S12 (green) highlighted. IF1 (purple) is shown as spacefill representation. (C) Close-up view showing that IF1 (purple ribbons) binding, causes A1492 and A1493 (red) to be flipped out of helix 44 (yellow) and the base-pair between A1413 (dark blue) and G1487 (light blue) to be broken.

in *Bacillus subtilis* led to alterations in protein synthesis, cell lysis upon entry into stationary phase, and the inability to sporulate [18]. Deletion of all three Csp proteins was lethal suggesting the importance of having at least one of this family present. Intriguingly, the defects caused by the double knock-out could be cured by the heterologous overexpression of *E. coli* IF1, suggesting that IF1 could assume some of the chaperone activities normally performed by the Csps. This raises the question if under some conditions the reverse could be true, especially for strains lacking the *infA* gene. Certainly, there is some evidence that members of the Csp family co-purify with ribosomes; however, this may be related simply to their chaperone activity and reflect their tendency to interact with RNA rather than their involvement in the initiation of protein synthesis.

Despite the low-sequence similarity between OB-fold family members, secondary-structure similarity is striking, as well as the localization of basic residues on one face of the OB-fold. The recent crystal structure of *T. thermophilus* IF1 bound to the 30S subunit [19] demonstrates that IF1 is no exception. Interactions with the 30S subunit associate with the highly basic surface of IF1, where conserved arginine residues (Arg46 and Arg64) stabilize RNA-binding interactions through stacking and electrostatic interactions. The IF1-binding site on the 30S subunit consists of a cleft formed by h44, the 530 loop and protein S12 (Fig. 7.1-2B; [19]). The loop between strands β3 and β4 is inserted into the minor groove of h44 and flips out residues A1492 and A1493 from their stacked position in h44 (Fig. 7.1-2C). This is reminiscent of the situation where these residues are flipped out due to binding of the antibiotic paromomycin to the decoding site (see Chap. 12) and also due to a cognate tRNA at the A-site (see Chap. 8.2). The major distinction being that during decoding, A1492 and A1493 are critically involved in direct monitoring of correct Watson–Crick pairing of the first two positions of the anticodon–codon duplex [20], whereas within the IF1:30S structure these residues are inaccessible, being protected by IF1 and S12. This suggests that although the binding sites of the A-site tRNA and IF1 overlap, there is little mimicry in their interaction.

Despite the expectation that IF1 would sterically occlude tRNA binding at the A-site [21], it is unlikely that this is the role of IF1 during initiation as there is only one tRNA-binding site on the 30S subunit, that of the prospective P-site [22, 23]; reviewed in Ref. [24]. It seems more probable that IF1 binding at the A-site induces conformational changes that promote subunit association during 70S initiation complex formation. Indeed, the flipping out of A1493 disrupts a base-pair with A1408, destabilizing the top of h44 and allowing lateral movement of bases C1412 and A1413 such that the base-pair between A1413 and G1487 is broken (see Fig. 7.1-2C). This lateral shift moves one strand of h44 with respect to the complementary strand generating "long distance" (up to 70 Å from the IF1-binding site) conformational changes within h44 [19]. The minor groove of h44 makes extensive contacts with the 50S subunit, one per helical turn, forming intersubunit bridges B3, B5 and the largest contact point between subunits, bridge B2a [25]. Thus, it is possible that these changes induced by IF1 binding may be responsible for the observed increase in association rates between 30S and 50S subunits [26]. The activation energy associated

with 70S formation is large, estimated at 80 kJ mol^{-1}, and is involved only in adaptation of the 30S subunit (not 50S subunit), rather than the association step itself [27]. Therefore it is tempting to speculate that the initiation factors, particularly IF1 because of the changes it induces in the 30S subunit, help to overcome the free-energy barrier for 50S subunit association with the 30S [28]. The functionally active 30S conformations are obtained by heat activation [29] and have been visualized by cryo-EM, which revealed that they bear a closer resemblance to the 50S-subunit-bound state than to the inactivated state [30]. This heat-activated "intermediate" state may reflect a physiological state [29], such as that induced *in vivo* by translational factors such as IF1. Indeed, the crystal structure of the IF1-bound 30S subunit [19] also exhibits more similarity to the 50S bound state [25, 31], than to that of the free 30S subunit [32].

Mutation of A1408G eliminates all indicators associated with IF1 binding to the 30S subunit, such as the "tell tale" footprints at A1492 and A1493, yet retains wild-type growth characteristic [33]. This is perplexing as IF1 interaction with the 30S subunit is essential for competent 70S formation [34] and cell survival [12]. The A1408G mutation would also be expected to disrupt the base pair with A1493, tempting speculation that by doing so it enables the 30S subunit to adopt a conformation mimicking that of the initiation complex, thus making IF1 dispensable for cell viability [33]. If this hypothesis would be correct, then direct interaction of IF1 and IF2 may not be necessary and that the 30S conformational change induced by IF1 is sufficient to stimulate IF2 binding.

7.1.4
The Domain Structure of Bacterial IF2

IF2 is the largest of all eubacterial translation factors and can be divided into three major domains based on primary sequence homology (Fig. 7.1-3A), an N-terminal domain (NTD) that is not conserved in sequence or length among bacteria, a central domain containing the guanine-nucleotide-binding motif (termed the G domain), and a C-terminal domain (CTD), which contains the entire fMet-$tRNA^{fmet}$-binding site (reviewed in Ref. [11]). *E. coli* IF2 has been divided further into subdomains by Sperling-Petersen and co-workers [35], such that the NTD consists of subdomains I–III, the G domain encompasses IV–VI-1 and the CTD, VI-2. In *E. coli*, the IF2 gene, *infB*, encodes three isoforms of IF2, termed IF2-1, -2 and -3 [36]. The latter two isoforms are smaller and result from translation at alternative initiation sites near the beginning of domain II (as indicated by arrows in Fig. 7.1-2A). The cellular level of all three isoforms is similar and the presence of all three isoforms has been shown in *E. coli* to be optimal for growth. However, the absence of multiple isoforms in the most of the bacteria suggests that they are not essential for survival; indeed many extremophilic species in bacteria, such as *Thermus,* or in Archea, such as *Sulfolobus* or *Methanococcus,* do not even have this NTD region [37]. A fragment consisting of subdomains I and II (but not subdomain I alone) was capable of binding the 30S subunit and IF2 lacking this region showed low binding affinity, suggesting the importance of this region for factor binding [38, 39]. Interestingly, this

same fragment has been shown to bind to the *infB* mRNA, hinting at the existence of an autoregulatory mechanism for IF2 [40]. NMR studies of the NTD subdomain I have revealed that residues 2–50 form a compact structure containing three short α-helices and three antiparallel twisted β-strands (Fig. 7.1-3A), the following residues 51–97 were unstructured and the rest of subdomain I (98–157) was of a highly helical nature [41]. The latter was suggested to act like a linker, much like that found between domains VI-1 and VI-2 of aIF5B (see the following). The compact core of subdomain I (IF2-DI) has structural similarity to the SC-fold domain of class Ia aminoacyl-tRNA synthetases (RS), such as that found in GlnRS (Fig. 7.1-3B). In the crystal structure of Gln-tRNA bound to GlnRS, the SC-fold domain contacts the inner side of the L-shaped tRNA, where it provides a connection between the RS domains that interact with the acceptor (green) and anticodon (red) of the tRNA (yellow in Fig. 7.1-3B). This suggests that the NTD of IF2 is probably associated with positioning of the anticodon stem loop of the fMet-tRNA into the P-site of the 30S subunit. Consistent with such a suggestion is the crosslink found between subdomain II of IF2 and the anticodon stem of fMet-tRNA as well as the similarity in the footprinting pattern found within this region in the presence of IF2 or MetRS.

The G domain and CTD of IF2 (IV–VI) are the most highly conserved regions, having homology across all kingdoms. In fact, IF2 from *Mycoplasma genitalium* and

Figure 7.1-3 The domain structure of *E. coli* initiation factor IF2. (A) Schematic diagram of the domain structure of *E. coli* IF2. There are two alternative initiation sites (arrowed) within subdomain II marked with IF2-2 and IF2-3. The structure of part of subdomain I of the N-terminal domain (NTD) of *E. coli* IF2 has been determined (pdb1euq; [101]) and the central (G domain) and CTD encompassing subdomains IV–VI have almost 50% similarity to the archeal IF2 homologue aIF5B, whose structure (pdb 1g7t; [102]) is also shown. (B) Part of the NTD of IF2 (residues 2–50; subdomain I) has structural homology with the SC-fold domain of Gln (and Met) aminoacyl-tRNA synthetases (GlnRS-SC). Within the complex GlnRS structure, this region contacts the anticodon stem of the tRNA (pdb1euq; [101]). The tRNA is colored yellow with the anticodon (red) and CCA end (green) highlighted for reference. The homologous region to domain I of IF2 in the GlnRS is colored purple. (C) Domain VI-2 of IF2 has structural homology with domain III of EF-Tu, leading to the proposal that IF2 recognition of the fMet of an initiator tRNA utilizes the equivalent surface. Shown here is domain VI-2 of IF2 (IF2-DVI-2; pdb 1d1n; [43]) compared with domain III of EF-Tu.Cys-tRNA^{Cys} (pdb1b23; [45]). In the latter structure, the Glu271 (red; stacks with A76) and His273 (green)/ Arg274 (purple) are in close proximity to the terminal adenine (A76) of the CCA end of tRNA (yellow) and the attached cysteine residue respectively. The equivalent positions are not conserved in IF2 suggesting that the details of recognition differ, however, residues on the equivalent surface as that used by EF-Tu to recognize tRNA predicted to participate in fMet recognition are indicated in red [43].
(D) Topology of IF2 (middle) on the 30S (left) and 50S subunit (right). Positions of IF2 used for site-directed hydroxyl-radical probing are shown on the structure of the aIF5B at the equivalent locations: cleavages from the pale and dark blue positions on IF2 map to the 30S subunit and include positions G35/G38-C40 and A397 of h3/h4 (brown), G423 (green; h16) and residues in h17/h18 (C443 and A498/A537, purple/red). Residues A1418 and A1483 (cyan) exhibited reactivity upon IF2 binding. On the 50S subunit, the L11-binding region (brown), the sarcin-ricin loop (purple) and H89 (green) have been also footprinted. The small and large subunits (shown as if the 70S ribosomes were opened like a book) are shown in ribbon format with rRNA and ribosomal proteins colored in pale and dark blue, respectively. Helix 44 (yellow) on the 30S subunit and the A- (red) and P-site (yellow) CCA-end substrates are highlighted for reference positioning.

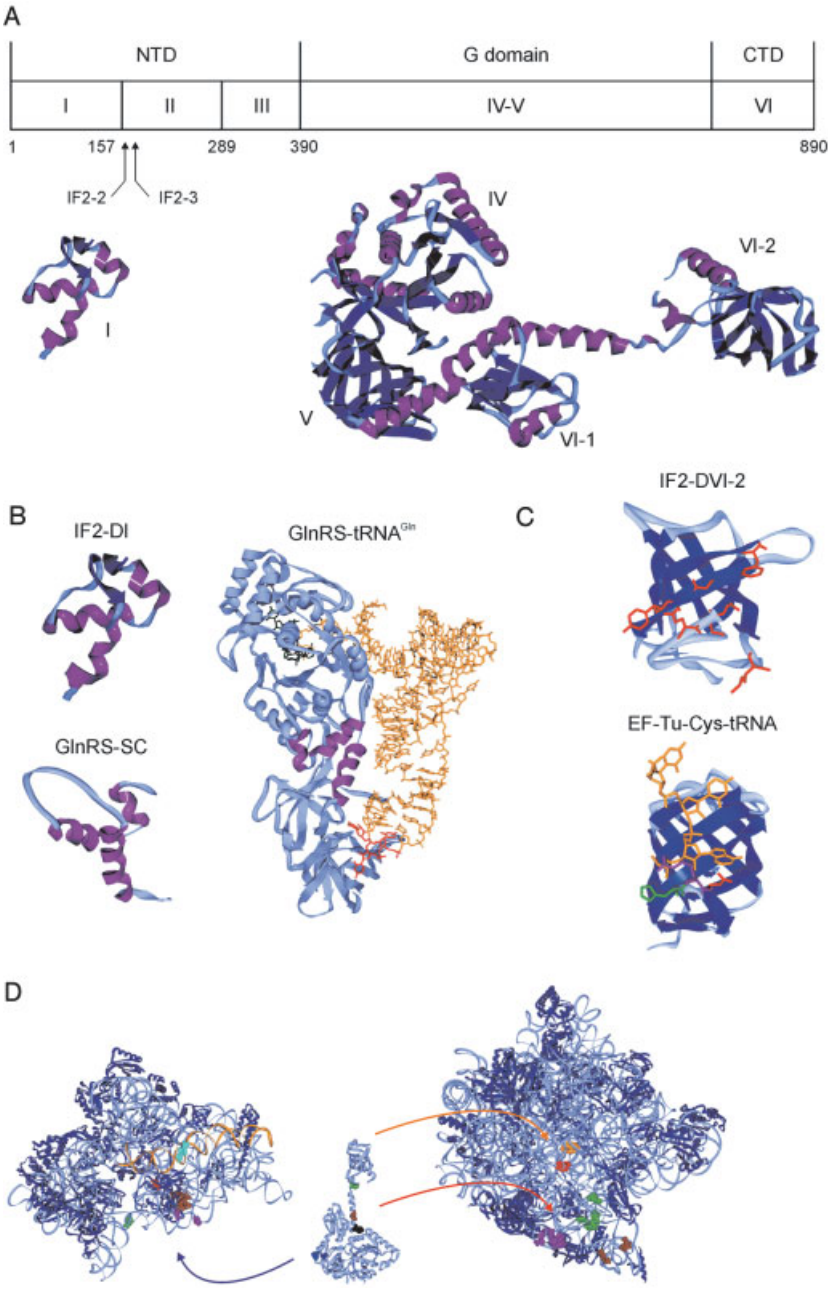

A
NTD
G domain
CTD
I
II
III
IV-V
VI
1
157
289
390
890
IF2-2
IF2-3
I
IV
VI-2
V
VI-1
B
IF2-DI
GlnRS-tRNA^Gln
GlnRS-SC
C
IF2-DVI-2
EF-Tu-Cys-tRNA
D

Figure 7.1-3

T. thermophilus are unusually small, consisting of only ~600 amino acids (rather than the usual 800–1000), the missing residues being absent from the NTD. Similarly, the IF2 homologs found in eukaryotes and archaea, termed eIF5B and aIF5B, respectively, are also smaller than their eubacterial counterparts and by comparison with their eubacterial counterparts contain only the G and CTD domains. The crystal structure of aIF5B from *Methanobacterium thermoautothrophicum* has been solved therefore providing a good homology model for the C-terminal region of IF2 (Fig. 7.1-3A). The crystal structure is described as being "chalice-shaped", where the "cup" region containing the G domain is identical to the corresponding regions in EF-Tu•GTP and EF-G•GDP. The stem of the chalice constitutes an α-helical linker region separating the G domain by over 40 Å from the base of the chalice (VI-2 in Fig. 7.1-3A). The structure of aIF5B was solved in the GTP, GDP and nucleotide-free state; however, surprisingly there was little significant difference between them and no observable change within domain IV, the formyl-methionine (fMet)-binding domain, relative to the functional state. This is consistent at least with the stable binding of fMet-tRNA^{fMet} observed with bacterial IF2 regardless of whether in the GTP or GDP form [42].

Recently, domain VI-2 of the CTD from *B. stearothermophilus* IF2 containing all the molecular determinants necessary and sufficient for fMet-tRNA^{fMet} recognition and binding [42] was determined by NMR [43] (Fig. 7.1-3C). The β-barrel structure of this domain shows remarkable similarity to domain II of EF-G and EF-Tu (Fig. 7.1-3C), despite having low-sequence identity (13%) and homology (17%). This high structural homology to EF-Tu and the availability of structures for two different aminoacyl–tRNA•EF–Tu complexes [44, 45] enabled a model to be proposed to define an IF2 recognition site for the fMet moiety and the acceptor stem (CCA-end) of the initiator tRNA [46, 43]. Although the mechanisms probably differ in their details since there is little comparative conservation in residues at equivalent positions between EF-Tu and IF2, it seems probable, however, that the same surfaces are used and a number of residues conserved within the IF2 family, such as R654, Q655, F657, G667 and E713, constitute the fMet-CCA-binding site (colored red in Fig. 7.1-3C).

7.1.5
Interaction Partners of IF2

On the ribosome, IF2 accelerates codon–anticodon base-pairing between the initiator fMet-tRNA^{fMet} and the start codon of the mRNA in the ribosomal P-site [47]. The specificity of the reaction is governed by the exclusive recognition of the fMet moiety of the initiator fMet-tRNA^{fMet} [48, 49]. In bacteria, IF2 can form a ternary complex with fMet-tRNA and GTP [50], whereas no evidence for the equivalent interaction between eIF5B and Met-tRNA has been found, although eukaryotes have an additional factor eIF2, which assumes this delivery role (see Sect. 7.2.7.4). Since the majority of the interactions of tRNA with programed 30S subunits depend on codon–anticodon interaction [51], the stimulation of this reaction by IF2 is probably due to the corresponding increase in stability of the initiator tRNA on the ribosome. In the presence of IF1, binding of the ternary complex is additionally stimulated,

suggesting some interplay between the factors. However, unlike the situation in eukaryotes, where eIF1A and eIF5B form a stable interaction in the absence of the ribosome [52], no evidence for such an interaction has been observed between bacterial IF1 and IF2. Consistent with this, the binding interface between eIF1A and eIF5B was recently determined to use the C-terminal region of each factor, i.e., regions that are not present in bacterial IF1 and IF2 [53]. Therefore, if an interaction between the bacterial counterparts exists it seems to occur only on the ribosome: IF2 has been crosslinked to IF1 on the ribosome [54] and subdomain II of IF2 has been proposed to interact with IF1 [38].

Manual sequence alignment of IF2 against EF-G suggested that IF2 could be aligned against all of EF-G, with the exception of the very N-terminal domain and the terminal region of domain IV of EF-G, i.e., the region mimicking the tRNA anticodon stem loop. Interestingly, this latter region was found to have similarity to IF1, leading to the suggestion that IF2 and IF1 may together mimic the structure of EF-G, thus extending the structural mimicry found between elongation factors EF-G and the ternary complex EF-Tu•tRNA•GTP (reviewed in [55]; see Chap. 8.2.5) to encompass the initiation factors [21]. However, the subsequent structure of IF1 bound to the 30S subunit does not support a direct structural mimicry of either domain IV of EF-G or the anticodon stem loop of a tRNA [19]. This aside, it does seem probable that the general topographies of the factors is consistent with this idea, since IF1 does bind in the A-site and interact with the bases A1492 and A1493 involved intimately in the decoding process. Furthermore, the strong homology between G domains of IF2/eIF5B and EF-G/EF-Tu and the ribosome-dependent activation of their GTPase activities suggests that IF2, in particular the G domain, is likely to occupy a similar position to EF–G, EF-Tu ternary complex and EF-1α at the A-site of the ribosome as visualized by a multitude of cryo-EM studies [56–62].

Early attempts to map the position of IF2 on the ribosome using chemical modification approaches have been relatively unsuccessful [63, 64]. The former study identified a large number of residues of the 16S rRNA spread throughout the 30S subunit [64], suggesting a weak or disperse association of IF2 with the 30S subunit or perhaps predominantly with ribosomal proteins. Recent base-specific probing studies also detected no protections of the 16S rRNA resulting from the binding of IF2 to 30S subunits; however, in loose couple ribosomes, binding of IF2 led to a decrease in reactivity of residues A1418 and A1483. Since these residues are located in the lower portion of h44 that makes contacts with the 50S subunit, the changes in reactivity are probably indirect and might indicate that IF2 has a "tightening effect" on the interaction between the subunits [65]. Site-specific hydroxyl-radical probing experiments suggested that the G domain (V) is in close proximity to the 16S rRNA, since cleavages of the 16S rRNA were observed from two positions in this region. The cleavages were localized to residues in h3/h4 (positions G35/G38-C40, A397), h16 (G423) and h17/h18 (C443, A498/A539), suggesting that the G domain of IF2 is in a similar position as that of EF-G. Consistently, cleavage of some 16S rRNA residues was observed when equivalent positions in EF-G to those in the IF2 study were used [66]. This raises the question as to why no protections

are observed when IF2 binds to the 30S subunit and invites the speculation that IF2 may undergo conformational changes upon 50S subunit association, such that stronger contacts are then made with the rRNA component of the 30S subunit in the 70S ribosome.

Two distinct regions of the 23S rRNA become protected from chemical probing upon binding of IF2 to 70S ribosomes; these include positions in the sarcin-ricin loop (G2655, G2661 and A2665 as well as an enhancement in the reactivity of A2660) and the loop at the end of H89, specifically positions A2476 and A2478 [65]. Residues in H89 (A2482 and U2474) were also cleaved using site-directed hydroxyl-radical probing when tethers were placed at positions within the CTD (VI-1) [67]. In addition, these tethers produced cleavages of nucleotides within the L11-binding region (G1068 and weakly at C1076 in H43). This set of protection and cleavage patterns for IF2 on the 50S subunit are similar to those determined for EF-G and EF-Tu [68], but not identical. This is certainly consistent with the observation that IF2 and EF-G compete for overlapping binding positions on the 70S ribosome [69]. Interestingly, in this study, the antibiotic micrococcin, which interacts with the L11 binding region, was shown not only to inhibit EF-G-dependent GTPase, but also to stimulate the IF2-dependent GTPase activity. These observations suggest that, although both factors interact with this region, they probably do so in a distinct manner. One of the largest differences is the implication of H89 in IF2 binding, since this region is not considered part of the binding site for the other elongation factors. Indeed the antibiotic evernimicin, which footprints within H89, has been proposed to act as an IF2-dependent translation-initiation inhibitor ([70]; see Chap. 12).

7.1.6
The Role of the IF2-dependent GTPase Activity

Early experiments suggested that the GTP form of IF2 was required for 70S ribosome formation by association of the component subunits and that hydrolysis of GTP released IF2 from the ribosome allowing translation to enter the elongation cycle by the binding of the ternary complex to the A-site [71, 72]. The situation was found to be the same in eukaryotes, where the GTP form of eIF5B was essential for subunit association and that hydrolysis of GTP releases eIF5B allowing peptide-bond formation to occur [73, 74]. This harmony was challenged when it was reported that the association of bacterial pre-initiation complexes with 50S subunits to form a post-initiation complex capable of peptide-bond formation required the same length of the time regardless of whether GTP or GDP was used [75]. The interpretation from these experiments was that the GDP form of IF2 catalyzes subunit association as efficiently as the GTP form and that GTP hydrolysis does not stimulate (i) the adjustment of fMet-tRNA in the P-site, (ii) the ejection of IF2 from the ribosome, or (iii) the formation of the initiation dipeptide (see Ref. [10]). However, an elegant series of experiments from Ehrenberg and co-workers [76] conclusively demonstrated that, in fact, the GTP form of IF2 (or with the non-hydrolyzable analog GDPNP), but not the GDP form, promotes rapid association of the ribosomal subunits during initiation (the K_a in the presence of GTP was over 20 times higher than

with GDP). In this study, the binding of GTP to IF2 was the most efficient in the presence of both fMet-tRNA and 30S subunits (with mRNA, IF1, and IF3), suggesting that the GTP form of IF2 stabilizes the binding of fMet-tRNA to the 30S subunit (and *vice versa*) and this promotes association. Furthermore, formation of the first dipeptide was also fast for IF2-GTP, but not IF2-GDP or IF2-GDPNP, indicating that GTP hydrolysis is necessary for rapid release of IF2 from the ribosome and therefore for dipeptide formation [76].

These observations led Ehrenberg and co-workers to propose a two-state model for IF2 action during initiation: the free form of IF2 is the GDP form, which has a low affinity for both the pre-initiation complex and 70S ribosomes. The presence of 30S subunits with *both* fMet-tRNA and mRNA promote nucleotide exchange, and stabilizes the pre-initiation complex since IF2 is now in the GTP form. The GTP form of IF2 has a high affinity for the 50S subunit and thus rapid association between the pre-initiation complex and the 50S subunit ensues. In this manner, only fully competent 30S pre-initiation complexes are converted into 70S initiation complexes. The high affinity of the GTP form of IF2 is consistent with the slow dissociation of IF2-GDPNP from the 70S ribosome. Subsequent to 70S association, GTP hydrolysis occurs and the low-affinity IF2-GDP dissociates from the ribosome, leaving the A-site free for the binding of the ternary complex. Interestingly, ternary complex binding was demonstrated to be possible in the presence of IF2, at least temporarily [76], which is consistent with the observation that ternary complex can bind to the 70S ribosome with the same kinetics in the presence or absence of IF2-dependent GTP hydrolysis [75]. Understanding of the exact binding position of IF2 will help to address the extent to the overlap in position with the ternary complex.

7.1.7
The Mystery of the IF3-binding Site on the 30S Subunit

IF3 was originally identified as an anti-association factor because it binds with high affinity to the 30S subunit (an association constant greater than 10^7 M^{-1} [77]) and thereby prevents re-association of the 30S and 50S subunits [26]. In addition to this function, IF3 is known in conjunction with IF2 to be involved in the discrimination between aminoacyl tRNAs, thereby permitting only the presence of an initiator tRNA at the P-site [34, 78]. Discrimination is based on recognition by IF3 of the anticodon loop and three base-pairs of the anticodon stem [79] and may even involve recognition of the start codon itself [80]. IF3 is encoded by the *infC* gene located at 37.5 min on the *E. coli* chromosome [81, 82], and has been shown to be essential for cell viability [83] and for protein synthesis [84]. IF3 consists of two distinct domains separated by a 20-residue-long (~45 Å) lysine-rich linker region [85]. It has been speculated that the two different IF3 functions described above could be attributable to each of the domains [86]. However, a number of other functions for IF3 have been assigned, such as the (i) adjustment of the mRNA from a so-called standby site to the P-site, (ii) stimulation initiator tRNA binding to the P-site, and (iii) dissociation of fMet-tRNA from the start codon of leaderless initiator tRNAs (see Ref. [87] and references therein).

The complete intact structure of IF3 has not been determined, but those of both NTD (with linker region) and CTD have been solved by X-ray crystallography [88] and NMR [89, 90] (Fig. 7.1-4A). The CTD has been shown to be capable of independently inhibiting 30S and 50S association, although it requires much higher excess over ribosomes compared with the full-length factor [90, 87]. *Thermotoga maritima* IF3 bound to the *T. thermophilus* 30S subunit has been visualized by cryo-EM at 27 Å resolution [91]. Density correlating to the CTD was found to locate to a region of the 30S subunit involved in forming bridges with the 50S subunit, suggesting that IF3 prevents subunit association by physically blocking the docking sites [91] in agreement with protection studies [63]. In contrast, recent X-ray crystallographic data of the CTD of *T. thermophilus* IF3 bound to the *T. thermophilus* 30S subunit revealed that IF3 was not bound at the subunit interface but at the upper end of platform on the solvent side [92] (Fig. 7.1-4B), where it makes contacts with h23, h26 and the 3′-proximal end of h45. In this position, the anti-association activity of IF3 cannot result from physically blocking subunit association, but probably derives from indirectly moderating the mobility of h45. Support for the latter model comes from an observation that a double mutant in h45 of the 16S rRNA reduces IF3 binding to the 30S subunit [93]. Furthermore, IF3 cannot dissociate 70S ribosomes carrying this double mutation even though the affinity of IF3 for the 70S ribosome is enhanced 30-fold over the wild-type 70S ribosomes [93]. A number of other biochemical data are more consistent with the platform localization of IF3, for example, regions in h45 (1506–1529) and h26 (819–859) have been crosslinked to IF3 [94], as have ribosomal proteins S7, S11, and S18 [95]. Docking the NTD of IF3 based on both the constraints of the position of the CTD on the 30S subunit and available biochemical data places the NTD in close proximity to the P-site [92]. This led to a model where the codon–anticodon recognition operates by space restrictions such that only correct-binding orientations are permissible, rather than by a direct interaction of the P-site tRNA with the NTD of IF3.

However, recent evidence has cast doubt of on both the position of the CTD of IF3 determined from the crystallographic analysis and the involvement of the NTD in any of the IF3 functions. It has been noted that 30S subunits are arranged within the crystal lattice such that they contact each other at the region where the cryo-EM has localized the CTD, leading to the suggestion that this could have masked the physiological binding site during the soaking of the crystals with the IF3 domain [96]. Furthermore, a comprehensive set of site-directed hydroxyl-radical probing experiments suggest that IF3 binds on the interface side of the 30S subunit and not on the solvent side as seen in the crystal structure. The model presented from these experiments places the CTD of IF3 at the subunit interface making contact with helices 23, 24 and 45 (cleavages colored red in Fig. 7.1-4C), whereas the linker and NTD were oriented towards the platform (colored yellow). This position for IF3 is in closer agreement with the cryo-EM localization but also the location of eukaryotic IF3 (eIF3) on the rat liver 40S ribosomal subunit to the interface surface by immuno-EM [97]. Indeed, mutation of G791A in h24 leads to a 10-fold decrease in the affinity of IF3 for the 30S subunit [98].

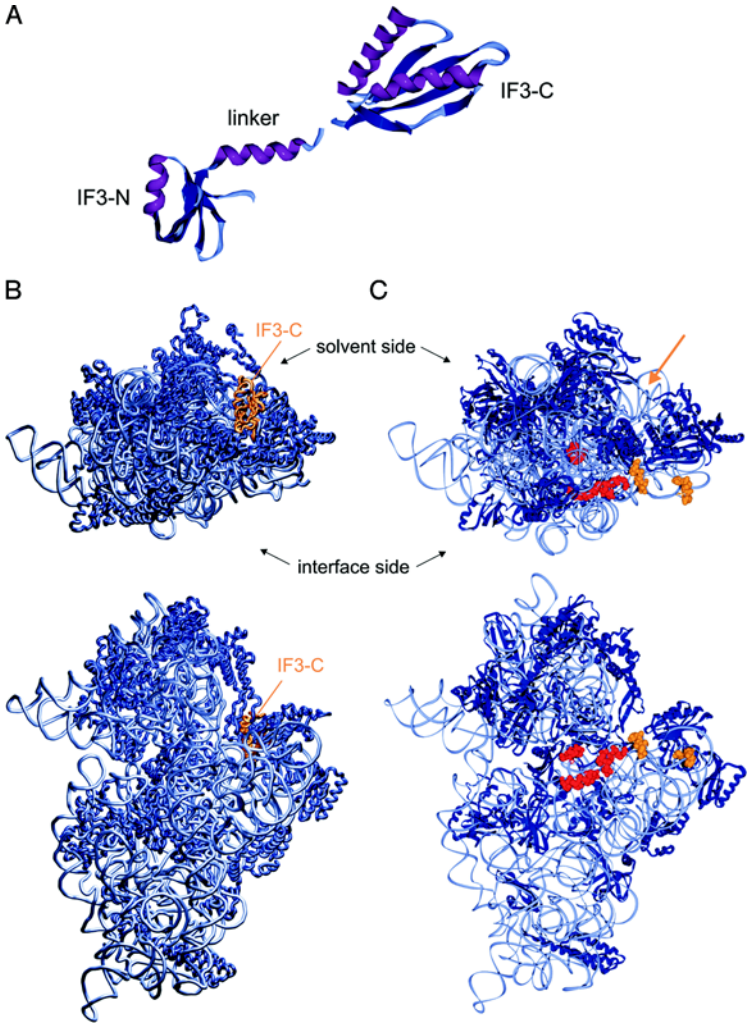


Figure 7.1-4 Where does IF3 really bind on the small ribosomal subunit? (A), IF3 is composed of an N-terminal domain linked by a long α-helix linker region (pdb 1TIF) to the C-terminal domains (pdbTIG) [88]. Both domains have a similar α/β topology with an exposed β-sheet that is reminiscent of several ribosomal and other RNA-binding proteins. (B), The CTD of IF3 was found to bind to the solvent side of the *T. thermophilus* 30S subunit by crystallography (pdb 1i96; [92]). Two views are shown: (i) a view from above looking down onto the head of the 30S subunit, with IF3C shown in orange, the 16S rRNA in light blue and ribosomal proteins in dark blue. (ii) View from 50S side onto the interface surface of the small subunit such that the IF3-C binding site is clearly on the back or solvent side. (C), Chemical cleavage of the 16S rRNA from specific sites on the C-terminal (cleavages colored red; locate mainly to 790 loop (h24) and 1400 region of h44) and linker region (yellow; found in predominantly in h23) of IF3 suggest that IF3 binds on the interface side of the 30S subunit [103], in contrast with the crystallography results. In the view from above the position of IF3-C found in the crystallography study is indicated.

Perhaps the most important observation is the recent report that the CTD of IF3 can perform all of the aforementioned activities attributed to the full-length molecule, when added in amounts (10–40 times) compensating for the reduced binding affinity [87]. In fact, the NTD alone displayed no affinity for the ribosome and no detectable functions even with high excess of protein, suggesting that the NTD is only stabilizing the interaction of the CTD with the 30S subunit. Since no conformational changes in the 30S subunit were observed upon binding of the CTD of IF3 in the crystal structure of the complex, it is difficult to envisage how IF3 can perform all its functions from a remote position on the solvent side of the 30S subunit. Certainly, further experiments need to be performed to address this issue, since if there is a second binding site for IF3, what is the function of this site?

References

1 J. Shine, L. Dalgarno, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 1342–1346.
2 J. A. Steitz, K. Jakes, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 4734–4738.
3 A. Hui, B. H. A. De, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4762–4766.
4 M. Brink, H. de Boer, *Meth. Mol. Biol.* **1998**, *77*, 117–128.
5 G. Z. Yusupova, M. M. Yusupov, J. H. Cate et al., *Cell* **2001**, *106*, 233–241.
6 B. Larsen, N. M. Wills, R. F. Gesteland et al., *J. Bacteriol.* **1994**, *176*, 6842–6851.
7 R. B. Weiss, D. M. Dunn, J. F. Atkins et al., *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 687–693.
8 V. Marquez, D. N. Wilson, K. H. Nierhaus, *Biochem. Soc. Trans.* **2002**, *30*, 133–140.
9 I. Moll, S. Grill, C. O. Gualerzi et al., *Mol. Microbiol.* **2002**, *43*, 239–246.
10 R. Boelens, C. O. Gualerzi, *Curr. Protein Pept. Sci.* **2002**, *3*, 107–119.
11 C. Gualerzi, L. Brandi, E. Caserta et al.: in *The Ribosome, Structure, Function, Antibiotics, and Cellular Interactions*, eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, 477–494.
12 H. S. Cummings, J. W. B. Hershey, *J. Bacteriol.* **1994**, *176*, 198–205.
13 A. Murzin, *EMBO J.* **1993**, *12*, 861–867.
14 D. E. Draper, L. P. Reynaldo, *Nucleic Acids Res.* **1999**, *27*, 381–388.
15 V. Ramakrishnan, S. W. White, *Trends Biochem. Sci.* **1998**, *23*, 208–212.
16 J. L. Battiste, T. V. Pestova, C. U. T. Hellen et al., *Mol. Cell* **2000**, *5*, 109–119.
17 A. Kay, M. Grunberg-Manago, *Biochim. Biophys. Acta* **1972**, *277*, 225–230.
18 M. H. Weber, C. L. Beckering, M. A. Marahiel, *J. Bacteriol.* **2001**, *183*, 7381–7386.
19 A. P. Carter, W. M. Clemons, Jr., D. E. Brodersen et al., *Science* **2001**, *291*, 498–501.
20 J. M. Ogle, D. E. Brodersen, W. M. Clemons Jr et al., *Science* **2001**, *292*, 897–902.
21 S. Brock, K. Szkaradkiewicz, M. Sprinzl, *Mol. Microbiol.* **1998**, *29*, 409–417.
22 A. Gnirke, K. H. Nierhaus, *J. Biol. Chem.* **1986**, *261*, 14506–14514.
23 D. Hartz, D. S. Mc Pheeters, L. Gold: in *The Ribosome: Structure, Function and Evolution*, eds W. E. Hill, A. Dahlberg, R. A. Garrett et al., American Society of Microbiology, Washington, DC 1990, 275–280.
24 K. H. Nierhaus, *Biochemistry* **1990**, *29*, 4997–5008.
25 J. H. Cate, M. M. Yusupov, G. Z. Yusupova et al., *Science* **1999**, *285*, 2095–2104.
26 M. Grunberg-Manago, P. Dessen, D. Pantaloni et al., *J. Mol. Biol.* **1975**, *94*, 461–478.
27 G. Blaha, N. Burkhardt, K. H. Nierhaus, *Biophys. Chem.* **2002**, *96*, 153–161.
28 I. Gabashvili, R. Agrawal, R. Grassucci et al., *EMBO J.* **1999**, *18*, 6501–6507.

29 A. Zamir, R. Miskin, D. Elson, *J. Mol. Biol.* **1971,** *60,* 347–364.
30 R. K. Agrawal, R. K. Lata, J. Frank, *Int. J. Biochem. Cell Biol.* **1999,** *31,* 243–254.
31 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001,** *292,* 883–896.
32 B. T. Wimberly, D. E. Brodersen, W. M. Clemons et al., *Nature* **2000,** *407,* 327–339.
33 K. D. Dahlquist, J. D. Puglisi, *J. Mol. Biol.* **2000,** *299,* 1–15.
34 D. Hartz, D. S. McPheeters, L. Gold, *Genes Dev.* **1989,** *3,* 1899–1912.
35 K. K. Mortensen, J. Kildsgaard, J. M. Moreno et al., *Biochem. Mol. Biol. Int.* **1998,** *46,* 1027–1041.
36 N. R. Nyengaard, K. K. Mortensen, S. F. Lassen et al., *Biochem. Biophys. Res. Commun.* **1991,** *181,* 1572–1579.
37 J. M. Moreno, H. P. Sorensen, K. K. Mortensen et al., *IUBMB Life* **2000,** *50,* 347–354.
38 J. M. Moreno, L. Drskjotersen, J. E. Kristensen et al., *FEBS Lett.* **1999,** *455,* 130–134.
39 J. M. Moreno, J. Kildsgaard, I. Siwanowicz et al., *Biochem. Biophys. Res. Commun.* **1998,** *252,* 465–471.
40 B. S. Laursen, de, A. S. S. A. et al., *Genes Cells* **2002,** *7,* 901–910.
41 B. S. Laursen, K. K. Mortensen, H. U. Sperling-Petersen et al., *J. Biol. Chem.* **2003,** *278,* 16320–16328.
42 R. Spurio, L. Brandi, E. Caserta et al., *J. Biol. Chem.* **2000,** *275,* 2447–2454.
43 S. Meunier, R. Spurio, M. Czisch et al., *EMBO J.* **2000,** *19,* 1918–1926.
44 P. Nissen, M. Kjeldgaard, S. Thirup et al., *Science* **1995,** *270,* 1464–1472.
45 P. Nissen, S. Thirup, M. Kjeldgaard et al., *Struct. Fold Des.* **1999,** *7,* 143–156.
46 M. Guenneugues, E. Caserta, L. Brandi et al., *EMBO J.* **2000,** *19,* 5233–5240.
47 C. O. Gualerzi, C. L. Pon, *Biochemistry* **1990,** *29,* 5881–5889.
48 U. L. RajBhandary, M. C. Chow: in *tRNA: Structure, Biosynthesis, and Functions,* eds D. Söll and U. L. RajBhandary, American Society for Microbiology, Washington, DC 1995, 511–528.
49 E. Schmitt, J. M. Guillon, T. Meinnel et al., *Biochimie* **1996,** *78,* 543–554.
50 C. Mayer, A. Stortchevoi, C. Kohrer et al., *Cold Spring Harb. Symp. Quant. Biol.* **2001,** *66,* 195–206.
51 M. A. Schäfer, A. O. Tastan, S. Patzke et al., *J. Biol. Chem.* **2002,** *277,* 19095–19105.
52 S. K. Choi, D. S. Olsen, A. Roll-Mecak et al., *Mol. Cell. Biol.* **2000,** *20,* 7183–7191.
53 A. Marintchev, V. G. Kolupaeva, T. V. Pestova et al., *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 1535–1540.
54 G. Boileau, P. Butler, J. W. B. Hershey et al., *Biochemistry* **1983,** *22,* 3162–3170.
55 P. Nissen, M. Kjeldgaard, J. Nyborg, *EMBO J.* **2000,** *19,* 489–495.
56 R. Agrawal, P. Penczek, R. Grassucci et al., *Proc. Natl.. Acad. Sci. USA* **1998,** *95,* 6134–6138.
57 R. K. Agrawal, A. B. Heagle, P. Penczek et al., *Nat. Struct. Biol.* **1999,** *6,* 643–647.
58 M. G. Gomez-Lorenzo, C. M. T. Spahn, R. K. Agrawal et al., *EMBO J.* **2000,** *19,* 2710–2718.
59 H. Stark, M. V. Rodnina, J. Rinkeappel et al., *Nature* **1997,** *389,* 403–406.
60 H. Stark, M. V. Rodnina, H. J. Wieden et al., *Cell* **2000,** *100,* 301–309.
61 H. Stark, M. V. Rodnina, H. J. Wieden et al., *Nat. Struct. Biol.* **2002,** *15,* 15–20.
62 M. Valle, J. Sengupta, N. K. Swami et al., *EMBO J.* **2002,** *21,* 3557–3567.
63 D. Moazed, R. R. Samaha, C. Gualerzi et al., *J. Mol. Biol.* **1995,** *248,* 207–210.
64 H. Wakao, P. Romby, S. Laalami et al., *Biochemistry* **1990,** *29,* 8144–8151.
65 A. La Teana, C. O. Gualerzi, A. E. Dahlberg, *RNA* **2001,** *7,* 1173–1179.

66 K. S. Wilson, H. F. Noller, *Cell* **1998**, *92*, 131–139.
67 S. Marzi, W. Knight, L. Brandi et al., *RNA* **2003**, *9*, 958–969.
68 D. Moazed, J. M. Robertson, H. F. Noller, *Nature* **1988**, *334*, 362–364.
69 D. M. Cameron, J. Thompson, P. E. March et al., *J. Mol. Biol.* **2002**, *319*, 27–35.
70 L. Belova, T. Tenson, L. Q. Xiong et al., *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3726–3731.
71 R. Benne, N. Naaktgeboren, J. Gubbens et al., *Eur J. Biochem.* **1973**, *32*, 372–380.
72 J. Dubnoff, A. Lockwood, U. Maitra, *J. Biol. Chem.* **1972**, *247*, 2884–2894.
73 J. Lee, T. Pestova, B. Shin et al., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16689–16694.
74 T. V. Pestova, I. B. Lomakin, J. H. Lee et al., *Nature* **2000**, *403*, 332–335.
75 J. Tomsic, L. A. Vitali, T. Daviter et al., *EMBO J.* **2000**, *19*, 2127–2136.
76 A. Antoun, M. Y. Pavlov, K. Andersson et al., *EMBO J.* **2003**, *22*, 5593–5601.
77 J. Weiel, J. W. Hershey, *Biochemistry* **1981**, *20*, 5859–5865.
78 G. Risuleo, C. Gualerzi, C. Pon, *Eur. J. Biochem.* **1976**, *67*, 603–613.
79 D. Hartz, J. Binkley, T. Hollingsworth et al., *Gene Dev.* **1990**, *4*, 1790–1800.
80 J. K. Sussman, E. L. Simons, R. W. Simons, *Mol. Microbiol.* **1996**, *21*, 347–360.
81 C. Sacerdot, P. Dessen, J. W. B. Hershey et al., *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 7787–7791.
82 C. Sacerdot, G. Fayat, P. Dessen et al., *EMBO J.* **1982**, *1*, 311–315.
83 C. L. Olsson, M. Graffe, M. Springer et al., *Mol. Gen. Genet.* **1996**, *250*, 705–714.
84 Y. Shimizu, A. Inone, Y. Tomari et al., *Nat. Biotechnol.* **2001**, *19*, 751–755.
85 J. H. Kycia, V. Biou, F. Shu et al., *Biochemistry* **1995**, *34*, 6183–6187.
86 E. de Cock, M. Springer, F. Dardel, *Mol. MicroBiol.* **1999**, *32*, 193–202.
87 D. Petrelli, A. LaTeana, C. Garofalo et al., *EMBO J.* **2001**, *20*, 4560–4569.
88 V. Biou, F. Shu, V. Ramakrishnan, *EMBO J.* **1995**, *14*, 4056–4064.
89 C. Garcia, P. L. Fortier, S. Blanquet et al., *Eur. J. Biochem.* **1995**, *228*, 395–402.
90 C. Garcia, P. L. Fortier, S. Blanquet et al., *J. Mol. Biol.* **1995**, *254*, 247–259.
91 J. P. McCutcheon, R. K. Agrawal, S. M. Philips et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4301–4306.
92 M. Pioletti, F. Schlunzen, J. Harms et al., *EMBO J.* **2001**, *20*, 1829–1839.
93 M. A. Firpo, M. B. Connelly, D. J. Goss et al., *J. Biol. Chem.* **1996**, *271*, 4693–4698.
94 C. Ehresmann, H. Moine, M. Mougel et al., *Nucleic Acids Res.* **1986**, *14*, 4803–4821.
95 L. A. MacKeen, L. Kahan, A. J. Wahba et al., *J. Mol. Biol.* **1980**, *255*, 10526–10531.
96 V. Ramakrishnan, *Cell* **2002**, *108*, 557–572.
97 S. Srivastava, A. Verschoor, J. Frank, *J. Mol. Biol.* **1992**, *226*, 301–304.
98 W. Tapprich, D. Goss, A. Dahlberg, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 4927–4931.
99 M. Sette, P. vanTilborg, R. Spurio et al., *EMBO J.* **1997**, *16*, 1436–1443.
100 H. Schindelin, W. Jiang, M. Inouye et al., *Proc. Natl. Acad. Sci. USA* **1994**, 91.
101 L. Sherlin, T. Bullock, K. Newberry et al., *J. Mol. Biol.* **2000**, *299*, 431–436.
102 A. Roll-Mecak, C. Cao, T. E. Dever et al., *Cell* **2000**, *103*, 781–792.
103 A. Dallas, H. F. Noller, *Mol. Cell* **2001**, *8*, 855–864.
104 H. U.Petersen, A.Danchin, M.Grunberg-Manago, *Biochemistry*, **1976**, *15*, 1357–1362.
105 K.Saito, L. C. Mattheakis, M. Nomura, *J. Mol. Biol.*, **1994**, *235*, 111–124.
106 A. G. Balakin, E. A. Skripkin, I. N. Shatsky, A. A. Bogdanov, *Nucl. Acids Res.*, **1992**, *20*, 563–571.
107 T. Udagawa, Y. Shimizu, T. Ueda, *J. Biol. Chem.*, **2004**, *279*, 8539–8546.
108 S. O'Donnell, G. Janssen, *J. Bacteriol.*, **2002**, *184*, 6730–6733.
109 M. A. Schäfer, A. O. Tastan, S. Patzke et al., *Biol. Chem.*, **2002**, *277*, 19095–19105.

110 M. M. Yusupov, G. Z. Yusupova, A.Baucom et al., *Science,* **2001,** *292,* 883–896.
111 S. Grill, I. Moll, D. Hasenohrl et al., *FEBS Lett.,* **2001,** *495,* 167–171.
112 I. Moll, U. Bläsi, *Biochem. Biophys. Res. Comm.,* **2002,** *297,* 1021–1026.
113 K.Tedin, A. Resch, U. Blasi, *Mol. Microbiol.,* **1997,** *25,* 189–199.
114 J. Sengupta, R. Agrawal, and J. Frank, *Proc. Natl Acad. Sci. USA.,* **2001,** *98,* 11991–11996.

7.2 Mechanism and Regulation of Protein Synthesis Initiation in Eukaryotes

Alan G. Hinnebusch, Thomas E. Dever, and Nahum Sonenberg

7.2.1 Introduction

7.2.1.1 Overview of Translation-initiation Pathways in Eukaryotes and Prokaryotes

The decoding of an mRNA transcript and synthesis of a protein by the ribosome requires the assistance of two highly conserved translation-elongation factors known as EF1A and EF2 in eukaryotes (and archaea) and EF-Tu and EF-G in eubacteria (reviewed by Gualerzi et al. [1] and in Chap. 7.1). Studies on the structure and mechanics of the ribosome and the roles of the elongation factors have revealed the intricate functions of these enzymes, and the complex reactions required for efficient protein synthesis of high fidelity. Prior to the elongation phase of translation, the ribosome must be loaded with the initiator methionyl tRNA (Met-tRNA$_i^{Met}$) and assembled on an mRNA at the start codon in a process referred to as initiation. The general scheme of translation initiation seems to be conserved throughout evolution, and a core set of *trans*-acting initiation factors that promote these reactions are similarly conserved. In eukaryotes, several embellishments to the core initiation pathway have evolved, and these new steps require the functions of additional initiation factors. The eukaryotic-specific factors not only increase the rate and fidelity of the process, but also provide a means to regulate protein synthesis in response to cellular or environmental signals.

The general schemes of translation initiation in prokaryotes and eukaryotes, summarized in Fig. 7.2-1A and B is based largely on the biochemical activities of the purified factors and other components of the protein synthesis machinery. (The relevant literature supporting the depiction of this pathway is not cited in this introductory note, but can be found in the body of this chapter for the eukaryotic pathway.) The small and large ribosomal subunits have distinct roles in protein synthesis. The small subunit is responsible for decoding and contains binding sites for the tRNAs and the mRNA, whereas the large subunit contains the active site of the peptidyltransferase. In both kingdoms, the starting point for assembly of the initiation complex is the production of free small ribosomal subunits. In eukaryotes, binding of the initiation factors (eIFs) 1, 1A and 3 to the small ribosomal subunit prevents its coupling with the large subunit. The initiation complex is then assembled on the free small subunit as follows.

The first step is the binding of Met-tRNA$_i^{Met}$ to the P-site of the small subunit. The Met-tRNA$_i^{Met}$ does not bind alone, rather it is delivered to the ribosome by eIF2. eIF2, similar to IF2 in eubacteria, is a GTP-binding (G) protein; however, only eIF2 requires GTP to bind Met-tRNA$_i^{Met}$. The eIF2 forms a stable ternary complex with

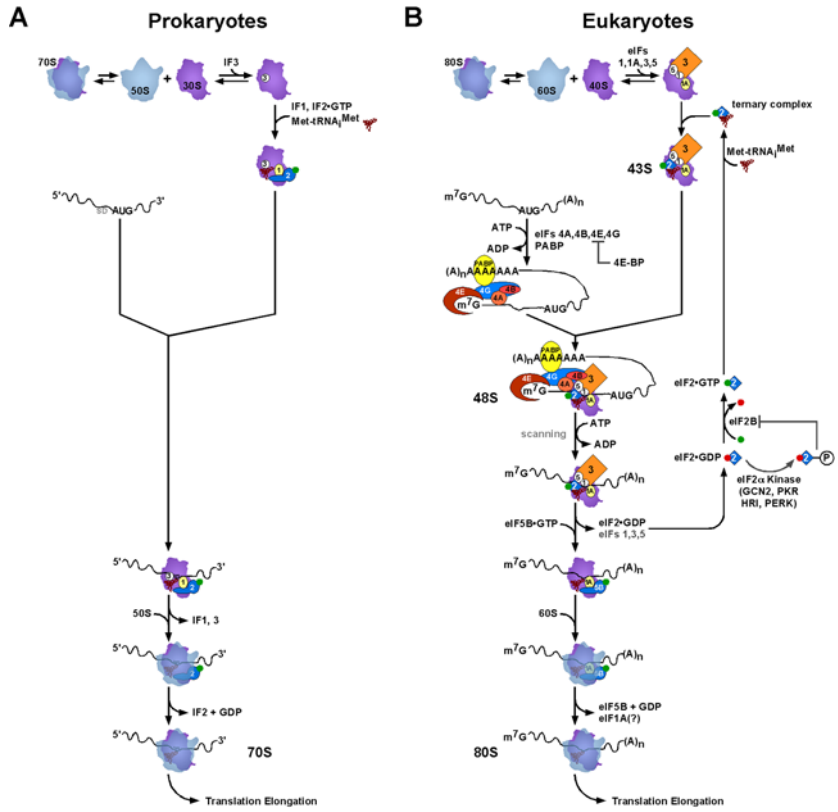


Figure 7.2-1 Pathways of translation initiation in prokaryotes and eukaryotes. The individual steps in the prokaryotic (A) and eukaryotic (B) pathways have been aligned to reflect the conservation of the reactions and functions of the factors. The three initiation factors (IF1-IF3) in prokaryotes and the various eukaryotic initiation factors (eIFs) are labeled. Biochemical studies have suggested that an alternate pathway in which mRNA binding to the 30S subunit precedes Met-tRNA$_i^{Met}$ binding may also function in prokaryotes (see Chap. 7.1). At the completion of the initiation pathway, Met-tRNA$_i^{Met}$ is bound to the ribosomal P site and the A site is vacant waiting for binding of the first elongating tRNA in an eEF1 A/EF-Tu•GTP•aminoacyl-tRNA ternary complex. The green dot represents GTP and the red dot is GDP.

GTP and Met-tRNA$_i^{Met}$, and this eIF2•GTP•Met-tRNA$_i^{Met}$ ternary complex (TC) binds to the 40S subunit. The factors eIF1, eIF1A, and eIF3 all stimulate binding of the TC to the 40S ribosome to form a 43S preinitiation complex. Binding of Met-tRNA$_i^{Met}$ to IF2 in bacteria is dependent on formylation of the methionine, a modification that does not occur in eukaryotes.

The next step in the pathway entails binding of the small ribosomal subunit to mRNA. Biochemical studies on translation initiation in prokaryotes revealed no fixed order of binding by fMet-tRNA$_i^{Met}$ and mRNA to the 30S subunit; however, genetic studies support the scheme shown in Fig. 1A wherein tRNA binding precedes mRNA binding (1). The 3′ end of 16S rRNA in the 30S ribosomal subunit binds dierctly to the Shine-Dalgarno (SD) sequence located just upstream of the start codon (see Chap. 7.1). In the cap-dependent initiation pathway found in eukaryotes, formation of the 43S preinitiation complex containing Met-tRNA$_i^{Met}$ is a prerequisite for mRNA binding and the formation of a 48S preinitiation complex (Fig. 7.2-1B). The m^{7}GpppN (where N is any nucleotide) cap structure at the 5′-end of the mRNA is the entry point for 40S ribosomes, and the start site is selected as the ribosome scans along the mRNA. Multiple initiation factors mediate the 5′-cap-binding specificity of the 48S complex. The eIF4F cap-binding complex consists of the cap-binding protein eIF4E, the ATP-dependent RNA helicase eIF4A, and the scaffold/adaptor protein eIF4G. The eIF4G is thought to facilitate ribosome binding near the 5′-cap by forming a bridge between the eIF4E•cap complex and eIF3, a constituent of the 43S complex. The eIF4G also has a binding domain for the poly(A)-binding protein (PABP) and thus can mediate binding of both ends of the mRNA to eIF3 and the 43S complex (Fig. 7.2-1B).

Following the formation of the eukaryotic 48S complex near the cap, the 40S ribosome scans in an ATP-dependent reaction in search of a start codon. It is generally thought that ATP is consumed by eIF4A in the removal of RNA secondary structures that impede sliding of 40S subunits along the mRNA. The eIF4B greatly stimulates the RNA helicase activity of eIF4A. Typically, translation initiates at the first AUG codon encountered during scanning from the cap, and several factors (eIF1, eIF1A eIF4G, and eIF2) play important roles in scanning and AUG recognition. Pairing of the anticodon of Met-tRNA$_i^{Met}$ with the AUG codon triggers GTP hydrolysis by eIF2, stimulated by the concerted action of the GTPase activating protein (GAP) eIF5 and the 40S ribosome itself. The eIF2•GDP and many, if not all, of the other initiation factors are subsequently released from the 48S complex (Fig. 7.2-1B).

In prokaryotes, IF1 and IF3 dissociate from the ribosome following AUG recognition, whereas IF2 remains bound (Fig. 7.2-1A). eIF5B, the eukaryotic homolog of bacterial IF2, promotes joining of the large ribosomal subunit to the preinitiation complex. Subunit joining triggers GTP hydrolysis by eIF5B, and release of these factors. (As eIF1A interacts with eIF5B, it is possible that these factors are released as a complex following GTP hydrolysis by eIF5B (Fig. 7.2-1B).) Finally, the 80S initiation complexes are ready to enter the elongation phase of protein synthesis and produce a protein. The eIF2•GDP must be recycled to eIF2•GTP by the guanine nucleotide exchange factor (GEF) eIF2B to permit a new round of initiation (Fig. 7.2-1B). In contrast, eIF5B (and IF2) are thought to recycle to the GTP-bound state without the assistance of a GEF.

A fraction of eukaryotic mRNAs is translated by an alternative mechanism, known as internal initiation, in which the 40S ribosome, with or without the assistance of eIFs, binds directly to an internal sequence in the mRNA upstream of the start codon. The 40S subunit is transferred from this internal ribosome entry site (IRES)

to the start codon either directly or through a short period of scanning. This mode of ribosome binding is similar to the SD/30S subunit interaction in bacteria. Consistently, IRES-dependent translation mechanisms dispense with some, or even all, of the canonical eukaryotic initiation factors.

7.2.1.2 Conservation and diversity of translation-initiation factors among bacteria, archaea and eukaryotes

Comparison of the translation-initiation factors in bacteria (Fig. 7.2-1A and Chap. 7.1) and in eukaryotes (Fig. 7.2-1B) reveals only three initiation factors in bacteria against at least 12 factors in eukaryotes. Several of the eukaryotic factors are composed of multiple polypeptides, and the 28 polypeptides that comprise the full complement of eukaryotic factors are listed in Tables 7.2-1 and 7.2-2, highlighting their functions and the conservation of sequences among human, plant and yeast homologs. Examination of the genome sequences from several archaea reveals orthologs of a subset of the eukaryotic factors, suggesting that archaea possess an initiation mechanism, which is intermediate in complexity between the prokaryotic and eukaryotic pathways.

The three bacterial factors IF1, IF2, and IF3 are functionally or structurally conserved in all three kingdoms, interact directly with the ribosome, and promote conserved steps in the initiation pathways (Fig. 7.2-1 and 7.2-2). The bacterial factors IF1 and IF2 perform analogous roles to the eukaryotic factors eIF1A and eIF5B, consistent with the structural conservation of these factors through evolution. Pairwise alignments among the bacterial, archaeal, and eukaryotic IF1/aIF1A/eIF1A factors reveal sequence identities ranging from 21 to 38% [2], and the solution structures of IF1 and eIF1A contain homologous oligonucleotide/oligosaccharide binding (OB) folds [3, 4]. The bacterial, archaeal and eukaryotic IF2/aIF5B/eIF5B sequences are 27–39% identical with aIF5B, excluding the nonconserved N-terminal domain (NTD) present in prokaryotic and eukaryotic factors [5]. Both IF2 and eIF5B have been implicated in promoting Met-tRNA$_i^{Met}$ binding to the ribosome and in subunit joining. In addition, IF2 and eIF5B physically and functionally interact with the homologous factors IF1 and eIF1A, respectively, although probably only on the ribosomes in the bacterial situation [6, 7]. Both bacterial IF3 and eukaryotic eIF1 play important roles in translation start site recognition. Intriguingly, eIF1 and the C-terminal domain (CTD) of IF3 have similar α/β-fold structures with a four- or five-stranded β-sheet packed against two α-helices (see Ref. [8]). The presence of a highly conserved (25–30% sequence identity) eIF1-like protein, distinct from IF3, in some, but not all, bacteria indicates that additional investigations are necessary to determine whether IF3 and eIF1 are true homologs.

The eIF5A in eukaryotes and EF-P in bacteria share approximately 20% sequence identity [2], and both proteins have been implicated in translation. They both have stimulatory activities in model assays of first-peptide-bond formation, especially the synthesis of methionyl puromycin; however, their precise roles in protein synthesis are unknown [9]. As depletion of eIF5A in yeast only slightly impaired protein

Table 7.2-1 Initiation factors from mammalian, plant, and yeast cells [a]

Molecular weight[b]						
Name	**Human**	**Arabidopsis**	**Saccharomyces**	**Yeast gene**	**% identity[c]**	**Functions[d]**
eIF1	12.6	12.6	12.3	*SU11*	58	AUG recognition; promotes TC and mRNA binding to 40S; 80S anti-association, binds eIF3c, eIF3a, and eIF5
eIF1A	16.5	16.6	17.4	*TIF11*	65	Promotes TC and mRNA binding to 40S; 80S anti-association; binds RNA, eIF5B, eIF2, eIF3, and eIF5B
eIF2α	36.2	41.6	34.7	*SUI2*	58	TC component; AUG recognition; mediates inhibitory interaction with eIF2B on phosphorylation of Ser^{51}, binds eIF2γ, binds eIF2B $\alpha/\beta/\delta$ subcomplex when phosphorylated
eIF2β	39.0	26.6	31.6	*SUI3*	42	TC component; GTP/Met-tRNA$_i^{Met}$ binding; AUG recognition; binds eIF2γ, eIF2Bε, eIF5, eIF3a, mRNA
eIF2γ	51.8	50.9	57.9	*GCD11*	71	TC component; GTP/Met-tRNA$_i^{Met}$ binding; GTPase; AUG recognition; binds eIF2α and eIF2β
eIF2Bα	33.7	39.8	34.0	*GCN3*	42	Nonessential in yeast; regulatory subunit that helps bind eIF2(αP) and inhibit GEF function; forms subcomplex with eIF2B β/δ
eIF2Bβ	39.0	43.6	42.6	*GCD7*	36	Regulatory subunit; helps bind eIF2(αP) and inhibit GEF function; forms subcomplex with eIF2B α/δ
eIF2Bγ	50.4		65.7	*GCD1*		Promotes GEF function of catalytic subdomain; forms subcomplex with eIF2Bε
eIF2Bδ	57.8	29.4	70.9	*GCD2*	36	Regulatory subunit; helps bind eIF2(αP) and inhibit GEF function; forms subcomplex with eIF2B α/β
eIF2Bε	80.2	81.9	81.2	*GCD6*	30	GEF catalytic subunit; forms subcomplex with eIF2Bγ
eIF4AI	44.4	46.7	45.1	*TIF1*	65	ATPase, RNA helicase
eIF4AII	46.3	46.8	44.6	*TIF2*		ATPase, RNA helicase

Molecular weight[b]						
Name	**Human**	**Arabidopsis**	**Saccharomyces**	**Yeast gene**	**% identity[c]**	**Functions[d]**
eIF4B	69.2	57.6	48.5	*TIF3*	22	Binds RNA and eIF3g; stimulates eIF4A helicase activity; nonessential in yeast
eIF4E	25.1	26.5	24.3	*CDC33*	33	Binds m^7G-cap of mRNA and eIF4G
eIFiso4E		22.5				
eIF4GI	175.6	153.2	107.1	*TIF4631*	22	Binds eIF4E, eIF4A, eIF3, eIF5, PABP, and kinase MNK1
eIF4GII	176.6	176.5	103.9	*TIF4632*	21	Binds eIF4E, eIF4A, eIF3, eIF5, PABP, and kinase MNK1
eIFiso4G		87.0				
eIF5	48.9	48.6	45.2	*TIF5*	39	AUG recognition; stimulates eIF2 GTPase in conjunction with 40S subunit; promotes TC and eIF3 binding to 40S; binds eIF2β, eIF1, and eIF3c
eIF5B	139.0		112.3	*FUN12*	70	Nonessential in yeast; GTPase; promotes subunit joining; stabilizes Met-$tRNA_i^{Met}$ binding to 40S; binds eIF1A
PABP	70.7	68.7	64.2	*PAB1*	59	Binds poly(A) tail of mRNA and eIF4G.
eIF3[e]						5 of 6 subunits essential in yeast; 10 subunits in human factor; 80S anti-association; promotes TC and mRNA binding to 40S; binds eIF5, TC, and eIF1 simultaneously in the MFC; binds eIF4G, eIF4B, and multiple 40S ribosome components

a Adapted from Ref. [174].
b The masses in kDa pertain to human or rat, Arabidopsis thaliana, and Saccharomyces cerevisiae proteins.
c Percent sequence identity shared by yeast and human proteins (from Ref. [174]).
d Some functions have been demonstrated only for the mammalian or yeast factor (see text for details).
e See Table 7.2-2 for detailed information on each subunit.

Table 7.2-2 eIF3 subunits from mammalian, plant, and yeast cells

Subunit	Human		Arabidopsis	Saccharomyces			Motifs/functions/comments[c]
	Name	MW[a]	MW	Name	Gene	MW	
eIF3a	p170	166.6	114.3	p110	*TIF32/ RPG1*	110.3	Contains PCI motif[b]; binds eIF3 subunits b, c, and j, eIF4B[d], eIF2*β*, eIF1, RPS0, RPS10, and 18S rRNA; promotes eIF2-eIF3 and eIF1-eIF3 interactions and 40S binding of eIF3; promotes TC and mRNA binding to 40S subunits in conjunction with eIF3b/c, plus a step(s) post-48S assembly
eIF3b	p116/ p110	92.4	81.9	p90	*PRT1*	88.1	Contains RRM; binds eIF3 subunits a, c, e[d], j, i, and g; promotes TC and mRNA binding to 40S subunits
eIF3c	p110	105.3	102.9	p93	*NIP1*	93.2	Contains PCI motif; binds eIF3 subunits a, b, and e[d], eIF1, eIF5, and RPS0; promotes eIF3 interactions with eIFs 1 and 5, eIF2-eIF3 interaction (via eIF5), and 40S binding of eIF3; promotes TC and mRNA binding to 40S subunits
eIF3d	p66	64.0	66.2	na[e]	na	na	Binds RNA[d], eIF3e[d]
eIF3e	p48/ INT-6	52.2	51.8	na	na	na	Contains PCI motif, binds eIF3 subunits a, b[d], c[d], d[d]
eIF3f	p47	37.6	31.9	na	na	na	Contains MDN[b] motif
eIF3g	p44	35.6	32.7	p33	*TIF35*	30.5	Contains RRM and Zn domain; binds eIF3 subunits b and i; binds RNA[d] and eIF4B
eIF3h	p40	39.9	38.4	na	na		Contains MDN[b] motif
eIF3i	p36/ TRIP-1	36.5	36.4	p39	*TIF34*	38.8	Contains 7 WD repeats; binds eIF3 subunits b and g
eIF3j	p35	29.1	na		*HCR1*	29.6	Nonessential in yeast; binds eIF3 subunits a and b; promotes MFC integrity and a step(s) post-48S assembly; 40S biogenesis
eIF3k	na	na	25.7	na	na		

a Masses in kDa calculated from deduced protein sequence.
b PCI motif has been called the PINT motif; MDN has been called the MPN motif [398, 399].
c Demonstrated for Saccharomyces subunit unless otherwise indicated.
d Demonstrated for human subunit only.
e na; not applicable, protein not found in eIF3 of this organism.

synthesis [10], the assignment of EF-P/eIF5A as a universally conserved translation-initiation factor should be viewed cautiously. This is especially true in the case of EF-P since it has been shown, using an *in vitro* translation system composed completely from purified factors, that EF-P was not necessary for translation and, furthermore, the addition of the factor did not improve the rate or efficiency of translation (T. Ueda, pers. comm.).

Building on this conserved core group of factors, the RNA-associated factor eIF4A and the Met-tRNA$_i^{Met}$-binding factor eIF2 are additionally present in archaea (Fig. 7.2-2). Whereas the translation-initiation pathway in archaea is not well understood, the identification of eIF2 and eIF4A suggests that a scanning-type mechanism for mRNA binding and AUG recognition may operate on some archaeal mRNAs. In fact, there is evidence that recognition and translation of the first open reading frame (ORF) on polycistronic mRNAs in some archaea occurs via scanning or direct binding of the ribosome to a 5′-terminal start codon, whereas subsequent ORFs are recognized by a bacterial-like SD interaction [11]. The absence of the cap-binding protein eIF4E as well as eIF4G from archaea is consistent with lack of m^{7}GpppN caps on archaeal mRNAs. Finally, several factors, including other members of the RNA-binding eIF4 family of proteins, eIF3, and the eIF2-interacting proteins eIF5 and eIF2B are restricted to eukaryotes (Fig. 7.2-2).

The presence of the GTPase eIF2, as well as the complex cap-dependent mRNA binding and scanning mechanisms in eukaryotic initiation, provide new opportunities for translational regulation. Phosphorylation of the α-subunit of eIF2 converts eIF2 from a substrate to competitive inhibitor of eIF2B, impairing the recycling of inactive eIF2•GDP to active eIF2•GTP, thereby inhibiting protein synthesis (Fig. 7.2-1B). Binding of 4E-binding proteins (4E-BPs) to eIF4E blocks eIF4E binding to eIF4G and prevents formation of the cap-binding complex eIF4F, thus impairing mRNA binding to the ribosome. Phosphorylation of the 4E-BPs prevents their binding to eIF4E and relieves translational inhibition. Regulation of translation-initiation factors has not been reported in bacteria. Thus, the appearance of new mechanisms and factors in evolution to facilitate both Met-tRNA$_i^{Met}$ and mRNA binding to the ribosome has provided powerful means to regulate initiation in eukaryotes.

7.2.1.3 **Genetic assays for *in vivo* functions of eIF2**

Many advances in our knowledge of the functions of eIF2, its GEF (eIF2B), and its GAP (eIF5) in recruitment of Met-tRNA$_i^{Met}$ and recognition of the start codon, have come from genetic analysis of translational control in yeast. Accordingly, these genetic systems are summarized briefly before considering the biochemical mechanisms of these steps in the pathway. As mentioned above, recycling of eIF2-GDP to eIF2-GTP by eIF2B is impaired by phosphorylation of eIF2 on Ser-51 of its α-subunit (eIF2[αP]) (Fig. 7.2-1B). As eIF2 is generally present in excess of eIF2B, and phosphorylation of eIF2-GDP increases its affinity for eIF2B, the recycling of eIF2 can be impaired by phosphorylation of only a fraction of eIF2 [12, 13]. Four eIF2α – Ser-51 kinases regulated by different signals have been identified in mammalian cells: HRI (heme deprivation), PKR (double-stranded RNA produced in virus-

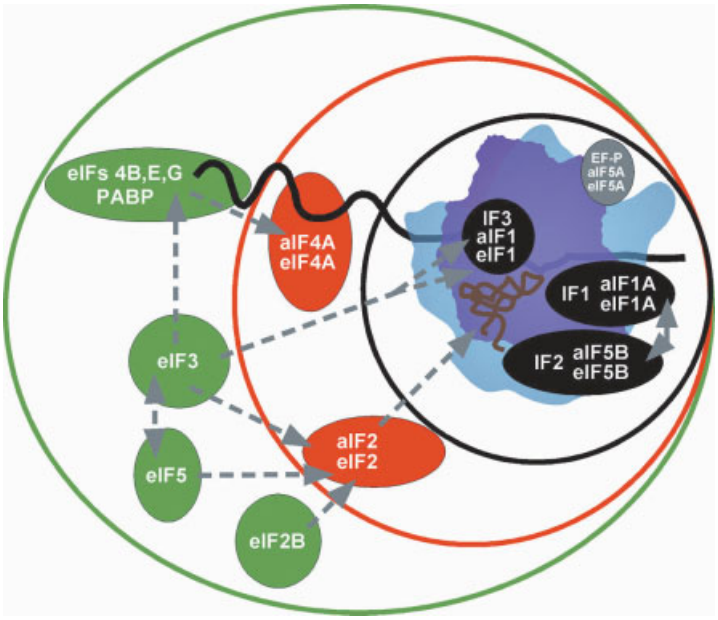


Figure 7.2-2 Conservation of a core set of translation-initiation factors through evolution. The translation-initiation factors identified in bacteria, archaea and eukaryotes are depicted based on their conservation through evolution. Depicted in black and enclosed within the innermost black circle, the three universally conserved initiation factors IF1/eIF1A, IF2/eIF5B, and IF3/eIF1 interact directly with the ribosome. The proposed grouping of IF3 and eIF1 is based on similar α/β-fold structures for the two factors, and their common function to insure accurate Met-tRNA$_i^{Met}$ and start site selection in the ribosome P-site. The factor EF-P/eIF5A, depicted in gray, is also universally conserved, however, questions have been raised regarding the assignment of this protein as a translation factor, and lowering the amount of eIF5A in yeast did not appear to impair translation initiation [10]. Building upon this core set of factors, the DEAD-box RNA helicase eIF4A and the tRNA delivery factor eIF2 were added in archaea and retained in eukaryotes (red circle). Finally, the eIF4 family of factors that function in mRNA binding, the eIF3 complex that facilitates both mRNA and tRNA binding to the 40S subunit, and the proposed GAP (eIF5) and GEF (eIF2B) for eIF2 were added in eukaryotes (green circle). Dashed gray arrows indicate protein–protein, protein–RNA (eIF2-tRNA), or factor–ribosome (eIF3-40S) interactions.

infected cells), PERK (unfolded proteins in the endoplasmic reticulum), and GCN2 (serum or amino acid starvation, UV irradiation) [14–18]. GCN2 (General Control Nonderepressible 2) is the only eIF2α kinase in *Saccharomyces cerevisiae,* where it is

activated by diverse starvation or stress conditions, including amino acid limitation. Physiological activation of GCN2 in amino acid-starved yeast cells does not generate eIF2[αP] at a level that prevents eIF2 recycling and blocks protein synthesis; instead, it specifically increases translation of *GCN4* mRNA, encoding the transcriptional activator of amino acid biosynthetic enzymes subject to the general amino acid control. The specific induction of *GCN4* translation by eIF2[αP] is mediated by four short open reading frames (uORFs) in the leader of *GCN4* mRNA [19].

According to the current model (Fig. 7.2-3), ribosomes scanning from the 5′-cap translate uORF1, and ~50% resume scanning as 40S subunits. Under nonstarvation conditions, all of these reinitiating ribosomes rebind the TC and reinitiate at uORFs 2–4, after which they dissociate from the mRNA and are prevented from translating *GCN4*. Phosphorylation of eIF2α by GCN2 in starved cells inhibits eIF2B and lowers the concentration of TC. Consequently, as many as ~50% of the 40S subunits scan-

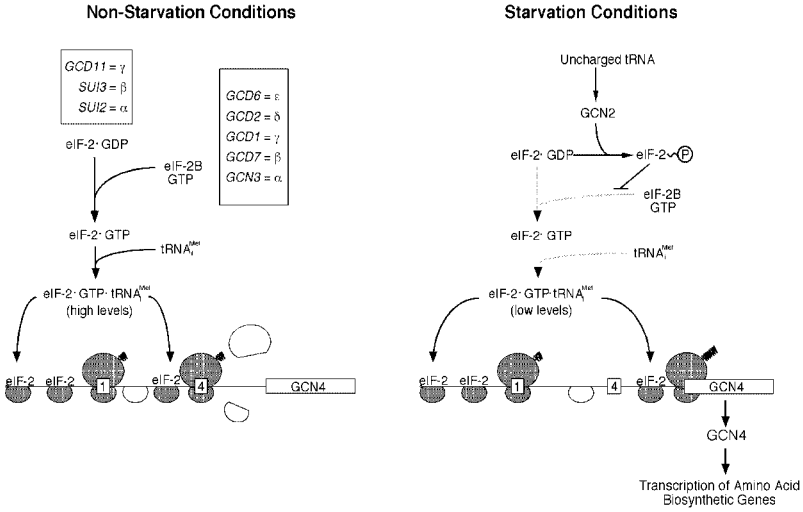


Figure 7.2-3 Molecular model for *GCN4* translational control. *GCN4* mRNA is depicted with uORF1 and 4 and the *GCN4* coding sequences shown as boxes. For simplicity, uORF2 and uORF3 were omitted because they are functionally redundant with uORF4. The 40S ribosomal subunits are shaded when associated with the ternary complex TC and competent to reinitiate at the next start codon they encounter. 80S ribosomes are shown translating uORF1, uORF4, or *GCN4* with the synthesized peptides depicted as coils. Free 40S and 60S subunits are shown dissociating from the mRNA following translation of uORF4. The three subunits of eIF2 and the five subunits of eIF2B are listed in the boxes on the left panel. Following translation of uORF1, the 40S ribosome remains attached to the mRNA and resumes scanning. Under nonstarvation conditions, the 40S quickly rebinds the TC and reinitiates at uORF4 because the TC concentration is high. Under amino acid starvation conditions, many 40S ribosomes fail to rebind the TC until scanning past uORF4, because the TC concentration is low, and reinitiate at *GCN4* instead. TC levels are reduced in starved cells due to phosphorylation of eIF2 by the kinase GCN2, converting eIF2 from substrate to inhibitor of its guanine-nucleotide exchange factor eIF2B.

ning from uORF1 reach uORF4 before rebinding the TC, and lacking Met-tRNA$_i^{Met}$, they bypass the uORF4 start codon. Most of these ribosomes rebind the TC before reaching the *GCN4* start codon. Thus, reducing TC levels by phosphorylating eIF2α allows a fraction of scanning 40S subunits to by-pass the inhibitory uORFs 2–4 and re-initiate at *GCN4* instead (see Refs. [15, 19] and references therein).

Mutants harboring lesions in eIF2γ [20] and the β, γ, δ, and ε subunits of eIF2B [19] were first isolated by their constitutive derepression of *GCN4* translation (general control derepressed, or Gcd$^-$ phenotype). These mutations also produce a slow-growth phenotype (Slg$^-$) and reduce rates of protein synthesis on rich medium, indicating nonlethal impairment of the essential functions of eIF2 or eIF2B [21–24]. Mutations in eIF2β and eIF2α can also produce Gcd$^-$ and Slg$^-$ phenotypes [25], as does reducing the copy number of *IMT* genes, encoding tRNA$_i^{Met}$ [26]. The derepression of *GCN4,* conferred by these Gcd$^-$ mutations is maintained in *gcn2Δ* cells [25, 27], suggesting that the mutations reduce TC levels independent of eIF2α phosphorylation. Consistently, overexpressing eIF2 prevents derepression of *GCN4* in starved wild-type cells (Gcn$^-$ phenotype) [26], presumably by offsetting the inhibitory effect of eIF2[αP] on TC formation (Fig. 7.2-3). Thus, the level of *GCN4* expression is a sensitive *in vivo* indicator of the functions of eIF2 and eIF2B in TC formation.

The genetic studies of Donahue and colleagues have provided a valuable entry into the mechanism of start codon selection by the TC. They have isolated mutations allowing expression of a defective *HIS4* gene harboring a non-AUG start codon. The isolation of mutations with this Sui$^-$ (suppressor of initiation codon) phenotype in one of the genes encoding tRNA$_i^{Met}$ showed that base-pairing between the start codon and Met-tRNA$_i^{Met}$ plays a dominant role in directing the 40S subunit to the initiation site [28]. The Sui$^-$ selection also yielded mutations in all three subunits of eIF2, eIF5 (the GAP for eIF2), and eIF1, implicating these factors in stringent selection of the start codon.

7.2.2 Generation of Free 40S Subunits and 40S Binding of Met-tRNA$_i^{Met}$

7.2.2.1 Dissociation of Idle 80S Ribosomes

Most ribosomal subunits that are not engaged in translation occur in idle 80S ribosomes, or "80S couples", which must be dissociated into 40S and 60S subunits to allow assembly of the 43S preinitiation complex. The eIF1A, eIF3, and eIF1 have all been implicated in this reaction, but the molecular mechanisms are unknown. The mammalian eIF3 can bind to 40S ribosomes in the absence of other factors [29], but its binding site on the 40S subunit, as visualized in EM images of negatively stained native 40S subunits, does not seem to preclude association of the 60S subunit [30, 31]. Most of the mass of eIF3 was found attached to the back lobes rather than to the 60S-interface side of the 40S subunit. Thus, eIF3 might function indirectly by

producing an allosteric change in the 40S subunit that inhibits 60S joining. It may also sterically impede 60S joining by stabilizing the binding of TC to the interface side of the 40S subunit. Consistent with the latter is a report that eIF3 does not exhibit ribosome dissociation activity alone, but can prevent 60S subunits from displacing the TC from 40S subunits in the absence of AUG or mRNA [32]. The eIF1A can augment this anti-association activity of eIF3, and can also function with eIF1 in the absence of eIF3 to prevent disruption of 43S complexes by 60S subunits [33]. The eIF3 in yeast is physically linked to eIF1, eIF5, and the TC in a multifactor complex (MFC) that can exist free of ribosomes [34]. Hence, all of these factors may bind coordinately to the 40S subunit and, together with eIF1A, produce a stable assembly that can resist displacement by a 60S subunit prior to mRNA binding. This would be consistent with previous observations that the stimulatory effect of eIF3 on TC binding to 40S subunits was greater when 60S subunits were present [35], and that binding of eIF3 itself is enhanced by simultaneous binding of the TC to 40S subunits [35–36].

7.2.2.2 Components of the eIF2/GTP/Met-tRNA$_i^{Met}$ Ternary Complex

Sequence Determinants of tRNA$_i^{Met}$ that Restrict it to the Initiation Pathway

Eukaryotic tRNA$_i^{Met}$ has sequence and structural characteristics that allow eIF2 to distinguish it from the elongator methionyl tRNA (tRNA$_e^{Met}$) and all other elongator tRNAs (reviewed in Ref. [15]). These include the A1:U72 base pair at the very end of the acceptor stem, several G:C base pairs in the anticodon stem, both of which were implicated in eIF2 binding, and (for yeast tRNA$_i^{Met}$) A54 in loop IV [39–42] (Fig. 7.2-4A). The A1:U72 base pair in tRNA$_i^{Met}$ also discriminates against its activity in elongation [40, 43], as do the A50:U64 and U51:A63 base pairs in the TΨC stem of human tRNA$_i^{Met}$, and the corresponding U50:A64 base pair in yeast tRNA$_i^{Met}$, which are believed to perturb the structure of this helix in a way that blocks eEF1α binding. The tRNA$_i^{Met}$ in fungi and plants additionally contains a unique 2′-*O*-phosphoribosyl modification of A64 in the TΨC helix that prevents elongator function [44, 45] and impedes binding to eEF1α-GTP [46]. Thus, structural perturbation of the TΨC stem seems to be a common strategy to block tRNA$_i^{Met}$ binding to eEF1α [43]. Inactivation of the yeast enzyme responsible for A64 modification (encoded by *RIT1*) showed that the modification is dispensable for initiator function and serves to block its activity in elongation [45]. This activity of RIT1 becomes essential in strains with mutations in eIF2 subunits or lacking a full complement of the *IMT* genes encoding tRNA$_i^{Met}$ [47]. The methionyl group attached to charged Met-tRNA$_i^{Met}$ may also increase the efficiency of translation initiation, as initiators charged with certain other amino acids function poorly in initiation [48, 49]. For more information about tRNA structure and modifications refer to Chap. 4.1.

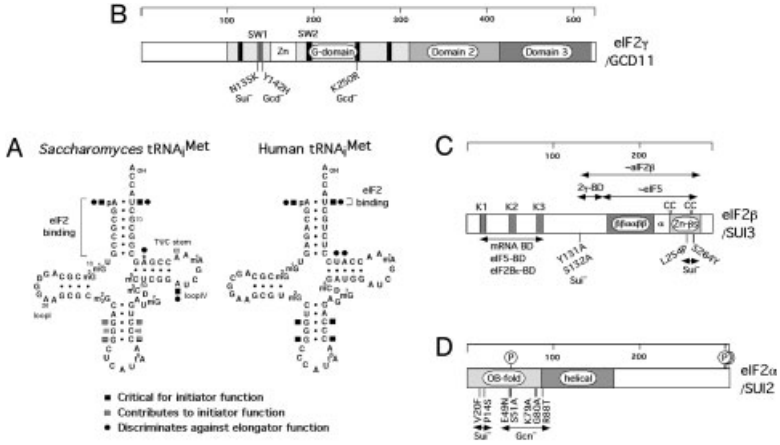


Figure 7.2-4 Schematic representations of the secondary structures of yeast and human initiator tRNA^{Met} and the primary structures of the subunits of yeast eIF2. **(A)** The sequences of the tRNAs and identities of modified bases are found in Ref. [395]. The asterisk at position 64 of yeast initiator designates the phosphoribosyl group attached to the ribose 2′-OH. See text for details. The numbering of bases shown for the yeast initiator also applies to human initiator. This figure has been adapted from RajBhandary and Chow [396] and is reprinted with permission from Ref. [15]. **(B)** The amino acid sequence of the γ subunit of eIF2 is depicted as a rectangle with amino acid positions shown above. Shading is used to depict the G-domain, and domains 2 and 3, all defined by the crystal structure of aIF2γ shown in Fig. 7.2-5(A). The four conserved motifs characteristic of G proteins are shown as black bars in the G-domain, one of which coincides with the switch 2 element (SW2). The locations and phenotypes of selected mutations are shown beneath the schematic using the one-letter code for amino acids. The abbreviation for the wild-type residue is followed by the position of the residue in the protein sequence and then the abbreviation for the substituting residue in the mutant. **(C)** The amino acid sequence of the β subunit of eIF2 is depicted with shading used to identify the lysine boxes (K1–K3) and the domains labeled $\beta\beta\alpha\alpha\beta\beta$ and Zn-βs (for Zn-binding β-sheet) whose 3D structures can be predicted from that of aIF2β shown in Fig. 7.2-5B. The four cysteine (C) residues that comprise the Zn-binding domain are indicated. Regions of similarity to eIF5 (~eIF5) or aIF2β (~aIF2β) are delimited with double-headed arrows, as are binding domains (BDs) for various factors or mRNA. The locations and phenotypes of selected mutations are shown beneath the schematic. **(D)** The amino acid sequence of the α subunit of eIF2 is depicted with shading used to identify the domains labeled OB-fold and helical whose 3D structures are shown in Fig. 7.2-5(C). The locations of the phosphorylation site for the kinases GCN2, HRI, PERK/PEK, and PKR at position 51, and the three casein kinase sites at the C-terminus are indicated with Ps. The locations and phenotypes of selected mutations are shown beneath the schematic. See text for more details.

eIF2γ Plays a Central Role in Binding Guanine Nucleotides and Initiator tRNA

The eIF2γ belongs to the superfamily of GTP-binding proteins and is closely related to the bacterial and eukaryotic elongation factors, EF-Tu and eEF1α, respectively, which deliver charged elongator tRNAs in ternary complexes with GTP to the ribosome during the elongation phase of protein synthesis. The molecular masses and sequences of eIF2γ proteins are well conserved among animals, plants, and fungi (Table 7.2-1), and orthologs also exist in archaea (Fig. 7.2-2). The sequence similarities between eukaryotic and archaeal eIF2γ proteins and EF-Tu extend throughout the G domain (domain 1), and into domains 2 and 3 [20, 50, 51] (Fig. 7.2-4B), consistent with the occurrence of binding sites for guanine nucleotides and Met-tRNA$_i^{Met}$ in eIF2γ. The crystal structure of the archaeal ortholog of eIF2γ from *P. abyssi* (Fig. 7.2-5A) shows three domains highly similar to domains 1–3 of EF-Tu, and the binding pocket for GDP-Mg^{2+} seen in the structure of the aIF2γ–GDP complex is superimposable on that of EF-Tu. Consistently, the *gcd11–K250R* mutation in yeast eIF2γ, which is predicted to alter the conserved Lys residue in the third consensus motif of the GDP-binding pocket [51] (Figs. 7.2-4B and 7.2-5A), increased the off-rates for GDP and GTP, without affecting Met-tRNA$_i^{Met}$ binding to purified eIF2. The eIF2–GTP complex is stabilized by Met-tRNA$_i^{Met}$, and addition of Met-tRNA$_i^{Met}$ overcame the GTP-binding defect of the *gcd11-K250R* lesion *in vitro* and suppressed its Slg$^-$ and Gcd$^-$ phenotypes *in vivo* [52]. Thus, there is little doubt that eIF2γ directly binds GTP.

In EF-Tu, the relative orientation of domain 1 versus domains 2 and 3 varies dramatically between the GDP-bound (inactive) and GDPNP-bound (active) states as a result of altered conformations of the switch-1 and switch-2 regions, which contact the γ-phosphate of GDPNP. In sharp contrast, the unliganded, GDP- and GDPNP-bound forms of aIF2γ all display close packing of domains 2–3 against domain 1 in the manner observed for GDPNP-bound EF-Tu. There is no difference in switch 1, and only a small conformational change in switch 2, between the GDP- and GDPNP-bound states, apparently because contacts with the γ-phosphate of GDPNP are lacking in the aIF2γ crystal structure. Therefore, it is difficult at present to account for the GTP requirement for Met-tRNA$_i^{Met}$ binding to eIF2 [51]. (It should be noted that the aIF2γ–GTP crystal structure was obtained for a mutant protein containing the G235D mutation in strand β8 of domain 2.)

A model of Met-tRNA$_i^{Met}$ docking on aIF2γ was constructed by superimposing the EF-Tu/GTP/Phe-tRNA^{Phe} complex on domains 2 and 3 of the aIF2γ–GDPNP structure. The β-hairpin in switch 1 is predicted to contact the acceptor stem and interact with the critical A1:U72 base pair in Met-tRNA$_i^{Met}$ [51]. Consistent with this model, the *gcd11-Y142H* mutation in yeast eIF2γ [24, 27], predicted to impair β-strand 2 in switch 1 [51] (Figs. 7.2-4B and 7.2-5A), produces Gcd$^-$ and Slg$^-$ phenotypes and a reduced polysome content, and is suppressed by overproducing tRNA$_i^{Met}$. Consistently, purified eIF2 containing the *gcd11-Y142H* subunit shows reduced Met-tRNA$_i^{Met}$ binding but normal off-rates for GDP and GTP [52]. The *N135K* mutation

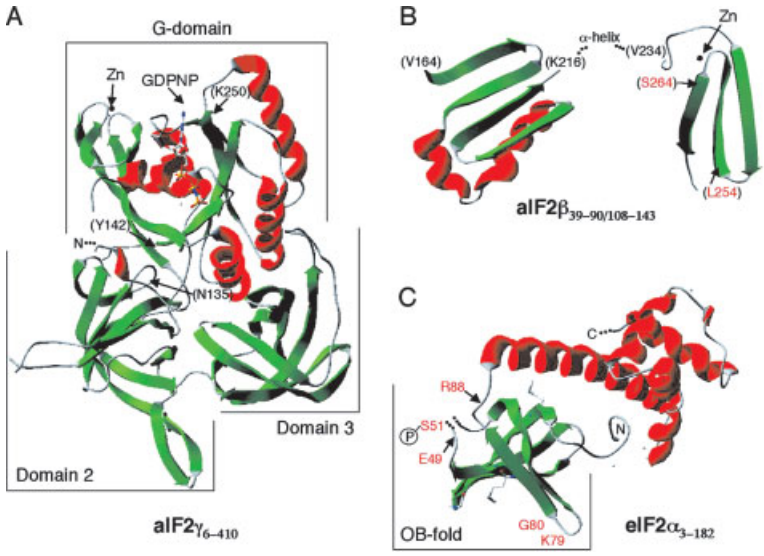


Figure 7.2-5 Ribbon diagram representations of the 3D structures of portions of aIF2γ, aIF2β, and eIF2α. **(A)** Co-crystal structure of residues 6–410 of aIF2γ from *P. abyssi* (shown with β-sheets in green and α-helices in red) in complex with GDPNP (shown in ball and stick representation) [51] (PDB ID: 1KK1). The predicted locations of selected residues in yeast eIF2γ (in parantheses) are indicated, as is the bound Zn atom and location of N-terminal residues not visualized in the structure (N...). **(B)** Solution structure of two domains of aIF2β from *M. jannaschii* joined by a predicted flexible α-helix, with a Zn atom bound to the C-terminal β-sheet domain [58] (PDB ID: 1K8B, and 1K81). The relative orientation of the two domains is unknown. The residues in yeast eIF2β corresponding to the N- and C-termini of the domains are shown in black with in parantheses. The predicted positions of two residues in yeast eIF2β which produce Sui$^-$ phenotypes when mutated are shown in red. **(C)** The crystal structure of residues 3–182 of human eIF2α [82] (PDB ID: 1KL9). The phosphorylation site at Ser^{51} is indicated in red as are the positions of residues in yeast eIF2α that produce Sui$^-$ phenotypes when mutated. All structures were drawn using the DeepView/Swiss-Pdb viewer (v. 3.7) using data obtained from the Protein Data Bank (www.pdb.org). For the NMR structures, the first of multiple solved structures stored in the PDB file was employed.

in yeast eIF2γ, isolated for its dominant Sui$^-$ phenotype [53], maps in the predicted β-hairpin of switch 1 [51] (Figs. 7.2-4B and 7.2-5A). *In vitro*, this lesion reduced TC formation, partly by increasing spontaneous GTP hydrolysis, but also by increasing the off-rate of Met-tRNA$_i^{Met}$ from eIF2 without affecting the affinity for GTP. Thus, there is strong evidence implicating switch 1 of yeast eIF2γ in Met-tRNA$_i^{Met}$ binding.

To account for the dominant Sui⁻ phenotype of this mutation, it was proposed that premature dissociation of Met-tRNA$_i^{Met}$ from the mutant eIF2–GTP complex during the scanning process allows incorrect pairing of the initiator with a near-cognate UUG codon [53].

The e/aIF2γ proteins contain some structural features not present in EF-Tu, including a disordered loop and β-hairpin in domain 2, and a zinc-binding "knuckle" containing four Cys residues appended to domain 1 [51] (Zn in Fig. 7.2-4B). Mutational analysis of yeast eIF2γ is consistent with the possibility that zinc binding to this domain is important, but not essential, for some aspect of eIF2 function [54]. However, there is no direct evidence for zinc binding by yeast eIF2γ. Moreover, only one of the four Cys residues is conserved in mammalian eIF2γ, making zinc-binding improbable for this protein. The e/aIF2γ proteins lack several residues in EF-Tu that help to clamp the 5′ phosphate group of the different elongator tRNAs [51]. Crosslinking and affinity-labeling experiments indicated that both the β and γ subunits of eIF2 are in close proximity to GTP and Met-tRNA$_i^{Met}$ in the TC [50, 55]. Moreover, an eIF2$\alpha\gamma$ dimer could bind GDP but was unable to form a stable TC with Met-tRNA$_i^{Met}$ [56]. These and other findings discussed below suggest that the β subunit contributes to Met-tRNA$_i^{Met}$ binding. The α and β subunits of both yeast and archaeal eIF2 interact directly with the γ subunit, but not with each other [51, 57], consistent with the notion that γ is the core subunit and that its functions in binding guanine nucleotides and Met-tRNA$_i^{Met}$, and in GTP hydrolysis, are augmented or regulated by the α and β subunits of eIF2.

eIF2β: Interactions with Met-tRNA$_i^{Met}$, mRNA, eIF5, and eIF3

eIF2β can be divided into three structural domains (Fig. 7.2-4C). The C-terminal half is closely related in sequence to the archaeal ortholog (aIF2β), and most probably has a two-domain structure similar to that solved by NMR for *Methanococcus Jannaschii* aIF2β [58] (Fig. 7.2-5B). The first domain in the latter consists of a four-stranded β-sheet with two helices packed against one face of the β-sheet. It is connected by an α-helical linker to the second domain, comprised of a three-stranded β-sheet with two CXXC clusters that form a Zn^{2+}-binding pocket at one end of the β-sheet. Both domains in aIF2β appear to be structurally independent units, and the C-terminal β-sheet is stabilized by Zn^{2+}. The N-terminus of eukaryotic eIF2β has an additional ~130 residues, not found in the archaeal orthologs, which contains three polylysine stretches (K-boxes 1–3) [59–61] (Fig. 7.2-4C).

Mutational analysis of yeast eIF2β (encoded by *SUI3*) shows that the Cys residues in the Zn^{2+}-binding pocket are critically required for eIF2β function *in vivo* [62]. A *SUI3* allele lacking the zinc-finger motif cannot support viability and has a dominant Gcd⁻ phenotype in otherwise wild-type cells, suggesting that the mutant protein forms an eIF2 complex defective for TC formation or 40S binding *in vivo*. Remarkably, all 13 dominant Sui⁻ alleles of *SUI3* alter conserved residues [59, 62] predicted to lie at one end or the other of the C-terminal β-sheet, in most cases near or within the loops connecting β-strands [58]. Biochemical analysis showed that two

such Sui⁻ mutations (*S264Y* and *L254P*) (Figs. 7.2-4C and 7.2-5B) increased GTPase activity by the purified TC, independent of the GAP function of eIF5 [53]. The *S264Y* mutation also led to increased dissociation of Met-tRNA$_i^{Met}$ from the TC independent of GTP hydrolysis, supporting a role for the β subunit in Met-tRNA$_i^{Met}$ binding. It was proposed that both defects increase the probability that the TC can dissociate during the scanning process and leave Met-tRNA$_i^{Met}$ inappropriately paired with a UUG start codon [53].

A segment of yeast eIF2β that is necessary and sufficient for eIF2γ binding was localized to residues 128–159 in eIF2β [63], just N-terminal to the region homologous to the well-defined structural domains in aIF2β [58] (2γ-BD (binding domain) in Fig. 7.2-4C). Alanine substitutions of the highly conserved Tyr^{131} and Ser^{132} residues in this region of yeast eIF2β (Fig. 7.2-4C) abolished *in vitro* binding to eIF2γ and impaired interaction of native eIF2β with the eIF2$\gamma\alpha$ dimer *in vivo*. The *SUI3-YS* allele containing both of these substitutions conferred Ts⁻ and Sui⁻ (or possibly Gcd) phenotypes and was synthetic lethal with the Sui⁻ *SUI3-S264Y* allele. Thus, by weakening β–γ interaction, the *SUI3-YS* mutation may exacerbate the hyperactive GTPase function of eIF2γ conferred by *S264Y*, reduce binding of Met-tRNA$_i^{Met}$ to eIF2, or both [63]. Interestingly, the C-terminal half of eIF2β shows strong similarity to the N-terminal portion of eIF5 [64] (~eIF5 in Fig. 7.2-4C), including the two CXXC clusters, raising the possibility that the homologous domains in eIF5 and eIF2β interact with one another, or compete for an interaction with the γ subunit, in a way that stimulates GTP hydrolysis by eIF2. It should be noted, however, that eIF5 lacks the major binding domain for eIF2γ between residues 128 and 159 in eIF2β [57, 63].

There are numerous reports that eIF2 binds mRNA and that this interaction can impede TC formation (reviewed in Ref. [55]) or stimulate translation [13, 65]. The β subunit has mRNA-binding activity [66] and seems to be required for mRNA binding by the eIF2 complex. A 4-thio-UTP-substituted viral mRNA was crosslinked to the C-terminal one-third of eIF2β containing the zinc-binding domain [67]; however, mutational analysis of the yeast protein suggests that the K-boxes in the NTD make an even larger contribution to mRNA binding (mRNA-BD in Fig. 7.2-4C). The third K-box was sufficient for nearly wild-type mRNA binding *in vitro*, even when altered to a run of arginine residues. Deletion of all three K-boxes was lethal, but *SUI3* alleles retaining any single K-box were viable, indicating functional redundancy for their essential function(s) *in vivo* [68].

The *SUI3* allele lacking all three K-boxes conferred dominant Slg⁻ and Gcd⁻ phenotypes, suggesting a defect in TC formation or binding to 40S ribosomes. Ostensibly at odds with this interpretation, the triple K-box mutations had no effect on TC formation by purified eIF2 *in vitro*. Moreover, eIF2 containing the β subunit lacking the K-boxes was found in 43S or 48S preinitiation complexes in yeast cells, indicating that the K-boxes are dispensable for TC formation and 40S binding *in vivo* [68]. Nevertheless, the dominant Gcd⁻ phenotype of this *SUI3* allele may signify a reduced rate of TC binding to 40S subunits that is sufficient to derepress *GCN4* translation because of the kinetic restrictions on re-initiation on *GCN4* mRNA. In fact, the K-boxes stabilize interaction between recombinant eIF2β or eIF2 holoprotein with the

catalytic subunit of eIF2B (ε/GCD6) *in vitro*, and all of the viable single and double K-box mutations in *SUI3* had a Gcd$^-$ phenotype. Thus, the K-box mutations most probably impede the recycling of eIF2-GDP to eIF2-GTP by eIF2B and diminish the rate of TC formation *in vivo* [69].

The K-boxes in eIF2β promote interaction between eIF2 and eIF5 in addition to their roles in mRNA binding and interaction with eIF2B (eIF5-BD in Fig. 7.2-4C). Mammalian eIF5 copurifies with eIF2 from lysates, and a 1:1 complex was formed *in vitro* with purified eIF2 and recombinant eIF5 [70]. Mammalian eIF5 binds specifically to eIF2β *in vitro*, dependent on the second K-box [64]. Mutational analysis of yeast eIF2β showed that at least one K-box was required for interaction with yeast eIF5, and that the K-boxes had additive effects on eIF5 binding *in vitro*. Similarly, K-boxes 1 or 3 were sufficient for association of eIF5 and eIF2β *in vivo*, but at levels approximately one-third of that seen with all K-boxes intact. As with mammalian eIF2β, the C-terminal half of yeast eIF2β (related in sequence to eIF5) contributed little to its interaction with eIF5 [69].

Interestingly, the K-box domain in eIF2β promotes binding to eIF5 and eIF2Bε through interactions with a conserved bipartite motif found at the C-termini of both proteins, dubbed AA-boxes 1 and 2 for the conserved aromatic and acidic residues they contain (Figs. 7.2-6A and 7.2-8D). Alanine substitutions of multiple residues in AA-boxes 1 or 2 of yeast eIF5 (12A and 7A, respectively, in Fig. 7.2-8D) impaired its interaction with recombinant eIF2β-NTD and purified eIF2 holoprotein *in vitro*. The *tif5-7A* allele (harboring the seven Ala mutations in AA-box2) likewise abolished native eIF5-eIF2 interaction and conferred a Ts$^-$ phenotype in yeast that was partially suppressed by overexpressing all three subunits of eIF2 and $tRNA_i^{Met}$. The *tif5-12A* allele (bearing the 12 Ala replacements in AA-box 1) is lethal [69]. Thus, the AA-boxes in eIF5 mediate an important interaction with TC *in vivo* that may facilitate the GAP function of eIF5 on base pairing of Met-$tRNA_i^{Met}$ with the start codon. Consistent with this idea, mutations in the AA-boxes of mammalian eIF5, which impair its interaction with eIF2β reduced the GAP activity of eIF5 *in vitro* and the eIF5-dependent formation of 80S initiation complexes [71] (M3 and M4 mutations in Fig. 7.2-8D). Surprisingly, a reduction in GAP activity *in vitro* was not observed in response to the more extensive mutations in the yeast eIF5 AA-boxes of *tif5-7A* and *tif5-12A* [72], possibly indicating that substrate binding is not rate-limiting in the model GAP assay established for yeast eIF5 [53]. As discussed below, the *tif5-7A* mutation also destabilizes a physical interaction between eIF2 and eIF3 that is bridged by the eIF5-CTD, impairing the binding of TC to 40S subunits *in vitro* and possibly impeding scanning or AUG recognition *in vivo* [34, 72].

The corresponding *12A* and *7A* mutations in the AA-boxes of eIF2Bε/GCD6 reduced its binding to the eIF2β–NTD and eIF2 holoprotein *in vitro*, just as observed for eIF5. Moreover, the corresponding *gcd6-7A* mutation reduced association between native eIF2 and eIF2B *in vivo* and conferred a Gcd$^-$ phenotype that could be suppressed by overexpressing eIF2 and initiator $tRNA^{Met}$, all consistent with a reduction in GDP–GTP exchange on eIF2. The *gcd6-12A* allele, bearing substitutions in the first AA-box was lethal, suggesting that the bipartite motif in GCD6

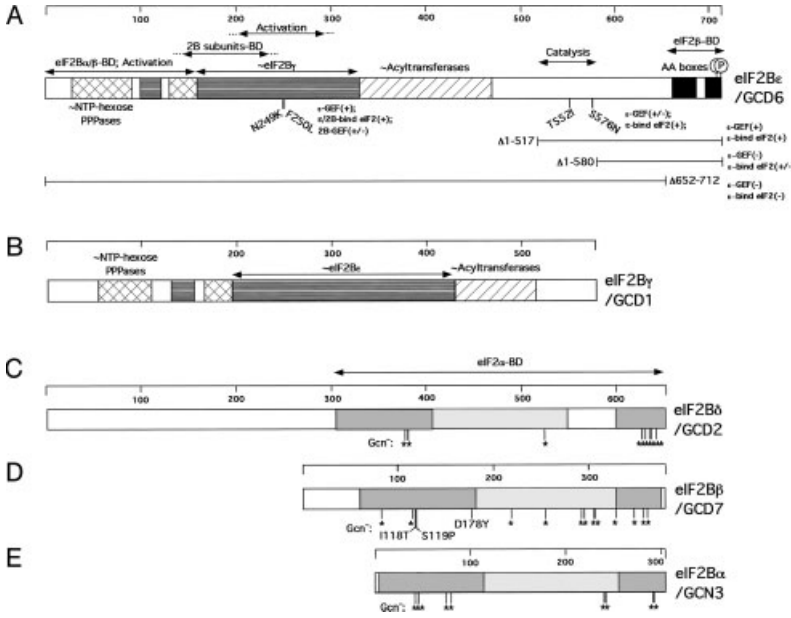


Figure 7.2-6 Schematic representations of the primary structures of the subunits of yeast eIF2B. **(A)** The amino acid sequence of the ε subunit of eIF2B is depicted as a rectangle with amino acid positions shown above. Hatching is used to depict domains with sequence similarity to the indicated proteins; the AA-boxes are shown as black boxes at the C-terminus. The boundaries of binding domains (BDs) for other factors or regions required for catalysis or activation of catalytic function (activation) are delimited with double-headed arrows. The locations and *in vitro* phenotypes of selected point mutations and deletions are shown beneath the schematic. *ε-GEF(+)* denotes wild-type GEF activity conferred by the isolated ε-subunit; *2B-GEF(+/–)* denotes reduced GEF activity conferred by the eIF2B holoprotein; *ε-* or *2B-bind eIF2(+)* denotes wild-type binding of eIF2 by the isolated ε-subunit or eIF2B holoprotein, respectively. **(B)** The amino acid sequence of the γ subunit of eIF2B depicted as for eIF2Bε in (A). **(C–E)** The amino acid sequences of the δ, β, and α subunits, respectively, of eIF2B with shading used to depict regions of similarity among the three proteins. The BD for eIF2α is delimited with a double-headed arrow. The locations of point mutations with Gcn⁻ phenotypes are shown beneath the schematic indicated with asterisks or the amino acid replacements. See text for further details.

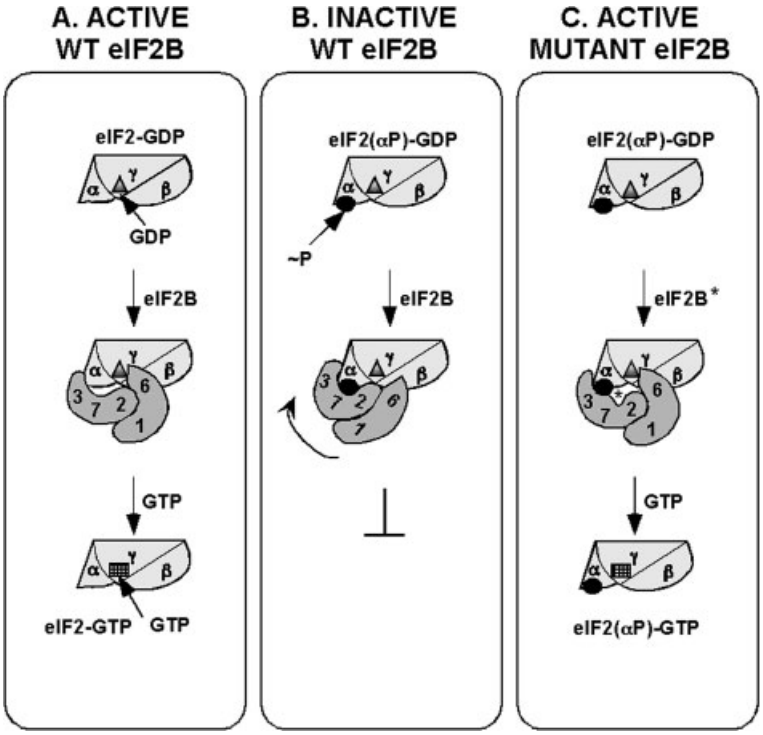


Figure 7.2-7 A mechanistic model for negative regulation of the guanine nucleotide exchange activity of eIF2B by eIF2(αP). **(A)** Unphosphorylated eIF2α promotes the GDP/GTP exchange activity of eIF2B. The heterotrimeric eIF2 (shown as different shapes labeled α, β, γ) complexed with GDP (shaded triangle) has two binding sites in eIF2B. The GCD2/δ-GCD7/β-GCN3/α regulatory subcomplex in eIF2B (labeled 2, 3, 7) binds to the α subunit of eIF2, whereas the GCD1/γ–GCD6/ε catalytic subcomplex in eIF2B (labeled 1, 6) interacts with the β and γ subunits of eIF2. Based on results with rat proteins, the GCD2 (δ) subunit of eIF2B may also interact with eIF2β. The binding interactions shown here position the catalytic subunit of eIF2B (GCD6/ε) in proximity to the bound GDP in the manner required to catalyze exchange of GDP for GTP (hatched rectangle) on eIF2. **(B)** Phosphorylated eIF2 inhibits the GDP/GTP exchange activity of wild-type eIF2B. Phosphorylation of eIF2α (●, labeled ~P in eIF2(αP)-GDP) leads to more extensive interactions between eIF2α and the eIF2B regulatory subcomplex, preventing productive interactions between GCD6/ε and the β/γ subunits of eIF2, thereby inhibiting nucleotide exchange. **(C)** A Gcn$^-$ mutation in the GCD7/β regulatory subunit of eIF2B weakens interaction between eIF2α(P) and the regulatory sub-complex of the mutant eIF2B complex (eIF2B*), permitting the productive interaction between GCD6/ε and eIF2(αP)-GDP necessary for GDP–GTP exchange. Reproduced from Ref. 84.

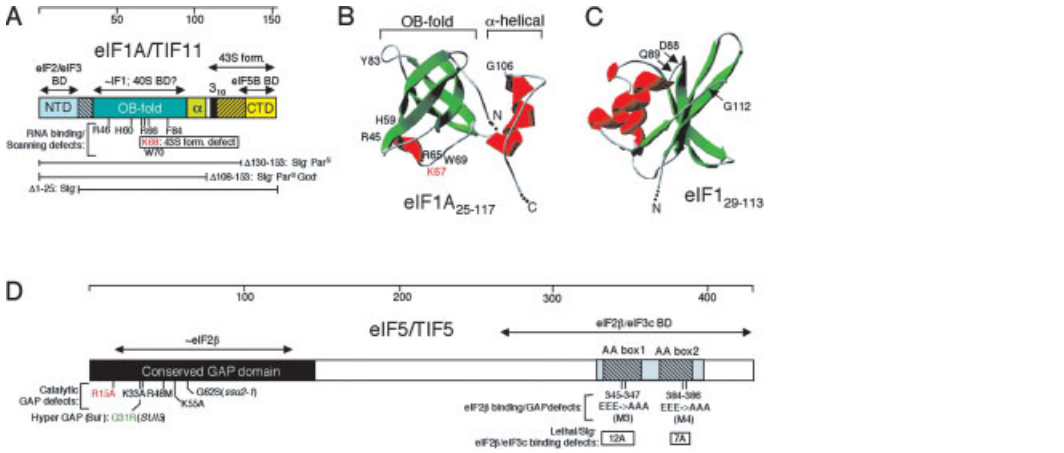


Figure 7.2-8 Schematic representations of the primary and tertiary structures of eIF1A, eIF1, and eIF5. **(A)** The amino acid sequence of yeast eIF1A (encoded by *TIF11*) is depicted as a rectangle with amino acid positions shown above. Different colors are used to depict the N-terminal (NTD), OB-fold, α-helical (α), 3_{10} helix, and C-terminal (CTD) domains, based on the 3D structure of the human protein shown in (B). The region of similarity to bacterial IF1 (~IF1) predicted to be the BD for the 40S ribosome, is delimited with a double-headed arrow, as are demonstrated BDs for eIF2, eIF3, and eIF5B, and the C-terminal region required for TC binding to 40S subunits and the formation of 43S complexes *in vivo* (43S form.) The locations and phenotypes of selected deletions are shown beneath the schematic, as are the residues homologous to those in human eIF1A whose substitution led to defects in RNA binding and AUG selection by the scanning 48S complex, and 43S complex formation (for K68 only), *in vitro*. **(B)** Solution structure of residues 25–117 of human eIF1A (191) (PDB ID: 1D7Q). **(C)** Solution structure of residues 29–113 of human eIF1 (PDB ID: 2IF1) [4]). The indicated residues D88, Q89, and G112, homologous to yeast residues 83, 84, and 107, respectively, give rise to Sui$^-$ (all three residues) and Mof$^-$ (G112 only) phenotypes when mutated in yeast. The structures in (B) and (C) were drawn using the DeepView/Swiss-Pdb viewer (v. 3.7) using data obtained from the Protein Data Bank (www.pdb.org). In each case, the first of multiple solved structures stored in the PDB file was employed. **(D)** The amino acid sequence of yeast eIF5 (encoded by *TIF5*) is depicted as a rectangle with amino acid positions shown above. Shading or hatching is used to depict the conserved N-terminal GAP domain and conserved CTD harboring AA-boxes 1 and 2 (hatched). The region of similarity to eIF2β is delimited with a double-headed arrow, as is the BD for eIF2β and eIF3c/NIP1. The locations and *in vitro* or *in vivo* phenotypes of selected point mutations are shown beneath the schematic.

is essential for the GEF function of eIF2B [69]. Interactions of the CTDs in eIF5 and eIF2Bε with the eIF2β–NTD appear to be mutually exclusive, as eIF2B and eIF5 are not found in the same complexes containing eIF2 in yeast cells [69]. The β subunit of archaeal eIF2 lacks the NTD containing the K-boxes and, consistently, archaea lack recognizable orthologs of eIF5 and eIF2Bε [73–75]. Thus, the K-boxes may have arisen during evolution, at least partly, to facilitate the interactions of eIF2 with the factors that regulate the status of its bound guanine nucleotide [69].

eIF2α Promotes and Regulates GDP–GTP Exchange by eIF2B

The α subunit of eIF2 (encoded by *SUI2* in yeast) contains the conserved Ser residue at position 51 whose phosphorylation converts eIF2-GDP from substrate to inhibitor of eIF2B [76, 77]. (Note that Ser-51 is actually the 52nd encoded residue in human and yeast eIF2α, as the N-terminal Met is removed posttranslationally; other residues in eIF2α also are typically numbered relative to the second encoded residue of the protein.) The sequence surrounding Ser^{51} is highly conserved in eukaryotic eIF2α proteins [78–80], but not in archaea [73], consistent with phosphorylation of this residue occurring only in eukaryotes. Interestingly, residues 14–93 in archaeal and eukaryotic eIF2α exhibit sequence similarities with the RNA-binding domain of *E. coli* ribosomal protein S1, a five-stranded antiparallel β-barrel called the OB-fold [81] (Fig. 7.2-4D). The crystal structure of the N-terminal segment of human eIF2α confirms the presence of the OB-fold in residues 1–87, with Ser-51 located in a long unstructured loop between β strands 3 and 4 (Fig. 7.2-5C). The OB domain of eIF2α lacks the clustered positively charged surface residues involved in RNA binding by other OB-fold proteins, and there is no evidence that eIF2α has RNA binding activity. Residues 88–182 comprise a helical domain that interacts with the OB domain, forming a highly conserved, negatively charged channel at the interface between the two domains [82] (Fig. 7.2-5C). Sui$^-$ mutations in yeast eIF2α alter residues in the NTD [78] (Fig. 7.2-4D). Thus, this region may contribute to Met-$tRNA_i^{Met}$ binding or an interaction with mRNA during scanning by eIF2. Other *sui2* mutations reduce the inhibitory effect of phosphorylated eIF2 on the GEF eIF2B (Gcn$^-$ phenotype) [83, 84] and alter residues in the eIF2α OB domain, including amino acids in the loop between β4 and β5 (G80 and K79), in the loop containing Ser-51 itself (E49), and in the loop connecting the OB-fold to the helical domain (R88) [82] (Fig. 7.2-5C). As discussed below, this portion of eIF2α most probably interacts with the regulatory subunits of eIF2B (α, β, and δ) and mediates inhibition of the GEF activity when Ser-51 in eIF2α is phosphorylated.

Recent studies of yeast eIF2α indicate that this subunit is dispensable for the essential functions of eIF2 in translation initiation and is required primarily to promote and regulate GDP–GTP exchange by eIF2B. While a *SUI2* deletion is lethal in otherwise wild-type cells, *sui2Δ* mutant cells can survive if eIF2[$\beta\gamma$] is overexpressed along with $tRNA_i^{Met}$. A nearly complete by-pass of eIF2α function was achieved by overexpressing the mutant protein eIF2γ-K250R along with eIF2β and $tRNA_i^{Met}$. The *K250R* mutation stimulates TC formation by eIF2-GDP *in vitro* in the absence of eIF2B by enhancing the spontaneous GDP–GTP exchange activity intrinsic to yeast eIF2. Consistently, overexpressing all three eIF2 subunits (with the K250R mutation in the γ subunit) and $tRNA_i^{Met}$ suppressed the lethality of deleting all four essential eIF2B subunits. These last findings imply that eIF2B has no essential functions beyond GDP–GTP exchange that cannot be by-passed by increasing the concentration of TC (85). In accordance with these genetic results, biochemical analysis of the eIF2[$\beta\gamma$] heterodimer showed that absence of the α subunit had no substantial effect on binding of guanine nucleotides, TC formation, binding of TC to purified 40S

subunits, or eIF5-catalyzed GTP hydrolysis by 43S complexes. The only defect observed was an 18-fold increase in the K_m of eIF2B for eIF2[$\beta\gamma$]-GDP versus eIF2[$\alpha\beta\gamma$]-GDP [86]. The latter suggests that eIF2α contributes to the binding of eIF2 by eIF2B, possibly through direct interactions with the eIF2B[$\alpha\beta\delta$] regulatory subcomplex [84, 87] (see below). In view of these findings, it is surprising that eIF2α is conserved in archaea, which lack eIF2B. Perhaps archaeal eIF2α performs a crucial function that is carried out redundantly by a eukaryotic-specific factor (e.g., eIF3) or a ribosomal protein.

Yeast eIF2α is phosphorylated *in vivo* on Ser residues 292, 294 and 301 at the extreme C-terminus (Fig. 7.2-4D). *In vitro* and *in vivo* results indicate that casein kinase II (CKII) phosphorylates one or all three residues. Whereas Ala substitutions of these residues did not confer any growth or Sui$^-$ phenotypes in wild-type cells, they exacerbated the growth defects of mutants in which eIF2B activity was inhibited by constitutive phosphorylation of Ser-51 in eIF2α (*GCN2^c* mutant) or by Gcd$^-$ mutations in eIF2Bα (*gcn3^c*) or eIF2Bδ (*gcd7*). Thus, lack of CKII phosphorylation reduces eIF2 activity significantly only when combined with a defect in eIF2 recycling [88]. CKII phosphorylation may promote productive interaction between eIF2-GDP and eIF2B. There is currently no evidence that this phosphorylation event is regulated in yeast cells. Mammalian eIF2α lacks the CKII sites and it is not a substrate for the mammalian kinase *in vitro* [13].

7.2.2.3 The GEF eIF2B regulates ternary complex formation

The Catalytic Function of eIF2B

1. *The mechanism of guanine nucleotide exchange*. Following recognition of the AUG codon and hydrolysis of the GTP bound to eIF2 in the TC, the resulting eIF2-GDP is released from the ribosome. At physiological Mg^{2+} concentrations, the eIF2–GDP complex dissociates slowly and the affinity of eIF2 is much greater for GDP than GTP [13]. Accordingly, the GEF eIF2B is required to displace the GDP bound to eIF2 and allow its replacement with GTP to regenerate the TC. The eIF2B contains five different subunits (α through ε), whose primary structures are well conserved between yeast and mammals (Table 7.2-1), and it occurs in a 1 : 1 complex with its substrate eIF2 in extracts [13,22].

The molecular mechanism of the exchange reaction is uncertain. Evidence supporting a substituted enzyme (ping-pong) mechanism involving a nucleotide-free eIF2B–eIF2 intermediate was presented [89]; however, it has been suggested that the high GDP concentration used in that study would have made it difficult to rule out a sequential mechanism involving a GTP–eIF2B–eIF2–GDP quaternary complex [90]. Indeed, kinetic data consistent with the sequential mechanism, have been reported for both mammalian [91] and yeast [92] eIF2B. Ostensibly at odds with the ping-pong mechanism is the fact that unlabeled GTP is required for displacement of radiolabeled GDP from eIF2 by eIF2B [91,93]. However, this would be expected for a ping-pong mechanism when eIF2B is present in catalytic amounts, as unlabeled

GTP will be needed to release eIF2B from the eIF2–eIF2B complex without reforming the starting substrate eIF2–[^{3}H]GDP. In fact, [^{3}H]GDP was released from eIF2 in the absence of GTP when stoichiometric amounts of eIF2B were employed [94]. Another observation, inconsistent with the ping-pong mechanism, namely that eIF2B cannot be displaced from eIF2 by GDP [93], also has been contradicted by the results of more recent experiments [94].

The sequential mechanism predicts that the eIF2–eIF2B complex should have two guanine nucleotide-binding sites in eIF2 and eIF2B, respectively. Dholakia and Wahba [91] reported that eIF2B binds GTP (but not GDP) with K_d of 4μM, and showed by photoaffinity labeling experiments that the β subunit of eIF2B contains a GTP-binding site. The latter was confirmed by Williams, et. al. [94], who found that eIF2Bβ (native or recombinant) can be crosslinked to GTP or ATP. Similarly, yeast eIF2B binds GTP with K_d of 1 μM [92]. The main difficulty with these last findings is that the β subunit is dispensable for GEF activity *in vitro* [87, 95]. In fact, the C-terminal ~25% of the ε/GCD6 subunit is sufficient for measurable eIF2B activity *in vitro* (see below). Manchester [96] suggested that the GTP-binding site in eIF2Bβ could increase the local concentration of the displacing nucleotide and thereby enhance the exchange reaction, effectively converting a basal ping-pong mechanism operative with eIF2Bε alone to the sequential mechanism seen for five-subunit eIF2B (holoenzyme) [91, 92]. This model seems at odds with the finding that the yeast eIF2B$\gamma\varepsilon$ binary complex and eIF2B holoenzyme were equally active [87]; however, the predicted difference in activity may be evident only at low GTP concentrations. It should also be noted that sequence motifs conserved in GTP-binding proteins do not occur in eIF2Bβ or in any other eIF2B subunit.

2. *Structures and functions of eIF2B subunits.* As indicated above, the eIF2B contains five different subunits. Recessive mutations in the yeast ε, δ, γ, and β subunits (encoded by *GCD6, GCD2, GCD1,* and *GCD7,* respectively) have Ts$^-$ and Gcd$^-$ phenotypes [19], indicative of reduced TC formation, and deleting any of these subunits is lethal. In contrast, deleting *GCN3* (encoding eIF2Bα) has a Gcn$^-$ phenotype (failure to induce *GCN4* in response to eIF2α phosphorylation) and no effect on cell growth [97]. Thus, eIF2Bα in yeast seems to be required primarily for inhibition of eIF2B by eIF2(αP). Similarly, a rat eIF2B complex devoid of the α subunit, either overexpressed in insect cell extracts or affinity-purified, had full GEF activity that was relatively insensitive to inhibition by eIF2(αP) [95, 98]. In contrast, a rabbit eIF2B complex lacking the α subunit did not co-purify with eIF2 and had only 20–25% of the activity of the five-subunit complex. Nearly full activity was recovered by adding recombinant eIF2Bα to the latter four-subunit preparation, leading to the conclusion that eIF2Bα is required for wild-type activity of rabbit eIF2B, perhaps by promoting substrate binding [94, 99].

Although four of the five subunits of yeast eIF2B are essential, the intrinsic GEF activity is lodged in the C-terminal ~25% of the ε/GCD6 subunit (Fig. 7.2-6A). The eIF2Bε from rat [95], *Drosophila* [100], and yeast [87, 101] can catalyze nucleotide exchange independently of the other subunits *in vitro*, albeit with 10–40-fold lower specific activity than that of eIF2B holoenzyme. In fact, a fragment of ε/GCD6

containing only the last 195 residues was comparable with full-length ε/GCD6 for eIF2 binding and GEF activity. Deletion of only the C-terminal 60 residues of ε/GCD6, containing the AA-boxes 1 and 2 mentioned above, destroyed eIF2 binding and GEF function by the isolated ε/GCD6 subunit, and greatly reduced the activity of the yeast eIF2B holoprotein (Fig. 7.2-6A) [101, 102]. Consistently, two serine residues in this segment of mammalian eIF2Bε S712 and S713 are phosphorylated *in vivo* and their replacement with nonphosphorylatable residues reduced both interaction of eIF2B with eIF2 and GEF activity in cell extracts. These sites are phosphorylated by casein kinase II *in vitro*, but it is unknown whether their phosphorylation is regulated as a means of controlling eIF2B activity *in vivo* [103]. A region N-terminal to the AA-boxes in ε/GCD6, between residues 518 and 581, is required for GEF activity but nonessential for eIF2 binding (Fig. 7.2-6A) [101, 102]. Hence, this region is predicted to contain the catalytic center in ε/GCD6, presumably responsible for distorting the GDP-binding pocket in eIF2γ to effect release of the bound GDP.

What are the functions of the other eIF2B subunits? As described above, mutations or deletions in the AA-boxes of ε/GCD6 do not abolish the activity of eIF2B holoprotein; hence, there must be additional contacts between eIF2B and eIF2 subunits. In fact, both the ε and δ subunits of eIF2B can interact with the C-terminal portion of mammalian eIF2β [104] and, as discussed below, the α, β, and δ eIF2B subunits form a stable subcomplex that can bind eIF2α. The yeast ε/GCD6–γ/GCD1 subcomplex has higher GEF activity than does ε/GCD6 alone, comparable with the eIF2B holoprotein [87, 95], and the stimulatory effect of γ/GCD1 is attributable partly to enhanced binding of eIF2 [87]. The ε and γ subunits have recognizable sequence similarity to one another and to NTP-hexose-pyrophosphorylases and acyltransferases [105] (Figs. 7.2-6A and B), but these similarities are of unknown significance. Point mutations in a highly conserved Asn-Phe-Asp motif at positions 249–251 in ε/GCD6 had no effect on GEF activity of the isolated subunit, but substantially reduced the activity of eIF2B holoprotein, nearly to the level of wild-type ε/GCD6 alone (Fig. 7.2-6A). These mutations did not impair complex formation with other eIF2B subunits, or eIF2 binding by the eIF2B holoprotein; hence, they seem to abrogate a stimulatory effect of γ/GCD1 or other eIF2B subunits on the catalytic function of ε/GCD6. Consistently, the mutations lie within a region of ε/GCD6 (defined by deletions Δ93–358 and Δ144–230), which is required for complex formation with other eIF2B subunits [101]. The extreme N-terminal 158 residues of rat eIF2Bε are required for association of eIF2Bα with the rest of the eIF2B holoprotein and also seem to promote GEF activity independent of their role in maintaining eIF2Bα in the complex. This segment of rat eIF2Bε also contains a strong binding site for the eIF2Bβ subunit [106] (Fig. 7.2-6A). Thus, the eIF2Bε-NTD is involved in interactions with other eIF2B subunits that influence the efficiency of the catalytic center in the C-terminal portion of the protein.

Inhibition of eIF2B by phosphorylated eIF2

The binary complex eIF2(αP)-GDP (phosphorylated on Ser^{51}) is a poor substrate for nucleotide exchange catalyzed by eIF2B in mammals [89, 93, 98, 107], *Drosophila*

[100], and yeast [87]. This does not reflect weak substrate binding, as phosphorylation of eIF2 increases its affinity for eIF2B [13], with estimates ranging from several-fold [93, 108] to more than 100-fold [89] increased affinity. It is frequently assumed that eIF2(αP)-GDP forms a nondissociable complex with eIF2B, physically sequestering eIF2B in an inactive state. At odds with this idea, it was found that eIF2B–eIF2(αP)–GDP complexes dissociate rapidly and that eIF2(αP)–GDP acts as a competitive inhibitor of eIF2B through an enhanced on-rate or decreased off-rate compared with unphosphorylated eIF2 [89]. Because eIF2 is generally present in molar excess of eIF2B, a moderate increase in affinity for eIF2B might account for the strong inhibition of translation that occurs in mammals [12, 109] and yeast [22, 77, 110] when only a fraction of eIF2 is phosphorylated. Studies in yeast showed that the degree of translation inhibition was correlated with the eIF2(αP) : eIF2 ratio instead of the absolute amount of eIF2(αP) present in cells, consistent with a competitive mode of inhibition and a relatively high dissociation rate for the inhibited eIF2B–eIF2(αP)–GDP complex [26].

As discussed above, phosphorylation of eIF2α by GCN2 in amino acid-starved yeast cells inhibits eIF2B and lowers the concentration of TC, reducing general translation initiation but specifically increasing *GCN4* translation. The fact that *gcn3Δ* mutants are defective for this response (Gcn$^-$ phenotype) [97] suggested that eIF2Bα/GCN3 mediates the inhibitory effect of eIF2(αP) on eIF2B function. The strong sequence similarity of eIF2Bδ/GCD2 and eIF2Bβ/GCD7 to α/GCN3 [111, 112] (Figs. 7.2-6C–E) suggested that δ/GCD2 and β/GCD7 also are involved in negative regulation of eIF2B by eIF2(αP). Consistently, overexpressing these three eIF2B subunits in yeast led to formation of a stable subcomplex that can reduce the toxic effect of high-level eIF2(αP) on cell growth [113] and can bind to purified eIF2 *in vitro* in a manner stimulated by eIF2α phosphorylation [87]. Hence, these authors proposed that the overexpressed subcomplex binds preferentially to eIF2(αP)–GDP and prevents it from interfering with the ability of endogenous eIF2B holoprotein to recycle the unphosphorylated eIF2-GDP. All three eIF2B subunits (α,β, and δ) are required for binding to eIF2(αP) [87]. Subsequently, it was shown that the eIF2B α/β/δ regulatory subcomplex, but not the individual subunits, can also bind recombinant eIF2α/SUI2 *in vitro*, dependent on phosphorylation of the latter at Ser^{51}. Thus, the eIF2B regulatory subcomplex directly interacts with eIF2α in a manner stabilized by phosphorylation of Ser^{51} [84]. There is genetic and biochemical evidence that the C-terminal portion of eIF2Bδ/GCD2, which is related in sequence to β/GCD7 and α/GCN3, is sufficient for complex formation with the latter two subunits *in vivo* [113]. Hence, a heterotrimeric structure comprised of the homologous segments of eIF2B α/β/δ is thought to be the binding domain for the phosphorylated NTD of eIF2α (eIF2α-BD in Figs. 7.2-6C, D and 7.2-7).

Additional genetic evidence implicating δ/GCD2 and β/GCD7 in negative regulation of eIF2B came from isolation of Gcn$^-$ point mutations in these two subunits that relieve the inhibitory effects of eIF2(αP) on translation without impairing GEF function, mimicking in both respects a *gcn3Δ* mutant [108, 114]. As these *GCD2* and *GCD7* mutations do not simply cause α/GCN3 to be lost from eIF2B, it appears that

δ/GCD2 and β/GCD7 act directly in the regulation of eIF2B [108]. The *GCD2* and *GCD7* mutations could decrease the affinity of eIF2B for eIF2(αP) or, alternatively, allow eIF2B to accept eIF2(αP)-GDP as a substrate. The latter mechanism is favored by the fact that nearly all eIF2α was phosphorylated in certain of these mutants [108], and was later confirmed biochemically for the *GCD7-S119P* and *GCD7-I118T, -D178Y* mutations in eIF2Bβ (Fig. 7.2-6D) and for the four-subunit complex lacking GCN3 (i.e., the *gcn3Δ* mutation), as follows. All three mutant eIF2B holoproteins were shown to catalyze nucleotide exchange at high levels using either phosphorylated or unphosphorylated eIF2-GDP as substrate, as did the ε/GCD6–γ/GCD1 catalytic subcomplex [87]. Based on these findings, it was proposed that the eIF2B $\alpha/\beta/\delta$ regulatory subcomplex is required to inhibit the ε/γ catalytic subcomplex when the substrate is phosphorylated. Tight binding of phosphorylated eIF2α to the eIF2B $\alpha/\beta/\delta$ subcomplex would prevent the productive interaction between the eIF2B ε/γ catalytic subcomplex and eIF2 β/γ, which is required for release of GDP from the latter (Fig. 7.2-7). Support for this model came from the fact that the Gcn$^-$ mutations *GCD7-S119P* and *GCD7-I118T,-D178Y* in eIF2Bβ decreased binding of the eIF2B $\alpha/\beta/\delta$ subcomplex, and also of eIF2B holoprotein, to phosphorylated recombinant eIF2α (Fig. 7.2-6D). These mutations also decreased interaction between the eIF2B and eIF2 holoproteins, even when the latter was unphosphorylated. Thus, contacts between the eIF2B $\alpha/\beta/\delta$ subcomplex and eIF2α probably contribute to the productive interaction of eIF2B with nonphosphorylated eIF2-GDP [84]. This is consistent with results showing that the K_m value of eIF2B for eIF2-GDP increased by an order of magnitude when the α subunit of eIF2 was missing [86]. Presumably, the presence of a phosphate group at Ser^{51} provides additional contacts with the eIF2B $\alpha/\beta/\delta$ subcomplex that interfere with the correct interaction between the eIF2B ε/γ subcomplex and the GDP-binding pocket in eIF2 β/γ (Fig. 7.2-7).

Two mutations were introduced into rat eIF2Bδ identical to substitutions in δ/GCD2 that individually rendered yeast eIF2B insensitive to eIF2(αP) *in vivo* (Gcn$^-$ phenotype) [108]. The rat eIF2B bearing the G377K, L381Q double substitution in the δ-subunit (eIF2B[δ^*]) was only minimally inhibited by preincubation with eIF2(αP), similar to what occurred with the four-subunit eIF2B lacking the α-subunit. Unlike the latter, however, the eIF2B(δ^*) complex was completely ineffective using eIF2(αP)-GDP as a substrate. Presumably, the eIF2B(δ^*) complex escapes inhibition primarily because it binds the phosphorylated inhibitor less tightly than the unphosphorylated substrate [98].

Most of the Gcn$^-$ mutations in eIF2B fall into two clusters located in regions of strong sequence similarity among the α/GCN3, β/GCD7 and δ/GCD2 subunits (Figs. 7.2-6C–E), leading to the suggestion that the structurally homologous segments in these subunits interact to form a binding pocket for the phosphorylated NTD of eIF2α [108]. As noted above, Gcn$^-$ point mutations were also isolated in the NTD of yeast eIF2α that eliminate the inhibitory effect of eIF2(αP)-GDP on eIF2B activity [83] (Fig. 7.2-4D). Consistently, a number of these mutations weaken binding of recombinant phosphorylated eIF2α to the eIF2B $\alpha/\beta/\delta$ subcomplex or eIF2B holoprotein [84], suggesting that at least some portion of the OB domain in eIF2α

(Fig. 7.2-5C) binds directly to the eIF2B $\alpha/\beta/\delta$ subcomplex (Fig. 7.2-7). The Ala substitution of Ser-48 has the same phenotype in mammalian cells when eIF2α is phosphorylated by PKR, HRI, or in response to heat shock [115–120]. Interestingly, addition of eIF2α-S48A to inhibited RRL reduced the abundance of 15S complexes containing eIF2, thought to represent inactive eIF2B-eIF2(αP)–GDP complexes [121]. This last finding supports the idea that Ala-48 reduces the affinity of eIF2(αP)–GDP for eIF2B [120]. Consistently, the Ala-48 mutation in yeast eIF2α partially suppressed growth inhibition by hyperactive GCN2c kinases without lowering Ser51 phosphorylation [77].

Interestingly, mutations in each of the five subunits of eIF2B have been associated with the human genetic disease leukoencephalopathy with vanishing white matter (VWM) [122, 123]. It is unknown whether these mutations lead to defects in eIF2B function or its regulation.

Additional Functions for eIF2B?

It was found that *gcd1* and *gcd2* mutations in yeast eIF2B subunits led to accumulation of eIF2 in 43–48S complexes [22, 23], implying that initiation was blocked at a step in the pathway following TC binding to the 40S subunit, rather than at TC formation. Similarly, in rabbit reticulocyte lysates (RRL) inhibited by eIF2α phosphorylation, the eIF2(αP) and exogenously added mRNA and $tRNA_i^{Met}$ accumulated in 48S complexes [124]. Other workers observed accumulation of 48S complexes and halfmer polysomes containing Met-$tRNA_i^{Met}$ in inhibited RRL that could be reversed by exogenous eIF2B. Because the 48S complexes lacked eIF2 and halfmers did not appear immediately, it was proposed that 80S initiation complexes could not proceed to elongation and dissociated into mRNA-bound 40S subunits (halfmers) [125, 126]. Several groups have observed eIF2-GDP bound to 60S [121, 125–128] or 40S subunits [129], which might represent physiological intermediates in the initiation pathway. Ribosome-bound eIF2-GDP could have a positive role in subunit joining, or it could arise following GTP hydrolysis and release of eIF2-GDP from the P-site on AUG recognition. In either case, eIF2B may be required to remove eIF2-GDP from the ribosome in addition to exchanging GDP for GTP, and phosphorylation of eIF2α could convert ribosome-bound eIF2-GDP into an inhibitor of subunit joining. Interestingly, deletion of the 40S protein RPS31/UBI3 in yeast suppressed the Gcd- and Ts- phenotypes of *gcd2* and *gcd1* mutations in eIF2B, prompting the suggestion that elimination of RPS31 partially overcomes a requirement for an eIF2B function on the 40S ribosome [130]. Consistently, eIF2B accumulated in 40S complexes in the *gcd1-101* mutant [22]. It has also been proposed that eIF2B stimulates TC formation [131] and TC binding to 40S subunits in the context of an eIF2B-eIF2-GTP-Met-$tRNA_i^{Met}$ quaternary complex [132]. As noted above, elimination of eIF2B is not lethal in yeast cells that are overexpressing the TC [85]; hence, all of these putative additional functions of eIF2B would have to be by-passed in yeast by artificially increasing the TC concentration.

7.2.2.4 Binding of Ternary Complex and mRNA to the 40S Ribosome is Stimulated by eIF3

eIF3 Promotes Ternary Complex Binding to 40S Ribosomes

The TC can bind to purified 40S subunits in the absence of other factors, and this interaction is stimulated by high, nonphysiological Mg^{2+} concentrations (greater than 2 mM) and the AUG triplet [36]. (Use of AUG in place of mRNA obviates the need for factors required for mRNA binding to the ribosome.) The stimulatory effect of the AUG triplet suggests that base pairing between the start codon and Met-$tRNA_i^{Met}$ stabilizes TC association with 40S ribosomes. High-level binding of the TC to 40S subunits under more physiological conditions requires eIF1, eIF1A and eIF3 [32, 33, 37, 133–135]. TC binding to purified 40S ribosomes can be stimulated by a factor of 2–3 by the addition of purified eIF3. The eIF3 can bind to 40S ribosomes in the absence of other factors, although this association is enhanced by simultaneous binding of the TC [35–38]. The majority of native free 40S subunits in mammalian extracts contain eIF3 [136]. Based on these results, it is generally considered that eIF3 binds to 40S subunits first and then helps to recruit the TC (Fig. 7.2-1). Consistently, a Ts^- lethal mutation in the yeast eIF3b subunit (encoded by *PRT1*) produced a severe initiation defect *in vivo* [137] and heat-treated *prt1-1* extracts were defective for TC binding to 40S subunits in a manner rescued by purified wild-type eIF3 [138, 139]. Relatively little is known about how eIF3 stimulates TC binding, although the physical connections linking yeast eIF3 to eIF2 in a MFC (described below) may permit cooperative binding of both factors to adjacent sites on the 40S subunit.

A Subunit Interaction Model for eIF3

Mammalian eIF3 is a complicated factor, containing 11 nonidentical subunits (Table 7.2-2). Purified plant eIF3 contains orthologs of 10 of these proteins, lacking only eIF3j/p35, and contains an 11_{th} subunit (eIF3l/p67) not found in the mammalian factor [140]. The eIF3 purified from budding yeast contains orthologs of only five mammalian eIF3 subunits as stoichiometric components (eIF3a/TIF32, eIF3b/PRT1, eIF3c/NIP1, eIF3g/TIF35, and eIF3i/TIF34) [139], all of which are essential proteins required for translation initiation *in vivo* [140–147]. A sixth ortholog, eIF3j/HCR1, is a nonessential, substoichiometric subunit of yeast eIF3 that promotes interactions between eIF3 and other eIFs in the 43S complex, and also has an independent function in 40S ribosome biogenesis [148–150]. A possible budding yeast ortholog of human eIF3e subunit, called PCI8, was found to interact with eIF3 holoprotein *in vivo* when overexpressed, and also can bind to recombinant eIF3b/PRT1 *in vitro*, as did human eIF3e/Int-6; however, a *pci8Δ* mutation had no effect on translation *in vivo* [151]. Together, these findings suggest that the essential yeast eIF3 subunits (a, b, c, g and i) constitute a conserved core that can execute the critical functions of this factor. Consistent with this idea, the eIF3g and eIF3i subunits are also essential in fission yeast, whereas the "non-core" subunits, Moe1/d and Int6/e,

are dispensable [152-155]. Deletion of the latter two proteins reduces the stability of the eIF3 complex in fission yeast extracts [156], but produces only a modest reduction in translation rates *in vivo* [152, 153, 156]. The *moe1Δ* and *int6Δ* mutations produce an assortment of phenotypes [152, 153, 156] that could arise from reduced translation of a subset of mRNAs or from a possible involvement of these noncore subunits in other biological processes.

When purified by its ability to stimulate Met-puromycin synthesis [157], yeast eIF3 preparations contained two additional proteins of 135 and 62 kDa, subsequently identified as TIF31 [147] and GCD10 [158] in addition to the core eIF3 subunits. Affinity purification of eIF3 confirmed the association of TIF31 with the complex [144], and recombinant TIF31 interacted with TIF35/eIF3g in several assays [147]. However, *TIF31* is nonessential and its deletion has no effect on cell growth or polysome profiles [147]. It was not possible to confirm a direct association of GCD10 with eIF3 by affinity purification or co-immunoprecipitation with tagged eIF3 subunits from cell extracts [139, 159, 160]. Moreover, GCD10 resides in a nuclear complex with the product of *GCD14* that is required for the formation of 1-methyladenosine at position 58 (m^1A58) in all tRNAs containing this modification, including $tRNA_i^{Met}$ [159, 161]. It is unclear whether GCD10 contributes to eIF3 function in the cytoplasm, and the requirement for GCD10 in translational repression of *GCN4* mRNA [27] can be explained at the level of $tRNA_i^{Met}$ biogenesis and TC formation.

Pairwise interactions among the yeast core eIF3 subunits have been studied extensively by yeast two-hybrid and *in vitro* binding assays, leading to a subunit interaction model for the complex (Fig. 7.2-9A) [139, 145–147, 149]. Many aspects of this model have been confirmed and refined *in vivo* by making deletions of predicted binding domains in affinity-tagged forms of the three largest eIF3 subunits and determining the subunit compositions of the resulting subcomplexes that were affinity-purified from yeast. In the latest model, each of the three largest subunits (TIF32/a, PRT1/b, and NIP1/c) contains separate binding domains for the other two proteins, whereas the smaller subunits (TIF34/i and TIF35/g) bind only to the CTD of PRT1/b and to one another. HCR1/j binds to both PRT1/b and TIF32/a [162]. In accordance with this model, PRT1/b can form two distinct subcomplexes *in vivo*, one containing TIF32/a, PRT1/b, and NIP1/c (a/b/c), and the other comprised of PRT1/b, TIF34/i and TIF35/g (b/i/g) (Fig. 7.2-9A). Whereas the larger a/b/c subcomplex could restore 40S binding of Met-$tRNA_i^{Met}$ and mRNA, and translation of a luciferase reporter mRNA in a *prt1-1* extract, the smaller b/i/g subcomplex was relatively inert for all three activities [163]. Consistent with the subunit interaction model, expression of N-terminally truncated PRT1/b lacking the predicted RNA recognition motif RRM; *Δ*RRM, Fig. 7.2-9D) sequestered TIF34 and TIF35 in an inactive subcomplex lacking TIF32/a and NIP1/c that could not associate with ribosomes in extracts and had a dominant-negative effect on cell growth [149, 164]. Similarly, overexpression of a truncated form of PRT1/b lacking the extreme C-terminus (*Δ7*, Fig. 7.2-9D) sequestered TIF32 and NIP1 in a defective subcomplex lacking TIF34/i and TIF35/g [162]. The deleterious effect on cell growth of producing the latter subcomplex underscores the fact that TIF34/i and TIF35/g are essential *in vivo* even though they are dispensable for measurable eIF3 activity *in vitro*.

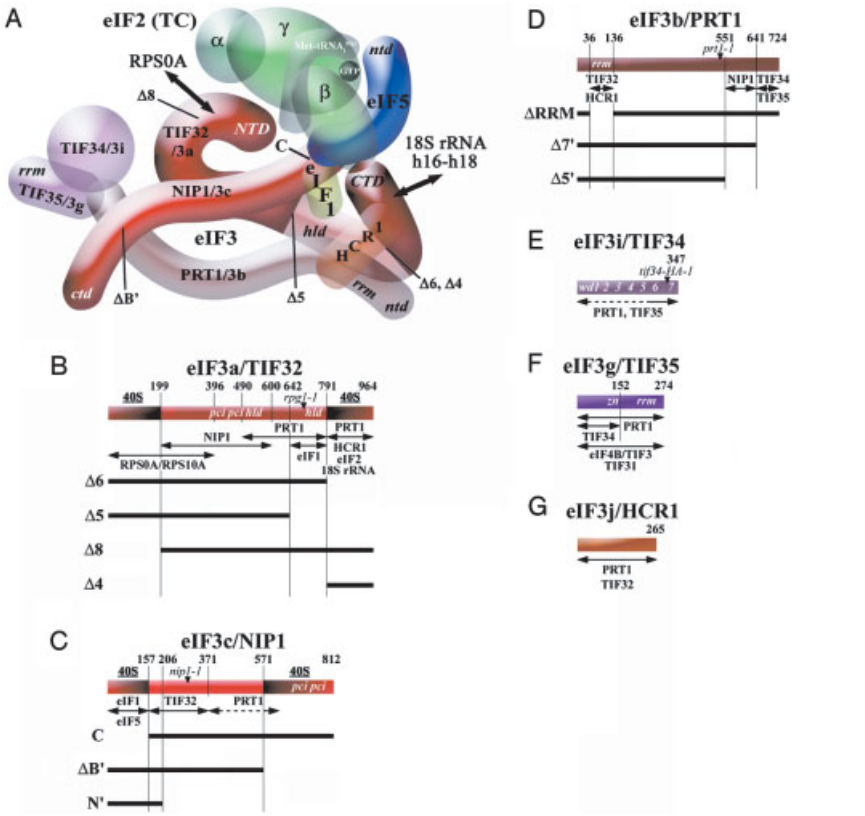


Figure 7.2-9 Schematic representations of the MFC and primary structures of eIF3 subunits of yeast. **(A)** A 3D model of the yeast MFC based on binary interactions between isolated recombinant subunits and affinity purifications of MFC subcomplexes produced by His_8-tagged subunits harboring deletions of predicted binding domains for other components of the complex. The various subunits of eIF3 (orange, red, and purple shapes) are labeled with their yeast (e.g., TIF32) and universal (e.g., 3a) designations. The subunits of eIF2 (green) are labeled α, β, and γ, with GTP and Met-tRNA$_i^{Met}$ bound (primarily) to eIF2γ to comprise the TC. The protein subunits and Met-tRNA$_i^{Met}$ are shown roughly in proportion to their molecular weights. The eIF5, NIP1, and termini of TIF32 are depicted as solid rather than partially transparent shapes to emphasize their importance in binding to 40S ribosomes. Specific inter-actions detected of eIF3 subunits with RPS0A and helices 16-18 of 18S rRNA are depicted (186). The locations of relevant deletion endpoints also are indicated. *NTD,* N-terminal domain; *CTD*, C-terminal domain; *hld*, HCR1-like domain; *rrm*, RNA recognition motif. **(B)–(G)** The amino acid sequences of the yeast eIF3 subunits are depicted as rectangles using the same color schemes as in (A) with selected amino acid positions and locations of selected point mutations shown above, and the locations of selected deletions shown below. The locations of binding domains for other MFC components, other eIFs, 40S ribosomal proteins RPS0A and RPS10A, and 18S rRNA are all indicated below the schematics with double-headed arrows. The locations of PCI homology domains in TIF32 and NIP1 (*pci*), the HCR1-like domain in TIF32 (*hld*), WD repeats in TIF34 (*wd 1, 2,..., 7*), predicted RNA recognition motifs (*rrm)* in PRT1 and TIF35, and a predicted Zn-binding domain in TIF35 (*zn*) are also indicated in the colored rectangles. See text for further details.

Although the PRT1/b CTD contains a NIP1/c-binding site (Fig. 7.2-9D), PRT1/b must interact with both TIF32/a and NIP1 for efficient incorporation into the eIF3 complex [162]. This can explain why a stable PRT1/b–NIP1/c binary complex was not formed *in vivo* by overexpressing these two subunits alone. By contrast, a stable TIF32/a–PRT1/b binary complex was purified from yeast and found to have low-level activity in promoting 40S binding of Met-tRNA$_i^{Met}$ and mRNA in a *prt1-1* extract [163]. Additionally, TIF32/a and NIP1/c can form a stable subcomplex in the absence of other eIF3 subunits [162]. The results of yeast two-hybrid and *in vitro* binding experiments suggest that the "noncore" eIF3 subunits eIF3e/INT-6 and eIF3d/Moe1 interact with one another and that eIF3e additionally binds to the three largest core subunits that comprise the stable a/b/c subcomplex described above [151, 165–167]. Consistently, disruption of eIF3d/Moe1 reduced the level of eIF3e/INT-6 in *S. pombe* extracts [167].

eIF3 Resides in a Multi Factor Complex with eIF1, eIF2, and eIF5

The eIF3 is physically associated with other essential eIFs in yeast. It co-purified with eIF1 [139, 168] and contained nearly stoichiometric amounts of eIF5 when purified by affinity chromatography [139]. *In vitro*, eIF1 and the eIF5-CTD can bind simultaneously to the NIP1/c-NTD [34, 69, 139]. Consistently, yeast eIF1 and eIF5 co-purified with the eIF3 a/b/c subcomplex, but not with the b/i/g subcomplex described above [163] (Fig. 7.2-9A). Interactions of eIF1 and eIF5 with eIF3c have also been observed for the mammalian factors [4, 169]. Interestingly, the yeast eIF5-CTD can interact simultaneously with NIP1/c and the β-subunit of eIF2 *in vitro* [34, 69], suggesting that eIF5 can bridge a physical interaction between eIFs-2 and -3. Indeed, a MFC containing eIF1, eIF2, eIF3, eIF5 and Met-tRNA$_i^{Met}$ (Fig. 7.2-9A) was shown to exist free of ribosomes and could be purified from yeast extracts. The seven alanine substitutions in AA-box 2 of the eIF5-CTD in the *tif5-7A* allele (described above) disrupt interactions of eIF5 with both eIF2β and the NIP1-NTD *in vitro* and dissociate eIF2 from eIF3 *in vivo*. This mutation confers a diminished rate of translation initiation and Slg$^-$ phenotype providing evidence that the MFC is an important initiation intermediate *in vivo* [34, 69]. Recent work indicates that TIF32/a mediates a second, direct contact between eIF3 and eIF2 in the MFC. The CTD of TIF32/a can bind to recombinant eIF2β *in vitro* and to eIF2 holoprotein *in vivo* in the absence of all other MFC components (Fig. 7.2-9B, Δ4). Consistently, a truncated form of TIF32/a lacking this binding domain (TIF32-Δ6) forms a MFC *in vivo* that lacks only eIF2. Overexpression of TIF32-Δ6 confers a dominant Slg$^-$ phenotype in otherwise wild-type cells and it exacerbates the translation-initiation defect in *tif5-7A* cells. Thus, the direct connection between eIF2β and eIF3 involving the TIF32-CTD and the indirect contact between eIF2β and NIP1/c via the eIF5-CTD seem to have additive stimulatory effects on a common step of translation initiation *in vivo* [162].

Formation of the MFC Stimulates Multiple Steps of Initiation

The presence of eIF2 and eIF3 in the MFC might be expected to enhance TC binding to ribosomes by cooperative binding of both factors to the 40S subunit. Three observations are consistent with this idea. First, TC binding to 40S subunits was defective in *tif5-7A* extracts in a manner rescued by purified wild-type eIF5 [72]. Second, the Slg$^-$ phenotypes of *tif5-7A* and high-copy *TIF32-Δ6* were partially suppressed by overexpression of the TC [69, 162]. Thus, at least one consequence of disrupting MFC integrity seems to be a reduction in TC binding to 40S subunits. Third, overexpression of the NIP1-NTD sequesters eIF2, eIF1 and eIF5 in a nonribosomal subcomplex lacking all eIF3 subunits (Fig. 7.2-9C, N) and produces a Gcd$^-$ phenotype. This phenotype is exacerbated by overexpressing eIF1 and eIF5, which enhances formation of the NIP1-NTD/eIF5/eIF1/eIF2 subcomplex *in vivo*, and also by overexpressing the TIF32-CTD, which sequesters eIF2 in a distinct binary complex. Because the Gcd$^-$ phenotype of the NIP1-NTD was suppressed by simultaneously overexpressing TC, it was concluded that TC binds to the 40S subunit inefficiently when it resides in the NIP1-NTD/eIF5/eIF1/eIF2 or the TIF32-CTD/eIF2 subcomplexes compared with intact MFC [162].

Paradoxically, no Gcd$^-$ phenotype was observed in *tif5-7A* mutant cells, even when the TIF32-Δ6 protein was being overexpressed. Moreover, there was an accumulation of 48S complexes containing eIF1, eIF2 and eIF3 but lacking eIF5 in *tif5-7A* cells [72]. These observations have been interpreted to indicate that the physical contacts among eIF2, eIF5 and eIF3 in the MFC are most critically required *in vivo* for a step(s) subsequent to TC binding to 40S subunits, such as scanning, AUG recognition, or GTP hydrolysis by eIF2. In this view, impairing one of the latter steps reduces the rate at which 48S complexes are consumed to produce 80S initiation complexes, compensating for the reduced rate of TC binding to 40S subunits that results from disrupting the MFC and suppressing the depletion of 43S complexes. The eIF5 stimulates GTP hydrolysis by eIF2 at AUG start codons and this reaction may be inhibited by eIF1 at non-AUG codons [53, 170] (see below). In addition, there is evidence that eIF1 promotes scanning and can destabilize 48S complexes at near-cognate start codons or at AUG triplets in a suboptimal sequence context [171]. As shown in Fig. 7.2-9(A), eIF1 is tethered to the MFC by interactions with the eIF5-CTD, the NIP1-NTD, and a C-terminal segment in TIF32 [34, 69, 139, 162, 163]. Thus, it is possible that MFC integrity is critically required to juxtapose eIF1, eIF2 and eIF5 in relation to one another and the P-site of the ribosome in a manner required for efficient scanning, AUG recognition, or GTP hydrolysis at the start codon.

To explain the absence of a Gcd$^-$ phenotype in the *tif5-7A* and *TIF32-Δ6* mutants, it could be proposed that a delay in scanning or GTP hydrolysis at the uORF4 start codon produced by these mutations would impede the progression of all 40S ribosomes scanning from uORF1, compensating for the reduced rate of TC binding expected to occur in these mutants. This would restore efficient reinitiation at uORF4 and suppress the Gcd$^-$ phenotype that normally results from a decreased rate of TC binding. By contrast, overexpressing the NIP1/c-NTD or TIF32/b-CTD

sequesters eIF2 in defective subcomplexes and reduces the concentration of intact MFC, but does not generate MFC subcomplexes with the defects in scanning, AUG recognition or GTP hydrolysis postulated above. Hence, overexpressing the NIP1/c-NTD or TIF32/b-CTD has the same outcome as mutations in eIF2B, merely reducing the rate of TC binding to 40S subunits, thus yielding a Gcd$^-$ phenotype [162]. An alternative possibility that cannot be discounted is that TC binding during reinitiation on *GCN4* mRNA does not involve eIF3 and the MFC, and that sequestering eIF2 in the subcomplex with eIF5, eIF1, and the NIP1/c-NTD interferes with its recycling by eIF2B or formation of the TC, rather than delaying TC binding to 40S subunits.

Possible Functions of eIF3 in mRNA Binding

In addition to its role in Met-tRNA$_i^{Met}$ recruitment, eIF3 also stimulates mRNA binding to the 40S subunit in mammalian and yeast extracts [29, 37, 133, 163]. Because TC binding stimulates mRNA binding to the 40S ribosome [37, 133], eIF3 could act indirectly through its role in TC recruitment. However, eIF3 also seems to have an additional function in mRNA binding independent of TC [133]. The latter is generally attributed to interactions between eIF3 and the mRNA-associated factors eIF4G [172] or eIF4B [173]. Whereas mammalian eIF4B interacts directly with the eIF3a/p170 subunit [174], the yeast homolog of eIF4B (encoded by *TIF3*) interacts with yeast TIF35/eIF3g [147] (Fig. 7.2-9F). Mammalian eIF3 contains three subunits that can bind RNA as isolated proteins (eIF3a/p170, eIF3d/p66, and eIF3g/p44) [144, 175–178] (Table 7.2-2) and thus eIF3 could interact directly with mRNA in the initiation complex. Indeed, the b, c, and d subunits of mammalian eIF3 were found crosslinked to globin mRNA in 48S preinitiation complexes [175]. RNA-binding activities of certain eIF3 subunits could mediate direct interactions with the 18S rRNA, as suggested by UV-crosslinking experiments for human eIF3d/p66 [179]. Deletion of the RRM from yeast eIF3g/TIF35 was not lethal but produced a Slg$^-$ phenotype. The nature of the RNA that interacts with this RRM is unknown.

Mammalian eIF3 can bind to the hepatitis C virus (HCV) and classical swine fever virus IRES elements, and the eIF3a/p170, eIF3b/p116, eIF3d/p66 and eIF3f/p47 subunits were found crosslinked to these mRNA sequences [180, 181]. The binding region for eIF3 in the HCV IRES has been localized to domains IIIa–b [180, 182] and the cryo-EM map of the IRES–40S complex places this domain extending from the platform side of the 40S subunit just below the mid-line of the particle [183]. This location is consistent with the binding site for eIF3 on 40S subunits visualized in three-dimensional (3D) reconstructions of electron micrographs of negatively stained native 40S subunits [31, 184]; however, eIF3 also makes extensive contacts with the solvent side of the 40S subunit in the model of Lutsch et al [184]. It is unclear whether conventional mRNAs translated by the scanning mechanism will interact with eIF3 in the same manner utilized by the HCV IRES, as the latter bypasses the requirement for the eIF4 factors in forming the 48S complex [185].

Binding of eIF3 to the 40S Ribosome

Recently, domains in eIF3 required for binding to 40S ribosomes were identified by investigating whether the MFCs formed by mutant versions of TIF32/a and NIP1/c, many of which lack numerous MFC components, can compete with native MFC for stable 40S binding *in vivo*. The results showed that the N-terminal half of TIF32, NIP1 and eIF5 comprise a minimal 40S binding unit (MBU) sufficient for 40S binding *in vivo* and *in vitro*. The N- and C-termini of NIP1 and the TIF32-NTD were required for 40S binding by otherwise intact MFC complexes (*TIF32-Δ8* mutation, Fig. 7.2-9B; *NIP1-ΔB′*, Fig. 7.2-9C), suggesting that these eIF3 segments make direct contact with the 40S ribosome. Consistently, the TIF32-NTD interacted specifically with 40S subunit proteins RPS0A and RPS10A, and NIP1 interacted with RPS0A and 18S rRNA *in vitro*. The NIP1-NTD may also contact the 40S subunit in addition to its role in tethering eIF5 to the MFC. eIF5 was necessary for 40S binding only when the TIF32-CTD was absent. Thus, whereas the *tif5-7A* mutation did not reduce 40S binding by any MFC components except eIF5, it reduced binding by the mutant subcomplexes formed by the C-terminally truncated proteins TIF32-Δ6 (lacking only eIF2) and TIF32-Δ5 (Δ5 and Δ6; Fig. 7.2-9B). Interestingly, a 140 nt segment of domain I in rRNA, encompassing helices 16–18, is necessary and sufficient for specific binding of 18S rRNA to the TIF32-CTD *in vitro*. Hence, the 40S binding activity of the TIF32-CTD may involve direct interaction with domain I of rRNA [186].

In the cryo-EM model of the yeast 40S subunit [187], RPS0A is on the solvent side of the 40S subunit between the protuberance (pt) and beak (bk). Hence, binding of the TIF32-NTD and NIP1 to RPS0A would place this portion of eIF3 on the solvent side of the subunit, consistent with the EM analyses of 40S–eIF3 complexes [31, 184] and the location of the HCV IRES (and its eIF3-binding domain) on the 40S subunit [180, 182]. Interaction between the TIF32-CTD and helices 16 and 18 of the rRNA would provide eIF3 with access to the 60S-interface side, as these helices are accessible from both sides of the 40S subunit. It was proposed that the bulk of eIF3 would bind to the solvent side of the 40S whereas the TIF32-CTD and NIP1-NTD would wrap around helix 16 or penetrate the cleft between the beak (bk) and shoulder (sh), respectively, gaining access to the interface side of the subunit. The P-site is located on the interface side ~50–55 Å from the binding sites for TIF32-CTD and NIP1-NTD predicted in this model [186]. This separation is comparable with the dimensions of the γ-subunit of eIF2 [51], making it reasonable to propose that the NTD of eIF2β can remain connected to the TIF32-CTD and the NIP1-NTD/eIF5 subassembly of the MFC while Met-tRNA$_i^{Met}$ is bound to the P-site. In contrast, the connections between eIF1 and the TIF32-CTD and NIP1-NTD might have to be severed to allow eIF1 to bind near the P-site [171].

7.2.2.5 eIF1A Stimulates Ternary Complex Binding to 40S Subunits and Participates in AUG Selection During Scanning

The ~17 kDa factor eIF1A has been implicated in ribosome dissociation, binding of TC and mRNA to 40S subunits, and also in scanning. The yeast and mammalian

eIF1A are similar in sequence and functionally interchangeable in supporting production of 80S initiation complexes and Met-puromycin synthesis using all mammalian components [188]. Yeast eIF1A is an essential protein *in vivo* and its depletion from cells impairs general translation initiation and leads to accumulation of 40S dimers [189]. The latter suggests that eIF1A is bound to native 40S subunits and prevents their dimerization *in vivo*. In early studies, mammalian eIF1A seemed to be less active than eIF3 in promoting TC binding to 40S subunits [37, 133], although a greater stimulation could be observed in the presence of 60S subunits and was attributed to a ribosome anti-association activity of eIF1A [190]. In a more recent study, purified eIF1A stimulated TC binding in the absence of mRNA or AUG triplet by almost 20-fold, whereas purified eIF3 conferred only a 3-fold stimulation [32, 134]. The eIF1 was found to augment the stimulatory effect of eIF1A on TC binding in the absence of AUG, even though it had little activity on its own, and the greatest level of TC binding occurred when eIF1, eIF1A and eIF3 were present simultaneously [33]. Studying the corresponding yeast factors in a reconstituted system, Lorsch et al. found that eIF1A could function in the absence of eIF3 to stimulate TC binding to 40S subunits in the presence of a model 43-nt unstructured mRNA and 60S subunits. In this system, eIF1A was strongly dependent on eIF1 for promoting TC binding, whereas eIF3 had no stimulatory activity in the presence or absence of eIF1 and eIF1A [135]. Thus, the relative importance of eIF1, eIF1A and eIF3 in promoting 43S complex formation *in vitro* seems to vary with the source and pre-paration of factors, ribosomes and assay conditions.

Maitra et al. [32] reported that eIF1A cannot stimulate TC binding in the presence of 60S subunits under conditions that promote subunit joining. The eIF3, by contrast, could function in the presence of 60S subunits, but its stimulatory effect disappeared with the addition of an AUG triplet. To account for these findings, they proposed that both eIF1A and eIF3 are required to form a stable 43S complex, with eIF1A catalyzing transfer of TC to 40S subunits harboring eIF3. The eIF3, in conjunction with TC, protects the 43S complex against disruption by a 60S subunit prior to mRNA binding but becomes dispensable for this function once Met-$tRNA_i^{Met}$ is base-paired to the AUG codon [32]. As mentioned above, they found more recently that eIF1A can enhance the anti-association activity of eIF3, and that eIF1A and eIF1 function together as effectively as eIF3 does alone in preventing disruption of 43S complexes by 60S subunits in the absence of AUG. As in the case of TC binding, the combination of all three factors conferred the greatest anti-association activity of all. Consistently, it was found that eIF3 (strongly) and eIF1A (moderately) enhanced stable 40S binding by eIF1, and that eIF1 (strongly) and eIF3 (moderately) enhanced 40S binding by eIF1A. Hence, eIF1, eIF1A and eIF3 probably cooperate in the formation of a stable 43S complex containing all three factors and TC prior to mRNA binding *in vitro* [33].

The eIF1A has an ortholog in Archaea and exhibits significant sequence similarity (21% identity) to bacterial initiation factor IF1 [2]. The three-dimensional structures of *E. coli* IF1 [3] and mammalian eIF1A [191] both contain the five-stranded antiparallel β-barrel known as the OB domain [3], whereas eIF1A contains an additional

α-helical domain and unstructured N- and C-terminal extensions not present in bacterial IF1 [191] (Fig. 7.2-8B). The archaeal orthologs of eIF1A contain abbreviated N- and C-terminal extensions and may lack the α-helical domain. As IF1 binds directly to the A-site of the 30S ribosome [192, 193], the OB-fold in eIF1A most probably binds to the A-site of 40S subunits in eukaryotes. The eIF1A shows nonspecific RNA-binding activity *in vitro* [194] with a K_d of ~15 mM [191], and NMR analysis has identified residues in the OB-fold and α-helical domains of eIF1A whose chemical shifts change in the presence of various RNAs, and thus may contact RNA directly. Consistently, mutations of several such residues reduced RNA binding by eIF1A. Interestingly, a *K67D* mutation of Lys^{67} also impaired eIF1A-stimulated TC binding to 40S subunits *in vitro*, leading to the suggestion that this residue is required for eIF1A binding to the rRNA in the 40S subunit (Figs. 7.2-8A and B) [191].

A C-terminal deletion that removes all of the unstructured CTD and a predicted 3_{10} helix in the helical domain of yeast eIF1A produced a Gcd^- phenotype in addition to the Slg^- phenotype observed for a smaller deletion that removes the eIF5B-binding domain (see below) (Fig. 7.2-8A). The fact that this Gcd^- phenotype was suppressed by overexpressing the TC suggests that it reflects diminished TC binding to 40S subunits scanning the *GCN4* mRNA leader after translating uORF1 [195]. The delayed rebinding of TC to these 40S subunits would allow a fraction of the latter to by-pass uORFs 2–4 and reinitiate at *GCN4* instead [19]. This provides the first *in vivo* evidence that eIF1A enhances TC binding.

Pestova et al. showed that eIF1A also acts in conjunction with eIF1 in the presence of TC, eIF3, and the mRNA-associated factors eIF4A, eIF4B and eIF4F, to promote formation of a stable 48S complex with the ribosome positioned at the AUG codon, as judged by toeprint analysis. In the absence of eIF1 and eIF1A, an unstable complex was formed close to the 5′-end, whereas addition of eIF1, in a manner enhanced by eIF1A, led to dissociation of this complex and the formation of the more stable, correctly positioned 48S complex. For EMCV RNA, where ribosome binding to the start codon is directed by an IRES, eIF1 could direct 40S ribosomes to the correct AUG without eIF1A. Thus, eIF1 may possess the critical activity for positioning a 40S ribosome at the start codon [196]. Interestingly, mutations in residues on the RNA-binding surface of eIF1A did not impair its ability to disrupt incorrect 48S complexes formed at the cap, but led to the stabilization of incorrect complexes located upstream from the start site [191] (Fig. 7.2-8A). These data are consistent with the idea that eIF1A acts from the A-site in conjunction with eIF1 to play a role in AUG selection by initiator tRNA during the scanning process.

eIF1A Interacts with the IF2 Ortholog eIF5B

As discussed above, eIF1A and eIF5B are structural and functionally similar to bacterial IF1 and IF2, respectively. In accordance with evidence that IF1 and IF2 interact on the 30S ribosome, eIF5B and eIF1A from yeast interact directly *in vitro* and are stably associated in cell extracts. The last 24 amino acids in the unstructured acidic tail of eIF1A and the C-terminal 153 residues in the eIF5B CTD are necessary and sufficient for strong interaction between the yeast factors [6, 195]. Concurrently,

NMR analysis showed that the C-terminal 14 residues of human eIF1A are sufficient for binding to the eIF5B-CTD *in vitro*, and provided evidence that the last five residues of eIF1A lie in a shallow hydrophobic groove between helices 13 and 14 in the eIF5B CTD [197]. This portion of eIF5B corresponds to the last 32 residues of the protein and is located in domain IV of the crystal structure of archaeal eIF5B that forms the base of the chalice-like molecule [198] (Figs. 7.2-16A and B). A deletion of the last 87 residues of yeast eIF5B impairs its function *in vivo* and *in vitro* [6], and deletion of the eIF5B-binding domain in eIF1A also reduces translation initiation *in vivo* [185, 195]. Thus, it seems likely that eIF1A–eIF5B association through their extreme C-termini enhances an important aspect of initiation. This interaction seems to be restricted to eukaryotes as the relevant domains in eIF1A and eIF5B are missing in bacterial IF1 and IF2, and archaeal eIF1A lacks the eIF5B-binding domain at the extreme C-terminus of eIF1A. There is evidence that the nonconserved NTD of yeast eIF5B makes an additional contact with eIF1A that can be observed only when both factors are bound to the same ribosome [195].

Overexpression of eIF1A exacerbated the growth defect of *fun12* mutants which either lack eIF5B entirely or contain a C-terminally truncated form of eIF5B. To explain this genetic interaction, it was suggested that eIF1A is partially dependent on eIF5B for release from the 80S initiation complex, such that eIF1A overexpression in a *fun12* mutant would prolong binding of eIF1A to the ribosome and impede entry of the first eEF1A-GTP–aminoacyl-tRNA complex into the A-site [6]. Given that IF1–IF2 association mutually stabilizes binding of these factors to the 30S ribosome [199, 200], interaction between the C-termini of eIF1A and eIF5B might also enhance their association with the 40S ribosome early in the pathway. Suggestive evidence for this possibility stems from the finding that deleting the eIF1A C-terminus confers sensitivity to paromomycin (Par^S phenotype) in a manner exacerbated by deleting the NTD of eIF5B [195]. This drug binds to the A-site of prokaryotic ribosomes in a region overlapping the binding site for IF1 [193]; hence, it may compete with eIF1A for A-site binding. In the absence of a strong interaction with eIF5B, eIF1A may compete less effectively with paromomycin for the A-site.

Wagner and co-workers [197] presented an intriguing structural model for eIF5B and eIF1A bound to the 40S ribosome (summarized in Fig. 7.2-16B, right), in which the IF1-related central domain of eIF1A is bound to the A-site and makes contact with domain II of eIF5B, in the manner proposed for bacterial IF1–IF2 interaction. The extreme C-terminus of eIF1A binds to domain IV in eIF5B, as described above, and the remainder of the unstructured eIF1A CTD is stretched out to span the 50 Å separating the eIF5B domains II and IV. Domain IV in eIF5B additionally makes contact with the methionine and accepter stem of Met-tRNA$_i^{Met}$ located in the P-site, as proposed for bacterial IF2 [197]. This model is consistent with the known physical interactions between eIF5B and eIF1A, and the genetic evidence that eIF5B enhances Met-tRNA$_i^{Met}$ binding to the P-site [201] and promotes the release of eIF1A from the A-site [6]. Interestingly, the helical domain of yeast eIF1A is oriented towards the P-site in the model of Wagner et al, consistent with its role in TC binding. It was shown that the unstructured basic NTD of eIF1A mediates direct interaction with eIF2 and

eIF3 *in vitro*, although these interactions appear to be confined to the surface of 40S subunits *in vivo*. Deleting this domain had a Slg$^-$ phenotype, especially at low growth temperatures, but did not confer Par^S or Gcd$^-$ phenotypes. Hence, its interactions with eIF2 or eIF3 may be most important for a step following recruitment of TC that involves isomerization of factors in the 43S or 48S complex [195]. If present in an extended conformation, the eIF1A NTD may be long enough to permit physical contact between eIF1A in the A-site and TC in the P-site (Fig. 7.2-16B, right).

7.2.3 Binding of Ribosomes to mRNA

7.2.3.1 The Ends of Eukaryotic mRNAs Contain Distinctive Conserved Structures

All nuclear-transcribed eukaryotic mRNAs contain the m^{7}GpppN cap structure (in which m is a methyl group and N is any nucleotide) [202]. The cap is added to the nascent pre-mRNA early during transcription, and plays important roles in mRNA metabolism in the nucleus and the cytoplasm, including splicing, nucleo-cytoplasmic export, translation and stability (see Ref. [203] for review). The cap is critical for efficient translation, being the primary mRNA structure recognized by the translation-initiation machinery, via eIF4E, for assembly of the 48S preinitiation complex. Some viral mRNAs do not contain a cap structure, and, thus, are not recruited to ribosomes via eIF4E, but use an alternative mechanism involving direct binding of the 40S ribosome to a specialized IRES (see below). A poly(A) tail is present on most eukaryotic cellular mRNAs (except for mammalian histones), and several viral mRNAs. The poly(A) tail plays an important role in mRNA stabilization and translation. Translation initiation is stimulated by the PABP, which binds to eIF4G and thus brings about circularization of the mRNA (see below).

7.2.3.2 Ribosome Binding to mRNA is Stimulated by the eIF4 Factors

The recruitment of ribosomes to eukaryotic mRNAs is catalyzed by the eIF4 group of factors, which includes eIF4A, eIF4B, eIF4G, and eIF4H, and requires the energy provided by hydrolysis of ATP (see Refs. [174, 204] for reviews). The eIF4A, eIF4E and eIF4G form a stable complex in mammalian cells, termed eIF4F, which interacts directly with the cap through eIF4E (Fig. 7.2-10A). In other species, including yeast, eIF4A is bound loosely to the eIF4F complex. The eIF4A is an RNA-dependent ATPase and RNA helicase, and these activities are stimulated by eIF4B and eIF4H. eIF4G is a large, modular scaffolding protein, containing binding sites for many other initiation factors (Fig. 7.2-11A).

In the most general mechanism for ribosome binding to mRNA, eIF4F binds directly to the cap structure via eIF4E. The 43S preinitiation complex is then recruited to the mRNA by an interaction between eIF4G and eIF3, to form the 48S pre-initiation complex (Figs. 7.2-10A and 7.2-1B). As discussed earlier and below, a

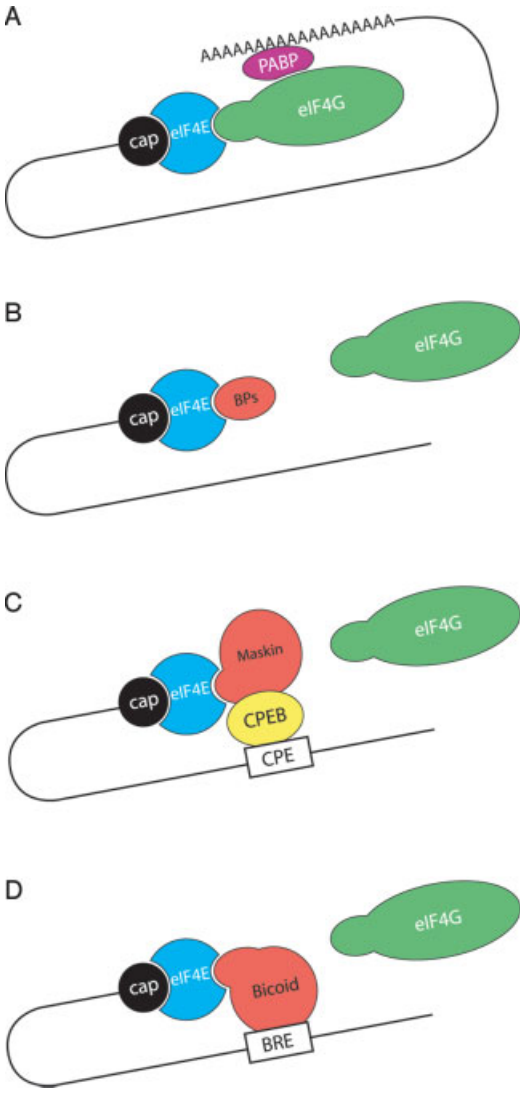


Figure 7.2-10 Regulation of translation via interactions with the mRNA 5′ cap structure **(A)** General model for 5′–3′ interactions. **(B)** Inhibition of cap-dependent by 4E-BPs binding to eIF4E and displacement of eIF4G. **(C)** CPEB-dependent displacement by maskin of eIF4G. CPEB interacts with the cytoplasmic polyadenylation element (CPE), which resides in the 3′-UTR and with eIF4E. This interaction results in inhibition of translation. **(D)** Bicoid binds directly to the Bicoid response element (BRE) in the 3′-UTR of *caudal* mRNA and displaces eIF4G from eIF4E, resulting in inhibition of translation. The common mechanism by which eIF4G is displaced from eIF4E by 4E-BPs, Maskin and Bicoid is competition for binding to eIF4E through the common motif YXXXXLΦ.In Maskin the tyrosine is substituted by threonine. Adapted with permission from Niessing et al. [293].

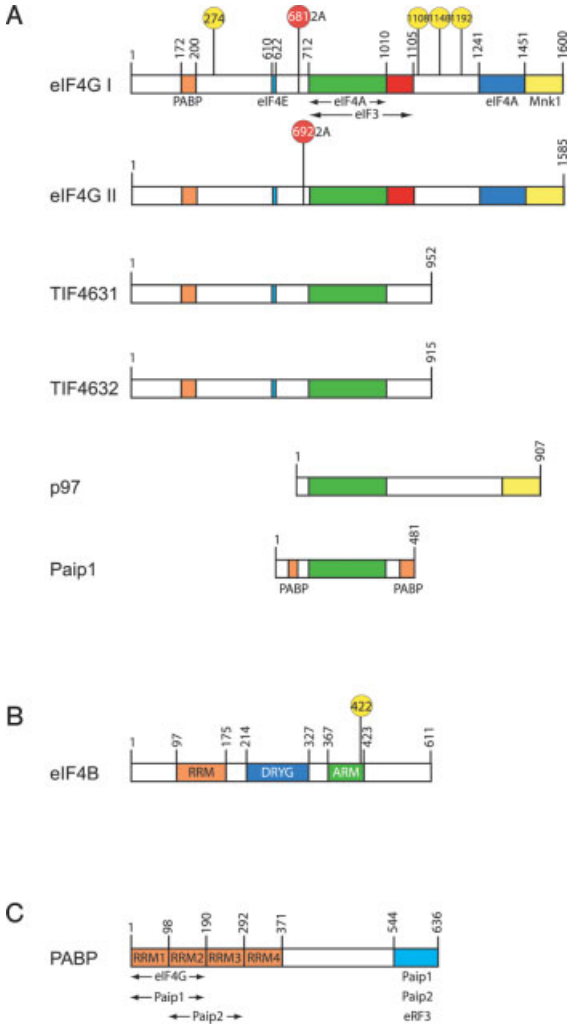


Figure 7.2-11 Schematic representations of the primary and domain structures of eIF4 factors and PABP. **(A)** Protein-binding domains in eIF4G family members are indicated by different colors. eIF4GI, eIF4GII, p97, and Paip1 are mammalian proteins, whereas TIF4631 and TIF4632 are the yeast eIF4G homologs. Phosphorylation sites in eIF4GI are indicated by their amino acid position (yellow circle), and the major cleavage site by the poliovirus 2A protease is also shown. **(B)** Functional domains in human eIF4B. RRM (RNA recognition motif) binds weakly to nonspecific RNA. DRYG (a region rich in aspartic acid, arginine, tyrosine and glycine) is important for dimerization and interaction with eIF3. ARM (arginine-rich motif) binds strongly non-specific RNA and is required for stimulation of eIF4A helicase activity. The S6 kinase phosphorylation site is indicated (yellow circle). Adapted from Methot, et. al. [271]. **(C)** Domain structure of human PABP. The boundaries of the conserved four RRM domains and the CTD (PABC) are shown. Also indicated are the binding sites of different proteins as described in the text.

number of other factors in the 43S complex participate in 48S complex assembly. It is noteworthy that a direct interaction between eIF4G and eIF3 has not been demonstrated in yeast, although eIF4G–eIF5 association has been detected and eIF5 may bridge an association between eIF4G and eIF3 in yeast cells [72]. In the following sections, structure–function relationships for the eIF4 initiation factors, followed by a discussion of alternative mechanisms of ribosome binding, ending with a discussion of the regulation of eIF4 activity are described.

eIF4E

eIF4E is the ~24 kDa cap-binding subunit of eIF4F. It was first identified by its ability to cross link to the cap structure [205] and was subsequently purified from a RRL [206]. eIF4E is essential for growth in yeast, and is highly conserved in primary sequence from yeast to human [207] (Table 7.2-1). Human and yeast eIF4E are 32% identical, and mammalian eIF4E can functionally substitute, albeit inefficiently, for yeast eIF4E [207]. The 3D structures of mammalian and yeast eIF4E bound to a cap analogue (m^7GDP) were solved by X-ray crystallography and NMR, respectively [208, 209] (Fig. 7.2-12). The protein resembles a cupped hand or baseball glove. It consists of a single α/β domain composed of an eight-stranded, antiparallel curved β-sheet, backed on its dorsal surface by three long α-helices. m^7GDP occupies a narrow slot on the concave surface of eIF4E, where m^7GDP binding is mediated by π–π stacking interactions between the base and indole side chains of two tryptophans, Trp^{56} and Trp^{102} in mammals or Trp^{58} and Trp^{104} in yeast. This binding is further stabilized by other interactions including hydrogen bonds, van der Waals contacts and electrostatic interactions. The amino acids involved in cap binding are conserved phylogenetically, demonstrating that the mechanism of cap-binding to eIF4E is also highly conserved.

eIF4E is phosphorylated on a single serine residue, Ser^{209} in mammals, which is conserved in all metazoans. Based on the mouse eIF4E-m^7GTP co-crystal structure, it was suggested that phosphorylated Ser^{209} forms a salt bridge with Lys^{159} [210]. The salt bridge was postulated to act as a clamp that brackets the proposed trajectory of the mRNA to stabilize mRNA–eIF4E interaction. However, two recent papers [211–212] demonstrated that eIF4E phosphorylated on Ser^{209} exhibits reduced, rather than increased, affinity (2–4-fold) for cap analogs. A recently described human eIF4E-m^7GpppA co-crystal structure might help to resolve some of these questions. It shows that the distance between the Cα positions of Ser^{209} and Lys^{159} is ~19 Å, which is too large for salt-bridge formation, thus arguing against a clamping mechanism [213, 214]. It was suggested that electrostatic repulsion between the penultimate adenosine of the cap structure and phosphorylated Ser^{209} might reduce eIF4E binding [211]. However, the distance might be too great (7 Å) for such a repulsion to take place. Thus, the mechanism by which phosphorylation of Ser^{209} lowers cap affinity is still in question. Regardless, the phosphorylation might stimulate the release of eIF4E from the cap structure so that the initiation complex could scan

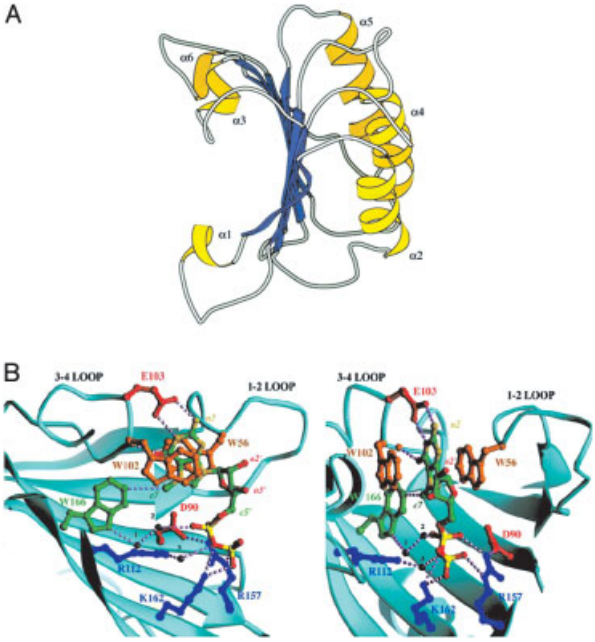


Figure 7.2-12 Ribbon diagram drawings of the 3D structure of yeast and murine eIF4E. **(A)** Solution structure of yeast eIF4E determined by NMR spectroscopy (shown with β sheets in blue and α helices in yellow). The cap structure binds to the convex surface of eIF4E, whereas the 4E-BP and eIF4G bind to a shared motif on the convex dorsal surface of eIF4E [209] (PDB ID: 1AP8). **(B)** X-ray co-crystal structure of eIF4E with m^7GDP. Only the cap-binding slot and the surrounding region of eIF4E is shown. The amino acids, which participate in m^7GDP binding, are indicated. Salt bridges, hydrogen bonds and van der Waals interactions are shown as dotted lines. Bridging water molecules are drawn in black (from Ref. [208]; PDB ID: 1EJ1). The two views of the molecule are rotated 90° relative to each other.

towards the initiation codon, in analogy to promoter clearance following phosphorylation of transcription complexes [211].

eIF4G

eIF4G is a modular scaffolding protein, and plays a major role in recruiting the ribosome to mRNA and coordinating the assembly of the 48S pre-initiation complex (Table 7.2-1). All eukaryotes contain two related eIF4G proteins (eIF4GI and eIF4GII in mammals are 46% identical; TIF4631 and TIF4632 in yeast are 53% identical). Neither TIF4631 nor TIF4632 is essential in yeast, but deletion of both is lethal [215]; hence, they execute partially overlapping functions. Mammalian cells

contain one additional, more distantly related eIF4G homologue p97/NAT1/DAP5 (see below). eIF4G is traditionally divided into three regions, which correspond to separate structural and functional domains connected by unstructured hinge regions that are sensitive to proteases (Fig. 7.2-11A). This stems from early reports on the cleavage by eIF4G into three fragments of roughly equal size by the picornavirus 2A protease [172]. The hinge regions are also susceptible to cleavage by the HIV-1 viral protease [216] and cellular proteases, such as the caspases involved in programed cell death (apoptosis) [217]. It is likely that the hinge regions provide the necessary flexibility to the independent domains to interact with each other and control eIF4G function [218].

The N-terminal fragment of eIF4G in all species examined contains binding sites for PABP and eIF4E [219–221] and the middle fragments contain the interaction domains for eIF4A, eIF3 and mRNA [222–224]. The C-terminal region, which exists only in metazoans (even in plants it is much shorter), contains an additional binding site for eIF4A [172, 224] and a binding domain for the serine–threonine protein kinases Mnk1 and Mnk2 [225, 226] (Fig. 7.2-11A).

It is well established that eIF4A binding to the mammalian eIF4G middle domain is sufficient for ribosome binding [227, 228]. Thus, a 43S pre-initiation complex can be assembled efficiently on an IRES-containing mRNA with only the eIF4G middle domain [228], or with the middle domain plus the eIF4E-binding site on a capped mRNA [218]. What then is the function of the second, C-terminal eIF4A binding site? Most evidence suggests that only one eIF4A molecule at a time interacts with one eIF4G molecule. This was initially suggested based on the finding that a mutant eIF4G defective for eIF4A binding to the C-terminal region exhibits a 3–4-fold reduction in translation, whereas deletion of the entire C-terminal fragment reduced translation by only 2-fold. It was therefore proposed that one molecule of eIF4A interacts with both the middle and C-terminal domains of eIF4G through two separate surfaces. This model was subsequently corroborated by showing that two differentially epitope-tagged eIF4A molecules failed to co-immunoprecipitate, consistent with nonsimultaneous binding of two molecules of eIF4A to the same eIF4G molecule [229]. Notwithstanding these conclusions, analysis of the ratio of eIF4A to eIF4G in the eIF4F complex led Korneeva et al. to conclude that the stoichiometry of eIF4A to eIF4G is 2:1 [230]. Clearly, further experiments are needed to resolve this disagreement. Why is there a requirement for two eIF4A-binding sites in eIF4G in mammals but not in plants or fungi? As noted above, the C-terminus might play a modulatory role in translation, which could involve phosphorylation of eIF4G via signaling pathways that affect growth and proliferation. Thus, in mammals, the CTD might mediate rearrangements of eIF4G interactions with other initiation factors following phosphorylation (see below). Recent findings show that the middle domain of eIF4G in yeast also interacts with eIF5 and this interaction could underlie the role of eIF5 in promoting 48S complex formation [72].

Several additional proteins have been reported to interact with eIF4G and could exert important physiological effects on translation or mRNA stability. However, these novel interactions have been less rigorously characterized in biochemical

assays than those described above. Pdcd4 is a tumor-suppressor protein that interacts with both eIF4A and eIF4G and inhibits cap-dependent translation [231]. The interaction site of Pdcd4 on eIF4G is in the middle domain. CBP80, together with CBP20, forms the nuclear cap-binding protein complex (nCBP) that is conserved from yeast to humans [232]. CBP20 interacts directly with the cap structure, whereas CBP80 was reported to interact with eIF4G both in yeast and in mammals [233, 234]. The interaction domain in yeast was mapped to the region between the eIF4E- and eIF4A-binding sites. There is evidence that the nCBP–eIF4G complex functions in the first round of translation [233, 235], and it was proposed that this "pioneer round" of translation is critical for degradation of mRNAs harboring premature stop codons by the nonsense-mediated mRNA decay (NMD) pathway [235]. Finally, the yeast decapping enzyme DCP1 was reported to interact with eIF4G and stimulate decapping of the mRNA [236]; however, the mechanism and characteristics of this interaction are yet to be determined.

The 3D structures of portions of the N-terminal and middle domains of eIF4G were resolved by X-ray crystallography and NMR spectroscopy. The structure of the eIF4E-binding site in the NTD of eIF4G was determined by NMR spectroscopy in yeast [237] and X-ray crystallography for mouse eIF4G [238]. In both species the eIF4E-binding domain is unfolded and becomes structured by an induced fit mechanism upon binding to eIF4E. A small unfolded fragment of mouse eIF4G (28 amino acids) assumes an α-helical structure with two turns when bound to eIF4E [238] (Fig. 7.2-13). A similar finding was made for an unstructured sequence of 88 amino acids in yeast eIF4G that becomes structured and resistant to proteases upon binding to eIF4E [237]. Strikingly, 4E-BPs, which are negative regulators of eIF4E, contain a closely related unstructured sequence, which interacts with the eIF4G-binding site on eIF4E and competes with eIF4G for binding to eIF4E (Fig. 7.2-10B). The sequence motif YXXXXLΦ (where Φ is a hydrophobic amino acid) is shared between 4E-BP and the 4E-binding domain in eIF4G. This motif is required for eIF4E binding by both proteins and makes direct contacts with amino acids in eIF4E. Consequently, the 3D structures of the 4E-binding domains in 4E-BP and eIF4G are superimposable (Fig. 7.2-13). Thus, 4E-BP competes with eIF4G by a *par-exellence* mechanism of molecular mimicry and inhibits the formation of eIF4F [239, 240]. This is a major pathway for regulating eIF4F function (see below).

The 3D structure of the middle domain of eIF4G, which binds eIF4A, eIF3 and RNA, also has been solved. This crescent-shaped domain consists of five HEAT repeats (Fig. 7.2-14A). Mutational analysis based on the structure has identified two adjacent patches of amino acids that bind eIF4A and the EMCV RNA, respectively [222].

p97/DAP5/NAT1[241–243], which plays a role in regulation of translation during apoptosis [241], exhibits homology to the C-terminal two-thirds of eIF4G. Consistently, p97 interacts with eIF3, eIF4A and Mnk1, but not with eIF4E or PABP [204] (Fig. 7.2-11A). Thus, p97 resembles the C-terminal fragment of eIF4G, which is generated during picornavirus infection and stimulates IRES-mediated translation (see below and Ref. [244]). In accordance with this, a fragment derived from p97 during

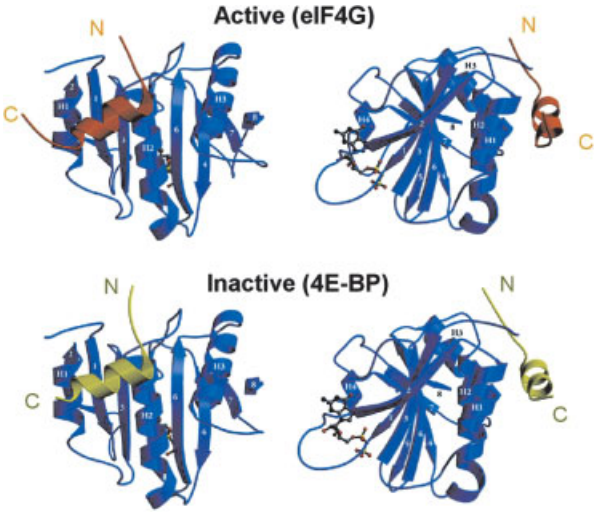


Figure 7.2-13 4E-BP1 is a molecular mimic of eIF4G. Ribbon diagrams of murine eIF4E (blue) in a complex with peptides of eIF4GII (orange), representing the active form of eIF4E, or 4E-BP1 (yellow), representing the inactive form of eIF4E, as determined by X-ray crystallography [238] (PDB IDs: 1EJ4 [4E-BP1] and 1EJH [eIF4GII]). m^7GDP is shown as a ball-and-stick representation bound to the convex surface of eIF4E. The structures on the right are rotated 90° vertically relative to each other.

apoptosis stimulates translation of IRES-containing mRNAs encoding proteins that function during apoptosis, including c-Myc, Apaf-1, XIAP, and p97 itself [204, 245, 246]. It is surprising that p97 inhibits, rather than stimulates, translation from the viral EMCV IRES [241–243]. It is therefore important to understand the different modes of action of p97 on the EMCV IRES (and perhaps other viral IRESs) versus the IRESs in cellular mRNAs.

eIF4A

The eIF4A is a ~45 kDa polypeptide that, is highly conserved in evolution (65% identity between yeast and human) (Table 7.2-1). Two different yeast genes (*TIF1* and *TIF2)* encode the same eIF4A protein, and either gene can provide the essential function of eIF4A *in vivo* [247]. The eIF4A is the most abundant translation-initiation factor, present at ~ 3 copies per ribosome as compared with 0.2–0.5 for other initiation factors [248, 249]. eIF4A is the prototype of the family of DEAD-box RNA helicases, named after one of the eight conserved motifs (D-E-A-D or Asp-Glu-Ala-Asp) shared amongst the family members. The DEAD-box family of proteins belongs to a much larger family, whose members contain the motif DExH/D (where x is any amino acid) [250]. Many of the DEAD-box proteins exhibit RNA-

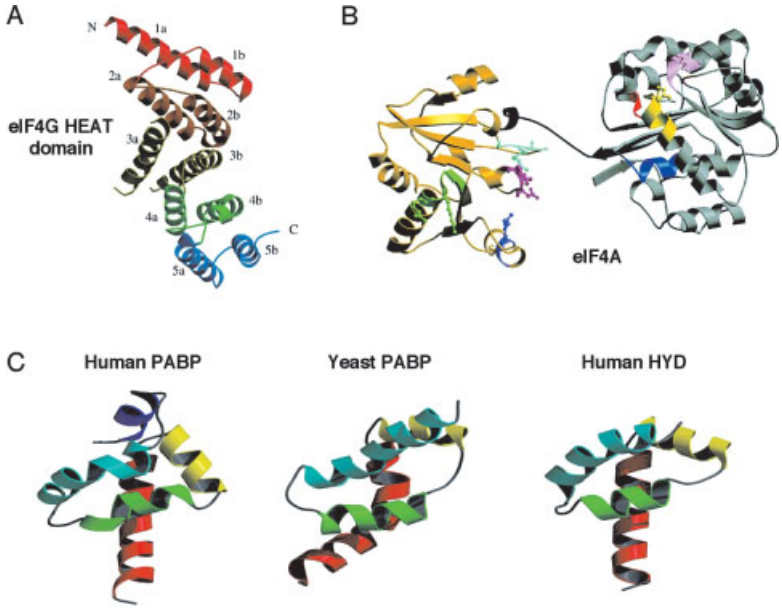


Figure 7.2-14 3D structures of mRNA binding factors. **(A)** Middle-region HEAT domain of eIF4G. Ribbon drawing of the HEAT domain of eIF4GII, which binds eIF4A and the EMCV IRES. The molecule is crescent shape, with the convex surface on the right and the concave on left. The view is along the cylindrical axis of the α helices. The HEAT domain consists of two antiparallel α helices, which are repeated in tandem five times. Unlike α helices in other HEAT domains, which are bent because they contain proline residues, those in eIF4G are predominantly straight. Each pair of helices is labeled by a different color [222]. **(B)** Model of eIF4A. The structure is shown as a ribbon diagram. The structure was constructed based on the separate crystallographic structures of the N- and C-terminal domains using molecular replacement [267]. According to the model, eIF4A, unlike other helicases, assumes a dumb-bell conformation, which consists of two compact domains. The two domains are connected through an extended, 11-residue linker. Conserved motifs are colored. **(C)** Comparison of the three PABC domains of human and yeast PABPs and human HYD. Ribbon diagrams of the 3D structures as determined by NMR spectroscopy for PABP, and X-ray crystallo-graphy for HYD. The α helices are differentially colored. Note that human PABP consists of five α helices, whereas yeast PABP and HYD contain only four helices. Adapted from Refs. [284, 285, 397]. (PDB IDs: 1GL9 (human PABC); 1IFW (yeast PABC); 1I2T (HYD)).

dependent ATPase activity, and a few show ATP-dependent RNA helicase activity. eIF4A has both activities [251, 252]; however, these activities are weak and (especially, the helicase activity) markedly stimulated by eIF4B or eIF4H [253, 254] The

mechanism of eIF4B stimulation is not well understood. eIF4A is a nonprocessive helicase, and unwinds by itself only 3–5 basepairs, but eIF4B prevents the reassociation of the unwound RNA, and thus promotes the unwinding of larger duplexes. Thus, the eIF4A (and subsequently all DEAD-box proteins) was hypothesized to function as a translocating motor, which removes mRNA secondary structure from the 5′-UTR to create a ribosome landing pad, and subsequently facilitate ribosome scanning [250, 253]. This is consistent with findings that secondary structures in the mRNA 5′-UTR potently inhibit translation [255, 256]. Also, inhibition of translation by dominant-negative mutants of eIF4A is directly proportional to the degree of secondary structure in the mRNA 5′-UTR [257]. There are three different eIF4A proteins in metazoans (eIF4AI, eIF4AII, and eIF4AIII). eIF4AI and eIF4AII are very similar in sequence (89% identity), and function [258]. eIF4AIII is 66% identical to eIF4AI in humans, but there is no evidence that eIF4AIII functions in translation [259]. It cannot substitute for eIF4AI in ribosome initiation complex formation. This might be explained by its ability to bind only to one of the eIF4A-binding sites on eIF4G (middle domain) [259]. A more distantly related eIF4A homolog, which functions in translation, is the yeast DED1 protein and its mouse homolog PL10. The *DED1* gene was initially isolated in a genetic screen using a temperature-sensitive mutant of eIF4E [260], and was subsequently shown to be required for translation initiation [260, 261]. DED1 is an RNA helicase, which functions independent of eIF4B [262]. Importantly, the murine homolog of DED1 (PL10), which is required for spermatogenesis [263], can substitute for its yeast counterpart [260]. An intriguing possibility is that DED1/PL10 is required for translation of a subset of mRNAs, perhaps those that contain extensive secondary structure. For example, Noueiry et al. [264] identified a mutant allele of *DED1*, which affects selectively the translation of a brome mosaic virus mRNA.

The eIF4A is composed of two domains, which were separately crystallized from yeast eIF4A fragments [265, 266]. A complete structure was assembled, based on results from the two studies [267] (Fig. 7.2-14B). A complete X-ray crystal structure of an eIF4A-like protein from *Methanococcus jannaschii* was also obtained [268]. This protein is similar in size to eIF4A. The eIF4A structure reveals that domain 1 contains the ATP-binding motifs, which are facing the cleft in the linker region separating the two domains. Domain 2 contains the binding site for RNA. The linker region also contains a motif (III), which links ATP binding with helicase activity. It is thought that the helicase activity is effected by conformational changes in the protein.

eIF4B

eIF4B was purified based on its ability to stimulate translation in an *in vitro* reconstituted rabbit reticulocyte translation system [133]. The best documented biochemical activity of eIF4B is its ability to stimulate the ATPase and RNA helicase activities of eIF4A in a highly specific manner (Table 7.2-1), as other RNA helicases are not stimulated by eIF4B. Also, genetic evidence demonstrated that eIF4A and eIF4B functionally interact [174, 269]. This striking biochemical and genetic specificity is all the

more surprising since there is no evidence that eIF4B interacts physically with eIF4A.

eIF4B is a 68 kDa polypeptide in mammals, but smaller in other species, and is conserved through evolution from yeast to humans [174]. However, the degree of sequence conservation is weak relative to other initiation factors (22% identity between yeast and human) (Table 7.2-1). Also, in contrast with most other initiation factors, eIF4B is not essential, as a yeast strain lacking the gene for eIF4B (*TIF3*) is viable [269]. Consistent with this, ribosome binding and translation are reduced, but not abrogated in the absence of eIF4B. Mammalian eIF4B is a dimer, whose dimerization is mediated by a region in the middle of the molecule called the DRYG domain (rich in aspartic acid, arginine, tyrosine and glycine) [173] (Fig. 7.2-11B). The DRYG region also interacts with eIF3 in mammalian cells and yeast, through eIF3g/p44, but an interaction was also reported with p170 in mammalian cells [270]. eIF4B also interacts strongly with RNA [271, 272]. A C-terminal arginine-rich motif (ARM) binds strongly to RNA in a sequence-independent manner. A second RNA-binding domain in the N-terminus containing an RRM [271] binds weakly to RNA in a sequence-independent manner. It was suggested that the RRM binds to the 18S rRNA, whereas the ARM binds to the mRNA [273]. Binding of eIF4B to the IRES of FMDV or EMCV is critical for translation [274, 275]. Thus, in addition to stimulating eIF4A helicase activity, eIF4B may promote mRNA–ribosome interaction during ribosome scanning, consistent with the finding that eIF4B exhibits RNA-annealing activity [276].

Mammalian cells contain an eIF4B-related protein termed eIF4H [277], which is 39% identical to eIF4B. eIF4H exhibits biochemical activities similar to eIF4B in that it stimulates the ATPase and helicase activities of eIF4A, and eIF4H can partially substitute for eIF4B in a highly fractionated rabbit reticulocyte translation system. The eIF4H is much smaller than eIF4B (~25 kDa), and lacks the DRYG domain, which mediates eIF4B dimerization. Consistent with this, eIF4H is a monomer. eIF4H contains an RRM, which is 45% identical to that of eIF4B, and binds weakly to RNA.

PABP

Although the PABP has not been traditionally viewed as an initiation factor, it is becoming clear that it plays an important role in translation initiation, primarily through its interaction with eIF4G. PABP is a phylogenetically conserved ~70 kDa polypeptide that is essential in yeast [278] (Table 7.2-1). PABP is an abundant protein, occurring in 6-fold excess of ribosomes, comparable with eIF4A levels [279]. A large number of studies have implicated PABP in mediating the stimulatory effects of the poly(A) tail on translation initiation (see e.g. Ref. [280]; see also review [281]).

An oligonucleotide of 12 adenosines (oligo $[A]_{12}$) is sufficient for binding to PABP, yet one PABP protects covers 27 adenosines on a poly(A) tail. PABP contains four highly conserved RRMs at the N-terminus (Fig. 7.2-11C). RRM1 and RRM2 form a contiguous binding site for oligo $(A)_{12}$ and can bind to poly(A) with an affinity equal

to that of wild-type PABP [282]. In addition, they are also bound by proteins (see below). By contrast, RRM3 and RRM4 bind to poly(A) with lower affinity (~10-fold), and exhibit weak affinity for non-poly(A) RNA. The proline-rich, C-terminal one-third of PABP serves as a docking site for several proteins, including PABP-interacting proteins 1 and 2 (Paip1 and Paip2) (Fig. 7.2-11C) (see below) [283–285].

Several distinct fragments of PABP stimulate translation in *Xenopus oocytes*, independent of their poly(A)-binding activity [280]. A fragment containing RRM1 and RRM2 of PABP, which binds eIF4G, strongly stimulates translation. RRM3, RRM4 and the C-terminus of PABP also stimulate translation, but to a lesser degree [280]. The mechanism by which RRM1 and RRM2 stimulate translation can be explained most probably through interaction with eIF4G, as discussed below. However, it is not clear how the other fragments stimulate translation.

The solution structure of the conserved ~75 amino acids at the C-terminus of PABP was determined by NMR spectroscopy [284]. The motif consists of five α-helices, which are arranged as an arrowhead (Fig. 7.2-14C). A deep hydrophobic pocket is formed between helices 2 and 3. This surface serves as binding site for Paip2, based on the chemical shift pattern induced upon ligand binding. Since these sequences are very highly conserved it is most probable that all proteins which interact with the PABP C-terminal conserved motif, such as Paip1 and eRF3, interact in a manner similar to Paip2. Consistent with this, Paip1 and Paip2 compete for binding to the CTD of PABP [283]. Similarly, Paip2 and eRF3 compete for binding (A. Kahvejian and N. Sonenberg, unpublished results). Surprisingly the human protein HYD (hyperplastic discs), which is unrelated to PABP in sequence, is structurally very similar to the PABP C-terminal conserved domain [285].

7.2.3.3 Circularization of mRNA via eIF4G–PABP Interaction

The cap and poly(A) tail of eukaryotic mRNAs are physically brought together by interactions between eIF4E and PABP with eIF4G to generate a circular mRNA (Fig. 7.2-10A). This was initially documented in yeast [219], and later shown for plants and mammals [220, 221]. Circularization of the mRNA provides a satisfactory explanation for the reported synergism between the cap structure and the poly(A) tail in stimulating translation initiation [286]. The eIF4G-binding site was mapped to the phylogenetically conserved RRM2 in yeast [287] and human PABP [220] (Fig. 7.2-11C), and the reciprocal binding site for PABP was mapped to the eIF4G N-terminus [220]. A stretch of 29 amino acids in the N-terminus of human eIF4G interacts with RRM1 and RRM2 of human PABP [220], as is the case for yeast PABP. Despite its high homology with yeast PABP, human PABP does not interact with yeast eIF4G [287], most probably because the PABP-binding sites in the human and yeast eIF4G proteins are quite divergent. However, the fact that this interaction has been conserved through evolution (or arose twice), despite the divergence of the protein sequences, underscores its paramount functional significance.

There are several mechanisms by which mRNA circularization could enhance translation initiation. First, circularization should increase the concentration of

terminating ribosomes in the vicinity of the mRNA 5′-cap structure and thereby facilitate ribosome recycling. This notion is bolstered by the finding that PABP also interacts with the termination factor eRF3. Thus, PABP may bridge eRF3 and eIF4G [288] (see below), looping out the 3′-UTR, and thus facilitating the direct shunting of ribosomes to the 5′-end of the mRNA. Secondly, PABP may participate in 60S subunit joining [289]. Thirdly, interaction of PABP with eIF4G may cause an allosteric effect that increases the affinity of eIF4E for the cap structure. It was also reported that in plants PABP enhances the helicase activity of eIF4F [290].

7.2.4 Translational Control by mRNA Circularization

mRNA circularization plays important roles in translational control through proteins unrelated to PABP, and many of these mechanisms are important during metazoan development. Whereas PABP circularizes the mRNA by interacting with eIF4G, other proteins cause circularization by interacting with eIF4E and thereby inhibiting eIF4E–eIF4G interaction and translation in a manner similar to the inhibitory mechanism for 4E-BPs. Certain mRNAs (such as cyclin B) contain cytoplasmic polyadenylation elements (CPEs) in their 3′-UTR, which are required for polyad-enylation and subsequent translation during *Xenopus* egg development. Polyadenylation is mediated by CPEB (cytoplasmic-polyadenylation binding protein). Maskin is a CPEB-interacting protein that inhibits polyadenylation and subsequent translation by interacting with eIF4E and displacing it from eIF4G [291]. Thus, by interacting with both eIF4E at the 5′-end and CPEB at the 3′-UTR, Maskin circularizes the mRNA and prevents complex formation with eIF4G (Fig. 7.2-10C). It is noteworthy that the motif in Maskin responsible for its binding to eIF4E is similar to that in eIF4G, but with the important difference that the tyrosine in the YXXXXLΦ motif is substituted by a threonine. This might explain why the interaction between maskin and eIF4E is rather weak [292].

Another recent example of 5′-cap-dependent translational control, which is mediated by the 3′-UTR and mRNA circularization, involves *caudal* mRNA translation in *Drosophila* [293]. *Caudal* is a transcription factor in the *Drosophila* embryo whose translation is specifically repressed in the anterior compartment by Bicoid, another transcription factor. Bicoid binds simultaneously to the Bicoid response element (BRE) in *caudal* mRNA and to eIF4E bound to the cap structure, consequently interfering with eIF4G binding (Fig. 7.2-10D). Thus, by competing with eIF4G binding to eIF4E through the common binding motif YXXXXLΦ, Bicoid inhibits specifically the translation of *caudal* mRNA.

Stimulation of translation via circularization occurs with other partners, which substitute for PABP, in mRNAs where a poly(A) tail is absent. For example, mammalian histone mRNAs are not polyadenylated and possess at their 3′-end a short stem-loop structure, which is required for optimal translation. This structure is bound by a protein, stem-loop binding protein (SLBP), which also binds to eIF4G, thus circularizing the mRNA. Consistent with this, SLBP stimulates translation. Curiously, SLBP does not bind to the PABP-overlapping binding site on eIF4G, but

rather at a site overlapping the eIF3-binding site [294]. Some viral mRNAs also do not contain a poly(A) tail. Rotavirus mRNAs are capped but not polyadenylated, and contain a sequence (UGACC) in their 3′-UTRs that is recognized by the viral NSP3 protein [295]. NSP3 interacts with the N-terminus of eIF4G to stimulate viral mRNA translation [296, 297]. In addition, NSP3 displaces PABP from its complex with eIF4G, and thus inhibits host protein synthesis [296]. In summary, these examples ascribe an important role for PABP as a translation factor that stimulates ribosome recruitement, and stress the importance of circularization for translational activation.

7.2.5 Regulation of eIF4 Function by Phosphorylation

7.2.5.1 eIF4E Phosphorylation

As discussed above, Ser^{209} is the only phosphorylation site on eIF4E that is conserved in metazoans. This residue is phosphorylated under many conditions including addition of growth factors, mitogens, hormones, and also in response to stress, such as arsenite or anisomycin treatment. The Mnk kinases which phosphorylate eIF4E in response to these conditions, are phosphorylated and activated by two different MAP kinase signaling cascades (Fig. 7.2-15) [298, 299]. The biological significance of eIF4E phosphorylation was clearly demonstrated in *Drosophila* where it was shown that alanine substitution of Ser^{251} (equivalent to mammalian Ser^{209}) caused slow development and resulted in smaller flies [300]. Significantly, the Asp^{251} mutation, which should mimic the phosphorylation state, rescued the small fly phenotype. It should be noted, however, that phosphorylation of eIF4E is not essential for general translation *in vivtro* and *in vitro* [301, 302]. Thus, the phosphorylation of eIF4E is probably critical for high-level translation of a subset of mRNAs, which encode proteins that function in cell growth and development.

7.2.5.2 eIF4E-4E-BPs

As discussed above, a major mechanism for the regulation of cap-dependent translation involves the family of proteins that inhibit translation initiation by binding to eIF4E, termed the 4E-BPs (see Ref. [204] for review. The family consists of three members (4E-BP1, 4E-BP2, and 4E-BP3 in mammals), which share 56–59% sequence identity. The conservation is especially prominent in the middle regions of the 4E-BPs, which contain the binding sites for eIF4E (including the YXXXXLΦ motif described above), and also in the flanking sequences that contain the phosphorylation sites, which regulate binding of 4E-BPs to eIF4E.

The phosphorylation state of several serine and threonine residues in the 4E-BPs regulates their affinity for eIF4E. Hypophosphorylated 4E-BPs bind strongly to eIF4E, but hyperphosphorylation reduces binding. Seven Ser/Thr phosphorylation sites were reported in the mammalian 4E-BP1 protein. Several of these sites were mapped by mass spectrometry, whereas others were inferred from mutagenesis studies. Two phosphorylated residues, Thr^{-37} and Thr^{-46}, lie on the N-terminal side

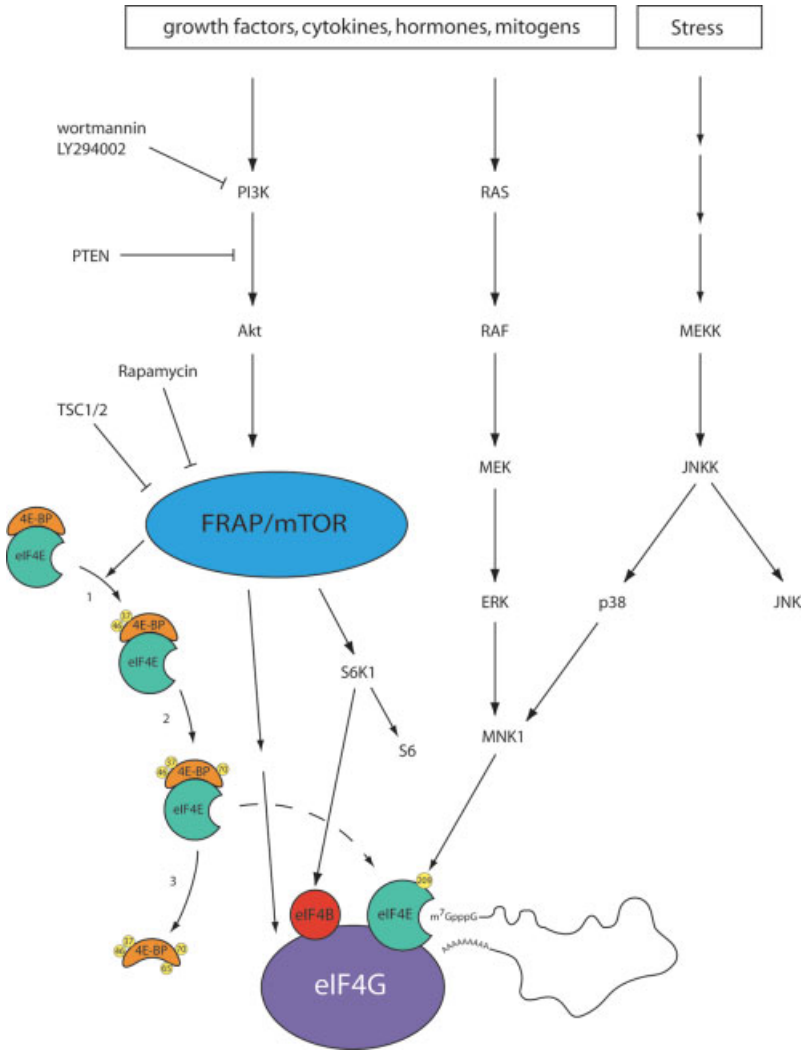


Figure 7.2-15 Signaling pathways leading to phosphorylation of translation-initiation components. Extracellular stimuli activate the PI3 kinase and Ras pathways. PI3 kinase signals through Akt and FRAP/mTOR to several translation components: 4E-BP, eIF4B, eIF4G and S6. Phosphorylation of 4E-BPs occurs in an ordered, hierarchical manner. Two tumor suppressor genes products: PTEN (phosphatase and tensin homolog deleted on chromosome ten), and TSC1/2 (tuberous sclerosis complex consisting of two proteins tuberin and hamartin), inhibit signaling through this pathway. Ras pathway activation leads to the phosphorylation and activation of Mnk, which directly phosphorylates eIF4E. Mnk is also phosphorylated by the stress-activated p38 MAP kinase.

of the eIF4E-binding motif, and five phosphorylated residues map on the C-terminal side: Ser-65, Thr-70, Ser-83, Ser-102, and Ser-112 (Ser-102 and Ser-112 exist only in 4E-BP1). Mutational studies combined with extensive 2D tryptic mapping analyses and the use of phospho-specific antibodies have demonstrated that 4E-BP1 phosphorylation is a highly ordered, hierarchical process [303]. Thus, dissociation of 4E-BP1 from eIF4E is a multistep process, in which phosphorylation of Thr-37 and Thr-46 acts as a "priming" event for Thr-780 and then Ser-65 phosphorylation. The role of Ser-83 phosphorylation remains to be elucidated. Ser-112 was reported to be phosphorylated by ATM (ataxia-telagengecia mutated) [304].

Crystallographic studies have suggested a mechanism to explain how 4E-BP phosphorylation may prevent binding to eIF4E. The presence of acidic patches on eIF4E flanking the bound 4E-BP peptide suggests that phosphorylation of the 4E-BPs on residues proximal to the eIF4E-binding site could induce electrostatic repulsion between the two proteins. The ordered phosphorylation of 4E-BP1 culminates in the phosphorylation of Ser-65 which is only 4 Å from Glu 70 in eIF4E. However, phosphorylation of this residue alone is not sufficient to disrupt eIF4E binding, suggesting that phosphorylation of other residues is also required [303]. The pathway that mediates 4E-BP phosphorylation relays signals from PI3K to Akt/PKB and FRAP/mTOR. The signaling pathway is illustrated in Fig. 7.2-15.

FRAP-mTOR (FKBP12-rapamycin-associated protein/mammalian target of rapamycin) is a member of the PIK family, whose members include cell-proliferation checkpoint proteins, such as ATM, ATR and DNA-PK, which function as protein kinases [305]. FRAP/mTOR is the mammalian homolog of yeast TOR proteins, which inhibit translation initiation and arrest yeast cells in the G1 phase in response to nutrient deprivation [306]. TOR genes were initially isolated by using a genetic screen for yeast mutants conferring resistance to rapamycin, an immunosuppressant drug that severely blocks T-cell growth at G1 [307]. Rapamycin acts intracellularly by binding to the immunophilin FK506-binding protein 12 (FKBP12). The FKBP12–rapamycin complex then specifically interacts with FRAP/mTOR and inhibits its activity. In mammalian cells, rapamycin blocks cap-dependent, but not IRES-mediated translation, through inhibition of 4E-BP phosphorylation [308]. In addition, expression of a rapamycin-resistant FRAP/mTOR mutant protein confers rapamycin resistance to 4E-BP1 phosphorylation. Whereas FRAP/mTOR directly phosphorylates only the "priming" sites, Thr-37 and Thr-46 in 4E-BP1 and 4E-BP2, the kinase probably plays an indirect, but critical, regulatory role in the phosphorylation of the downstream sites, Ser-65 and Thr-70 [309]. The kinases(s) responsive to FRAP/mTOR activation, and responsible for Ser-65 and Thr-70 phosphorylation remain(s) to be identified.

7.2.5.3 **eIF4G Phosphorylation**

Both eIF4GI and eIF4GII are phosphoproteins [310, 311], but their phosphorylation is differentially regulated in the cell [311]. Two clusters of phosphorylation sites were demonstrated in eIF4GI (Fig. 7.2-11A). One cluster maps to the N-terminus and

contains Ser^{314} (numbering is according to the full-length eIF4GI cDNA clone [312]), but it is unclear what conditions promote this phosphorylation [311]. Another cluster of serum-stimulated phosphorylation sites was mapped to the hinge region between the middle and C-terminal domains. These consist of Ser^{1148}, Ser^{1188}, and Ser^{1232} and are sensitive to PI3K and FRAP/mTOR inhibitors [311]. The effect of eIF4GI phosphorylation on its activity is unclear as no evidence for changes in activity or association with other initiation factors has been reported following phosphorylation. However, it is expected that eIF4GI phosphorylation would engender a conformational change in the protein that affects its activity. Clearly, the generation of "knock-in" mice with substitutions in the phosphorylation sites will be essential for assessing the biological significance of eIF4G phosphorylation. It is interesting that total phosphorylation of eIF4GII and p97 is lower than eIF4GI, and is not modulated by serum or mitogens. This is consistent with the fact that the phosphorylated C-terminal region in eIF4GI is not conserved in eIF4GII and p97, and this region is not phosphorylated in the latter proteins [311].

7.2.5.4 eIF4B Phosphorylation

eIF4B is phosphorylated in response to a variety of extracellular stimuli that induce cell growth and proliferation, such as serum, insulin and phorbol esters [313, 314]. One of the sites, Ser^{422} (Fig. 7.2-110B), is phosphorylated by S6K1 (S6 kinase 1), both *in vivo* and *in vitro*. S6K1 is a direct phosphorylation target for FRAP/mTOR (Fig. 7.2-15). This is consistent with the sensitivity of Ser^{422} phosphorylation to wortmannin and LY92900, which inhibit PI3K activity [315]. Thus, the PI3K/Akt-PKB/FRAP-mTOR signaling pathway regulates the phosphorylation state of eIF4B, eIF4G and 4E-BPs, underscoring its importance in controlling translation rates.

7.2.6 Translational Control by Paips – PABP Interacting Proteins

Two proteins that strongly interact with PABP and affect translation were identified by Far-Western interaction screening [316, 317]. Paip1 is a ~56 kDa protein, which activates translation *in vivo*. It is homologous to the central region of eIF4G and interacts with eIF4A [316] (Fig. 7.2-11A). Paip1 is also involved in mRNA turnover, as it is found in a protein complex with PABP that recognizes the major protein-coding region determinant of instability (mCRD) of the c-*fos* proto-oncogene mRNA [318]. Paip2, a much smaller protein (MW = 14 kDa), preferentially inhibits the translation of mRNAs containing a poly(A) tail, and also IRES-containing mRNAs that are eIF4G-dependent [317]. Paip2 inhibits the formation of 80S initiation complexes [317]. Although it is possible that Paip2 inhibits the conversion of 48S preinitiation complexes to 80S initiation complexes [289], Paip 2 also partially inhibits the formation of 48S preinitiaition complexes (Kahvejian and Sonenberg, unpublished). In an RNA-binding assay, Paip2 strongly hinders the interaction of PABP with poly(A) and disrupts the poly(A)-organizing activity of PABP [283, 317]. These last

findings suggest that Paip2 impedes initiation by disrupting the circular conformation of the mRNA. Thus, Paip2 may inhibit translation at several levels, including the recycling of ribosomes on the same mRNA, 48S preinitiation complex formation, and 40S-60S subunit joining. Paip1 binds to PABP with a 1:1 stoichiometry [319], whereas Paip2 binds with a 2:1 stoichiometry [317]. Paip1 interacts with RRM1+RRM2, and the C-terminus of PABP [319], whereas Paip2 interacts with RRM2+RRM3, and the C-terminus of PABP (Fig. 7.2-11C) [317].

The C-terminus of PABP interacts with Paip1 and Paip2, and with other proteins (see below) that contain the PABP C-terminus-binding motif, a 15-amino-acid stretch with the consensus sequence vxxsxLnpNAkeFvp [284, 285]. The translation-termination factor eRF3 contains this 15-amino-acid motif in its N-terminus and interacts with the C-terminus of PABP, as discussed above [288, 320]. An interesting protein that contains the PABP C-terminal binding motif, Ataxin-2, is implicated in spinocerebellar ataxia [321], and interacts with another RRM containing protein A2BP1 [322]. The function of Ataxin-2 is unknown, but its interaction with PABP may shed new light on spinocerebellar ataxia. Paip2 may therefore compete with C-terminus-binding partners of PABP, and modulate aspects of PABP function, which are unrelated to translation. Paip1 and Paip2 also interact with the C-terminus of the HYD ubiquitin ligase via the same 15-amino-acid stretch [285], potentially targeting these proteins for degradation. Perhaps ubiquitination serves as a regulator of mRNA circularization and translation. It is interesting to note that Paip1 and Paip2 exist only in metazoa. They may thus represent a higher-order mechanism for translational control in multicellular eukaryotes.

7.2.7 AUG Recognition during Scanning

7.2.7.1 AUG is the Predominant Signal for Initiation and is Selected by Proximity to the 5′-end by the Scanning Mechanism

Evidence for the Scanning Mechanism

Most mRNAs in eukaryotes are translated by the scanning mechanism, and genetic analysis in yeast played an important role in uncovering this process. Sherman and co-workers [323–325] isolated revertants of mutations in the start codon of *CYC1*, encoding cytochrome *c*, and found that translation could be restored by creation of an AUG at any of six locations in a span of 25 nucleotides near the 5′-end of the gene. This finding implied that AUG is the only sequence critically required for initiation. They also showed that an AUG could function efficiently as a start codon only if it occurred as the 5′-proximal AUG triplet in the mRNA. The null allele *cyc1-341*, containing AUGs at positions 1 and 5 plus the UAA terminator at position 3, could revert to a functional allele by elimination of the AUG at codon 1. Thus, the AUG at codon 5 could function efficiently as a start site so long as it was not preceded by another AUG [325]. Similarly, the nonfunctional *cyc1-362* allele was found to contain a single base pair change that introduced a new AUG 20 nt upstream of

the normal start codon, initiating a short upstream ORF (uORF). The fact that the uORF in *cyc1-362* blocked initiation at the normal start codon also implied that ribosomes were incapable of re-initiating downstream following termination of translation at the uORF [326]. This behavior is in sharp contrast with the ability of bacterial ribosomes to re-initiate translation efficiently on polycistronic mRNAs. A genetic analysis of initiation codon mutations at *HIS4* showed that for this gene also, 5′-proximal location is much more important than surrounding sequence context in determining whether an AUG triplet can function as a start codon [327].

These genetic findings were in accordance with previous observations by Kozak that most eukaryotic mRNAs are monocistronic and lack AUG codons upstream of their initiation sites. Combining these facts with observations that the m^7G cap stimulates translation and experiments showing that mammalian ribosomes cannot bind circular mRNAs [328], but will migrate on mRNA after binding at the 5′-end [329], Kozak proposed the scanning model. According to this hypothesis, the 40S subunit binds to the mRNA at the 5′-end and threads along the mRNA until reaching an AUG, whereupon an 80S initiation complex is assembled [330]. The scanning hypothesis is consistent with observations that insertions of secondary structure into the mRNA leader inhibit translation initiation [255, 331]. In addition, extensive mutational analysis of preproinsulin mRNA demonstrated that 5′-proximal AUG triplets are utilized preferentially as start sites and that insertion of an uORF inhibits initiation at the downstream cistron [332, 333], just as observed for yeast *CYC1*.

The scanning hypothesis was modified after the discovery that sequences immediately surrounding the start codon, particularly at the –3 and +4 positions (where A is designated the +1 base), can have a strong effect on initiation frequency, to the point where a 5′-proximal AUG can be by-passed if it occurs in an unsuitable sequence context. This exception to the first AUG rule was called "leaky scanning" [334]. The "first AUG rule" also can be violated in mammalian cells when an uORF is located a considerable distance upstream from the protein coding sequences, and Kozak hypothesized that ribosomes can resume scanning after translating the uORF and re-initiate downstream if they have sufficient time to reassemble a 48S preinitiation complex before reaching the next start codon. A separation of ~80 nt was sufficient for a high rate of re-initiation on preproinsulin mRNA [335].

Initiation Factor Requirements for Scanning

Although there is considerable evidence, both genetic and biochemical, that ribosomes move along the mRNA by a scanning process, the mechanism of scanning is neither understood at the molecular level, nor have scanning ribosomes been observed by any imaging technique. Nonetheless, as discussed below, important contributions to scanning of eIF4F, eIF1, and eIF1A, have begun to emerge. It is clear that energy derived from ATP hydrolysis is required for scanning; however, a critical issue is whether the 48S preinitiation complex actively uses the energy to traverse the mRNA or whether ATP-dependent unwinding of the mRNA by eIF4F is necessary solely to facilitate ribosome diffusion [336].

Recent support for the idea that ribosomes can diffuse on the 5′-UTR of an mRNA was obtained by Pestova and Kolupaeva [171]. The authors used an artificial mRNA with an unstructured 5′-UTR and demonstrated that significant ribosome binding to the initiation codon occurred in the absence of ATP and the eIF4 factors, provided that eIF1 was present along with eIF3 and the TC. These results are consistent with earlier data indicating that the requirement for ATP, eIF4A, and eIF4B is significantly reduced for mRNAs with diminished secondary structure in their 5′-UTR (such as alfalfa mosaic virus RNA4) [337]. The eIF1 was not required for 48S preinitiation complex formation at the start codon on the unstructured mRNA provided that eIF4F was present. From these findings, it was concluded that the 43S complex (40S/eIF3/TC) can bind to mRNA but requires either eIF1 or eIF4F to locate the AUG start codon. In this view, eIF1 and eIF4F have overlapping functions in ribosomal scanning. The results suggested that eIF1A also increases the processivity of scanning [171]. However, it seems possible that binding to mRNA, rather than scanning, was the critical step stimulated by eIF1 in these studies. A requirement for eIF1 in mRNA binding was discounted because the 40S–eIF3–TC complex could bind to mRNA and form a stable complex on an AUG triplet located only 1–2 nt from the 5′-end in the absence of both eIF1 and eIF4F factors. However, mRNA binding in this latter case could be facilitated by direct base-pairing of Met-tRNA$_i^{Met}$ with the 5′-proximal AUG codon.

The study of Pestova and Kolupaeva [171] provided clear evidence that eIF4F is involved in scanning, since it was not required for ribosome binding at the start codon in the unstructured mRNA leader, but was essential for this reaction when stable secondary structure was introduced into the leader, or in the case of native β-globin mRNA. The requirement of eIF4A for scanning must be mediated through its binding to eIF4G, as the former functions only as a subunit of eIF4F, whereby it cycles through the complex during the initiation process [258]. Thus, the function of eIF4G in scanning (see below) must be dependent on its ability to bind eIF4A.

Translational Control by Leaky Scanning

The phenomenon of leaky scanning has several important consequences for gene expression. First, translation of an mRNA can be down-regulated by a naturally occurring upstream AUG codon in the mRNA leader, in inverse proportion to the probability of leaky scanning past the first AUG. Secondly, Kozak showed that a 5′-proximal uORF can reduce the inhibitory effect of a second uORF located further downstream that is too close to the first uORF for efficient re-initiation following translation of the latter. In this way, translation of the first uORF promotes leaky scanning past the second uORF and enables subsequent reinitiation at the coding sequences downstream [335]. This is the mechanism employed in *GCN4* translational control, where translation of uORF1 allows ribosomes to by-pass uORFs 2–4 and reinitiate at *GCN4* when TC levels are reduced by eIF2α phosphorylation [338].

Leaky scanning can also allow the production of multiple proteins from the same mRNA. Interesting examples of this last phenomenon involve proteins that function

in mitochondria and cytoplasm and require a leader peptide for import into mitochondria. Ribosomes initiating at the first AUG produce the longer protein containing the leader peptide, whereas the shorter protein is translated by ribosomes that leaky-scan past the first AUG and initiate at the second start site [339]. Finally, there are numerous instances of translational repression by uORFs in which the uORF-encoded peptide blocks translation termination at the uORF stop codon and produces a barrier to the progression of scanning 40S subunits attempting to leaky scan past the uORF start site and reach the coding sequences downstream. In some cases, the inhibitory effect of the uORF-encoded peptide on termination can be modulated by nutrients, providing a mechanism for translational control of the downstream coding sequences (see Refs. [340, 341] for recent reviews).

7.2.7.2 The Anticodon of $tRNA_i^{Met}$, eIF2 Subunits, eIF1, and eIF5 are Determinants of AUG Selection during Scanning

Genetic experiments by Donahue and co-workers showed that the anticodon of $tRNA_i^{Met}$ plays a key role in the recognition of an AUG start codon by the scanning 40S ribosome [28]. They showed that overproducing a mutant form of $tRNA_i^{Met}$ containing an anticodon of 3′-UCC-5′ versus 3′-UAC-5′ restored expression of a *his4* allele with AGG in place of the AUG start codon. Other *his4* alleles with different start codons were not suppressed by $tRNA_i^{Met}$ (UCC), showing that codon–anticodon base pairing was required for suppression. Moreover, introduction of an extra AGG codon upstream of the AGG start site abolished suppression, indicating that the upstream AGG was recognized preferentially by the mutant $tRNA_i^{Met}$ (UCC) – a hallmark of the scanning process. This work established that perfect base-pairing between the anticodon of the initiator and the start codon, regardless of their exact sequences, is a fundamental requirement for efficient initiation in yeast. As discussed earlier, Donahue *et al.* also isolated mutations in eIF1, the three subunits of eIF2, and eIF5 that increased expression of a defective *his*4 allele containing AUU in place of AUG at the start codon. These Sui$^-$ mutations allow a UUG present at the third codon of *his4* to be recognized as a start site by Met-$tRNA_i^{Met}$, despite the predicted U-U mismatch at the first position of the codon–anticodon duplex [59, 78, 170, 342]. Biochemical analysis of Sui$^-$ mutants led to the proposal that the intrinsic rate of GTP hydrolysis by eIF2, and its modulation by the GAP eIF5, are important determinants of AUG recognition during scanning (see Ref. [343] and references therein).

7.2.7.3 eIF1 plays a role in TC binding, scanning, and AUG selection

The eIF1 is a ~12.5 kDa polypeptide (Table 7.2-1), essential in yeast [170], which plays an important role in assembly of preinitiation complexes and selection of AUG start codons. Early biochemical studies indicated that mammalian eIF1 has a weak stimulatory effect on binding of TC and mRNA to 40S or 80S initiation complexes in the presence of other factors [37, 133, 344]. As noted above, more recent studies showed that yeast eIF1 is critically required along with eIF1A for 48S complex formation in a

reconstituted system using a model mRNA [135]. In addition, mammalian eIF1 was found to augment the functions of eIF3 and eIF1A in promoting TC binding in the absence of mRNA [33]. The finding that mammalian eIF1 stimulated Met-puromycin synthesis by 80S initiation complexes only in the absence of AUG suggested that eIF1 can partially substitute for the start codon in positioning TC in the P-site [344]. Mammalian eIF1 also prevented 60S subunit joining in the absence of mRNA [133], and it enhanced the activities of eIF3 and eIF1A in this regard [33], consistent with a ribosome anti-association activity for eIF1.

The results of the yeast Sui$^-$ mutant selection first revealed that eIF1/SUI1 is required for stringent selection of AUG as the start codon [170]. As described above, mammalian eIF1 participates with eIF1A in promoting stable 48S complex formation at the start codon on globin mRNA [196]. When using a synthetic mRNA with a leader devoid of secondary structure, eIF1 was not required for scanning in the presence of eIF4F. However, in the absence of eIF1, there was an increase in the selection of (i) near-cognate start codons (e.g., AUU), (ii) AUG triplets surrounded by a suboptimal sequence context, and (iii) AUG triplets located within 4 nt of the 5′-end. Pestova et al proposed that eIF1 binds to the 40S ribosome near the P-site, similar to bacterial IF3, and influences the positions of mRNA, Met-tRNA$_i^{Met}$, or both, such that the initiator interacts strongly only with cognate AUG triplets in the proper context. Thus, in the absence of eIF1, the 43S complex can scan an unstructured leader (when eIF4F, 4A, 4B, and 1A are present), but the decoding site improperly accepts noncognate start codons or AUGs in a poor sequence context [171].

The solution structure of eIF1 has been solved by NMR (Fig. 7.2-8C) and the fold resembles that of certain ribosomal proteins and RNA-binding proteins; however, there is no evidence for direct interaction of eIF1 with RNA. The Sui$^-$ alleles of yeast *SUI1* (*D83Y, D83G*, *Q84P*, and *G107R*) [170, 345] alter residues predicted to be clustered together on the surface of eIF1 (Fig. 7.2-8C) and thus may comprise an important domain for eIF1 function in AUG selection [4]. Interestingly, the *SUI1* allele known as *mof2-1* (*G107R*) increases programmed –1 ribosomal frame-shifting on yeast L-A virus mRNA (maintenance of frame, or Mof phenotype) in addition to its Sui$^-$ phenotype, and the *sui1-1* allele (*D83G*), but not Sui$^-$ alleles of *SUI2* or *SUI3* affecting eIF2α or eIF2β, respectively, has a weak Mof$^-$ phenotype. The Mof$^-$ phenotype was recapitulated in *mof2-1* translation extracts and rescued with recombinant eIF1/SUI1. Thus, it was proposed that eIF1 plays a role in accurate decoding during the elongation phase. This unexpected activity seems to be conserved in humans, as human eIF1 cDNA complemented the Mof$^-$ phenotype of the *mof2-1* mutant [345]. Ostensibly at odds with this model, it was found that eIF1 cannot destabilize 80S initiation complexes, but only 48S complexes, formed on AUG triplets located at the extreme 5′-end of the mRNA [171].

7.2.7.4 eIF5 Functions as a GTPase Activating Protein for eIF2 in AUG Selection and Subunit Joining

The eIF5 is a 45–49 kDa polypeptide that stimulates hydrolysis of the GTP bound to TC in a 48S preinitiation complex positioned at the AUG start codon (Table 7.2-1).

This leads to release of eIF2-GDP, leaving Met-tRNA$_i^{Met}$ base-paired to AUG in the P-site of a 40S initiation complex that is competent to join with a 60S subunit in a manner stimulated by eIF5B (see below). The eIF5 functions catalytically and cannot stimulate GTP hydrolysis by TC that is free of ribosomes [38, 346, 347]. Thus, eIF5 may be regarded as a 40S-ribosome-dependent-GAP for eIF2 whose function is contingent upon a perfect codon–anticodon match between Met-tRNA$_i^{Met}$ and the AUG. The gene encoding the factor in yeast, *TIF5*, is essential [348] and depletion of eIF5 from yeast cells impairs translation initiation *in vivo*, leading to accumulation of vacant 80S couples [349].

As discussed above, the *TIF5* allele *SUI5-G31R* increases the rate of eIF5-stimulated GTP hydrolysis by the TC in a model assay for eIF5 GAP function, providing a plausible explanation for its defect in AUG selection (Sui$^-$ phenotype) [53]. By contrast, the *ssu2-1* mutation in the N-terminus of eIF5 (G62S) (Fig. 7.2-1D) led to a substantial reduction in eIF5 GAP activity *in vitro*, in accordance with its Ts$^-$ phenotype [72]. The catalytic domains of the GAPs for Ras and Rho have been shown to contain a critical arginine residue flanked by conserved hydrophobic residues that plays a catalytic role in stablilizing the transition state of the GTP hydrolysis reaction, known as an "arginine finger" [350]. It was shown that mutating such an invariant arginine residue in mammalian eIF5 (Arg-^{15}) destroys its GAP function in model 48S complexes and its ability to support both eIF5-dependent translation in a yeast extract and growth of a *tif5Δ* yeast strain. Mutations of this residue did not diminish interaction of eIF5 with recombinant eIF2β, with eIF2 holoprotein in the presence of GTPγS, or with model 43S complexes; hence, the mutation impairs eIF5 catalytic function and not substrate binding [351, 352]. Furthermore, the eIF2-GDP/eIF5 complex is stabilized by aluminum fluoride (AlF_4^-), a compound that combined with GDP acts as a structural mimic of the transition state of the GTPase reaction by G proteins [350]. These findings are consistent with the idea (but do not prove) that eIF5 functions as a GAP by inserting Arg^{15} into the GTP-binding pocket of eIF2γ to stabilize the transition state for GTP hydrolysis. Mutation of a second conserved arginine residue in eIF5 (Arg-^{48}) to methionine had only a modest effect on GAP function, but it did destabilize the eIF2–GDP–AlF_4^-–eIF5 complex [352]. Conserved Lys-^{33} and Lys-^{55} residues also contribute to catalytic function [351]. Because GTP hydrolysis requires pairing of Met-tRNA$_i^{Met}$ with AUG in the P-site, some component of the 48S complex besides eIF5, possibly a segment of the ribosome itself, most probably interacts with the switch-I or switch-II segments in eIF2γ to trigger GTP hydrolysis. This prediction is also supported by the fact that eIF5 is not present in archaea, suggesting that it evolved to enhance or regulate the GAP activity of the ribosome.

It was mentioned earlier that the AA-boxes in the CTD of eIF5 are required for stable binding of eIF5 to the β-subunit of eIF2 and to eIF2 holoprotein, both *in vitro* and *in vivo*, and also to the NIP1/c subunit of eIF3. Mutations in the AA-boxes of mammalian eIF5 that impair its interaction with the NTD of eIF2β (without reducing eIF5–eIF3c association), reduced the GAP activity of eIF5 *in vitro* [71] and its ability to support both eIF5-dependent translation in a yeast extract and growth of a

tif5Δ yeast strain. These findings suggest that the eIF5-CTD–eIF2*β* interaction is important for anchoring eIF2 to eIF5 in a manner that facilitates productive interaction of the catalytic domain of eIF5 (in the NTD) with the GTP-binding pocket of eIF2*γ*. Surprisingly, even more extensive mutations in the AA-boxes of yeast eIF5, encoded by *tif5-7A* and *tif5-12A*, did not impair eIF5 GAP activity in model 48S complexes [72] even though they eliminated stable binding between eIF5 and eIF2. The *tif5-7A* mutations weaken eIF5–NIP1/eIF3c interaction and destabilize the MFC *in vivo* [34, 69], and they impair binding of Met-tRNA$_i^{Met}$ and mRNA to 40S ribosomes in a yeast extract [72]. The latter is consistent with the idea that stabilization of the MFC by the eIF5-CTD promotes 48S complex assembly. The mRNA-binding defect in the *tif5-7A* extract could result indirectly from reduced TC binding, but there is also evidence that the AA-boxes in the eIF5-CTD promote interaction between eIF3 and eIF4G *in vivo* [72]. In the same study, it was shown that the C-terminal half of recombinant yeast eIF4G can bind directly to the eIF5-CTD, dependent on the AA-boxes, and this interaction can occur simultaneously with the eIF5-CTD–NIP1 interaction *in vitro*. Thus, the eIF5-CTD may bridge an interaction between eIF3 and eIF4G and thereby enhance mRNA binding to 40S ribosomes in yeast cells.

In spite of the impaired 48S assembly observed in *tif5-7A* cell-free extracts, the rate-limiting defect produced by this mutation *in vivo* lies downstream of 48S formation. Thus, 48S complexes containing eIF1, eIF2 and eIF3, but lacking eIF5, accumulated in the polysome fraction of *tif5-7A* cells. It is possible that dissociation of eIF5 from 48S complexes simply impairs the ability of eIF5 to perform its GAP function effectively on recognition of the start codon, even though this defect was not observed *in vitro* with model 48S complexes. Ostensibly at odds with this idea, 48S complexes accumulated in the *tif5-7A* mutant but not in *ssu2-1* (*tif5-G62S*) cells. Considering that the mutant eIF5 encoded by *ssu2-1* is defective for GAP activity *in vitro*, it appears that 48S complexes positioned at the AUG decay to free 40S subunits if GTP hydrolysis does not occur immediately following AUG recognition [72]. If so, then accumulation of stable 48S complexes in *tif5-7A* cells may signify a delay in reaching the start codon during the scanning process. Given that both eIF1 and eIF4G interact with the eIF5-CTD [34], and both are implicated in scanning and AUG recognition [171, 196, 343], a reduced rate of scanning may be the rate-limiting defect in *tif5-7A* cells.

7.2.8
Joining of 60S Subunits to 40S Ribosomal Complexes

Following scanning of the 40S subunit and selection of the AUG start codon, the next step in translation initiation is the joining of a 60S subunit containing the peptidyl-transferase active site, to form an 80S ribosome that is competent for elongation. Whereas the TC and factors eIF1, eIF1A, eIF3 and eIF5 are thought to be associated with the scanning 40S subunit, none of these factors are present on the 80S ribosome [36, 37]. Hydrolysis of the GTP in the TC leads to release of eIF2•GDP, and presumably the other factors, from the 48S preinitiation complex. However, the precise timing and requirements for factor release from the 48S complex has not

been determined, and interpretation of results from previous work is complicated by the recent realization that two factors, eIF5 and eIF5B, are necessary for the conversion of the 48S complex to an 80S ribosome. As described earlier, eIF5 promotes GTP hydrolysis by the TC leading to release of eIF2•GDP. The Met-tRNA$_i^{Met}$ remains bound in the ribosomal P-site with its anticodon base-paired to the start codon on the mRNA. Previously, it was thought that the 60S ribosomal subunit binds passively to this 48S complex following release of eIF2 and the other factors. However, reconstitution of the subunit joining step of protein synthesis using an authentic mRNA and the full complement of known mammalian initiation factors revealed the requirement for an additional factor termed eIF5B.

7.2.8.1 eIF5B Catalyzes a Second GTP-dependent Step in Translation Initiation

The eIF5B is a 112 kDa (yeast) to 139 kDa (mammals) polypeptide, with an electrophoretic mobility in SDS/PAGE closer to 150 kDa, probably owing to a highly charged N-terminus containing several runs of polylysine, polyaspartate and polyglutamate. The *FUN12* gene encoding yeast eIF5B is nonessential for viability; however, *Δfun12* strains exhibit a severe slow-growth phenotype with a doubling time 3-fold greater than wild-type [201]. The eIF5B is an ortholog of prokaryotic IF2, and contains at its center a consensus GTP-binding domain [5] (Fig. 7.2-16A). Similar to IF2, eIF5B binds GTP and GDP with similar affinities [353, 354], consistent with the lack of requirement for a guanine-nucleotide exchange factor for these factors. Early biochemical studies indicated that eIF5B (previously referred to as eIF-5 or IF-M2A, 355) stimulated the GTPase activity of eIF2, the role now ascribed to eIF5. However, with the realization that eIF5B is a GTPase and that eIF5 promotes GTP hydrolysis by eIF2, the proposed GAP-like function of eIF5B should be viewed cautiously.

The determination that both eIF2 and eIF5B are GTPases suggests that there are two GTP-dependent steps in eukaryotic translation initiation as opposed to the single GTP requirement in bacteria. The requirement for eIF5B to promote subunit joining was revealed when 48S complexes were assembled on β-globin mRNA using the full complement of known mammalian translation-initiation factors, including eIF1, eIF1A, eIF2, eIF3, eIF4A, eIF4B, eIF4F and radioactively labeled Met-tRNA$_i^{Met}$. Addition of recombinant eIF5 was insufficient to convert the 48S complexes in to 80S complexes, and the factor eIF5B was isolated from the ribosomal salt wash and shown to promote 80S complex formation [353]. Consistent with the latter, eIF5B was necessary to convert 48S complexes to 80S ribosomes competent for methionyl-puromycin (MP) synthesis [353]. This assay mimics the formation of the first peptide bond and monitors the transfer of labeled Met from Met-tRNA$_i^{Met}$ to puromycin, an aminoacyl-tRNA analog. In assays containing eIF1, eIF1A, eIF2 and eIF3, both eIF5 and eIF5B were necessary for MP synthesis. Interestingly, if 48S complexes were formed in the presence of only eIF1A and eIF2, then eIF5 was sufficient for MP synthesis [353]. Thus, eIF5 may promote GTP hydrolysis by eIF2 leading to release of eIF2•GDP from the 48S complex, and eIF5B may be required for the subsequent removal of eIF1, eIF3, and eIF5, which is probably necessary for 60S subunit joining.

When 48S complexes are mixed with 60S subunits and eIF5B in the presence of nonhydrolyzable GDPNP in place of GTP, 80S complex formation is unaffected; however, the 80S complexes are not able to synthesize MP [353]. Thus, GTP hydrolysis by eIF5B is not required for 80S complex formation, but rather for conversion of the 80S complex to a translation competent state. Consistent with this idea, kinetic analysis of translation initiation in RRL revealed a requirement for GTP hydrolysis late in the pathway to prepare the 80S complex for elongation [356].

Both yeast and mammalian eIF5B possess potent ribosome-dependent GTPase activities. The 60S subunit was reported to stimulate weakly the GTPase activity of human eIF5B [353] with the 40S and 60S subunits required for maximal GTPase activity. In contrast, the GTPase assays of yeast eIF5B [354] and rabbit eIF5B (called IF-M2A at the time) [355] revealed an absolute requirement for both ribosomal subunits. The G4 sequence motif (NKxD) in G domains interacts with the guanine base of the nucleotide establishing guanine specificity. Thus, the side chain of Asp-759 in human eIF5B (Asp-533 in yeast eIF5B) is predicted to interact via a hydrogen bond with the 2-amino group of GTP. Substitution of Asn for Asp-759 in human eIF5B severely impaired the ribosome-dependent GTPase and the subunit joining activities of the factor. These results indicate that GTP binding by eIF5B is necessary for subunit joining. The D759N mutation in human eIF5B endowed the factor with XTPase activity [357] (Fig. 7.2-16A). This nucleotide-specificity switch can be explained by the side chain of Asn-759 interacting via a hydrogen bond with the 2-keto group on XTP. Consistent with this explanation, human eIF5B-D759N stimulated subunit joining and MP synthesis when the assays were supplemented with XTP in addition to the GTP required by eIF2 to form a TC and bind Met-tRNA$_i^{Met}$ to the ribosome [357]. Thus, the requirement for both GTP and XTP in assays employing the human eIF5B-D759N mutant demonstrates that there are two nucleotide (GTP)-dependent steps in eukaryotic translation initiation.

7.2.8.2 GTPase Switch Regulates Ribosome Affinity of eIF5B and Governs Translational Efficiency

The eIF5B has a homolog in archaea, aIF5B, which functionally substitutes for yeast eIF5B both *in vivo* and *in vitro* [5, 198]. The crystal structure of eIF5B from the archaea *Methanobacterium thermoautotrophicum* (*M. therm.*) revealed a four-domain protein resembling a chalice [198] (Fig. 7.2-16B). All archaea lack the nonconserved NTD found in eukaryotic eIF5B and some bacterial IF2 proteins. As deletion of the NTD has no obvious effect on eIF5B function in yeast [6], the crystal structure represents the functionally important regions of the factor. Domain I of aIF5B is the GTP-binding domain and it resembles the G domains found in other G proteins like Ras, EF-Tu, and the heterotrimeric G proteins. Domain II of aIF5B is a β-barrel fold that, together with domain I, can be nearly superimposed on domains I and II of eIF2γ and the translation-elongation GTPases EF-Tu and EF-G. Domains I–III form the cup of the chalice-shaped factor and they are connected via a 40 Å long α-helical stem to domain IV, a second β-barrel fold, which represents the base of the chalice-

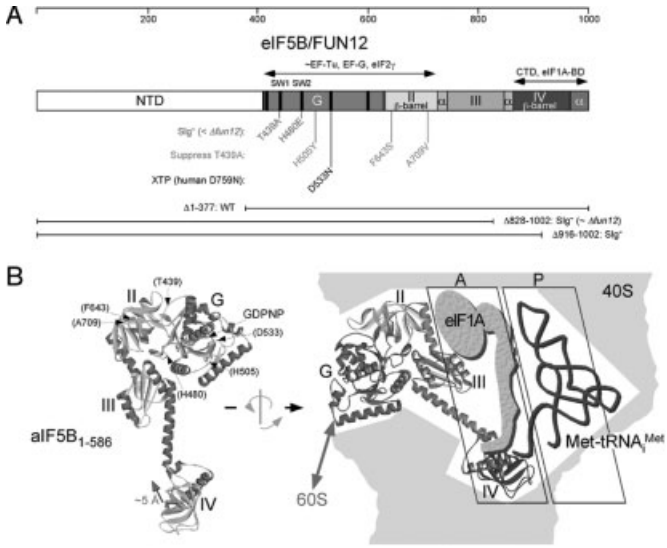


Figure 7.2-16 Structural and functional properties of eIF5B. **(A)** The 1002-amino-acid yeast eIF5B, encoded by the *FUN12* gene, is depicted schematically and divided into the following structural domains: nonconserved NTD, GTP-binding domain (G, red), β-barrel domain II (yellow), domain III (green), and β-barrel domain IV (blue). Domain IV plus the C-terminal α-helices (purple) comprise the C-terminal domain (CTD) containing the binding domain (BD) for eIF1A [6]. Recent NMR studies mapped the binding interface to the C-terminal α-helices in eIF5B (purple) and the CTD of eIF1A [197]. The G domain and domain II of eIF5B and aIF5B are conserved in both sequence and structure with the corresponding domains in eIF2γ (Fig. 7.2-5A), and the translational GTPases EF-Tu and EF-G [198]. Locations of key point and deletion mutations are indicated below the schematic. Depicted in red are dominant-negative mutations in the conserved Switch 1 (SW1) and Switch 2 (SW2) regions of the G domain that confer a growth defect in *fun12Δ* cells greater than that of the *fun12Δ* allele itself [354, 357]. Three mutations that suppress the toxic effects of the SW1 T439A mutation are indicated in green [354], and the G4 sequence motif mutation that alters the nucleotide specificity of eIF5B from GTP to XTP is labeled in black [357]. eIF5B alleles lacking the NTD retain full activity *in vivo*, whereas, partial or full removal of the CTD abolishes eIF5B function [6]. (B) Structure of *M. therm.* aIF5B. (Left) Ribbons representation of the ventral, GTP-binding face of aIF5B in complex with GDPNP (PDB ID code: 1G7T, 198). α-helices are depicted in red, and β-sheets are colored blue. Locations of the four domains of the protein and the key point mutations described in (A) are indicated. The arrow adjacent to domain IV reflects the ~5 Å movement of this domain away from the reader when the factor is bound to GDP [198]. (Right) Model of eIF5B binding and interactions on the 40S subunit. The dorsal face of aIF5B is presented in ribbons representation with the domains colored as in (A): red, G domain; yellow, domain II; green, domain III; blue, domain IV; purple, C-terminal α-helices. The structure has been rotated 180 about a vertical axis and then tilted relative to the structure in the left panel to reflect the predicted mode of eIF5B binding to the 40S ribosome (tan). Domain IV contacts the CTD of eIF1A (blue) in the A-site and is positioned near the aminoacyl end of the Met-tRNA$_i^{Met}$ (brown) in the P-site, domain II contacts the small ribosomal subunit, and the G domain contacts the GTPase center on the large subunit [197, 360, 361]. The aIF5B images in both panels were generated using the program WebLab Viewer Lite (v. 3.2, Accelrys, Inc.).

shaped molecule (Figs. 7.2-16A and B). As mentioned above, the eIF1A-binding site maps to domain IV of eIF5B [6] and an eIF5B allele lacking domain IV is functionally impaired *in vivo* (Fig. 7.2-16A), suggesting that the eIF5B–eIF1A interaction is important. Moreover, NMR structural analysis and chemical-shift experiments suggest that the C-terminal 14 residues of eIF1A bind to a narrow hydrophobic groove on the surface of domain IV formed by the two α-helices at the extreme C-terminus of eIF5B [197] (Fig. 7.2-16B). The eIF5B homolog in *Drosophila melanogaster* has been reported to bind to the DEAD-box RNA helicase VASA, encoded by the gene *vas* [358]; however, the VASA-binding site on eIF5B has not been mapped. Interestingly, mutations in *Drosophila* eIF5B enhance the embryonic patterning and germ cell specification defects of *vas* mutants, suggesting that the physical interaction between eIF5B and VASA is important for the translation of at least a subset of mRNAs encoding proteins that play important roles in early development [358].

Comparison of the structures of aIF5B in its active GTP-bound state and inactive GDP-bound state revealed that modest conformational changes in the active site of the G domain resulted in rotations of domains II and III that trigger lever-type motions of the long helical stem and domain IV [198] (see Fig. 7.2-16B, left panel). This lever-type motion amplifies over a distance of 90 Å, the intricate molecular rearrangements distinguishing GTP from GDP in the G domain active site, and results in an ~5 Å swing of domain IV. Interestingly, these domain rearrangements share some resemblance to the stroke-like motions observed in the ATPase motors like myosin. This similarity raised the possibility that eIF5B functions as a molecular motor, as proposed for the translation elongation-factor GTPase EF-G [359]; however, this possibility now seems very unlikely (see below).

Several studies have addressed the role of GTP hydrolysis by eIF5B. Pestova et al. [353] showed that GTP hydrolysis was not required for subunit joining *per se*, but for formation of a functional 80S competent for translation (MP synthesis). In addition, GTP hydrolysis was essential for the catalytic activity of eIF5B to promote subunit joining. The eIF5B functioned equally well in the presence of GTP versus nonhydrolyzable GDPNP to generate stoichiometric amounts of 80S complexes; however, GDPNP blocked the catalytic reutilization of eIF5B. In addition, the 80S complexes formed in the presence of GDPNP were defective for MP synthesis. Examination of the 80S complexes formed in these assays revealed the stable binding of eIF5B to 80S complexes formed using GDPNP, but not GTP [353]. Further evidence that GTP hydrolysis by eIF5B is not required for subunit joining, but is necessary for a later step in the translation pathway was obtained from studies of eIF5B mutants. Mutation of the conserved Thr-439 in Switch 1 of yeast eIF5B to Ala, or substitution of Glu for the conserved Switch 2 His480 in yeast eIF5B and the corresponding His706 in human eIF5B lowered the ribosome-dependent GTPase activity of the factor to below background levels [354, 357]. The mutant factors retained subunit-joining activity; however, they were defective for stimulating protein synthesis both *in vivo* and *in vitro*. In addition, the mutant factors exhibited a dominant-negative phenotype in yeast causing a slow-growth phenotype in strains co-expressing wild-type eIF5B, and severely impairing the growth of strains lacking eIF5B [354] (Fig. 7.2-16A).

Finally, the human eIF5B-H706E mutant mimicked wild-type eIF5B in the presence of GDPNP and was stably bound to the 80S product of the subunit joining reaction [357].

The results of the above studies suggest that GTP hydrolysis by eIF5B is required for dissociation of the factor from the ribosome following subunit joining. However, the experiments did not distinguish whether GTP hydrolysis by eIF5B was necessary for mechanical work on the ribosome required for protein synthesis, or if GTP hydrolysis and the domain rearrangements in eIF5B simply functioned as a switch lowering the ribosome-binding affinity of eIF5B. Three mutations in yeast eIF5B, one in the G domain (H505Y) and two in domain II (F643S and A709V), were isolated as suppressors of the eIF5B-T439A mutation [354] (Figs. 7.2-16A and B, left panel). The H505Y suppressor mutation lowered the ribosome-binding activity of eIF5B yet restored nearly wild-type translation *in vivo* without restoring eIF5B GTPase activity [354]. The uncoupling of eIF5B GTPase and translational activities by the suppressor mutations revealed that eIF5B GTPase activity is not essential for mechanical work in translation initiation. Instead, the eIF5B GTPase switch regulates the ribosome-binding affinity of the factor. These results suggest that similar to eIF2, eIF5B is performing both regulatory and catalytic roles in translation initiation.

Yeast lacking eIF5B or expressing eIF5B alleles that lack GTPase activity exhibit a Gcn$^-$ phenotype [354]. This Gcn$^-$ phenotype was attributed, at least in part, to a defect in translating uORF1 on the *GCN4* mRNA. Translation of uORF1 on the *GCN4* mRNA is necessary to restrict ribosomes to the reinitiation mode where the levels of TC determine whether ribosomes translate or scan past the inhibitory uORF3 and uORF4. Removal of uORF1 prevents ribosomes from bypassing uORF2, uORF3, or uORF4, and blocks *GCN4* expression. A 2–5-fold increase in the number of ribosomes leaky scanning past uORF1 without initiating translation was observed in strains lacking eIF5B or expressing GTPase-deficient forms of the factor [354]. The increased leaky scanning is consistent with the subunit joining function of eIF5B. Ribosomes scanning from the 5′-cap will stop at the uORF1 start codon, release eIF2, and await eIF5B-catalyzed subunit joining. In the absence of eIF5B or in cells expressing GTPase-deficient mutants of eIF5B, subunit joining may be inefficient or the 80S complexes formed may be aberrant, and ribosomes will either resume scanning or disengage from the mRNA.

Bacterial IF2 binds fMet-tRNA^{Met} to the 30S subunit and, similar to eIF5B, is released from the ribosome following subunit joining. The binding site for fMet-tRNA$_i^{Met}$ has been mapped to domain IV of IF2, and the tRNA recognition involves primarily fMet and perhaps a few residues of the acceptor stem (reviewed in Refs. [1, 7] and Chap. 7.1). Formylation of the amino acid is a critical determinant for binding to IF2; and neither aIF5B nor eIF5B has been shown to bind Met-tRNA$_i^{Met}$, consistent with the lack of formylation of Met-tRNA$_i^{Met}$ in archaea and eukaryotes. However, eIF5B may stabilize the binding of Met-tRNA$_i^{Met}$ to the 80S complex. Substantially greater amounts of Met-tRNA$_i^{Met}$ were bound to 80S complexes formed by wild-type eIF5B in the presence of GDPNP versus GTP [354]. As eIF5B would

remain bound to the 80S complexes formed using GDPNP, the factor may stabilize Met-tRNA$_i^{Met}$ binding. In a similar manner, eIF5B may stabilize 48S complexes following the release of eIF2•GDP. Polysomes with halfmers containing a 48S complex bound at the start codon of an mRNA accumulate in yeast with reduced amounts of 60S subunits. These halfmers fail to form in strains lacking eIF5B [357], suggesting that eIF5B stabilizes this intermediate in the translation-initiation pathway. Consistent with the proposed role of eIF5B to stabilize Met-tRNA$_i^{Met}$ binding to 48S and 80S complexes, the slow-growth phenotype of yeast strains lacking eIF5B can be partially suppressed by overexpressing tRNA$_i^{Met}$ [201].

Synthesizing the results of the various *in vivo* and *in vitro* studies on eIF5B, the following model can be proposed for eIF5B function. Following scanning, GTP hydrolysis by eIF2, and release of eIF2•GDP, eIF5B•GTP binds to the 48S complex. The binding of eIF5B may stimulate release of eIF1, eIF1A, or eIF3, or alter the 40S structure to permit 60S subunit joining. In the 80S complex, eIF5B is positioned such that domain IV is near the top of the A-site in close proximity to the Met of the P-site-bound Met-tRNA$_i^{Met}$ (Fig. 7.2-16B, right panel). Based on the cryo-EM images of EF-Tu and EF-G bound to the ribosome, the GTP-binding domain of eIF5B is likely to contact the GTPase-activating center on the large subunit, whereas domain II contacts the small subunit (see Refs. [360, 361]). Joining of the 60S subunit triggers the GTPase activity of eIF5B, the factor changes conformation, and dissociates from the ribosome. Finally, the 80S ribosome is prepared to enter the elongation phase of protein synthesis with Met-tRNA$_i^{Met}$ in the P-site and a vacant A-site ready to accept the first elongator tRNA in complex with eEF1A and GTP.

7.2.9
IRES-mediated Translation Initiation

Although it is clear that canonical cap-dependent initiation plays a central role in the recruitment of ribosomes to a large number of mRNAs, translation can initiate by a cap-independent mechanism on some mRNAs, and under specific conditions. This mode of initiation entails the binding of 40S ribosomes to an internal ribosome entry site, or IRES, which is mostly (but not always) part of the 5′-UTR of the mRNA. IRES-mediated translation was first discovered in picornavirus mRNAs (poliovirus [362] and encephalomyocarditis virus [363]). Unlike all nuclear-transcribed cellular mRNAs, these viral mRNAs do not possess a cap structure and, therefore, must be translated by a cap-independent mechanism. All picornavirus RNAs (including enteroviruses, rhinoviruses, aphthoviruses and others) contain an IRES. Other viruses that do not contain a cap structure, such as HCV [364] and cricket paralysis virus (CrPV) [365], also translate by an IRES-mediated mechanism. However, there are striking differences among the mechanisms of IRES-mediated translation initiation on picornaviruses, HCV, and CrPV mRNAs. Picornaviruses require the same canonical set of translation-initiation factors as required for cellular mRNAs, with the exception of eIF4E. (The picornavirus, hepatitis A virus is the sole exception in this regard because its translation is dependent on eIF4E) [366–

368]. The function of the cap structure and eIF4E in translation of picornavirus mRNA is supplanted by a direct interaction of eIF4G with the IRES, as shown for EMCV RNA. This interaction is enhanced by eIF4A [223]. Efficient translation of picornavirus mRNAs *in vitro* and possibly *in vivo* requires the participation of non-canonical initiation factors, which are termed ITAFs (IRES *trans*-acting factors) (see Ref. [244] for review).

Another class of viruses, exemplified by HCV, uses a much simpler mechanism of internal ribosome binding. This involves the direct binding of the 40S subunit, assisted by eIF2 and eIF3, to the IRES, which includes the initiator AUG [185]. This relatively simple, prokaryotic-like, mode of ribosome binding was a key factor for the successful solution of the structure of the HCV–IRES–40S ribosome complex by cryoelectron microscopy [183]. Yet, even a more basic and extraordinary mechanism of translation initiation that does not involve any initiation factors or initiator tRNA was described for some insect viruses such as cricket paralysis virus (CrPV) [365] and Plautia stali intestine virus [369]. In this case, 40S ribosomes bind directly to a non-AUG initiation codon. The large 60S subunit joins the complex without any requirement for eIF5 or eIF5B. Most remarkably, the P-site CCU triplet is not decoded, as it base-pairs with a pseudoknot sequence in the 18S rRNA. Thus, the first decoded triplet is in the A-site.

Internal ribosome binding is not restricted to viruses, which do not contain a cap structure, but also occurs on both viral and cellular mRNAs that contain a cap structure. Among the capped viral mRNAs that contain an IRES are those of retroviruses, including HIV-1 [370]. There are approximately 50 cellular mRNAs, which have been reported to contain an IRES (a database that lists these mRNAs is accessible at http://ifr31w3.toulouse.inserm.fr/IRESdatabase/). However, this list is certain to grow, as the estimated number of IRES-containing mRNAs is ~10% of the total mRNA population. This estimate is based on the number of mRNAs that are resistant to inhibition of translation following poliovirus infection, which causes a dramatic shut-off of cap-dependent host protein synthesis [371]. An important common feature of IRES-containing cellular mRNAs is that they encode for proteins which play key roles in cell growth, proliferation, and differentiation, e.g., fibroblast growth factor-2 (FGF-2), Myc, vascular endothelial growth factor (VEGF) and inhibition of apoptosis factor (XIAP).

A very interesting and evolving aspect of cellular IRESs is their function in translational control. IRES-containing cellular mRNAs can be translated to differing extents by the cap-dependent mechanism, but the IRES allows for translation under conditions where cap-dependent translation is inhibited. Inhibition of cap-dependent translation occurs under diverse conditions of stress, such as apoptosis, serum starvation, hypoxia and γ-irradiation [372]. Thus, proteins, which are important for cell function under these conditions are often encoded by IRES-containing mRNAs. For example, VEGF and the transcription factor known as hypoxia-inducible factor-1 (HIF-1) are required for cell growth under hypoxic conditions, where cap-dependent translation is inhibited. Similarly, Myc and Apaf-1 are required during apoptosis. IRES-mediated translation also can be regulated during the cell cycle. Cap-dependent

translation is repressed during mitosis as a result of 4E-BP dephosphorylation [373]. Short-lived proteins that are essential for cell survival must be synthesized during mitosis. Two such proteins whose mRNAs contain a mitosis-restricted IRES were identified: ornithine decarboxylase [370, 374] and PITSLRE protein kinase [375]. Retroviruses stimulate cells to become arrested in mitosis. Consistent with this, translation of retrovirus and HCV mRNAs, which is IRES-dependent, is enhanced in mitotically arrested cells [376].

Finally, there are additional mechanisms of translation initiation that seem to be hybrids of cap- and IRES-dependent mechanisms, known collectively as ribosomal shunting. Ribosome binding to these mRNAs is cap-dependent and entails a period of scanning; however, the scanning ribosomes then physically by-pass relatively large stretches of the 5′-UTR and rebind to the mRNA in the vicinity of the start codon. Examples of initiation by shunting include the 35S mRNA of the cauliflower mosaic virus, major late mRNAs of adenovirus and Sendai virus mRNA [377]. Just as in the case of IRES function, there appears to be a diversity of *cis*-acting sequences and structures in the mRNA that stimulate shunting on these viral mRNAs. It is probable that shunting will also be found to operate in cellular mRNAs.

7.2.10 Future Prospects

There is now a firm fundamental understanding of the mechanism of translation initiation in eukaryotes. Much of this knowledge has been traditionally obtained through the use of genetic analysis in yeast and biochemical assays in mammalian systems. More recently, a great deal of progress has been made by solving the 3D structures of individual translation factors, collecting vast new information on the interactions among initiation factors, and the analysis of functional multi-subunit complexes. Indeed, it is possible that recruitment of ribosomes to mRNAs occurs in just a two-step process, whereby ribosomes first interact with eIF1A and a multi-factor initiation factor complex that includes eIF1, eIF2, eIF3, eIF5, and Met-tRNA$_i^{Met}$, which is then recruited to the mRNA, via bridging with eIF4F. The new data provide a strong basis for studying structure–function relationships in the translation-initiation machinery. The next important advance in structural studies of translation will be solving the 3D structures of stable multi-subunit initiation factors (such as eIF3 and eIF2). Cryoelectron microscopy studies are already underway to determine the structures of these factors. It would certainly be more challenging to obtain structures of transient complexes including eIF3 in a complex with eIF5, eIF1, and eIF2 or the eIF4F complex. The ultimate goal of the structural studies is undoubtedly to obtain the 3D structure of the ribosomes complexed with initiation factors and mRNA.

An important, but difficult, issue to address in translational control mechanisms is how are the phosphorylation states of different initiation factors integrated to effect the translation of a specific set(s) of mRNAs. For example, serum simultaneously stimulates the phosphorylation of eIF4E, eIF4B, eIF4G, the eIF4E-repressor proteins 4E-BPs and ribosomal protein S6. However, it is unclear, except for 4E-BPs,

how (and if) phosphorylation of other translation components affects translation in serum-stimulated cells. The ultimate manner to assess the importance of phosphorylation of these factors is to generate 'knock-in' mutant mice and cells expressing nonphosphorylatable forms of the relevant factors. It is probable that general translation rates would not be affected by the abolition of eIF4B, and eIF4G phosphorylation, as was documented for eIF4E: as described above, eIF4E phosphorylation can be abrogated in mammalian cells without any effect on global translation rates [302]. However, mutation of the single phosphorylation site on eIF4E in flies causes a reduction in size and development of the flies [300]. The identification of mRNAs, whose translation is affected by the phosphorylation of initiation factors, would necessitate polysome profiling in combination with microarray analysis [378].

Another important area of future research involves uncovering the full range of biological processes, which are affected by eIF2α phosphorylation. It is now abundantly clear that the paradigm established in yeast for the mechanism of stimulation of *GCN4* translation upon eIF2α phosphorylation [15] applies to all organisms. One of the intriguing examples is the selective increase in the translation of activating transcription factor 4 (ATF4) following the phosphorylation of eIF2α by PERK under ER stress conditions. ATF4 in turn regulates the unfolded protein response (UPR) to stress [379]. The stress response is implicated in the etiology of many diseases, including diabetes. Knockout mice for PERK develop diabetes before reaching 1 month of age. Indeed, a familial diabetes mellitus disease known as Wolcott–Rallison syndrome (WRS), an autosomal recessive disorder, is caused by a mutation in PERK [380]. A fascinating connection between stress, eIF2α phosphorylation and disease was recently proposed to explain the pathophysiology of two linked inherited diseases: leukoencephalopathy with VWM, which is a brain disease, and premature ovarian failure (OF). These diseases can be caused by mutations in each of the five different subunits of eIF2B [122, 123, 381]. It stands to reason that the mutations in eIF2B increase the susceptibility of particular organs such as brain and ovaries to cellular stress. Other eIF2α kinases also are involved in disease development. In addition to functioning as a major player in the host antiviral defense arsenal, PKR also appears to function as a tumor suppressor [382]. Importantly, investigators are taking advantage of these attributes to target preferentially cancer cells, which have lost PKR or its upstream effectors due to mutations, for killing by oncolytic viruses [383].

With regard to human disease, phosphorylation of 4E-BPs has been extensively studied in relation to cell growth and its possible relevance to cancer development. Rapamycin, which inhibits 4E-BP phosphorylation is an anticancer drug candidate, whose antitumorigenic activity could be mediated by its activity on 4E-BP. Consistent with this idea, eIF4E and eIF4G are overexpressed in a large number of tumors, and they can oncogenically transform rodent cells in culture [384, 385].

A very promising and nascent research area in translational control concerns synaptic plasticity. This term refers to the ability of individual synapses in a neuron to undergo enduring changes in strength in response to experience. These changes play key roles in learning and memory. Synaptic plasticity is affected by local translation

(see Refs. [386, 387] for reviews). It has been known for two decades that ribosomes and translation factors are localized in dendrites beneath the post-synaptic sites. Local protein synthesis is required for the development of long-term potentiation (LTP) and long-term depression (LTD), which are associated with memory. The exact mechanism of translational up-regulation in synapses is not known, but several signaling pathways that are responsible for the phosphorylation of translation-initiation factors are involved. In particular, the PI3K/Akt-PKB/FRAP-mTOR pathway has been implicated, as rapamycin inhibits both this signaling pathway and LTP and conversely both are stimulated by brain-derived growth factor (BDNF) [388, 389]. An alternative mechanism to activate local translation in synapses is via increased polyadenylation of αCaMKII in response to different stimuli, such as light (for dark-reared rats) [390] or NMDA [391].

One of the most revolutionary, but controversial, concepts in translation is the possibility of nuclear translation in eukaryotes. The current dogma is that the mRNA can be translated only after it is exported from the nucleus to the cytoplasm. However, a recent report concluded that translation occurs in the nucleus [392]. This idea is attractive because it provides the simplest explanation for the mechanism of nonsense-mediated decay (NMD) in mammalian cells and its association with nuclear events. NMD of a subset of mRNAs in mammalian cells is physically associated with the nucleus and is affected by splicing. Thus, the first round of nuclear translation could provide a surveillance or proofreading mechanism prior to mRNA export [235]. However, there is a raging debate as to whether nuclear translation does indeed occur [393, 394]. One of the arguments consistent with nuclear translation is the existence of a pool of initiation and elongation factors in the nucleus. However, not all of the known initiation factors are found in the nucleus [393]. Thus, if translation takes place in the nucleus it is imperative to identify the factors that are involved. If translation happens not to occur in the nucleus then it would be important to understand the nuclear functions of initiation factors, such as eIF4E and eIF4G.

Based on the immense assembled knowledge on the architecture of the translation-initiation apparatus and the mechanistic insights into translation, it is anticipated that the next decade will reveal further progress in understanding the important contributions of this machinery to complex biological processes ranging from cell growth and development to memory and metabolism. Such progress will undoubtedly impinge on the efforts to cure major diseases such as cancer, diabetes, obesity and Alzheimer.

Acknowledgements

We thank Leos Valasek, Klaus Nielsen and Christie Hamilton for helpful comments on the paper and assistance in preparation of figures. We also thank Felecia Johnson for help in preparing the paper.

References

1 C. Gualerzi, L. Brandi, E. Caserta et al.: in *The Ribosome. Structure, Function, Antibiotics, and Cellular Interactions,* eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, 477–494.

2 N. C. Kyrpides, C. R. Woese, *Proc. Natl. Acad. Sci. USA* **1998**, *95,* 224–228.

3 M. Sette, P. van Tilborg, R. Spurio, et al., *EMBO J.* **1997**, *16,* 1436–1443.

4 C. M. Fletcher, T. V. Pestova, C. U. T. Hellen et al., *EMBO J.* **1999**, *18,* 2631–2639.

5 J. H. Lee, S. K. Choi, A. Roll-Mecak et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96,* 4342–4347.

6 S. K. Choi, D. S. Olsen, A. Roll-Mecak et al., *Mol. Cell. Biol.* **2000**, *20,* 7183–7191.

7 R. Boelens, C. O. Gualerzi, *Curr. Protein Pept. Sci.* **2002**, *3,* 107–119.

8 N. Sonenberg, T. E. Dever, *Curr. Opin. Struct. Biol.* **2003**, *13,* 56–63.

9 M. C. Ganoza, M. C. Kiel, H. Aoki, *Microbiol. Mol. Biol. Rev.* **2002**, *66,* 460–485.

10 H. A. Kang, J. W. Hershey, *J. Biol. Chem.* **1994**, *269,* 3934–3940.

11 N. Tolstrup, C. W. Sensen, R. A. Garrett et al., *Extremophiles* **2000**, *4,* 175–179.

12 R. J. Jackson: in *Translation in Eukaryotes,* ed. Trachsel, CRC Press, Boca Raton, FL 1991, 193–230.

13 C. G. Proud, *Curr. Toic Cell. Regul.* **1992**, *32,* 243–369.

14 J.-J. Chen: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 529–546.

15 A. G. Hinnebusch: in *Translational Control of Gene Expression,* eds N. Sonenberg, J.W.B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 185–243.

16 R. Kaufman: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 503–527.

17 D. Ron, H. P. Harding: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 547–560.

18 J. Deng, H. P. Harding, B. Raught et al., *Curr. Biol.* **2002**, *12,* 1279–1286.

19 A. G. Hinnebusch: in *Translational Control,* eds J. W. B. Hershey, M. B. Mathews and N. Sonenberg, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1996, 199–244.

20 E. M. Hannig, A. M. Cigan, B. A. Freeman et al., *Mol. Cell. Biol.* **1992**, *13,* 506–520.

21 D. Tzamarias, I. Roussou, G. Thireos, *Cell* **1989**, *57,* 947–954.

22 A. M. Cigan, M. Foiani, E. M. Hannig et al., *Mol. Cell. Biol.* **1991**, *11,* 3217–3228.

23 M. Foiani, A. M. Cigan, C. J. Paddon et al., *Mol. Cell. Biol.* **1991**, *11,* 3203–3216.

24 D. R. Dorris, F. L. Erickson, E. M. Hannig, *EMBO J.* **1995**, *14,* 2239–2249.

25 N. P. Williams, A. G. Hinnebusch, T. F. Donahue, *Proc. Natl. Acad. Sci. USA* **1989**, *86,* 7515–7519.

26 T. E. Dever, W. Yang, S. str•m et al., *Mol. Cell. Biol.* **1995**, *15,* 6351–6363.

27 S. Harashima, A. G. Hinnebusch, *Mol. Cell. Biol.* **1986**, *6,* 3990–3998.

28 A. M. Cigan, L. Feng, T. F. Donahue, *Science* **1988**, *242,* 93–97.

29 R. Benne, J. W. B. Hershey, *Proc. Natl. Acad. Sci. USA* **1976**, *73,* 3005–3009.

30 U. A. Bommer, G. Lutsch, J. Stahl et al., *Biochimie* **1991**, *73,* 1007–1019.

31 S. Srivastava, A. Verschoor, J. Frank, *J. Mol. Biol.* **1992**, *220,* 301–304.

32 J. Chaudhuri, D. Chowdhury, U. Maitra, *J. Biol. Chem.* **1999**, *274,* 17975–17980.

33 R. Majumdar, A. Bandyopadhyay, U. Maitra, *J. Biol. Chem.* **2003**, *278,* 6580–6587.

34 K. Asano, J. Clayton, A. Shalev et al., *Genes Dev.* **2000,** *14,* 2534–2546.
35 H. Trachsel, T. Staehelin, *Biochim. Biophys. Acta* **1979,** *565,* 305–314.
36 D. T. Peterson, W. C. Merrick, B. Safer, *J. Biol. Chem.* **1979,** *254,* 2509–2519.
37 R. Benne, J. W. B. Hershey, *J. Biol. Chem.* **1978,** *253,* 3078–3087.
38 J. Chaudhuri, A. Chakrabarti, U. Maitra, *J. Biol. Chem.* **1997,** *272,* 30975–30983.
39 U. von Pawel-Rammingen, S. Aström, A. S. Byström, *Mol. Cell. Biol.* **1992,** *12,* 1432–1442.
40 S. U. Astrom, U. von Pawel-Rammingen, A. S. Bystrom, *J. Mol. Biol.* **1993,** *233,* 43–58.
41 H. J. Drabkin, B. Helk, U. L. RajBhandary, *J. Biol. Chem.* **1993,** *268,* 25221–25228.
42 D. Farruggio, J. Chaudhuri, U. Maitra et al., *Mol. Cell. Biol.* **1996,** *16,* 4248–4256.
43 H. J. Drabkin, M. Estrella, U. L. Rajbhandary, *Mol. Cell. Biol.* **1998,** *18,* 1459–1466.
44 S. Kiesewetter, G. Ott, M. Sprinzl, *Nucleic Acids Res.* **1990,** *18,* 4677–4682.
45 S. U. Astrom, A. S. Bystrom, *Cell* **1994,** *79,* 535–546.
46 C. Forster, K. Chakraburtty, M. Sprinzl, *Nucleic Acids Res.* **1993,** *21,* 5679–5683.
47 S. U. Astrom, M. E. Nordlund, F. L. Erickson et al., *Mol. Gen. Genet.* **1999,** *261,* 967–976.
48 T. Wagner, M. Gross, P. B. Sigler, *J. Biol. Chem.* **1984,** *259,* 4706–4709.
49 H. J. Drabkin, U. L. RajBhandary, *Mol. Cell. Biol.* **1998,** *18,* 5140–5147.
50 N. J. Gaspar, T. G. Kinzy, B. J. Scherer et al., *J. Biol. Chem.* **1994,** *269,* 3415–3422.
51 E. Schmitt, S. Blanquet, Y. Mechulam, *EMBO J.* **2002,** *21,* 1821–1832.
52 F. L. Erickson, E. M. Hannig, *EMBO J.* **1996,** *15,* 6311–6320.
53 H. Huang, H. Yoon, E. M. Hannig et al., *Genes Dev.* **1997,** *11,* 2396–2413.
54 F. L. Erickson, L. D. Harding, D. R. Dorris et al., *Mol. Gen. Genet.* **1997,** *253,* 711–719.
55 H. Trachsel: in *Translational Control,* eds J. W. B. Hershey, M. B. Mathews and N. Sonenberg, Cold Spring Harbor Laboratory Press, Plainview, NY 1996, 113–138.
56 A. Flynn, S. Oldfield, C. G. Proud, *Biochim. Biophys. Acta* **1993,** *1174,* 117–121.
57 G. M. Thompson, E. Pacheco, E. O. Melo et al., *Biochem. J.* **2000,** *347,* 703–709.
58 S. Cho, D. W. Hoffman, *Biochemistry* **2002,** *41,* 5730–5742.
59 T. F. Donahue, A. M. Cigan, E. K. Pabich et al., *Cell* **1988,** *54,* 621–632.
60 V. K. Pathak, P. J. Nielsen, H. Trachsel et al., *Cell* **1988,** *54,* 633–639.
61 X. Ye, D. R. Cavener, *Gene* **1994,** *142,* 271–274.
62 B. Castilho-Valavicius, G. M. Thompson, T. F. Donahue, *Gene Expr.* **1992,** *2,* 297–309.
63 N. N. Hashimoto, L. S. Carnevalli, B. A. Castilho, *Biochem. J.* **2002,** *367,* 359–368.
64 S. Das, T. Maiti, K. Das et al., *J. Biol. Chem.* **1997,** *272,* 31712–31718.
65 H. Rosen, G. Di Segni, R. Kaempfer, *J. Biol. Chem.* **1982,** *257,* 946–952.
66 R. Gonsky, D. Itamar, R. Harary et al., *Biochimie,* **1992,** *74,* 427–434.
67 A. Flynn, I. N. Shatsky, C. G. Proud et al., *Biochim. Biophys. Acta* **1994,** *1219,* 293–301.
68 J. P. Laurino, G. M. Thompson, E. Pacheco et al., *Mol. Cell. Biol.* **1999,** *19,* 173–181.
69 K. Asano, T. Krishnamoorthy, L. Phan et al., *EMBO J.* **1999,** *18,* 1673–1688.
70 J. Chaudhuri, K. Das, U. Maitra, *Biochemistry* **1994,** *33,* 4794–4799.
71 S. Das, U. Maitra, *Mol. Cell. Biol.* **2000,** *20,* 3942–3950.
72 K. Asano, A. Shalev, L. Phan et al., *EMBO J.* **2001,** *20,* 2326–2337.
73 C. J. Bult, O. White, E. J. Olsen et al., *Science* **1996,** *273,* 1058–1073.
74 H.-P. Klenk, R. A. Clayton, J.-F. Tomb et al., *Nature* **1997,** *390,* 364–370.
75 D. R. Smith, L. A. Doucette-Stamm, C. Deloughery et al., *J. Bacteriol.* **1997,** *179,* 7135–7155.
76 J. W. B. Hershey, *Annu. Rev. Biochem.* **1991,** *60,* 717–755.

77 T. E. Dever, L. Feng, R. C. Wek et al., *Cell* **1992**, *68*, 585–596.
78 A. M. Cigan, E. K. Pabich, L. Feng et al., *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2784–2788.
79 S. Qu, D. R. Cavener, *Gene* **1994**, *140*, 239–242.
80 H. Ernst, R. F. Duncan, J. W. B. Hershey, *J. Biol. Chem.* **1987**, *262*, 1206–1212.
81 M. Bycroft, T. J. P. Hubbard, M. Proctor et al., *Cell* **1997**, *88*, 235–242.
82 M. C. Nonato, J. Widom, J. Clardy, *J. Biol. Chem.* **2002**, *277*, 17057–17061.
83 C. R. Vazquez de Aldana, T. E. Dever, A. G. Hinnebusch, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7215–7219.
84 T. Krishnamoorthy, G. D. Pavitt, F. Zhang et al., *Mol. Cell. Biol.* **2001**, *21*, 5018–5030.
85 F. L. Erickson, J. Nika, S. Rippel et al., *Genetics* **2001**, *158*, 123–132.
86 J. Nika, S. Rippel, E. M. Hannig, *J. Biol. Chem.*, **2000**, *276*, 1051–1056.
87 G. D. Pavitt, K. V. A. Ramaiah, S. R. Kimball et al., *Genes Dev.* **1998**, *12*, 514–526.
88 L. Feng, H. Yoon, T. F. Donahue, *Mol. Cell. Biol.* **1994**, *14*, 5139–5153.
89 A. G. Rowlands, R. Panniers, E. C. Henshaw, *J. Biol. Chem.* **1988**, *263*, 5526–5533.
90 K. L. Manchester, *Biochem. Biophys. Res. Commun.* **1997**, *239*, 223–227.
91 J. N. Dholakia, A. J. Wahba, *J. Biol. Chem.* **1989**, *264*, 546–550.
92 J. Nika, W. Yang, G. D. Pavitt et al., *J. Biol. Chem.* **2000**, *275*, 26011–26017.
93 D. J. Goss, L. J. Parkhurst, *J. Biol. Chem.* **1984**, *259*, 7374–7377.
94 D. D. Williams, N. T. Price, A. J. Loughlin et al., *J. Biol. Chem.* **2001**, *276*, 24697–24703.
95 J. R. Fabian, S. R. Kimball, N. K. Heinzinger et al., *J. Biol. Chem.* **1997**, *272*, 12359–12365.
96 K. L. Manchester, *Biochem. Biophys. Res. Commun.* **2001**, *289*, 643–646.
97 E. M. Hannig, A. G. Hinnebusch, *Mol. Cell. Biol.* **1988**, *8*, 4808–4820.
98 S. R. Kimball, J. R. Fabian, G. D. Pavitt et al., *J. Biol. Chem.* **1998**, *273*, 12841–12845.
99 B. L. Craddock, C. G. Proud, *Biochem. Biophys. Res. Commun.* **1996**, *220*, 843–847.
100 D. D. Williams, G. D. Pavitt, C. G. Proud, *J. Biol. Chem.* **2001**, *276*, 3733–3742.
101 E. Gomez, G. D. Pavitt, *Mol. Cell. Biol.* **2000**, *20*, 3965–3976.
102 E. Gomez, S. S. Mohammad, G. D. Pavitt, *EMBO J.* **2002**, *21*, 5292–5301.
103 X. Wang, F. E. Paulin, L. E. Campbell et al., *EMBO J.* **2001**, *20*, 4349–4359.
104 S. R. Kimball, N. K. Heinzinger, R. L. Horetsky et al., *J. Biol. Chem.* **1998**, *273*, 3039–3044.
105 E. V. Koonin, *Protein Sci.* **1995**, *4*, 1608–1617.
106 T. G. Anthony, J. R. Fabian, S. R. Kimball et al., *Biochim. Biophys. Acta* **2000**, *1492*, 56–62.
107 N. S. B. Thomas, R. L. Matts, R. Petryshyn et al., *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 6998–7002.
108 G. D. Pavitt, W. Yang, A. G. Hinnebusch, *Mol. Cell. Biol.* **1997**, *17*, 1298–1313.
109 A. G. Rowlands, K. S. Montine, E. C. Henshaw et al., *Eur. J. Biochem.* **1988**, *175*, 93–99.
110 A. M. Cigan, J. L. Bushman, T. R. Boal et al., *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5350–5354.
111 C. J. Paddon, E. M. Hannig, A. G. Hinnebusch, *Genetics* **1989**, *122*, 551–559.
112 J. L. Bushman, A. I. Asuru, R. L. Matts et al., *Mol. Cell. Biol.* **1993**, *13*, 1920–1932.
113 W. Yang, A. G. Hinnebusch, *Mol. Cell. Biol.* **1996**, *16*, 6603–6616.
114 C. R. Vazquez de Aldana, A. G. Hinnebusch, *Mol. Cell. Biol.* **1994**, *14*, 3208-3222
115 M. V. Davies, M. Furtado, J. W. B. Hershey et al., *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9163–9167.

116 R. J. Kaufman, M. V. Davies, V. K. Pathak et al., *Mol. Cell. Biol.* **1989,** *9,* 946–958.
117 S. Y. Choi, B. J. Scherer, J. Schnier et al., *J. Biol. Chem.* **1992,** *267,* 286–293.
118 P. Murtha-Riel, M. V. Davies, B. J. Scherer et al., *J. Biol. Chem.* **1993,** *268,* 12946–12951.
119 K. V. Ramaiah, M. V. Davies, J. J. Chen et al., *Mol. Cell. Biol.* **1994,** *14,* 4546–4553.
120 A. Sudhakar, T. Krishnamoorthy, A. Jain et al., *Biochemistry* **1999,** *38,* 15398–15405.
121 N. S. B. Thomas, R. L. Matts, D. H. Levin et al., *J. Biol. Chem.* **1985,** *260,* 9860–9866.
122 P. A. J. Leegwater, G. Vermeulen, A. A. M. Konst et al., *Nat. Genet.* **2001,** *29,* 383–388.
123 M. S. van der Knaap, P. A. Leegwater, A. A. Konst et al., *Ann. Neurol.* **2002,** *51,* 264–270.
124 A. De Benedetti, C. Baglioni, *J. Biol. Chem.* **1983,** *258,* 14556–14562.
125 M. Gross, R. Redman, D. A. Kaplansky, *J. Biol. Chem.* **1985,** *260,* 9491-9500.
126 M. Gross, M. Wing, C. Rundquist et al., *J. Biol. Chem.* **1987,** *262,* 6899–6907.
127 K. V. A. Ramaiah, R. S. Dhindsa, J. J. Chen et al., *Proc. Natl. Acad. Sci. USA,* **1992,** *89,* 12063–12067.
128 A. Chakrabarti, U. Maitra, *J. Biol. Chem.* **1992,** *267,* 12964–12972.
129 P. Raychaudhuri, U. Maitra, *J. Biol. Chem.* **1986,** *261,* 7723–7728.
130 P. P. Mueller, P. Grueter, A. G. Hinnebusch et al., *J. Biol. Chem.* **1998,** *273,* 32870–32877.
131 N. K. Gupta, A. L. Roy, M. K. Nag et al.: in *Post-transcriptional Control of Gene Expression,* eds J. E. G. McCarthy and M. F. Tuite, Springer,- Berlin, Heidelberg 1990, 521–526, *Vol. H49*
132 M. Salimans, H. Goumans, H. Amesz et al., *Eur. J. Biochem.* **1984,** *145,* 91–98.
133 H. Trachsel, B. Erni, M. H. Schreier et al., *J. Mol. Biol.* **1977,** *116,* 755–767.
134 J. Chaudhuri, K. Si, U. Maitra, *J. Biol. Chem.* **1997,** *272,* 7883–7891.
135 M. A. Algire, D. Maag, P. Savio et al., *RNA* **2002,** *8,* 382–397.
136 H. A. Thompson, I. Sadnik, J. Scheinbuks et al., *Biochemistry* **1977,** *16,* 2221–2230.
137 L. H. Hartwell, C. S. McLaughlin, *Proc. Natl. Acad. Sci. USA* **1969,** *62,* 468–474.
138 P. Danaie, B. Wittmer, M. Altmann et al., *J. Biol. Chem.* **1995,** *270,* 4288–4292.
139 L. Phan, X. Zhang, K. Asano et al., *Mol. Cell. Biol.* **1998,** *18,* 4935–4946.
140 E. A. Burks, P. P. Bezerra, H. Le et al., *J. Biol. Chem.* **2001,** *276,* 2122–2131.
141 J. R. Greenberg, L. Phan, Z. Gu et al., *J. Biol. Chem.* **1998,** *273,* 23485–23494.
142 L. Valásek, H. Trachsel, J. Haek et al., *J. Biol. Chem.* **1998,** *273,* 21253–21260.
143 T. Naranda, M. Kainuma, S. E. McMillan et al., *Mol. Cell. Biol.* **1997,** *17,* 145–153.
144 P. Hanachi, J. W. B. Hershey, H. P. Vornlocher, *J. Biol. Chem.* **1999,** *274,* 8546–8553.
145 M.-H. Verlhac, R.-H. Chen, P. Hanachi et al., *EMBO J.* **1997,** *16,* 6812–6822.
146 K. Asano, L. Phan, J. Anderson et al., *J. Biol. Chem.* **1998,** *273,* 18573–18585.
147 H. P. Vornlocher, P. Hanachi, S. Ribeiro et al., *J. Biol. Chem.* **1999,** *274,* 16802–16812.
148 L. Valásek, J. Haek, H. Trachsel et al., *J. Biol. Chem.* **1999,** *274,* 27567–27572.
149 L. Valásek, L. Phan, L. W. Schoenfeld et al., *EMBO J.* **2001,** *20,* 891–904.
150 L. Valásek, J. Haek, K. H. Nielsen et al., *J. Biol. Chem.* **2001,** *276,* 43351–43360.
151 A. Shalev, L. Valásek, C. A. Pise-Masison et al., *J. Biol. Chem.* **2001,** *276,* 34948–34957.
152 Y. Akiyoshi, J. Clayton, L. Phan et al., *J. Biol. Chem.* **2000,** *276,* 10056–10062.
153 A. Bandyopadhyay, T. Matsumoto, U. Maitra, *Mol. Biol. Cell* **2000,** *11,* 4005–4018.
154 C. R. Chen, Y. C. Li, J. Chen et al., *Proc. Natl. Acad. Sci. USA* **1999,** *96,* 517–522.

155 R. Crane, R. Craig, R. Murray et al., *Mol. Biol. Cell* **2000,** *11,* 3993–4003.
156 A. Bandyopadhyay, V. Lakshmanan, T. Matsumoto et al., *J. Biol. Chem.* **2002,** *277,* 2360–2377.
157 T. Naranda, S. E. MacMillan, J. W. B. Hershey, *J. Biol. Chem.* **1994,** *269,* 32286–32292.
158 M. T. Garcia-Barrio, T. Naranda, R. Cuesta et al., *Genes Dev.* **1995,** *9,* 1781–1796.
159 J. Anderson, L. Phan, R. Cuesta et al., *Genes Dev.* **1998,** *12,* 3650–3662.
160 O. Calvo, R. Cuesta, J. Anderson et al., *Mol. Cell. Biol.* **1999,** *19,* 4167–4181.
161 J. Anderson, L. Phan *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 5173–5178.
162 L. Valásek, K. H. Nielsen, A. G. Hinnebusch, *EMBO J.* **2002,** *21,* 5886–5898.
163 L. Phan, L. W. Schoenfeld, L. Valásek et al., *EMBO J.* **2001,** *20,* 2954–2965.
164 D. R. H. Evans, C. Rasmussen, P. J. Hanic-Joyce et al., *Mol. Cell. Biol.* **1995,** *15,* 4525–4535.
165 C. Morris-Desbois, V. Bochard, C. Reynaud et al., *J. Cell Sci.* **1999,** *112,* 3331–3342.
166 A. Yahalom, T. H. Kim, E. Winter et al., *J. Biol. Chem.* **2000,** *276,* 334–340.
167 H. C. Yen, E. C. Chang, *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 14370–14375.
168 T. Naranda, S. E. MacMillan, T. F. Donahue et al., *Mol. Cell. Biol.* **1996,** *16,* 2307–2313.
169 A. Bandyopadhyay, U. Maitra, *Nucleic Acids Res.* **1999,** *27,* 1331–1337.
170 H. J. Yoon, T. F. Donahue, *Mol. Cell. Biol.* **1992,** *12,* 248–260.
171 T. V. Pestova, V. G. Kolupaeva, *Genes Dev.* **2002,** *16,* 2906–2922.
172 B. J. Lamphear, R. Kirchweger, T. Skern et al., *J. Biol. Chem.* **1995,** *270,* 21975–21983.
173 N. Methot, M. S. Song, N. Sonenberg, *Mol. Cell. Biol.* **1996,** *16,* 5328–5334.
174 J. W. B. Hershey, W. C. Merrick: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 33–88.
175 P. Westermann, O. Nygard, *Nucleic Acids Res.* **1984,** *12,* 8887–8897.
176 K. L. Block, H. P. Vornlocher, J. W. B. Hershey, *J. Biol. Chem.* **1998,** *273,* 31901–31908.
177 K. Asano, H.-P. Vornlocher, N. J. Richter-Cook et al., *J. Biol. Chem.* **1997,** *272,* 27042–27052.
178 K. Asano, T. G. Kinzy, W. C. Merrick et al., *J. Biol. Chem.* **1997,** *272,* 1101–1109.
179 O. Nygard, P. Westermann, *Nucleic Acids Res.* **1982,** *10,* 1327–1334.
180 D. V. Sizova, V. G. Kolupaeva, T. V. Pestova et al., *J. Virol.* **1998,** *72,* 4775–4782.
181 E. Buratti, S. Tisminetzky, M. Zotti et al., *Nucleic Acids Res.* **1998,** *26,* 3179–3187.
182 J. S. Kieft, K. Zhou, R. Jubin et al., *RNA,* **2001,** *7,* 194–206.
183 C. M. Spahn, J. S. Kieft, R. A. Grassucci et al., *Science,* **2001,** *291,* 1959–1962.
184 G. Lutsch, R. Benndorf, P. Westermann et al., *Eur. J. Cell Biol.* **1986,** *40,* 257–265.
185 T. V. Pestova, I. N. Shatsky, S. P. Fletcher et al., *Genes Dev.* **1998,** *12,* 67–83.
186 L. Valásek, A. Mathew, B. S. Shin et al., *Genes Dev,* **2003,** *17,* 786–799.
187 C. M. Spahn, R. Beckmann, N. Eswar et al., *Cell,* **2001,** *107,* 373–386.
188 C. L. Wei, M. Kainuma, J. W. Hershey, *J. Biol. Chem.* **1995,** *270,* 22788–22794.
189 M. Kainuma, J. W. B. Hershey, *Biochimie,* **2001,** *83,* 505–514.
190 A. Thomas, H. Goumans, H. O. Voorma et al., *Eur. J. Biochem.* **1980,** *107,* 39–45.
191 J. B. Battiste, T. V. Pestova, C. U. T. Hellen et al., *Mol. Cell,* **2000,** *5,* 109–119.
192 D. Moazed, R. R. Samaha, C. Gualerzi et al., *J. Mol. Biol.* **1995,** *248,* 207–210.
193 A. P. Carter, W. M. Clemons, Jr. D. E. Brodersen et al., *Science,* **2001,** *291,* 498–501.
194 C. L. Wei, S. E. MacMillan, J. W. B. Hershey, *J. Biol. Chem.* **1995,** *270,* 5764–5771.

195 D. S. Olsen, E.M. S. et al., *EMBO J.* **2003,** *22,* 193–204.
196 T. V. Pestova, S. I. Borukhov, C. U. T. Hellen, *Nature,* **1998,** *394,* 854–859.
197 A. Marintchev, V. G. Kolupaeva, T. V. Pestova et al., *Proc. Natl. Acad. Sci. USA,* **2003,** *100,* 1535–1540.
198 A. Roll-Mecak, C. Cao, T. E. Dever et al., *Cell,* **2000,** *103,* 781–792.
199 C. O. Gualerzi, C. L. Pon, *Biochemistry,* **1990,** *29,* 5881–5889.
200 J. M. Palacios Moreno, L. Drskjotersen, J. E. Kristensen et al., *FEBS Lett.* **1999,** *455,* 130–134.
201 S. K. Choi, J. H. Lee, W. L. Zoll et al., *Science,* **1998,** *280,* 1757–1760.
202 A. J. Shatkin, *Cell,* **1976,** *9,* 645–653.
203 Y. Furuichi, A. J. Shatkin, *Adv. Virus Res.* **2000,** *55,* 135–184.
204 A. C. Gingras, B. Raught, N. Sonenberg, *Ann. Rev. Biochem.* **1999,** *68,* 913–963.
205 N. Sonenberg, M. A. Morgan, W. C. Merrick et al., *Proc. Natl. Acad. Sci. USA,* **1978,** *75,* 4843–4847.
206 N. Sonenberg, M. A. Morgan, D. Testa et al., *Nucleic Acids Res.* **1979,** *7,* 15–29.
207 M. Altmann, P. P. Mueller, J. Pelletier et al., *J. Biol. Chem.* **1989,** *264,* 12145–12147.
208 J. Marcotrigiano, A. C. Gingras, N. Sonenberg et al., *Cell,* **1997,** *89,* 951–961.
209 H. Matsuo, H. Li, A. M. McGuire et al., *Nat. Struct. Biol.* **1997,** *4,* 717–724.
210 J. Marcotrigiano, A.-C. Gingras, N. Sonenberg et al., *Nucleic Acids Symp. Ser.* ed. **1997,** 8–11.
211 G. C. Scheper, B. van Kollenburg, J. Hu et al., *J. Biol. Chem.* **2002,** *277,* 3303–3309.
212 J. Zuberek, A. Wyslouch-Cieszynska, A. Niedzwiecka et al., *RNA,* **2003,** *9,* 52–61.
213 K. Tomoo, X. Shen, K. Okabe et al., *Biochem. J.* **2002,** *362,* 539–544.
214 A. Niedzwiecka, J. Marcotrigiano, J. Stepinski et al., *J. Mol. Biol.* **2002,** *319,* 615–635.
215 C. Goyer, M. Altmann, H. S. Lee et al., *Mol. Cell. Biol.* **1993,** *13,* 4860–4874.
216 T. Ohlmann, D. Prevot, D. Decimo et al., *J. Mol. Biol.* **2002,** *318,* 9–20.
217 M. Bushell, D. Poncet, W. E. Marissen et al., *Cell Death Differ.* **2000,** *7,* 628–636.
218 S. Morino, H. Imataka, Y. V. Svitkin et al., *Mol. Cell. Biol.* **2000,** *20,* 468–477.
219 S. Z. Tarun, A. B. Sachs, *EMBO J.* **1996,** *15,* 7168–7177.
220 H. Imataka, A. Gradi, N. Sonenberg, *EMBO J.* **1998,** *17,* 7480–7489.
221 H. Le, R. L. Tanguay, M. L. Balasta et al., *J. Biol. Chem.* **1997,** *272,* 16247–16255.
222 J. Marcotrigiano, I. B. Lomakin, N. Sonenberg et al., *Mol. Cell,* **2001,** *7,* 193–203.
223 I. B. Lomakin, C. U. Hellen, T. V. Pestova, *Mol. Cell. Biol.* **2000,** *20,* 6019–6029.
224 H. Imataka, N. Sonenberg, *Mol. Cell. Biol.* **1997,** *17,* 6940–6947.
225 S. Pyronnet, H. Imataka, A. C. Gingras et al., *EMBO J.* **1999,** *18,* 270–279.
226 A. J. Waskiewicz, J. C. Johnson, B. Penn et al., *Mol. Cell. Biol.* **1999,** *19,* 1871–1880.
227 E. De Gregorio, T. Preiss, M. W. Hentze, *EMBO J.* **1999,** *18,* 4865–4874.
228 T. V. Pestova, I. N. Shatsky, C. U. T. Hellen, *Mol. Cell. Biol.* **1996,** *16,* 6870–6878.
229 W. Li, G. J. Belsham, C. G. Proud, *J. Biol. Chem.* **2001,** *276,* 29111–29115.
230 N. L. Korneeva, B. J. Lamphear, F. L. Hennigan et al., *J. Biol. Chem.* **2001,** *276,* 2872–2879.
231 H. S. Yang, A. P. Jansen, A. A. Komar et al., *Mol. Cell. Biol.* **2003,** *23,* 26–37.
232 J. D. Lewis, E. Izaurralde, *Eur. J. Biochem.* **1997,** *247,* 461–469.
233 P. Fortes, T. Inada, T. Preiss et al., *Mol. Cell,* **2000,** *6,* 191–196.
234 L. McKendrick, E. Thompson, J. Ferreira et al., *Mol. Cell. Biol.* **2001,** *21,* 3632–3641.
235 Y. Ishigaki, X. Li, G. Serin et al., *Cell,* **2001,** *106,* 607–617.
236 C. Vilela, C. Velasco, M. Ptushkina et al., *EMBO J.* **2000,** *19,* 4372–4382.

237 P. E. C. Hershey, S. M. McWhirter, J. Gross et al., *J. Biol. Chem.* **1999,** *274,* 21297–21304.

238 J. Marcotrigiano, A. C. Gingras, N. Sonenberg et al., *Mol. Cell,* **1999,** *3,* 707–716.

239 S. Mader, H. Lee, A. Pause et al., *Mol. Cell. Biol.* **1995,** *15,* 4990–4997.

240 A. Haghighat, S. Mader, A. Pause et al., *EMBO J.* **1995,** *14,* 5701–5709.

241 N. Levy-Stumpf, L. P. Deiss, H. Berissi et al., *Mol. Cell. Biol.* **1997,** *17,* 1615–1625.

242 H. Imataka, S. Olsen, N. Sonenberg, *EMBO J.* **1997,** *16,* 817–825.

243 S. Yamanaka, K. S. Poksay, K. S. Arnold et al., *Genes Dev.* **1997,** *11,* 321–333.

244 C. U. Hellen, P. Sarnow, *Genes Dev.* **2001,** *15,* 1593–1612.

245 S. Henis-Korenblit, N. L. Strumpf, D. Goldstaub et al., *Mol. Cell. Biol.* **2000,** *20,* 496–506.

246 S. Henis-Korenblit, G. Shani, T. Sines et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 5400–5405.

247 P. Linder, P. P. Slonimski, *Nucleic Acids Res.* **1988,** *16,* 10359.

248 R. Duncan, J. W. Hershey, *J. Biol. Chem.* **1983,** *258,* 7228–7235.

249 T. von der Haar, J. E. McCarthy, *Mol. Microbiol.* **2002,** *46,* 531–544.

250 N. K. Tanner, P. Linder, *Mol. Cell,* **2001,** *8,* 251–262.

251 G. W. Rogers, Jr. N. J. Richter, W. C. Merrick et al., *J. Biol. Chem.* **1999,** *274,* 12236–12244.

252 G. W. Rogers, Jr. A. A. Komar, W. C. Merrick et al., *Prog. Nucleic Acid Res. Mol. Biol.* **2002,** *72,* 307–331.

253 F. Rozen, I. Edery, K. Meerovitch et al., *Mol. Cell. Biol.* **1990,** *10,* 1134–1144.

254 G. W. Rogers, Jr. N. J. Richter, W. F. Lima et al., *J. Biol. Chem.* **2001,** *276,* 30914–30922.

255 J. Pelletier, N. Sonenberg, *Cell,* **1985,** *40,* 515–526.

256 S. B. Baim, D. F. Pietras, D. C. Eustice et al., *Mol. Cell. Biol.* **1985,** *5,* 1839–1846.

257 Y. Svitkin, A. Pause, A. Haghighat et al., *RNA,* **2001,** *7,* 382–394.

258 J. Yoder-Hill, A. Pause, N. Sonenberg et al., *J. Biol. Chem.* **1993,** *268,* 5566–5573.

259 Q. Li, H. Imataka, S. Morino et al., *Mol. Cell. Biol.* **1999,** *19,* 7336–7346.

260 J. de la Cruz, I. Iost, D. Kressler et al., *Proc. Natl. Acad. Sci. USA* **1997,** *94,* 5201–5206.

261 R. Y. Chuang, P. L. Weaver, Z. Liu et al., *Science* **1997,** *275,* 1468–1471.

262 I. Iost, M. Dreyfus, P. Linder, *J. Biol. Chem.* **1999,** *274,* 17677–17683.

263 P. Leroy, P. Alzari, D. Sassoon et al., *Cell* **1989,** *57,* 549–559.

264 A. O. Noueiry, J. Chen, P. Ahlquist, *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 12985–12990.

265 J. Benz, H. Trachsel, U. Baumann, *Struct. Fold Des.* **1999,** *7,* 671–679.

266 E. R. Johnson, D. B. McKay, *RNA* **1999,** *5,* 1526–1534.

267 J. M. Caruthers, E. R. Johnson, D. B. McKay, *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 13080–13085.

268 R. M. Story, H. Li, J. N. Abelson, *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 1465–1470.

269 R. Coppolecchia, P. Buser, A. Stotz et al., *EMBO J.* **1993,** *12,* 4005–4011.

270 N. Methot, E. Rom, H. Olsen et al., *J. Biol. Chem.* **1997,** *272,* 1110–1116.

271 N. Methot, A. Pause, J. Hershey et al., *Mol. Cell. Biol.* **1994,** *14,* 2307–2316.

272 T. Naranda, W. B. Strong, J. Menaya et al., *J. Biol. Chem.* **1994,** *269,* 14465–14472.

273 N. Methot, G. Pickett, J. Keene et al., *RNA* **1996,** *2,* 38–50.

274 E. Martinez-Salas, S. L. Quinto, R. Ramos et al., *Biochimie,* **2002,** *84,* 755–763.

275 K. Ochs, L. Saleh, G. Bassili et al., *J. Virol.* **2002,** *76,* 2113–2122.

276 M. Altmann, B. Wittmer, N. Methot et al., *EMBO J.* **1995,** *14,* 3820–3827.

277 N. J. Richter-Cook, T. E. Dever, J. O. Hensold et al., *J. Biol. Chem.* **1998,** *273,* 7579–7587.

278 A. B. Sachs, R. W. Davis, R. D. Kornberg, *Mol. Cell. Biol.* **1987,** *7,* 3268–3276.

279 M. Gorlach, C. G. Burd, G. Dreyfuss, *Exp. Cell Res.* **1994,** *211,* 400–407.

280 N. K. Gray, J. M. Coller, K. S. Dickson et al., *EMBO J.* **2000,** *19,* 4723–4733.

281 A. Sachs: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 447–465.

282 R. C. Deo, J. B. Bonanno, N. Sonenberg et al., *Cell* **1999,** *98,* 835–845.

283 K. Khaleghpour, A. Kahvejian, G. De Crescenzo et al., *Mol. Cell. Biol.* **2001,** *21,* 5200–5213.

284 G. Kozlov, J. F. Trempe, K. Khaleghpour et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 4409–4413.

285 R. C. Deo, N. Sonenberg, S. K. Burley, *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 4414–4419.

286 D. R. Gallie, *Genes Dev.* **1991,** *5,* 2108–2116.

287 L. J. Otero, M. P. Ashe, A. B. Sachs, *EMBO J.* **1999,** *18,* 3153–3163.

288 N. Uchida, S. Hoshino, H. Imataka et al., *J. Biol. Chem.* **2002,** *277,* 50286–50292.

289 A. Searfoss, T. E. Dever, R. Wickner, *Mol. Cell. Biol.* **2001,** *21,* 4900–4908.

290 C. C. Wei, M. L. Balasta, J. Ren et al., *Biochemistry* **1998,** *37,* 1910–1916.

291 B. Stebbins-Boaz, Q. Cao, C. H. de Moor et al., *Mol. Cell* **1999,** *4,* 1017–1027.

292 Q. Cao, J. D. Richter, *EMBO J.* **2002,** *21,* 3852–3862.

293 D. Niessing, S. Blanke, H. Jackle, *Genes Dev.* **2002,** *16,* 2576–2582.

294 J. Ling, S. J. Morley, V. M. Pain et al., *Mol. Cell. Biol.* **2002,** *22,* 7853–7867.

295 D. Poncet, C. Aponte, J. Cohen, *J. Virol.* **1993,** *67,* 3159–3165.

296 M. Piron, P. Vende, J. Cohen et al., *EMBO J.* **1998,** *17,* 5811–5821.

297 P. Vende, M. Piron, N. Castagne et al., *J. Virol.* **2000,** *74,* 7064–7071.

298 B. Raught, A.-C. Gingras, N. Sonenberg: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Plainview, NY 2000, 245–294.

299 C. G. Proud, *Essays Biochem.* **2001,** *37,* 97–108.

300 P. E. Lachance, M. Miron, B. Raught et al., *Mol. Cell. Biol.* **2002,** *22,* 1656–1663.

301 L. McKendrick, S. J. Morley, V. M. Pain et al., *Eur. J. Biochem.* **2001,** *268,* 5375–5385.

302 S. J. Morley, S. Naegele, *J. Biol. Chem.* **2002,** *277,* 32855–32859.

303 A.-C. Gingras, B. Raught, S. P. Gygi et al., *Genes Dev.* **2001,** *15,* 2852–2864.

304 D. Q. Yang, M. B. Kastan, *Nat. Cell Biol.* **2000,** *2,* 893–898.

305 C. T. Keith, S. L. Schreiber, *Science* **1995,** *270,* 50–51.

306 J. L. Crespo, M. N. Hall, *Microbiol. Mol. Biol. Rev.* **2002,** *66,* 579–591.

307 R. T. Abraham, G. J. Wiederrecht, *Annu. Rev. Immunol.* **1996,** *14,* 483–510.

308 L. Beretta, A.-C. Gingras, Y. V. Svitkin et al., *EMBO J.* **1996,** *15,* 658–664.

309 A. C. Gingras, S. P. Gygi, B. Raught et al., *Genes Dev.* **1999,** *13,* 1422–1437.

310 P. T. Tuazon, W. C. Merrick, J. A. Traugh *J. Biol. Chem.* **1989,** *264,* 2773–2777.

311 B. Raught, A.-C. Gingras, S. P. Gygi et al., *EMBO J.* **2000,** *19,* 434–444.

312 M. P. Byrd, M. Zamora, R. E. Lloyd, *Mol. Cell. Biol.* **2002,** *22,* 4499–4511.

313 R. Duncan, J. W. Hershey, *J. Biol. Chem.* **1985,** *260,* 5493–5497.

314 S. J. Morley, J. A. Traugh, *J. Biol. Chem.* **1990,** *265,* 10611–10616.

315 B. Raught, F. Peiretti, A.-C. Gingras et al., **2003** HEMBO J. in press.

316 A. W. B. Craig, A. Haghighat, A. T. K. Yu et al., *Nature* **1998,** *392,* 520–523.

317 K. Khaleghpour, Y. V. Svitkin, A. W. Craig et al., *Mol. Cell* **2001,** *7,* 205–216.

318 C. Grosset, C. Y. Chen, N. Xu et al., *Cell* **2000,** *103,* 29–40.

319 G. Roy, G. De Crescenzo, K. Khaleghpour et al., *Mol. Cell. Biol.* **2002,** *22,* 3769–3782.

320 S. Hoshino, M. Imai, T. Kobayashi et al., *J. Biol. Chem.* **1999,** *274,* 16677–16680.

321 D. Lorenzetti, S. Bohlega, H. Y. Zoghbi, *Neurology* **1997,** *49,* 1009–1013.

322 H. Shibata, D. P. Huynh, S. M. Pulst, *Hum. Mol. Genet.* **2000**, *9*, 1303–1313.
323 F. Sherman, J. W. Stewart, A. M. Schweingruber, *Cell* **1980**, *20*, 215–222.
324 J. W. Stewart, F. Sherman, N. A. Shipman et al., *J. Biol. Chem.* **1971**, *246*, 7429–7445.
325 F. Sherman, J. W. Stewart: in *The Molecular Biology of the Yeast Saccharomyces Metabolism and Gene Expression*, eds J. N. Strathern, E. W. Jones and J. R. Broach, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1982, 301–334.
326 J. I. Stiles, J. W. Szostak, A. T. Young et al., *Cell* **1981**, *25*, 277–284.
327 T. F. Donahue, A. M. Cigan, *Mol. Cell. Biol.* **1988**, *8*, 2955–2963.
328 M. Kozak, *Nature* **1979**, *280*, 82–85.
329 M. Kozak, A. J. Shatkin, *J. Biol. Chem.* **1978**, *253*, 6568–6577.
330 M. Kozak, *Cell* **1978**, *15*, 1109–1123.
331 M. Kozak, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 2850–2854.
332 M. Kozak, *Cell* **1983**, *34*, 971–978.
333 M. Kozak, *Nucleic Acids Res.* **1984**, *12*, 3873–3893.
334 M. Kozak, *Cell* **1986**, *44*, 283–292.
335 M. Kozak, *Mol. Cell. Biol.* **1987**, *7*, 3438–3445.
336 N. Sonenberg, *Gene Expr.* **1993**, *3*, 317–323.
337 L. Gehrke, P. E. Auron, G. J. Quigley et al., *Biochemistry* **1983**, *22*, 5157–5164.
338 J. P. Abastado, P. F. Miller, B. M. Jackson et al., *Mol. Cell. Biol.* **1991**, *11*, 486–496.
339 C. L. Wolfe, Y. C. Lou, A. K. Hopper et al., *J. Biol. Chem.* **1994**, *269*, 13361–13366.
340 D. R. Morris, A. P. Geballe, *Mol. Cell. Biol.* **2000**, *20*, 8635–8642.
341 A. P. Geballe, M. S. Sachs: in *Translational Control of Gene Expression*, eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 595–614.
342 B. Castilho-Valavicius, H. Yoon, T. F. Donahue, *Genetics*, **1990**, *124*, 483–495.
343 T. Donahue, in *Translational Control of Gene Expression*, eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 487–502.
344 A. Thomas, W. Spaan, H. van Steeg et al., *FEBS Lett.* **1980**, *116*, 67–71.
345 Y. Cui, J. D. Dinman, T. G. Kinzy et al., *Mol. Cell. Biol.* **1998**, *18*, 1506–1516.
346 P. Raychaudhuri, A. Chaudhuri, U. Maitra, *J. Biol. Chem.* **1985**, *260*, 2132–2139.
347 A. Chakrabarti, U. Maitra, *J. Biol. Chem.* **1991**, *21*, 14039–14045.
348 D. Chakravarti, U. Maitra, *J. Biol. Chem.* **1993**, *268*, 10524–10533.
349 T. Maiti, U. Maitra, *J. Biol. Chem.* **1997**, *272*, 1833–18340.
350 K. Scheffzek, M. R. Ahmadian, A. Wittinghofer, *Trends Biochem. Sci.* **1998**, *23*, 257–262.
351 S. Das, R. Ghosh, U. Maitra, *J. Biol. Chem.* **2001**, *276*, 6720–6726.
352 F. E. Paulin, L. E. Campbell, K. O'Brien et al., *Curr. Biol.* **2001**, *11*, 55–59.
353 T. V. Pestova, I. B. Lomakin, J. H. Lee et al., *Nature* **2000**, *403*, 332–335.
354 B. S. Shin, D. Maag, A. Roll-Mecak, et al., *Cell* **2002**, *111*, 1015–1025.
355 W. C. Merrick, W. M. Kemper, W. F. Anderson, *J. Biol. Chem.* **1975**, *250*, 5556–5562.
356 J. R. Lorsch, D. Herschlag, *EMBO J.* **1999**, *18*, 6705–6717.
357 J. H. Lee, T. V. Pestova, B. S. Shin, et al., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16689–16694.
358 P. Carrera, O. Johnstone, A. Nakamura et al., *Mol. Cell* **2000**, *5*, 181–187.
359 M. V. Rodnina, A. Savelsbergh, V. I. Katunin et al., *Nature* **1997**, *385*, 37–41.
360 A. Roll-Mecak, B. S. Shin, T. E. Dever et al., *Trends Biochem. Sci.* **2001**, *26*, 705–709.
361 V. Ramakrishnan, *Cell* **2002**, *108*, 557–572.
362 J. Pelletier, G. Kaplan, V. R. Racaniello et al., *Mol. Cell. Biol.* **1988**, *8*, 1103–1112.
363 S. K. Jang, H.-G. Kr•usslich, M. J. H. Nicklin et al., *J. Virol.* **1988**, *62*, 2636–2643.

364 K. Tsukiyama-Kohara, N. Iizuka, M. Kohara et al., *J. Virol.* **1992,** *66,* 1476–1483.

365 J. E. Wilson, T. V. Pestova, C. U. Hellen et al., *Cell* **2000,** *102,* 511–520.

366 A. M. Borman, K. M. Kean, *Virology* **1997,** *237,* 129–136.

367 I. K. Ali, L. McKendrick, S. J. Morley et al., *J. Virol.* **2001,** *75,* 7854–7863.

368 T. V. Pestova, C. U. T. Hellen, I. V. Shatsky, *Mol. Cell. Biol.* **1996,** *16,* 6859–6869.

369 J. Sasaki, N. Nakashima, *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 1512–1515.

370 A. Brasey, M. Lopez-Lastra, T. Ohlmann et al., *J. Virol.* **2003,** *77,* 3939–3949.

371 G. Johannes, M. S. Carter, M. B. Eisen et al., *Proc. Natl. Acad. Sci. USA* **1999,** *96,* 13118–13123.

372 M. Holcik, N. Sonenberg, R. G. Korneluk, *Trends Genet.* **2000,** *16,* 469–473.

373 S. Pyronnet, J. Dostie, N. Sonenberg, *Genes Dev.* **2001,** *15,* 2083–2093.

374 S. Pyronnet, L. Pradayrol, N. Sonenberg, *Mol. Cell* **2000,** *5,* 607–616.

375 S. Cornelis, Y. Bruynooghe, G. Denecker et al., *Mol. Cell* **2000,** *5,* 597–605.

376 M. Honda, S. Kaneko, E. Matsushita et al., *Gastroenterology* **2000,** *118,* 152–162.

377 R. J. Schneider: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Plainview, NY 2000, 581–593.

378 M. S. Carter, K. M. Kuhn, P. Sarnow, in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews Cold Spring Harbor Laboratory Press, NY 2000, 615–636.

379 H. P. Harding, I. Novoa, Y. Zhang et al., *Mol. Cell* **2000,** *6,* 1099–1108.

380 M. Delepine, M. Nicolino, T. Barrett et al., *Nat. Genet.* **2000,** *25,* 406–409.

381 A. Fogli, D. Rodriguez, E. Eymard-Pierre et al., *Am. J. Hum. Genet.* **2003,** *72,* 6.

382 A. E. Koromilas, S. Roy, G. N. Barber, et al., *Science* **1992,** *257,* 1685–1689.

383 J. C. Bell, K. A. Garson, B. D. Lichty et al., *Curr. Gene. Ther.* **2002,** *2,* 243–254.

384 A. Lazaris-Karatzas, K. S. Montine, N. Sonenberg, *Nature* **1990,** *345,* 544–547.

385 J. W. B. Hershey, S. Miyamoto: in *Translational Control of Gene Expression,* eds N. Sonenberg, J. W. B. Hershey and M. B. Mathews, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2000, 637–654.

386 C. Job, J. Eberwine, *Nat. Rev. Neurosci.* **2001,** *2,* 889–898.

387 O. Steward, E. M. Schuman, *Annu. Rev. Neurosci.* **2001,** *24,* 299–325.

388 N. Takei, M. Kawamura, K. Hara et al., *J. Biol. Chem.* **2001,** *276,* 42818–42825.

389 H. Tang, G. Reis, H. Kang et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 461–472.

390 S. Wu, D. Wells, J. Tay et al., *Neuron* **1998,** *21,* 1129–1139.

391 Y. S. Huang, M. Y. Jung, M. Sarkissian et al., *EMBO J.* **2002,** *21,* 2139–2148.

392 F. J. Iborra, D. A. Jackson, P. R. Cook, *Science* **2001,** *293,* 1139–1142.

393 M. T. Bohnsack, K. Regener, B. Schwappach et al., *EMBO J.* **2002,** *21,* 6205–6215.

394 J. E. Dahlberg, E. Lund, E. B. Goodwin, *RNA,* **2003,** *9,* 1–8.

395 M. Sprinzl, C. Horn, M. Brown et al., *Nucleic Acids Res.* **1998,** *26,* 148–153.

396 U. L. RajBhandary, C. M. Chow: in *tRNA Structure, Biosynthesis, and Function,* eds D. Soll and U. L. Raj Bhandary, American Society for Microbiology Press, Washington, DC 1995, 511–528.

397 G. Kozlov, N. Siddiqui, S. Coillet-Matillon et al., *J. Biol. Chem.* **2002,** *277,* 22822–22828.

398 L. Aravind, C. P. Pontings, *Protein Sci.* **1998,** *7,* 1250–1254

399 K. Hofmann, P. Bucher, *Trends Biochem. Sci.* **1998,** *23,* 204–205

8
The Elongation Cycle

Knud H. Nierhaus

The elongation cycle (parts of previous reviews of the Nierhaus group [1–6] have been integrated in this chapter) is the heart of protein synthesis. Each "heart beat" of the ribosome prolongs the nascent polypeptide chain by a single amino acid. The elongation cycle can be divided into the following three basic reactions: (i) Occupation of the A-site by the incoming tRNA. This step can be further subdivided into (a) the decoding reaction, which mainly restricts the aminoacyl-tRNA and ribosome interactions to codon–anticodon interactions (low-affinity interaction) and (b) the accommodation reaction, where high-affinity binding of the whole tRNA to the A-site results in the docking of the aminoacyl residue into the peptidyl transferase (PTF) center of the large subunit. (ii) Peptide-bond formation, where the nascent polypeptide chain is transferred from the P-site tRNA onto the aminoacyl moiety of the A-site tRNA. This leaves a deacylated (or uncharged) tRNA at the P-site and a peptidyl-tRNA at the A-site (with the nascent chain extended by one amino acid). (iii) Translocation involves the movement of mRNA•$tRNA_2$ on the ribosome by one codon length so as to place the deacylated tRNA into the E-site and the peptidyl-tRNA into the P-site, thus freeing the A-site for the next incoming aminoacyl-tRNA. Progression of the ribosome through these various stages of the elongation cycle is catalyzed by protein factors called elongation factors, specifically elongation factor G (EF-G) and elongation factor Tu (EF-Tu) in bacteria. These factors transiently interact with the ribosome at specific points during the elongation cycle to facilitate movement onto the following stage of the cycle, e.g., EF-Tu facilitates delivery of the aa-tRNA to the A-site (step 1), whereas EF-G mediates movement or translocation of the tRNAs from the A- and P-sites to the P- and E-sites, respectively (step 3). Translocation shifts the ribosome from a pre-translocational (PRE) state to a post-translocational state (POST). Both factors have been visualized by cryo-EM in functional complexes with the ribosome at various stages during the elongation cycle [7–10]. By combining these reconstructions with those of PRE and POST tRNA•70S ribosome images, a comprehensive overview of the elongation cycle has been constructed from cryo-EM images [11] (Fig. 8-1).

The ribosomal state before translocation (PRE in Fig. 8-1c) is characterized by tRNAs at the A- and P-sites and, following translocation (POST as seen in Fig. 8-1f), by tRNAs at the P- and E-sites. The PRE and POST states represent the main states

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

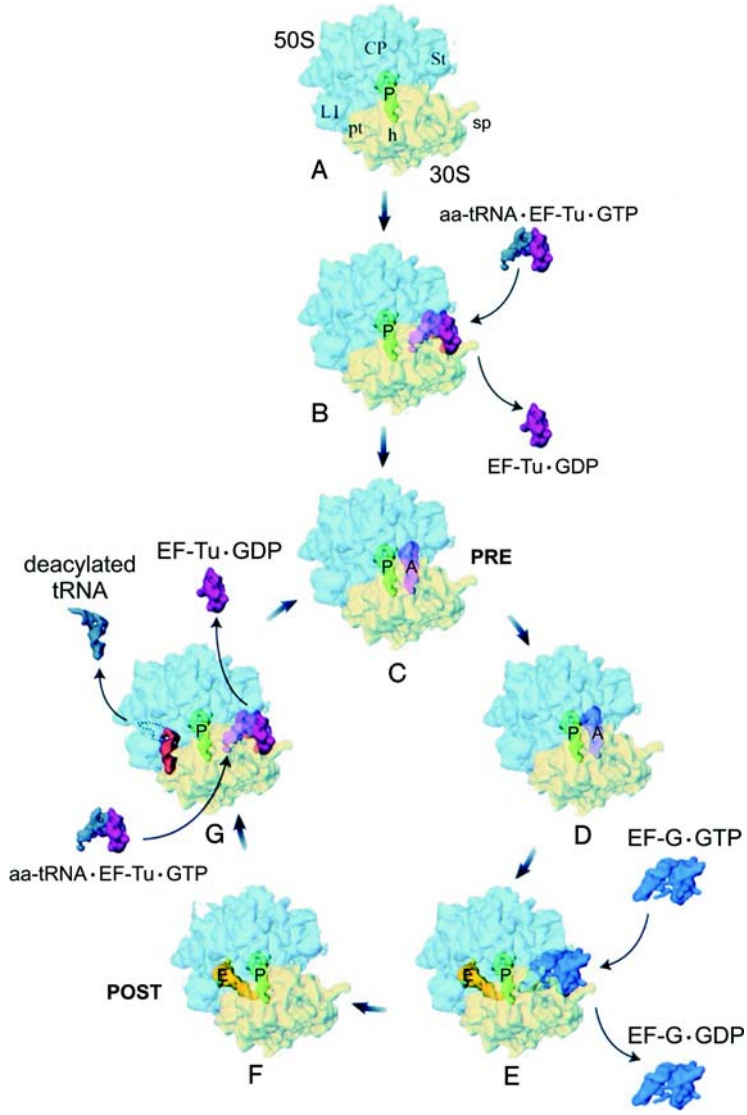


Figure 8-1 Overview of the translation cycle. Multiple cryo-electron microscopic studies have determined the tRNA and elongation factor-binding positions on the 70S ribosome during different stages of the elongation cycle (see Ref. [11] and references therein). These positions of the ribosomal ligands have been overlaid on to an 11.5 Å resolution three-dimensional (3D) map of the ribosome to generate a schematic overview of the elongation cycle, the details of which are provided in the text. The small 30S subunit is in yellow, the 50S large subunit in blue. Adapted from Agrawal et al. [11].

of the active ribosome and are separated by high activation-energy barriers (about 80–90 kJ mol^{-1}; [12]). The POST state probably represents a lower energy level of the ribosome than the PRE state. This is indicated by the fact that after peptide-bond formation an incubation at 37° C for 2 min is sufficient to promote translocation from the PRE to the POST state in the absence of the translocation factor EF-G and GTP ("spontaneous translocation", [13–15]). In contrast, the reverse translocation has never been observed with an isolated POST state.

Therefore, during the elongation cycle, the ribosome can be thought of as oscillating between the two main states, namely the PRE and POST states. Even ribosomes carrying only a P-site tRNA (e.g., AcPhe-tRNA), a state that is referred to as the P_i state, seems to exist in two different coformations, as evident from the differential effects of puromycin, tetracycline, viomycin, and thiostrepton on the two subpopulations of ribosomes [16]. Finally, even the empty ribosome can adopt two different conformations, one resembling the PRE, the other the POST conformation, as judged by synergistic effects of EF-Tu and EF-G on their respective uncoupled GTPases [17].

By reducing the activation energy barrier that exists between the PRE and POST states, the elongation factors significantly accelerate protein synthesis by more than four orders of magnitude (104-fold; a spontaneous translocation in the absence of EF-G lasts about 2 min [13], and an enzymatic translocation in the presence of EF-G about 30 μs [18]). Therefore, they resemble enzymes which also lower the activation energy of a reaction and, due to the enormous acceleration factor of 10^6–10^{12} achieved, just enable a reaction to occur [19]. But enzymes accelerate a reaction only until the equilibrium is reached, i.e., enzymes can usually catalyze both the forward and reverse reactions. In this respect, the elongation factors are more specific since they not only accelerate the reaction rate but also determine the direction of the reaction: EF-Tu catalyzes the POST→PRE transition and EF-G the reverse PRE→POST reaction. This unidirectional mechanism of action of the elongation factors explains why the presence of two elongation factors is universally conserved. Only the higher fungi such as *Saccharomyces cerevisiae* or *Candida albicans* require a third factor, EF-3, which is essential for protein synthesis playing a role in removal of deacylated tRNA from the E-site of the ribosome ([20]; see Sect. 8.1.2 for the ATP-dependent functions of EF-3).

Both the universal elongation factors are prototypes of the large superfamily of G-proteins. Like all G-proteins, these elongation factors are GTPases and follow a cycle of activation in the presence of GTP (the "ON" conformation) and deactivation when GDP is bound ("OFF" conformation) [21]. In the "ON" or GTP conformation, G-proteins bind to their target substrate, be it a protein or complex, to trigger their specific reaction. Subsequently, the GTPase center of the G-protein is activated and the terminal phosphate residue of the bound GTP molecule is cleaved off to yield GDP. The G-protein now adopts the deactivated "OFF" conformation, which has a lower affinity for the target substrate and therefore dissociates from it. For the cycle to be repeated, GDP must be exchanged for GTP on the G-protein. This means for the elongation factors, the energy liberated by the elongation-factor-dependent GTP

hydrolysis is used for the release of the elongation factors after they have done their job rather than for the reaction triggered by the elongation factors in their "on" state (a different view for EF-G was proposed by the "motor protein" hypothesis; see Sect. 8.4.1).

8.1 Models of the Elongation Cycle

Before we describe the molecular details of the individual reactions of the elongation cycle, we will briefly and critically consider the two prevailing models of the elongation cycle, namely the hybrid site and α–ε model.

8.1.1 The Hybrid-site Model for Elongation

This model is based on the observation that bases of rRNA were protected against chemical modification when tRNAs were bound specifically to the ribosomal A-, P- and E-sites. Each tRNA position on the ribosome was correlated with a specific protection pattern, e.g., binding of AcPhe-tRNA (a simple mimic of a peptidyl-tRNA) to the P-site defined the "P-site pattern", whereas binding of a ternary complex Phe-tRNA·EF-Tu·GTP to the A-site (after pre-filling the P-site with a deacylated-tRNA) produced an "A-site pattern". The "E-site pattern" was derived simply from the binding of deacylated tRNA to the E-site, but was not very well defined in the poly(U)-dependent system used [22].

When AcPhe-tRNA was bound to the P-site and the protection patterns were assessed both before and after "peptide-bond formation" (or more accurately, the transfer of the AcPhe to puromycin in the A-site), the P-site pattern shifted to an E-site pattern on the 23S rRNA after peptide-bond formation, whereas an unaltered P-site pattern remained on the 16S rRNA. The conclusion was that the tRNA was in a P/P state before and in a P/E state after the puromycin reaction (where the letter before and after the slash indicate the site bound by the tRNA on the 30S and 50S subunit, respectively). In another experiment, ternary complex was added to ribosomes carrying an AcPhe-tRNA at the P-site, which following peptide-bond formation would put an AcPhe-Phe-tRNA at the A-site and a deacylated tRNA at the P-site. An "A plus P" pattern was observed on the 16S rRNA, whereas the 23S rRNA exhibited a "P plus E" pattern. These results were interpreted such that the analog of the peptidyl-tRNA was in an A/P-site and the deacylated tRNA in a P/E site. Following translocation, the protection patterns suggested that the tRNAs were in the P/P and E-sites, where the E-site is located solely on the 50S subunit rather than on the 30S subunit.

The hybrid-site model [22] in its original form is depicted in Fig. 8-2(A) with its diagnostic feature of hybrid sites after peptide-bond formation: the tRNAs pass through the ribosome with a "creeping" movement. On the 30S the picture follows the classical scheme outlined in the preceding section; however on the 50S, the

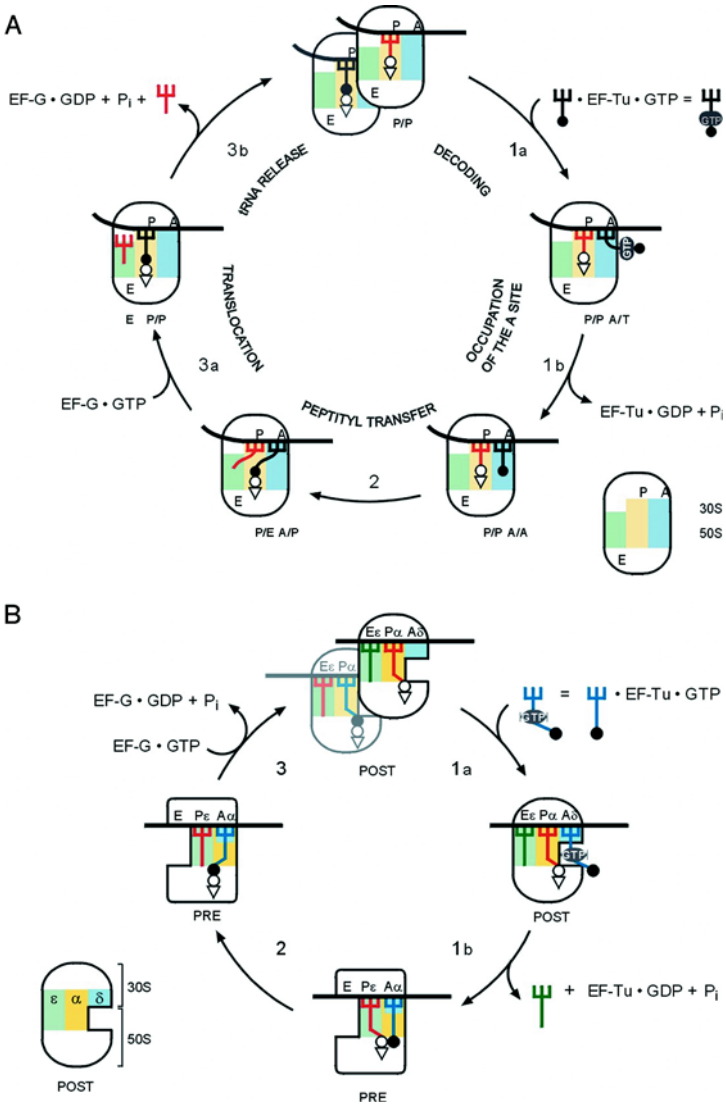


Figure 8-2 Models of the elongation cycle: (A) The hybrid-site model according to Ref. [22]. For explanation see text. The essence of this model is a creeping movement of the tRNAs through the ribosome. (B) The α–ε model of the elongation cycle according to Ref. [49]. The essential feature is a movable ribosomal α–ε domain that connects both subunits through the intersubunit space, binds both tRNAs of an elongating ribosome and carries them from the A- and P-sites to the P- and E-sites, respectively, during translocation. The model keeps all the features of the allosteric-three-site model (see text), but explains the reciprocal linkage between A- and E-sites by the fact that the α–ε domain moves out of the A-site during translocation leaving the decoding center alone at the A-site rather than by an allosteric coupling. In (B) Yellow and green represent the two binding regions of the α–ε domain; blue, the decoding center at the A-site.

movement is proposed to start *after* peptide-bond formation but *before* EF-G-dependent translocation. The model has been modified in light of the crystal structure [23], where the E-site tRNA shows intensive contacts with the small subunit, such that now the tRNA is found in the E/E-site after translocation [24]. A further change was introduced so that after peptide-bond formation the tRNAs initially remain at the classical A/A- and P/P-sites, before they shift, after an undefined time period, into the hybrid A/P- and P/E-sites, respectively [24]. These changes were necessary to account for the observation that no hybrid sites were observed in a systematic study of the elongation cycle ([11]; also see below) and the CCA ends did not move after peptide-bond formation in a 50S crystal [25].

In functional studies, the P-site is operationally defined as the site where an acylated-tRNA can react with puromycin, in contrast with an A-site location where it is for all intense and purposes puromycin unreactive. The observation that an acylated tRNA can in fact undergo the puromycin reaction [26] does not seriously challenge this definition, since the latter reaction differs qualitatively from that of the P-site reaction in that it is extremely slow. For example, under conditions of 6 mM Mg^{2+} and in the presence of polyamines, the A-site puromycin reaction proceeds 200 times slower than that in the P-site (A. Potapov, C. Spahn and K. H. Nierhaus, unpublished results). Note that the hybrid-site model uncouples the functional definition of the P-site from the structural one, because this model locates the peptidyl residue at the P-site of the PTF center (on the 50S subunit) after the peptide-bond formation (because the peptidyl-tRNA is at the A/P-site), yet the peptidyl residue is not puromycin-reactive. Consequently, this model is forced to distinguish between two P-site positions on the 50S subunit, one that is puromycin-reactive and one that is not. Finally, the problem of moving the $(tRNA)_2$•mRNA complex has not become significantly simpler, even though the tRNA movement from one site to the other now occurs in two steps.

There are two major criticisms that should be considered when applying the hybrid-site model:

1. The concept of hybrid sites rests solely on the protection patterns of the 23S rRNA. However, the protections of 16 out of 17 bases of 23S rRNA were dependent on the ultimate A_{76} residue or CA_{76}-3′ residues of the universal CCA-3′ terminus of the tRNAs [27]. In contrast with the hybrid-site perception, the crystal structure of the 50S subunit after peptide-bond formation demonstrates that the CCA ends at A- and P-sites do not move [25].

2. The experiments from which the model is derived employs a vast range of Mg^{2+} concentrations ranging from 5 to 25 mM [28, 27]. However, it is well known that the binding properties and the interdependencies of the various sites are extremely sensitive to changes in Mg^{2+} concentration [29]. This sensitivity probably reflects an increasing distortion of the ribosome with increasing Mg^{2+} concentration, which one would expect affects a fine-structure analysis such as the chemical probing of the rRNA bases. In fact, a deacylated tRNA bound to programed ribosome under conventional ionic conditions (Table 8-1) was found exclusively at the P/E hybrid site, whereas under more *in vivo* ionic conditions, the deacylated tRNA was

found at the classical P-site, suggesting that the hybrid-site concept may simply be a buffer artefact [30]. A systematic analysis of tRNA-binding sites during elongation did not provide any evidence for hybrid states of a tRNA [11]. Recently, cryo-EM of PRE state ribosomes prepared in a different laboratory also showed no evidence for the presence of hybrid states [31]. In this study, the post-peptide-bond formation of PRE state ribosome complexes had a dipeptidyl tRNA at the classical A-site and not at the A/P hybrid site as would be predicted by the hybrid-site model. From these independent studies, it is at least clear that a ribosome containing a hybrid site does not represent a significant proportion of the elongating ribosome population. Furthermore, the crystal structure of a programed ribosome containing three deacylated tRNAs at 5.5 Å resolution identified a deacylated tRNA at the classical P-site [23], rather than at the P/E hybrid site as might be expected from the hybrid-site model.

Contrary to the belief of some, the ratchet model, where the subunits twist relative to each other in a forward-and-back movement by about 4.5° during translocation [32], does not in fact support the hybrid-site model. This is because the forward-and-back ratchet movements occur *during* one translocation step, i.e., the tRNAs are at A- and P-sites before translocation and at P- and E-sites after translocation (see Sect. 8.4.2). However, it could be that during one translocation reaction the two tRNAs do transiently pass through a hybrid position when they go from the A- and P-sites to the P- and E-sites, respectively. In fact, this interpretation is supported by a cryo-EM analysis of the translocation movement [31], suggesting that hybrid states are indeed translocation intermediates.

8.1.2 The Allosteric Three-site Model (α–ε Model; Reciprocal Coupling between the A- and E-sites)

Under unfavorable buffer conditions the dissociation rate of a tRNA at the E-site was reported to be 0.3 s^{-1} [33], whereas the elongation rate, determined under comparable conditions, was reported to be 10 times faster (3 s^{-1}) [34]. This alone suggests that an active mechanism must exist to eject the deacylated tRNA from the E-site. Under near *in vivo* conditions, the situation is even more significant: the dissociation rate of an E-site tRNA from an isolated POST state is much lower and can be measured in hours rather than seconds. POST states can be isolated via overnight centrifugation through sucrose cushions without loss of any deacylated tRNA from the E-site (see, e.g., Ref. [35]), and native polysomes isolated using a procedure lasting longer than 24 h contain an occupied E-site almost quantitatively [36]. It is clear that the release of a deacylated tRNA from the E-site must be an active process and cannot occur via a simple diffusion process as considered by some authors (see, e.g., Ref. [37]). In fact, crystal structures of 70S•tRNA complexes [23] and cryo-EM studies [31] have demonstrated stable and tight E-site binding, although a mechanism for the E-tRNA release was not deduced.

A reciprocal linkage between A- and E-sites has been identified as being the active mechanism for E-site ejection: the E-site affinity drops markedly upon occupation of the A-site and *vice versa*, i.e., the occupation of the A-site is coupled with the tRNA

release from the E-site [38, 39]. In the years following the identification of the reciprocal linkage between A- and E-sites, much data accumulated providing support for this model: (i) A direct and unequivocal demonstration of the reciprocal linkage was achieved using a heteropolymeric mRNA displaying three different codons at the three sites together with three respective cognate tRNAs, each labeled with a different isotope [40]. (ii) The activation energy for occupation of the A-site depends on the charging state of the E-site: when a tRNA is present at the E-site the activation energy is twice as large as that observed when the E-site is free [12]. (iii) Thiostrepton, viomycin and all types of aminoglycosides severely impair A-site binding only if the E-site is occupied [41]. (iv) The reciprocal coupling of A- and E-sites has also been observed with the ribosomes of organisms from other evolutionary domains, viz. with ribosomes of *Halobacterium halobium* (archea; [42]) and yeast (eukarya; [20]), suggesting that this relationship is universally conserved. Indeed, in yeast, the reciprocal coupling was observed when the functions of the third elongation factor (EF-3) were studied [20]. In the POST state, yeast 80S ribosomes bound the E-tRNA so tightly that a ternary complex aa-tRNA•EF-1•GTP could not bind to the A-site. First, the binding of EF-3 to the ribosome and ATP cleavage was necessary to free the E-site tRNA, perhaps by "opening" up the E-site by movement of the L1 protuberance of the 50S subunit. Only then was it possible for A-site occupation and the concomitant release of the E-tRNA to occur. These results illustrate the bi-directionality of the reciprocal linkage: if the E-tRNA cannot be released, the A-site remains in its low-affinity state and cannot be occupied; if EF-3 "opens" the E-site, A-site occupation triggers release of the E-tRNA.

The reciprocal linkage between A- and E-sites led to the "allosteric three-site model", which is characterized by three basic features (the last version is the α–ε model, see Fig. 8-2B; see Ref. [43] for review):

1. Ribosomes contain three tRNA-binding sites [44, 45].
2. The first and the third sites, A- and E-sites, respectively, are coupled in a reciprocal fashion: occupation of the A-site decreases the affinity at the E-site and *vice versa* [40, 39, 20].
3. Both tRNAs that are present at the A- and P-sites before translocation and at the P- and E-sites after translocation are linked to the mRNA via codon–anticodon interaction [46, 20]. Codon–anticodon interaction at the E-site seems to be essential for establishing the POST state containing the P- and E-tRNAs [47]. Only the POST state is the proper substrate for the ternary complex aa-tRNA·EF-Tu·GTP.

The reciprocal linkage model was extended to the so-called α–ε model when the accessibility of the phosphate groups of a tRNA at the various ribosomal sites was tested in PRE and POST states. These experiments demonstrated that a tRNA in the A-site of a PRE state ribosome had a strikingly different pattern when compared with the corresponding tRNA in the P-site [48]. However, after translocation to the P- and E-sites, the protection patterns of both tRNAs hardly changed [49]. The conclusion was that the tRNAs bound to a structural domain of the ribosome, and that this structural domain moved during translocation from the A- and P-sites to the P- and E-sites while maintaining contact with both tRNAs. Therefore, the structural

domain was considered to contain two binding regions, which were termed α and ε (Fig. 8-2B). A tRNA bound to the α binding region is located at the A-site before translocation and at the P-site after translocation. Similarly, a tRNA bound to ε binding region is located at P-site before translocation and at the E-site after translocation. Thus, only α can appear at the A-site and only ε at the E-site (hence the nomenclature), whereas at the P-site either α or ε can be present depending on the translocation state. Support for the model was recently provided by results obtained using a completely different approach: site-specific Pb^{2+} cleavage was applied to trace tertiary alterations of tRNAs and rRNAs in PRE and POST state ribosomes [50]. Deacylated tRNAs and AcPhe-tRNA produced the same cleavage pattern in solution but very different ones when bound to the ribosome. Consistent with phosphorothiote experiments, the specific and distinct patterns for the bound tRNAs did not change during translocation. This again led to the conclusion that while the tertiary structure of the adjacent tRNAs at A- and P-sites are different, the fact that they do not change during translocation argues for a ribosomal "conveyor" that binds both tRNAs and moves them during translocation. Comparing contact patterns of tRNAs obtained with isolated subunits and 70S ribosomes using the phosphorothioate method indicated that the two parts of the "conveyor", both of which bind a tRNA, probably move in a concerted fashion but not strictly side-by-side [51]. Therefore, the recent observation that the deacylated tRNA seems to move from the P- to the E-site via a hybrid position [31] is not in conflict with the α–ε model.

The α–ε model integrates the well-documented fact that the post-translocational ribosome with a low-affinity A-site is capable of selecting the aminoacyl-tRNA cognate to the codon at the A-site, i.e., the decoding process occurs before the α–ε domain of the ribosome flips back from the P-E-site to the A-P-site. According to the α–ε model, the decoding site, being exclusively on the 30S subunit, is stably located at the A-site where it is called δ. It follows that α is superimposed on δ in the PRE state and separated from the δ domain in the POST state (Fig. 8-2B). This feature of the α–ε model predicts that the ribosome has two tRNA-binding sites in the PRE state and three in the POST state (the two high-affinity sites α–ε at the P- and E-sites and the low-affinity site δ at the A-site).

All three features of the allosteric three-site model are also valid for the α–ε model, although they are extended or re-interpreted, such that (i) the three tRNA-binding sites still exist but only in the POST state. Saturation of 70S ribosomes with deacylated tRNAs levels off at three tRNAs per ribosome. First, the P- and E-sites are filled, thus establishing a POST state, thereby creating a low-affinity A-site (δ-site at the A-site). This is illustrated by the requirement of excess (> 6-fold) of deacylated tRNAs over ribosomes to fill the third site [45]. (ii) The A- and E-sites have an inverse relationship in that an occupied E-site is accompanied by a low-affinity δ-site at the A-site and an occupied A-site with no affinity at the E-site. The new interpretation of this relationship is that allostery is not involved: during translocation the α region moves from the A-site to the P-site leaving the decoding center δ in the A-site ("low-affinity" A-site), and during aminoacyl-tRNA binding to the A-site the α–ε domain jumps from the P-E to the A-P, leaving the E-site without a tRNA-binding capacity. (iii) The

two tRNAs have a similar mutual arrangement relative to each other. In fact, the angles between the tRNAs in the PRE and POST states are almost identical, as shown by cryo-EM analysis (39° and 35° in the PRE and POST states, respectively [11]). Also the arrangement relative to the mRNA, i.e., the codon–anticodon interactions, is maintained before, during and after translocation with regard to the α–ε binding regions. Thus, the movement of the tRNAs occurs simultaneously on the two subunits in a coordinated fashion, and during translocation the tRNAs might transiently move through hybrid positions as has been made evident for the deacylated tRNA moving from the P-site over the hybrid position P/E to the E-site [31], whereas the predictions of the hybrid-site model, namely that the tRNAs swing into a hybrid site after peptide-bond formation could not be confirmed [11, 31]. However, the CCA ends of the two tRNAs present at A- and P-sites are directly adjacent at the PTF center at the PRE state (required for peptide-bond formation), whereas they are separated substantially in the POST state. After the peptide bond has been formed, there is no need to keep them together (see Figs. 6-2A and B in Chap. 6). This fact indicates that the α and ε regions do not strictly move side-by-side during translocation.

Ribosomal candidates for a movable domain have been identified as bridge B2a in 70S ribosomes via cryo-EM [52] and X-ray crystallography [23], and probable ribosomal components such as the upper region of the h44 of the 16S rRNA [53], H69 of the 23S rRNA and parts of the ribosomal protein L2 [54, 55].

The α–ε model provides a dynamic picture of the translating ribosome. The translocation reaction is explained in a (possibly too) simple fashion: the α–ε domain moves together with both tRNAs and the corresponding codons. The required movement of about 10 Å, the length of a codon, is not unusual for molecular movements of enzyme substructures. For example, the distance between the first and the second domains of EF-Tu is enlarged by up to 40 Å upon GTP cleavage [56]. According to the α–ε model, the critical step during the elongation cycle is not the translocation reaction but rather the binding of the new aminoacyl tRNA in the α-site at the A-site, since during the latter the α–ε domain has to release the tRNAs in order to switch from the E-A-sites to the P-A sites (see Fig. 8-2B). This is in agreement with the finding that occupation of the A-site is the rate-limiting step of elongation, not the translocation reaction [57, 12].

To date, the most comprehensive analysis of the translocation reaction by cryo-EM [31] challenges an essential feature of the α–ε model, namely, PRE state ribosomes carrying an fMet-Phe-tRNA at the A-site and a deacylated $tRNA^{Phe}$ at the P-site, also carried a tRNA at the E-site. This was particularly surprising since deacylated tRNA was not actually included in the reaction; however, the authors assumed that the source of the E-tRNA was the free pools of deacylated tRNA in solution, which enabled direct binding to the "high-affinity" E-site. Regardless of the source, the presence of simultaneously occupied A- and E-sites is in direct contradiction with the α–ε model. However, to contradict the α–ε model directly the stoichiometry of the A- and E-sites needs to be assessed. The authors made no attempts to resolve this conflict between their conclusions and the findings that an E-site release upon A-site occupation has been observed with both prokaryotic and eukaryotic ribosomes

(see Ref. [58] for review). In any case, it is clear that we are far from having a detailed picture of the translocation reaction and the presence of high-resolution structures of PRE and POST states would certainly go some way to resolving one of the central enigmas of ribosome research.

8.2
Decoding and A-site Occupation

8.2.1
Some General Remarks about Proofreading

The term "proofreading" stems from the glossary of the printing arts. The first outprint of a newspaper, e.g., was checked by the print master, and if he found a misprint, the letter was exchanged and the mistake corrected for the second print that was delivered to the clients. In molecular biology, this term has a similar meaning, in that the last amino acid building block that is to be added to the growing polypeptide chain is checked for correctness before it is permanently incorporated into the synthesized protein.

Proofreading is a common phenomenon in polymerases that synthesize DNA [59] and RNA [60]. In fact, polymerization of nucleic acids is well suited for proofreading, since its basic mechanism is that of "tail growth" (Fig. 8-3). This means the new building block that is to be added to the nascent chain is providing the energy-rich bond (the phosphoric acid anhydride bond between the α and the β phosphate residues of an NTP) for the link (the 3′ ester bond) to the nascent chain. If the addition was wrong, the proofreading center hydrolyzes the last four nucleotides, thus enabling another chance for the correct addition.

In contrast, the situation is very different in the case of protein synthesis since it follows the principle of "head growth" (Fig. 8-3): the nascent chain provides the energy-rich ester bond for the formation of the peptide bond that links the newly added amino acid. The peptidyl residue is added to the newly arrived amino acid, and the ester bond of the latter is used for the next round of elongation. Therefore, in *strictu sensu*, proofreading cannot exist in protein synthesis, since such a process would sacrifice the already synthesized chain via removal of the last incorporated amino acid. Not surprisingly, therefore, a "proofreading center" was not detected in the atomic structures of the ribosome. Nevertheless, the term proofreading is widely used in the field of protein synthesis, but here in a broader sense, such as "kinetic proofreading" that will be discussed in Sect. 8.2.3. We should also note here that synthetases of both class I and class II have "proofreading centers" that play an important role for achieving the accuracy of charging tRNAs [61–63] (see also Chap. 4.2).

8.2.2
Discrimination against Noncognate aa-tRNAs

A new elongation cycle begins when an aa-tRNA enters the ribosomal A-site as the ternary complex aa-tRNA·EF-Tu·GTP. The codon displayed in the A-site is specific

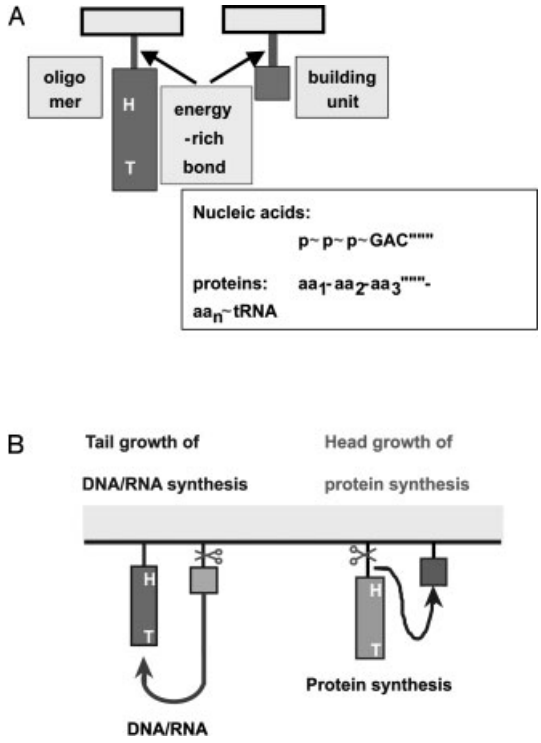


Figure 8-3 Principles of tail- and head-growth: (A) Definitions of head (H) and tail (T) in the synthesis of nucleic acids and proteins. Tail growth means that the energy-rich bond of the new building unit is used for incorporation, whereas during head growth the energy-rich bond of the nascent chain is used for incorporation of a new building block. (B) Tail growth is exemplifed by the synthesis of nucleic acids (activities of replicases and transcriptases), and head growth to protein synthesis on the ribosome. See text for details.

for a single species of tRNA, termed the *cognate* tRNA, which has an anticodon that perfectly complements the A-site codon. However, there are many other tRNA competitors that can interfere with this selection process: 41 in *E. coli* and even more in the eukaryotic cell. To make matters worse, 4–6 of these tRNAs, termed *near-cognate* tRNAs, will have an anticodon similar to the cognate tRNA. The remaining 90% have a dissimilar anticodon and are termed *noncognate* tRNAs. The problem is compounded further when one considers that the aa-tRNAs are delivered in the form of a ternary complex, i.e., in complex with the elongation factor EF-Tu and GTP. The ribosome must therefore discriminate between relatively large ternary complexes (72 kDa), which present multiple potential interaction sites with the ribosome, on the basis of a small discrimination area, the anticodon (1 kDa). The discrimination

potential of the discrimination energy can only be reached under equilibrium conditions and, in this case, the free energy of binding is relatively large, with only a tiny fraction being discrimination energy. This means that equilibrium can only be reached after long time periods, i.e., this process must be slow to be accurate. Since we know that protein synthesis is a relatively fast and accurate process, the ribosome must overcome this hurdle. The question is how?

A model has been proposed which overcomes this problem by simply dividing A-site occupation into two distinct events; a decoding step followed by an accommodation step (reviewed in Ref. [64]). During the initial decoding step, the A-site is in a low-affinity state, which reduces interaction of the ternary complex to mainly codon–anticodon interactions, thus excluding general contacts of the tRNA and elongation factor. By restricting the binding surface of the ternary complex to the discriminating feature, i.e., the anticodon, the binding energy is both small and equivalent to the discrimination energy. In addition, since the binding energy is small, equilibrium can be rapidly attained, thus ensuring the efficiency of the reaction is retained. The second step, accommodation of the A-site, requires release of the aa-tRNA from the ternary complex into the A-site. This step utilizes the nondiscriminatory binding energy to dock the tRNA precisely into the A-site and the attached amino-acyl residue into the PTF center on the 50S subunit in preparation for peptide-bond formation. As we have already seen in the previous section, accommodation of the aa-tRNA in the A-site is accompanied by release of the E-tRNA. Evidently, this second step of A-site binding involves gross conformational changes within the ribosome [12] and thus can be thought of as a relatively slow process in comparison with the first or decoding step. It follows that A-site binding occurs via a coupled reaction system, consisting of a fast initial decoding and a slow second accommodation reaction. This has the important consequence that the initial reaction operates at equilibrium even when the whole system runs under steady-state conditions. It is this feature that enables the discriminatory potential of codon–anticodon interaction to be efficiently exploited.

Recently, the first step of A-site binding (low-affinity A-site) was viewed using cryo-EM by analyzing ternary complexes stalled at the A-site using the antibiotic kirromycin [65, 10]. Although kirromycin allows GTP hydrolysis, it inhibits the associated conformational changes in EF-Tu that are necessary for dissociation from the ribosome. The cryo-EM reconstructions suggest that the anticodon-stem loop of the tRNA is kinked to allow codon–anticodon interaction and thus overcomes the unfavorable incoming angle of the tRNA to the A-site dictated by the ternary complex (see Fig. 8-4; [10]).

As accommodation of an aa-tRNA into the A-site involves the dissociation of EF-Tu·GDP from the ribosome, which is in turn coupled with the hydrolysis of GTP, it is interesting to note that in *E. coli* up to 2 GTPs are hydrolyzed during the incorporation of cognate-tRNAs and up to 6 for near-cognate-tRNAs, whereas noncognate-tRNAs do not trigger EF-Tu-dependent GTP hydrolysis at all [66]. The acceptance of a near-cognate aminoacyl-tRNA consumes three times more GTP than that of a cognate one, thus improving the accuracy by a factor of 3 only. Since the total

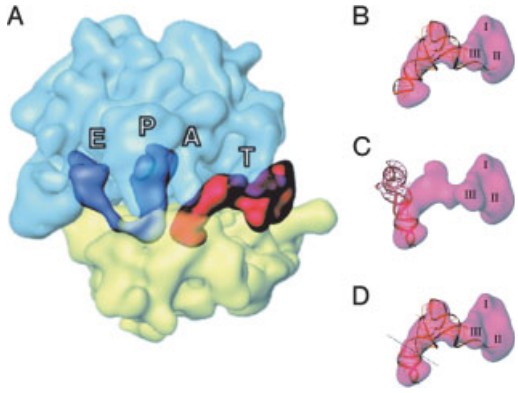


Figure 8-4 The ternary complex on the ribosome during the decoding process. (A) Binding of the ternary complex aa-tRNA•EF-Tu•GTP (red or pink) during the first step of A-site occupation (T position), with tRNAs present at P- and E-sites. Ribosomal subunits are in blue (large subunit) and yellow (small subunit). (B-D) Fitting of the crystal structure of the ternary complex into the difference mass corresponding to the ternary complex on the ribosome. To fit satisfactorily the crystal structure of a tRNA into the corresponding cryo-EM density requires the introduction of a kink in the anticodon stem of the amino-acyl-tRNA of about 40°. From Ref. [10]; for details see text.

accuracy is characterized by a factor of about 3000 (one mis-incorporation in 3000 incorporation events), it is clear that this observation adds further weight to the argument that the tRNA discrimination is governed predominantly by anticodon–codon recognition during the initial binding step (see also the next section).

Why is the low-affinity A-site during the decoding step important for preventing the interference of noncognate tRNAs with the decoding reaction? Preventing this interference by noncognate tRNAs is not trivial since they represent the majority of about 90% of the ternary complexes. If the decoding step reduces the interactions of the ternary complex with the ribosome mainly to codon–anticodon interaction, then this will prevent noncognate ternary complexes interfering with decoding, since interaction of the ternary complex outside of the anticodon is not possible and the anticodon of noncognate tRNAs cannot form efficient base-pairs with the codon displayed at the A-site. In other words, the A-site codon does not really exist for the noncognate complexes. This fact reduces the selection problem by an order of magnitude: instead of selecting one out of 41 tRNA species (cognate versus near- plus noncognate tRNAs) only 1 out of 4–6 tRNAs have to be selected (cognate versus near-cognate tRNAs). The selection problem is comparable with that of a transcriptase selecting the correct nucleotides out of 4 possible ones during RNA synthesis. This process occurs with a precision of better than one mistake in 60 000 incorporations without proofreading [60]. The fact that the noncognate ternary complexes do not interfere with the selection process has been demonstrated by a number of different

approaches: (i) When an E-tRNA induces a low-affinity A-site, a noncognate ternary complex is not incorporated into the nascent peptide chain [47]. (ii) The addition of an excess of noncognate ternary complexes does not slow down the rate of poly(Phe) synthesis in fast systems ([67] and A. Bartetzko and K.H. Nierhaus, unpublished observations). (iii) As mentioned previously, noncognate ternary complexes also show no traces of GTP turnover, whereas near-cognate complexes have a turnover about three times higher than that of cognate ternary complexes [66].

The next obvious question is how the cognate and near-cognate tRNAs are discriminated. This is a question that can now be answered at the molecular level, as discussed in the next section.

8.2.3
Decoding of an aa-tRNA (Cognate versus Near-cognate aa-tRNAs)

A model for the discrimination between cognate and near-cognate aa-tRNAs was proposed by Potapov about 20 years ago [68]. According to this model, the decoding center of the ribosome recognizes the anticodon–codon duplex, in particular sensing the stereochemical correctness of the partial Watson–Crick base-pairing and the positioning of the phosphate-sugar backbone within this structure. A test of this hypothesis was performed with an mRNA that carried a DNA codon (deoxycodon) at one of the three ribosomal sites. If the stability of the base pairs, i.e., the hydrogen bonding between codon–anticodon bases of the Watson–Crick pairs, is the sole requirement for the recognition step, then a 2′-deoxy base in the codon should not affect the decoding process. If, however, the stereochemical correctness of the base pairing is tested, i.e., including the positioning of the sugar pucker, then a 2′-deoxy base should impair the decoding process. It was found that a deoxycodon at the A-site was disastrous for tRNA binding at this site, whereas a deoxycodon at the P-site had no effect on tRNA binding to the P-site. This observation also explains previous results according to which a DNA cannot take over the function of an mRNA (see Ref. [69] and references therein).

The components of the ribosome directly involved in decoding were identified to 3.1 Å by crystallography [70]. Crystal packing of the *Thermus thermophilus* 30S subunit fortuitously placed the spur (h6) of one subunit into the P-site of another, thus mimicking the anticodon stem loop (ASL) of a P-tRNA. Another surprise was that the base-pairing partner to the P-tRNA mimic was the 3′-end of the 16S rRNA, which folding back on itself extended into the decoding center. This situation, with the P-site filled, enabled Ramakrishnan and co-workers [70] to soak an ASL fragment (ASL-tRNA) and a complementary mRNA fragment into these crystals to study aa-tRNA decoding at the A-site.

The results can be summarized as follows: the binding of mRNA and cognate aa-tRNA induces two major rearrangements within the ribosomal decoding center: the universally conserved residues A1492 and A1493 flip out of the internal loop of h44, whereas the universally conserved base G530 switches from *syn* to *anti* conformation. A1493 recognizes the minor groove of the first base pair (ASL-tRNA position A36–U1 of the mRNA) via a type I A-minor motif (Fig. 8-5A). A1493 establishes

three hydrogen bonds (H-bonds) with the first position of the codon–anticodon duplex: two with 2′-OH groups from both A36 and U1 and another with the O2 of U1. It is noteworthy that the latter H-bond is not sequence-specific as might be expected, since the O2 position of pyrimidines and the N3 of purines occupy equivalent positions in the minor groove of a double helix and both are H-bond acceptors.

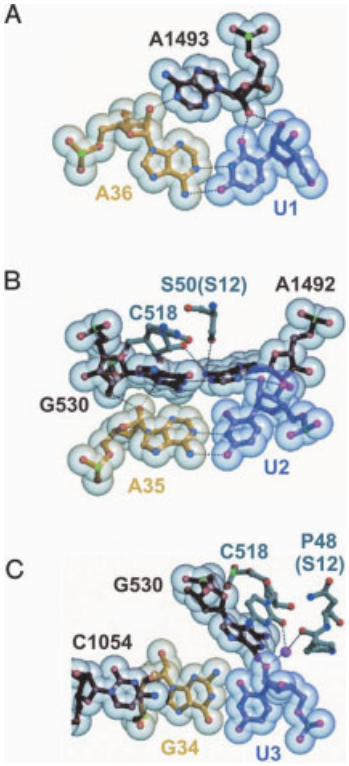


Figure 8-5 The principles of decoding in the A-site of the ribosome. (A) The first base pair of codon–anticodon interaction (position 1) exemplifies a type I A-minor motif: A1493 binds to the minor groove of the A36–U1 base pair via H-bonds. (B) The middle position illustrates a type II A-minor motif: A1492 and G530 acting in tandem to recognize the stereochemical correctness of the A35–U2 base pair using H-bonds. (C) The third (or wobble) base pair (G34–U3) is less rigorously monitored. C1054 stacks against G34, whereas U3 interacts directly with G530 and indirectly with C518 and proline 48 of S12 through a magnesium ion (magenta). All nucleotides involved in monitoring positions 1 and 2 are universally conserved. Adapted from Ogle et al. [70].

The second base pair (A35–U2) is also monitored via 2′-OH interactions, but the job is split between two bases, namely A1492 and G530 (this type II interaction of an A-minor motif is seen in Fig. 8-5B). A1492 and G530 are locked in position by secondary interactions with S12 (serine 50) and another universally conserved residue, C518. Monitoring the second base pair seems to occur more rigidly than that of the first pair in accordance with the fact that the second base pair is of utmost importance for decoding followed by the first pair, whereas the third pair of the codon–anticodon duplex is of least importance [71]. In other words, correct positioning of the 2′-OH groups of the first and second base pairs is critical in forming A-minor interactions and thus efficient duplex sensing. In contrast, the third position is less rigorously monitored, allowing latitude for wobble interactions (Fig. 8-5C). This is evident in the third base pair (G34-U3), where the minor groove remains exposed despite direct interactions with C1054, G530 and indirect metal-mediated interactions with C518 and proline 48 of S12. Taken together, these results confirm the Potapov hypothesis and explain how decoding operates through the recognition of the correct stereochemistry of the A-form codon–anticodon duplex. Furthermore, the fact that the components involved are universally conserved suggests that the mechanism of decoding is probably similar for ribosomes from all kingdoms.

Important extensions to this picture could be made when the known crystal data of *T. thermophilus* 30S subunit showing cognate codon–anticodon interactions between an UUU codon and an ASL structure of the $tRNA^{Phe}$ [70] was complemented with a data set obtained after soaking the programed 30S subunits with near-cognate ASLs from $tRNA_2^{Leu}$ and $tRNA^{Ser}$. The cognate $tRNA^{Phe}$ has the anticodon 3′-AAG-5′, the near-cognate $tRNA_2^{Leu}$ and $tRNA^{Ser}$ contain the anticodons 3′-GAG-5′ and 3′-AGG-5′ with a G:U mismatch at the first or second position, respectively, whereas such a mismatch is allowed at the wobble position [72]. The most important conclusion from this work was that only the cognate codon–anticodon interaction could induce a "closed" form of the 30S subunit involving a movement of the shoulder and head towards each other, whereas near-cognate ASLs could only induce this conformation in presence of the misreading-inducing aminoglycoside antibiotic pactamycin. The complete movement can be seen under http://www.cell.com/cgi/content/full/111/5/721/DC1, and has a number of important consequences. Various nonpolar interactions between the ribosomal proteins S4 and S5 are broken upon transition from the closed to the open form, whereas h44, h27 and S12 come together and can form additional salt bridges (between residues K57 and the phosphate of C1412 of h44 or K46 and the phosphate of A913 of h27). It has already been noted previously that mutations which would be predicted to break the nonpolar interactions between S4 and S5 induce a "ram" phenotype (ram, ribosomal ambiguity mutations: ribosomes with a defect that is characterized by a high-level amino acid mis-incorporation [73, 74]). In other words, these mutations facilitate the transition into the open form and thus reduce translation fidelity by inducing mis-incorporation. On the other hand, mutations of S12 that destabilize the closed form would impair the transition to the closed form and thus increase the fidelity of aa-tRNA selection at the A-site. In fact, mutations of K57 of S12 result in the most-

accurate ("hyperaccurate") phenotype known. Therefore, we have gained for the first time a molecular understanding of mutants that increase or decrease the level of mis-incorporations of the ribosome. Since an occupied E-site that has been shown to induce a low affinity of the A-site and improve the accuracy (see Ref. [75] for review), also makes an important contribution to the ribosomal power for A-site tRNA discrimination, this suggests that in the frame of the open-closed model of Ramakrishnan and coworkers [72], the extensive contacts the E-site tRNA with both the 30S and 50S subunit might increase the energetic costs for the transition from the open to the closed form.

Prior to the Potapov hypothesis, it had been proposed that the ribosome utilized a "proofreading mechanism" to improve the accuracy of translation [76, 77]. This mechanism was suggested to operate by re-selection of the correct substrate during a so-called "discarding step", after the initial binding of the A-tRNA. Because re-selection is dependent upon release of the tRNA from EF-Tu and is accompanied by GTP cleavage, the GTP consumption for the incorporation of a cognate and near-cognate amino acid provides a measure of the power of proofreading. Insofar as the crystal structure of EF-Tu and the ribosome are concerned, the ribosomal proofreading mechanism lacks its own active center (see Sect. 8.2.1 for a principal comparison of the synthesis of nucleic acids and proteins concerning proofreading). Instead, the term "proofreading" has been broadened by introducing "kinetic proofreading" that occurs after the release of the binary complex EF-Tu·GDP [78]. A simple model for kinetic proofreading is the following: the binding energy during the decoding step (first step of A-site binding, see 8.2.2) is for the near-cognate aa-tRNA lower than for the cognate one. Therefore, the probability of triggering the gross-conformational change required for the accommodation of the aa-tRNA into the A-site (second step of A-site binding) is lower than for the near-cognate. This in turn prolongs the resting time of the near-cognate aa-tRNA at the low-affinity A-site and provides an additional chance for the near-cognate aa-tRNA to fall off the low-affinity A-site [79] thus increasing the accuracy. Re-binding of this near-cognate aa-tRNA is unlikely in the presence of competing ternary complexes that have a 2–3 orders of magnitude higher affinity for the A-site than the naked aa-tRNA [80].

The importance of the kinetic proofreading step can be quantitatively determined by taking advantage of the fact that the kinetic proofreading mechanism requires EF-Tu-dependent GTP hydrolysis. Accuracy of aa-tRNA selection in the presence of EF-Tu and a noncleavable GTP analog was determined to be 1:1000 [81], an accuracy only three times lower than that seen *in vivo* (1:3000). Exactly a threefold difference was also determined for the GTP consumption per incorporation of cognate versus near-cognate amino acids [66]. Thus it is clear that the significant contribution to the accuracy of translation (1000-fold) lies within the stereo-chemical monitoring of the codon–anticodon duplex by the ribosome as predicted by Potapov and that the "kinetic proofreading mechanism" plays only a minor role, conferring a 3-fold improvement in the accuracy. This view was qualitatively confirmed by a recent direct measurement of the discrimination power of the initial binding without proofreading, where the binding of cognate and near-cognate ASL-tRNA fragments

to the A-site of 70S ribosomes were compared. The accuracy was found to be between 1 : 350 and 1 : 500, thus also demonstrating that the lion's share of the ribosomal accuracy is carried by the initial binding [72].

8.2.4 Roles of EF-Tu

The following functions of EF-Tu can be distinguished: (i) EF-Tu within the ternary complex aa-tRNA•EF-Tu•GTP reduces the activation energy barrier between the POST and the PRE states by about 120 kJ mol^{-1} [12] and thus allows the transition from the POST to the PRE state with a high rate. (ii) EF-Tu binds an aa-tRNA at the amino acid acceptor stem thus shielding the labile ester bond between the aminoacyl residue and the tRNA. (iii) A third function (related to function (i) but not identical) is the carrier role of EF-Tu, namely, to deliver the aa-tRNA to the A-site: the ternary complex has an affinity for the A-site, 2–3 orders of magnitude higher than that of the corresponding aminoacyl-tRNA [80]. (iv) Another role for EF-Tu was identified by Uhlenbeck and co-workers [82]. Measuring the affinities of various cognate aa-tRNA (e.g., Val-$tRNA^{Val}$) and some mis-pairs (e.g., Ala-$tRNA^{Val}$) they recognized that either the amino acid or the tRNA binds to EF-Tu with high affinity to form stable ternary complexes aa-tRNA · EF-Tu · GTP. For example, EF-Tu · GTP easily forms a ternary complex with Asp-$tRNA^{Asp}$ or Asn-$tRNA^{Asn}$, but not with the mis-charged Asp-$tRNA^{Asn}$, since in the latter case both moieties bind with low affinities. This observation explains an important scenario that was an enigma hitherto: in most organisms there are only 18 or 19 synthetases, i.e., not the 20 different synthetases corresponding to the 20 natural amino acids. For example, many organisms do not contain a synthetase specific for asparagine (AsnRS). In this case, AspRS is also charging $tRNA^{Asn}$ with aspartic acid, yielding a mis-charged Asp-$tRNA^{Asn}$, which is recognized by enzymes that amidate Asp to Asn on the tRNA. The mis-charged Asp-$tRNA^{Asn}$ does not form a stable ternary complex with EF-Tu · GTP and thus Asp is not incorporated at codons specifying Asn. This discrimination process via EF-Tu was termed thermodynamic compensation and adds to the accuracy of the translational process [82].

8.2.5 Mimicry at the Ribosomal A-site

The A-site is not restricted to binding tRNAs exclusively. During the various stages of the elongation cycle, a number of translational factors interact at the A-site. The first structures determined for these translational factors were those of EF-G [83, 84] and EF-Tu [83, 85]. Interestingly, the structure of the latter, in the form of a ternary complex EF-Tu·GTP·tRNA [86], had a striking similarity to that of EF-G·GDP, such that domains 3–5 of EF-G closely mimic the tRNA in the ternary complex (Figs. 8-6A and B; reviewed by Nissen et al. [87]). This suggested that the binding pocket of the A-site constrains the translational factors binding there to conform to a tRNA-like shape. In the last few years, solution structures for various termination factors, such

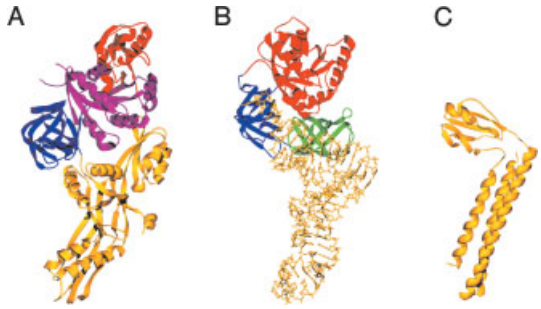


Figure 8-6 Molecular mimicry of tRNAs by translation factors. Comparison of the crystal structures for (A) EF-G GDP with domains 3–5 in gold (pdb1fmn) [148], (B) EF-Tu GTP tRNA (pdb1ttt) [86], (C) RRF (pdb1eh1) [149], figures of crystal structures were generated with Swisspdb viewer [150] and rendered with POVRAY.

as RF2, eRF1 and especially RRF (Fig. 8-6C) that also interact at the ribosomal A-site, have generally supported this concept. However, recent studies on the conformation and orientation of these factors on the ribosome suggest that the translation mimicry hypothesis has been over-extrapolated and that an overall tRNA shape does not necessarily suggest a binding orientation analogous to that of the tRNA (see Chap. 12 on termination for more details).

Mimicry of RNA by protein may be a more common feature in ribosomes than first realized. Organellar ribosomes generally have shorter rRNA components when compared with *E. coli*. Recent analyses of the chloroplast and mitochondrial ribosome components suggest that these rRNA losses are compensated for by both increases in size of the ribosomal protein homologs and the presence of additional organelle-specific ribosomal proteins [88–92]. Mitochondria represent an extreme example in that the protein component of the ribosomes represents two-thirds of the mass instead of one-third as in *E. coli* ribosomes. It is worth mentioning that the rRNA does consist predominantly of universally conserved residues that locate to the active centers of the ribosome, i.e., the decoding center on the 30S subunit and the PTF center on the large subunit [93], thus reinforcing the importance of these regions.

8.2.5 Translational Errors

Three types of ribosomal errors can be distinguished, the underlying mechanisms of which, not only partially overlap but are also intimately related: (i) a simple mistake in the decoding of a codon, (ii) a processivity error, and (iii) a loss of the correct reading frame (frameshift). A decoding error can lead to incorporation of an incorrect aminoacyl residue into the nascent peptide chain. A processivity error is defined as the release of a prematurely short peptidyl-tRNA from the ribosome. A shift in the

reading frame usually means the immediate loss of the genetic information, and will lead to the release of the synthesized peptide from the ribosome due to the appearance of a premature stop codon in the A-site. From the frameshift, peptide release will occur statistically within about 20 decoding steps in the incorrect reading frame since three of the 64 codons are stop codons.

Incorrect incorporation of an amino acid at a sense codon is termed a missense substitution and occurs with a global average of $\sim 3 \times 10^{-4}$ [94]. The third nucleotide of a codon is misread the most often, followed by the first one; for the decoding process the middle nucleotide of a codon seems to be the most important and is misread with the lowest detectable frequency. The codon lexicon is arranged in a way that an error in the reading of the third nucleotide of a codon results in the incorporation of either the same or a similar amino acid into the nascent chain, i.e., one with chemical properties similar to the correct amino acid such as charge or hydrophilicity. In this way, the effects of the missense substitution are buffered and usually do not lead to disastrous malfunctions of the corresponding protein. According to a rough estimate, only 1 in 400 missense events will completely inactivate the product [95].

Mis-incorporations are not only due to ribosomal decoding errors, but can be caused at the synthetase level via mis-charging. However, the charging mistakes of synthetases are usually below the level of the ribosomal mis-reading and can reach a precision of 1 mis-charging in 100 000 charging events [96]. Owing to the generally high accuracy of the synthetases, the mis-charging effects are negligible for protein synthesis (see Chap. 4 for charging mechanisms of the synthetases). EF-Tu also participates in preventing mis-incorporations in the frame of the thermodynamic compensation mechanism described in Sect. 8.2.4, whereas the discrimination cognate versus noncognate is discussed in Sect. 8.2.2 and that between cognate versus near-cognate in Sect. 8.2.3.

In the rare cases of processivity errors, a stop codon can be translated (termed readthrough) by a ternary complex aminoacyl-tRNA·EF-Tu·GTP leading to an extension on the protein product. However, usually processivity errors are premature drop-offs of the peptidyl-tRNAs, the frequency of which has been estimated to be around 4×10^{-4} (i.e., four drop-offs in 10 000 amino acid incorporations; [97]). The ester bond linking the peptidyl residue to the tRNA is more stable than the corresponding bond of an aminoacyl-tRNA. This means that peptidyl-tRNAs would accumulate in the cell over time, thereby sequestering the tRNAs and prohibitively restricting protein synthesis. Therefore, the existence of the enzyme peptidyl-tRNA hydrolase is essential for cell viability, since it cleaves the relevant ester bond (with the exceptions of fMet-tRNA and aminoacyl-tRNA), thus recycling the tRNAs. In studies with a protein of more than 1000 amino acids (β-galactosidase), the fraction of initiating ribosomes that did not complete the synthesis was estimated to > 20%, and up to half of the effect was caused by a premature drop-off, the other half by truncated mRNAs resulting from an abortive transcription of the *lacZ* gene [98]. The probability of a premature drop-off is probably not identical for every codon, but is rather sensitive to context effects and occurs more often with short peptidyl-tRNAs.

Nevertheless, the energy impact of processivity errors seems to be more severe than that of missense errors, the latter being the "standard" mistake during the decoding process. An estimation of the energy loss caused by processivity errors amounts to 3% of the total energy turnover of rapidly dividing cells.

Truncated mRNA may trap a synthesizing ribosome since a stop codon necessary to provide an organized termination event is absent. Bacteria contain a stable RNA of about 350 nt that rescues these trapped ribosomes. This RNA (10Sa RNA or tmRNA) can be charged with alanine by the corresponding synthetase, occupy the A-site, and after the nascent peptide has been transferred to the alanyl residue (tRNA function) can function as a mRNA and by doing so add a 10-amino-acid peptide tag to the nascent chain, allowing an ordered termination event via a programed stop codon. Owing to dual tRNA and mRNA functions of this RNA, it has acquired the name, tmRNA. The functions of the tmRNA are (i) to tag the abortive peptides with an additional sequence at the C-terminus that destines the peptide for efficient degradation and (ii) to recycle trapped ribosomes [99]. Although truncated mRNAs are not rare, the rescue of trapped ribosomes does – at least in some organisms – not depend solely on the presence of tmRNA, since null mutants are viable in *E. coli* although not in *Baccillus subtilis* at higher temperatures [100]. The precise mechanism of tmRNA action is not known (see Chap. 11 for more details).

What causes processivity errors to occur? Several mechanisms can cause a processivity error but the predominant cause is probably an event shortly after the onset of protein synthesis, i.e., the insertion of the growing peptide chain into the ribosomal tunnel. This tunnel can harbor a sequence of about 30 amino acids before the growing peptide chain emerges from the back of the 50S subunit into the cytosol surrounding the ribosome. The macrolide antibiotics, such as erythromycin, cause accumulation of short oligo-peptidyl-tRNAs ranging in size from 1 to 8 amino acids long depending on the macrolide (see Chap. 12). This occurs because these antibiotics, by binding within the tunnel, prevent egress of the nascent polypeptide chain and therefore induce drop-off. Another mechanism is a false stop, i.e., a sense codon is incorrectly recognized by a release factor leading to termination of protein synthesis. A false stop is a rare event with a probability of about 10^{-6} per codon [101]. Finally, frameshifts can lead to protein fragments as mentioned previously.

Since loss of the reading frame would lead to an immediate loss of the genetic information, maintaining the reading frame is an essential task of the ribosome. Usually, a loss in the reading frame occurs only once in 30 000 elongation cycles [98]; however, at the recoding site of the RF2 mRNA, where a + 1 frameshift at the 26th codon is essential for production of the full-length and active RF2 protein, loss of reading frame occurs with an efficiency of between 30 and 50%. Under certain conditions the frameshifting frequency can reach almost 100% efficiency, i.e., four orders of magnitude more often than that observed with other mRNAs. Obviously, there must be a ribosomal mechanism for maintaining the reading frame that is switched-off during RF2 synthesis. A detailed analysis has revealed that it is the presence of a cognate tRNA at the E-site that is essential for maintaining the reading frame. *In vitro* experiments show that the absence of an E-tRNA allows frameshift events to occur with a frequency of up to 20%, whereas in the presence of an E-site

tRNA no frameshifting was observed (V. Marquez, D.N. Wilson, W.P. Tate, F. Triana-Alonso and K.H. Nierhaus, in press). This does not mean that all recoding events are triggered by a pre-mature release of the E-tRNA; a detailed discussion of other aspects of recoding events can be found in Chap. 10.

8.3 The PTF Reaction

The PTF reaction is the central enzymatic activity of the large subunit. It occurs when a peptidyl-tRNA is located in the P-site and an aa-tRNA is in the A-site, termed a PRE state. Both L-shaped tRNAs at P- and A-sites form an angle of about 40° [11, 102, 23], whereas the acceptor stems are related by a translational movement, i.e. the CCA ends of both tRNAs at the PTF are related by an angle of approximate 180° and are, thus, in effect, mirror images of one another. The twist to accomplish this reflection occurs almost entirely between nucleotides 72 and 74 [103].

During PTF the α-amino group of the A-tRNA attacks the carbonyl group of the peptidyl residue of the P-tRNA, which is linked by an ester bond to the tRNA moiety (Fig. 8-7). This forms a tetrahedral intermediate, which resolves to yield a peptidyl bond. As a result, the aa-tRNA becomes a peptidyl-tRNA prolonged by one aminoacyl residue, and the former peptidyl-tRNA is stripped of its peptidyl residue to become a deacylated tRNA without a significant change of the place of the tRNA moieties [11, 104].

A long-standing debate within the translation field concerned whether or not the PTF reaction is catalyzed by proteins or rRNA. The PTF center was identified by using a putative transition state analog of the PTF reaction, which was soaked into crystals of the 50S subunit from *Haloarcula marismortui* [105]. This analog, which has been introduced by the Yarus group and hence termed the Yarus inhibitor, is a mimic of the CCA end of a P-tRNA attached to puromycin in the A-site (inset in Fig. 8-7) and is a strong competitive inhibitor of the A site substrate [106]. The region moulding the binding site of the inhibitor is densely packed with highly conserved bases of the 23S rRNA, mainly derived from the so-called PTF ring of domain V. The PTF ring structure with 41 nucleotides (Fig. 8-8) is one of the most highly conserved in rRNA, and its PTF involvement is supported by crosslinking studies from the acyl residues of tRNAs at A- and P-sites (see, e.g., Ref. [107]) as well as by mutations that render cells resistance against many antibiotics blocking peptide-bond formation (see Chap. 12 and Ref. [108] for review).

Although there are 15 proteins that interact with domain V of the 23S rRNA, only the extensions of proteins L2, L3, L4, and L10e come within 20 Å of the active site (Fig. 8-9). That the active center of the ribosome is made exclusively from RNA, implies that the ribosome is a true ribozyme.

Each of the four proteins L2, L3, L4, and L10e (a homologue of bacterial L16) has a globular domain connected to a long extension that penetrates deeply into domain V and approaches the active site. Such a long extension is quite common to ribosomal proteins and it is thought to play a role as a "glue" for the quaternary structure of the

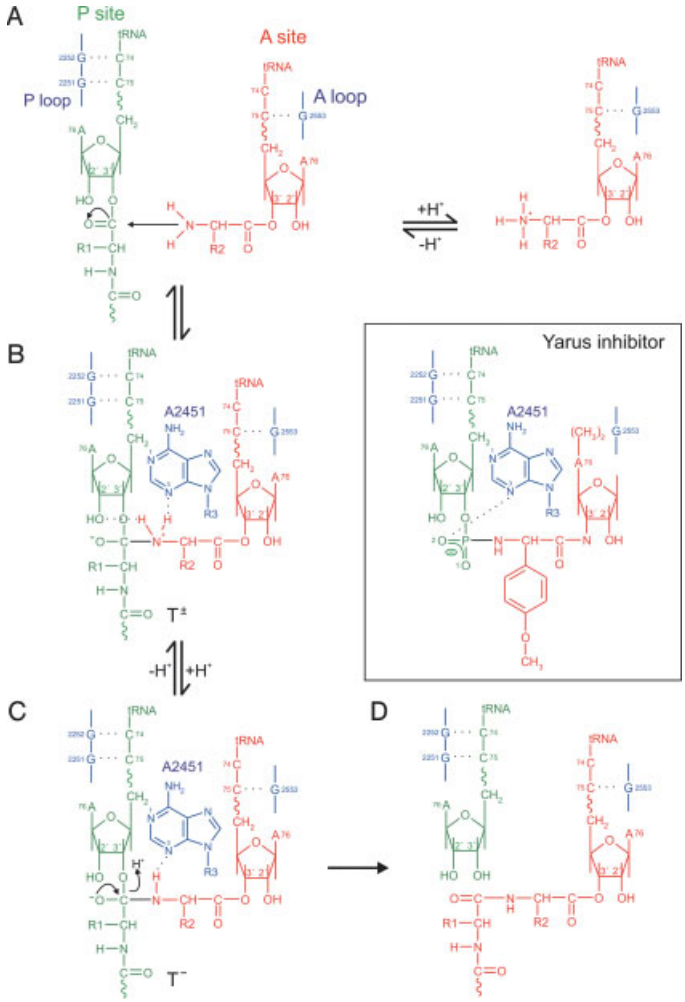


Figure 8-7 The PTF reaction. The figure shows the four possible steps of peptide-bond formation according to recent crystallographic and biochemical data [152, 103, 105, 122]. The essential features are (A) C74 and C75 of the P-site tRNA (green) are Watson–Crick paired with G2252 and G2251, respectively, of the P loop (blue). Similarly, C75 from the A-site substrate (red) forms a Watson–Crick base pair with G2553 (A-loop). The α-NH_2 function of the A-site aminoacyl-tRNA is an ammonium ion at pH 7 [153]. (B) Deprotonation of the ammonium ion triggers the nucleophilic attack of the α-amino function on the carbonyl group of the P-site substrate, which results in the tetrahedral intermediate T±. The secondary α-NH_2 group forms a hydrogen bond with N3 of A2451 and a second with either the 2′-OH of the A76 ribose at the P-site (shown here) or alternatively with the 2′-OH group of A2451. The oxyanion of the tetra-hedral intermediate points away from the N3-A2451 [103] and thus cannot, in contrast with the previous proposal [105], form a H-bridge. (C) Further deprotonation of the secondary α-NH_2 group leads to the tetra-hedral inter-mediate T- and the PTF reaction is completed by an elimination step. (D) The peptidyl residue is linked to the aminoacyl-tRNA at the A-site via a peptide bond.

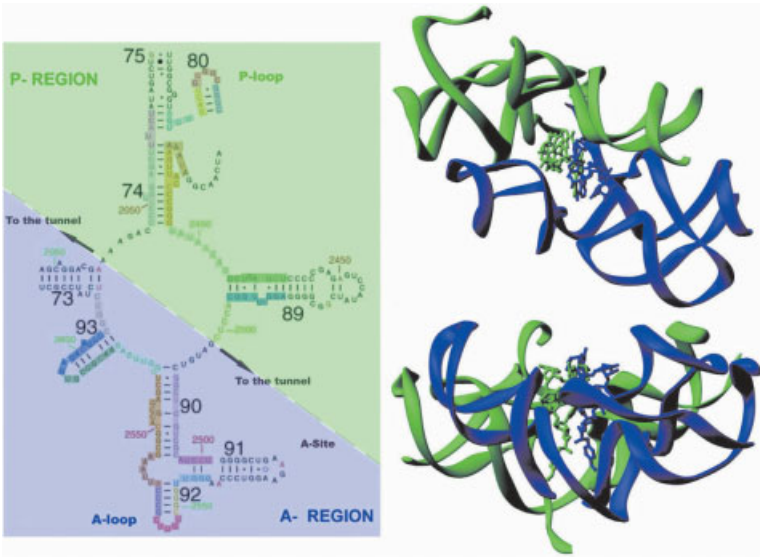


Figure 8-8 Secondary structure of the domain V of the *E. coli* 23S rRNA. Left: the A-site (blue) and P-site (green) regions that are related by 2-fold symmetry, where the symmetry-related residues within these regions are highlighted with the same color. Right, The 2-fold symmetry is illustrated from two different views using ribbon representations of the PTF center from *D. radiodurans* 50S subunit, with the A- and P-site CCA-end ligands indicated in the corresponding colors. This figure was taken from Ref. [146] with permission.

ribosome (see Chap. 1). Interestingly, three out of the four proteins in the vicinity of the PTF center are also present in eubacterial ribosomes, viz. L2, L3, and L4. These proteins have been identified previously together with 23S rRNA as major candidates for the PTF activity by single-omission tests in a total reconstitution system of the large subunit [109, 110].

One significant difference between the *Deinococcus radiodurans* 50S structure and that from *H. marismortui* is the presence of protein L27, a protein that has no homolog counterpart in the latter archaea organism [111]. L27 is one of the few proteins that are present in the interface region of the 50S subunit. It has been proposed that L27 plays a role in placement of the CCA ends of the A- and/or P-site tRNAs, and based on docking of the tRNAs from the *T. thermophilus* 70S:tRNA$_3$

Figure 8-7 contd. The inset is the Yarus inhibitor CCdAp-puro-mycin (CCdApPmn), which was used to identify the PTF center of the ribosome. The interactions of the Yarus inhibitor with the rRNA were deduced from 50S crystals of *H. marismortui* ribosomes after soaking the inhibitor into the crystals. Note that it was concluded that the protonated N3 of A2451 makes a H bridge to O2, which was thought to mark the position of the oxyanion of the tetrahedral intermediate (transition state) formed during peptide-bond formation [105] (cf. with the probably correct representation in step (B).

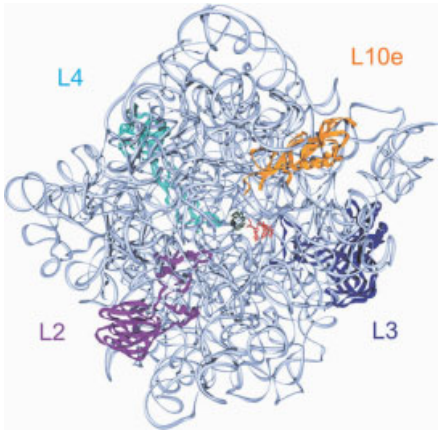


Figure 8-9 The A- and P-site products in red and green, respectively, bound at the PTF center of the 50S subunit. The proteins that reach within ~20 Å of the PTF center include proteins L2 (purple), L3 (blue), L4 (cyan) and L10e (brown). This figure was generated from pdb file 1KQS [25] using Swisspdb viewer [150] and rendered with POVRAY.

structure into the *D. radiodurans* 50S structure, contact of the CCA ends with L27 were predicted [111] Indeed, photoreactive derivatives of yeast NAc-tRNA^{Phe} containing 2-azidoadenosine at their 3′-termini could be crosslinked to L27 when bound at the ribosomal P-site [112]. Recently, Zimmermann and co-workers [113] have shown that it is the very N-terminal of L27 which is crosslinked, since deletion of the 3–6 amino acids severely reduced the crosslinking efficiency and deletions of more than nine amino acids totally abolished crosslinking altogether. Therefore, L27 is more probably involved in tRNA positioning rather than in the PTF reaction itself and, furthermore, seems to be specific for only eubacteria.

The debate has now turned to whether the PTF reaction follows a physical or in addition also a chemical principle.

8.3.1 A Short Intermission: Two Enzymatic Principles of PTF Activity

There are two main principles associated with enzymatic reactions: a chemical and a physical. A number of examples exist where one or other principle is predominant but they need not be mutually exclusive (see Ref. [1] for review).

8.3.1.1 Chemical Concept: A Transient Covalent Bond between Active Center and Substrate(s)

Although the serine proteases subtilisin in bacteria and chymotrypsin in mammals have arisen independently during evolution, both have an identical activation center

for clearing the peptide bond containing a triad of Asp, His and Ser (Fig. 8-10). The three amino acid residues participate directly in the catalysis through a transient covalent event. The Asp–His module executes a general acid–base catalysis consisting of a proton donation (the acidic step), and a proton-accepting step (the basic step). The nucleophilic serine residue attacks the first substrate (which can be a peptide or an ester) forming a covalent acylated enzyme intermediate, i.e., after cleavage of the peptide bond the Ser residue binds the carbonyl residue transiently forming a seryl ester. The serine residue is then displaced via a nucleophilic attack of the second substrate (which can be an amine, a water molecule, or an alcohol). Importantly, the His residue has a pK_a of about 7, making it excellently suited to function in this type of catalysis which requires steps of both proton donation and acceptance at pH 7.

In the case of the PTF reaction, we can replace the intermediate seryl ester by a peptidyl-tRNA, where the peptidyl residue is also linked to the tRNA body via an ester bond. After nucleophilic attack by the α-amino group, a covalent intermediate is formed between the peptidyl and aa-tRNA. This intermediate complex with a tetrahedral carbon and a negatively charged oxygen is unstable and decomposes to give a peptidyl moeity (the nascent chain) linked to the aminoacyl-tRNA via a peptide bond in the A-site and a deacylated tRNA at the P-site. Both reactions occur equally well with an alcohol [114] or, with a water molecule instead of an aminoacyl-tRNA, in the case of termination reaction. The mechanism requires (i) the activation (deprotonation) of the nucleophilic α-amino group of the aa-tRNA by a general base catalyst, e.g., a His–Asp system as in the serine proteases, (ii) the stabilization of the

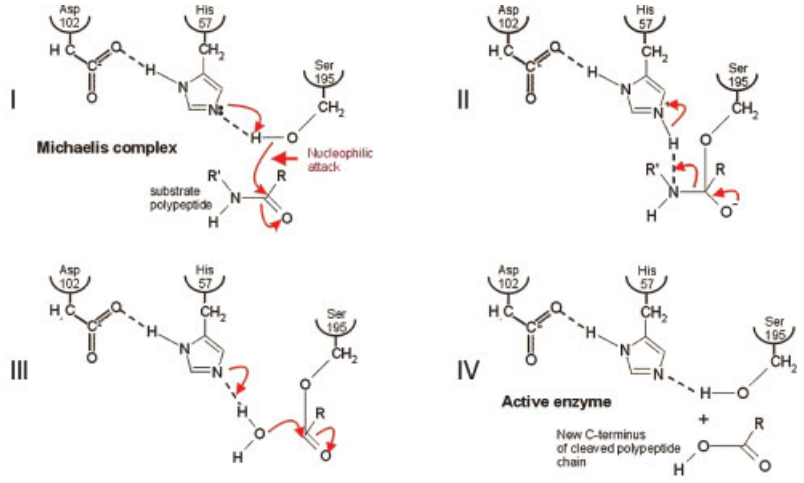


Figure 8-10 The mechanism of peptide-bond hydrolysis of serine proteases. Three amino acids Ser, His and Asp participate in this reaction. (I) The nucleophilic hydroxyl anion of the serine residue attacks the carbonyl group of the substrate (peptide bond) forming (II) a covalent acyl-enzyme intermediate. (III) The serine molecule is then displaced via a nucleophilic attack by a water molecule. (IV) Both acyl formation and breakdown proceed via a normally high-energy tetrahedral intermediate.

tetrahedral intermediate resulting from the nucleophilic attack of the aa-tRNA on the ester linkage of peptidyl-tRNA, and (iii) the activation of the tetrahedral intermediate, which then breaks down due to proton donation from the general acid catalyst. To avoid a side reaction with a water molecule, i.e., hydrolysis of the peptidyl-tRNA during PTF reaction, the PTF center must be in a hydrophobic pocket. However, in the crystal structure of 50S (1JJ2.pdb released upon the publication of Ref. [115]) the residue A2451, which is in proximity of the PTF catalytic center (see the next section), is surrounded by 56 water molecules within a distance of 10 Å. This finding questions the assumption that the identified region for peptide-bond formation is in its fully active state, although it is competent to form a peptide bond [25].

The essential involvement of general acid–base catalysis in peptide-bond formation is consistent with the following experimental data: (i) If an enzymatic reaction is retarded in the presence of heavy hydrogen D (D_2O) instead of H (H_2O), the obvious conclusion is that a general acid–base catalysis is involved, since the migration of D versus H is slower. Precisely, such an effect has been observed for peptide-bond formation [1]. (ii) The pH dependence of peptide-bond formation peaks at pH 7, similar to the pK_a curve for a His residue [114, 116].

It was shown that the PTF activity can be blocked by His-modification reagents and that inactivation follows a one-hit kinetics of modification of His residues on 50S subunits [117], indicating that one His residue is essential for peptide-bond formation. Furthermore, phenyl-boric acid which reacts specifically with His residues, blocks peptide-bond formation [118]. Taken together, the interpretation was that a His residue might mediate a general acid–base catalysis as is known from the case of serine proteases. These suggestions of a catalytic mechanism, based on general acid–base catalysis and a possible involvement of a His residue, lead to the idea that the mechanism of serine proteases is exploited by the ribosome as well. A candidate for this His residue is His^{229} of L2 (see Ref. [54] for discussion and references therein). The recent X-ray analysis of 50S crystals demonstrates that a His–Asp module as in the case of serine proteases does not seem to apply, since no proteins are within 18 Å of the active site (see the next section; [105]). The role of a His residue as being seemingly critical for peptide-bond formation is therefore still unclear.

8.3.1.2 **Physical Concept: The Template Model**

The essence of this concept is that an enzyme organizes a defined stereochemical arrangement of the two substrates that are to be covalently bonded. The stereochemical arrangement is sufficient to allow for a dramatic acceleration of the reaction rate by 10^6–10^9-fold. This concept does not require any direct chemical involvement in the catalysis of the reaction such as a transient covalent binding of the substrate(s) to the enzyme.

Let us consider the acceleration factor provided by the ribosome. The upper limit of the rate for a ribosome-free environment can be estimated from a reaction between NH_2OH and AcPhe-tRNA (see Ref. [1]). The rate for the nucleophilic attack to form an ester or a peptide bond is $6 \times 10^{-5}\ M^{-1}\ s^{-1}$, which means one peptide bond

is formed in 30 h. On the ribosome, one peptide bond is made in 50 ms (15–20 peptide bonds per ribosome per second [119]). From these numbers (15–20) $s^{-1}/6 \times 10^{-5}\ M^{-1}\ s^{-1}$, an acceleration factor of 3×10^5 M can be calculated; in other words, the ribosome accelerates the reaction by a factor of 3×10^5 M.

Kinetic data from organic chemistry demonstrate that a rate factor of this magnitude can be obtained from simple model compounds where the reactants are appropriately juxtaposed. Bruice and Benkovic [120] have shown that the rate of intramolecular amine attack in phenyl-4(dimethylamino) butyrate (I in Fig. 8-11) is 1.3×10^3 M faster than bimolecular trimethylamine attack on phenyl acetate; a similar enhancement of the rate has been observed for intramolecular reaction in succinate half ester anion (II; Fig. 8-11) as compared with bimolecular reaction of acetate ion with phenyl acetate. A further rate enhancement is seen with an intramolecular reaction in the rigid ester anion (III) that proceeds 2.3×10^2 times more rapidly than that in (II), because in (III) the oxyanion nucleophile and ester are more rigidly fixed and thus better suited for the reaction, whereas (II) has rotational freedom to adopt nonproductive conformations. These results allow the prediction that if the ribosome were simply to hold the α-amino nitrogen of an aminoacyl-tRNA in the same position relative to the ester linkage of a peptidyl-tRNA as the carboxylate anion is held relative to the ester linkage in (III), peptide-bond formation would occur spontaneously with a rate comparable to the *in vivo* rate of ribosomes. The more rigid the reactants are fixed in a favorable stereochemistry, the faster the reaction proceeds. In other words, the rate of peptide-bond performance of the ribosome can be explained exclusively using the template model.

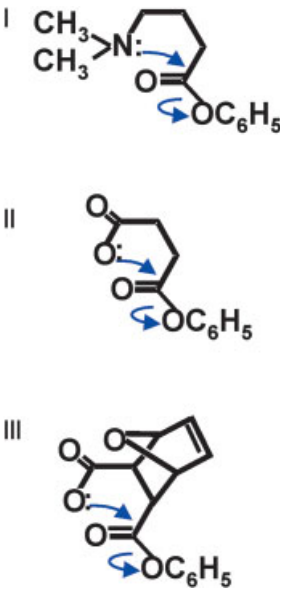


Figure 8-11 Peptide-bond formation in model compounds with appropriate juxtaposed nucleophile. See text for explanations.

8.3.2 Data from the Crystal Structures

The universally conserved residue A2451 of domain V is the nearest base to the Yarus-transition analog. It was thought to be a good candidate for a general acid–base catalyst, since its N3 is about 3 Å from the oxygen and 4 Å from the nitrogen of the phosphoamide in the Yarus inhibitor (Fig. 8-12A). This proposal was strengthened when the p*K* value of A2451 was found to be abnormally high at neutral pH (> 7, which is 6 pH units higher than expected) [121, 122], a property essential for acid–base catalysis as it allows for easy donation and withdrawal of a proton from the α-amino group of the aa-tRNA at the A-site. According to the same model [105], protonation of A2451 would also allow formation of a hydrogen bond with the carbonyl oxyanion of the tetrahedral transition state analog (inset of Fig. 8-7). However, the pH dependence of DMS modification at position 2451, with which the abnormally high p*K* of this residue was demonstrated, was subsequently shown to be displayed only by inactive ribosomes [123]. Shortly following, several groups reported that A2451 was not essential for peptide-bond formation, since ribosomes bearing mutations at position 2451 exhibited only modest (2–14-fold) decreases in the rate of peptidyl transfer [124] and were instead shown to be defective in substrate binding [125].

The next glimpse of a stage in the PTF reaction was that of the final products, obtained by soaking of A- and P-site substrates into enzymatically active *H. marismortui* 50S crystals [25]. The activity of the crystals was demonstrated by addition of the native 50S crystals to a solution containing A- and P-site substrate analogs which resulted in product formation, and ceased upon removal of the crystals. The structure determined to 2.4–3.0 Å was post-peptide-bond formation but pre-translocation and showed that the deacylated CCA bound to the P-site had its 3′-OH in close proximity to the N3 of A2451 (Fig. 8-12B) [25].

Counterevidence against the involvement of A2451 in stabilizing the transition state analog followed: If the oxyanion of the tetrahedral intermediate is hydrogen-bonded to the N3 of A2451, then this N3 must be protonated at around pH 7 and therefore should lose its proton at pH > 7.3. In this case, one would expect a strong pH dependence of the affinity of the Yarus inhibitor, since the hydrogen bond would contribute significantly to the affinity. To test this hypothesis, Strobel and co-workers [122] determined the affinity of the Yarus inhibitor for the 50S subunit at all pH values between 5 and 8.5 and found that it remained unchanged, a result inconsistent with the idea that the oxyanion is stabilized via a H-bond to the N3 of A2451. The same conclusion was drawn by subsequent crystallographic studies showing that the oxyanion of the tetrahedral intermediate points away from N3 of A2451, thus excluding a possible hydrogen bonding between these two atoms [103].

Furthermore, the Yarus inhibitor is not an honest mimic of the transition state: the distance between the O2 of the Yarus inhibitor and the 2′-C of the deoxy-A76 (dA76) ribose at the P-site is only 2.8 Å (arrowed in Fig. 8-12A). A physiological P-site substrate contains a 2′-OH at this 2′-C atom, which is essential for peptide-bond formation [126], and would sterically clash with the O2 in the position of the Yarus

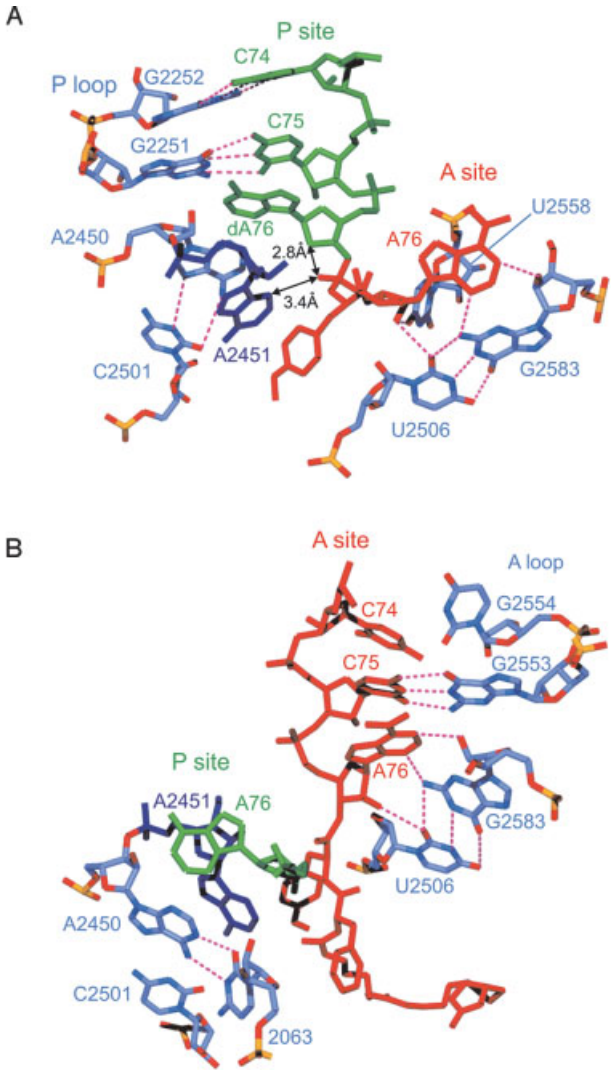


Figure 8-12 Tight fixation of the CCA ends of the P- and A-tRNAs observed in 50S subunit from *H. marismortui* in complex with (A) the Yarus inhibitor, and (B) the products following peptide-bond formation. The CCA ends of the tRNAs in the A- and P-sites are colored red and green, respectively. The N3 of A2451 (dark blue) is 3.4 Å from the O2 of the Yarus inhibitor (see also inset in Fig. 8-7), whereas the same O2 is only 2.8 Å from the 2′-deoxy of A76 (arrowed). Selected rRNA residues of domain V of the 23S rRNA are colored light blue, including the A- and P-loop bases that participate in A- and P-site CCA end fixation (*E. coli* numbering). In (B) the P-site C74 and C75 have been omitted for clarity. Dashes indicate H-bonding and rRNA nucleotides use the following color scheme: oxygen, red; phosphorus, yellow; nitrogen, blue; carbon, dark blue. (A) and (B) were generated from pdb files 1FFZ [105] and 1KQS [25], respectively, using Swisspdb viewer [150] and rendered with POVRAY.

inhibitor observed in the crystal. The essential nature of this 2′-OH might be explained by the observation that the α-NH_2 possibly forms a hydrogen bond with this OH (illustrated in Fig. 8-7B).

The general base catalysis debate flared up again, when Rodnina and co-workers [127] presented evidence that peptide-bond formation depends on two ionizable groups, one with a pK_a of 6.9 and the other with a pK_a of 7.5. The former was shown to be associated with the α-NH_2 group of puromycin used in the kinetic experiments, whereas the latter seemed to be ribosome-associated. The ionizable group evidently belongs to A2451 since a ribosome bearing an A2451U mutation catalyzed peptide-bond formation ~130 times slower than normal and had lost the pH dependence associated with the titratable group at a pK_a of ~7.5. However, an alternative explanation suggested by the authors was that the protonated group is part of the A2450:C2063 base pair lying directly behind A2451 (seen in Fig. 8-12B) – the candidate-ionizable group being the N1 of A2450. Although a distance of 7 Å from the N1 to the a-NH_2 group is too long for hydrogen transfer, a postulated conformational change of the PTF center might bring A2450 within the range [127]. The assumption of conformational changes broadens again the number of possible candidates that might play a role in the kind of chemical catalysis that is advocated here. We note that evidence has been presented implicating His^{229} of protein L2 in this catalysis [54], although current maps place this residue more than 20 Å from the tetrahedral intermediate of the transition state. The best that can be said at the moment is that a direct role of A2451 in a general acid–base catalysis is hard to be reconciled with the observation that A2451 in active ribosomes does not contain a titratable group at this pK_a in contrast with inactive ribosomes [123].

In fact, the ribosome need not directly involve chemically in the catalysis of the PTF reaction, such as the formation of a transient covalent interaction between the substrate (tRNAs) and the enzyme (the ribosome or, more specifically, in this case the rRNA). The template model predicts that tight stereochemical arrangement of substrates relative to one another would be sufficient to provide the dramatic acceleration of the reaction rate needed for peptide-bond formation (see the previous section). In this case, the role of A2451 would be to withdraw a proton from free nucleophilic α-NH_2 group of the A-site substrate or form a hydrogen bond with the α-NH_2 group, thus promoting peptide-bond formation via proper positioning of the NH_2 group. The reaction scheme would follow similar to that presented in Fig. 8-7 and described in more detail in the corresponding legend.

Tight fixation of the CCA ends of the P- and A-tRNAs is exactly what is observed both in the analog soaked 50S crystal structure (Fig. 8-12A) and also with the 50S structure containing the products of the PTF reaction following soaking of the A- and P-site substrates (Fig. 8-12B). In the P-site, the CCA end is locked into position by two Watson–Crick base pairs, viz. C74 and C75 with G2252 and G2251. A76 stacks on the ribose of A2451 (seen clearly in Fig. 8-12B). In the A-site, the CCA end of the aa-tRNA is fixed by (i) Watson–Crick base-pairing between C75 and G2553, (ii) a type-I A-minor motif between A76 and the G2583–U2506 base pair, and (iii) an additional H-bond interaction between the 2′-OH of A76 with U2585. Tight fixation

of the CCA ends of both A- and P-tRNAs at the PTF center underlines the importance of the template model in peptide-bond formation, viz. that precise stereochemical fixation is predominantly responsible for the enormous acceleration of the reaction. The rate of peptide-bond formation on the ribosome of ~50 s^{-1} was estimated to be ~10^5 faster than the uncatalyzed reaction (the rate in the absence of ribosomes) [1]. The PTF reaction without chemical catalysis (i.e., ribosomal ionizing group is protonated at pH < 7) occurs with a rate of ~0.5 s^{-1} [127], which is still >1000 faster than the uncatalyzed reaction. If this estimation is correct then the physical mode of peptide-bond formation represents approximately 90% of the reaction rate, with the chemical mode making up the remaining 10%.

8.3.3
Why both the Physical and Chemical Concepts for Peptide-bond Formation?

If the physical mode sufficiently explains the overall rate of protein synthesis, why does the chemical mode exist in addition? A possible explanation is that the template model requires a maximal rigid surface for the ligands, whereas the ribosome has to be flexible to allow the movement of the mRNA•$tRNA_2$ complex during translocation. To achieve both a fast rate and a flexible structure of the ribosome, both physical and chemical catalysis might need to be involved for an optimal rate of protein synthesis, although it is already clear that the template (physical) mode probably plays the dominant role in peptide-bond formation.

8.4
The Translocation Reaction

Following peptide-bond formation, the positions of the tRNAs remain unchanged. This has been demonstrated by cryo-EM analyses of *E. coli* ribosome complexes [11, 104] and – concerning the CCA ends at the PTF center – by soaking A- and P-site substrates into active 50S *H. marismortui* crystals and solving the structure of the reaction products after peptide-bond formation [25]. Now the ribosome must transfer the products, the peptidyl-tRNA in the A-site and deacylated tRNA in the P-site, to the P- and E-sites respectively, i.e., shifting the ribosome from PRE to POST state (Figs. 8-1(c) and (f), respectively). This process is termed translocation. It must be extremely accurate at both ends of the tRNA molecule: the anticodon–codon complex must be moved exactly 10 Å (the length of one codon), longer or shorter movements will change the reading frame. At the other end of the A-site peptidyl-tRNA, the CCA end must also be precisely moved into the P-site so as to set up the next peptidyl-transferase reaction with the incoming A-tRNA. Incorrect placement of the peptidyl-tRNA at the P-site could be disastrous for peptide-bond formation and lead to abortion of translation.

Ribosomes have an innate translocase activity [14], but it is more than one order of magnitude slower than that of the EF-G catalyzed reaction [13]. This implies that the structures necessary to move the tRNAs reside in the ribosome and that the role of

EF-G (and EF2 in eukaryotes) is to reduce the activation energy barrier that separates the two sets of tRNA positions. Important questions remaining unanswered are: How does EF-G mediate translocation and what ribosomal components are involved in transfer of the tRNAs?

8.4.1 Conservation in the Elongation Factor-G Binding Site

The crystal structure of EF-G has been solved in the nucleotide-free form [83] and in complex with GDP [84]. The GTP form is the active one, which binds to the ribosome and triggers translocation. Hydrolysis of GTP inactivates EF-G and dissociates it from the ribosome (reviewed in Ref. [128]). EF-G belongs to the same subfamily of G proteins as IF2, RF3, and EF-Tu, the latter of which has been crystallized in both GTP (active) and GDP (inactive) forms. Comparison of these EF-Tu structures reveals that they exhibit large domain shifts relative to one another [85].

Cryo-EM reconstructions of EF-G bound to bacterial 70S ribosomes at 17.5–20 Å [7, 129, 9] and EF2 to eukaryotic 80S ribosomes at 17.5 Å [130] show similar binding sites for both factors (Figs. 8-13A and B). In these complexes, antibiotics were used

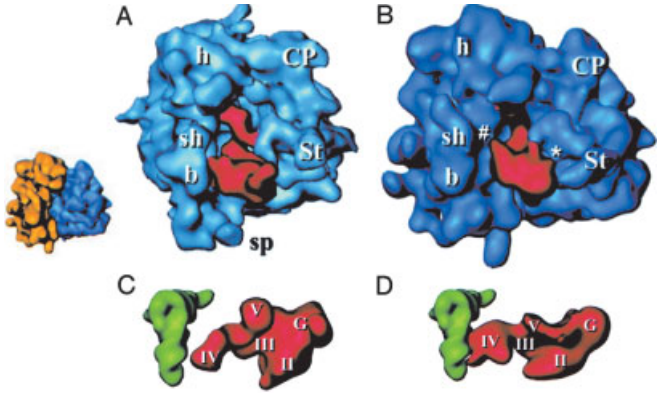


Figure 8-13 Comparison of cryo-electron microscopic analyses of EF-G 70S complex from *E. coli* and EF2 80S complex from *S.cerevisiae*. Side view of (A) EF-G 70S complex and (B) EF2 80S complex with small subunit on left and large subunit on right. The same orientation is seen, in the small inset on the left, of an empty 80S ribosome, where the 40S and 60S subunits are colored yellow and blue, respectively. Relative arrangements of (C) EF-G and P-tRNA and (D) EF2 and P-tRNA (where the P-tRNA was placed into the 80S density map on the basis of the position observed in a P-tRNA-bound 70S ribosome reconstruction). Landmark abbreviations are for the small subunit: b, body; bk, beak; h, head; sh, shoulder and sp, spur. Landmarks for the large subunit: CP, central protuberance; St, L7/L12 stalk. The roman numerals on EF-G/EF2 refer to the domains. Figures modified from Gomez-Lorenzo et al. [130].

to trap the elongation factor on the ribosome. EF-G was trapped on the 70S ribosome using the antibiotic fusidic acid, which allows translocation and GTP hydrolysis, but blocks the switch into the GDP conformer of the factor, preventing dissociation from the ribosome. The eukaryotic eEF2 was locked on the yeast ribosome using the antifungal sordarin, which is thought to function analogously to fusidic acid [130]. The EF-G complex was formed with a typical PRE state, i.e., A- and P-tRNAs were present. As expected, the tRNAs were translocated to the P- and E-sites, but of special interest is that the tip, domain IV, of EF-G was shown to occupy the position of the A-site. Similarly, the position for EF2 seen in the EF2:ribosome complex also occupied the A-site, and came very close to the position of the P-tRNA (Figs. 8-13C and D). EF-G-mediated translocation is also possible in the presence of nonhydrolyzable GTP analogs, such as GDPNP, thus suggesting that binding of EF-G alone is sufficient for translocation and that hydrolysis is necessary for conformational change and release of EF-G·GDP [128]. We repeat that the fact that translocation is an intrinsic activity of the ribosome has been shown by the factor free "spontaneous" translocation [131]. Furthermore, a dipeptidyl-tRNA that has been formed at the A-site via peptide-bond formation shows a much faster spontaneous translocation than a chemically made dipeptidyl-tRNA that has been bound to the A-site [13]. The energy for translocation therefore might arise by the peptide-bond formation, at least in part. The surprising observation that the antibiotic sparsomycin, an efficient blocker of peptide-bond formation, is able to effectively trigger a single translocation has corroborated the view that it is not the energy from the EF-G-dependent GTP hydrolysis that drives the translocation [132].

Precisely the opposite view, namely that EF-G-dependent GTP hydrolysis limits the translocation reaction thus accelerating the translocation, has been put forward by one group [37]. In this case, EF-G would be considered a "motor protein", although this view has not yet been accepted by the wider scientific community. One reason might be that the same group has published just the opposite results about 15 years ago: Also applying stop-flow measurements, they showed that the translocation reaction limits the EF-G-dependent GTP hydrolysis and not *vice versa*, namely that EF-G-dependent GTP cleavage occurs *after* translocation and thus the correspondingly released energy cannot directly flow into the translocation reaction [34].

For classical G-proteins, a GTPase-activating protein (GAP) stimulates the G-protein-mediated hydrolysis of GTP. In the case of EF-G/EF2, the GAP is provided by components of the ribosome. There are certainly gross conformational changes visible upon binding of each elongation factor to the ribosome. One of the most striking changes is seen within the stalk region; density for this region is absent in the empty 70S and 80S ribosomes but becomes more ordered upon EF-G/EF2 binding [7, 129, 130], supporting its universal role in factor binding. The best candidates for the GAP role include the pentameric stalk complex, which consists of the ribosomal proteins L10 $(L7/L12)_4$, as well as a region of the 23S rRNA termed the sarcin-ricin loop (SRL). The SRL is so named because cleavage after G2661 of the bacterial 23S rRNA within this region by the highly specific RNase α-sarcin inhibits all elongation

factor-dependent activities [133] and similar effects are seen after removing the neighboring base A2660 (*E. coli* nomenclature) in the 26S rRNA of yeast by the N-glycosidase activity of the ricin A-chain [134]. Furthermore, this region contains the longest (12 nucleotides) universally conserved stretch of rRNA underlining its functional importance. Recently, hybrid ribosomes were constructed where the proteins at the GTPase center from *E. coli*, L7/L12 and L10, were replaced with their eukaryotic counterparts from rat P1/P2 and P0, respectively [135]. Both the *in vitro* translation and GTPase activity of the resultant hybrid ribosomes were strictly dependent on the presence of the eukaryotic elongation factors, EF2 and EF1a. This reflects not only the specificity of the interaction between the stalk proteins and the elongation factors from each kingdom, but also the importance of the stalk proteins in mediating elongation factor GTPase activity.

The ribosomal protein L11 (and associated L11-binding site of the 23S rRNA) is often considered as a candidate for a GAP role and is often referred to as the GTPase-associated center (GAC) alone and collectively with the SRL (however to avoid confusion we will refrain from using this term, especially since L11 is unlikely to be a true GAC of the ribosome, see following). The reason that L11 has been assigned a GAP role is because mutations in both L11 and its binding site on the 23S rRNA can confer resistance against the antibiotic thiostrepton, a potent inhibitor of EF-G- and EF-Tu-dependent GTPase activities [136]. However, the direct involvement of L11 in the factor-dependent GTPase is not very probable, since (i) mutants lacking L11 are viable, although extremely compromised as indicated by about 6-fold growth retardation [137], and (ii) the IF-2-dependent GTPase is stimulated rather than blocked by thiostrepton [138]. Furthermore, replacement of either the L10 $(L7/L12)_4$ complex or L11 with the equivalent rat protein showed that the P0 $(P1\ P2)_2$ complex, but not the eukaryotic counterpart to L11 (RL12), was responsible for factor specificity and associated GTPase dependence, although addition of L11 or RL12 did stimulate protein synthesis significantly [135]. Consistently, L11 is certainly in close proximity to the elongation factors: cryo-EM analyses of EF-G bound to 70S ribosomes revealed that upon binding of EF-G, the N-terminal domain of L11 is shifted so as to form an arc-like connection with the G-domain of EF-G [139]. This arc-like connection is also observed in the EF2–80S complex although it is broader and more fused [130]. However, from cryo-EM reconstructions of EF-Tu ternary complex ribosome complexes stalled with the antibiotic kirromycin, it is the α–sarcin/ricin loop that makes direct contact with the G domain of the EF-Tu [104] and not the L11 region. Although residue 1067 within H43 of the L11-binding site rRNA makes definite contact with the highly conserved region of the T arm of the tRNA, no contact between L11 and the tRNA is observed in the recent 9 Å reconstruction [104], in contrast with earlier proposals that were based on lower resolution reconstructions [140]. The most striking result to emerge from the improved resolution was an observed movement of 7 Å in the L11 region upon ternary complex binding to bring this region into contact with the tRNA [104]. This leads the authors to propose a model, whereby the conformation change in the L11 region perturbs the orientation of the tRNA such that codon–anticodon interaction can be established,

which in turn provides the trigger for GTPase activation of EF-Tu. In this model, the L11 region would not act as a GAP directly, but would stimulate the activity consistent with the available biochemical data.

Lastly, it is noteworthy that L11 also stimulates the stringent response factor RelA in the presence of a deacylated tRNA at the ribosomal A-site, which leads to the RelA-dependent synthesis of (p)ppGpp, an essential effector during the stringent response (see Chap. 11).

8.4.2 Dynamics within the Ribosome

The α–ε model for translocation hypothesizes a moveable domain within the ribosome that carries the A- and P-tRNAs during translocation (reviewed in Ref. [43]). Evidence for this model comes from testing the accessibility of phosphate groups on the tRNAs in the PRE and POST states. The essential observation was that the protection patterns of A- and P-tRNAs differ from one another, but the corresponding tRNAs exhibit the same protection patterns in the PRE state as they do in the POST. This suggests that distinct ribosomal components are involved in carrying the tRNAs from the PRE to the POST state. In contrast with the intensive contact patterns observed with tRNAs bound to programed ribosomes, the 40 nucleotides of mRNA covered by the ribosome [141] have almost no specific contacts with the ribosome, except two positions upstream of the two decoding codons (i.e., those codons involved in codon–anticodon interactions; [142]). This immediately suggests that the tightly bound tRNAs are the handle to move the $tRNA_2$•mRNA complex during translocation, whereas the mRNA follows because of the direct connection with the tRNAs via the two adjacent codon–anticodon interactions. This is one important reason why the two adjacent tRNAs maintain their simultaneous codon–anticodon interaction before (A- and P-sites) and after translocation (P- and E-sites).

There are a number of candidates that may play a role in translocation of the tRNAs or even constitute portions of the moveable domains. Distinct regions within the crystal structures are disordered, most probably reflecting flexibility in these components. A classic example is the stalk region, which, as already mentioned, only becomes ordered upon factor binding. Another is the L1 region, the flexibility of which may regulate E-tRNA release. There are also certain structures that become either ordered or rearranged upon subunit association. Most of these elements are constituents of intersubunit bridges. One striking example is the universally conserved H69 in domain IV of the large subunit rRNA. H69 is the major element of bridge B2a, the largest intersubunit bridge (see Fig. 8-14), and is disordered in the *H. marismortui* 50S subunit structure but ordered in the *D. radiodurans* 50S subunit. Comparing the latter structure with the *T. thermophilus* 70S structure suggests that upon association H69 swings out towards h44, another very flexible element. In this extended conformation, H69 would be predicted to make contact with both A- and P-tRNAs [111]. Another element that is not fully resolved in either of the 50S structures is H38 of domain II, a constituent of bridge B1a, often called the "A-site

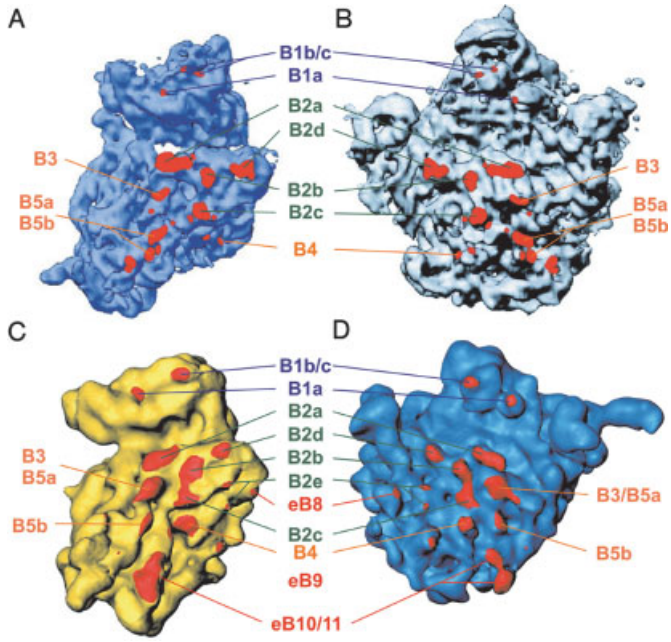


Figure 8-14 Comparison of intersubunit bridge positions between bacteria and yeast ribosomes. (A, B): The 30S (blue) and 50S (gray) ribosomal subunits of *T. thermophilus* are shown from their intersubunit sides. Intersubunit bridges are marked in red and are annotated according to the nomenclature from Gabashvili et al. [52], where the lettering of the bridges B1, B2 and B3–B5 are labeled in blue, green and orange, respectively. Figure adapted from Cate et al. [154]. (C, D): The 40S (yellow) and 60S (blue) ribosomal subunits of *S. cerevisiae* are also shown from the interface side with the intersubunit bridges in red. The lettering common to *T. thermophilus* are labeled in blue (B1), green (B2) and orange (B3–B5), whereas those of additional intersubunit connections in the yeast ribosome are labeled eB8–eB11 in red. Note that bridges B6 and B7 are not included for simplicity. Adapted from Spahn et al. [155].

finger" because it contacts the A-tRNA. As previously mentioned, both B1a and B2a bridge elements have corresponding counterparts within the 80S yeast ribosome, strengthening their candidacy for a role in translocation.

The most significant progress of our understanding of the translocation reaction comes from a detailed cryo-EM study of this reaction [31], together with a biochemical analysis [143]. EF-G induces a ratchet-like movement of the 30S subunits as first shown with empty ribosomes [144], but here the movement could be coupled with translocation of tRNAs at higher resolution, yielding further insight into this reaction: EF-G•GTP does not bind to ribosomes in a POST state or to ribosomes that

carry a peptidyl-tRNA at the P-site, but does so with high affinity if the ribosome carries only a deacylated tRNA at the P-site as it is the case in the PRE state after peptide-bond formation (see Fig. 8-2B). However, it difficult to imagine how EF-G binding is influenced by a deacylated tRNA at the P-site, when the A-site is occupied with a peptidyl-tRNA.

The former state is denoted "locked" for binding of EF-G•GTP, the latter "unlocked", although no structural differences could be detected at the resolution of 10–13 Å. The following picture emerged: (i) After occupying the A-site with an aa-tRNA and peptide-bond formation, the peptidyl-tRNA is in the A-site and a deacylated tRNA at the P-site, no hybrid site is visible. (ii) EF-G•GTP binds and induces the first forth-movement of the ratchet motion of the 30S subunit, a turn of about 20°. During this movement, the deacylated tRNA is seen in a hybrid position P/E and it is believed (although not yet observed) that the peptidyl-tRNA would also occupy a hybrid position A/P. (iii) The ribosome triggers the EF-G-dependent GTP hydrolysis and the authors postulate that the resulting EF-G conformational change into the GDP conformer completes the translocation reaction in that (i) the 30S subunit moves back (second part of the ratchet movement), (ii) the tRNAs continue their movement to E- and P-sites, respectively, and (iii) EF-G•GDP dissociates from the ribosome. Until now, EF-G•GTP has not been successfully crystallized, whereas EF-G•GDP has. A comparison of the structure of EF-G•GDPNP (a GTP analog) on the ribosome with that of the crystallized EF-G•GDP revealed a striking shift of domains 3–5 resulting in a movement of about 35 Å of the tip of domain 4. The authors postulated that this dramatic movement, similar in extent to that observed between the GTP and GDP conformers of EF-Tu (40 Å), promotes the back movement of the ratchet motion and might be also essential for the release of EF-G•GDP from the ribosome. If the striking conformational change is causing the release of EF-G rather than the second half of the ratchet movement, then the EF-G-dependent GTP hydrolysis has nothing to do with a motor protein function of this elongation factor. If, however, the GTP hydrolysis is in fact promoting the back-swing of the 30S subunit, then this could be interpreted as the molecular-correlate to a motor-protein role of EF-G. The authors [31] also detect a substantial movement of the L1 protuberance during the translocation and postulate an active participation of this structure in the translocation of the deacylated tRNA from the P-site to the E-site.

During the translocation reaction a movement of a peptidyl-tRNA from the A-site to the P-site via a P/E hybrid position can be easily reconciled with the α–ε model as has been suggested previously [51]. Note the difference to the hybrid-site model. This model postulates that after peptide-bond formation and before translocation the tRNAs move on the large subunit but stay on the 30S subunit, i.e., the peptidyl-tRNA at the A-site moves from A/A-site to the A/P-site, and only the translocation movement brings the tRNA into the P/P-site. In contrast, the cryo-EM study suggests that during translocation the peptidyl-tRNA moves from the A- to the P-site via a transient A/P position.

Despite the asymmetry of the ribosome, an internal symmetry within the large ribosomal subunit has been identified ([145]; reviewed in Refs. [146, 147]). The fact

that the CCA ends of tRNAs at the A- and P-sites are related by 180° rotation rather than by a translational movement like the rest of the tRNA [103, 105, 25], suggested that at least a minimum of symmetry might exist within the PTF center of the ribosome. However, Yonath and co-workers [145] realized that this symmetry extended beyond the A and P loops, and identified two regions of approximately 90 nucleotides that are related by rotational symmetry in the *D. radiodurans* 50S subunit (Fig. 8-8). The closest residues to the center of rotation include a number of residues of the PTF ring, strands of H89 and H93 and the stem loops of H92 and H80, which are the A and P loops, respectively (symmetry-related residues are colored with corresponding colors in Fig. 8-8). A similar symmetry was also discovered within the large subunit from archaea *H. marismortui* and thermophilic eubacteria *T. thermophilus* suggesting that it is a conserved feature of ribosomes. This is in itself not surprising since this region of the ribosome is highly conserved, but does add an extra dimension. But what is the significance of the symmetry?

Following peptide-bond formation, the peptidyl-ACC of the A-site tRNA must move into the P-site of the PTF center, i.e., a transition from making interactions with the A loop to making interactions with the P loop. Since the symmetry-related residues within the PTF center effectively provide a lining for the walls of the cavity, the idea arose that the CCA ends of the tRNA should follow a rotation movement during translocation. By computational modelling of such a transition, it was possible to accomplish a rotation from the A- to the P-site without any steric clashes. Interestingly, the symmetry-related residues within the P-site are located slightly deeper in the tunnel than the corresponding A-site residues, leading to the proposal that the rotational movement during translocation drives the peptide into the tunnel [147].

One of the most exciting discoveries related to the rotational symmetry was that the motion is centered on residue A2602. This residue is known to be highly flexible, a fact emphasized by the different orientation of the adenine base of this residue in almost all the different crystal structure complexes solved to date. Indeed, upon binding of many ligands, such as antibiotics and CCA-end mimics, A2602 assumed a different position [146]. This flexibility and the location of A2602 at the center of rotation may suggest that A2602 has a role in guiding the CCA end of the tRNA from the A- to the P-site. Another conserved residue predicted to maintain interaction throughout the rotational motion was U2585, which has also been identified as being very flexible.

The ribosome research is in exciting upheaval phase where the structure begins to explain the function. For the first time, we can imagine how the ribosome is using its complicated structure to perform its complicated function.

References

1 K. H. Nierhaus, H. Schulze, B. S. Cooperman, *Biochem. Int.* **1980**, *1*, 185–192.
2 K. H. Nierhaus, *Angew. Chem. Int. Ed.* **1996**, *35*, 2198–2201.
3 K. H. Nierhaus: in *Ribosomal RNA and Group I Int*, eds R. Green, R. Schroeder, R. G. Landes, Georgetown, TX 1996, 69–81.
4 K. H. Nierhaus: in *Nature Encyclopedia of Life Sciences*, Nature Publishing Group, London 1999, www.els.net.
5 D. N. Wilson, G. Blaha, S. R. Connell et al., *Curr. Protein Pept. Sci.* **2002**, *3*, 1–53.
6 D. N. Wilson, K. H. Nierhaus, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3464–3486.
7 R. Agrawal, P. Penczek, R. Grassucci et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6134–6138.
8 H. Stark, E. V. Orlova, J. Rinke-Appel et al., *Cell* **1997**, *88*, 19–28.
9 H. Stark, M. V. Rodnina, H. J. Wieden et al., *Cell* **2000**, *100*, 301–309.
10 M. Valle, J. Sengupta, N. K. Swami et al., *EMBO J.* **2002**, *21*, 3557–3567.
11 R. K. Agrawal, C. M. T. Spahn, P. Penczek et al., *J. Cell. Biol.* **2000**, *150*, 447–459.
12 S. Schilling-Bartetzko, A. Bartetzko, K. H. Nierhaus, *J. Biol. Chem.* **1992**, *267*, 4703–712.
13 K. Bergemann, K. H. Nierhaus, *J. Biol. Chem.* **1983**, *258*, 15105–15113.
14 L. P. Gavrilova, A. S. Spirin, *FEBS Lett.* **1971**, *17*, 324–326.
15 A. R. Cukras, D. R. Southworth, J. L. Brunelle et al., *Mol. Cell* **2003**, *12*, 321–328.
16 U. R. Kutay, C. M. T. Spahn, K. H. Nierhaus, *Biochim. Biophys. Acta* **1990**, *1050*, 193–196.
17 J. R. Mesters, A. P. Potapov, J. M. de Graaf et al., *J. Mol. Biol.* **1994**, *242*, 644–654.
18 A. Savelsbergh, V. I. Katunin, D. Mohr et al., *Mol. Cell* **2003**, *11*, 1517–1523.
19 J. M. Berg, J. L. Tymoczko, L. Stryer: *Biochemistry*, 5th edition, Freeman, New York **2002**.
20 F. J. Triana-Alonso, K. Chakraburtty, K. H. Nierhaus, *J. Biol. Chem.* **1995**, *270*, 20473–20478.
21 H. R. Bourne, D. A. Sanders, F. McCormick, *Nature* **1990**, *348*, 125–132.
22 D. Moazed, H. F. Noller, *Nature* **1989**, *342*, 142–148.
23 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001**, *292*, 883–896.
24 H. F. Noller, M. M. Yusupov, G. Z. Yusupova et al., *FEBS Lett.* **2002**, *514*, 11–16.
25 T. M. Schmeing, A. C. Seila, J. L. Hansen et al., *Nat. Struct. Biol.* **2002**, *9*, 225–230.
26 Y. Semenkov, T. Shapkina, V. Makhno et al., *FEBS Lett.* **1992**, *296*, 207–210.
27 D. Moazed, H. F. Noller, *Cell* **1989**, *57*, 585–597.
28 A. Gnirke, K. H. Nierhaus, *J. Biol. Chem.* **1986**, *261*, 14506–14514.
29 H. J. Rheinberger, K. H. Nierhaus, *J. Biomol. Struct. Dyn.* **1987**, *5*, 435–446.
30 R. K. Agrawal, P. Penczek, R. A. Grassucci et al., *J. Biol. Chem.* **1999**, *274*, 8723–8729.
31 M. Valle, A. Zavialov, J. Sengupta et al., *Cell* **2003**, *114*, 123–134.
32 C. M. Spahn, G. Blaha, R. K. Agrawal et al., *J. Mol. Cell.* **2001**, *7*, 1037–1045.
33 J. M. Robertson, H. Paulsen, W. Wintermeyer, *J. Mol. Biol.* **1986**, *192*, 351–360.
34 J. M. Robertson, C. Urbanke, G. Chinali et al., *J. Mol. Biol.* **1986**, *189*, 653–662.
35 J. Wadzack, N. Burkhardt, R. Jünemann et al., *J. Mol. Biol.* **1997**, *266*, 343–356.
36 J. Remme, T. Margus, R. Villems et al., *Eur. J. Biochem.* **1989**, *183*, 281–284.
37 M. V. Rodnina, A. Savelsbergh, V. I. Katunin et al., *Nature* **1997**, *385*, 37–41.

38 H.-J. Rheinberger, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 4213–4217.
39 H.-J. Rheinberger, K. H. Nierhaus, *J. Biol. Chem.* **1986**, *261*, 9133–9139.
40 A. Gnirke, U. Geigenmüller, H.-J. Rheinberger et al., *J. Biol. Chem.* **1989**, *264*, 7291–7301.
41 T. P. Hausner, U. Geigenmüller, K. H. Nierhaus, *J. Biol. Chem.* **1988**, *263*, 13103–13111.
42 H. Saruyama, K. H. Nierhaus, *Mol. Gen. Genet.* **1986**, *204*, 221–228.
43 K. H. Nierhaus, C. M. T. Spahn, N. Burkhardt: et al., in *The Ribosome. Structure, Function, Antibiotics, and Cellular Interactions*, eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, 319–335.
44 R. A. Grajevskaja, Y. V. Ivanov, E. M. Saminsky, *Eur. J. Biochem.* **1982**, *128*, 47–52.
45 H.-J. Rheinberger, H. Sternbach, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5310–5314.
46 H.-J. Rheinberger, H. Sternbach, K. H. Nierhaus, *J. Biol. Chem.* **1986**, *261*, 9140–9143.
47 U. Geigenmüller, K. H. Nierhaus, *EMBO J.* **1990**, *9*, 4527–4533.
48 M. Dabrowski, R. Junemann, M. A. Schafer et al., *Biochemistry (Moscow)* **1996**, *61*, 1402–1412.
49 M. Dabrowski, C. M. T. Spahn, M. A. Schäfer et al., *J. Biol. Chem.* **1998**, *273*, 32793–32800.
50 N. Polacek, S. Patzke, K. H. Nierhaus et al., *Mol. Cell* **2000**, *6*, 159–171.
51 M. A. Schäfer, A. O. Tastan, S. Patzke et al., *J. Biol. Chem.* **2002**, *277*, 19095–19105.
52 I. S. Gabashvili, R. K. Agrawal, C. M. T. Spahn et al., *Cell* **2000**, *100*, 537–549.
53 M. S. VanLoock, R. K. Agrawal, I. S. Gabashvili et al., *J. Mol. Biol.* **2000**, *304*, 507–515.
54 G. Diedrich, C. M. T. Spahn, U. Stelzl et al., *EMBO J.* **2000**, *19*, 5241–5250.
55 R. Willumeit, S. Forthmann, J. Beckmann et al., *J. Mol. Biol.* **2001**, *305*, 167–177.
56 H. Berchtold, L. Reshetnikova, C. O. A. Reiser et al., *Nature* **1993**, *365*, 368.
57 N. Bilgin, L. A. Kirsebom, M. Ehrenberg et al., *Biochimie* **1988**, *70*, 611–618.
58 G. Blaha, K. H. Nierhaus, *Cold Spring Harbor Symposia on Quantitative Biology*, **2001**, 135–145 *Vol. 65*.
59 H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477–511.
60 R. T. Libby, J. L. Nelson, J. M. Calvo et al., *EMBO J.* **1989**, *8*, 3153–3158.
61 K. Beebe, L. Ribas De Pouplana, P. Schimmel, *EMBO J.* **2003**, *22*, 668–675.
62 A. C. Bishop, K. Beebe, P. R. Schimmel, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 490–494.
63 T. L. Hendrickson, T. K. Nomanbhoy, V. de Crecy-Lagard et al., *Mol. Cell* **2002**, *9*, 353–362.
64 K. H. Nierhaus, *Mol. Microbiol.* **1993**, *9*, 661–669.
65 H. Stark, M. V. Rodnina, H. J. Wieden et al., *Nat. Struct. Biol.* **2002**, *15*, 15–20.
66 A. Weijland, A. Parmeggiani, *Science* **1993**, *259*, 1311–1314.
67 N. Bilgin, M. Ehrenberg, C. Kurland, *FEBS Lett.* **1988**, *233*, 95–99.
68 A. P. Potapov, *FEBS Lett.* **1982**, *146*, 28–33.
69 A. P. Potapov, F. J. Triana-Alonso, K. H. Nierhaus, *J. Biol. Chem.* **1995**, *270*, 17680–17684.
70 J. M. Ogle, D. E. Brodersen, W. M. Clemons Jr et al., *Science* **2001**, *292*, 897–902.
71 S. Osawa, T. H. Jukes, K. Watanabe et al., *Microbiol. Rev.* **1992**, *56*, 229–264.
72 J. M. Ogle, F. V. Murphy, M. J. Tarry et al., *Cell* **2002**, *111*, 721–732.
73 A. P. Carter, W. M. Clemons, D. E. Brodersen et al., *Nature* **2000**, *407*, 340–348.
74 W. M. Clemons, J. L. C. May, B. T. Wimberly et al., *Nature* **1999**, *400*, 833–840.
75 K. H. Nierhaus, *Biochemistry* **1990**, *29*, 4997–5008.

76 J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4135–4139.
77 J. Ninio, *Biochimie* **1975**, *57*, 587–595.
78 M. Ehrenberg, D. Andersson, K. Mohman et al.: in *Structure, Function and Genetics of Ribosomes*, eds B. Hardesty, G. Kramer, Springer, New York 1986, 573–585.
79 M. V. Rodnina, T. Daviter, K. Gromadski et al., *Biochimie* **2002**, *84*, 745–754.
80 S. Schilling-Bartetzko, F. Franceschi, H. Sternbach et al., *J. Biol. Chem.* **1992**, *267*, 4693–4702.
81 A. M. Karim, R. C. Thompson, *J. Biol. Chem.* **1986**, *261*, 3238–3243.
82 F. J. LaRiviere, A. D. Wolfson, O. C. Uhlenbeck, *Science* **2001**, *294*, 165–168.
83 A. Aevarsson, E. Brazhnikov, M. Garber et al., *EMBO J.* **1994**, *13*, 3669–3677.
84 J. Czworkowski, J. Wang, T. A. Seitz et al., *EMBO J.* **1994**, *13*, 3661–3668.
85 H. Berchtold, L. Reshetnikova, C. O. A. Reiser et al., *Nature* **1993**, *365*, 126–132.
86 P. Nissen, M. Kjeldgaard, S. Thirup et al., *Science* **1995**, *270*, 1464–1472.
87 P. Nissen, M. Kjeldgaard, J. Nyborg, *EMBO J.* **2000**, *19*, 489–495.
88 K. Yamaguchi, K. von Knoblauch, A. R. Subramanian, *J. Biol. Chem.* **2000**, *275*, 28455–28465.
89 E. C. Koc, W. Burkhart, K. Blackburn et al., *J. Biol. Chem.* **2001**, *276*, 19363–19374.
90 E. C. Koc, W. Burkhart, K. Blackburn et al., *J. Biol. Chem.* **2001**, *276*, 43958–43969.
91 T. Suzuki, M. Terasaki, C. Takemoto-Hori et al., *J. Biol. Chem.* **2001**, *276*, 33181–33195.
92 K. Yamaguchi, A. R. Subramanian, *J. Biol. Chem.* **2000**, *275*, 28466–28482.
93 M. R. Sharma, E. C. Koc, P. P. Datta et al., *Cell* **2003**, *115*, 97–108.
94 F. Bouadloun, D. Donner, C. G. Kurland, *EMBO J.* **1983**, *2*, 1351–1356.
95 J. Langridge, J. H. Campbell, *Mol. Gen. Genet.* **1969**, *103*, 339–347.
96 F. Cramer, U. Englisch, W. Freist et al., *Biochimie* **1991**, *73*, 1027–1035.
97 J. R. Menninger, *J. Biol. Chem.* **1976**, *251*, 3392–3398.
98 F. Jorgensen, C. G. Kurland, *J. Mol. Biol.* **1990**, *215*, 511–521.
99 K. C. Keiler, P. R. H. Waller, R. T. Sauer, *Science* **1996**, *271*, 990–993.
100 A. Fujihara, H. Tomatsu, S. Inagaki et al., *Genes Cells* **2002**, *7*, 343–350.
101 F. Jorgensen, F. M. Adamski, W. P. Tate et al., *J. Mol. Biol.* **1993**, *230*, 41–50.
102 K. H. Nierhaus, J. Wadzack, N. Burkhardt et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 945–950.
103 J. L. Hansen, T. M. Schmeing, P. B. Moore et al., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11670–11675.
104 M. Valle, A. Zavialov, W. Li et al., *Nat. Struct. Biol.* **2003**, *10*, 899–906.
105 P. Nissen, J. Hansen, N. Ban et al., *Science* **2000**, *289*, 920–930.
106 M. Welch, J. Chastang, M. Yarus, *Biochemistry* **1995**, *34*, 385–390.
107 G. Steiner, E. Kuechler, A. Barta, *EMBO J.* **1988**, *7*, 3949–3955.
108 C. M. T. Spahn, C. D. Presscott, *J. Mol. Med.* **1996**, *74*, 423–439.
109 F. Franceschi, K. H. Nierhaus, *J. Biol. Chem.* **1990**, *265*, 16676–16682.
110 H. Schulze, K. H. Nierhaus, *EMBO J.* **1982**, *1*, 609–613.
111 J. Harms, F. Schluenzen, R. Zarivach et al., *Cell* **2001**, *107*, 679–688.
112 S. Kirillov, J. Wower, S. Hixson et al., *FEBS Lett.* **2002**, *514*, 60–66.
113 B. Maguire, A. Beniaminov, P. Ramu et al.: in *8th Annual Meeting of the RNA Society*, RNA Society, Vienna 2003, 199.
114 S. R. Fahnestock, H. Neumann, V. Shashua et al., *Biochemistry* **1970**, *9*, 2477–2483.
115 D. J. Klein, T. M. Schmeing, P. B. Moore et al., *EMBO J.* **2001**, *20*, 4214–4221.
116 B. E. Maden, R. E. Monro, *Eur. J. Biochem.* **1968**, *6*, 309–316.

117 K. K. Wan, N. D. Zahid, R. M. Baxter, *Eur. J. Biochem.* **1975,** *58,* 397–402.
118 I. Rychlik, J. Cerna, *Biochem. Int.* **1980,** *1,* 193–200.
119 H. Bremer, P. P. Dennis: in *Escherichia coli and Salmonella,* eds F. C. Neidhardt, R. C. III, J. L. Ingraham, et al., ASM Press, Washington, DC 1996, 1553–1569.
120 T. Bruice, S. Benkovic, *J. Am. Chem. Soc.* **1963,** 85.
121 G. W. Muth, L. Ortoleva-Donnelly, S. A. Strobel, *Science* **2000,** *289,* 947–950.
122 K. M. Parnell, A. Seila, S. A. Strobel, *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 11658–11663.
123 M. A. Bayfield, A. E. Dahlberg, U. Schulmeister et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 10096–10101.
124 J. Thompson, D. F. Kim, M. O'Connor et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 9002–9007.
125 N. Polacek, M. Gaynor, A. Yassin et al., *Nature* **2001,** *411,* 498–501.
126 K. Quiggle, G. Kumar, T. W. Ott et al., *Biochemistry* **1981,** *20,* 3480–3485.
127 V. I. Katunin, G. W. Muth, S. A. Strobel et al., *Mol. Cell* **2002,** *10,* 339–346.
128 Y. Kaziro, *Biochim. Biophys. Acta* **1978,** *505,* 95–127.
129 R. K. Agrawal, A. B. Heagle, P. Penczek et al., *Nat. Struct. Biol.* **1999,** *6,* 643–647.
130 M. G. Gomez-Lorenzo, C. M. T. Spahn, R. K. Agrawal et al., *EMBO J.* **2000,** *19,* 2710–2718.
131 L. P. Gavrilova, A. S. Spirin: in *Methods in Enzymology,* eds K. Moldave and L. Grossmann, vol. 30 1974, 452–462.
132 K. Fredrick, H. F. Noller, *Science* **2003,** *300,* 1159–1162.
133 T. P. Hausner, J. Atmadja, K. H. Nierhaus, *Biochimie* **1987,** *69,* 911–923.
134 Y. Endo, K. Tsurugi, *J. Biol. Chem.* **1987,** *262,* 8128–8130.
135 T. Uchiumi, S. Honma, T. Nomura et al., *J. Biol. Chem.* **2002,** *277,* 3857–3862.
136 E. Cundliffe: in *The Ribosome: Structure, Function and Evolution,* eds W. E. Hill, A. Dahlberg, R. A. Garrett et al., ASM Press, Washington, DC 1990, 479–490.
137 G. Stöffler, E. Cundliffe, M. Stöffle-Meilicke et al., *J. Biol. Chem.* **1980,** *255,* 10517–10522.
138 D. M. Cameron, J. Thompson, P. E. March et al., *J. Mol. Biol.* **2002,** *319,* 27–35.
139 R. K. Agrawal, J. Linde, J. Sengupta et al., *J. Mol. Biol.* **2001,** *311,* 777–787.
140 H. Stark, M. V. Rodnina, J. Rinkeappel et al., *Nature* **1997, 389,** 403–406.
141 D. Beyer, E. Skripkin, J. Wadzack et al., *J. Biol. Chem.* **1994,** *269,* 30713–30717.
142 E. V. Alexeeva, O. V. Shpanchenko, O. A. Dontsova et al., *Nucleic Acids Res.* **1996,** *24,* 2228–2235.
143 A. V. Zavialov, M. Ehrenberg, *Cell* **2003,** *114,* 113–122.
144 J. Frank, R. K. Agrawal, *Nature* **2000,** *406,* 318–322.
145 A. Bashan, I. Agmon, R. Zarivach et al., *Mol. Cell* **2003,** *11,* 91–102.
146 I. Agmon, T. Auerbach, D. Baram et al., *Eur. J. Biochem.* **2003,** *270,* 2543–2556.
147 A. Bashan, R. Zarivach, F. Schluenzen et al., *Biopolymers* **2003,** *79,* 19–41.
148 M. Laurberg, O. Kristensen, K. Martemyanov et al., *J. Mol. Biol.* **2000,** *303,* 593–603.
149 T. Toyoda, O. F. Tin, K. Ito et al., *RNA* **2000,** *6,* 1432–1444.
150 N. Guex, M. C. Peitsch, *Electrophoresis* **1997,** *18,* 2714–2723.
151 U. B. Rawat, A. V. Zavialov, J. Sengupta et al., *Nature* **2003,** *421,* 87–90.
152 R. Green, J. R. Lorsch, *Cell* **2002,** *110,* 665–668.
153 J. M. Berg, J. R. Lorsch, *Science* **2001,** *291,* 203.
154 J. H. Cate, M. M. Yusupov, G. Z. Yusupova et al., *Science* **1999,** *285,* 2095–2104.
155 C. M. Spahn, R. Beckmann, N. Eswar et al., *Cell* **2001,** *107,* 373–386.

9
Termination and Ribosome Recycling

Daniel N. Wilson

9.1
Introduction

The translation-elongation cycle is brought to an abrupt halt by the appearance of a stop codon in the ribosomal A site, rather than the usual sense codon of the mRNA. If the mRNA being translated is thought of as a sentence then the stop codons generally act as a "full-stop", i.e. they signal to the ribosome that the protein has been fully translated and must now be released from the ribosome. The process of polypeptide release is supposed to be the third (and generally the final) stage of translation, with initiation and elongation (see Chaps. 7 and 8) being the first and second, respectively. However, it should be acknowledged that the process of recycling the ribosomes following release of the polypeptide is often referred to as a post-termination event and, therefore, can be thought as the fourth stage of translation. There are numerous reviews attempting to keep up with the rapid pace that the termination researchers have attained in the past few years and demonstrating the unexpected discoveries that are hiding around every corner (for a recent review see [1]).

So how does the presence of a stop codon in the ribosomal A site trigger entry into the termination and ribosome recycling stages of the translation cycle? The answer to this question relates directly to the make-up of the genetic code; usually all codons, but three (although this can sometimes be one or two for particular organisms or organelles, see later), correspond to a particular amino acid and therefore with an aminoacylated or charged tRNA. Clearly, in the situation where a ribosome bears a stop codon (or any codon for that matter) in the A site for which there is no corresponding aminoacyl-tRNA, the elongation cycle cannot continue. This type of reasoning led to the proposal for the existence of a "terminator-tRNA", i.e., a special (uncharged) tRNA that recognized the stop codons and somehow mediated release of the completed polypeptide. In light of what we now know about the interaction between the anticodon of the tRNA and the codon of the mRNA and how the stereochemistry of this interaction is monitored by the ribosome to ensure its correctness, the idea of a terminator-tRNA would seem like an obvious choice to initiate termination events. However, as so often seen before, nature had evolved another quite distinct mechanism, which does not rely on RNA–RNA interaction at all.

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

No terminator tRNAs were discovered, instead it turned out that stop codons or termination signals are recognized by protein factors, termed termination release factors (RFs), so named because they mediate the *release* of the nascent polypeptide from the ribosome. However, it soon became clear that there were two sets of RFs involved in the termination process. The original factors, which decode the stop codons and actually release the polypeptide from the ribosome, were therefore termed class I or decoding release factors. In bacteria, such as *Escherichia coli*, there are two decoding factors, RF1 and RF2. The class II RFs operate after the decoding factors (and perhaps more important for factor nomenclature were also identified subsequent to the decoding factors) and are therefore termed RF3s. RF3 is involved in removal of the decoding RFs from the ribosome and therefore stimulates release of the polypeptide by recycling of the decoding factors. In eukaryotes (and archea), there is only ever one class I decoding RF, termed eRF1 (aRF1) and by analogy the class II factor is termed eRF3 (aRF3). The class II factors are G-proteins and therefore their affinity for the ribosome is regulated by the guanine nucleotide state of the factor (see Chap. 8). For bacterial RF3, the GDP form of RF3 has been shown to have a low affinity for the ribosome following release of the decoding factor and therefore falls off the ribosome. This post-termination ribosome, however, still bears the mRNA it was translating, a deacylated or uncharged tRNA at the P site (the tRNA from which the nascent polypeptide was released) and probably an additional deacylated tRNA at the E site. These components need to be removed and the ribosome dissociated into its component subunits in preparation for the next round of translation. This essential process is termed ribosome recycling. In bacteria, three factors are involved in this process, one that is specific for this stage termed the ribosome recycling factor (RRF) and the others, which are active during elongation and initiation, namely elongation factor G (EF-G) and initiation factor 3 (IF3), respectively. There is some debate at present as to the exact details of the steps mediated by these three factors and this is discussed in more detail in Sect. 9.5. Since RRF is not found in eukaryotes (or archaea), except in mitochondria or chloroplasts, there must be an alternative system operating to fulfill or circumvent the need for this process in the cytoplasm.

9.2 Stop Codon Recognition and Release of the Nascent Polypeptide Chain

The importance of the termination phase of protein synthesis is emphasized by the universal presence of class I decoding release factors, i.e., all archea and eukaryotic genomes sequenced so far have the presence of aRF1 and eRF1, respectively, and all bacterial genomes have at least one of the two decoding factors RF1 or RF2. For example, in the genome of *Mycoplasma genitalium*, the *prfA* gene encoding RF1 is present whereas *prfB*, the gene encoding RF2, is absent. Since *M. genitalium* is often referred to as the bacteria containing the "minimal complement" of genes necessary for survival, its small genome size arising from having dispensed with most of the

non-essential genes [2, 3], this illustrates the importance of the decoding termination factors. Indeed, although eukaryotic and archeal RF1 perform the same role as the bacterial RF1 (and RF2), there is no obvious sequence or structural homology between the two proteins, suggesting that they have arisen independently and thus represent examples of functionally convergent evolution. Despite this evolutionary independence, some remarkable similarities between the two factors have emerged, suggesting that their mechanism of action may also be similar. When one considers the extreme conservation in the target or substrate of their reaction, namely, the ribosome with a peptidyl-tRNA in the P site and a stop codon in the A site, it is not unforeseeable that the similarities in the constraints imposed by the ribosome in terms of binding site are reflected by similarities within the bacterial and eukaryotic factors. With regard to the mechanism of action of the bacterial decoding factors on the ribosome, a number of surprises have recently been brought to light and it remains to be seen whether the eukaryotic factors really operate through a similar mechanism.

9.3 The Bacterial Class I Decoding Release Factors

The eubacterial decoding factors, RF1 and RF2, exhibit overlapping specificities with regard to stop-codon discrimination: for example, in *E. coli*, RF1 decodes the stop codons UAG and UAA, whereas RF2 decodes UGA and UAA [4, 5]. Organisms such as *M. genitalium*, which have dispensed with the *prfB* gene, do not require the corresponding RF2 factor because UGA is not regarded as a stop codon, being decoded by Trp-tRNA^{ACU} and thus incorporating tryptophan at these codons [6]. In organelles, where codon reassignment is common (reviewed in Refs. [7, 8]), there is almost always a loss of the *prfB* gene, for example, yeast mitochondria also decode UGA by tryptophan and therefore have only a single RF of the RF1-type [9]. This raises the intriguing question of why it is always the *prfB* gene and not the *prfA* gene that is continuously lost from these genomes?

9.3.1 The Structure of RF2 and Translational Mimicry

A comparison of the *E. coli prfA* and *prfB* genes [10, 11] and in particular the protein products revealed extensive sequence homology. Based on secondary-structure predictions, the decoding factors have been discussed in terms of five [12] or seven [13] domain models. However, on the basis of functional data, Tate and co-workers [14, 15] proposed a simple two-domain "tRNA-analogue model", where the decoding RFs were thought to span from the stop codon in the decoding site of the 30S subunit to the PTF center on the 50S subunit. In this model, each domain of the RF was associated with a function, one domain with stop codon recognition and the other with release of the polypeptide, in analogy with a tRNA where the anticodon stem loop is associated with codon recognition and the acceptor stem of the tRNA with peptide-bond formation. This tRNA mimicry concept was brought to another level with the

arrival of the structure of the ternary complex of EF-Tu•GTP•Phe-tRNA^{Phe} [16], which had striking overall similarity with the translation elongation factor EF-G [17, 18]. In particular, the tertiary structure of domains III-V of EF-G resembled that of the tRNA in the ternary complex (see Figs. 8.6A and B). This suggested that the A site of the ribosome, the binding site for the tRNA, but also for the translation factors such as EF-G and the EF-Tu ternary complex, imposed constraints on the factors to take-on the shape of tRNA to bind within A-site region. Since the initiation factors IF1 and IF2, as well as the termination factors RF1, RF2, RF3 and RRF, were also known to bind within this region, the idea of macromolecular mimicry of tRNAs by protein factors was extended to encompass all phases of the translation cycle (reviewed by Nissen et al. [19]). Indeed Sprinzl and co-workers [20] proposed, based on sequence alignments, that IF2 would have a structure similar to EF-G except that the missing region of complementarity in IF2, namely of domain IV of EF-G, would be compensated for by homology found in IF1, i.e. together IF2 and IF1 would also mimic the shape of the aa-tRNA•EF-Tu•GTP ternary complex. However, the recent crystal structure of the IF1 bound to the 30S subunit [21] suggests that the concept of molecular mimicry may have been overextrapolated in this case, since the structure and binding position of IF1 bears little resemblance to that of the anticodon stem loop of a tRNA, nor to that of domain IV of EF-G (see Chap. 7.1).

The long-awaited structure of a decoding RF was that of the eubacterial *E. coli* RF2 [22], rather than from *Thermus thermophilus* as might have been first thought [23]. Since the vast majority of studies into RFs had come from studies on *E. coli*, this provided the perfect opportunity to correlate structure and function (see the next section). The *E. coli* RF2 structure was solved to 2.3 Å and revealed a four-domain arrangement (Fig. 9.1A). The N-terminal domain (domain I) contains four α-helices that form an α-helical bundle, whereas domains II, III and IV form a very compact

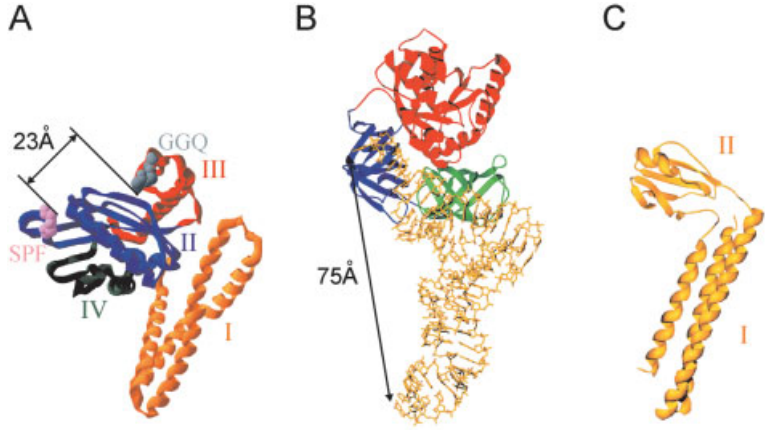


Figure 9.1 Molecular mimicry between tRNA and termination factors. The crystal structures of (A) RF2 [22], (B) tRNA (yellow) in form of ternary complex, (C) RRF [28] are compared illustrating the overall tertiary structure similarities.

structure stabilized by multiple interdomain interactions [22]. Indeed, the structure could be interpreted as having an overall two-domain arrangement, since domain I and domains II–IV are relatively separate from one another. This is supported by the observation that the recent structure of *T. thermophilus* RF2 was almost identical to that of *E. coli* RF2, except that there was a different orientation of domain I with respect to domains II–IV, which similar to the *E. coli* structure, also formed a distinct compact superdomain [24]. Despite the overall two-domain topology, there is little resemblance between the RF2 structures and a tRNA (Fig. 9.1B; reviewed in Refs. [25–27]); however, the previous structures of the ribosome recycling factor (RRF) had reinforced the idea of tRNA mimicry, since this small factor was composed of two domains that were arranged in a definite L-shape with similar dimensions and orientations as a tRNA (Fig. 9.1C; [28]). On the basis of the similarity between domain I of RF2 and domain I of RRF, which was proposed to mimic the anticodon stem loop of the tRNA, domain I of RF2 was docked into the A site. However, this docking, nor in fact any other single orientation, could simultaneously satisfy all the available biochemical data, which had defined particular regions of the decoding RFs as being associated with particular functions, such as codon recognition or peptidyl-tRNA hydrolysis.

9.3.2
The Two-domain Functional Model for RF2

Before the arrival of the RF2 crystal structures, there existed very strong evidence that the decoding factors were intimately involved in stop-codon recognition and peptidyl-tRNA hydrolysis. In fact, distinct regions and even specific amino acids within these regions had been implicated in each function. If the decoding RFs were to mimic tRNAs, then they would be required to interact at one end of the molecule with the stop codon in the A site and at the other end to mediate the transfer of the polypeptide chain from the P-site tRNA to water, thereby releasing it from the ribosome. So what evidence exists for these functional domains in the RFs and how do they relate to the available crystal structure?

9.3.3
Identifying Functional Important Regions within the Decoding RFs

Perhaps the first indication as to which regions of the decoding RFs were functionally significant arose through the use of suppression studies, where competition at stop codons between the decoding RFs and so-called suppressor tRNAs was monitored. The first of these studies analyzed the effect of decoding RFs at UAA stop-codon contexts by utilizing a three-plasmid system, one expressing either RF1 or RF2, one expressing a UAA stop-codon suppressor tRNA and a reporter plasmid expressing a UAA-containing *LacZ* reporter construct [29, 30]. The application of this type of system was used to identify mutant RFs with reduced activity, i.e., the situation that resulted in more efficient competition of the suppressor tRNAs at the stop

signal (see Table 9.1 and Fig. 9.2). Although these genetic screens provided a powerful technique for identifying functional regions in RFs, most of the mutations conferred a recessive phenotype and thus without complementary *in vitro* data provide limited insight into the true nature of the mutation. Indeed, a number of the RF mutants were temperature-sensitive [31–36], suggesting that the defects in these factors may have been due to folding perturbations.

RF fusion proteins were the first demonstration of partly functional release factors, where partial activity in ribosome binding was detected in the absence of peptidyl-tRNA hydrolysis (or release) activity [37]. The separation of the decoding and release activities of the decoding RFs was in agreement with the two-domain model of the RFs. A more directed approach to discovering regions of importance within the decoding RFs was undertaken by generating chimeric RF constructs by exchanging regions between *E. coli* RF1 and RF2 [14]. One of the chimeric constructs generated was identical to RF1 except for the replacement of a small region within domain III by the corresponding region in RF2. Although this in effect yielded only 10 amino acid substitutions, many of which were either conservative or in positions of low conservation, the effect was significant: *in vivo*, the expression of this chimeric factor was toxic to the *E. coli* cell, reproducibly inducing complete cessation of growth upon induction and the strong selection against the plasmid expressing the chimera and often leading to deletions within the plasmid [14]. *In vitro*, this chimera was shown to have significant reduced release activity, whereas the codon-dependent binding activity remained unaffected [14]. This was the first

Table 9.1 Decoding Release Factor Mutations

Factor	Mutation[a]	Reference
RF1 (*E. coli*)	G168D	35
	R137P	36
RF1 (*S. typhimurium*)	G180S	31
	H182Y	
mtRF1 (*S. cerevisiae*)	R190K	9, 12
	P192L	38
RF2 (*E. coli*)	L63F and D79G	151
	E89K	34
	D143N	
	L328F	
	F207T	13
	R213I	
	T246A	63
	E167K	136
RF2 (*S. typh.*)	Y144UGA	33

a All numbering corresponds to the equivalent residue in *E. coli*.

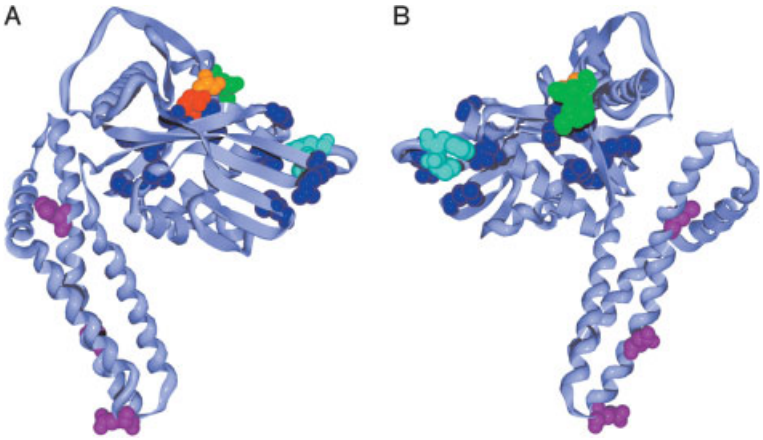


Figure 9.2 The crystal structure of *E. coli* RF2 with mutants from Table 9.1 mapped onto the equivalent positions in *E. coli* RF2. The tripeptide (cyan) and GGQ motif (light green) as well as position 246 (yellow) and the omnipotent E167K (red) are also indicated. Note that most of the mutations map within domains II-IV (dark blue), whereas only few (L63, D79 and E89 colored purple) are in domain I.

evidence implicating domain III as being associated with release of the polypeptide and is discussed in more detail in Sect. 9.2.6.

Around the same time, two mutations in the mitochondrial RF1 gene were shown to be responsible for a splicing defect in the yeast *S. cerevisiae* [9, 12]. The mutations were identified as being in domain II of RF1 and were subsequently shown to impair the factor's ribosome binding ability *in vivo* [38] and *in vitro* [39]. That these sites in domain II were directly associated with ribosome binding was strengthened, when second site suppressors were identified in h44 of the 16S rRNA that restored the binding of the mutant factors. Three of these second site suppressors disrupt a base-pair within h44 and confer resistance to the aminoglycoside antibiotic paromomycin, which is known to bind directly within the decoding site and induce translational misreading (see Chap. 12.3.1.2). Another mutation mapped within the so-called switch region (h27) that has been implicated in translational fidelity [40] and constitutes the binding site for the classic misreading antibiotic streptomycin (see Chap. 12.3.1.2). This is very suggestive of a direct interaction between domain II of the RF and regions within the decoding site, if not the stop codon itself. Indeed, one of the second site suppressor locations (A1408) has been crosslinked from the first position of a UAA stop codon located in the A site [41] and normally base-pairs with A1493, one of the universally conserved residues that is intimately involved in monitoring the correctness of codon–anticodon interaction in the A-Site (see Chap. 8.2.3).

9.3.4 Codon Recognition Domain of Bacterial RFs: the Termination Signal

It seems very probable that the decoding release factors recognize a stop codon in the A site of the ribosome by directly interacting with it, as opposed to recognizing the presence of a stop codon indirectly through putative conformational changes that occur specifically in the ribosome during the termination phase. The latter seems less probable – it would require conformational changes that are specific for the stop codon, since RF1 and RF2 discriminate between UAG and UGA, although such changes have been proposed [42]. In addition, close proximity between stop codon and RF is evident from the zero-length crosslinks observed from the first position of a stop codon (using thiouridine) to RF2 [41, 43]. Moreover, it has been shown that the context of the stop codon played an important role in the efficiency of the termination reaction (reviewed in Refs. [44–46]). This was perhaps first evident from the markedly differing levels of stop codon readthrough by suppressor tRNAs depending on the context of the stop codon, both upstream and downstream. Tate and co-workers went on to demonstrate that the position immediately following (3′ to) the stop codon (termed the +4 position) had the most significant effect on termination efficiency [47], but additional influences were also seen as far downstream as the +6 position [48]. Since RF decoding of stop codons involves protein–RNA interactions, rather than RNA–RNA (tRNA–mRNA) interaction, and requires no further reading frame maintenance, there is no reason to restrict the interaction to triplet stop codon. Instead RFs can be thought to decode termination signals consisting of the stop codon and its surounding context. Evidence for such a proposal comes from striking correlation between the relative efficiency of the termination signals seen with *in vivo* reporter assays, used by Tate and co-workers for example, and from analysis of the termination signals present in the genome (Fig. 9.3). This is especially true in the case of UGAN and UAAN signals utilized by *highly expressed genes*, where in the +4 position (N), pyrimidine is preferred to purine, with uridine usually the strongest. The presence of tandem stop codons, UGAUGA for example, was thought to be advantageous, since it appeared to be over-represented in the genomes of bacteria, however, from analysis using the reporter systems, it seems that tandem stop signals confer no additional advantage other than having the +4 position uridine [49]. Further support for a direct interaction between RFs and the extended termination signal came when crosslinks from the +4 position [50] and weaker crosslinks from the +5 and +6 positions (but not +7 to +10 positions) [48] with the RF were demonstrated. Attempts to identify the region within the RF to which the crosslink was attached were largely unsuccessful, although there was some hint that domain II was involved [51].

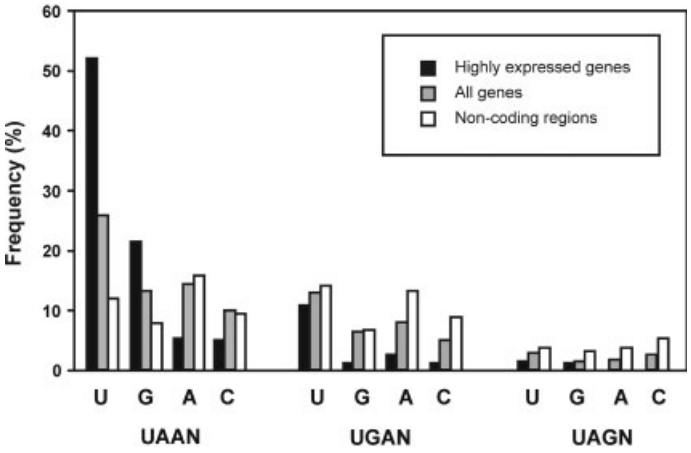


Figure 9.3 The frequency (in percentage) of occurrence for base (N is either U, G, A or C) following each of the three stop codons UAA, UGA and UAG in the higly expressed genes (black bars), all genes (grey bars) and within non-coding regions (white bars) of the *Escherichia coli* genome.

9.3.5 Codon Recognition Domain of Bacterial RFs: the "Tripeptide Motif"

To identify the domain within decoding RFs involved in codon recognition, genetic screening of mutant RF2s that could complement both conditional-lethal RF1 and RF2 mutants was undertaken. One of the factors identified had a mutation E167K and could terminate translation at all three stops, much similar to an eRF1, although the specificity of the interaction was questionable, since termination at some sense codons was also observed [52]. Following this, Nakamura and co-workers performed an elegant series of experiments identifying tripeptide motifs, PAT and SPF, within domain II of the decoding RFs, RF1 and RF2 respectively, which discriminate specifically the three different stop codons ([53], reviewed in [54, 55]). The first and third positions of the tripeptide motif discriminate the second and third positions of the stop codon (Fig. 9.4, [55]), such that the motif SXT is omnipotent (including UGG) and PXF is UAA restricted [53]. From a biochemical point of view, the location of the tripeptide motif within domain II was consistent with many other mutations located within this region that were identified using various readthrough assays. In this respect, it is noteworthy that one of the locations of one of the mitochondrial RF1 mutants corresponded with the first position of the tripeptide motif. Furthermore, the so-called "charge-flip" changes at multiple Glu (E) residues located in close proximity to the tripeptide motif also interfered with codon recognition [56].

However, from a structural point of view, it was surprising that the tripeptide motif was located within domain II and not domain I. Because domain I of RF2

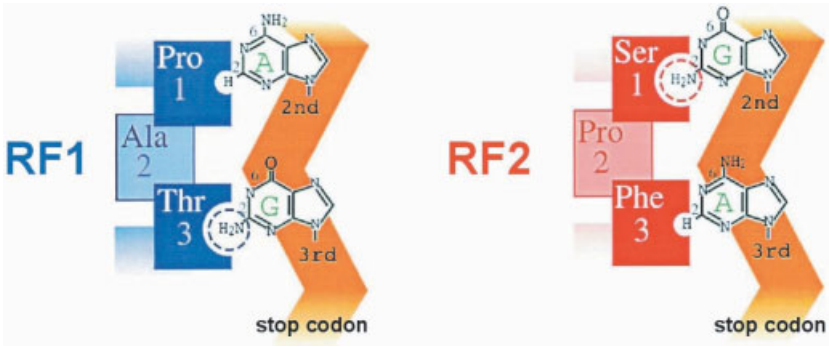


Figure 9.4 A model showing how the tripeptide motif PAT in RF1 and SPF motif in RF2 decode the second and third positions of the stop codon. Reprinted with permission from Ref. [55].

resembled domain I of RRF, which in turn based on a proposed tRNA mimicry was equivalent to the anticodon stem of a tRNA (see Fig. 9.1), the thinking at the time was that the region associated with stop-codon recognition should be located within the loop located between α3 and α4. Although there was little support biochemically for this presumption, a model was presented where the RF2 structure was docked into the programed 70S crystal structure with the tip of domain I in the decoding site. This model necessitated that the tripeptide motif function indirectly to decode the stop codon; however it made contact with h44 and therefore was supposed to monitor stop-codon recognition indirectly through conformational changes induced within h44 [22]. In support of this model was that the so-called GGQ motif (see following sections), located in domain III was located in close proximity to the PTC in accord with biochemical data and there were no spatial clashes. In contrast, when the tripeptide motif was placed in the decoding site, there were numerous clashes between other regions of the RF2 and components of the ribosome. Furthermore, in this orientation, the GGQ motif was located far from the PTC.

9.3.6 Peptidyl-tRNA hydrolase function of bacterial RFs: domain III and the GGQ motif

The ultimate role of the class I RFs is to mediate the release of the polypeptide from the peptidyl-tRNA bound at the P site of the ribosome. During elongation, the peptidyl moeity of the P-site tRNA is transferred to the α-amino group of the aminoacyl-tRNA bound at the A site (Fig. 9.5A; see Chap. 8.3 for more details). By analogy with this situation, the termination reaction was proposed to involve the transfer of the peptidyl moeity to a water molecule, as illustrated in Fig. 9.5(B) [57–59]. The question is to what extent do the class I RFs play a role in this reaction? Are the decoding factors simply messengers passing on a signal from the decoding site to

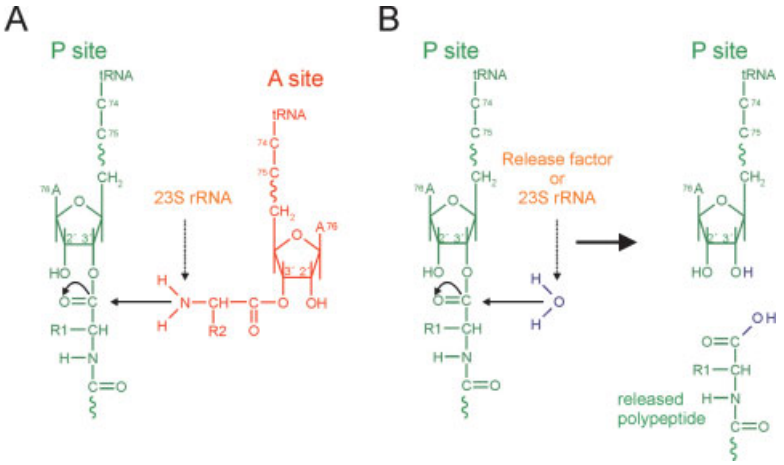


Figure 9.5 Scheme for (A) peptidyl-transferase and (B) termination reactions on the ribosome. In (A) there is a nucleophillic attack by the α-amino group of the A site tRNA on the carbonyl group of P site peptidyl-tRNA (indicated by solid arrow). There is controversy over the extent that components of the ribosome, such as bases of the 23S rRNA, contribute to this reaction (see Chapter 8). In (B) the water molecule replaces the A site ligand to make a nucleophillic attack on P site ligand. The question has been raised as to whether this water molecule is co-ordinated by the termination release factor or some component of the 23S rRNA (indicated by dashed arrow).

the PTF center that mediates peptide release or does a region of the decoding factor play an active role in the release of the peptide by, for example, co-ordinating the water molecule needed for the reaction?

One of the first regions implicated in peptide release came from proteolytic studies with the decoding factors. A protease-sensitive site was identified in a similar position in RF1 and RF2 [15], which maps to what we now know is the GGQ-containing loop within domain III (Fig. 9.2). Limited proteolytic cleavage of RF2 with chymotrypsin produced two relatively stable fragments of approximately 24 and 15 kDa. N-terminal sequencing demonstrated that the cleavage occurred between tyrosine and arginine residues at positions 244 and 245, respectively. Interestingly, the chymotryptic fragments of RF2 remained associated, when isolated using anion-exchange and gel-permeation chromatography under non-denaturing conditions and while the nicked RF2 factor retained the ability to bind to the ribosome and discriminate stop codons, it was completely inactive for peptide release [15]. This suggested that this region within domain III was essential for release activity of the factors, but not binding or codon recognition.

At around the same time, a chimera of RF1 was constructed where a small section of domain III (including the proteolytic site) was replaced with the homologous region from RF2 [14]. As mentioned, this chimera was inactive for release activity

and unlike the exogenous expression of RF1 in bacteria, overexpression of the chimera was very toxic to the cell. Indeed, the observed phenotype was remarkably reminiscent to that observed when *E. coli* RF2 is exogenously overexpressed. In addition, the *in vitro* characteristics were also similar, namely that the purified protein was active for codon recognition and ribosome binding but severely impaired for release activity [60]. What was intriguing was that there was a correlation between the level of overexpression and the loss of release activity, such that the higher the expression, the lower the specific activity of the RF2 protein [61, 62]. However, this phenotype seemed to be specific for *E. coli* since overexpression of *S. typhimurium* was not toxic to the cell and this recombinant RF2 protein was fully active *in vitro* (see [63] and references therein). Since there were only 16 differences between *S. typhimurium* and *E. coli* RF2, fragment swap and site-directed mutagenesis experiments were employed to identify which residues were responsible for the characteristic phenotypes. One of the 10 RF2 residues in the RF1 chimera, Thr246, was identified, which is normally serine in *E. coli* RF1 and alanine in *S. typhimurium* RF2 at the equivalent position. Replacing Thr246 in *E. coli* RF2 or in the RF1 chimera with Ser or Ala restored peptide-release activity to the purified factor and alleviated the growth toxicity associated with factor expression [63, 64]. In fact, the class I decoding RFs of most bacteria have Ser or Ala at position 246. The Thr246 seemed to be specific for the K12 strains of *E. coli*, since *E. coli* strains such as MRE600 or BL21 were shown to have Ala246 [65]. Ehrenberg and co-workers [65] went on to show using an *in vitro* system that although the termination efficiency of overexpressed RF2 with Ala246 was considerably improved compared with the RF2 with Thr246, it was still significantly less efficient than the endogenous RF2 factor, which also has Ala246. Characterization of the two proteins revealed that the overexpressed RF2 factor lacked an N^5-methylation on glutamine at position 252 of a universally conserved GGQ motif (see below). Indeed, the presence of the methylation stimulated the termination activity of the RF2 factor regardless of the amino acid at position 246, which suggested that the effects are cumulative [65]. It should be noted that RF1 also undergoes methylation although the effect of the modification of the termination efficiency by RF1 does not seem to be as significant as that seen for RF2. Interestingly, the methylase that modifies RF1 and RF2 was identified independently by two different groups [66, 67] and renamed *prmC* [66] from *hemK*, after its mis-assignment as an enzyme involved in the heme biosynthetic pathway (reviewed in [68]). In *E. coli*, the *prmC* gene is located directly downstream of *prfB*, the gene encoding RF2, such that the AUG of the start codon of the former overlaps the UGA stop codon of the latter. Knock-out of *prmC* gene is not lethal but the growth rate is severely reduced, probably due to defects in translational termination. Both RF1 and RF2 isolated from the *ΔprmC* strain lack the methylation at Gln252 as expected. Suppressor strains that partially overcome the growth defect were characterized as having Thr246Ala mutations in the RF2 gene. One of the questions remaining is how does methylation at Gln252 affect the peptide release activity of the decoding RFs?

If the decoding RFs really directly mediate release of the polypeptide, then some part of the factor needs to interact with the PTF center of the ribosome. Since this region consists entirely of highly conserved rRNA nucleotides made up from domain V of the 23S rRNA, a corresponding region of high conservation between all RFs should exist. Despite the lack of sequence homology between bacterial and eukaryotic decoding RFs, a universally conserved sequence of Gly-Gly-Gln (usually referred to as the "GGQ motif") in domain III was identified [69]. The Gln residue of the GGQ motif is the same residue identified by Ehrenberg and co-workers as being methylated in *E. coli*. The GGQ motif has been demonstrated to be essential for peptide release activity of the factor: mutations in either of the glycine residues abolished activity in bacterial RFs *in vivo* and *in vitro* [70]. Indeed, one of these mutants, RF2-GAQ, was shown to be 4–5 orders of magnitude less efficient in the termination reaction than wild-type RF2-GGQ, despite the fact that the binding of both factors to the ribosome were similar [71]. While mutations of Gln (Q) of the GGQ motif to Gly or Ala yielded RFs that retained some *in vitro* termination activity (~20%), these factors could not however rescue the appropriate thermosensitive RF mutants at the non-permissive temperature *in vivo* [70]. Indeed the RF2-GGA mutants were much more active than the RF2-GAQ mutants, by approximately one order of magnitude [71]. Within the RF2 crystal structure the electron density for the loop containing the GGQ motif was poorly resolved suggesting that this region is highly flexible. In eukaryotes, the GGQ motif was also shown to be important for peptide release and was similarly located within a loop at the end in domain I (NTD) of eRF1 (see Sect. 9.3.1 for more details).

9.3.7
Large Conformational Changes Associated with RF2 Binding to the Ribosome

Now there were two sets of very compelling biochemical data, one associating the codon-recognition function of RF2 with the tripeptide motif in domain II and the other associating the peptide-release function of RF2 with the GGQ motif, both of which were located only 45 amino acids away in sequence from one another. Although theoretically this would be sufficient to span from the decoding site on the 30S subunit to the PTF center on the 50S which is a distance of ~70 Å, in the RF2 crystal structure, these motifs are located only 23 Å apart (as seen in Fig. 9.2). To resolve this paradox, the decoding RFs needed to be analyzed in terms of their interaction with the ribosome.

A successful technique for mapping the position of translation factors on the ribosome has been hydroxyl-radical probing. The technique involves the tethering of Fe(II) to sulfydryl groups of cysteine residues on the surface of a protein via a chemical linker and has been applied to define the binding sites of tRNA [72], IF3 [73], EF-G [74], RRF [75] as well as RF1 [76] and RF2 [77]. The results of these latter studies for the RFs demonstrated that tethers placed in close proximity to the tripeptide motif (Fig. 9.6A), cleaved only 16S rRNA in the vicinity of the decoding

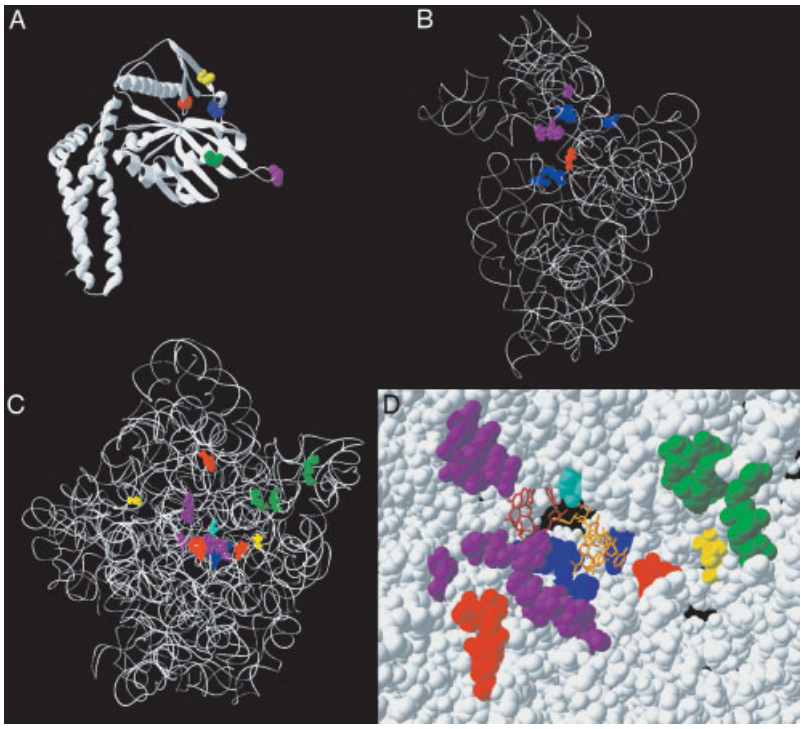


Figure 9.6 The ribosomal binding site of *E. coli* RF2. (A) Identification of Fe(II)-tethering sites on the structure of *E. coli* RF2. Space-fill representation of tethered residues are as follows: Val243 (yellow), Thr246 (purple), Cys274 (red), Leu201 (green) and Ser209 (magenta). (B) Mapping the footprint sites of 201 and 209 of RF2 on the 16S rRNA of the *T. thermophilus* 30S structure: cleavages from position 209 alone are green, whereas those from 201 and 209 are yellow. For reference, the conserved A1492 and A1493 are shown in red. (C) Mapping the footprint sites of 243, 246 and 274 of RF2 on the 23S rRNA of the *H. marismortui:* cleavages from positions 243, 246 and 274 are colored yellow, blue and red, respectively. Where there are cleavages from two positions, the bases are colored green (243 and 246) or purple (246 and 274). (D) The region of the 50S subunit containing the PTF center has been amplified and presented as a space-fill representation with the cleaved residues colored as in (C). For reference, peptide-bond formation products in the A and P sites are colored brown and orange, respectively, and A2451 is colored cyan. Mapping of hydroxyl radical probing data from various tethers placed on RF2 (A) to the 16S rRNA of the 30S (B) and the 23S rRNA of the 50S (C). (D) Close-up of the PTC cleavages (data taken from Scarlett et al. [77]).

site, namely the 790 loop (h24), the 530 loop (h18) and h33 (Fig. 9.6B). These cleavages were similar to those generated from a tether placed at the tip of domain IV of EF-G [74], i.e., the domain that mimics the anticodon stem-loop region of a tRNA and has been shown by cryo-EM to approach the decoding region of the A site (see Chap. 8.4.1). What was surprising was that unlike the probing with EF-G no cleavage in the decoding region (1400 region) was observed. This may reflect the difference in action of the factors, such that RFs physically contact and protect this region during stop signal decoding, whereas EF-G exhibits sequence-independent ribosome binding. Likewise, tethers flanking the $_{250}$GGQ$_{252}$ motif (Fig. 9.6A) cleaved only the 23S rRNA and predominantly within of the PTF center and the so-called GTPase-associated center (GAC; Figs. 9.6C and D). For example, tethers located at either position 246 and/or 273 cleaved position; (i) A2602, a universally conserved bulged nucleotide that becomes protected from chemical probing by the aminoacyl moiety of an A-site-bound tRNA [78]. Cleavage of A2602 was also obtained from a tether placed at the 5′-ACC-end of a deacylated tRNA, (ii) nucleotides 2253–2254 within the P loop of domain V which form Watson–Crick base-pairs with the C74 and C75 positions of a P-site-bound tRNA, and (iii) A2451, which has been proposed to be intimately involved in facilitating peptide-bond formation (see Chap. 8.3.2).

Again, the picture emerged that the tripeptide and GGQ motifs were distinctly associated with the 30S and 50S subunit, respectively, and suggested that RF2 must undergo significant conformational rearrangement upon binding to the ribosome to fulfill this criteria. However, since "seeing is believing", not until the cryo-electron microscopic reconstruction of the RF2-ribosome termination complex was accomplished was the matter finally put beyond reasonable doubt [79, 80]. What these two reconstructions revealed was that the tripeptide motif was in fact located in the decoding site and that domain III came away from domains II/IV such that the GGQ motif located within the loop region now reached into the PTF center (Fig. 9.7). The RF was oriented on the ribosome such that the α-helical bundle of domain I was contacting the GAC, consisting of the base of the stalk region, L11-binding region (H43 and H44 of the 23S rRNA), sarcin-ricin loop (H93) and to some extent L7/L12.

This finding fits nicely with the wealth of biochemical data implicating the GAC as an interaction site for RF1 and RF2. Preliminary footprinting studies performed in the early 1990s using RF2 identified a change in the protection pattern within the 2660 region (H93) of the 23S rRNA (C.M. Brown and W.P. Tate, unpublished results). Cleavages were also identified within this region from tethers located at position 229 in RF1 [76] and 650 of EF-G [74]. In addition, position 229 of RF1 cleaves within the L11-binding region, both in H43 and H44 [76], as do tethers from positions 650 and 655 in domain V of EF-G, which cleaved within the 1095 and 1067 stem-loops, respectively [74]. This region is the target for a family of thiazole antibiotics such as thiostrepton and micrococcin, both of which have been shown to inhibit elongation (see Chap. 12.3.3.1) and termination factor activity [81]. When 70S ribosomes were treated with anti-L11 (and anti-L16), RF-dependent peptidyl-tRNA

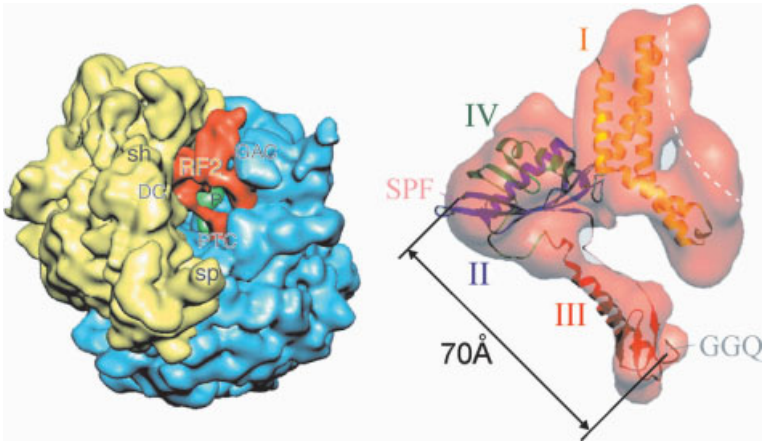


Figure 9.7 Cryo-EM reconstruction of the *E. coli* RF2 termination complex, revealing the significant conformational changes induced in RF2 upon ribosome binding. The approximate distance between the tripeptide motif (SPF in RF2) and the GGQ motif located in the loop of domain III is indicated. Reprinted with permission from Rawat et al. [80].

hydrolysis was partially inhibited whereas binding was not [82–86]. Specifically, the N-terminal region of L11 seems critical for modulating RF function, as antibodies to nucleotides 1–64 as opposed to 65–102 strongly inhibited *in vitro* termination [83].

Although it seems probable that RF1 will also bind analogously to RF2 on the ribosome, it is noteworthy that RF1 is slightly shorter than RF2 and therefore is probably missing the first α-helix of domain I. Consistently, many of the differential effects between RF1 and RF2 have been related to the interaction between the factors with the GAC region. For example, the absence of L11 exhibits a differential effect on RFs, such that RF1 activity is reduced and RF2 activity is enhanced several fold [87, 83]. The presence of L11 is required for RF1 function, specifically the N-terminal domain (NTD) [88], but somehow suppresses RF2 function [85]. Modification of Tyr7 (and negligibly at Tyr61) in the NTD of L11 and reincorporation into 70S ribosomes significantly reduced RF1 but not RF2 termination activity [84]. In contrast, mutations in the 23s rRNA of the L11-binding region seem to influence RF2-mediated termination without affecting RF1 [89]. For example, mutation at G1093A results in UGA-specific suppression [90] and was independent of the disruption of the base pair with A1098 [91]. *In vitro,* this mutation was shown to reduce the association constants of RF1, but more markedly of RF2, with the ribosome [92], suggesting that this region provides more of a binding site for RF2 than RF1. Interestingly, deletion of two nucleotides within domain V of 23S rRNA (G2046 and C2049) could compensate for the conditional lethality caused by the G1093A mutation. These deletions decreased the UGA suppression *in vivo* associated with the G1093A phenotype (reviewed in [93]). Mutations in nucleotides neighboring G1093 also exhibit

UGA-specific suppression, as do substitutions or deletions at A1067 [93]. Taken together, these data suggest that RF1 and RF2 have overlapping binding sites but interact with the ribosomal components in a distinct manner.

9.3.8 The Trigger for RF-mediated Release of the Nascent Chain and the Outcome

If the arrival of the stop codon in the A site is the discriminating factor for decoding RFs, it is logical to presume that the cognate stop codon is the initial signal for the RF to accommodate into the A site by undergoing a conformational transformation. An analogous situation might be the selection process of cognate tRNAs (see Chap. 8.2): here codon–anticodon interaction at the A site presumably transfers a signal to the GTPase center inducing EF-Tu-dependent GTP hydrolysis, which in turn leads to the dissociation of EF-Tu from the ribosome and the release of the aminoacyl-tRNA (aa-tRNA), enabling it to accommodate into the A site on the 50S subunit, i.e., the accommodation of the CCA-end into the PTF center. Recent cryo-EM reconstructions of the EF-Tu•GTP•aa-tRNA stalled with the antibiotic kirromycin reveal that the GAC region moves 7 Å to contact the tRNA [94]. Since the bacterial decoding factors are not delivered to the ribosome and have no GTPase activity, the mechanism to ensure that termination does not occur at non-cognate or sense codons might be a conformational change rather than the factor-mediated GTP hydrolysis. It has been shown that mutations at position 246 in the GGQ-containing loop of domain III could influence both codon-dependent binding as well as peptidyl-tRNA hydrolysis activity of the factor. This led to the suggestion that recognition of the correct stop codon by the RF may induce a conformational change in the factor, a so-called switch [64], so as to bring the GGQ motif into the PTF center. Therefore, the signal, which originates from correct detection of the stop codon in the A site, is relayed through the RF to the tip of domain III. In fact, domain III of RF2 shows structural homology with ribosomal protein S5, the structure of which has been solved for the isolated protein (pdb1pkp, [95]) and on the 30S subunit (1fjf, [96]). The loop of S5, equivalent to the GGQ loop in RF2, adopts a straight β-hairpin upon assembly into the ribosome, i.e., binding to its rRNA segment. By analogy, the tip of the β-hairpin that contains the GGQ motif in RF2 may also insert into the PTF center to modulate release of the polypeptide chain. The question is whether some residue(s) within the decoding RF modulates this reaction, perhaps by co-ordinating a water molecule or whether the RFs mediate this reaction indirectly via residues of the rRNA in the PTF center. It had been proposed that the glutamine residue (Q of the GGQ motif) was involved in co-ordinating the water molecule [69, 97]. However, mutation of GGQ to GGA in both bacterial RF1 and RF2 [70] and in eRF1 [98] produced factors that still retained partial peptidyl-tRNA hydrolysis activity, disproving this notion – indeed, bacterial factors containing GAQ were more severely affected than the GGA mutants [70, 71]. Similarly, mutations of either glycine of the GGQ motif in human eRF1 abolished release activity without affecting ribosome-binding activity of the factor [69]. It should be pointed out that the context of the GGQ motif

in domain III is GxGGx, which is in fact not unique to bacterial RFs, being present also in the loop or linker (α4 and β1) located between domains I and II–IV. Nor is it in fact unique to decoding factors generally, since the GxGGx motif is a well-known turn motif found in, for example, the HIV proteases. This suggests that the GGQ motif may in fact be more important for providing flexibility, i.e., the flexibility between domains I and II–IV could be important for interaction with the GAC and stabilizing the open position of the factor, whereas in domain III, the GxGGx motif may be important for positioning of other residues in this region that directly or indirectly participate in the release reaction. If this is the case and the GGQ motif is not functionally important for the release reaction *per se*, then the focus should be turned to the multitude of highly conserved residues that are found in the region following the GGQ motif, for example, the asparagine (N) residue, located three amino acids C-terminal to the GGQ motif, which is conserved in all bacterial RF sequences known to date [99].

Recently, reconstituted 50S subunits from *T. thermophilus* with 23S rRNA containing mutations at position A2602 within the PTF center were shown to exhibit differential effects with regard to catalysing the peptidyl-transferase reaction in comparison with the termination reaction [100]. In particular, mutations at position A2602, or deletion of this residue altogether, did not significantly affect the former reaction but severely reduced RF1-mediated peptide release, completely abolishing it in some cases. The authors demonstrated that this effect was not due to a decrease in the binding of the decoding factors since (i) increasing the excess of the decoding factors did not compensate for the loss of activity and (ii) the A2602 mutants were inactive under conditions which induce an RF-independent peptidyl-tRNA hydrolysis assay (the presence of 30% acetone [101, 57]) but were active for peptide-bond formation under the same conditions. In the RF-dependent peptidyl-tRNA hydrolysis assay, the authors could demonstrate that the addition of deacylated tRNA cognate to the A site significantly stimulated the reaction. These results suggest that different features of the PTF center are responsible for the PTF reaction during elongation and the peptidyl-tRNA hydrolysis reaction during termination. Furthermore, this was exemplified by the differential effects that a subset of PTF inhibitors had on the two reactions, in particular, the lincosamides, lincomycin and clindamycin completely abolished the PTF reaction at 100 μM, whereas the termination remained unaffected [100]. Thus, this suggests that the release of the peptide by the decoding factors would be indirectly governed through A2602. Since A2602 has been shown to be highly flexible, being in a different conformation in almost every 50S ribosome structure to date, the binding of the RFs may alter the positioning of this residue such that reactive groups can activate a water molecule, thus enabling nucleophilic attack on the carbonyl carbon atom of the peptidyl-tRNA ester bond (Fig. 9.8) [100].

9.4 Eukaryotic Class I Termination Factors

Although a single eRF1 was detected as early as 1971 [102], partially purified [103] and shown to recognize all three stop codons [104], only more recently was the gene

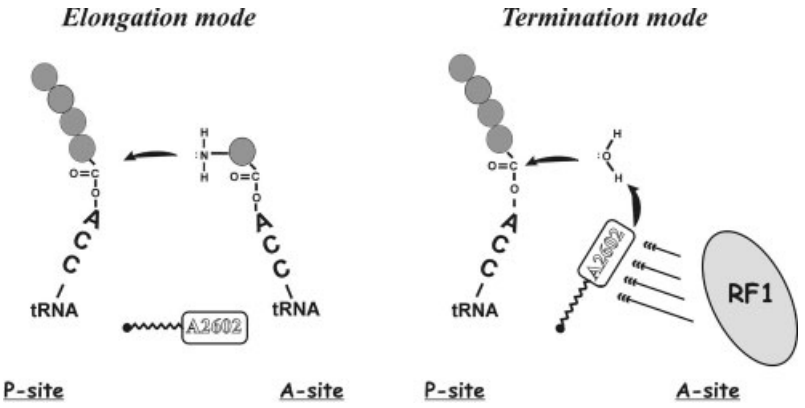


Figure 9.8 The putative conformational switch at A2602 as a trigger for changing the mode of activity of the ribosomal peptidyl transferase center. Orientation of A2602 during translation elongation allows for proper positioning of peptidyl- and aminoacyl-tRNAs in the peptidyl transferase center that makes peptidyl transfer and a new peptide-bond formation possible. Binding of the class 1 release factor (RF1 in the figure) in response to the presence of a stop codon in the decoding site reorients A2602. This places it in a position where its reactive groups can potentially activate a water molecule, facilitating its nucleophilic attack on the carbonyl carbon atom of the peptidyl-tRNA ester bond, thereby accelerating the rate of peptidyl-tRNA hydrolysis. Reprinted with permission from Polacek et al. [100].

correctly identified [105]. Standard methods of sequence comparison showed no significant similarity between the bacterial decoding RFs and the eRF1 family [105]. Although conservation of some sequence elements suggesting a common evolutionary origin has been claimed [106], this study did not even identify the universally conserved GGQ motif associated with peptidyl-tRNA hydrolysis in both prokaryotic and eukaryotic decoding factors ([69]; see Sect. 9.2.6). Furthermore, there is no sign of evolutionary conservation when comparing the crystal structure for *E. coli* RF2 [22] (Fig. 9.2) and human eRF1 (Fig. 9.9; [97]; comment by Kisselev [107]). In contrast, RFs from archeal species (aRF1) show high amino acid sequence homology to eRF1s and have several sequence motifs in common. Indeed, aRF1 from *Methanococcus jannaschii* was shown to be active with mammalian ribosomes, terminating translation at all three stop codons [108], strengthening predictions that aRF1s and eRF1s are descendent from a common evolutionary ancestor [69].

The crystal structure of human eRF1 presents a three-domain molecule with an asymmetric Y-shaped formation for which each of the three domains can be tentatively assigned a function (Fig. 9.9A; [97]): The stem, domain 1 (N-terminal domain), and one arm, domain 2, are proposed to represent the codon recognition and peptidyl-tRNA hydrolysis domains, respectively, whereas the other arm, domain 3 (C-terminal domain), is the site of interaction with eRF3 (reviewed in [1];

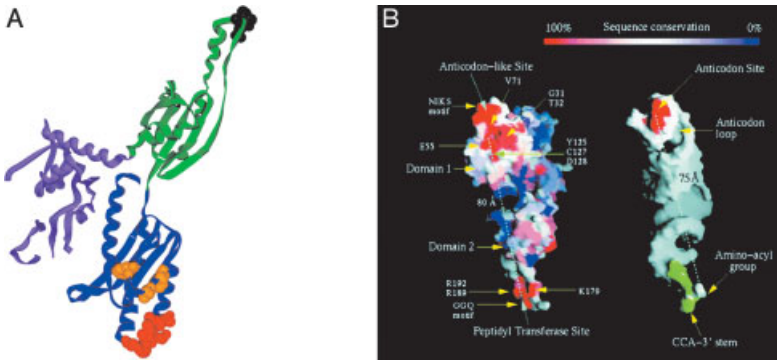


Figure 9.9 The crystal structure of human eRF1 reveals a three-domain structure. (A) Ribbons view of human eRF1 [97] with domain I (blue) with TASNIKS motif (red) and the YxCxxxF (yellow), domain II (green) with GGQ motif in dark green and domain III in purple. (B) Surface representation with sequence conservation: reprinted with permission from Song et al. [97].

see Sect. 9.4.2). The proposed mimicry of a tRNA molecule by eRF1 is certainly not convincing, particular considering the distance between the regions proposed to mimic the anticodon and CCA end, domains I and II, respectively, are approximately ~100 Å apart [22]. This exceeds the 75 Å distance measured from the anticodon stem loop to the CCA end of a tRNA, suggesting that if domains I and II are really associated with the decoding site on the 30S subunit and the PTF center on the 50S subunit, then conformational change upon ribosome binding is necessary to account for this discrepancy. So what is the evidence that associates domains I and II with the aforementioned regions? First, by simply looking at the sequence conservation of human eRF1, it was immediately obvious that there are regions located in domains I and II that are highly conserved (Fig. 9.9B) and also positively charged. These are prime candidates for regions to interact with the decoding and PTF center, since these regions of the ribosome are also highly conserved and composed almost entirely of rRNA, which is negatively charged.

9.4.1 Stop-codon Recognition is Associated with Domain I of eRF1

Reassignment of a stop codon in bacteria can simply involve loss of the appropriate RF, as previously mentioned for the absence of RF2 in *Mycoplasma genitalium*. However, in eukaryotes, there is only one factor that decodes all the stop codons. Therefore, changes in codon reassignment, i.e., the reassignment of a stop codon as sense, should be reflected by changes in the sequences of the eRF1, since this factor should no longer recognize the reassigned stop codon as sense.

In this respect, the use of alternative nuclear genetic codes makes the ciliates perfect for this type of analysis, for example, some ciliate species, such as the *Euplotes*

have reassigned UGA as Cys (C), whereas other species, such as *Tetrahymena*, translate both UAA and UAG with Glu (E) ([109]; reviewed in [110]). Validating this assumption, it was recently demonstrated that, indeed, the ciliate eRF1 does not recognize the reassigned stop codons *in vitro* [111, 112]. A flurry of sequencing activity saw a rapid increase in the number of available ciliate eRF1 gene sequences [100, 113–116]. These types of analyses revealed that most of the convergent changes were indeed associated with domain I; however, the number of positions decreased significantly as the number of ciliate eRF1 gene sequences available increased, challenging the reliability of this method.

Consistent with the assignment of domain I as the codon-recognition domain, random mutagenesis of yeast eRF1 identified numerous locations scattered through domain I, which altered the stop-codon recognition specificity [117]. The mutations located to a groove formed by two helices (α2 and α3), which was proposed to form a binding pocket into which a triplet stop codon was modeled [117]. The involvement of domain I in codon recognition was convincingly demonstrated when hybrid eRF1, containing the domain I from *Tetrahymena* eRF1 (which recognizes only UGA) and domains II and III from *Saccharomyces cerevisiae* eRF1 (which recognizes all stop codons), terminated only at UGA stop codons [111]. Consistently, it was demonstrated recently that a combination of four substitutions in two different regions of domain I had a profound effect on the stop-codon specificity of human eRF1 *in vitro*, such that it only terminated efficiently at UGA stop codons, similar to ciliate eRF1s [118]. This result suggests that in fact two distinct regions within domain I are involved in codon recognition, and that the protein-anticodon mimicry concept [55] may, in contrast with the situation in bacteria where decoding occurs through a simple tripeptide motif, be far too simplistic to describe the situation in eukaryotes and bacteria.

The two prime candidates thought to be involved in stop-codon recognition in eRF1 are two loop regions located in domain I, one containing a heptapeptide sequence $_{58}$TASNIKS$_{64}$ (human eRF1 numbering) and the other a consensus sequence $_{125}$YxCxxxF$_{131}$ (reviewed in [1]). The close proximity of the TASNIKS sequence to the stop codon was confirmed when the Lys (K) residue was found to be crosslinked when synthetic mRNAs containing 4-thiouridine at the first position of the stop codon were used [119]. This suggests that like the situation in bacteria, the protein factor may be directly decoding the stop codon; however, the mechanism may differ significantly.

It is noteworthy that the conditional lethality associated with a mutation in domain I of the yeast eRF1 (P86A) is rescued by compensating mutations A1491G and U1949 located in helix 44 of the decoding region [120]. Interestingly, the mutation G1491 creates a base-pair with C1409 yielding yeast cells that are extremely sensitive to paromomycin. Furthermore, second-site mutations were identified in the switch region at U912C and G886A of the 18S rRNA [120]. Whether this region actually represents a universal switch has recently been brought into doubt by the creation of the equivalent switch mutants in yeast, which did not exhibit the predicted ram or restrictive phenotypes although they did support the involvement of this region in ribosomal fidelity [336]. In any case, the complexity of stop-codon

recognition seems to be a conserved feature between eukaryotes and eubacteria and will require dissection of the (e)RF:termination complex by cryo-EM and crystallization to understand the mechanism fully.

9.4.2
eRF1-mediated Polypeptide Release

As mentioned in Sect. 9.2.6, a universally conserved GGQ motif was identified in all decoding RFs [69]. In eRF1 and aRF1s, the GGQ motif is located in the extremity of domain II, forming a highly exposed minidomain ([97]; Fig. 9.9A). Mutations in the GGQ motif results in loss of peptidyl-tRNA hydrolysis activity, particularly in the first two Gly (G) positions [69, 97], whereas eRF1 with mutations at the Gln (Q) position still retain some activity *in vitro* [98, 121]. If the Lys (K) of the TASNIKS motif is located at the decoding site, then the GGQ motif is separated by 100 Å and therefore is too far apart to fit nicely into the PTF center. Of course, the binding of eRF1 to the ribosome may, in analogy with the bacterial RF2 situation, result in conformational changes in the eRF1 such that the distance between the K and the GGQ is reduced to the optimal 75 Å. This may not be as significant as that observed for RF2 and may simply involve the movement of domain II as proposed in Klaholz et al. [79]. Alternatively, if the YxCxxxF motif, rather than the TASNIKS motif, was directly in contact with the stop codon in the A site, then little or no conformational change in eRF1 would be required since the distance measured between the former motif and the GGQ is precisely 75 Å [1].

9.5
Dissociation of the Post-termination Complex

9.5.1
Eubacterial RF3 Dissociates the Class I Termination Factors

RF3 activity was identified over 30 years ago and the corresponding protein was termed "S", for its ability to stimulate the termination efficiency of RF1 and RF2 [122, 123]. RF3 is not essential for cell survival, since a gene knock-out of *prfC* is viable and the Mycoplasma species have dispensed with this gene. However, the importance of RF3 is illustrated by its necessity for translational fidelity, especially under stress conditions [124]. Utilizing an *in vitro* translation system, RF3 was shown to decrease the recycling time of the decoding RFs [125] by accelerating the dissociation of the decoding RFs from the ribosome [126]. RF3 was shown to have a particularly pronounced effect at strong stop signals [127], where the association rate of the decoding factor for the ribosome is much higher [125]. There is a cost associated with this increased recycling rate, namely a slight reduction in the fidelity of decoding [128].

RF3 contains a GTP-binding motif and thus belongs to the large family of G proteins. It has more sequence similarity with EF-G than to EF-Tu, supporting the contention that RF3 plays a dissociative rather than a delivery role for the decoding RFs.

In fact, the homology between RF3 and EF-G was proposed, based on a dot plot and threading analysis, to extend beyond domains I, II and the G domain of EF-G to include domain II and part of domain IV (Fig. 9.10; [129]). By analogy with EF-G, this would suggest that the N-terminal region of RF3 extends towards the 30S subunit and may exert its dissociative effect on the decoding factors through this extension, perhaps by physically levering the decoding factor from the ribosome.

A recent analysis of the role of guanine nucleotides during RF3 action supports a dissociative role for RF3 [130]. These results suggest that it is the RF3•GDP form that binds a post-termination ribosome complex, i.e., a ribosome that has released the nascent chain but still contains a stop codon and corresponding decoding RF at the A site. Nucleotide exchange occurs on the ribosome and is activated by the post-termination complex. The finding that RF3•GDPNP competes for a binding site with the decoding factors suggests that it is the RF3•GTP form that is responsible for dissociating the decoding factors from the ribosome. This implies that it is the nucleotide exchange (GDP for GTP) that dissociates the decoding factors from the ribosome and not hydrolysis of the GTP to GDP. Instead, hydrolysis of

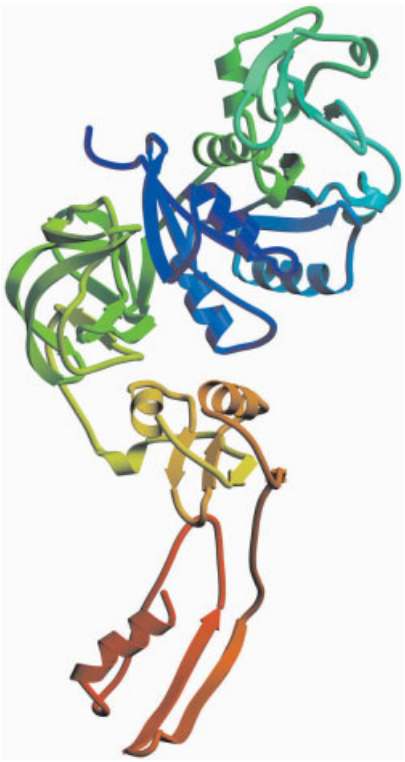


Figure 9.10 Threading model of *E. coli* RF3 based on sequence similarity with EF-G. Reprinted with permission from Wilson et al. [129].

GTP functions to dissociate RF3 since the RF3•GDP form has less affinity for the ribosome due to the absence of the decoding factors, i.e., RF3•GDP has high affinity for ribosome in the *absence* of the peptidyl moiety and the presence of the decoding factors [130]. This ensures that RF3 cannot bind and dissociate the decoding factors until they have completed their job, i.e., release of the nascent polypeptide.

9.5.2 Eukaryotic RF3: Dissociation versus Delivery of eRF1

In contrast with bacterial RF3, eRF3 is an essential gene [131, 132] and has been shown to interact physically and functionally with eRF1 [133, 134]. This interaction involves the C-terminal domain 3 of eRF1, although the exact residues involved appear to differ for different organisms and methods used to determine their interaction [135–137]. Progressive deletion of the C-terminal 6–19 amino acids in *S. cerevisiae* eRF1 [135] and 17 amino acids of *S. pombe* [136] results in a corresponding loss of eRF3 binding. C-terminal deletions disrupt a conserved motif, which with the most recent eRF1 sequences added to the alignment has become GFGGxGG/AxxR and remove a high number of acidic amino acids, mainly glutamic and aspartic acids, which, when mutated to alanine, significantly reduce eRF3 binding [137]. Within the crystal structure for eRF1, the last 15 amino acids, constituting the acidic region, are disordered suggesting some flexibility [97]. These C-terminal residues appear to be dispensable in *Homo sapiens* eRF1 as deletion of the last 22 amino acids (which includes all the acidic residues and part of the conserved motif) did not significantly reduce eRF3-binding capability [138, 137]. Instead, further deletions were necessary to loose eRF3 binding [137]. In any case, the core eRF3-binding region identified for *H. sapiens* eRF1 by these deletion studies (residues 281–415) correlates well with domain 3 from the crystal structure [97].

Another distinction between eubacterial and eukaryotic RF3s is that eRF3 has higher amino acid sequence homology to elongation factor EF-1a (the EF-Tu equivalent present in eubacteria) than to EF2 (the EF-G equivalent), implying that the mode of action of these factors may be different. Heterodimer formation between eRF1 and eRF3 is also suggestive of a delivery mechanism for eRF3, analogous to the delivery of a tRNA to the ribosome by EF-1a. It is probable that this interaction is not strictly necessary as deletion of residues within domain 3 of eRF1 results in the loss of eRF3 interaction, while retaining termination activity and maintaining the viability of the yeast cell [138, 137]. However, it should be noted that these cells exhibit a nonsense suppression phenotype, suggesting a reduction in the efficiency of termination [135, 136]. Although consistent with results where overexpression of both eRF1 and eRF3 was necessary for efficient termination [139], it seems that overexpression of eRF1 alone can also rescue a nonsense suppressor phenotype *in vitro* [140] and *in vivo* [141]. Perhaps the most compelling evidence for the dispensability of eRF3 comes from analyses of the situation in archea. A number of genomes from the archea kingdom have been completely sequenced and all aRF1 genes identified have a shorter C-terminal region of domain 3 lacking the acidic

residues and have a less conserved motif (xFxGxxG/AxLRY/F). Correspondingly, no equivalent gene to eRF3 or bacterial RF3 has been identified in these genomes [142].

Thus, the importance of eRF3 is unclear. Perhaps it derives from another role of eRF3. It has been well documented that the first 114 amino acids of eRF3 are not required for termination activity [143], instead they are implicated with a prion-like form [PSI+] of inheritance (reviewed in [144]). The nonsense suppression phenotype of [PSI+] cells results from the sequestering of eRF1 into large oligomers of eRF3. Under stress conditions, modulation of the cellular levels of solubilized eRFs by chaperones was demonstrated to confer a selective advantage to the yeast cells [145]. Intriguingly, eRF3 has also been shown to play a role in mRNA stability (see Chap. 5.3.2).

9.6 Ribosome Recycling

9.6.1 RRF Mediates Ribosome Recycling in Eubacteria

Following release of the nascent polypeptide and dissociation of the decoding factors by RF3, the cell must recycle the mRNA, deacylated tRNA at the P site and dissociate the 70S ribosome into the constituent subunits, in preparation for the next round of translation. This process is mediated by three factors, a ribosome recycling factor (RRF) that was identified over 30 years ago [146, 147], EF-G and IF3. RRF is an essential gene, the dependence on RRF for cell growth is exemplified again by the Mycoplasma species, which have dispensed with RF3 but retained RRF [2]. RRF and RF3 are both necessary for fast ribosome recycling times. Although their effects are additive, the larger contribution comes from RRF [148]. The exact role of RRF in ribosome recycling is unclear. A model proposed by Ehrenberg and co-workers [149] suggests that RRF, in concert with EF-G, dissociates the 70S subunit but does not release the mRNA or the deacylated tRNA from the P site. Instead, dissociation of the tRNA and mRNA from the 30S subunit is proposed to be a role undertaken by IF3. This model proposes that EF-G has another role to that performed during elongation, namely a dissociative rather than a translocative role, and together with RRF generates a high-energy state necessary for subunit dissociation.

An alternative hypothesis from Kaji and co-workers derives from the remarkable similarity between the crystal structure of RRF and that of a tRNA (see Fig. 9.1C; [28]). In this model, RRF would bind to the ribosomal A site, and EF-G, analogous to its function during elongation, would translocate RRF and the deacylated P site tRNA, from the A and P sites, to the P and E sites respectively. Furthermore, Kaji and co-workers advocate that the role of RRF and EF-G is the removal of the deacylated tRNAs and mRNA from the ribosome but not the dissociation of the 70S ribosome into the component subunits. This latter step is proposed to be mediated by IF3, which is well-known to fulfill this role (see Chap. 7.1). Inhibitors of translocation, such as thiostrepton, aminoglycosides and viomycin, also inhibit ribosome

recycling suggesting that the translocative role of EF-G is also important for the post-termination step [150]. Recently, hydroxyl-radical-probing data [75] have suggested that the mimicry by RRF of a tRNA is in fact misleading, since RRF binds to the ribosome with domain I (the region that was proposed to mimic the anticodon stem loop of a tRNA) extending into the PTF center on the 50S subunit; in fact, RRF makes little contact with the 30S subunit. Since there is some overlap in the position of domain II and the binding site of EF-G, this suggests that EF-G binding may propel RRF through the ribosome to clear out the tRNAs and mRNAs.

RRF is not present in the cytoplasm of eukaryotes (or archaea); the only forms present in eukaryotes are either mitochondrial or plastid in the case of plants; therefore, it will be interesting to see how ribosome recycling is mediated in this case.

References

1 L. Kisselev, M. Ehrenberg, L. Frolova, *EMBO J.* **2003,** *22,* 175–182.
2 C. M. Fraser, J. D. Gocayna, O. White et al., *Science* **1995,** *270,* 397–403.
3 C. A. Hutchison, S. N. Peterson, S. R. Gill et al., *Science* **1999,** *286,* 2165–2169.
4 T. Caskey, E. Scolnick, R. Tompkins et al., *Cold Spring Harbor Symp. Quant. Biol.* 1969, *34,* 479–488.
5 E. Scolnick, R. Tompkins, T. Caskey et al., *Proc. Natl. Acad. Sci. USA* **1968,** *61,* 768–774.
6 J. M. Inamine, K. Ho, S. Loechel et al., *J. Bacteriol.* **1990,** *172,* 504–506.
7 S. Osawa, T. H. Jukes, *J. Mol. Evol.* **1995,** *41,* 247–249.
8 J. M. O'Sullivan, J. B. Davenport, M. F. Tuite, *Trends Genet.* **2001,** *17,* 20–22.
9 H. J. Pel, C. Maat, M. Rep et al., *Nucleic Acids Res.* **1992,** *20,* 6339–6346.
10 C. T. Caskey, W. C. Forrester, W. Tate et al., *J. Bacteriol.* **1984,** *158,* 365–368.
11 R. B. Weiss, J. P. Murphy, J. A. Gallant, *J. Bacteriol.* **1984,** *158,* 362–364.
12 H. J. Pel, M. Rep, L. A. Grivell, *Nucleic Acids Res.* **1992,** *20,* 4423–4428.
13 K. Ito, K. Ebihara, M. Uno et al., *Proc. Natl. Acad. Sci. USA* **1996,** *93,* 5443–5448.
14 J. G. Moffat, B. C. Donly, K. K. Mccaughan et al., *Eur. J. Biochem.* **1993,** *213,* 749–756.
15 J. G. Moffat, W. P. Tate, *J. Biol. Chem.* **1994,** *269,* 18899–18903.
16 P. Nissen, M. Kjeldgaard, S. Thirup et al., *Science* **1995,** *270,* 1464–1472.
17 A. Aevarsson, E. Brazhnikov, M. Garber et al., *EMBO J.* **1994,** *13,* 3669–3677.
18 J. Czworkowski, J. Wang, T. A. Seitz et al., *EMBO J.* **1994,** *13,* 3661–3668.
19 P. Nissen, M. Kjeldgaard, J. Nyborg, *EMBO J.* **2000,** *19,* 489–495.
20 S. Brock, K. Szkaradkiewicz, M. Sprinzl, *Mol. Microbiol.* **1998,** *29,* 409–417.
21 A. P. Carter, W. M. Clemons, Jr., D. E. Brodersen et al., *Science* **2001,** *291,* 498–501.
22 B. Vestergaard, L. Van, G. Andersen et al., *Mol. Cell* **2001,** *8,* 1375–1382.
23 O. Tin, A. Rykunova, T. Muranova et al., *Biochimie* **2000,** *82,* 765–772.
24 S. Vörtler, H. Dobbek, I. Dieser et al., Translation Termination: Structure of T. th. RF2 and its Interaction with RNA, in *8th Annual tRNA Meeting,* Banz, Germany 2003.
25 D. E. Brodersen, V. Ramakrishnan, *Nat. Struct. Biol.* **2003,** *10,* 78–80.
26 M. Kjeldgaard, *Mol. Cell* **2003,** *11,* 8–10.
27 Y. Nakamura, M. Uno, T. Toyoda et al., *Cold Spring Harb. Symp. Quant. Biol.* **2001,** *66,* 469–475.
28 M. Selmer, S. Al-Karadaghi, G. Hirakawa et al., *Science* **1999,** *286,* 2349–2352.
29 R. Martin, M. Hearn, P. Jenny et al., *Mol. Gen. Genet.* **1988,** *213,* 144–149.
30 R. Martin, M. Weiner, J. Gallant, *J. Bacteriol.* **1988,** *170,* 4714–4717.

31 T. Elliott, X. H. Wang, *J. Bacteriol.* **1991,** *173,* 4144–4154.
32 K. Kawakami, T. Inada, Y. Nakamura, *J. Bacteriol.* **1988,** *170,* 5378–5381.
33 K. Kawakami, Y. Nakamura, *Proc. Natl. Acad. Sci. USA* **1990,** *87,* 8432–8436.
34 O. Mikuni, K. Kawakami, Y. Nakamura, *Biochimie* **1991,** *73,* 1509–1516.
35 O. Olafsson, J. U. Ericson, R. Vanbogelen et al., *J. Bacteriol.* **1996,** *178,* 3829–3839.
36 S. M. Ryden, L. A. Isaksson, *Mol. Gen. Genet.* **1984,** *193,* 38–45.
37 B. C. Donly, C. D. Edgar, F. M. Adamski et al., *Nucleic Acids Res.* **1990,** *18,* 6517–6522.
38 H. J. Pel, M. Rep, H. J. Dubbink et al., *Nucleic Acids Res.* **1993,** *21,* 5308–5315.
39 M. E. Askarian-Amiri, H. J. Pel, D. Guevremont et al., *J. Biol. Chem.* **2000,** *275,* 17241–17248.
40 J. S. Lodmell, A. E. Dahlberg, *Science* **1997,** *277,* 1262–1267.
41 W. Tate, B. Greuer, R. Brimacombe, *Nucleic Acids Res.* **1990,** *18,* 6537–6544.
42 V. Ivanov, A. Beniaminov, A. Mikheyev et al., *RNA* **2001,** *7,* 1683–1692.
43 C. M. Brown, W. P. Tate, *J. Biol. Chem.* **1994,** *269,* 33164–33170.
44 W. P. Tate, S. A. Mannering, *Mol. Microbiol.* **1996,** *21,* 213–219.
45 W. P. Tate, E. S. Poole, M. E. Dalphin et al., *Biochimie* **1996,** *78,* 945–952.
46 W. P. Tate, E. S. Poole, J. A. Horsfield et al., *Biochem. Cell Biol.* **1995,** *73,* 1095–1103.
47 E. S. Poole, C. M. Brown, W. P. Tate, *EMBO J.* **1995,** *14,* 151–158.
48 E. Poole, L. Major, S. Mannering et al., *Nucleic Acids Res.* **1998,** *26,* 954–960.
49 L. L. Major, T. D. Edgar, P. Yee Yip et al., *FEBS Lett.* **2002,** *514,* 84–89.
50 E. S. Poole, R. Brimacombe, W. P. Tate, *RNA* **1997,** *3,* 974–982.
51 L. Poole, W. Tate, *Encyclopedia of Life Sci.* **1999.**
52 K. Ito, M. Uno, Y. Nakamura, *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 8165–8169.
53 K. Ito, M. Uno, Y. Nakamura, *Nature* **2000,** *403,* 680–684.
54 Y. Nakamura, K. Ito, *FEBS Lett.* **2002,** *514,* 30–33.
55 Y. Nakamura, K. Ito, M. Ehrenberg, *Cell* **2000,** *101,* 349–352.
56 M. Uno, K. Ito, Y. Nakamura, *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 1819–1824.
57 C. T. Caskey, E. Scolnick, R. Tompkins et al., *Methods Enzymol.* **1971,** *20,* 367–375.
58 B. E. Maden, R. E. Monro, *Eur. J. Biochem.* **1968,** *6,* 309–316.
59 W. P. Tate, C. M. Brown, *Biochemistry* **1992,** *31,* 2443–2450.
60 F. M. Adamski: *Regulation of Prokaryote Release Factors 1 and 2* University of Otago 1992.
61 J. A. Horsfield, D. N. Wilson, S. A. Mannering et al., *Nucleic Acids Res.* **1995** *23,* 1487–1494.
62 W. P. Tate, F. M. Adamski, C. M. Brown et al.: in *The Translational Apparatus: Structure, Function, Regulation, Evolution.,* eds K. H. Nierhaus, F. Franceschi, A. R. Subramanian et al., Plenum New York 1993, 253–262.
63 M. Uno, K. Ito, Y. Nakamura, *Biochimie* **1996,** *78,* 935–943.
64 D. N. Wilson, D. Guevremont, W. P. Tate, *RNA* **2000,** *6,* 1704–1713.
65 V. Dincbas-Renqvist, A. Engstrom, L. Mora et al., *EMBO J.* **2000,** *19,* 6900–6907.
66 V. Heurgue-Hamard, S. Champ, A. Engstrom et al., *EMBO J.* **2002,** *21,* 769–778.
67 K. Nakahigashi, N. Kubo, S. Narita et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 1473–1478.
68 S. Clarke, *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 1104–1106.
69 L. Frolova, R. Tsivkovskii, G. Sivolobova et al., *RNA* **1999,** *5,* 1014–1020.
70 L. Mora, V. Heurgue-Hamard, S. Champ et al., *Mol. Microbiol.* **2003,** *47,* 267–275.
71 A. V. Zavialov, L. Mora, R. H. Buckingham et al., *Mol. Cell* **2002,** *10,* 789–798.
72 S. Joseph, H. F. Noller, *EMBO J.* **1996,** *15,* 910–916.

73 A. Dallas, H. F. Noller, *Mol. Cell* **2001**, *8*, 855–864.
74 K. S. Wilson, H. F. Noller, *Cell* **1998**, *92*, 131–139.
75 L. Lancaster, M. C. Kiel, A. Kaji et al., *Cell* **2002**, *111*, 129–140.
76 K. S. Wilson, K. Ito, H. F. Noller et al., *Nat. Struct. Biol.* **2000**, *7*, 866–870.
77 D. J. Scarlett, K. K. McCaughan, D. N. Wilson et al., *J. Biol. Chem.* **2003**, *278*, 15095–15104.
78 D. Moazed, H. F. Noller, *Cell* **1989**, *57*, 585–597.
79 B. P. Klaholz, T. Pape, A. V. Zavialov et al., *Nature* **2003**, *421*, 90–94.
80 U. B. Rawat, A. V. Zavialov, J. Sengupta et al., *Nature* **2003**, *421*, 87–90.
81 N. Brot, W. P. Tate, C. T. Caskey et al., *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 89–92.
82 W. P. Tate, C. T. Caskey, G. Stoffler, *J. Mol. Biol.* **1975**, *93*, 375–389.
83 W. P. Tate, J. Dognin, M. Noah et al., *J. Biol. Chem.* **1984**, *259*, 7317–7324.
84 W. P. Tate, K. K. McCaughan, C. D. Ward et al., *J. Biol. Chem.* **1986**, *261*, 2289–2293.
85 W. P. Tate, H. Schulze, K. H. Nierhaus, *J. Biol. Chem.* **1983**, *258*, 12816–12820.
86 W. P. Tate, H. Schulze, K. H. Nierhaus, *J. Biol. Chem.* **1983**, *258*, 12810–12815.
87 K. K. McCaughan, C. D. Ward, C. N. A. Trotman et al., *FEBS Lett.* **1984**, *175*, 90–94.
88 N. Van Dyke, E. Murgola, *J. Mol. Biol.* **2003**, *330*, 9–13.
89 W. Xu, F. Pagel, E. Murgola, *J. Bacteriol.* **2002**, *184*, 1200–1203.
90 D. K. Jemiolo, F. T. Pagel, E. J. Murgola, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12309–12313.
91 W. B. Xu, E. J. Murgola, *J. Mol. Biol.* **1996**, *264*, 407–411.
92 A. L. Arkov, D. V. Freistroffer, M. Y. Pavlov et al., *Biochimie* **2000**, *82*, 671–682.
93 A. Arkov, E. Murgola, *Biochemistry (Moscow)* **1999**, *64*, 1354–1135.
94 M. Valle, A. Zavialov, W. Li et al., *Nat. Struct. Biol.* **2003**, *10*, 899–906.
95 V. Ramakrishnan, S. W. White, *Nature* **1992**, *358*, 768–771.
96 B. T. Wimberly, D. E. Brodersen, W. M. Clemons et al., *Nature* **2000**, *407*, 327–339.
97 H. Song, P. Mugnier, A. Das et al., *Cell* **2000**, *100*, 311–321.
98 A. Seit Nebi, L. Frolova, N. Ivanova et al., *Mol. Biol. (Moscow)* **2000**, *34*, 899–900.
99 P. V. Baranov, O. L. Gurvich, A. W. Hammer et al., *Nucl. Acids Res.* **2003**, *31*, 87–89.
100 N. Polacek, M. J. Gomez, K. Ito et al., *Mol. Cell* **2003**, *11*, 103–112.
101 C. T. Caskey, A. L. Beaudet, E. M. Scolnick et al., *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 3163–3167.
102 A. L. Beaudet, C. T. Caskey, *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 619–624.
103 C. T. Caskey, A. L. Beaudet, W. P. Tate, *Methods Enzymol.* **1974**, *30*, 293–303.
104 D. S. Konecki, K. C. Aune, W. P. Tate et al., *J. Biol. Chem.* **1977**, *252*, 4514–4520.
105 L. Frolova, X. Le Goff, H. H. Rasmussen et al., *Nature* **1994**, *372*, 701–703.
106 Y. Nakamura, K. Ito, L. A. Isaksson, *Cell* **1996**, *87*, 147–150.
107 L. Kisselev, *Structure (Cambridge)* **2002**, *10*, 8–9.
108 M. Dontsova, L. Frolova, J. Vassilieva et al., *FEBS Lett.* **2000**, *472*, 213–216.
109 R. D. Knight, L. F. Landweber, *Cell* **2000**, *101*, 569–572.
110 C. A. Lozupone, R. D. Knight, L. F. Landweber, *Curr. Biol.* **2001**, *11*, 65–74.
111 K. Ito, L. Frolova, A. Seit-Nebi et al., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8494–8499.
112 S. Kervestin, L. Frolova, L. Kisselev et al., *EMBO Rep.* **2001**, *2*, 680–684.
113 Y. Inagaki, W. F. Doolittle, *Nucleic Acids Res.* **2001**, *29*, 921–927.
114 A. Karamyshev, Z. Karamysheva, K. Ito et al., *Biochemistry (Moscow)* **1999**, *64*, 1391–1400.

115 S. Kervestin, O. Garnier, A. Karamyshev et al., *J. Eukaryot. Microbiol.* **2002,** *49,* 374–382.
116 A. Liang, C. Brunen-Nieweler, T. Muramatsu et al., *Gene* **2001,** *262,* 161–168.
117 G. Bertram, H. A. Bell, D. W. Ritchie et al., *RNA* **2000,** *6,* 1236–1247.
118 A. Seit-Nebi, L. Frolova, L. Kisselev, *EMBO Rep.* **2002,** *3,* 881–886.
119 L. Chavatte, A. Seit-Nebi, V. Dubovaya et al., *EMBO J.* **2002,** *21,* 5302–5311.
120 I. V. Velichutina, J. Dresios, J. Y. Hong et al., *RNA* **2000,** *6,* 1174–1184.
121 A. Seit-Nebi, L. Frolova, J. Justesen et al., *Nucleic Acids Res.* **2001,** *29,* 3982–3987.
122 M. R. Capecchi, H. A. Klein, *Cold Spring Harbor Symp. Quant. Biol.* 1969, *34,* 469–477.
123 G. Milman, J. Goldstein, E. Scolnick et al., *Proc. Natl. Acad. Sci. USA* **1969,** *63,* 183–190.
124 O. Mikuni, K. Ito, J. Moffat et al., *Proc. Natl. Acad. Sci. USA* **1994,** *91,* 5798–5802.
125 M. Y. Pavlov, D. V. Freistroffer, V. Heurgue-Hamard et al., *J. Mol. Biol.* **1997,** *273,* 389–401.
126 D. V. Freistroffer, M. Y. Pavlov, J. MacDougall et al., *EMBO J.* **1997,** *16,* 4126–4133.
127 D. J. Crawford, K. Ito, Y. Nakamura et al., *EMBO J.* **1999,** *18,* 727–732.
128 D. V. Freistroffer, M. Kwiatkowski, R. H. Buckingham et al., *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 2046–2051.
129 D. N. Wilson, M. E. Dalphin, H. J. Pel et al.: in *The Ribosome. Structure, Function, Antibiotics, and Cellular Interactions,* eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, 495–508.
130 A. V. Zavialov, R. H. Buckingham, M. Ehrenberg, *Cell* **2001,** *107,* 115–124.
131 V. V. Kushnirov, M. D. Ter-Avanesyan, M. V. Telkov et al., *Gene* **1988,** *66,* 45–54.
132 P. G. Wilson, M. R. Cuthbertson, *J. Mol. Biol.* **1988,** *199,* 559–573.
133 I. Stansfield, L. Eurwilaichitr, Akhmaloka et al., *Mol. Microbiol.* **1996,** *20,* 1135–1143.
134 G. Zhouravleva, L. Frolova, X. Legoff et al., *EMBO J.* **1995,** *14,* 4065–4072.
135 L. Eurwilaichitr, F. M. Graves, I. Stansfield et al., *Mol. Microbiol.* **1999,** *32,* 485–496.
136 K. Ito, K. Ebihara, Y. Nakamura, *RNA* **1998,** *4,* 958–972.
137 T. I. Merkulova, L. Y. Frolova, M. Lazar et al., *FEBS Lett.* **1999,** *443,* 41–47.
138 L. Y. Frolova, T. I. Merkulova, L. L. Kisselev, *RNA* **2000,** *6,* 381–390.
139 I. Stansfield, K. M. Jones, V. V. Kushnirov et al., *EMBO J.* **1995,** *14,* 4365–4373.
140 G. Drugeon, O. JeanJean, L. Frolova et al., *Nucleic Acids Res.* **1997,** *25,* 2254–2258.
141 X. LeGoff, M. Philippe, O. JeanJean, *Mol. Cell. Biol.* **1997,** *17,* 3164–3172.
142 Y. Nakamura, K. Ito, *Genes Cells* **1998,** *3,* 265–278.
143 M. D. Ter-Avanesyan, V. V. Kushnirov, A. R. Dagkesamanskaya et al., *Mol. Microbiol.* **1993,** *7,* 683–692.
144 M. F. Tuite, *Cell* **2000,** *100,* 289–292.
145 S. S. Eaglestone, B. S. Cox, M. F. Tuite, *EMBO J.* **1999,** *18,* 1974–1981.
146 A. Hirashima, A. Kaji, *Biochem. Biophys. Res. Commun.* **1970,** *41,* 877–883.
147 A. Hirashima, A. Kaji, *J. Mol. Biol.* **1972,** *65,* 43–58.
148 M. Y. Pavlov, D. V. Freistroffer, J. MacDougall et al., *EMBO J.* **1997,** *16,* 4134–4141.
149 R. Karimi, M. Pavlov, R. Buckingham et al., *Mol. Cell* **1999,** *3,* 601–609.
150 M. Kiel, V. Raj, H. Kaji et al., *J. Biol. Chem.* **2003,** *278,* 48041–48050.
151 E. D. Wu, H. Inokuchi, H. Ozeki, *Jpn J. Genet.* **1990,** *65,* 115–119.

10
The Mechanism of Recoding in Pro- and Eukaryotes

Elizabeth S. Poole, Louise L. Major, Andrew G. Cridge, and Warren P. Tate

10.1
Introduction

During protein synthesis, the ribosome uses mechanisms that maintain the translational frame and the nature of the interactions between the RNA participants are critical to this process. Of the RNA visitors to the ribosome, the tRNA occupies two of three possible sites [1] at the active center depending on whether the ribosome is in the pre-translocational or post-translocational state of the polypeptide-chain-elongation cycle [2]. In the pre-translocation state, the tRNAs are at the A-site (as the reaction substrate) and the P-site (as the reaction intermediate). In the post-translocation state, the tRNAs occupy the P-site and the E-site (as the reaction product). These pairs of tRNAs make unique interactions with the host structural rRNAs in their particular environments, and the tRNAs themselves undergo some conformational flexing during these interactions which are important for maintaining canonical events. The other key RNA visitor to the ribosome during protein synthesis is the mRNA that occupies a specific channel in the neck of the small ribosomal subunit as it threads through the decoding site during triplet decoding [3]. This threading does not impose a stress on the triplet code reading frame unless there is a 'tangle' of some kind in the downstream region of the mRNA. Such ordered tangles, commonly in the form of stem-loops or pseudoknots, can impose sufficient pressure to facilitate a non-canonical or 'recoding' event in the form of a change in frame in the mRNA or a change in the interpretation of a particular codon. Generally, this will occur only when additional particular sequence motifs occur in the mRNA itself. Under these conditions, there is a recoding event, since the expected translational event does not occur because of the new meaning of the codon or because of a subtle or more significant reading frame change [4].

Genetic recoding initially only seemed to happen in the test tube under special conditions, or, at most, to be the domain of viral RNAs that use extraordinary means to subvert the host ribosomes' fidelity for their own purposes. Clearly, viruses are major users of recoding mechanisms on host-cell ribosomes, but, in addition, it is now well established that cells themselves use recoding strategies as another layer of regulating the expression of a small subset of their genes. Interest in genetic recoding has grown as its importance as a mainstream mechanism of

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

regulating gene expression has become apparent and as its absolute importance in the biology of some pathogenic viruses, like HIV-1, has become obvious. Speculation is now occurring as to whether recoding is widespread as a mechanism to produce minor products in addition to standard proteins from a particular mRNA. This could contribute to the proteome in ways that are beyond comprehension at present [5]. However, genetic recoding is not just restricted to changes in the translational frame. It is a term used to encompass other non-canonical events that give unexpected results from the translation of a signal in the mRNA. For example, included in this definition are the following:

1. Selenocysteine incorporation at a stop codon in a small subset of UGA stop codons in pro- and eukaryotic genes [6].
2. Incorporation of amino acids by the decoding of stop codons by near-cognate tRNAs (commonly referred to as readthrough) [7].
3. Disengaging and slipping through a section of the mRNA (bypassing) [8].
4. Frameshifting on the mRNA, either + or – (slipping forward or slipping backward) [9].

10.2 Maintaining Decoding Accuracy and the Reading Frame

Ribosomes are not absolute in their avoidance of error during translation of an mRNA; they incorporate an incorrect amino acid only occasionally, perhaps 1 in $10^3 – 10^4$ times and they lose the reading frame of the mRNA maybe 1 in $10^4 – 10^5$ times. Fortunately, this level of error does not compromise the ability of the ribosome to make a protein of 1000 amino acids. This means that if recoding is to occur, then there must be active mechanisms to promote recoding rather than relying on natural error, especially if it is to occur at a specific site. We know now from the work of Ramakrishnan and co-workers [10] that the rRNA uses a sensing mechanism of the codon–anticodon interaction in each of the three nucleotide positions when the substrate tRNA is in the A-site. For example, interaction of a 16S rRNA base (A1493) with the 2′-hydroxyl in each ribose of the nucleotides in the first base pair precludes non-Watson–Crick interaction at this position. In contrast, the wobble base pair is not as constrained as the other two base pairs and allows for more variety in the kinds of tRNA:mRNA interactions at this position. Similarly, the structures of the 70S ribosome with a P-site tRNA and the 30S subunit, where part of a second subunit molecule was found to mimic the decoding stem of the P-site tRNA, indicate that there are constraints on interactions at the P-site. At this site, the mRNA is forced to adopt a kinked formation allowing the anticodon stems of the A- and P-site tRNAs to be relatively far apart. The ribosome structures have revealed much detail and allowed insight into how decoding accuracy can occur, resulting in speculation of constraints that prevent a shift in reading frame.

Clearly, there are mechanisms and constraints through structural interactions of the tRNAs with the rRNA and the mRNA that are strong determinants for the canonical events of protein synthesis and ensure that non-canonical events are an exception rather than the rule. For a non-canonical event to occur there have to be

extraordinary circumstances that overcome the normal restraints on such events. These can be primary sequence *cis* signals and/or specific secondary structures in the mRNA spaced at particular distances from a primary signal. These mRNA elements somehow perturb the normal kinetics of protein synthesis, often causing pauses at decoding sites that allow competition between canonical and non-canonical events. The non-canonical result will occur in a proportion of the ribosomal passages through the signal with the frequency dependent on competition strength between the canonical and non-canonical events.

10.3 The Use of a Stop Signal for both Elongation and Termination of Protein Synthesis

Although the stop codons (UAA, UAG, and UGA) were once thought to be used universally as stop signals in protein synthesis, there are now many specific examples where they have been captured to encode amino acids. However, in most of these instances they only signal stop or sense. For example, in mitochondria and in mycobacterium species, UGA is frequently used to code for tryptophan but in these cases does not signal stop. Also, unicellular eukaryotic organisms such as Tetrahymena use UGA for stop whereas UAA and UAG code for glutamine. These are all examples where there has been codon 'takeover' or, perhaps, 'reassignment' in different organisms although the events that they signal are still canonical processes of protein synthesis.

An exception to this kind of promiscuity for stop codons is when a UGA signals selenocysteine (Sec) in a small number of genes but still signals stop in the same organism in the vast majority of occurrences. This implies that elongation and termination must be in competition at the signal. This competition occurs in a wide range of organisms in the eubacteria, archaea and eukarya kingdoms. The mechanisms for incorporation of Sec at UGA sites are distinct in prokaryotic and eukaryotic organisms, although there are some similarities. For example, there is a secondary structural element that is critical for the signal to function in both but in prokaryotes it is within the coding region and close to the primary signal UGA, whereas in eukaryotes it is quite distant and found in the 3′-hychroxyl untranslated region. In both cases, there is a special tRNA for Sec that is a minor isoacceptor of a serine tRNA where the serine has been modified by specific proteins.

10.4 The Mechanism for Sec Incorporation at UGA Sites in Bacterial mRNAs

Elegant studies from Böck and co-workers [11, 12] through the 1980s and 1990s defined the genes that were responsible for incorporation of Sec into proteins and largely defined the mechanism of how this occurred at the ribosome. Four genes controlling this mechanism were defined *selA–D*.

10.4.1 The Gene Products

1. The *selD* gene product is a 37 kDa monomeric protein, selenophosphate synthetase [13]. Although the identity of the physiological selenium substrate that is phosphorylated by ATP is still uncertain, there is evidence for an enzyme-bound phosphoryl intermediate [14], which is then attacked *in vitro* by selenide at the active site releasing the selenophosphate product. Selenide is a highly reactive molecule and may not be the physiological substrate or, if so, may be sequestered by another molecule for this reaction.
2. *SelC* is the gene encoding a minor serine tRNA that is first aminoacylated with serine by the normal synthetase [15] and after conversion of Ser to Sec is subsequently used during translation to deliver Sec into the elongating polypeptide. Interestingly, the tRNA is not well recognized by the typical elongation factor, EF-Tu, that delivers aminoacyl-tRNAs to the ribosome. This implies that the Ser-$tRNA^{Sec}$ would have structural features unlike all other tRNAs and, indeed, it does have subtle differences in its structure. At a length of 95 nucleotides, it is one of the longest tRNAs known and this is largely because it has a large variable arm. Moreover, several invariant residues in other tRNAs are different in the *selC* gene product and, significantly, the amino acceptor arm has eight rather than the seven base pairs of other tRNAs [16].
3. *SelA* encodes a 50 kDa subunit of an oligomeric protein comprising 10 subunits. Each subunit has a pyridoxyl phosphate moiety. Conversion of Ser-$tRNA^{Sec}$ to the Sec derivative is catalyzed by this enzyme using the selenophosphate as a substrate donor of selenium. The whole conversion takes place on this enzyme [17]. The serine is converted into amino acrylyl derivative by elimination of water from the seryl moiety first and then the activated selenium derivative is added to this intermediate to complete the conversion. The enzyme has a high degree of specificity for the Ser-tRNA, with one tRNA bound per two subunits [18]. Using electron microscopy, it was determined that the enzyme comprises a double ring of five subunits each, consistent with the stoichiometry of tRNA binding (five per enzyme) [19]. The extra long amino acceptor stem of the tRNA and its large variable loop are both important for this binding.
4. *SelB* encodes the specific elongation factor that recognizes the Sec-$tRNA^{Sec}$ and is clearly important for the delivery of Sec to the elongating polypeptide. It is a protein of ~69 kDa and exhibits a high degree of sequence similarity to both EF-Tu and the initiation factor, IF-2, within its N-terminal region (244 amino acids). SELB is much bigger than EF-Tu (69 kDa versus 43 kDa) and the C-terminal extension on SELB not shared with EF-Tu may have some other function such as recognizing the mRNA context of the UGA recoding site [20]. The protein cannot bind Ser-$tRNA^{Sec}$ in contrast with the Sec derivative and this explains why Ser is not incorporated at a UGA recoding site. The major determinant on the tRNA for binding to SELB is the eight base amino acceptor stem and conversion to the typical seven base pair stem abolishes binding [21].

10.4.2
The Mechanism of Sec Incorporation

Sec incorporation involves an intriguing mechanism in which Sec-tRNA^{Sec} and SELB are major players. As well, SECIS (selenocysteine insertion sequence) elements are critical in a small number of specific mRNAs such as that for formate dehydrogenase F (FDHF) [22]. These mRNAs contain UGA and a stem-loop (the SECIS element). In *fdhF,* it was originally predicted that the stem-loop closely followed the UGA codon but further studies have suggested that the UGA is within the stem-loop structure that forms the SECIS element before it approaches the ribosomal decoding site [23]. SELB binds to the stem at a specific site in the apical loop and upper helical region. The structure of the loop rather than its primary sequence seems to be the important determinant. A bulged region on the upper 5′-arm of the stem and nucleotides in the apical loop are protected by SELB from hydroxyl radical cleavage in footprint experiments.

Over-expression of SELB and SELC does not lead to a misincorporation of Sec at typical UGA stop codons. This indicates that delivery of SELB.GTP.Sec-tRNA^{Sec} to the ribosome is different than that for other tRNAs in ternary complexes. Hüttenhofer and Böck [24] have obtained evidence that suggests the ternary complex may be in a 'pre-competent state' before binding to the mRNA stem-loop. A variety of approaches indicate that SELB must be complexed with the SECIS element for a productive interaction with the ribosome to occur. These studies suggest that binding to the SECIS element induces a conformational switch in SELB that facilitates the formation of an anticodon:codon interaction between the Sec-tRNA^{Sec} and the UGA codon as it reaches the ribosomal A-site. The SECIS element would then act like a safety switch, preventing normal UGA termination codons being decoded as Sec by the SELB.GTP.Sec-tRNA^{Sec} ternary complexes [24, 25]. Once the switch converts SELB into a 'competent state', it would give SELB a strong selective advantage when it reaches the ribosomal A-site and is in competition with the decoding release factor, RF2, to decode the now A-site UGA. In this way, the SECIS element acts not only as a functional switch for protection, but also as a facilitator to send the ternary complex along a kinetic path whereby Sec is incorporated into the polypeptide chain.

10.4.3
The Competition between Sec Incorporation and Canonical Decoding of UGA by RF2

Factors affecting the competition between the RF2 and Sec incorporation *in vivo* during translation of the *fdhF* Sec (UGA) recoding site have been defined with wild-type and modified *fdhF* sequences [26]. Altering sequences surrounding the UGA codon to create more or less efficient UGA-containing stop signals without affecting the secondary structure of the SECIS element, have indicated that the kinetics of stop signal decoding have a significant influence on Sec incorporation efficiency. The UGA codon in the specific *fdhF* sequence remains 'visible' to the decoding RF2 that *in vitro* can form a site-directed 'zero-length' crosslink to it when the secondary

structure of the mRNA created by the stem-loop is absent [27]. Increasing the cellular concentration of either the RF2 decoding molecule for termination, or the $tRNA^{Sec}$ decoding molecule for elongation (for Sec incorporation), showed that these molecules are able to compete for the UGA by a kinetic competition that is dynamic and dependent on the growth rate of *Escherichia coli*. The $tRNA^{Sec}$-mediated decoding can compete more effectively for the recoding site UGA at lower growth rates, consistent with the well-established anaerobic induction of *fdhF* expression, when, presumably, Sec-containing enzymes are in an environment protected from oxidation.

How is the competition between the RF2 and $tRNA^{Sec}$-decoding molecules mediated? There is a reciprocal relationship between termination and Sec incorporation efficiencies at the *fdhF*-recoding site UGA, providing compelling evidence for competition between the canonical and non-canonical decoding events. Mansell et al. [27] have proposed a 'helical approach' mechanism for how this competition might be mediated. The competitiveness of either decoding molecule at the UGA can change according to the relative concentrations of the participating molecules. The SELB complex carrying $tRNA^{Sec}$ is bound to the apical loop of the *fdhF* stem-loop as the sequence approaches the ribosomal decoding site. If the complex is to remain bound, it must rotate about the axis of the helical stem as the secondary-structure unwinds. There is a likelihood of a ribosomal pause or translational slowing because of the increased torsional load imposed by the unwinding hairpin. The SELB complex is ideally positioned to deliver the $tRNA^{Sec}$ just as the UGA reaches the A-site and this would apparently give the $tRNA^{Sec}$ a significant advantage over the decoding RF2 to reach the inner cavern of the ribosomal active center. Indeed, this may be why the relatively efficient termination context of the UGA performs poorly against this competition. However, the creation of a translational slowing [28] by the 'helical approach' of the SELB complex may also provide the window of opportunity for the RF2 to remain relatively competitive for decoding the UGA (Fig. 10.1A).

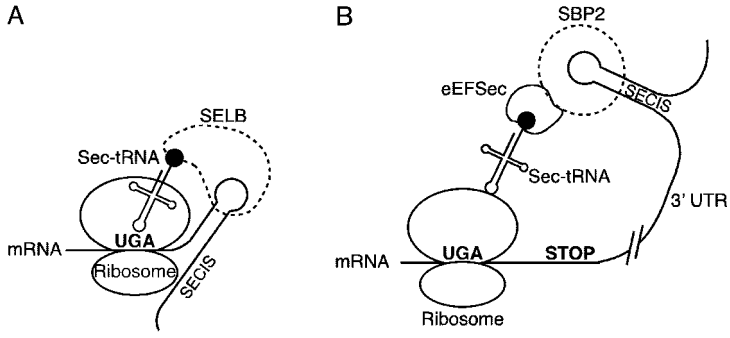


Figure 10.1 The *cis* elements and *trans* factors critical for Sec insertion during selenoprotein synthesis. The mechanism for Sec incorporation is shown for prokaryotes in (A) and for eukaryotes in (B).

10.5 Mechanism for Sec Incorporation at UGA Sites in Eukaryotic and Archaeal mRNAs

Less is known about the mechanism of Sec incorporation into eukaryotes and archaea. The SECIS elements in eukaryotes are also stem-loops but differ in structure and location from their bacterial counterparts. They are distant from the UGA site where the incorporation of Sec is to occur. Indeed, in eukaryotes, there is a minimal spacing of 60 nucleotides between the two for the SECIS element to function, but it can be as far away as several thousand bases. Although these elements are downstream of the UGA in most cases, there is one report of a SECIS element in the archaeon, *Methanococcus jannaschii* that is upstream of the UGA recoding site [29].

10.5.1 The Gene Products

Are there equivalent genes to those found in bacteria that mediate Sec incorporation into archaea and eukaryotes?

1. **SELA**, the oligomeric selenocysteine synthase has been found in archaea but not, as yet, in eukaryotes although putative homologs have been suggested [6].
2. **SELB.** The search for this specific elongation factor protein has been protracted and has resulted in several false leads. Initially, a protein, selenocysteine-binding protein 2 (SBP2), essential for Sec incorporation into rabbit reticulocyte lysate was thought to be the eukaryotic equivalent of SELB but it lacked elongation factor function [30]. Eventually, after database searches of increasing complexity to look for Sec-specific elongation factor homology from the archaea through to the eukaryotes, specific candidates for the SELB protein were identified [31]. One of these candidates, eEFSec, expressed as a recombinant protein, exhibited all the necessary and expected characteristics of the required factor being highly specific in its binding of Sec-$tRNA^{Sec}$ and associating with SPB2 to form a complex with the SECIS element during selenoprotein synthesis.
3. **SELC** was identified first as a minor serine-specific tRNA in mammals [32] and later was shown to be the $tRNA^{Sec}$ in eukaryotes [33]. It is present in all eukaryotic species examined and, almost exclusively, the gene is present as a single copy. The human tRNA can substitute for the bacterial SELC in Ser to Sec conversion. An important recognition determinant is a 13 base-pair coaxial helix involving an extended acceptor stem of 9 base pairs and a shortened T stem of 4 base pairs probably present in both bacterial and human tRNAs. Although the 9/4 arrangement rather than a 7/5 structure is somewhat controversial, the archaeal SELC can fold only into a 9/4 structure and provides an evolutionary reason for the presence of this 9/4 coaxial helix.
4. **SELD** was identified in humans by Berry and co-workers [34] and, although having a low similarity to the bacterial SELD, could complement a bacterial *selD* mutation. The bacterial protein is also functional in mammalian cells

suggesting a strong commonality of mechanism. The mouse and human enzymes themselves are selenoproteins with selenium at the active site and, therefore, may be involved in autoregulation of selenocysteine metabolism.

10.5.2 The Mechanism of Sec Incorporation at Specific UGA Stop Codons

Incorporation of Sec into a mammalian or archaeal protein during UGA decoding is dependent on the presence of specific structures in the mRNAs encoding these proteins. They can have different structures and certainly different primary sequences. Two consensus classes have been defined [35]. The first consists of a 9–11 base-pair stem separating a conserved SECIS element core at the base of the stem, from a 10–14 nucleotide loop with three adenosines at the 5′-side of it. In the second class, there are three adenosines comprising an internal bulge in the stem before it continues to a smaller 3–6 nucleotide loop at the tip.

How does the location of this SECIS element affect the mechanism, given that the decoding complex in bacteria is placed immediately following the UGA codon at the decoding site and the same mechanism is not feasible in mammals or archaea? Unlike the situation in prokaryotes, there are two proteins involved rather than simply a specific elongation factor to carry the Sec-tRNA^{Sec}. The first to be discovered was a SECIS element binding protein, SBP2 [30]. This protein binds selenoprotein mRNAs specifically, and Sec incorporation depends on its presence. It is speculated that the protein may play a role in excluding the eukaryotic release factor from the UGA site since it has homology to a yeast omnipotent termination suppressor of protein biosynthesis, SUP1. The eEFSec binds both isoforms of Sec-tRNA^{Sec} but not its serylated precursor or other tRNAs and, as well, binds GTP to show the classic characteristics of an elongation factor. Indeed, this protein also interacts with SBP2 and the two proteins function together for Sec incorporation into selenoproteins. This implies that the delivery complex consists of the SECIS element, SBP2, and the Sec-tRNA^{Sec} bound to eEFSec (Fig. 10.1B). How this spans the distance to the upstream UGA is not clear, but if there were a kinetic exclusion mechanism to prevent eRF1 from decoding the UGA as stop, the complex could position itself optimally for the decoding event. For example, if SBP2 were to make an association with the UGA before it entered the ribosomal A-site, then the complex already would be positioned for decoding and the eRF1 would be compromised. As both eEFSec and SBP2 have nuclear localization sites, it is speculated that the complex might be assembled on the mRNA in the nucleus ensuring the first round of translation is primed for Sec incorporation.

10.6 Why does Recoding Occur at Stop Signals?

Incorporation of selenocysteine at UGA stop codons could be explained simply by the presence of specific SECIS elements. However, the immediate context of the

UGA stop signal at most Sec incorporation sites influences competition by the Sec incorporation machinery. For example, the nucleotide immediately following the UGA recoding site in eukaryotic selenoprotein mRNAs is usually either pyrimidine U or C. Changing this nucleotide from C to a purine in the type I iodothyronine 5′-deiodinase mRNA decreases the amount of complete product and significantly increases the premature chain-termination product. This suggests that there is kinetic competition between termination and Sec incorporation at the site [36]. The context of the nucleotides surrounding stop codons clearly has a major influence on whether a particular stop codon is efficient and whether competing non-canonical events can occur (Fig. 10.2). If stop codons were decoded at different kinetic rates according to the context of the surrounding nucleotides, then there is opportunity for near-cognate tRNAs to be more competitive in some circumstances. The stop codon could be decoded as sense, or even facilitate a frameshift event to occur during a translational pause when the kinetics of stop signal decoding are particularly slow.

What is the evidence that upstream and downstream contexts can affect stop codon efficiency? As early as 1981, Kohli and Grosjean [37] highlighted an apparent bias in both the codon immediately prior to stop codons and in the nucleotide immediately following the stop codon in the very limited data set of gene sequences available at the time. Analysis of nucleotide bias surrounding stop codons became possible as more sequencing data emerged and, more recently, as the sequences of whole genomes have been completed. An algorithm was created to extract sequences around stop codons [38] and a TransTerm database constructed as a resource for translational signal analysis [39]. It was concluded from a study of nearly 1000 *E. coli* genes and lesser numbers of genes from other bacteria that the stop signal was actually a tetranucleotide rather than a triplet codon and, as well, there were clear upstream and downstream contextual biases [40]. This was most apparent for stop signals used in the most highly expressed genes (the top 10%) where, in addition to a limited subset of sense codons known to be used in these mRNAs, the stop signals had U following the stop codon almost exclusively. More recent analyses of larger data sets of both prokaryotic and eukaryotic genes indicate that there is a clear 'signature' comprising a sequence element that starts two codons before the stop codon itself and extends beyond for another six nucleotides or so [41]. This strongly hinted that there may be a hierarchy of stop signals of varying decoding efficiencies and that this may exert a significant influence on whether a recoding event could occur.

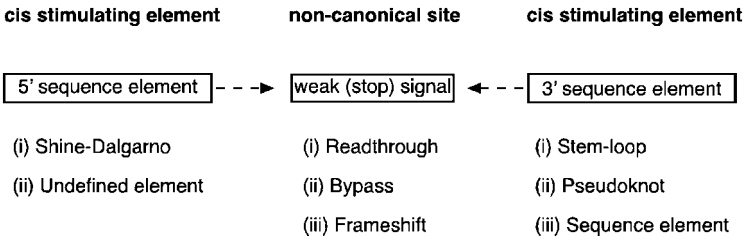


Figure 10.2 Sequence elements that influence recoding sites.

10.6.1 The Stop Signal of Prokaryotic Genomes – Engineered for High Efficiency Decoding?

Important to study this question, has been the availability of the complete genome sequence of two *E. coli* strains that had been separated by approximately 4.5 million years of evolution. The 0157:H7 pathogenic strain has acquired or retained 25% more genes than strain K12 and each has acquired or retained unique genes not found in the other strain, with 1387 genes unique to 0157:H7 and 528 unique to K12 [42]. Nevertheless, the bias of the molecular signature at the stop codons in the genes from these two strains was highly similar for TAA and TGA. In addition, the signatures were somewhat different for each of the three codons. Moreover, when this bias was analyzed to determine which individual nucleotides contributed, there was a high degree of concordance between the genes from the two strains despite there being greater then 75 000 polymorphisms in the 'homologous backbone' of the DNA sequences between the two strains. These data provide further evidence that there is a preferred sequence element for high efficiency stop signal decoding at least for UAA and UGA, and, by inference, the occurrence of non-canonical error events is minimized.

Sequence biases upstream and downstream from stop codons must reflect different mechanisms. The upstream sequences in the mRNA (the last two codons) are already involved in decoding events through ribosomal P- and E-site interactions with tRNAs and, therefore, their influence on RF stop-codon decoding must be indirect. On the other hand, the downstream sequences are not involved in mRNA–rRNA interactions according to current understanding and would be available to make direct interactions with the decoding RF protein itself. The fact that the stop signal is decoded by a protein rather than a tRNA as for sense codons, means that there is no intrinsic reason why more nucleotides than just the triplet codon might make contact with the RF. Indeed, with this in mind, zero-length site-directed crosslinking from specific positions in the mRNA to the *E. coli* RF protein using ^{4}thio-U instead of U in the RNA sequence has been carried out. Crosslinks were obtained from the first position of the stop codon (+1) and the three positions following the stop codon (+4 to +6) but not beyond, suggesting close physical contact between these nucleotide positions and the protein factor [43]. These data support the concept of direct interaction between the RF and stop signal that extends beyond just the three nucleotides from the stop codon.

What might be occurring upstream of the stop codon? Here, there are two features that are created by the tRNA-decoding events. First, the two tRNAs in the P- and E-site positions have specified the ultimate and penultimate amino acids of the completed polypeptide positioned at the peptidyltransferase center and at the beginning of the ribosomal exit tunnel, respectively. Secondly, the tRNAs themselves have a common three-dimensional shape but have micro stereochemical detail in the bases and modifications in each position of their sequence. Therefore, a particular tRNA in the P-site can create a three-dimensional environment against which the decoding RF, spanning between the decoding center and the peptidyltransferase center in the

ribosomal A-site, has the potential to make contact. In this way, both the amino acids specified and the tRNAs themselves may influence the stop signal decoding rate by the RF. This protein is occupying the binding site (A-site) that has been carefully crafted for an aminoacyl-tRNA and, by analogy, is like a cuckoo in the nest of another bird. Just as the cuckoo can be too big for the nest, so the RF may be constrained in the ribosomal A-site. If this were the case, then the upstream sequences coding for specific tRNAs and amino acids may contribute to a three-dimensional binding site for the RF that results in altered kinetics during stop codon decoding.

To determine whether the C-terminal amino acid of a protein might be having an effect on the efficiency of stop signal decoding, it is possible to examine whether a particular amino acid is abundant and whether its occurrence is highly biased. It is interesting that the two *E. coli* strains have similar proportions of each amino acid at the C-terminal positions of their proteins. The abundance and bias in the use of amino acids at this position was highly similar over all genes within the two strains and particularly within the TAA and TGA terminating genes. There were global trends in bias both for and against amino acids with certain characteristics that were still evident when the abundance of the amino acids at the C-terminal position were analyzed. What was significant from this study was that the biases and abundances of the amino acids were not identical for each of the specific stop codons, suggesting that the biases were related to the termination phase of protein synthesis and not simply some other unrelated translational process.

The conclusion from these analyses is that the last amino acid of the protein may have some stereochemical or charge-related effect on the efficiency of RF-mediated stop signal decoding when this amino acid is positioned at the ribosomal peptidyl-transferase center through the P-site tRNA. Although the stop codon is decoded by the RF positioned at the A-site in the decoding centre, for successful termination to occur a signal must be transmitted through the RF structure to the peptidyltransferase center so that release of the polypeptide can occur [44]. Further studies are needed to investigate this possibility.

What is the situation with the penultimate amino acid that was brought to the ribosome by the tRNA now positioned at the E-site? It is difficult to understand a direct effect of this amino acid on termination efficiency as it is likely to be entering the exit tunnel and less likely, therefore, to affect the RF stereochemically. However, there may be an indirect influence on the stereochemical position of the C-terminal residue by the preceding amino acid through its interactions with the surrounding ribosomal architecture and this could explain an influence on stop signal decoding efficiency. The analysis has shown that fewer residues show bias at this position than at the ultimate position but there were still preferences for amino acids with certain characteristics [42]. The picture that has emerged from these studies is that there is a much weaker influence on termination efficiency from the penultimate amino acid than the last amino acid, but still with a suggestion of some indirect influence yet to be understood.

Bias in the two codons upstream of the stop codon may not necessarily reflect the amino acid at all or solely, but may be a more direct effect of the tRNAs that are bound into the two ribosomal sites and their interactions with other RNAs. As a

particular amino acid can be carried by different tRNA isoacceptors often recognizing different codons (e.g., Leu_1 CUG; Leu_2 CUC/U; Leu_3 CUA/G; Leu_4 UUG; and Leu_5 UUA), the codon abundance in the last position before the stop codon has been analyzed in detail. This has allowed a determination of whether there is a subset of tRNAs in the termination complex during stop signal decoding. As for analysis of amino acid frequency, both abundance and bias of codons are important criteria and may not necessarily correlate. There were both specific codons preferred and specific codons that were rare in the last position. These were different from the biases found in the penultimate codon but, as with the amino acid analysis, effects in this position were much less marked and gave a relatively weak signature.

What do these trends in codon use at the last amino acid position mean for the selection of specific tRNAs? There were several striking consequences. All but one of the codons common in this position were decoded by only one isoacceptor tRNA species, whereas codons selected against were often decoded by several isoacceptor tRNAs. There were specific sequence characteristics of the abundant tRNAs and some of these related to particular stop codons. For example, before UAA, the most common four codons were decoded by only two tRNAs that were two (of only three) tRNAs with a modified mnm^5s^2U at the anticodon wobble position 34. These modified tRNAs were also used abundantly before the other stop codons as well. This suggests that the ultimate tRNA is contributing to the efficiency of stop signal decoding and may reflect, as indicated above, its contribution to the binding site architecture for the decoding RF in the A-site or the maintenance of stable peptidyl–tRNA interactions. Structurally, the tRNA would line one side of the space that the RF occupies during decoding and there is a potential for interaction between these two macromolecules. Indeed, site-directed crosslinks from the elbow of the tRNA (position 8) and the anticodon loop (position 32) to the RF has been achieved with zero-length crosslink moieties in the tRNA suggesting that there is very close contact at several positions [45].

Is there any evidence that bioinformatic analyses have revealed important features of the stop signal that are physiologically important for both non-canonical recoding and canonical decoding in protein synthesis? First, classic recoding sites where stop signals are involved do have contexts that are rarely found at natural termination sites of genes. The best example is the frameshift site in the bacterial RF2 gene where the downstream context UGACUA is found in only three other genes. This context is the weakest of all 64 UGANNN sequences tested at this site *in vivo* allowing the non-canonical recoding event to occur in approximately 90% of ribosomal passages. In contrast, the strongest context of the 64 possibilities was UGAUUA with substitution of U for C in the +4 position and consistent with the predictions from bioinformatic analysis. Similarly, the downstream context UGACAC at the selenocysteine incorporation site in the *fdhF* gene was also shown to contribute to a relatively weak termination signal. At each of the three stop codons, the data from the 64 contexts of the +4 to +6 bases gave a hierarchy of signal efficiencies with the +4 base highly influential. This supports the original suggestion that the termination 'codon' should be thought of as comprising four and not three bases [39].

Indeed, in eukaryotic *in vitro* experiments, the RF has been shown to respond to a minimum of a four-base stop signal [46].

Does the termination signal efficiency of particular downstream contexts determined experimentally correlate with signal abundance in *E. coli* genes? For TGA contexts, signal strength correlated well with signal abundance and six-base bias, implying the most efficient signals are those that are most frequently used and that use of inefficient signals is avoided except at recoding sites. However, the minor set of TAG signals in *E. coli* continue to be an enigma as there was a range of decoding efficiencies (Fig. 10.3) but with a negative correlation between abundance and bias with efficiency.

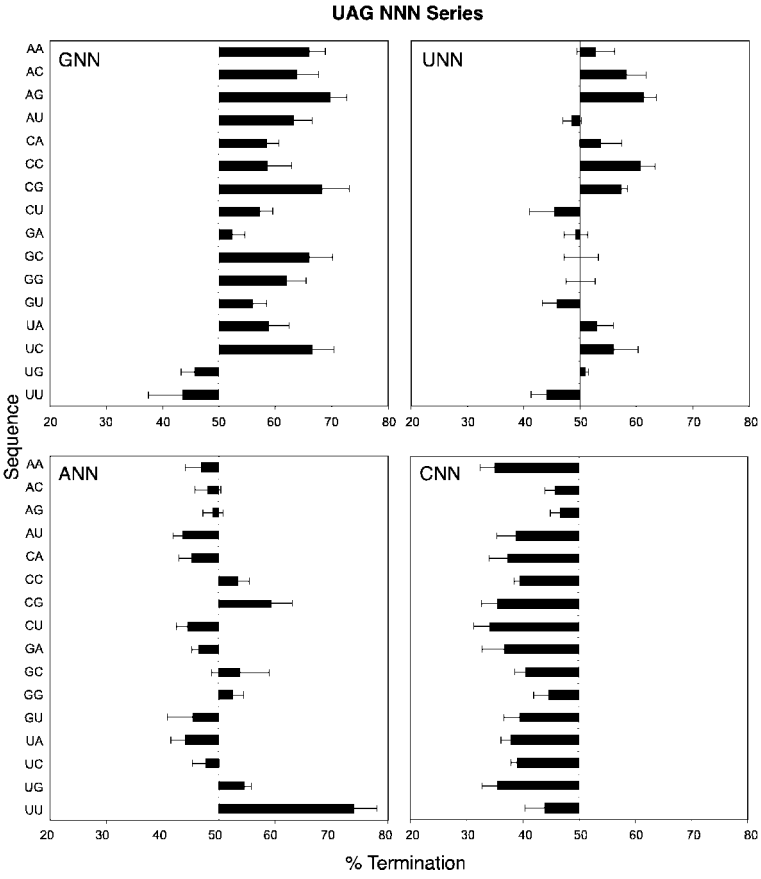


Figure 10.3 Termination efficiencies of UAG-NNN signals in competition with frameshifting. The termination efficiencies were measured as described in Ref. [78]. Efficiencies are shown relative to a point of equal competition (50% termination midline) with signals divided into graphs according to the identity of the +4 base immediately following the UAG codon. The majority of signals with a +4 G support > 50% termination, signals with a +4 C all gave < 50% termination and signals with a +4 U or A fall on either side of this midline. The error bars are the S.E.M. for at least six determinations of each signal.

Apart from suggestions from the bioinformatic analyses described above, earlier experimental work had shown effects of the last two sense codons on termination signal strength at selected contexts, as measured by the failure of the signal to specify stop but instead allow incorporation of an amino acid through near-cognate decoding [47–49]. More recently, termination signals spanning 12 nucleotides comprising the last two sense codons, the stop codon, and the three nucleotides following, were constructed to reflect predicted weak, strong and hybrid (strong upstream and weak downstream, and vice versa) signals (Fig. 10.4) [42]. Again, for UGA signals it was possible to predict correctly which would be strong and which would be weak (those that allow significant stop codon readthrough) with upstream and downstream sequences acting co-operatively. This became more obvious when tested in *E. coli* strains carrying suppressor tRNAs where competition was stronger with a cognate tRNA present. With UAA signals, competition from suppressor tRNAs was sufficient to reveal relative strengths of the stop signals only in the suppressor strain. With the enigmatic UAG signals, the 5′ contexts behaved according to prediction, but the effects of 3′ contexts did not correlate with bias or abundance.

These studies clearly indicate that context both upstream and downstream of the stop codon in bacteria has effects on the efficiency of stop codon decoding. Certain contexts not frequently found with UAA and UGA stop codons are assumed to increase the translational pause at the A-site stop and allow for competing events such as near-cognate decoding or translational frameshifting to occur. They provide the capacity for a recoding site to evolve at a stop codon allowing non-canonical events to occur during protein synthesis, thus recruiting additional complexity to the regulation of gene expression. This is clearly the case at the frameshift site of the *prfB* gene encoding RF2 [50–52] and at the Sec incorporation site of the *fdhF* gene [27]. In both cases, the balance of the canonical and non-canonical events can be significantly altered *in vivo* by changing the stop codon sequence context despite not altering the specific *cis* elements that favor the non-canonical event. Physiologically, it is events *in trans* that alter competition such as the concentration of the decoding RF in the case of the frameshift event [52] and the relative concentrations of the two decoding molecules RF and Sec-tRNA^{Sec} in the case of Sec incorporation [27]. On the other hand, such competitions are precluded at most stop signals found at the ends of the coding

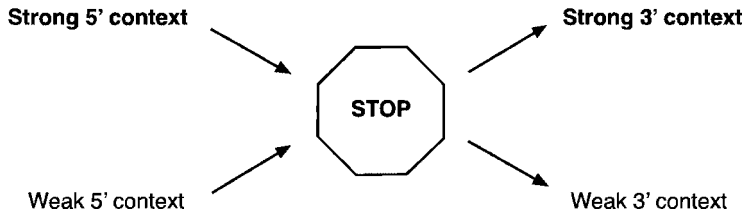


Figure 10.4 The strategy for testing the strength of termination signals spanning 12 nucleotides. Constructs were designed that contained TAA, TGA, or TAG with a predicted strong or weak sequence element 5' or 3' to the stop codon to give 'strong', 'weak' and hybrid signals.

regions of mRNAs because the context makes the stop signal so competitive for termination against non-canonical events that recoding is insignificant.

10.6.2
The Stop Signal of Eukaryotic Genomes – Diversity Contributes to Recoding

When nucleotide bias is examined in the non-redundent cDNA sequences from a number of eukaryotic genomes such as *Saccharomyces cerevisiae*, *Drosophila melanogaster, Caenorhabditis elegans*, or *Homo sapiens*, a characteristic pattern is seen [53]. A scan towards the stop signal reveals a characteristic increase in bias a few nucleotides before the codon as is seen for prokaryotic genomes and following the stop codon there is a gradual decrease of bias for up to nine nucleotides downstream (Figure 10.5A). This is the classic signature of a sequence element, with nucleotides upstream and downstream from the stop codon having the potential to contribute to the strength of the signal. When the nucleotides contributing to the bias are analysed, the pattern is similar amongst the various eukaryotic genomes but different from those of the *E. coli* strains. For example, where U (especially) and G in the +4 position are highly abundant and contribute to highly efficient signals in bacteria, in eukaryotes, the purines G (predominately) and A are favored (Figure 10.5B).

Highly expressed genes in these eukaryotic genomes have been classified by analysis of two-dimensional protein gels and through the use of the codon adaption index (CAI; a measure of the codon subset used within the gene and strongly correlated with level of expression) together with a number of other criteria. The genes identified as highly expressed have a more significant bias in the nucleotides

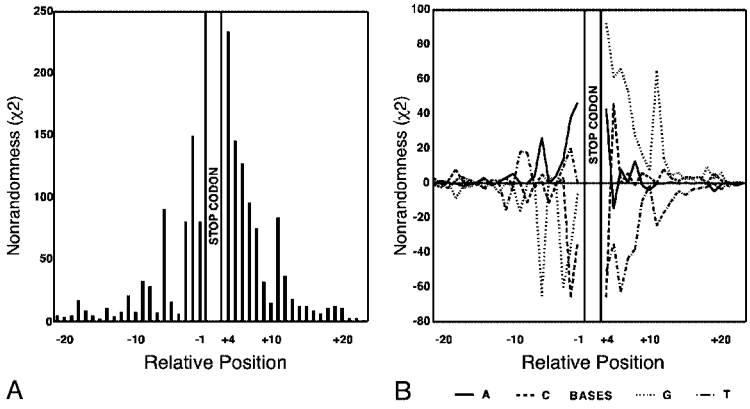


Figure 10.5 A statistical analysis of the nucleotide bias surrounding stop codons in the *Saccharomyces cerevisiae* genome. The bias (nonrandomness) was determined (A, B) by calculating the χ^2 value [(observed-expected)^2/expected]. Bias is shown for nucleotide position (A) and in the individual bases contributing to bias at each position (B).

around the stop codon compared with those genes that are expressed at lower levels. This indicates that there is some translational advantage to the bias in the sequence element specifying termination. The bias and abundance of the amino acid found in the C-terminal position and the particular codon used is less marked than in prokaryotic genomes although there are residues that are still over-represented (His, Lys) and under-represented (Gly, Pro). There are also codon biases, for example, the lysine codon AAA is over-represented with respect to AAG in the yeast genome. This may relate to the fact that the tRNA for AAA is the hypermodified $tRNA_2^{Lys}$, which may improve stability of peptidyl–tRNA interaction at the P-site during stop signal decoding with a corresponding decrease in the rate of readthrough at the following stop codon. Various features of a tRNA contribute to an ability to suppress a stop codon and read it as sense. Suppressor tRNAs, apart from $tRNA^{Sec}$, are normal cellular tRNAs with a primary role to decode cognate sense codons but also they can decode stop codons that are near-cognate. There must be some enhanced codon:anticodon stability in the near-cognate interaction perhaps through unconventional base-pairings. Supporting this, is the extent of tRNA modification in or 3′ to the anticodon important to enhance or depress readthrough rate [54]. Clearly, an unmodified tRNA coupled with a weak termination context would most probably enhance stop-codon suppression and allow readthrough to compete with the termination event.

Bioinformatic analysis of the eukaryotic genomes has enabled prediction of strong and weak stop signal elements. Strong elements favor penultimate and ultimate codons that use tRNAs modified in their anticodons, coupled with G in the +1 position immediately 3′ to the stop codon. In contrast, weak elements favor penultimate and ultimate tRNAs that are unmodified in their anticodons and a stop signal with T or C in the +1 position. These predicted weak signals are similar to the sequences found in viral recoding sites [53]. Experiments in mammalian *in vitro* and *in vivo* systems have confirmed that sequences over-represented upstream and downstream from the stop codon support only low levels of readthrough compared with that supported by under-represented sequences. Sequences found at viral recoding sites supported elevated levels of readthrough in the test system and the patterns were similar for all three stop codons and in contrast with that found at prokaryotic stop codons. This can be explained by the fact that the eukaryotic RF, eRF1, recognizes all three stop codons in contrast with the two bacterial factors that each recognize UAA but are then specific for either UAG or UGA.

Another recent study using mammalian cells *in vivo* showed that the –1 base (the base immediately prior to the stop codon) of a UAG stop signal element influenced recoding at the site mediated either through an effect from the characteristics of the P-site tRNA or interactions of the anticodon with the nucleotide [55]. In yeast, termination efficiency can vary significantly according to context and specific sequence motifs supporting readthrough have been identified [56]. Rousset and co-workers [57] looking for 'weak termination contexts', found readthrough motifs in eight yeast open reading frames within the genome. Studies with one of these, *PDE2* encoding a cAMP phosphodiesterase, found a 20-fold difference in readthrough depending on

whether the strain was [*PSI*$^-$] (an epigenetic element resulting in decreased accuracy of translation termination) or [*psi*$^-$] (normal strain) and the extended protein had lower stability.

Just as is the case with prokaryotes, in eukaryotes, the sequences both upstream and downstream from the stop codon can have a profound influence as to how efficiently the codon signals stop. The nature of the sequence element opens up opportunities for recoding to occur at the stop codons and provides a subtle layer of gene regulation in specific circumstances where defined amounts of a protein are required.

10.7
Readthrough of a Stop Signal: Decoding Stop as Sense

The fact that the sequence context of the stop codon can produce a hierarchy of stop signals of varying efficiencies gives the potential for subtle redefinition of the signal so as to regulate amounts of a protein dependent on a readthrough event or for provision of a balanced ratio of two proteins from the same mRNA sequence. The extreme case is where the codon itself is defined for two purposes as with Sec incorporation at specific UGA codons discussed above. In this case, there is a cognate species for both events rather than a near-cognate competitor. However, it requires extra sequence elements for one of the cognate decoding species, the Sec-tRNA^{Sec}, to be competitive at the site. Indeed, Sec incorporation could be regarded as a canonical event and not as a recoding event where the UGA codon is reclassified in the same genetic system for another purpose. Rather than modifying one of 64 codons to provide a specific codon for Sec with a modified base to provide unique structure, one of the existing codons is utilized. Given that the origin of Sec may have been ancient and its presence particularly relevant as a catalytic residue when oxygen was not such a dominant physiological molecule, then UGA may have originally encoded Sec but subsequently has been captured as a stop codon. It is interesting that in other genetic systems where a stop codon is used to specify an amino acid, for example, UGA for Trp in mammalian mitochondria, there is no duality of meaning for the codon. In mitochondria the definition of the codon as stop has most probably been lost during evolution.

Readthrough of the UGA stop codon at the Sec insertion site in the *fdhF* mRNA is prevented even in the absence of selenium [58] and this depends on sequences upstream and downstream from the UGA. This could be interpreted either as a special protective element or as the sequence element of the stop signal now simply out-competing near-cognate events. The natural sequences were better at preventing readthrough than alternatives tested. This context may have evolved so that competition for decoding was only between Sec incorporation and termination, excluding the third possible event of another amino acid being incorporated at the site. This would be potentially possible if the decoding rate by the competing RF was slowed. However, such a protein would be non-functional without the key Sec residue at the active site thereby providing a rationale to exclude this possibility.

Most cases of stop codon readthrough would involve near-cognate tRNAs that become competitive either because of the sequence of the stop signal or because there are other *cis* elements that favor the near-cognate over the cognate event. Given the potential for readthrough to be used as a mechanism of regulating gene expression and creating more diversity in the proteome, it is surprising that more examples have not been found. To date, there are a small number of examples where readthrough seems to be important and these span viruses, bacteria, and eukaryotes. Readthrough occurs at UGA and UAG stop codons with the most common amino acids incorporated, Trp and Gln, respectively. Although the efficiency of the event is relatively low (1–10%) in competition with termination, this is still 100- to 1000-fold above the error rate. The *cis* elements that influence readthrough can be well beyond the boundaries of the stop signal. For example, similar to the SECIS element in eukaryotes for Sec incorporation, an element several hundred nucleotides downstream from the stop codon influences readthrough at a specific stop codon in barley yellow dwarf virus [59]. In addition, secondary-structural elements are also important. A classic example is found in the synthesis of the murine leukemia virus gag-pol precursor protein, where a pseudoknot is an important mediator of the recoding event [60]. On the other hand, other viral examples do not seem to have *cis* elements beyond the immediate environment of the stop codon that is under recoding pressure. For UGA recoding sites in Sindbis virus [61], and in an *E. coli* bacteriophage RNA Qβ [62] only the +4 base of the stop signal seems critical. In both cases, the nucleotide following the stop codon is C contributing significantly to a poor context for termination. These situations would be classic candidates for readthrough based on what we know now about the stop signal.

Atkins and co-workers [63] have examined 91 unique viral sequences where readthrough of stop signals is known to occur. It is of interest that 90% had one of six tri-nucleotide sequences downstream from the site (out of the possible 64). The authors make the point that the identity restriction of six nucleotides following the stop codon is remarkable given they come from RNA viruses where mutation rates are high. In other words, there has been strong pressure to retain the contexts. While readthrough may reflect the strength of the stop signal that these contexts create and the rate of decoding by the RF, in the case of RNA viruses where evolution of optimum sequences is likely to occur quite rapidly, other equally or more important features may also have evolved. These may allow for an enhanced rate of aminoacyl-tRNA binding during near-cognate decoding (perhaps mediated by non-canonical interactions between particular context sequences of the stop-codon mRNA and the rRNA in the environment of the A-site) apart from any secondary structural elements that may be enhancing these effects. Recoding in these circumstances utilizes a stop codon as the marker for the site and in most cases a stop signal that is decoded more slowly by the RF so that there is a favorable site for building a recoding signal with the addition of other sophisticated and specific elements in *cis* or in *trans*.

The examples in prokaryotic and eukaryotic genomes where readthrough appears to be important are generally less well studied but are potentially very interesting. For example, in *D. melanogaster,* at least three genes seem to be regulated via stop codon

readthrough. Readthrough at a UGA in the *kel* gene is regulated both in a tissue and developmentally specific manner with maximal readthrough during metamorphosis [64]. A topoisomerase gene in *Bacillus firmus* is the only documented bacterial gene not derived from bacteriophages that is supposed to use readthrough as a means of regulation [65].

10.8 Bypassing of a Stop Codon: 'Free-wheeling' on the mRNA

In the last decade or so, the more RNA is studied the more remarkable mechanisms are discovered associated with its biology. Many of these could not have been anticipated or dreamt by scientists as possibilities. One such recoding event is stopcodon bypassing, where a section within the mRNA coding region is missed out and recoding starts again further down the mRNA. The classic example is for the bacteriophage T4 *gene 60*, a topoisomerase subunit gene, where 50 nucleotides are omitted from the decoding process for the correct full-length protein to be produced [66]. In this case, the flanking codons are matching GGAs decoded by $tRNA^{Gly2}$ and the next codon to be decoded after the first GGA is a stop signal with a weak downstream context (UAGCCU). In this case, the choice of events is for the peptidyl-tRNA in the P-site to detach from the mRNA before the UAG in the A-site is decoded and synthesis terminated. This detachment (called 'take-off' by Atkins and colleagues) occurs with very high efficiency and under physiological conditions it appears that termination is out-competed. There are additional *cis* and *trans* elements that drive the event that includes in addition to detachment of the peptidyl-tRNA, scanning of the mRNA and re-attachment (or 'landing' according to the Atkins' nomenclature) where canonical protein synthesis resumes.

As has become the familiar pattern at recoding sites, the stop codon forms the basic platform for the event and other sophisticated elements have been put in place to ensure its efficiency. The detachment site and the stop codon are within the stem of a stem-loop secondary structure but, intriguingly, a sequence of charged and hydrophobic amino acids in the nascent peptide synthesized up to this point also acts as mediator of the event [67]. Clearly, to initiate the event there must be competition between detachment and termination. The elements favoring detachment overwhelm stop codon decoding by RF1, the cognate decoder of UAG, as most ribosomes initiate bypassing. A detailed study has been undertaken to try to assign the importance of the various *cis* and *trans* elements in the three stages of the event; detachment, scanning and landing [68]. Structures of the bacterial ribosome suggest the peptidyl-tRNA is held at the P-site with a number of interactions between it and rRNA in the vicinity of the decoding site where codon–anticodon interaction is occurring [69]. This is in contrast with the A-site tRNA, where there is a paucity of apparent interactions near the site of codon–anticodon interaction. Slippage of the peptidyl-tRNA on the mRNA involving both detachment and movement to prevent re-attachment must require some significant disruption to these normal contacts between tRNA and rRNA. It is assumed that the decoding rate of UAG by RF1 is sufficiently slowed so that the force for detachment can predominate. Indeed, even with

a temperature-sensitive RF1 strain it was not possible to demonstrate competition between termination and detachment because of the overwhelming advantage of the recoding event and it was not until functional RF1 was over-expressed that a decrease in detachment of the peptidyl-tRNA was observed [68]. Presumably, this was because the local concentration of RF1 at the recoding site was higher, the rate of decoding the stop codon was enhanced and, therefore, the kinetic pause was shortened.

This study supports a simple pause model at the termination codon for providing the right conditions for 'take off' and modest stability of the peptidyl-tRNA interaction at the P-site. The current model for RF-mediated release of the completed polypeptide from the P-site tRNA invokes a conformational change in the RF with altered ribosomal interactions and positioning in the A-site after successful cognate decoding of the stop codon. This is proposed to trigger a signal to the peptidyltransferase center that initiates hydrolysis of the ester bond between the ultimate tRNA and growing polypeptide [70, 44]. If such a conformational change is prevented or altered following successful decoding at the recoding site, this might also favor detachment of the peptidyl-tRNA before hydrolysis can occur. However, as high concentrations of RF1 can compete successfully, the bypassing elements can, at most, decrease the likelihood of a successful signal being triggered to the peptidyltransferase center by the RF. On the other hand, the initial binding rate of RF1 to the A-site might be decreased significantly leaving the A-site empty and an empty A-site might be required for detachment.

The role of the *cis*-acting stem-loop can be compensated for by the removal of ribosomal protein L9. This suggests L9 may have a role in preventing slippage of the peptidyl-tRNA at the P-site, or a role in A-site decoding. However, recent data suggest defects in L9 may enhance mRNA movement through the ribosome [71]. Therefore, the ability of ribosomes lacking L9 to complement mutations in the stem-loop might be through this mechanism. On the other hand, the loss in bypass efficiency by mutant $tRNA^{Gly2}$ can be compensated only by mutations in the nascent *gene 60* protein before the site. This suggests the two *trans* elements, the tRNA and the nascent peptide, are operating through different mechanisms. The data support the nascent peptide either enhancing the peptidyl-tRNA dissociation and the stem-loop occupancy of the A-site, or indirectly enhancing movement of a dissociated peptidyl-tRNA. Presumably, all of the elements will be impinging on the ribosomal architecture and interactions in different ways to loosen structurally mediated tight controls on frame maintenance that are important for canonical decoding events during translation.

The mechanism of bypassing seems bizarre to contradict all reason as a logical mechanism of making a viable protein but this is characteristic of RNA. That these highly unusual mechanisms are present, is an important example of how a complicated RNA machine involving three types of RNA can evolve even more sophisticated processes beyond the extremely intricate procedures required for normal canonical events.

10.9 Frameshifting Around Stop or Sense Codons

Frameshifting involves a disruption and slippage of the interactions between mRNA and tRNA. There is forward (+) or backward (–) movement of the mRNA with respect to tRNA anticodon:codon binding. However, with frameshifting, the movement is of only one or two bases before re-engagement of the tRNA and disruption of the original reading frame is a consequence of this slippage. This is in contrast with 'bypassing' mRNA, where a greater length of the mRNA is avoided with respect to the tRNA. Whereas disruption of the reading frame can occur here as well, translation still has the potential to resume in the same frame as the prior synthesis. Gallant and Lindsay [72] have shown that ribosomes can slide over 'hungry codons' (described as such where an aminoacyl-tRNA is limiting) and then resume translation at a cognate codon many nucleotides downstream similar to the classical bypass event shown with bacteriophage T4 *gene 60*. Slippage over 'hungry codons' is not 'programed' but occurs in the mRNA under a special set of physiological circumstances [72]. In both frameshifting and bypassing, the existing interactions between tRNA and mRNA and possibly also rRNA are perturbed in the initial event. The difference is in the re-engagement process. It occurs almost immediately in the case of frameshifting and is mostly dictated by the particular mRNA sequence and the opportunity for the tRNA to re-engage in the new frame through anticodon:codon interactions. In this way, frameshifting is an example of a programed translational event.

Thus far, discussion of recoding in this chapter has focussed on sites that have stop codons as an essential framework (readthrough, Sec incorporation, bypassing). In the case of frameshifting, + or – slippage events are found not only at stop codons but also at regions of the mRNA where a stop codon is not present. Clearly, at these sites the advantage of having a slowly decoded stop codon to allow for kinetic competition at the translational pause is not needed because there are alternative ways in which the pause is created at a sense codon or kinetic competition sufficiently favors the non-canonical event. Non-programed frameshifting can occur naturally at sense codons but at a very low frequency (less than 5×10^5 per codon) [73].

Farabaugh and co-workers [74] have described how frameshifting can disrupt the reading frame by a number of possible mechanisms such as:

1. Expansion or contraction of codon size to four or two bases instead of three (this would not require disengagement of existing interactions).
2. Orientation of the incoming tRNA to allow recognition of three bases but not the next three that are in-frame.
3. Translocation of the mRNA after peptide-bond formation to move the mRNA two or four bases forward instead of three.
4. Disruption of existing RNA–RNA interactions to facilitate slippage after translocation.

Debate on which of these mechanisms operate physiologically has been lively, but there is no necessity to explain all frameshift events by a universal mechanism.

For example, a mutant $tRNA^{Gly}$ is able to correct frameshift mutations [75] and a possible mechanism is the recognition of GGGN rather than GGN. However, it is hard to accommodate this model with the recent structural information on how the rRNA elegantly 'senses' each of the three bases of the correct codon in the ribosomal A-site [10].

10.9.1 Forward Frameshifting: the +1 Event

The first discovery of frameshifting in a cellular gene initially appeared to be a 'one-off' discovery because it was an exquisite example of how a specific gene could regulate its own synthesis by a unique mechanism. The gene was *prfB* encoding the bacterial RF2 that recognizes the UGA stop codon. The frameshift site within the RF2 mRNA contained a UGA stop codon, the very codon recognized and decoded by RF2 as stop. This immediately suggested that here was a unique mechanism by which RF2 could control its own synthesis. The stop codon is in-frame at position 26 and it was discovered from sequencing of the first 44 amino acids of the protein that there had to be a +1 frameshift event at the stop codon during translation to obtain the derived sequence of the full-length protein [50]. This was a programed event in that while the stop codon provided the framework of the site, there were other *cis* elements that facilitated the event. These elements included the codon immediately before the stop codon, CUU, that can detach from the anticodon of the peptidyl-tRNA in the P-site allowing it to re-pair in the +1 frame with UUU, comprising the last two bases and the first base of the stop codon UGA. When this occurs, the codon in the new +1 frame A-site is GAC allowing for Asp to be incorporated as the next amino acid.

Not only is translational frameshifting conserved in the *prfB* genes from a wide range of bacteria (approximately 70% of those sequenced so far) but the CUUUGA motif is also retained implying a conservation of mechanism [76]. Just 5' of the CUUUGA motif, an internal Shine–Dalgarno sequence normally found 5–6 bases in front of start codons in bacterial genes, base-pairs with a complementary region of the 3′-terminus of the 16S rRNA to facilitate the frameshift event [77]. As well, the stop signal contains the least efficient downstream sequence, UGACUA, of any found in bacteria [78, 42]. This implies that stress is placed on interactions between the mRNA, tRNA and rRNA during decoding of the CUU. At the same time as the 'misplaced' Shine–Dalgarno interaction is occurring, a weak stop signal is present in the A-site giving a decoding pause and allowing time for the tRNA to disengage from the peptidyl-tRNA and re-engage with the next base in the mRNA.

The *trans* element for frameshifting in the RF2 mRNA is, of course, the RF2 protein itself (Fig. 10.6). Modulation of its concentration can shift frameshift efficiency from 0–100% [51, 52, 79], illustrating how effective the mechanism can be to regulate the cellular concentration of RF2. Under normal physiological conditions of bacteria growing in rich media in log phase, the frameshift efficiency is approximately 30–50%. Frameshifting will also occur if the stop codon is replaced by a sense codon but then the efficiency is inversely related to the rate of amino-

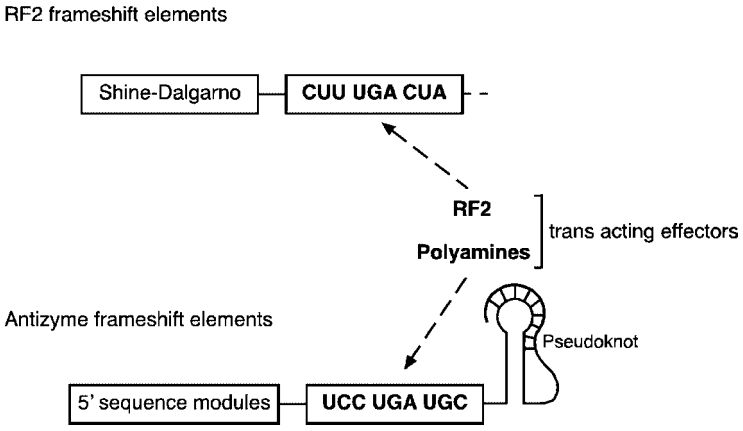


Figure 10.6 The cis elements and trans-acting effectors critical for +1 frameshifting at the RF2 and antizyme recoding sites.

acyl-tRNA selection and supports the importance of the translational pause [80]. Perhaps a U:G wobble base pair in the pre-shift codon:anticodon that has been associated with high-frequency frameshifting facilitates the event at the RF2 site [81].

A similar +1 programed frameshifting event with a novel autoregulatory mechanism was discovered some 10 years later in the mammalian ornithine decarboxylase antizyme gene [82]. Antizyme synthesis is induced by polyamines (see Fig. 10.6). Antizyme binds to and induces a conformational change in the key enzyme in polyamine synthesis, ornithine decarboxylase, targeting it for turnover. This was another example of a remarkable autoregulatory circuit where the key molecule was not a protein this time but the polyamine metabolites, spermidine, spermine and putrescine. Polyamines are known to interact with RNA and probably influence key interactions at the decoding site that facilitate the recoding event. The recoding site in the antizyme mRNA is also at the codon preceding a stop codon (UCCUGAU). Key *cis* elements are responsible for enhancing frameshift efficiency; at least 50 nucleotides in modules just 5′ of the frameshift site are responsible for 2- to 3-fold stimulation of frameshift efficiency, the stop codon increases efficiency 15- to 20-fold and there is a downstream pseudoknot just three nucleotides distant from the motif that contributes 2.5- to 5-fold. The upstream sequences may be equivalent in some way to the Shine–Dalgarno element of the *prfB* site in function. In contrast with this site, whereas findings from site-directed mutagenesis studies suggested that re-pairing between the tRNA and the +1 shift CCU was unlikely, phylogenetic analysis and observations of –2 frameshifting at the site in yeast are supportive of a re-pairing mechanism [83].

Two mammalian paralogues of what is now called antizyme 1 have been found [84] each with a frameshift site, with antizyme 3 being tissue and cell-type specific. Atkins and co-workers [83] have studied these antizyme sites in a wide range of organisms and this has enabled them to make several conclusions about recoding

events. First, recoding sites can be selected over very long evolutionary periods. Secondly, some plasticity is characteristic of the site and, thirdly, once an initial site is established a range of stimulators acting in *cis* can evolve to give specific characteristics and optimize recoding efficiencies.

An example of a mechanism for + 1 frameshifting, which is different from that of tRNA slippage coupled with slow decoding of the codon in the A-site, seems to occur with the yeast retrotransposon Ty3 [73]. In this case, frameshifting occurs at the sequence GCGAGUU as a result of the incoming tRNA pairing with the out-of-frame GUU without slippage of the peptidyl-tRNA. Interestingly, over-expressed near-cognate P-site tRNAs were able to induce frameshifting generally at the Ty3 site and is in contrast with cognate tRNAs that decreased frameshift efficiency [84]. This suggests that near-cognate tRNAs may be decoding GCG in the Ty3 site and, when coupled with a slowly decoded codon like AGU, allow take-over by the tRNA recognizing the +1 codon, GUU.

Yet another example of +1 frameshifting has been found in the yeast telomerase gene, *EST3*. Yeast use telomerase to maintain the ends of their chromosomes and the +1 translational frameshift event required to produce fully functional telomerase occurs at a motif, CUUAGUUGAG [85]. This motif is similar to the Ty1, Ty2 and Ty4 frameshift site, CUU AGG C, in the underlined portion but, as well, contains a stop codon. These additional examples indicate there may be more sites yet to be found where + 1 frameshifting plays an important part in gene regulation.

10.9.2
Programed –1 Frameshifting: A Common Mechanism used by Many Viruses During Gene Expression

Diverse virus groups have evolved a recoding site to express coat proteins and enzymes in a carefully balanced ratio. The mechanism involves switching frame during the synthesis of a polyprotein (gag) so that the extended product (gag:pol) from the new frame contains the enzyme sequences that can be excised from the protein. For most passages (approximately 90%), the ribosome does not frameshift with the result that the coat protein subunits are produced in 10-fold higher amounts than the enzymes. This subunit : enzyme ratio is important for productive infection of the virus. For such viruses, the recoding site has clear motifs and stimulatory elements consisting of two *cis* acting sequences, a heptanucleotide motif XXXYYYZ (a slippery sequence) and a secondary structural element, usually a pseudoknot.

It was a seminal paper in 1985 that indicated quite clearly that frameshifting was not going to be restricted to the example crafted so elegantly for RF2 gene expression [86]. The Rous Sarcoma Virus used a recoding site that included a stop codon (AAAUUUAUAG) similar to that used in expression of RF2 but here the shift was backwards rather than forwards. The new codon in the A-site became AUA rather than UAG and the stop codon was avoided. Jacks and Varmus [86] noticed that the AAAUUUA motif allowed slippage of the two tRNAs (Asn and Leu) simultaneously to re-pair in a near-cognate manner with AAA UUU. They proposed a simultaneous-

slippage model from the A- and P-sites meaning that slippage would occur before peptide transfer and translocation. This mechanism became the generally accepted model for –1 frameshifting.

In hindsight, there were two aspects of this mechanism that required further investigation. First, the proposed mechanism would not involve the stop codon in the recoding event within the decoding center and yet the importance of the stop codon as a part of recoding sites as discussed here is clearly established. While the codon following the slippery motif in viruses is often not a stop codon, for example, it is GGG in HIV-1, there is still a restricted subset of codons used in this position. This is highly suggestive that the UAG codon in the Rous Sarcoma Virus frameshift site (or the GGG codon in the HIV-1 site) might be involved in the recoding event and may have a role at the decoding site (Figure 10.7). How can this occur? A simultaneous-slippage model where slippage occurs at the P- and E-sites rather than the A- and P-sites, would allow the next codon to be occupying the A-site. Then the stop or sense codon could contribute to a translational pause and facilitate the slippage event. Secondly, the unusual simultaneous-slippage mechanism proposed had slippage occurring in competition with peptide-bond formation that is normally a kinetically rapid event in protein synthesis. It is more likely that frameshifting is in competition with a kinetically slower event. While others have favored this to be a step after peptide-bond formation but prior to translocation so that slippage could still occur from the A- and P-sites, we believed an equally similar step occurred after translocation, i.e., slippage from the P- and E-sites in competition with the new event in the now empty A-site. This could be the decoding of the stop or sense codon depending on the sequence composition of the particular viral site.

For this reason, we attempted to determine whether the codon after the slippery site was being decoded in the A-site before slippage had occurred. Initially, because the genes for the bacterial RFs were available within plasmids we used the HIV-1

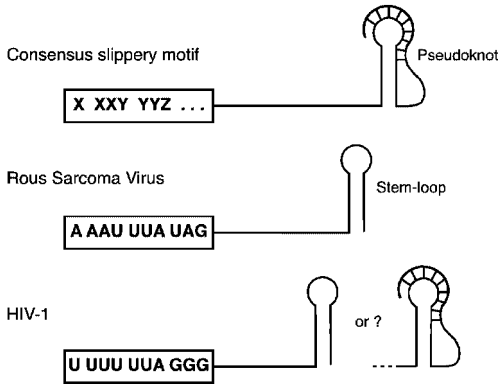


Figure 10.7 Sequence motifs and cis elements at viral –1 frameshift sites.

recoding site with the GGG codon replaced with each of the stop codons and bacterial ribosomes. We then tested whether over-expression *in vivo* of the bacterial RFs would affect frameshifting at the recoding site that occurred readily on these bacterial ribosomes. There were two key observations. First, changing the codon from the natural GGG to each of the stop codons, UAG, UGA and UAA, markedly lowered frameshifting, indicating that the codon was having a major effect on the mechanism. More significantly, over-expression of the factors eliminated frameshifting in a codon-specific manner (RF1 at UAG and UAA, RF2 at UGA and UAA). This was compelling evidence that the codon following the recoding site was being decoded before slippage had occurred and supported a P–E site simultaneous-slippage model [87]. Now that the genes for the eukaryotic RFs are available we have repeated this study *in vivo* using mammalian cells. The results from these studies using mammalian ribosomes concur with our *in vivo* studies that used bacterial ribosomes. First, the stop codon depressed frameshift efficiency. Secondly, over-expression of eRF1 caused a further reduction of efficiency as it did also at the antizyme recoding site used as a control [88]. This is provocative evidence that this –1 frameshift event is occurring in a similar manner to a +1 frameshift where recoding is in competition with a canonical decoding event at the ribosomal A-site.

Recently, Dinman and co-workers [89] have proposed an elegant integrated model to explain programed frameshifting that defines why in some circumstances +1 and in others –1 frameshifting occurs. A key element of the model is that the occupancy states of the ribosome are different for each event. For +1 frameshifting, the A-site is empty and the shift occurs in the post-translocational state (P- and E-sites occupied with tRNAs), but for –1 frameshifting, the A-site is proposed to be already occupied and therefore the ribosome is in the pre-translocational state (A- and P-sites occupied with tRNAs). They have proposed this model as a result of studies that used antibiotics with normal and mutant yeast strains and how under these circumstances frameshifting efficiency was affected. A consistent set of conclusions was drawn from the differential antibiotic sensitivities of the +1 and –1 frameshift events to support the model. For example, translocation inhibitors interfere with +1 but not –1 frameshifting, although neither do they enhance it. As predicted by the model, a peptidyltransferase inhibitor, sparsomysin, that would increase a ribosomal pause when the A- and P-sites are occupied enhances –1 but not + 1 frameshifting.

These conclusions are consistent with what is known about +1 frameshifting from the examples described in this chapter. However, it conflicts with conclusions drawn from our recent evidence that frameshifting at the classic viral –1 site also occurs at the post-translocational state of the ribosome at which time the A-site is empty. In our model, +1 and –1 frameshift events would generally start from the same post-translocational state. It is the specific sequence of the site, the nature of the constraining *cis*-acting elements and how they impinge on interactions between the tRNAs, mRNA and rRNA and, in particular, the stability of the peptidyl-tRNA:mRNA interaction that will determine whether the shift is forwards or backwards. It invokes the common element of a translational pause at the decoding A-site, dislocation in most cases of the codon:anticodon pairing interactions at the P- and E-sites and re-pairing

close by where possible (+1 and –1 events) or bypassing larger tracts of sequence if immediate re-pairing is not favorable (Fig. 10.8).

How can these apparently conflicting positions be accommodated? It could be that if there were a weakening of the already compromised interactions at recoding sites by any effector such as an antibiotic or a mutation creating a new pause in a particular ribosomal state, then this would have the potential to facilitate a shift in frame. Frameshifting may be able to occur in different states of the ribosome under these circumstances. If the partially inhibited process still does not become the rate-limiting step, then it may have no effect on frameshifting. However, if it now makes that process the slowest step, then frameshifting could be enhanced (e.g., inhibiting –1 but not +1 frameshifting by the peptide-bond formation inhibitor, sparsomycin). Moreover, as the detailed new structural information on the ribosome has become available it is revolutionizing our understanding on how antibiotics act and our understanding on the various steps of protein synthesis deduced from biochemical studies both *in vitro* and *in vivo*. Further experiments will be needed to resolve some of these apparent paradoxes and it may not be possible to invoke a completely integrated model to explain all programed frameshift events.

In addition to the classic viral –1 frameshift events of the retroviruses and other viral groups including bacteriophages, there are also examples identified in cellular genes. In the prokaryotic examples, such as that found in the gene *dnaX* encoding a subunit of DNA polymerase III [90–92], there are different features to the recoding site that reflect how elements can be added to a basic recoding framework. The Shine–Dalgarno sequence documented for the *prfB* gene (+1 frameshifting) is found in the *dnaX* gene as an upstream element, but at a different and precise spacing for acting as a stimulator. The slippery heptamer sequence, AAAAAAG, allows two Lys–tRNAs paired with AAAAAG to slip backwards and re-pair with AAAAAA so that the following G becomes the first base of the next codon. Before slippage, this G would be paired with a modified U (mnm^5S^2U) in the anticodon of the tRNA, the same modification found often in the P-site tRNA at termination sites. However, the modification prevents stable base pairing with G and thereby weakens the codon:anti-

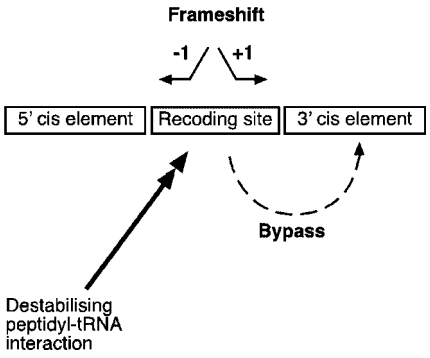


Figure 10.8 Alternative recoding events that could result from destabilization of the peptidyl-tRNA interaction.

codon interaction. Downstream from the slippery site there is a hairpin loop that can also act as a stimulator independent of the upstream element.

Does –1 frameshifting occur in cellular genes in eukaryotes? As the features of a –1 viral recoding site are quite clear-cut, it has been possible to create an algorithm to search genomes for genes that might use this strategy. While a number of possible genes have been flagged, there is, as yet, neither any evidence that they use this strategy nor a rationale as to why they might do so [93]. Nevertheless, recently, an EST was discovered that represented a gene, *edr*, that appeared to have a retrovirus as its ancestor. The recoding site was GGGAAAC with a pseudoknot located downstream. It was expressed temporally and spatially during development although it is not known whether expression is essential for development. However, it seems to use a –1 frameshift mechanism for expression [94]. As programed –1 frameshifting is used in the viral biology of pathogenic viruses like HIV-1 and is a potential target for an antiviral agent, it is important to identify putative human genes that might use this mechanism as this might preclude the viral recoding site as a place of attack in an antiviral strategy. So far, there is no definitive data to preclude this approach although further study is required on the importance of expression of the *edr* gene.

10.10 Conclusion

When RNA molecules interact there can be amazing consequences. We have known for some time that protein synthesis is likely to be the consequence of a highly accurate and fast RNA machine that incorporates specific features to ensure that speed and accuracy can occur together. Recent developments have heightened this appreciation now that the RNA and protein components can be seen in atomic detail. For the first time, it is possible to visualize how some of these special features function. The sensing by the rRNA that the correct codon:anticodon interactions between the tRNA and the mRNA are occurring at the decoding A-site is an exquisite example. The 'enzyme' (the rRNA) holds the 'substrates' (the tRNAs) with unique interactions to maintain the canonical three-base reading frame to ensure they are read with an acceptable level of accuracy. It is not surprising then that this high degree of precision can be overridden. Sometimes, this is a result of a low-frequency error. Sometimes, it is as a result of an atypical physiological perturbation such as starvation for an amino acid but as we know now, sometimes, it is a result of a non-canonical recoding event that has evolved for highly specialized physiological reasons.

As each interaction between the mRNA, rRNA and the tRNAs is so critical for maintenance of precision, there is potential for elements acting in *cis* or *trans* to perturb one or more of these specific pairings. Sequences in the mRNA, or structures that can form within it, have the capacity to form new interactions with the rRNA and put strain on normal interactions. As a consequence, the highly ordered co-ordination between structure and function can be disrupted and kinetic parameters can be slowed giving a chance for alternative events that normally would not be competitive to be significant. *Trans* activators of non-canonical recoding can be of a diverse nature such as the protein RF2 acting at the recoding site within its own mRNA, or

metabolites like polyamines in the antizyme frameshift site or a unique tRNA like $tRNA^{Sec}$ in the Sec incorporation site. Sites where recoding events have evolved often have a stop codon as a central core with upstream and downstream elements that help to ensure that stop codon is more slowly decoded. These sequence elements exert influence in two ways. First, sequences contribute to the decoding efficiency of the stop signal itself. Upstream elements likely to be contributing to the three-dimensional architecture of the RF binding site and downstream elements are probably facilitating stable interactions between the factor and mRNA. Secondly, the upstream and downstream elements can independently perturb other parts of the decoding center and affect parameters independent of the stop codon. While having a stop codon at the recoding site is not universal, it does provide an ideal environment for the creation of translational pauses and provides a vulnerable 'Achilles heel' of canonical protein synthesis where recoding events can evolve.

It is not surprising that most examples of recoding occur in the highly specialized genetic system of the virus. Both bacteriophages and eukaryotic viruses have evolved and maintained highly successful recoding mechanisms. The classic –1 frameshift site found in many eukaryotic viruses has allowed multiple proteins to be synthesized from a single RNA with economy and in balanced amounts. It is surprising, now that we know recoding can be an important translational control mechanism for gene expression, that more examples have not been found in prokaryotic and eukaryotic cellular genes. Those that are known and relatively well characterized such as +1 frameshifting in both the *prfB* gene for RF2 and in the gene for ornithine decarboxylase antizyme and Sec incorporation at selected UGA stop codons, are highly elegant mechanisms for gene expression regulation. These events add a layer of fine control and subtlety to the more typical control mechanisms. Now that new genome and proteome information is appearing rapidly, will the opportunity to examine gene function in detail reveal more examples of the recoding types that are already known and, perhaps, even new events? It is likely that new events will be discovered but these probably will represent rare exclusive examples where some niche advantage is gained. RNA provides 'flexible moulding clay' from which new mechanisms and functions can emerge [95]. There are sure to be more surprises.

References

1 H. J. Rheinberger, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5310–5304.
2 K. H. Nierhaus, *Mol. Microbiol.* **1993**, *9*, 661–669.
3 J. Frank, J. Zhu, P. Penczek et al., *Nature* **1995**, *376*, 441–444.
4 R. F. Gesteland, J. F. Atkins, *Annu. Rev. Biochem.* **1996**, *65*, 741–768.
5 S. L. Alam, J. F. Atkins, R. F. Gesteland, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14177–14179.
6 M. J. Berry: in *Translational Control of Gene Expression, eds* N. Sonenberg, J. W. B. Hershey, M. B. Mathews, Cold Spring Harbor Laboratory Press, New York 2000.
7 P. V. Baranov, R. F. Gesteland, J. F. Atkins, *Gene* **2002**, *286*, 187–201.
8 A. J. Herr, J. F. Atkins, R. F. Gesteland, *Annu. Rev. Biochem.* **2000**, *69*, 343–372.
9 P. J. Farabaugh, *Microbiol. Rev.* **1996**, *60*, 103–134.

10 J. M. Ogle, D. E. Brodersen, W. M. Jr. Clemons et al., *Science* **2001**, *292*, 897–902.
11 A. Böck, K. Forchhammer, J. Heider et al., *Mol. Microbiol.* **1991**, *5*, 515–520.
12 A. Böck, K. Forchhammer, J. Heider et al., *TIBS* **1991**, *16*, 463–467.
13 W. Leinfelder, K. Forchhammer, B. Veprek et al., *Proc. Natl. Acad. Sci. USA* **1990**, *87* 543–547.
14 L. S. Mullins, S-B. Hong, G. Gibson et al., *J. Am. Chem. Soc.* **1997**, *119*, 6684–6685.
15 W. Leinfelder, E. Zehelein, M. A. Mandrand-Berthelot et al., *Nature* **1988**, *331*, 723–725.
16 C. Baron, E. Westhof, A. Böck etal., *J. Mol. Biol.* **1993**, *231*, 274–292.
17 K. Forchhammer, W. Leinfelder, K. Boesmiller et al., *J. Mol. Biol.* **1991**, *266*, 6318–6323.
18 K. Forchhammer, A. Böck, *J. Biol. Chem.* **1991**, *266*, 6324–6328.
19 H. Engelhardt, K. Forchhammer, S. Muller et al., *Mol. Microbiol.* **1992**, *6*, 3461–3467.
20 K. Forchhammer, W. Leinfelder, A. Böck, *Nature* **1989**, *342*, 453–456.
21 C. Baron, A. Böck, *J. Biol. Chem.* **1991**, *266*, 20375–20379.
22 F. Zinoni, J. Heider, A. Böck, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 4660–4664.
23 A. Hüttenhofer, E. Westhof, A. Böck, *RNA* **1996**, *2*, 354–366.
24 A. Hüttenhofer, A. Böck, *Biochemistry* **1998**, *37*, 885–890.
25 S. J. Klug, A. Huttenhofer, M. Kromayer et al., *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6676–6681.
26 J. B. Mansell, *PhD Thesis*, University of Otago, New Zealand, 1999.
27 J. B. Mansell, D. Guévremont, E. S. Poole et al., *EMBO J.* **2001**, *20*, 7284–7293.
28 S. Suppmann, B. C. Persson, A. Böck, *EMBO J.* **1999**, *18*, 2284–2293.
29 R. Wilting, S. Schorling, B. C. Persson et al., *J. Mol. Biol.* **1997**, *266*, 637–641.
30 P. R. Copeland, J. E. Fletcher, B. A. Carlson et al., *EMBO J.* **2000**, *19*, 306–314.
31 R. M. Tujebajeva, P. R. Copeland, X.-M. Xu et al., *EMBO Rep.* **2000**, *1*, 158–163.
32 D. Hatfield, A. Diamond, B. Dudock, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6215–6219.
33 B. J. Lee, M. Rajagopalan, Y. S. Kim et al., *Mol. Cell Biol.* **1990**, *10*, 1940–1949.
34 S. C. Low, J. W. Harney, M. J. Berry, *J. Biol. Chem.* **1995**, *270*, 21659–21674.
35 E. Grundner-Culemann, G. W. Martin III, J. W. Harney et al., *RNA* **1999**, *5*, 625–635.
36 K. K. McCaughan, C. M. Brown, M. E. Dalphin et al., *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5431–5435.
37 J. Kohli, H. Grosjean, *Mol. Gen. Genet.* **1981**, *182*, 430–439.
38 C. M. Brown, P. A. Stockwell, C. N. A. Trotman et al., *Nucleic Acids Res.* **1990**, *18*, 2079–2086.
39 C. M. Brown, M. E. Dalphin, P. A. Stockwell et al., *Nucleic Acids Res.* **1993**, *21*, 3119–3123.
40 C. M. Brown, P. A. Stockwell, C. N. A. Trotman et al., *Nucleic Acids Res.* **1990**, *18*, 6339–6345.
41 W. P. Tate, E. S. Poole, J. A. Horsfield et al., *Biochem. Cell Biol.* **1995**, *73*, 1095–1103.
42 L. L. Major, *PhD Thesis*, University of Otago, New Zealand, 2001.
43 E. S. Poole, L. L. Major, S. A. Mannering et al., *Nucleic Acids Res.* **1998**, *26*, 954–960.
44 D-J. G. Scarlett, K. K. McCaughan, D. N. Wilson et al., *J. Biol. Chem.* **2003**, *278*, 15095–15104.
45 L. L. Major, E. S. Poole, W. P. Tate, in preparation.
46 D. S. Konecki, K. C. Aune, W. Tate et al., *J. Biol. Chem.* **1977**, *252*, 4514–4520.
47 S. Mottagui-Tabar, A. Björnsson, L. A. Isaksson, *EMBO J.* **1994**, *13*, 249–257.
48 S. Zhang, M. RydénAulin, L. A. Isaksson, *J. Mol. Biol.* **1996**, *261*, 98–107.

49 A. Björnsson, S. Mottagui-Tabar, L. A. Isaksson, *EMBO J.* **1996,** *15,* 1696–1704.
50 W. J. Craigen, R. G. Cook, W. P. Tate et al., *Proc. Natl. Acad. Sci. USA* **1985,** *82,* 3616–3620.
51 W. J. Craigen, C. T. Caskey, *Nature* **1986,** *322,* 273–275.
52 B. C. Donly, C. D. Edgar, F. M. Adamski et al., *Nucleic Acids Res.* **1990,** *18,* 6517–6522.
53 A. G. Cridge, E. S. Poole, W. P. Tate, unpublished results.
54 H. Beier, M. Grimm, *Nucleic Acids Res.* **2001,** *29,* 4767–4782.
55 M. Cassan, J.-P. Rousset, *BMC Mol. Biol.* **2001,** *2,* 3.
56 B. Bonetti, L. W. Fu, J. Moon et al., *J. Mol. Biol.* **1995,** *251,* 334–345.
57 O. Namy, G. Duchateau-Nguyen, J.-P. Rousset, *Mol. Microbiol.* **2002,** *43,* 641–652.
58 Z. Liu, M. Reches, H. Engelberg-Kulka, *J. Mol. Biol.* **1999,** *294,* 1073–1086.
59 C. M. Brown, S. P. Dinesh-Kumar, W. A. Miller, *J. Virol.* **1996,** *70,* 5884–5892.
60 N. M. Wills, R. F. Gesteland, J. F. Atkins, *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 6991–6995.
61 G. Li, C. M. Rice, *J. Virol.* **1993,** *67,* 5062–5067.
62 A. M. Weiner, K. Weber, *J. Mol. Biol.* **1973,** *80,* 837–855.
63 L. Harrell, U. Melcher, J. F. Atkins, *Nucleic Acids Res.* **2002,** *30,* 2011–2017.
64 D. N. Robinson, L. Cooley, *Development* **1997,** *124,* 1405–1417.
65 D. M. Ivey, J. Cheng, T. A. Krulwich, *Nucleic Acids Res.* **1992,** *20,* 4928.
66 W. M. Huang, S. Z. Ao, S. Casjens et al., *Science* **1988,** *239,* 1005–1012.
67 R. B. Weiss, W. M. Huang, D. M. Dunn, *Cell* **1990,** *62,* 117–126.
68 A. J. Herr, R. F. Gesteland, J. F. Atkins, *EMBO J.* **2000,** *19,* 2671–2680.
69 J. H. Cate, M. M. Yusupov, G. Zh. Yusupova et al., *Science* **1999,** *285,* 2095–2104.
70 D. N. Wilson, D. Guévremont, W. P. Tate, *RNA* **2000,** *6,* 1704–1713.
71 A. J. Herr, C. C, Nelson, N. M. Wills et al., *J. Mol. Biol.* **2001,** *309,* 1029–1048.
72 J. Gallant, D. Lindsey, *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 13771–13776.
73 P. J. Farabaugh, *Annu. Rev. Genet.* **1996,** *30,* 507–528.
74 G. Stahl, G. P. McCarty, P. J. Farabaugh, *Trends Biochem. Sci.* **2002,** *27,* 178–183.
75 D. Riddle, J. Carbon, *Nature New Biol.* **1973,** *242,* 230–234.
76 P. V. Baranov, R. F. Gesteland, J. F. Atkins, *EMBO Rep.* **2002,** *3,* 373–377.
77 R. B. Weiss, D. M. Dunn, A. E. Dahlberg et al., *EMBO J.* **1988,** *7,* 1503–1507.
78 E. S. Poole, C. M. Brown, W. P. Tate, *EMBO J.* **1995,** *14,* 151–158.
79 F. M. Adamski, C. Donly, W. P. Tate, *Nucleic Acids Res.* **1993,** *21,* 5074–5078.
80 J. F. Curran, M. Yarus, *J. Mol. Biol.* **1989,** *209,* 65–77.
81 J. F. Curran, *Nucleic Acids Res.* **1993,** *21,* 1837–1843.
82 S. Matsufuji, T. Matsufuji, Y. Miyazaki et al., *Cell* **1995,** *80,* 51–60.
83 I. P. Ivanov, R. F. Gesteland, J. F. Atkins, *Nucleic Acids Res.* **2000,** *28,* 3185–3196.
84 A. Sundararajan, W. A. Michaud, Q. Qian et al., *Mol. Cell* **1999,** *4,* 1005–1015.
85 D. K. Morris, V. Lundblad, *Current Biol.* **1997,** *7,* 969–976.
86 T. Jacks, H. E. Varmus, *Science* **1985,** *30,* 1237–1242.
87 J. A. Horsfield, D. N. Wilson, S. A Mannering et al., *Nucleic Acids Res.* **1995,** *23,* 1487–1494.
88 R. S. Graves, C. Z. M. McKinney, E. S. Poole et al., unpublished results.
89 J. W. Harger, A. Meskauskas, J. D. Dinman, *Trends Biochem. Sci.* **2002,** *27,* 448–453.

90 A. L. Blinkowa, J. R. Walker, *Nucleic Acids Res.* **1990,** *18,* 1725–1729.
91 A. M. Flower, C. S. McHenry, *Proc. Natl. Acad. Sci. USA* **1990,** *87,* 3713–3717.
92 Z. Tsuchihashi, A. Kornberg, *Proc. Natl. Acad. Sci. USA* **1990,** *87,* 2516–2520.
93 A. B. Hammell, R. C. Taylor, S. W. Peltz et al., *Genome Res.* **1999,** *9,* 417–427.
94 K. Shigemoto, J. Brennan, S. E. Walles et al., *Nucleic Acids Res.* **2001,** *29,* 4079–4088.
95 D. Kennedy, *Science* **2002,** *298,* 2283.

11
Regulation of Ribosome Biosynthesis in *Escherichia coli*

Madina Iskakova, Sean R. Connell, and Knud H. Nierhaus

Overview of Ribosome Biosynthesis Regulation

In all living organisms, protein biosynthesis, a central process important for cellular growth and development, is catalyzed by the ribosome. The ribosome represents a cellular paradigm for a macromolecular machine and, essentially, translates genetically encoded information into functionally active proteins. Structurally, the ribosome is composed of both RNA and proteins interwoven in a complex three-dimensional-fold (see Chap. 2). This complex structure necessitates an ordered assembly process (see Chap. 3), as well as coordinated production of the constituent RNA and protein molecules (discussed below). Specifically, the regulated synthesis of the ribosomal components and the cooperativity of the assembly process ensure that each ribosome contains a full compliment of it constitutive parts.

Additionally, the ribosome is physically large (2.5 MDa in *Escherichia coli*), its 3 RNA chains comprise 4566 nucleotides [1] and the 54 proteins, 7343 amino acids [2]. In bacteria, ribosomes can account for as much as 50% of the cell dry mass [3–5], whereas in eukaryotes the corresponding value is less than 5% [6]. Taking into account that the translational apparatus comprises, in addition to ribosomes, elongation factors (e.g., EF-Tu is the most common protein in the bacterial cell contributing ~10% of the total protein mass), synthetases and tRNAs, one can estimate that up to 60% of the total cell energy is expended on the synthesis of the translational apparatus. This enormous energetic commitment by the bacterial cell requires a coordinated synthesis of rRNA and ribosomal proteins. Accordingly, an intricate network of regulations exists in bacteria (i) to ensure a balanced synthesis of rRNAs and ribosomal proteins and (ii) to adapt ribosome synthesis to the cells nutritional environment. Only under unfavorable conditions and during the stationary phase, are ribosomes present in excess. During stationary phase, 70S ribosomes are present as inactive 100S dimers due to a ribosome-associated protein (ribosome modulation factor, [7]). Consistently, null mutations of this factor affect viability of cells at stationary, but not log phase [8].

The tight coupling of ribosome synthesis to growth rate of the cell is termed *growth rate control* and is defined as "the number of ribosomes per unit amount of cellular protein in *E. coli* is proportional to the growth rate (μ), and the rate of ribosome biosynthesis is proportional to μ^2 [9]". One important control mechanism not

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

observed in eukaryotes is termed the stringent control, which results in an immediate stop of RNA synthesis upon a shortage of amino acids. Further bacterial control mechanisms for ribosome biosynthesis are present at the level of ribosomal RNA (rRNA) transcription (see Ref. [9] for a review of rRNA regulation) as well as translation of ribosomal proteins. We will see that the latter, called "translational control", is regulated in response to the rRNA levels (see Ref. [10] for a review of r-protein regulation), whereas rRNA synthesis is directly regulated (growth rate control).

In this review, we will outline the fundamental ideas governing the regulation of ribosome biosynthesis and where established and proven models exist we will delve into specific examples. In the case where more detailed knowledge is required we would direct the reader to the comprehensive reviews given by Refs. [9–11].

11.1 Regulation of rRNA Synthesis

11.1.1 Organization of rRNA Operons and Elements of rRNA Promoters

In *E. coli* there are seven copies of the rRNA operons arranged within the first half of the chromosome relative to the origin of replication (Fig. 11.1A). As is characteristic for highly expressed genes the rRNA operons are transcribed in the direction of DNA synthesis (Fig. 11.1A) to avoid clashes between replicase and transcriptase. The reason is that only under these conditions can transcriptase transcribe the operon continuously from initiation to termination in harmony with the replicase synthesizing the leading DNA strand, rather than being interrupted by gaps between the Okazaki pieces of the lagging strand of the replication fork. Furthermore, the clustering of the rRNA operons around the origin of replication (Fig. 11.1A) ensures that their relative gene dosage is much greater then the absolute gene dosage, since replication eyes around the replication origin can lead to partial di- and tetra-diploidy (Fig. 11.2). This may be important in rapidly growing cells where initiation of DNA synthesis is faster than cell division, ensuring that multiple copies of the genome around the origin of replication exist in the cell.

In *E. coli* the rRNA operons are designated *rrnA*, *rrnB*, *rrnC*, *rrnD*, *rrnE*, *rrnG*, and *rrnH*. In the rRNA operons, the 16S, 23S, and 5S rRNA, as well as several tRNA genes are co-transcribed as a single transcript (Fig. 11.1B), which is subsequently processed to generate the individual rRNA (see Chap. 3) and tRNA molecules. Transcription of the 5S, 16S, and 23S molecules as a single transcript ensures stoichiometric production of the rRNAs; however, it should be noted that in a few prokaryotic species the rRNA operons are fragmented, e.g., in the bacteria *Thermoplasma* 23S, 16S, and 5S rRNA are all transcribed from separate promoters, whereas the bacterium *Thermus thermophilus* and archaeon *Desulfococcus*, the 16S and (23S+5S) have individual promoters (Fig. 11.3)

The rRNA genes are transcribed from tandem promoters called the *rrn* P1 and *rrn* P2 promoters (Fig. 11.1B). Both promoters have similar core sequences consisting of –10 and –35 regions that are recognized by the σ^{70} subunit of the RNA polymerase

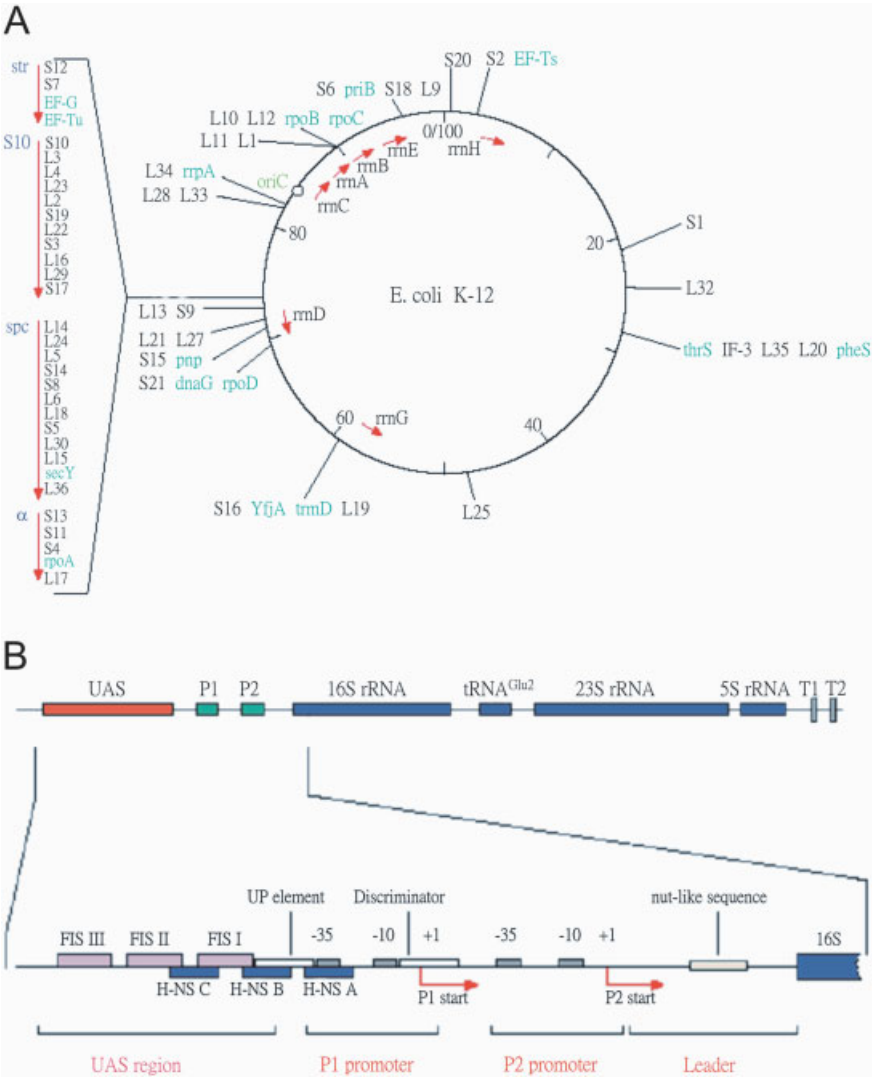


Figure 11.1 (A) Genetic organization of the rRNA and r-protein operon within the *E. coli* chromosome. The cluster of the four large protein operons is shown on the left (str, S10, spc and α operon). Non-ribosomal components are colored. The red arrows indicate the direction of transcription. OriC, origin of replication. (B) The architecture of the rRNA operons, where promoter and regulatory regions are enlarged. See text for details. (A) and (B) have been reproduced from Refs. [50] and [24] respectively.

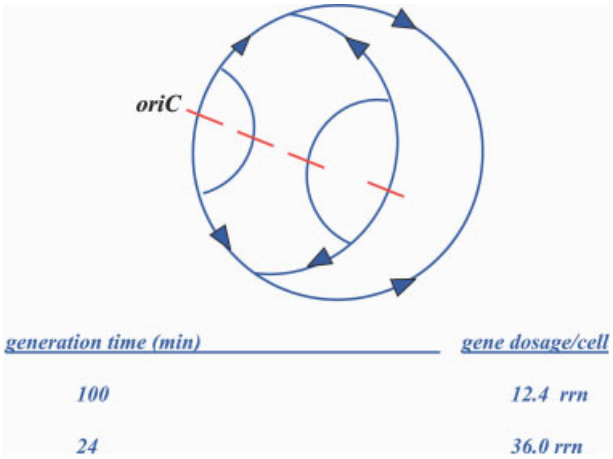


Figure 11.2 Relative gene dosage: during DNA synthesis, transcription units closer to oriC exist in a higher relative copy numbers compared with more distal operons. The relative copy numbers of rrnC, A, B, and E will exceed those of rrnH, D, or G, when cells grow in a rich medium.

(RNAP), although they do not have perfect σ^{70} consensus sequences. Additionally, both have a G+C-rich sequence downstream of the –10 region called a discriminator sequence, which plays a role in regulation during the stringent response (see Sect. 11.1.3). Until recently, it was believed that the *rrn* P1 promoter was regulated whereas the *rrn* P2 was constitutively active at low levels. However, recent work suggests that the P2 promoter is regulated similar to the P1 but not nearly to the same extent [12]. Table 11.1 compares structural features and the extent of regulation with that of the *lac* promoter.

The fact that in fast growing *E. coli* cells the majority of the RNAP initiates transcription at the *rrn* P1 is due to recruitment by the upstream activating sequences (UAS; Fig. 11.1B; Ref. [13]). The UAS includes the UP element which is a A+T-rich sequence upstream of the –35 element as well as 3–5 binding sites for a *trans*-acting

Table 11.1 Features of the rrn P1 and P2 promoters and comparison of their relative transcription intensities with that of the *lac* promoter

	rrn		
	P1	**P2**	***lac* promoter**
–35 region	TTGTC	TTGACT	TTGACA
–10 (TATA box)	TATAATG	TATATA	TATAAT
Discriminator	CGCC(T/A)CC	C(G/A)C(C)ACC	–
Start transcription	A/G	C/T	A
Regulation	+++	+	++ (*lacI*)
Transcription	3-10	1	0.25-1

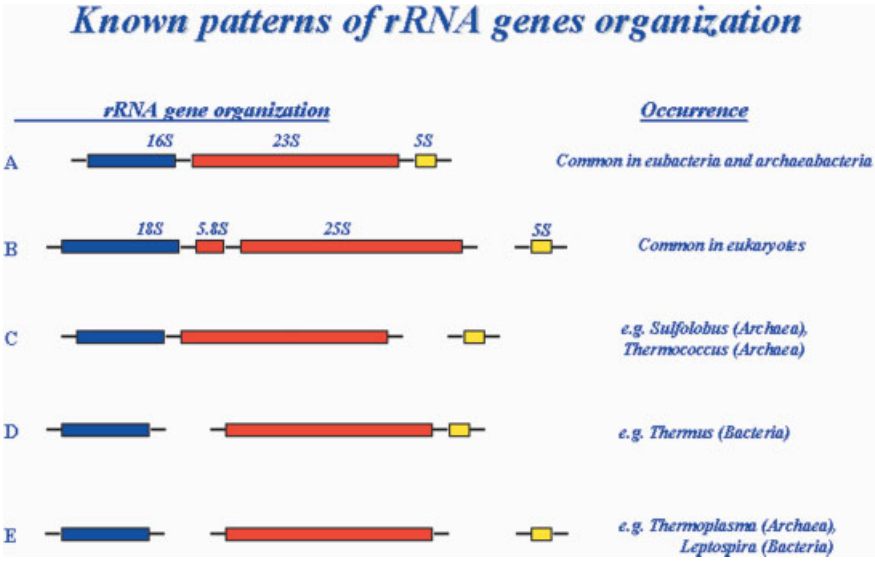


Figure 11.3 Various patterns of organization of rRNA genes in different organisms. In most prokaryotes, rRNA genes are expressed as one operon, and in eukaryotes, the 5S rRNA genes are expressed independently. Some exceptional organizations of rRNA genes in a few prokaryotes are presented in this figure. Similar to *Sulfolobus* [85] and *Thermococcus* [86], 16S and 23S rRNA genes are linked, whereas the 5S rRNA gene is not. In *Thermus thermophilus* [87], the 16S rRNA gene is separated from and transcribed independently of the 23S and 5S genes. *Thermoplasma* contains one copy of each of the rRNA genes, dispersed in the genome and separated by at least 52 kbp [88]. Also in *Leptospira* the three rRNAs are transcribed separately [89].

protein called Fis. In the UAS of the *rrn* P1 promoter, the UP element is the most important factor for activating transcription (20–50-fold activation; Ref. [13]). This activation is promoted through an interaction with the C-terminal domain of the α subunit of RNAP (αCTD) and the minor groove of the DNA representing the UP element [14, 15]. Similar to the UP element, Fis also interacts with the αCTD but instead uses protein–protein interactions [16, 17] to stimulate transcription 3–4-fold [13]. The binding sites of Fis on the DNA are spaced in a way that on the DNA helix the Fis proteins bind on the same side of the helix and induce a DNA bending, which facilitates the downstream TATA box to be melted [16, 18]. Fis is not essential for growth, but plays an important role in the feedback mechanism of growth rate control [11]. The binding of Fis proteins is counteracted by the H-NS proteins that bind downstream, partially overlapping the Fis-binding sites (Fig. 11.1B; Ref. [19]).

11.1.2 Models for rRNA Regulation

Regulation of rRNA synthesis is believed to be primarily due to a feedback mechanism that relates the excess translational ability of the cell to the production of rRNA. This is logical because the translation ability of the cell is synonymous with the amount of ribosomes present, which in turn is influenced by the production of RNA. Although this feedback mechanism was demonstrated in several ways, it can be nicely elucidated from experiments that show overproduction of the rRNA inhibits transcription from the *rrn* promoters, but only if the overproduced rRNA was competent to form active ribosomes that could engage in protein synthesis [20].

Although the Fis and UP elements are responsible for the strength of the *rrn* promoters [13], they are not probably involved in growth rate regulation, since promoters with deletions in the UAS are still actively regulated [21]. Although there is still disagreement over the source and mechanism of growth rate regulation (see Ref. [22] for a discussion), the most probable molecules to be involved in regulating rRNA synthesis are ppGpp and iNTPs [23, 24]. The ability of ppGpp to regulate rRNA synthesis during extreme amino acid starvation or stringent response has been well studied (see Sect. 11.1.3). ppGpp is proposed to regulate transcription by decreasing the half-life of the open complex formed by RNAP during initiation of transcription. The half-life of the open complex at rRNA promoters is normally much shorter than that of other cellular promoters, e.g., promoters of genes involved in amino acid biosynthesis [25]. This marked difference in half-life can then explain the extreme sensitivity of rRNA promoters to the destabilizing effects of ppGpp [25]. The intercellular levels of ppGpp vary with growth rate in agreement with the idea that they couple rRNA synthesis with the growth rate [26].

In addition to (p)ppGpp, iNTPs (the NTPs that represent the initiating nucleotide in rRNA synthesis: GTP in *rrn*D P1 and ATP in the other *rrn* P1 promoters; Fig. 11.1B and Table 11.1) are presumed to confer growth rate control, because initiation of transcription from the RNA promoters is strongly dependent on their concentration. Gaal *et al.* [24] propose that the intercellular levels of GTP and ATP would vary depending on (i) the nutritional environment and (ii) on the translational capacity of the cell, since the process of translation would consume ATP and GTP. In the model presented in Fig. 11.4 by Gaal *et al.* [24], these properties are combined to explain both feedback regulation and growth rate control of rRNA synthesis. It should be noted however that this model is largely contested, as it is not known if the intercellular concentrations of NTPs vary significantly [27].

Schneider *et al.* [26, 28] observe that the roles of ppGpp and iNTPs in rRNA regulation may be complementary to "increasing the regulatory robustness of the system" and function during different growth phases.

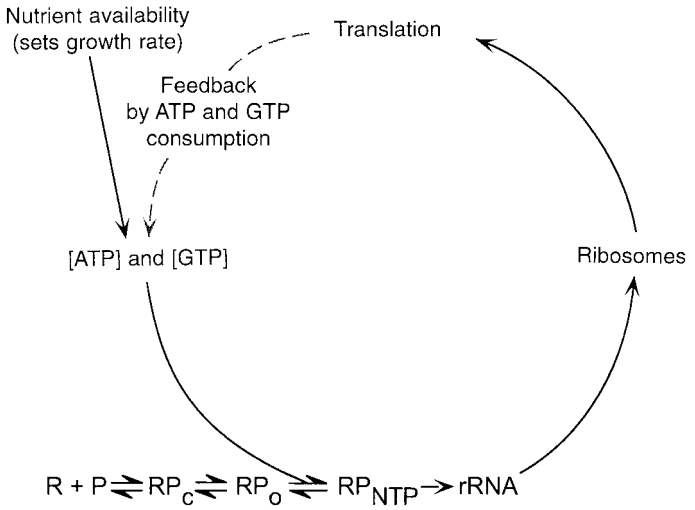


Figure 11.4 NTP-sensing model for RNA regulation. The intercellular ATP and GTP concentrations regulate the stability of the complex of the RNAP at the *rrn* promoters. When the concentrations are sufficient, initiation of rRNA transcription ensues leading to the production of ribosomes. Upon formation, the ribosome will engage in protein synthesis (translation) and in doing so consume GTP and ATP. If the translational activity of the cell is too high with respect to nutrient availability, the GTP and ATP levels will fall, which will in turn limit the initiation of rRNA transcription and therefore couple ribosome production to the growth rate of the cell. This figure has been reproduced from Ref. [24].

11.1.3 Stringent Response

Curtailment of nutrient supply results in an adaptation response in many bacteria termed the *stringent* response. The stringent response is a unique case of rRNA regulation that is triggered by amino acid starvation. It stringently couples protein synthesis with that of RNA and is characterized by (i) a rapid shut down of stable RNA (rRNA, tRNA) synthesis, and (ii) a sudden and significant increase in the levels of ppGpp and pppGpp. The basic level of (p)ppGpp in the cell is about 60 μM and can increase to mM values, i.e., virtually the whole cell content of GDP and GTP is converted to (p)ppGpp [23].

The stringent response provides the cell with a prominent regulatory means to control gene expression. The effect is two-fold: (i) transcriptional repression of genes associated with the translational apparatus, e.g., genes encoding tRNAs, rRNAs [29], ribosomal proteins, translational factors and synthetases [30]. (ii) Up-regulation of genes encoding metabolic enzymes, especially those involved in amino acid biosynthesis [31].

Activation of the stringent response initially stems from the shortage of one (or more) amino acid(s), which in turn produces a significant increase in uncharged-tRNA (deacylated tRNA) for the corresponding amino acid(s). In log-phase bacterial

cells, deacylated tRNA constitutes approximately 15% of the total tRNA, the majority of which is present in a bound state, namely, bound either to ribosomes or synthetases. Under conditions of amino acid starvation, the deacylated tRNA fraction can increase to over 80% of the total tRNA [32]. The scarcity of the aminoacylated tRNA, compounded by the large pools of free deacylated tRNA, enables deacylated tRNA to bind an empty ribosomal A site, conditional to the presence of a cognate codon. The presence of a deacylated tRNA at the A site triggers RelA-dependent synthesis of guanosine 5′-triphosphate 3′ diphosphate (pppGpp) and guanosine 3′, 5′ bisphosphate (ppGpp), collectively referred to as (p)ppGpp. The reaction catalyzed by RelA utilizes ATP and GTP, or GDP, to produce AMP and pppGpp, or ppGpp, respectively (Fig. 11.5, Refs. [33–35]). The products, (p)ppGpp, most probably exert a regulatory effect on transcription via an interaction with the β-subunit of the transcriptase [36–38]. (p)ppGpp is again converted to GTP and GDP, respectively, by the enzyme SpoT (Fig. 11.5; Ref. [39]).

The inhibition of transcription, resulting from the starvation-induced inhibition of translation, is almost immediate (within a minute) and is termed stringent control [40]. This stringent coupling between translation and transcription can be relieved by mutations in either *relA*, the gene encoding the stringent factor RelA [40, 41], or in *relC*, the gene for ribosomal protein L11 ([42, 43]; *relC* ≡ *rplK*). Uncoupling the activities establishes a relaxed phenotype, where RNA synthesis can continue for a period of more than 1 h following translation inhibition [44].

Early studies demonstrated that RelA binding to 70S ribosomes is essential for the production of (p)ppGpp synthesis [45–47] and that RelA binding is enhanced by the presence of a poly(U)-mRNA [48]. The synthesis of (p)ppGpp has been shown to be dependent on a deacylated tRNA at the A site [33] and inhibited in the absence of L11 *in vivo* [42].

Recently, the mechanism of RelA-mediated (p)ppGpp synthesis has been dissected *in vitro* [49]. It was shown that binding of RelA to the ribosome is predominantly influenced by mRNA and not by deacylated tRNA or L11. In contrast, RelA-catalyzed (p)ppGpp synthesis is strictly dependent on L11. Furthermore, it has been clearly demonstrated that it is the release of RelA from the ribosome, not release of the deacylated tRNA, that is concomitant with (p)ppGpp synthesis. Figure 11.6 illustrates

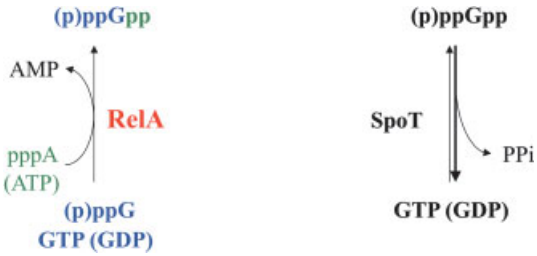


Figure 11.5 Synthesis of (p)ppGpp by RelA and its degradation by SpoT.

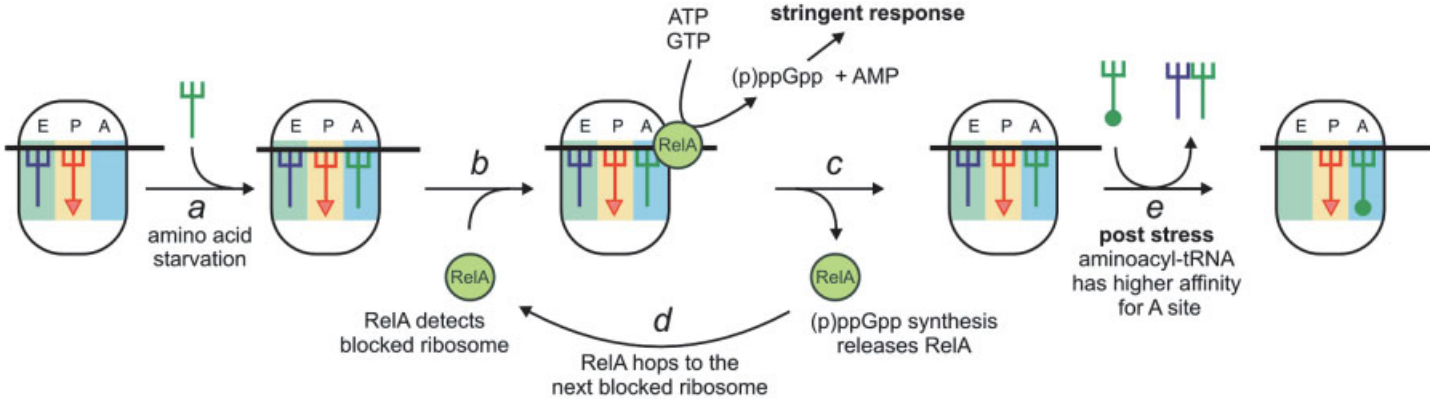


Figure 11.6 Model for mechanism of RelA-mediated (p)ppGpp synthesis: (a) amino acid starvation generates large pools of deacylated tRNAs, which binds to the ribosomal A site with low affinity and blocks the ribosome, (b) RelA detects a blocked ribosome with a 3′-extension of the mRNA, (c) RelA mediates the conversion of ATP and GTP(GDP) to AMP and (p)ppGpp in the presence of a deacylated tRNA at the A site. Release of RelA but not the A-site-bound deacylated tRNA occurs simultaneously with RelA-mediated (p)ppGpp synthesis. (d) RelA 'hops' to the next blocked ribosome and the synthesis of (p)ppGpp is repeated. High levels of (p)ppGpp activate the stringent response. (e) Aminoacylated tRNAs are replenished following post-stress conditions. The higher affinity of an aminoacylated tRNA over deacylated tRNAs for the A site enables displacement of the deacylated tRNAs, which rescues blocked ribosomes and reactivates translation. Note that binding of an aminoacylated tRNA at the A site also results in concomitant release of the E-site-bound tRNA (reviewed in Ref. [90]), thus two deacylated tRNAs are released. Reproduced from Wendrich *et al.* [49].

the present understanding of the mode of action of RelA during the stringent response.

11.2 Regulation of r-protein Synthesis

11.2.1 Some General Remarks

In *E. coli* there are 54 proteins comprising the ribosome. Unlike the rRNA genes, which generally exist in multiple copies within the ribosome, the r-protein genes are present in only a single copy [50]. About half the r-proteins are arranged in the *spc*, S10, *str*, and α operons, whereas the remaining r-proteins are scattered throughout the *E. coli* chromosome in operons containing 1–4 genes (Fig. 11.1A; Ref. [10, 50]). The r-protein operons are often named based on the fact that they harbor genes whose mutations confer resistance to an antibiotic [10]. The r-proteins operons are depicted in Fig. 11.7; however, it should be noted that the division of the operons is not so clear, as often an upstream operon will transcribe into a downstream operon [10]. For example, the α operon can be transcribed from both its own promoter and by RNAP initiating on the *spc* operon promoter [51, 52].

As mentioned above, regulated production of the r-proteins with respect to the nutritional state of the cell is believed to be mediated indirectly by the cellular rRNA levels (reviewed in Ref. [10]). Namely, r-proteins are produced to a level which matches the production of rRNA and when r-protein production exceeds this level, then the accumulation of free ribosomal proteins *feedbacks* and negatively regulates r-protein expression. This is termed *autogenous control* and dictates that, after synthesis from a polycistronic mRNA, a single protein acts both as a r-protein and as a regulatory protein such that its accumulation in an rRNA-free form leads to inhibition of expression of the entire mRNA (Fig. 11.8).

As seen in Fig. 11.7, the regulatory proteins (red boxes) are very often primary rRNA-binding proteins, i.e., that they are capable of binding to the free ribosomal RNA. In contrast with a widespread assumption [10, 53, 54], this does not mean that they associate with the ribosome very early in its synthesis – in fact, only two proteins, L24 and L3, are capable of initiating the assembly of prokaryotic ribosomes (see Chap. 3.2), whereas all other proteins can bind to the rRNAs with high affinity only via the help of other proteins. Nevertheless, if a ribosomal protein can bind already to the naked rRNA, its affinity to a partially assembled subunit is even higher.

Inherent in the *autogenous control* model is the idea that the rRNA and the mRNA targets compete for binding to the regulatory protein. Direct binding to the mRNA targets has been shown for many of the regulatory proteins including L10–L12 [55], S4 [56], S8 [57], S15 [58] and most recently L4 [59]. Additionally, it has been proposed that a *molecular mimicry* exists between the mRNA and rRNA targets and, in some cases, e.g., L1 [60], S8 [57], S15 [61], L20 [62] and L4 [59], this has been established.

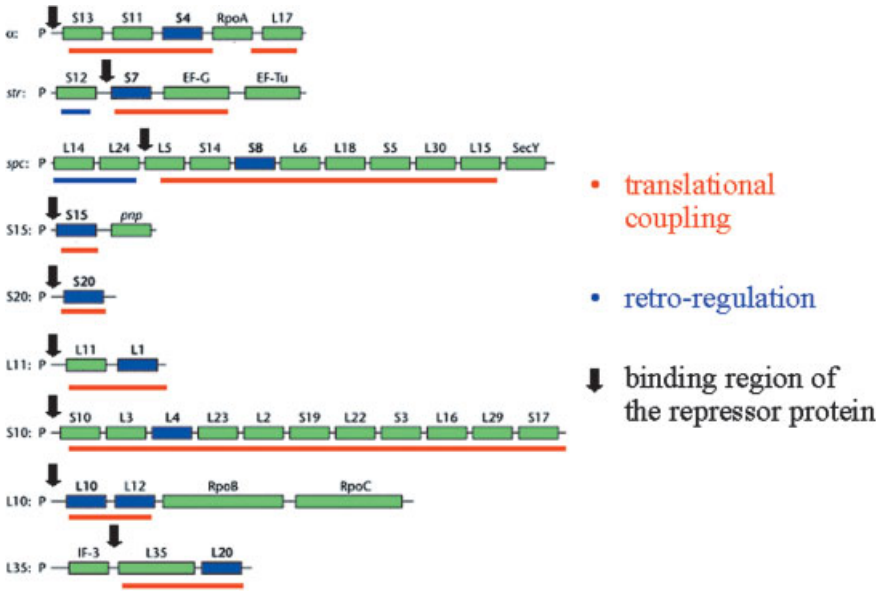


Figure 11.7 The operon structure of the r-proteins. Ribosomal protein operons regulated by translational feedback. The name of each operon is given. P denotes the transcription start site. Individual genes of the operon are shown as green boxes and labeled according to the gene product. The regulatory product is indicated by a blue box and the respective mRNA binding site is shown by an arrow. Genes underlined red are under translational feedback regulation; genes underlined blue are not. In case of the L10 operon the regulator is a complex of L10(L7/L12)$_4$. This figure has been modified from Ref. [50].

The *autogenous control* of r-protein operons is believed to be regulated at various levels:

1. *Transcription*: The binding of the regulatory r-protein effects the elongation of the mRNA. This control mechanism is seen in the S10 operon, where L4 functions in translational attenuation (Sect. 11.2.2.2).
2. *Translation*: The simplest assumption is that the regulatory ribosomal protein binds to the ribosomal binding site of the first cistron and thus interferes with the 30S *de novo* initiation. However, this is not what is observed; instead quite often two alternative secondary structures can form including the ribosomal binding site; one conformation preventing the formation of a productive initiation complex, the other allowing initiation. The regulatory protein is stabilizing the non-productive conformer. In particular, the binding of the regulatory r-protein can (i) entrap the ribosome–mRNA complex in a state that is not competent to initiate translation as seen with the α-operon (called entrapment; see Sect. 11.2.2.3), and (ii) trigger a conformational change in the

mRNA that obscures the RBS in secondary structure so that initiation is inhibited.

3. *mRNA stability*: The binding of the regulatory r-protein destabilizes the mRNA such that it is more readily degraded and therefore cannot be translated (see Sects. 11.2.2.1 and 11.2.2.4).

In many cases, the regulatory protein binds to the 5′ leader sequence directly inhibiting translation of the first gene of the operon, whereas the translation of the following genes is inhibited indirectly by disrupting *translational coupling*. Translational coupling is the phenomenon, where the ribosomal binding sites (RBS) are buried in secondary structures, which are unfolded by ribosomes translating the preceding cistron. When translation of the preceding cistron is finished, the empty 70S ribosome does not necessarily fall off the mRNA but can scan until it reaches the ribosomal binding site of the downstream cistron. fMet-tRNA is bound to the P site with the help of IF2, and translation of the downstream cistron commences. The initiation type is called "70S type initiation" in contrast with the canonical 30S *de novo* initiation (see Chap. 7.1). This "70S type initiation" is a speciality of polycistronic mRNAs encoding ribosomal proteins, and it is this type of initiation that requires N-blocked methionyl-$tRNA_i$ rather than the 30S *de novo* initiation [63, 64]. Since eukaryotic mRNAs are monocistronic and therefore require 40S *de novo* initiation, the initiator Met-$tRNA_i$ does not need to be formylated.

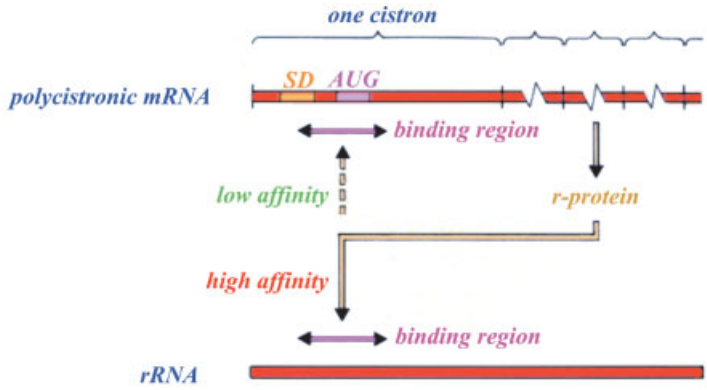


Figure 11.8 Scheme of the translational control of ribosomal proteins. Under fast growth conditions ribosomal proteins do not have a free pool, therefore synthesized ribosomal proteins flow directly into ribosome assembly. However, under some unfavorable conditions, a pool of ribosomal proteins does exist, so that the regulatory proteins indicated below (see also Fig. 11.7) can now bind to their respective mRNA and block the translation of the whole polycistronic mRNA. If the synthesis of ribosomal proteins and rRNA reaches again a molar balance, the ribosomal proteins will dissociate from the mRNA and bind to the higher-affinity binding site on the partially assembled ribosome.

11.2.2 Various Models for r-protein Regulation

11.2.2.1 *spc* operon

The regulatory protein S8 binds between the second and third cistrons, just in front of the L5 cistron of the *spc* mRNA (see Fig. 11.7). Therefore, the downstream genes are regulated via translational coupling. What happens with the genes L14 and L24?

They are repressed by S8 indirectly: when the binding of S8 blocks the translation of the L5 cistron and the downstream cistrons, the mRNA is freed of ribosomes and is thus a target for endonucleolytic RNases. Afterwards, 3′ to 5′ exonucleases will digest the L24 and L14 cistrons ("retroregulation", Ref. [65]).

S8 binds to an mRNA structure similar to the rRNA-binding site with a 5-fold lower affinity [57]. The L5 initiation codon AUG is bulged out but not the Shine–Dalgarno sequence [66, 67]. The ribosomes might be trapped on the SD in a non-productive complex, if S8 is bound in the front of the L5 cistron.

11.2.2.2 **S10 operon**

The S10 operon encodes 11 proteins (Fig. 11.7) and expression of the r-proteins from this operon is regulated by L4 [54, 68]. The regulation by L4 is unique among the r-proteins because it acts at both the transcriptional and translational levels, where both contribute about equally to attain a maximal repression of 25-fold [69]. The determinants for the transcriptional and translational controls are both located within the 172 nucleotide 5′ untranslated region seen in Fig. 11.9. Within this region helices HD and HE are required for transcriptional control, whereas helix HE and the unstructured downstream sequence are required for translational control ([69–72]; Fig. 11.9). Within this region, L4 is proposed to form specific interactions with the loop of helix HD and non-sequence specific interactions with helix HE (Fig. 11.9A; Ref. [59]).

L4-mediated transcription termination occurs on the descending side of helix HE at a string of U's [73] that resemble a rho-independent terminator. It is believed that as the RNAP bound by the transcription factor NusA – a protein that helps RNAP correctly recognize termination signals – transcribes the S10 operon it briefly pauses at a NusA-dependent site before resuming transcription ([74, 75]; Fig. 11.9B). However, under conditions where there is an excess of free cellular L4, this r-protein would bind to its target on the nascent mRNA transcript resulting in structural changes that are propagated to the RNAP–NusA complex [59]. This would prolong the NusA-dependent pause and seemingly stimulate transcription termination (Refs. [74, 75]; Fig. 11.9B), thereby down-regulating r-protein expression from the S10 operon.

The mechanism of L4-mediated translational control is not so well studied; however, it may also operate using an entrapment-based mechanism. This is because the mRNA target for L4 does not include the RBS [59] and, therefore, is not probably

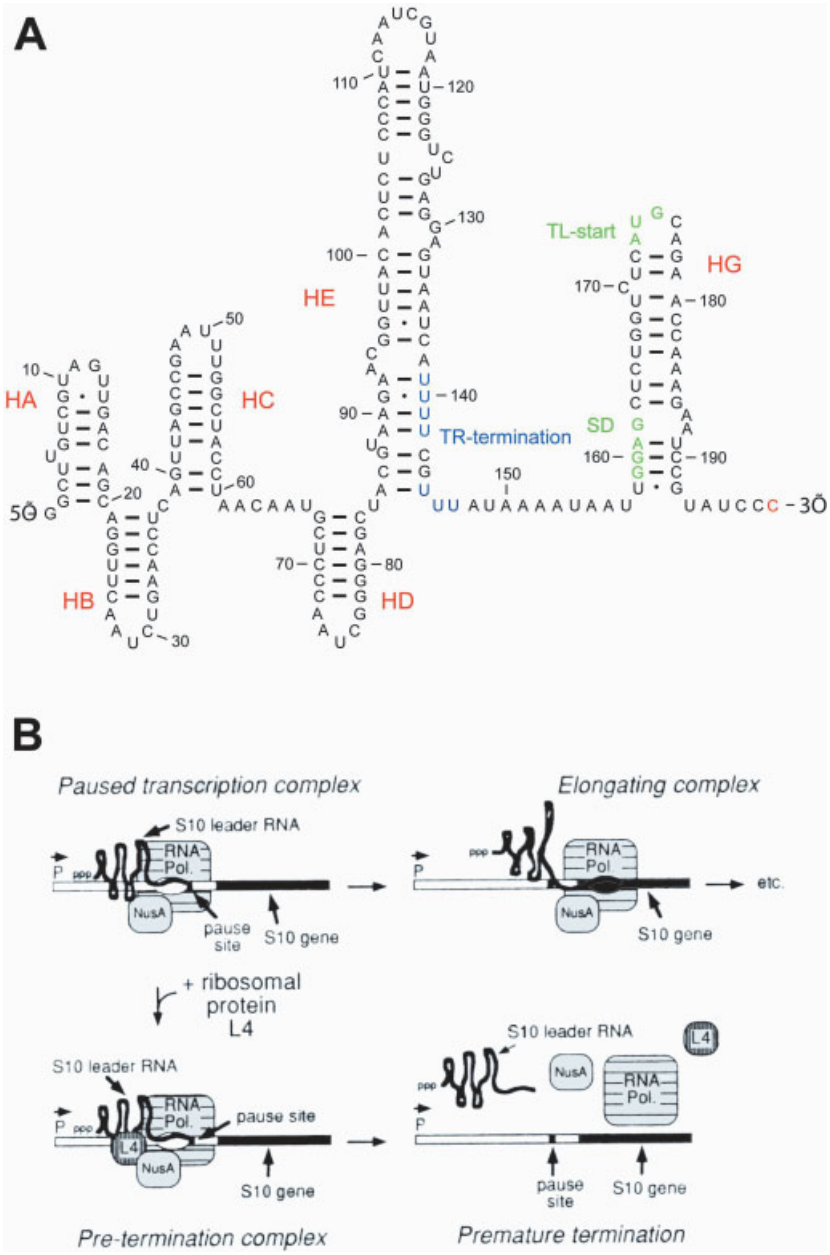


Figure 11.9 Transcriptional regulation of the S10 operon. (A) The secondary structure of leader sequence is illustrated with the L4-dependent transcriptional termination site colored blue, and the translational start site and SD sequence for the S10 gene colored green. (B) A model for L4-dependent transcriptional attenuation is illustrated. Panel A and B have been reproduced from Refs. [59] and [10], respectively.

competing with the ribosome for binding to the mRNA. Additionally, chemical probing experiments show no structural rearrangements in the mRNA leader sequence in the vicinity of the RBS [59] and therefore it is unlikely that L4 binding results in the RBS becoming sequestered in secondary structure. In this case, it is probable that L4 binding stabilizes the ribosome–mRNA complex in a pre-initiation state that is not competent to continue along the initiation pathways as described below for the α operon. The inhibition of translation initiation would block translation of the first gene in this operon, S10, whereas expression of the other downstream genes would be inhibited by disrupting translational coupling [76].

11.2.2.3 α operon

The α operon comprises four r-proteins (S13, S11, S4, and L17), as well as the gene encoding the α-subunit of the RNAP (Fig. 11.7). Expression of the r-proteins is regulated by S4 [77]. S4 binds to a nested pseudoknot structure (Fig. 11.10A) in the 5′ leader sequence upstream of S13 [56, 78]. This pseudoknot structure is believed to exist in two folded states, one that is active and one that is inactive for formation of a ternary initiation complex (Fig. 11.10B; Refs. [79, 80]). In the scheme presented in Fig. 11.10B, the 30S subunit is capable of binding both forms with equal affinity, whereas S4 binds and stabilizes the inactive form driving the equilibrium towards the inactive state [81]. In the absence of free S4, the 30S is capable of binding the mRNA in either state and eventually form an initiation complex, since the active and inactive forms are inter-convertible (Fig. 11.10B). The S4-bound inactive mRNA is free to interact with the 30S subunit (i.e., it does not prevent 30S binding), but the 30S-mRNA–S4 complex cannot be converted to an active state that is capable of progressing through the initiation pathway and bind fMet-tRNA^{fMet} (Fig. 11.10B; Ref. [81]). Therefore, translation of S13 is blocked by 'entrapping' the 30S subunit on the RBS in a non-productive state and thereby prevents subsequent initiation attempts by a new incoming 30S subunit [81]. Spedding and Draper [79] state that "an advantage of an entrapment mechanism is that it does not demand that the repressor bind tightly enough to displace the ribosome, which has substantial affinity for the mRNA".

11.2.2.4 *str* operon

This operon is named *str* operon, since a mutation in the first gene encoding S12 (*rpsL*) can confer resistance against the antibiotic streptomycin (see Chap. 12). The *str* operon comprises genes for, in the order, S12, S7, elongation factor G, and one of the two genes coding for elongation factor Tu (*tufA*). The repressor protein is S7, which binds in front of its own cistron after the S12 cistron. The expression of S12 is also regulated by S7, obviously by "retroregulation" similar to the regulation of L14 and L24 syntheses (see Sect. 11.2.2.1), whereas overproduction of S7 partially represses EF-G synthesis and that of EF-Tu only weakly. Expression of S7 is translationally coupled to the synthesis of the preceding S12, but interestingly S12 is also expressed independent of S7, which is depressing exclusively the coupled translation of both

proteins [82]. The fact that coupled and independent translation follows distinct mechanisms is convincing evidence that for translational coupling the ribosome translating the preceding cistron is continuing the synthesis of the downstream cistron without mixing with the pool of free ribosomal subunits [82, 83].

11.2.2.5 **IF3 operon**

IF3 operon contains three genes coding for IF3, and ribosomal proteins L35 and L20. L20 can act as a repressor of the translation of both the cistron encoding L35 and its own cistron by translational coupling. L20-mediated repression requires a long base-pairing interaction of its polycistronic mRNA, namely between nucleotide residues within the IF3 cistron and residues just upstream of the L35 cistron. This interaction results in the formation of a pseudoknot. Springer and co-workers [62, 84]

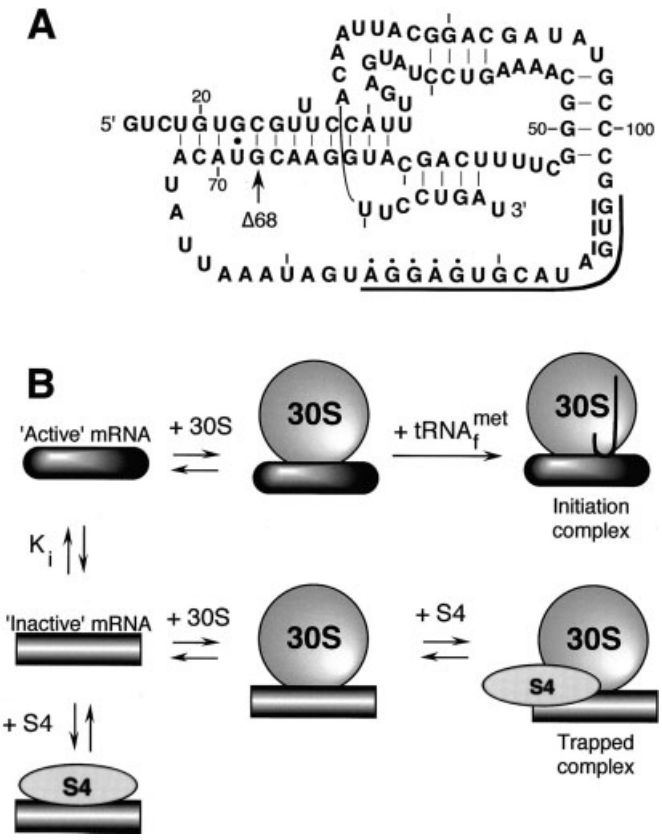


Figure 11.10 Regulation of the α-operon by S4. (A) The secondary structure of the nested pseudoknot structure in the 5′ leader sequence of the α-operon is illustrated. The RBS is underlined with the SD indicated with *dots* and the GUG initiation codon marked with *dashes*. (B) A schematic showing an 'entrapment' model for regulation of the α-operon. This figure has been reproduced from Ref. [81].

showed that L20 causes protection of nucleotide residues in two regions *in vitro*: the first region is the pseudoknot itself and the second lies in an irregular stem located upstream of the L35 cistron (Fig. 11.11). Both regions bind independently to L20 *in vitro*, and mutation and deletion studies demonstrated that they are essential for repression *in vivo*. Both sites are similar to the L20-binding site on 23S rRNA. This observation suggests that L20 recognizes its mRNA and its binding site on the 23S rRNA in a similar way.

11.3 Conclusion

These few examples demonstrate that the regulation of ribosomal proteins (1) depends on the synthesis of rRNA, (2) follows a general scheme whereby the repressor protein binds to its own mRNA and stabilizes one of two conformers, which prevents the formation of a productive 30S *de novo* initiation, and (3) that the recognition mechanism that the repressor protein uses to bind to its own mRNA is unique in every case, as are the protein–protein interactions. Furthermore, the polycistronic mRNAs coding for ribosomal proteins are distinct in that the cistron blocked by the repressor protein initiates via the canonical 30S *de novo* initiation, whereas the downstream cistrons are translationally coupled, as characterized by a 70S initiation type, where the ribosome translating the preceding cistron melts a secondary structure that sequesters the ribosomal binding site of the downstream cistron (see Sect. 3.1).

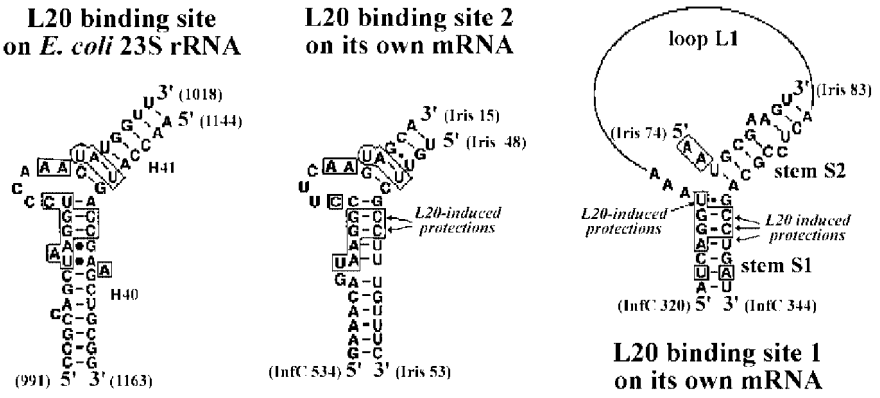


Figure 11.11 Secondary-structure similarities between the L20-binding site on the *E. coli* 23S rRNA and on the *rpm*I mRNA. The putative L20 binding site on *E. coli* 23S rRNA was deduced from the L20-binding site on 23S rRNA in the large ribosomal subunit of *D. radiodurans*. Thin lines and small dots indicate canonical and G+U base pairings, respectively; non-canonical base-pairings are indicated with large dots. Numbering of the terminal nucleotide residues on each strand of L20-binding site on 23S rRNA is that of *E. coli* 23S rRNA. Regions of sequence similarity between the L20-binding site on *E. coli* 23S rRNA and its own mRNA are boxed. Nucleotide residues that contain phosphate groups protected by L20 in iodine footprinting experiments are indicated by black arrows. The relevant features of the pseudoknot structure of L20-binding site 1 (stems 1 and 2 and loop L1) are also indicated [62].

References

1 H. F. Noller, *Annu. Rev. Biochem.* **1984**, *53*, 119–162.
2 B. Wittmann-Liebold: in *Structure, Function and Genetics or Ribosomes*, eds B. Hardesty and G. Kramer, Springer, Berlin, Heidelberg, New York 1986, 326–361.
3 A. Tissières, J. Watson, D. Schlessinger et al., *J. Mol. Biol.* **1959**, *1*, 221–233.
4 H. Bremer, P. P. Dennis: in *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, eds F. Neihardt, J. Ingraham, K. Low, B. Magasanik, M. Achaechter and H. Umbarger, American Society for Microbiology, Washington, DC 1987, 1527–1542.
5 N. O. Kjeldgaard, K. Gausing: in *Ribosomes*, eds A. T. M. Nomura and P. Lengyel, Cold Spring Harbor Laboratory, Cold Spring Harbor 1974, 369–392.
6 G. Blobel, V. R. Potter, *J. Mol. Biol.* **1967**, *26*, 279–292.
7 A. Wada, Y. Yamazaki, N. Fujita et al., *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2657–2661.
8 M. Yamagishi, H. Matsushima, A. Wada et al., *EMBO J.* **1993**, *12*, 625–630.
9 C. G. Condon, C. Squires, C. L. Squires, *Microbiol. Rev.* **1995**, *59*, 523–645.
10 J. M. Zengel, L. Lindahl, *Prog. Nucl. Acid Res. Mol. Biol.* **1994**, *47*, 331–370.
11 J. Keener, M. Nomura: in *Escherichia Coli and Salmonella Typhimurium*, ed. F. C. Neidhardt, American Society for Microbiology, Washinghton, DC 1996, 1417–1431.
12 H. D. Murray, J. A. Appleman, R. L. Gourse, *J. Bacteriol.* **2003**, *185*, 28–34.
13 C. A. Hirvonen, W. Ross, C. E. Wozniak et al., *J. Bacteriol.* **2001**, *183*, 6305–6314.
14 W. Ross, A. Ernst, R. L. Gourse, *Genes Dev.* **2001**, *15*, 491–506.
15 K. Yasuno, T. Yamazaki, Y. Tanaka et al., *J. Mol. Biol.* **2001**, *306*, 213–225.
16 S. E. Aiyar, S. M. McLeod, W. Ross et al., *J. Mol. Biol.* **2002**, *316*, 501–516.
17 S. M. McLeod, S. E. Aiyar, R. L. Gourse et al., *J. Mol. Biol.* **2002**, *316*, 517–529.
18 A. J. Bokal, V. Ross, T. Gaal et al., *EMBO J.* **1997**, *16*, 154–162.
19 M. Falconi, A. Brandi, A. LaTeana et al., *Mol. Microbiol.* **1996**, *19*, 965–975.
20 S. Jinks-Robertson, R. L. Gourse, M. Nomura, *Cell* **1983**, *33*, 865–876.
21 M. S. Bartlett, R. L. Gourse, *J. Bacteriol.* **1994**, *176*, 5560–5564.
22 X. Zhang, P. Dennis, M. Ehrenberg et al., *Biochimie* **2002**, *84*, 981–996.
23 H. Bremer, P. P. Dennis: in *Escherichia Coli and Salmonella*, eds F. C. Neidhardt, R. C. III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikow, M. Riley, M. Schaechter and H. E. Umbarger, ASM Press, Washington, DC 1996, 1553–1569.
24 T. Gaal, M. S. Bartlett, W. Ross et al., *Science* **1997**, *278*, 2092–2097.
25 M. M. Barker, T. Gaal, C. A. Josaitis et al., *J. Mol. Biol.* **2001**, *305*, 673–688.
26 D. A. Schneider, R. L. Gourse, *J. Bacteriol.* **2003**, *185*, 6185–6191.
27 C. Petersen, L. B. Moller, *J. Biol. Chem.* **2000**, *275*, 3931–3935.
28 D. Schneider, W. Ross, R. Gourse, *Curr. Opin. Microbiol.* **2003**, *6*, 151–156.
29 R. Lazzarini, A. Dahlberg, *J. Biol. Chem.* **1971**, *246*, 420–429.
30 P. Dennis, M. Nomura, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 3819–3823.
31 M. Cashel, D. R. Gentry, V. J. Hernandez et al., in *Escherichia Coli and Salmonella Thyphimurium: Cellular and Molecular Biology*, eds F. C. Neidhardt, R. I. Curtis, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger, ASM Press, Washington, DC 1996, 1458–1496.

32 C. D. Yegian, G. S. Stent, E. M. Martin, *Proc. Natl. Acad. Sci. USA* **1966**, *55*, 839–846.
33 W. A. Haseltine, R. Block, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 1564–1568.
34 W. A. Haseltine, R. Block, W. Gilbert et al., *Nature* **1972**, *238*, 381–384.
35 J. Sy, F. Lipmann, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 306–309.
36 D. Chatterji, N. Fujita, A. Ishihama, *Genes Cells* **1998**, *3*, 279–287.
37 A. Travers, *Mol. Gen. Genet.* **1976**, *147*, 225–232.
38 A. J. J. van Ooyen, M. Gruber, P. Jorgensen, *Cell* **1976**, *8*, 123–128.
39 E. A. Heinemeyer, D. Richter, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 4180–4183.
40 G. S. Stent, S. Brenner, *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 2005–2014.
41 R. Block, W. Haseltine, *J. Biol. Chem.* **1975**, *250*, 1212–1217.
42 J. D. Friesen, N. P. Fiil, J. M. Parker et al., *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 3465–3469.
43 J. Parker, R. J. Watson, J. D. Friesen et al., *Mol. Gen. Genet.* **1976**, *144*, 111–114.
44 X. Yang, E. Ishiguro, *J. Bacteriol.* **2001**, *183*, 6532–6537.
45 S. Ramagopal, B. D. Davis, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 820–824.
46 D. Richter, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 707–711.
47 D. Richter, P. Nowak, U. Kleinert, *Biochemistry* **1975**, *14*, 4414–4420.
48 E. G. Wagner, C. G. Kurland, *Biochemistry* **1980**, *19*, 1234–1240.
49 T. M. Wendrich, G. Blaha, D. N. Wilson et al., *Mol. Cell* **2002**, *10*, 779–788.
50 R. Wagner, *Encycl. Life Sci.* 1999. www.els.net
51 L. Lindahl, F. Sor, R. H. Archer et al., *Biochim. Biophys. Acta. Biochimica et Biophysica Acta* **1990**, *1050*, 337–342.
52 D. P. Cerretti, D. Dean, G. R. Davis et al., *Nucleic Acids Res.* **1983**, *11*, 2599–2616.
53 M. Nomura, J. L. Yates, D. Dean et al., *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 7084–7088.
54 J. M. Zengel, D. Mueckl, L. Lindahl, *Cell* **1980**, *21*, 523–535.
55 M. Johnsen, T. Christensen, P. P. Denis et al., *EMBO J.* **1982**, *1*, 999–1004.
56 I. C. Deckman, D. E. Draper, *Biochemistry* **1985**, *24*, 7860–7865.
57 R. J. Gregory, P. B. F. Cahill, D. L. Thurlow et al., *J. Mol. Biol.* **1988**, *204*, 295–307.
58 C. Philippe, C. Portier, M. Mougel et al., *J. Mol. Biol.* **1990**, *211*, 415–426.
59 U. Stelzl, J. M. Zengel, M. Tovbina et al., *J. Biol. Chem.* **2003**, *278*, 28237–28245.
60 B. Said, J. R. Cole, M. Nomura, *Nucleic Acids Res.* **1988**, *16*, 10529–10545.
61 A. Serganov, E. Ennifar, C. Portier et al., *J. Mol. Biol.* **2002**, *320*, 963–978.
62 M. Guillier, F. Allemand, S. Raibaud et al., *RNA* **2002**, *8*, 878–889.
63 H. U. Petersen, A. Danchin, M. Grunberg-Manago, *Biochemistry* **1976**, *15*, 1357.
64 H. U. Petersen, A. Danchin, M. Grunberg-Manago, *Biochemistry* **1976**, *15*, 1362–1369.
65 L. Mattheakis, L. Vu, F. Sor et al., *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 448–452.
66 D. P. Cerretti, L. C. Mattheakis, K. R. Kearney et al., *J. Mol. Biol.* **1988**, *204*, 309–329.
67 M. Mougel, F. Eyermann, E. Westhof et al., *J. Mol. Biol.* **1987**, *198*, 91–107.
68 J. L. Yates, M. Nomura, *Cell* **1980**, *21*, 517–522.
69 L. P. Freedman, J. M. Zengel, R. H. Archer et al., *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6516–6520.
70 X. Li, L. Lindahl, J. M. Zengel, *RNA* **1996**, *2*, 24–37.
71 Y. Sha, L. Lindahl, J. M. Zengel, *J. Mol. Biol.* **1995**, *245*, 486–498.
72 J. M. Zengel, L. Lindahl, *J. Bacteriol.* **1996**, *178*, 2383–2387.
73 J. M. Zengel, L. Lindahl, *J. Mol. Biol.* **1990**, *213*, 67–78.
74 J. M. Zengel, L. Lindahl, *Biochimie* **1991**, *73*, 719–727.
75 J. M. Zengel, L. Lindahl, *Genes Dev.* **1992**, *6*, 2655–2662.

76 L. Lindahl, R. H. Archer, J. R. McCormick et al., *J. Bacteriol.* **1989,** *171,* 2639–2645.

77 M. S. Thomas, D. M. Bedwell, M. Nomura, *J. Mol. Biol.* **1987,** *196,* 333–345.

78 C. K. Tang, D. E. Draper, *Cell* **1989,** *57,* 531–536.

79 G. Spedding, D. E. Draper, *Proc. Natl. Acad. Sci. USA* **1993,** *90,* 4399–4403.

80 G. Spedding, T. C. Gluick, D. E. Draper, *J. Mol. Biol.* **1993,** *229,* 609–622.

81 P. J. Schlax, K. A. Xavier, T. C. Gluick et al., *J. Biol. Chem.* **2001,** *276,* 38494–38501.

82 K. Saito, L. C. Mattheakis, M. Nomura, *J. Mol. Biol.* **1994,** *235,* 111–124.

83 K. Saito, M. Nomura, *J. Mol. Biol.* **1994,** *235,* 125–139.

84 C. Chiaruttini, M. Milet, M. de Smit et al., *Biochimie* **1996,** *78,* 555–567.

85 Q. She, R. K. Singh, F. Confalonieri et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 7835–7840.

86 D. E. Culham, R. N. Nazar, *Mol. Gen. Genet.* **1988,** *212,* 382–385.

87 R. K. Hartmann, V. A. Erdmann, *J. Bacteriol.* **1989,** *171,* 2933–2941.

88 A. Ruepp, W. Graml, M. L. Santos-Martinez et al., *Nature* **2000,** *407,* 508–513.

89 M. Fukunaga, I. Mifuchi, *J. Bacteriol.* **1989,** *171,* 5763–5767.

90 K. H. Nierhaus, *Biochemistry* **1990,** *29,* 4997–5008.

12
Antibiotics and the Inhibition of Ribosome Function

Daniel N. Wilson

12.1
Introduction

The protein synthetic machinery is a highly complex apparatus that offers many potential sites for functional interference and therefore represents a major target for antibiotics. The knowledge of the structure and function of the ribosome and associated translation factors has progressed enormously in the past five years, which has, in turn, accelerated our understanding of the mechanism of drug action. Conversely, drugs have been used as tools to probe the translation cycle thus providing a means to dissect further the multitude of steps involved in protein synthesis. In an era where bacteria are showing an ever-increasing resistance to many clinically relevant antibiotics the importance to understand their mechanism of inhibition is essential to the development of novel and more effective replacements. Here we attempt to provide a summary of the current understanding of how antibiotics functionally disrupt translation, with an emphasis on antibiotics that have been well-characterized biochemically, and in particular, structurally.

12.1.1
The Inhibition of Protein Synthesis in Bacteria

The translation machinery ensures accurate conversion of the genetic information of the messenger RNA (mRNA) into the corresponding polypeptide sequence. The ribosome provides the platform on which the mRNA can be decoded by transfer RNAs (tRNAs). Each tRNA carries a specific amino acid, which is faithfully incorporated into the growing polypeptide chain. Three tRNA-binding sites exist on the ribosome: The *A-site* is the site at which decoding occurs; here the correct *aminoacyl-tRNA* (aa-tRNA) is selected on the basis of the mRNA codon displayed at this site. Before peptide-bond formation, the *P-site* carries the *peptidyl-tRNA*, the tRNA bearing the elongating polypeptide chain. The *E-site* binds exclusively deacylated tRNAs, i.e., those tRNAs that having undergone peptide-bond formation are ready to *exit* from the ribosome (see Chap. 6 for details).

Protein synthesis can be divided into three distinct phases (Fig. 12.1). The initiation phase results in the necessary binding of the first or initiator tRNA to the P-site

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

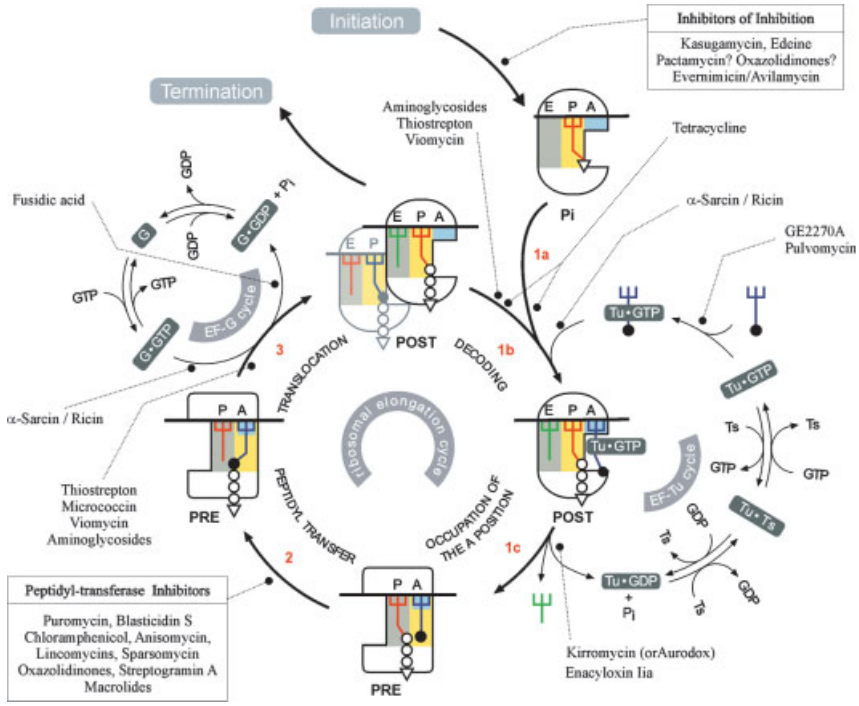


Figure 12.1 Inhibition of the initiation and elongation phases of translation by antibiotics. This figure was updated and modified from [64].

of the ribosome (Chap. 7). The second phase of translation, elongation, involves the movement of tRNAs in a cyclic fashion through the three binding sites on the ribosome, where the number of cycles is dictated by the length of the polypeptide being synthesized (Chap. 8). The first step in the cycle involves binding of the aa-tRNA to the A-site, which is facilitated by a protein factor EF-Tu. During subsequent elongation cycles, binding of the aa-tRNA to the A-site releases the E-site tRNA, maintaining two tRNA per ribosome at any given stage. Peptide-bond formation proceeds, transferring the entire polypeptide chain from the peptidyl-tRNA in the P-site to the aminoacyl moiety of the A-site tRNA. Now the ribosome has a peptidyl-tRNA at the A-site and a deacylated-tRNA at the P-site, a situation that is restored by a process termed translocation. The translocation reaction is catalysed by a second elongation factor, EF-G, and returns the *peptidyl-tRNA* to the *P-site* (although the peptidyl moiety is now extended by one amino acid) and the *deacylated tRNA* to the *E-site* – the outcome being that the *A-site* is now free to bind the next *aa-tRNA*. Thus, in the course of an elongation cycle, the ribosome can be thought to oscillate between two functional states; the pre- (PRE) and post-translocation (POST) states. This oscillation continues until a stop signal in the mRNA

enters the A-site triggering the third and final phase, termination. Stop signals are not generally decoded by tRNAs, but are recognized by protein termination factors, which function to hydrolyse the peptidyl-tRNA bond and release the translated polypeptide from the ribosome (termination is discussed in more detail in Chap. 9).

Every step described above provides a potential target for antibiotics. Indeed, antibiotics have been discovered that target almost every step of translation (as illustrated in Fig. 12.1), although with differing degrees of specificity. Antibiotics are defined as low-molecular-weight metabolic products, usually below 2 kDa, which are produced by microorganisms and inhibit at low concentrations (typically in the μM range) the growth of other microorganisms. The term antibiotic is used here in the broader sense to encompass non-natural chemical compounds that exhibit inhibitory effects against particular microorganisms, such as synthetic and semi-synthetic compounds.

Studies into translation as a target for antibiotics are not recent undertakings and were well advanced early on, as exemplified by the 100-page review of Gale et al. [1] in the early 1970s. Since then few new antibiotics have been uncovered, although numerous "second- and third-generation" derivatives have been developed that are much more potent than their "first-generation" forefathers. However, our understanding of antibiotic action has benefited recently from the exponentially increasing knowledge and understanding of the ribosome itself. Atomic resolution structures for the small and large subunits alone and in complex with ribosomal-substrate-mimics, reveal that the functional centers on the respective subunits are composed almost entirely of rRNA (reviewed in Refs. [2–6]). Thus, the importance of rRNA seems paramount for function, whereas ribosomal proteins may have been acquired somewhat later to "fine-tune" the process.

This is further emphasized by the structures of many small- and large subunit-antibiotic complexes (summarized in Table 12.1; reviewed in Refs. [7, 8]), which have revealed that the binding sites of antibiotics targeting the ribosome are in most cases composed entirely of rRNA, i.e., there is little or no interaction with ribosomal proteins. This is in accord with the observation that antibiotics generally target the functional centers of the ribosome. Bearing in mind the complexity of the ribosome, it is surprising that antibiotics target such a limited number of sites on the ribosome – a generous approximation would be that the antibiotics target 1% of the total volume of the ribosome (Figs. 12.2A and B). In any case, these ribosome–antibiotic complexes have furthered our understanding of both ribosome function and antibiotic inhibition mechanisms enormously. They will be analyzed in more detail as we follow the process of translation, chronologically with respect to the reaction sequence of a synthesizing ribosome. For each of the antibiotics discussed, the chemical structures are presented in Appendices A–F in the order they are discussed in.

Table 12.1 Compilation of pdb files for the antibiotic–ribosome structures

Antibiotic	Class	Species[a]	Res.[b] (Å)	Pdb[c] ID	Reference
Small subunit					
Edeine A	Edeine	T.t	4.5	1I95	24
Compound 2	Aminoglycoside		2.5	1O9M	235
Gentamycin C1a[d]	Aminoglycoside			1BYJ	74
Geneticin[d]	Aminoglycoside		2.4	1MWL	75
Hygromycin B	Aminoglycoside	T.t	3.3	HNZ	29
Pactamycin	Pactamycin	T.t	3.4	1HNX	29
Paromomycin[d]	Aminoglycoside			1A3M	236, 70
Paromomycin[d]	Aminoglycoside		2.5	1J7T	72
Paromomycin	Aminoglycoside	T.t	3.3, 3.0	1IBK, 1FJG[e]	67, 68
Paromomycin + ASL^{Phe}/U_6-mRNA	Aminoglycoside	T.t	3.11	1IBL	68
Paromomycin + ASL^{leu2}/U_6-mRNA	Aminoglycoside	T.t	3.0	1N32	66
Paromomycin + ASL^{Ser}/U_6-mRNA	Aminoglycoside	T.t	3.35	1N33	66
Spectinomycin		T.t	3.0	1FJG[e]	67
Streptomycin	Aminoglycoside	T.t	3.0	1FJG[e]	67
Tetracycline	Tetracycline	T.t	3.4	1HNW	29
Tetracycline	Tetracycline	T.t	4.5	1I97	24
Tobramycin[d]	Aminoglycoside		2.5	1LC4	76
Large subunit					
ABT-773	Ketolide	D.r	3.5	1NWX	143
Anisomycin	PTF inhibitor	H.m	3.0	1K73	112
Azithromycin	Azalide	D.r	3.2	1NWY	143
Azithromycin	Azalide	H.m	3.2	1M1K	145
Blasticidin S	PTF inhibitor	H.m	3.0	1KC8	112
Carbomycin A	Macrolide	H.m	3.0	1K8A	145
Chloramphenicol	PTF inhibitor	D.r	3.5	1K01	127
Chloramphenicol	PTF inhibitor	H.m	3.0	1NJI	112
Clarithromycin	Macrolide	D.r	3.5	1K00	127
Clindamycin	Lincosamide	D.r	3.1	1JZX	127
Dalfopristin	Streptogramin A	D.r	3.4	1SM1	274
Erythromycin	Macrolide	D.r	3.5	1JZY	127
Puromycin[f]	PTF inhibitor	D.r	3.7	1NJ0	108
Puromycin[g]	PTF inhibitor	H.m	3.0	1FG0	104
Puromycin[h]	PTF inhibitor	H.m	3.2	1FFZ	104
Puromycin[i]	PTF inhibitor	H.m	3.1	1KQS	104
Quinupristin	Streptogramin B	D.r	3.4	1SM1	274

Antibiotic	Class	Species[a]	Res.[b] (Å)	Pdb[c] ID	Reference
Roxithromycin	Macrolide	D.r	3.8	1JZZ	127
Sparsomycin	PTF inhibitor	D.r	3.7	1NJN	108
Sparsomycin + ASM	PTF inhibitor	D.r	3.6	1NJM	108
Sparsomycin + CCA-pcb	PTF inhibitor	H.m	2.8	1M90	112, 106
Spiramycin	Macrolide	H.m	3.0	1KD1	145
Telithromycin	Ketolide	D.r	3.4	1P9X	144
Troleandromycin	Macrolide	D.r	3.4	1OND	142
Tylosin	Macrolide	H.m	3.0	1K9M	145
Virginiamycin M	Streptogramin A	H.m	3.0	1N8R	112

a T.t, D.r and H.m correspond to the thermophilic bacterium *T. thermophilus*, the radiation-resistant eubacterium *Deinnoccus radiodurans* and the archaebacterium *Haloarcula marismortui*, respectively.

b Res. value corresponds to the maximum resolution attained in the high-resolution bin. The resolution values for NMR investigations are not given.

c Protein data bank (pdb) files for each antibiotic complex can be downloaded at http://www.rcsb.org/pdb/ and easily viewed with rasmol (http://www.bernstein-plus-sons.com/software/rasmol_2.7.7/ [237]), swiss-pdb-viewer (http://www.expasy.ch/spdbv/ [238]) or VMD (http://www.ks.uiuc.edu/Research/vmd/ [239]).

d The structures for these antibiotics were solved in complex with an RNA fragment mimic of the decoding center of the 30S subunit.

e These antibiotics have the same pdb number since their structures were determined at the same time by soaking crystals in a solution containing a mixture of all three antibiotics, streptomycin, paromomycin and spectinomycin.

f Puromycin is in the form of ACC-puromycin

g Puromycin attached to a 13 bp minihelix and thus mimics a tyrosyl-tRNA acceptor stem

h Puromycin in the form of an analog of A-site aa-tRNA and P-site peptidyl-tRNA covalently linked by the tetrahedral carbonyl carbon intermediate during peptide-bond formation (Yarus inhibitor)

i The products of the PTF reaction where the A-site has CCA and the P-site contains puromycin in the form of CC–Puromycin–phenylalanine–caproic acid–biotin.

12.2 Inhibitors of Initiation

The initiation phase in bacteria differs significantly from that in eukaryotes (see Chap. 7); bacterial transcription and translation are coupled whereas they are compartmentalized into the nucleus and cytosol of the eukaryotic cell, respectively. This segregation is responsible for the increased complexity of events preceding translation initiation itself, these include such processes as mRNA splicing and transport. The complexity of eukaryotic translation initiation is also reflected by the

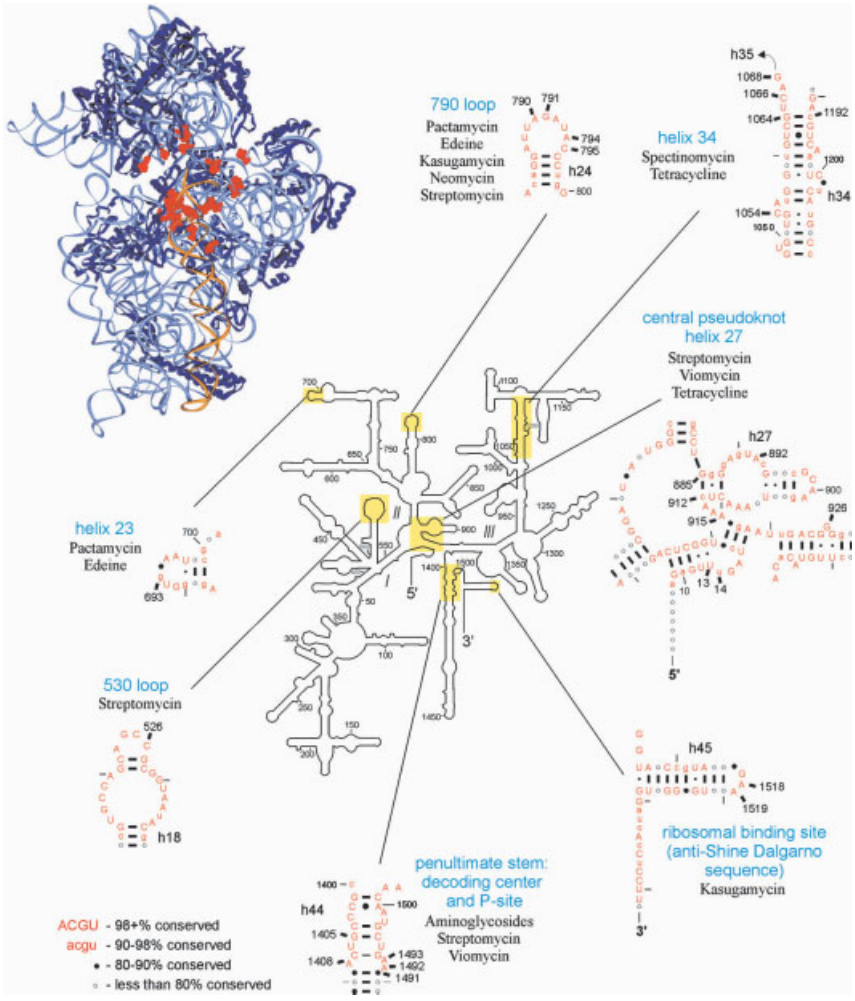


Figure 12.2

large number of protein factors required for this event, at least 13 have been identified so far, whereas in bacteria only three are necessary. In theory, one would think that this disparity could provide an ideal target for drugs, i.e., specific inhibitors of bacterial initiation, which should not have any adverse affects on eukaryotic cells. However, in practice, there are few antibiotics known today that *specifically* block initiation and those that have been identified seem to be universal inhibitors of translation initiation. Surprisingly, it is the drugs that target the conserved functional centers within the ribosome, such as the decoding site or the peptidyl transferase (PTF) center that exhibit differential effects across the three kingdoms (several reasons are discussed below).

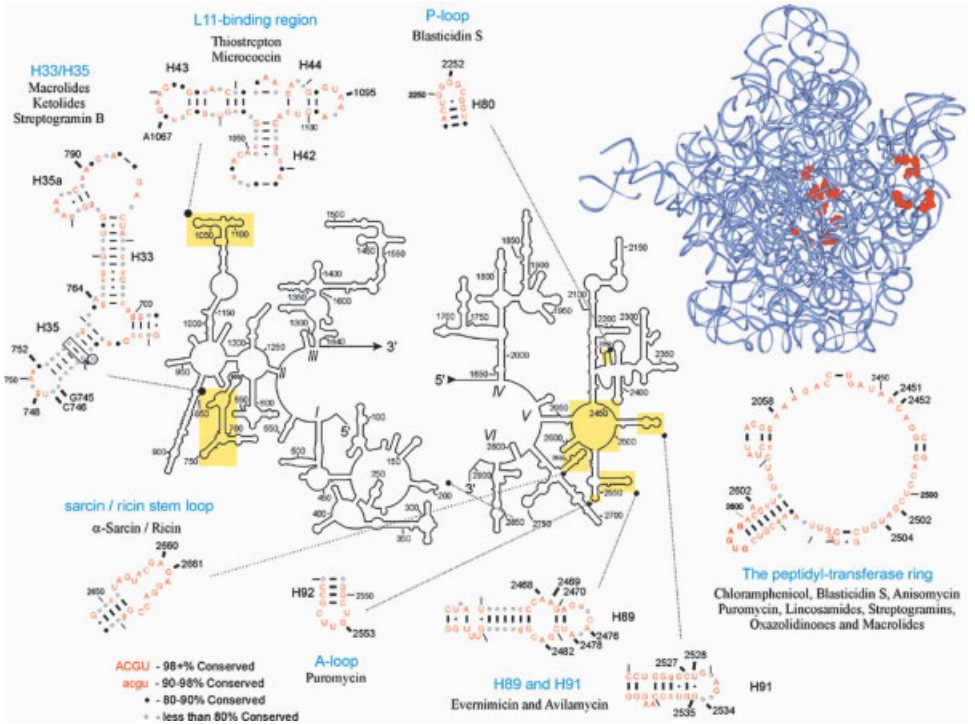


Figure 12.2 Regions of the small and large rRNA associated with various antibiotics. The (A) 16S rRNA and (B) 23S rRNA are schematically represented in the center of each figure, with regions of specific interest enlarged and bases associated with listed antibiotics indicated. The secondary structures were obtained from URL: http://www.rna.icmb.utexas.edu/ and illustrate the conservation in the 16S and 23S rRNA throughout the bacterial phyla, mapped on to the *E. coli* secondary structures [270].

Translation initiation by bacterial ribosomes operates through a pre-initiation complex, consisting of the small ribosomal subunit, mRNA, the initiator fMet-tRNA and three initiation factors, IF1, IF2, and IF3. Association with the large subunit releases the remaining IFs, leaving the initiator tRNA at the P-site, ready for A-site binding (reviewed in Ref. [9]).

There are at least five antibiotics that are commonly referred to as translation-initiation inhibitors. Kasugamycin and edeine both act to inhibit initiator tRNA binding to the 30S subunit but probably do so via different mechanisms. The exact step of initiation by pactamycin is unclear: early reports suggest that it allows initiator tRNA binding to the 30S subunit but prevents association of the pre-initiation complex with the 50S subunit, although recent analysis suggests that pactamycin is in fact a translocation inhibitor. The oligosaccharide antibiotics evernimicin and avilamycin bind the 50S subunit and appear to inhibit association of the pre-initiation complex with the antibiotic bound 50S subunit, thereby preventing 70S ribosome formation.

12.2.1 Kasugamycin

Kasugamycin (Ksg) is thought to inhibit the initiation phase of protein synthesis without affecting elongation [10]. By preventing the binding of fMet-tRNA^{fMet} to the prospective P-site on the 30S subunit, Ksg prevents the formation of the pre-initiation complex, although the exact mechanism by which this is accomplished remains to be determined.

In the early 1970s, resistance to Ksg was shown to arise from mutations of a particular gene, termed *ksgA* because of the resistance phenotype, later shown to encode a methylase responsible for post-transcriptional modification of two adenine residues near the 3′-end of 16S rRNA, namely the universally conserved bases A1518 and A1519 (*Escherichia coli* numbering is used throughout this chapter, unless otherwise indicated) [11–13]. These modifications are the only universally conserved modifications of the rRNAs. A recent study demonstrated that almost any base mutation at position A1519, but none at A1518, could confer resistance to Ksg [14]. This same study also identified two other universally conserved 16S rRNA positions A794 and G926 as conferring Ksg resistance, in agreement with earlier studies demonstrating that binding of Ksg to ribosomes protected these same bases from chemical probing (also the reactivity of C795 was found to be enhanced [15]; Fig. 12.2A). It seems unlikely that a single molecule of Ksg could contact all three regions simultaneously since G926 is located some 15–20 Å from A794 and A1519 in the 30S crystal structure, although contact with the latter two would be possible since they are only 6–7 Å apart (Figs. 12.3A and B).

Ksg does not inhibit translation of leaderless mRNAs at concentrations where translation of canonical mRNAs is abolished [16, 17]. (Canonical mRNAs in prokaryotes are those mRNAs that contain an AUG start codon, upstream of which, is the so-called Shine and Dalgarno (SD) sequence, a region of the mRNA complementary to part of the 3′-end of the 16S rRNA called the anti-SD. Base pairing between these two sequences is thought to help in positioning the start codon in the ribosomal P-site.)

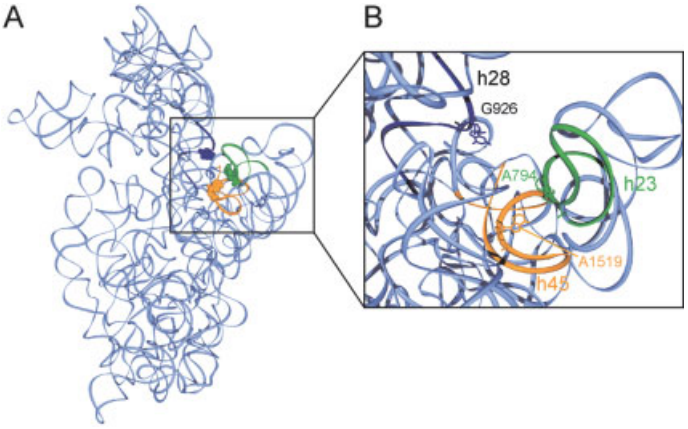


Figure 12.3 The nucleotides of the 16S rRNA associated with the antibiotic kasugamycin. (A) The rRNA of the *T. thermophilus* 30S subunit is shown as a ribbons representation (pale blue) with h28 (blue), the 790-loop (green) and the penultimate helix, h45 (yellow) highlighted. Bases associated with kasugamycin are shown as space fill representation and are detailed in B. (B) Close-up of bases associated with kasugamycin: G926 of h28 (blue), A794 of h24 (green) and A1519 of h45 (yellow). These figures were made from pdb 1fjf [271].

The implication here is that Ksg cannot inhibit P-site tRNA binding when there is no mRNA sequence upstream of the start codon. It is interesting to note that addition of Ksg could remove pre-bound initiator tRNA from 30S subunits, but not from 70S ribosomes [18]. However, although this is an attractive hypothesis, we note that Ksg had no effect on SD-anti-SD interactions nor did it influence the function of IF3 [17].

In addition, Ksg may affect the full maturation of the 30S subunit, since overexpression of the *KsgA* methylase (that when knocked out gives rise to Ksg^R) rescued a temperature-sensitive strain resulting from deletion of a small GTPase called Era [19]. Era has recently been implicated in a final step in the assembly of the 30S subunit, namely the processing of the precursor 17S rRNA to the mature 16S rRNA [20]. Intriguingly, 61S particles are formed in the presence of low concentrations of Ksg that contain normal 50S ribosomal subunits but reduced 30S particles. The latter contain 11 instead of 21 S-proteins and are competent in translating leaderless mRNAs, but not normal SD-sequence carrying mRNAs [21]. Clearly, determination of the Ksg binding site on the 30S subunit would provide much needed help to functional biochemists in elucidating the action of this antibiotic.

12.2.2 Edeine

Edeine A_1 (Ede) is the major active component of one of a number of isomers produced by the bacterium *Bacillus brevis* Vm4 and is composed of an N-terminal

β-tyrosine residue attached to a C-terminal spermidine-like moiety (see Appendix A1). Ede is effective *in vitro* but not *in vivo*. *In vitro* it blocks ribosomes derived from organisms of all kingdoms. Ede has been shown to inhibit specifically mRNA-directed binding of aa-tRNAs to both 30S subunits and 70S ribosomes [22]. Consistent with these observations, the footprints on the 16S rRNA corresponded exactly with those produced by a P-site bound tRNA [23, 15]. Recently, the crystal structure for Ede bound to the *Thermus thermophilus* 30S subunit revealed a single binding site for this antibiotic on the solvent side of the platform sandwiched between h24, h28, h44, and h45 [24]. Binding of Ede induces base-pair formation between C795 and G693 at the tips of h24a and h23b, respectively [24] (see Figs. 12.4A and B). Exactly these same residues were protected from chemical probing upon tRNA binding to the ribosomal P-site [23]. In light of the 70S crystal structure with tRNAs bound at A-, P- and E-sites [25], it is now clear that these bases are located in the E-site and not the P-site as first thought. The conclusion being that the protections most probably result from conformational changes in the rRNA upon tRNA binding and not through direct interaction with the P-tRNA. This suggests (i) that tRNA binding at the P-site requires an open conformation regarding the C795-G693 base-pair, and (ii) because Ede also induced conformational changes in this region, the

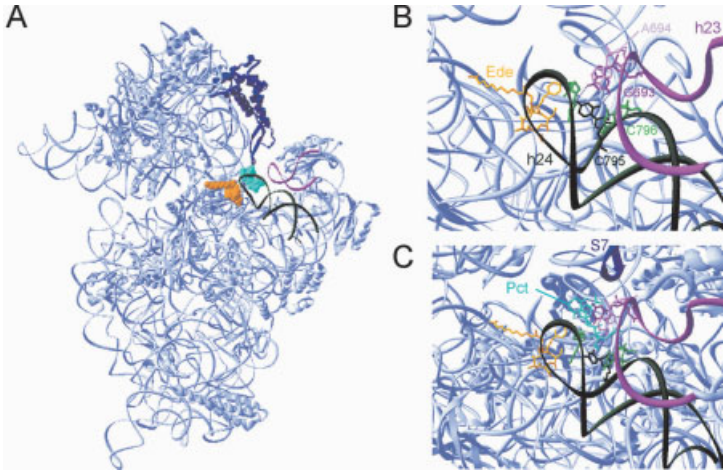


Figure 12.4 The binding site of edeine and pactamycin on the *T. thermophilus* 30S subunit. (A) The 16S rRNA is shown in ribbons (pale blue) with h23 (crimson) and h24 (dark green) and ribosomal protein S7 (blue) highlighted. The relative positions of edeine (yellow) and pactamycin (cyan) are shown with spacefill representations. (B) Close-up view of the G693-C795 base-pair (dark green and crimson, respectively) induced upon edeine (Ede, yellow) binding. Other bases indicated are A694 (pink) and C796 (pale green). Hydrogen bonding is indicated with a dashed blue line. (C) Close-up view of the pactamycin (Pct, cyan) bound between the tip of h23 and h24. The two distal rings of Pct can be seen stacking onto each other and the base of G693, whereas the third ring inserts into the crevice between the helices where it interacts with C795 and C796. All bases and helices are coloured as in (A) and (B). These figures were made from pdb files 1hnx [29] and 1i95[24].

mode of inhibition of Ede may be indirect, i.e., Ede binding may, by inducing the closed conformation (C795-G693 base-pair formation) and locking h23-h24 together, mimic a ribosome already containing a P-site tRNA and therefore prevent the associated conformational changes necessary for stable binding of tRNA to the P-site.

Furthermore, the observation that the predominant contacts between the P-site tRNA and the 30S subunit are due to codon–anticodon interactions [25, 272], suggests that the mechanism of action of Ede is to prevent interaction between codon and anticodon. This is supported by the observation that Ede inhibits the binding of the encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) to the ribosome but not that of the cricket paralysis virus (CrPV), since the former initiates out of the P-site using Met-tRNAi and eIF2-GTP, whereas the latter initiates out of the A-site [273].

One effect of Ede that had been previously overlooked was its ability to induce translational misreading, at levels comparable with those of the classic misreading antibiotic streptomycin [22]. This represents the first example of an antibiotic that induces misincorporation events from the E-site, supporting the link between E-site and translational fidelity proposed by the α–ε model for elongation (see Chap. 8.1). Consistently, the introduction of site-specific mutations within S7 or S11, both of which are located within the E-site connecting the head and platform, also severely reduced translational fidelity, promoting not only misreading but also frameshifting and nonsense suppression [26]. The site-specific mutations were located at the interface between S7 and S11 and are therefore involved in forming the channel through which the mRNA passes, suggesting that the loss of translational fidelity may operate partly by disturbing the path of the mRNA. Similarly, the base-pair induced by Ede binding would also be expected to disrupt the path of the mRNA through the E-site.

12.2.3 Pactamycin

Pactamycin (Pct) was isolated as a potential anti-tumor agent [27], but subsequently shown to be equally effective against intact cells of both bacteria and eukaryotes therefore limiting its clinical use. Although Pct was extensively studied during the 1960s and 1970s, the exact mode of Pct action remained unclear (reviewed in Ref. [28]). Pct is termed an initiation inhibitor since, under certain conditions Pct has been observed to cause an accumulation of putative pre-initiation complexes when in the presence of crude initiation factors. The conclusion from these experiments was that fMet-tRNA binding to the 30S subunit was not inhibited but that association of the pre-initiation 30S complex with the 50S subunit to form a 70S ribosome was prevented, possibly due to a non-functional P-tRNA orientation. However, these observations were inconsistent with reports where pre-initiation complexes associate to form 70S ribosomes although they were non-functional for translation. Furthermore, in eukaryotes, multiple studies observed accumulation of di- and tripeptides suggesting that Pct targeted, not initiation, but an early elongation step (see Ref. [28]).

The crystal structure of the 30S subunit from *T. thermophilus* in complex with Pct has been resolved to 3.4 Å [29]. A single binding site determined on the 30S subunit revealed that the two distal rings of Pct stack upon each other and with G693 at the tip of h23b of the 16S rRNA, whereas the central ring interacts with C795 and C796 in h24a [29] (Fig. 12.4C). Pct, bound in this position, mimics a dinucleotide of the mRNA in the E-site leading to the proposal that Pct disrupts the path of the mRNA through the ribosome. The implication being that translocation of tRNAs into the E-site may be prevented or may cause the tRNAs to drop off the ribosome.

To understand further the target of Pct inhibition, a systematic study analyzing the effect of Pct on each step of initiation and elongation was performed recently [22]. Surprisingly, Pct exhibited no inhibitory effect during the initiation stage of translation, subunit association nor on A-site binding. The site of action seemed to be translocation of the A and P-tRNA to the P and E-sites. What was unusual was that translocation inhibition was markedly influenced by the tRNA species, such that translocation of Met-tRNA, Val-tRNA or Lys-tRNA was significantly inhibited yet when Phe-tRNA was present little or no inhibition was observed. This was then shown to be consistent with the lack of inhibition of Pct on poly(U)-dependent poly(Phe) synthesis but the severe inhibition of Pct on poly(A)-dependent poly(Lys). In conclusion, these results suggest that Pct should not be referred to as an initiation inhibitor, but in fact as an early elongation or translocation inhibitor.

Since the binding site of Pct maintains the C795 and G693 bases in an open conformation, whereas the presence of Ede induces a closed conformation, viz. a Watson–Crick base-pair, the relationship between these two antibiotics was investigated [22]. This study could demonstrate that the inhibition by Ede of fMet-tRNA binding to 30S subunits could be relieved with increasing concentrations of Pct. Similarly, the inhibition of binding of AcPhe-tRNA to the P-site of 70S ribosomes by Ede could also be alleviated by addition of Pct. This led to the proposal that P-site tRNA binding is regulated by the conformation of the C795-G693 basepair and illustrates how studies into antibiotic action yield insight into fundamental mechanisms of ribosome function.

12.2.4 Evernimicin and Avilamycin

The orthosomycins evernimicin (Evn) and avilamycin (Avn) are oligosaccharide antibiotics that exhibit excellent activity against a broad range of Gram-positive bacteria. Evernimicin (SCH27899) was isolated from *Micromonospora carbonaceae* and was trialed as therapeutic agent by Schering-Plough under the name Ziracin. Resistance to these antibiotics has resulted from mutations in ribosomal protein L16 [30–33] and in H89 and H91 of the 23S rRNA [34–36], suggesting that both antibiotics bind the large ribosomal subunit. In agreement, chemical footprinting of Avn on the 23S rRNA identified residues A2482 in H89 and A2534 in H91 [36] whereas the same technique identified these bases, amongst others, within the same helices using Evn [35]. Furthermore, position G2470 is methylated by EmtA,

a methyltransferase – the effect of which is to confer resistance to both Avn and Evn [37] (see Table 12.2 and Fig. 12.2B). The locations of H89, H91 and L16 are in close proximity to each other in the crystal structures of the large subunit [38, 39], being located towards the base of the ribosomal stalk, some distance from the tunnel and PTF center – the "hotspots" for antibiotic interference on the 50S subunit (Figs. 12.5A and B). These are the only known antibiotics to interact with this region of the ribosome, thus explaining the previous observations that bacteria resistant to numerous other antibiotics show no cross-resistance to Evn or Avn as well as the observation that other ribosomal antibiotics do not compete with Evn for ribosome binding [40].

Interestingly, H89 has also been associated with initiation factor IF2, namely, the protection of positions A1476 and A2478 by IF2 from chemical probing when the modifying agent dimethylsulfate was used [41]. Belova et al. [35] proposed that Evn specifically inhibited IF2-dependent formation of a 70S initiation complex, but since the assay that was employed monitors only the transfer of f[^{3}H]Met to puromycin, it was not possible to determine whether Evn actually inhibits subunit association or whether association occurs, but the fMet is not in the correct orientation for transfer to puromycin. The former seems most probable given that *in vivo* brief incubation of bacterial cells with Evn reduced the amount of 70S ribosomes. In agreement with

Table 12.2 Evernimicin- and avilamycin-resistant mutations in L16 and 23S rRNA

Antibiotic	Ribosomal component	Mutation position	Detection[a]	Reference
Evn	L16	R51H, I52T, R56H	Spontaneous	33
Evn	L16	I52S, I52T, I52N, R51C	Spontaneous, engineered	31
Evn	L16	R51C, R51H	Spontaneous	32
Evn/Avn	L16	R56H, R56H, I52T, I52S	Spontaneous	30
Evn	23S rRNA	A2469C, C2480T, G2535A, G2536C	Spontaneous	34
Evn	23S rRNA	G2535A	Spontaneous	33
Evn	23S rRNA	A2471G, A2471C, A2478C, U2479C, C2480A, C2480U, G2527A, U2528C, and G2535A	Spontaneous, Two were also engineered	35
Evn	23S rRNA	A2468, A2469, A2476, A2478, A2482 in H89 and A2534 was protected in H91	Footprinting	35
Avn	23S rRNA	G2470U, A2471G, G2472U, U2479C, C2480U	Spontaneous	36
Avn	23S rRNA	A2482 (H89) A2534 (H91)	Footprinting	36

a Although *E. coli* numbering is given for convenience, the spontaneous resistance mutations were generally detected in other organisms such as *Streptococcus pneumoniae, S. aureus and H. halobium.*

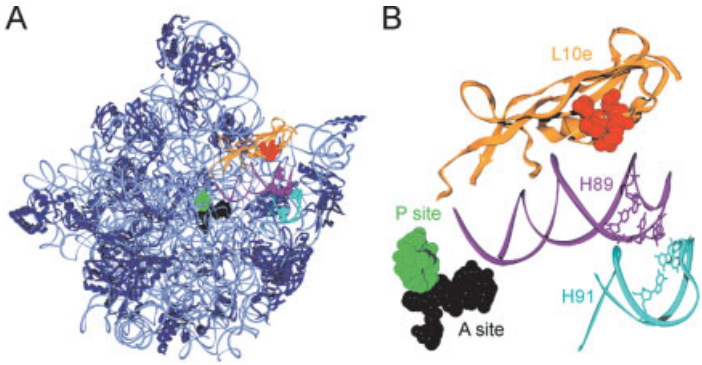


Figure 12.5 The putative binding site of Evernimicin and Avilamycin on the 50S subunit. (A) Regions associated with resistance to Evn and Avn are highlighted on the *H. marismortui* 50S subunit (PDB1KQS; [107]). Ribosomal RNA shown in ribbons with H89 (purple) and H91 (cyan) and ribosomal proteins coloured dark blue except L10e (L16 homolog) which is coloured yellow. Positions of L10e and H89/H91 associated with resistance to Evn or Avn are indicated in red spacefill. The products of the PTF reaction in the A site (dark green) and P site (light green) are provided as a reference for the PTF center. (B) Close-up of (A), showing only H89, H91, L10e and, as a positional reference, the PTF products bound at the PTF centre. Coloured as in (A), except here the bases of H89 and H91 that are associated with resistance to Evn/Avn are coloured in purple and cyan respectively.

this proposal, *in vitro* assays with Evn did not inhibit peptidyl-transferase activity in the absence of IF2 nor did Evn prevent pre-initiation complex formation [35].

It should also be noted that Evn has also been shown to inhibit specifically assembly of the 50S subunit. However, this assembly inhibition required a 50% inhibitory dose some 10 times higher than that required to inhibit protein synthesis indicating that ribosome function is the primary target of the drug [42]. There are numerous antibiotics that are much more efficient inhibitors of ribosome assembly and they are the subject of the next section.

12.2.5 Antibiotic Inhibitors of Ribosome Assembly

It is becoming clear that a large number of antibiotics that have been well characterized as protein synthesis inhibitors, also double as inhibitors of ribosomal subunit assembly (reviewed in Ref. [43]). In the past few years the list has been growing rapidly and now includes representatives of the entire MLS_B class of 50S subunit inhibitors, i.e., macrolides, lincosamides and streptogramin B compounds [42], as well as many antibiotics that target the 30S subunit, such as members of the aminoglycoside family [44]. A number of clinical relevant antibiotics fall into these categories

emphasizing the importance of understanding the dual action of these drugs to develop more effective inhibitors; examples include both 14-/15- (erythromycin and semi-synthetic derivatives azithromycin, clarithromycin, and roxithromycin) and 16-membered macrolides (tylosin and spiramycin), the ketolide antibiotics, the lincosamides lincomycin and clindamycin, the oxazolidone linezolid and also the medically important aminogylcosides (see later sections for more information on these antibiotics).

The importance of the inhibition of subunit assembly is determined by measuring the effect of each antibiotic on both protein synthesis and subunit assembly over a range of antibiotic concentrations. In almost all cases the IC_{50} for translation inhibition was precisely half the IC_{50} required for blocking subunit assembly (see Table 12.3; [43]), suggesting that the inhibitory effects on translation and subunit assembly are equivalent. The reasoning is that since subunit assembly is a prerequisite for active ribosomes, an IC_{50} for subunit assembly will reduce the number of translationally active ribosomes by half and therefore the IC_{50} for translation will be exactly half that observed for subunit assembly. The exception was evernimicin, which, as mentioned previously, required a 13-fold higher concentration to inhibit subunit assembly than that required for translation inhibition.

Furthermore, most of the antibiotics tested specifically inhibited assembly of only one ribosomal subunit, such that antibiotics that inhibited 50S subunit formation exhibited no inhibition on 30S subunit assembly and *vice versa*. Subunit inhibition also correlated with the known inhibitory action on translation of each antibiotic, for example, the macrolide erythromycin, which has been shown to obstruct the progression of the nascent chain by binding within the tunnel of the 50S subunit (see Sect. 12.3.2.4), inhibited 50S subunit formation without influencing 30S subunit formation [45]. Likewise, the aminoglycoside neomycin and a closely related derivative paramomycin (both of which induce translational misreading by binding within the decoding center of the 30S subunit; see Sect. 12.3.1.2), were shown *in vivo* to reduce 30S subunit formation without effecting 50S subunit assembly [44]. In contrast, antibiotics, such as chloramphenicol, a potent peptidyl-transferase inhibitor, were shown to inhibit formation of both ribosomal subunits in a non-specific fashion. Although this is also true of streptogramin A class (e.g. virginiamycin M and

Table 12.3 Antibiotics IC_{50} for translation and subunit assembly [43]

	Half-inhibitory concentration IC_{50} ($\mu g\ ml^{-1}$)	
Antibiotic	**Translation**	**Assembly**
Erythromycin	0.17	0.36
Azithromycin	2.5	5.0
Clarithromycin	0.075	0.15
ABT773	0.02	0.035
Telithromycin	0.04	0.08
Evernimicin	0.03	0.4
Linezolid	0.3	0.6
TAN1057A	4.5	9.0

CP36926), the streptogramin B antibiotics (e.g. pristinamycin I_A, virginiamycin S and CP37277) inhibit only 50S subunit formation – a difference not necessarily unexpected since the streptogramin A and B class are not structurally related.

Recently, experiments were undertaken to determine the mechanism of inhibition on subunit assembly [46]. Cells were treated with radiolabelled erythromycin and then the 70S ribosomes and subunits were separated from one another on a sucrose gradient and their radioactive content determined. The results revealed that erythromycin bound both the mature form of the 50S subunit as expected but also an assembly intermediate found to contain the 23S and 5S rRNAs and 18 of the 34 ribosomal proteins found in the *E. coli* 50S subunit. Importantly, the binding stoichiometry of each interaction was 1:1. Therefore, it seems likely that during an early stage in the assembly of the 50S subunit, a binding site is formed, perhaps the same or similar one to that in the mature subunit, to which erythromycin binds. Binding of erythromycin might prevent either conformational changes in the precursor particle, or the binding of additional ribosomal proteins, necessary for further assembly. This mode of inhibition is likely to be transferable to other antibiotics, especially other macrolides (and ketolides), and also to the structurally unrelated lincomycins, all of whose binding sites overlap to a large extent (see Sect. 12.3.2). Furthermore, 30S subunit assembly progresses through intermediates (see Sect. 3.1); therefore, by applying the same model for assembly inhibition, one would predict that small subunit assembly is stalled at a precursor stage, perhaps formation of the 21S particle.

12.3 Inhibitors of the Elongation Cycle

Following initiation, the ribosome is "primed" with an initiator tRNA at the P-site (fMet-tRNA in bacteria and Met-$tRNA_i$ in eukaryotes and archaea) and displays a codon at the A-site specific for one species of tRNA. (A species of tRNAs is defined by the anticodon it bears as opposed to the amino acid. This is because there are a set of tRNAs that are charged with the same amino acid although they bear distinct anticodons. See Chap. 4.1 for more information.) This is a unique situation for a ribosome since it has only a single tRNA bound, i.e., the A and E-sites are free. Binding of a tRNA to the A-site in this situation is termed A-site occupation of the initiation type (i-type). In all subsequent rounds of elongation, the ribosome will always carry two tRNAs, either an A- and P-tRNA in the PRE or a P- and E-tRNA in the POST state. Binding of a tRNA to the A-site of the latter is termed A-site binding of the elongation type (e-type). For some antibiotics, this distinction is irrelevant, for example, tetracycline inhibits both states, but for others it is not, as exemplified by the differential inhibition observed by various antibiotics including aminoglycosides, thiostrepton and viomycin [47].

The elongation cycle can be thought of as the heart of protein synthesis and as such is the prime target of the majority of antibiotics identified to date. Because of the diversity and "ingenuity" of the action of the vast array of studied antibiotics, each step of inhibition will be subdivided into distinct categories for convenience,

although it is acknowledged that some antibiotics can be classified into more than one category.

12.3.1 Antibiotic Action and A-site Occupation

The first step of the elongation cycle involves A-site occupation. Binding of tRNA to the A-site can be separated into two consecutive steps: (i) An initial step involving the binding of the ternary complex aa-tRNA•EF-Tu•GTP to the ribosome, during which only parts of the anticodon stem-loop of the incoming tRNA is in the A-site proper (discussed in more detail in Chap. 8.2), and (ii) a second step involving the hydrolysis of GTP to GDP by EF-Tu and dissociation of EF-Tu from the ribosome, which in turn releases the CCA end of the tRNA enabling it to move into position on the 50S subunit. The outcome being that the tRNA is now fully "accommodated" into the A-site.

Antibiotics inhibit or impair A-site occupation in a number of ways. Tetracycline inhibits the conversion between the first and second binding steps, i.e., prevents full accommodation of the A-site tRNA, which results in the loss of the tRNA from the A-site. In contrast, streptomycin and the aminoglycoside family of antibiotics could be said to "encourage" binding of tRNAs to the A-site, regardless of whether they are cognate for the codon being displayed there. Finally, there is a subset of antibiotics that interfere with EF-Tu function in a significantly different way; for example, the ribotoxins sarcin and ricin prevent binding of EF-Tu to the ribosome, while the antibiotic kirromycin has the exact opposite effect to prevent EF-Tu dissociation.

12.3.1.1 Tetracycline: An Inhibitor of A-site Occupation

Although tetracycline (Tet) was introduced into medicine as early as 1948, it is only recently that the inhibitory mode of action of this antibiotic has become clear. During protein synthesis, Tet is an inhibitor of the elongation cycle, where it specifically prevents binding of tRNA to the A-site. Though non-enzymatic binding (without EF-Tu) of tRNA to the A-site is totally inhibited, the first step of tRNA binding in the form of the ternary complex aa-tRNA•EF-Tu•GTP is possible. However, upon hydrolysis of GTP to GDP by EF-Tu the incoming tRNA is lost from the ribosome.

Recently, two independent crystallography groups solved the structure of the 30S subunit in complex with Tet [29, 24]. In one of the studies two binding sites were identified [29], whereas the other identified six, two of which were equivalent to those of the first study [24] (Fig. 12.6A). Since the inhibitory effect of Tet is presumed to result from binding at a single high-affinity site on the ribosome the task now is to decide which of the identified sites is the biologically relevant inhibitory site. Mounting evidence suggests that the highest occupancy site in both structures is the relevant one and will be referred to hereafter as the primary binding site (Tet-1 in Ref. [24]). In contrast, it seems probable that the contribution made by the secondary (Tet-5 in Ref. [24]) and lesser-occupied sites to inhibition of A-site

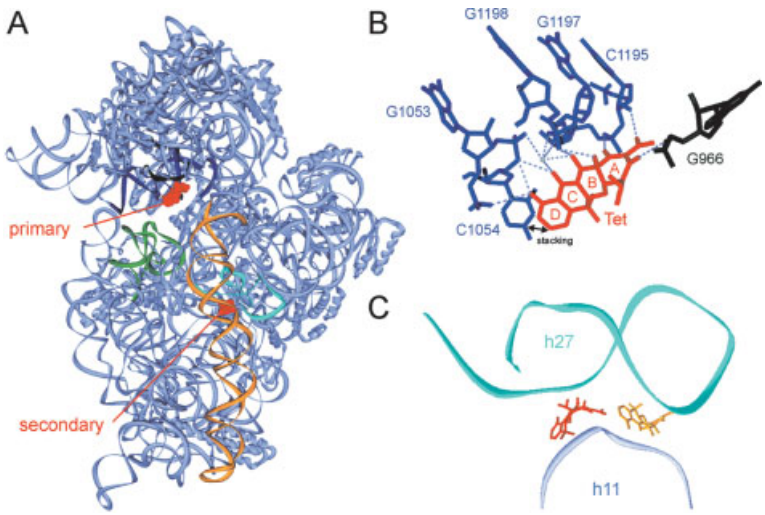


Figure 12.6 The primary and secondary tetracycline binding sites on the 30S subunit. (A) Overview of the primary and secondary tetracycline binding sites (pdb 1hnw) [29], with h34 (blue), h31 (dark green), h18 (light green) h27 (cyan) and h44 (yellow) high-lighted. (B) In the primary binding site the charged or polar face of Tet makes contact exclusively with the phosphate backbone of the 16S rRNA (positions G1053, C1054, 1195-1198 of h34 (blue) and 966 of h31 (green); note that position A965 and the base of U1196 are omitted for clarity), except for stacking interaction with the base of G1054. (C) The secondary Tet binding sites differ significantly between the two independent studies and are shown in red [29] and yellow [24] with h11 (light blue) and h27 (cyan).

binding are negligible, although they do explain the previously inexplicable and often contradictory data for Tet obtained via a variety of biochemical techniques.

In both structures, the primary binding site encompasses the irregular minor groove of h34 and the loop of h31, thus placing Tet directly in the decoding center of the A-site (Fig. 12.6A, B). From this position, Tet was proposed to interfere sterically with a tRNA bound at the A-site. Since the binding mode of the aa-tRNA when in the form of a ternary complex was notably different in orientation to that of the final accommodated state (see Sect. 12.3.1.3 for more details), this explains why the initial binding step of the ternary complex was not inhibited by Tet. Therefore, it is conceivable that upon release of the aa-tRNA from the ternary complex (concomitant with GTP hydrolysis and EF-Tu dissociation from the ribosome) the aa-tRNA falls off the ribosome, since it is only bound weakly to the ribosome through codon–anticodon interactions. The position of Tet in the primary binding site sterically prohibits the accommodation step causing the aa-tRNA to fall-off of the ribosome. This also explains why non-enzymatic binding is inhibited directly.

Moreover, the primary binding site best fulfils the plethora of biochemical data supporting it as the biological inhibitory one:

1. In the primary Tet binding site some interactions between Tet and the 16S rRNA are mediated through a bound magnesium ion (Fig. 12.6B) which may explain the magnesium dependence that has been observed with Tet inhibition [48].
2. Tet interacts with the rRNA using the charged face of the molecule whereas the more hydrophobic face protrudes into the intersubunit space (Fig. 12.6B). Consistently, substitutions generated along the hydrophobic face are generally tolerated (with the exception of the 4-dimethylamino group), while substitutions along the charged face result in loss of antibacterial activity [49].
3. While Tet derivatives that bind the ribosome and inhibit protein synthesis always enhance the DMS reactivity of bases associated with the primary site (for example, C1054 and U1052 in the 16S rRNA), only a subset protects bases associated with the secondary sites (A892) [50].
4. Single-site mutations that confer resistance to Tet are in close proximity to the primary binding site, such as the 16S rRNA mutations discovered in *Helicobacter pylori* and *Propionibacterium acnes* [51-53].
5. The ribosome protection protein Tet(O) binds to the ribosome in the vicinity of the primary binding site and confers resistance to Tet by chasing Tet specifically from this site [54] (this is discussed in more detail in Sect. 12.5.6).

It should be noted that the position of the secondary binding site is often regarded as the other most probable candidate for the inhibitory action of Tet because of its position in the so-called switch region of the ribosome (discussed in more detail in Sect. 12.3.1.2). This region was initially thought to be important for translational fidelity [55], but similar experiments with yeast ribosomes could not confirm this conclusion [56]. In any case, for the reasons mentioned above, its contribution to antimicrobial activity is likely to be less significant. In addition, it is noteworthy that the binding positions of the secondary sites, although being similar between both studies, are upon closer inspection notably different (Fig. 12.6C).

With regard to the broad spectrum of action of Tet, i.e., the observed inhibition of *in vitro* translation across all kingdoms, there is agreement with the interaction of Tet on the 30S, at least for the primary binding site, since the interactions of the Tet molecule are almost entirely composed of interactions with the sugar-phosphate backbone of the rRNA. The single interaction with the base of a nucleotide is via stacking interactions, and therefore does not discriminate between the type of base (Fig. 12.6B). As a corollary, the base mutations that give rise to Tet resistance are usually not associated with the strand containing the stacked bases, but the opposite strand and therefore resistance clearly results from conformational changes that disturb the Tet binding pocket. Other mechanisms that give rise to Tet resistance are reviewed in Ref. [57] and discussed in more detail in Sect. 12.5.6. In spite of its universal inhibition of ribosomes from all organisms, its wide use in medicine is due to the fact that the drug penetrates bacterial cell walls easily in contrast with those of eukaryotic cells.

The increasing incidence of bacterial resistance to the Tet group of antibiotics has led to a decline in their medical usage, which, in turn, has led to a new drive to find novel Tet derivatives. Although attempts were well underway since the discovery of chlortetracycline (reviewed in Ref. [49]), only the recently developed third-generation glycylcyclines look to fulfill their promise (reviewed in Ref. [58]). The glycylcyclines are derivatives of minocycline (7-demethylamino-6-demethyl-6-deoxy-tetracycline) and contain various substitutions at position 9 of the molecule (see Appendix B1 for structure comparisons). Of particular interest are DMG-MINO (9-(*N,N*-dimethylglycylamido)-minocycline), DMG-doxycycline, DMG-DMDOT (9-(*N,N*-dimethylglycylamido)-6-demethyl-6-deoxytetracycline), all of which retained antimicrobial activity and were even effective against some bacterial strains bearing *tet* resistance genes. Reports showed that these derivatives bound more effectively to the primary binding site than Tet, resulting in a 10-fold increase in potency for inhibition [59, 60], perhaps explaining why the ribosomal protection proteins are ineffective against these drugs. Recent interest has focused on one particular derivative, tigecycline or GAR-936 (also called TBG-MINO or 9-(*t*-butylglycylamido-minocycline)), which is currently undergoing Phase II clinical trials. In a recent study, strains resistant to both DMG-MINO and DMD-DOT were discovered (all with mutations in genes encoding Tet efflux proteins). These strains were susceptible to tigecycline; however, screening for resistant mutants uncovered strains with mutant efflux proteins conferring up to 4-fold higher resistance to tigecycline [61].

12.3.1.2 **Antibiotics Affecting the Fidelity of Translation**

The ribosome misreads 1 in ~3000 codons [62]. This effectively means that for every 3000 correct amino acids introduced into nascent polypeptides only one single erroneous amino acid is incorporated. This intrinsically low rate of misincorporation ensures that almost every protein produced by the ribosome is functionally active. The mechanism by which the ribosome successfully accomplishes this feat of translational fidelity has been solved at atomic resolution (see Sect. 8.2). In short, the ribosome monitors the stereochemical interactions between the A-site codon of the mRNA and the anticodon of the tRNA, to distinguish correct (cognate) from incorrect (near- or non-cognate) codon–anticodon interactions.

The aminoglycosides, i.e., streptomycin and the gentamycin, kanamycin, and neomycin families, interfere primarily with A-site occupation of the e-type, but not with that of the i-type [47]. These antibiotics commonly stimulate misreading, resulting in the incorporation of an incorrect amino acid. The increased misincorporation (=1:100) is not responsible for the bactericidal effect *per se*. (Note that a drug is bactericidal if it kills the cells rather than the usual blocking of growth, which is termed a bacteriostatic effect.) This is indicated by the fact that some mutants of the ribosomal protein S4 impair the accuracy of protein synthesis to a similar extent without affecting cell viability. Instead, the bactericidal effect is probably due to the blockage of A-site occupation of the e-type (both P and E-sites are occupied, see Fig. 12.1), i.e., the transition from the POST to the PRE state of the ribosome is blocked. Neomycin and hygromycin additionally impede the movement in the reverse direction

(from PRE to POST), viz. the translocation reaction. Similarly, blockage of the transition between the PRE and POST states in either direction has also been observed with the translocation inhibitors thiostrepton and viomycin [47].

Streptomycin

Streptomycin (Stp) is one of the most extensively investigated antibiotics known to directly interact with the ribosome. Stp is structurally related to the aminoglycoside family of antibiotics and exhibits the same classical hallmark, i.e., it induces translational misreading. Despite these common features, streptomycin binds to a distinct site on the ribosome and therefore mediates its inhibitory and misreading effects by an unrelated mechanism. For this reason streptomycin is treated separately from the other aminoglycosides. The amount of biochemical and structural data relating to this antibiotic that have accumulated over the past 50 years of study is immense, yet the mechanism of inhibition remains to be completely deciphered.

Stp has been co-crystallized in a complex with the 30S subunit of *T. thermophilus* [63]. In excellent agreement with much of the biochemical data (discussed in detail in Ref. [64]) Stp has a single binding site on the 30S subunit that connects helices from all four different domains of the 16S rRNA, namely h1 (nts 13), h18 (526), h27 (915) and h44 (1490) and makes interactions with ribosomal protein S12 (Fig. 12.7). Interestingly, streptomycin interacts only with the sugar-phosphate backbone of the 16S rRNA, i.e., there are no base-specific interactions. A number of the mutations in the ribosomal protein S12 that confer resistance to, and in some cases even dependence on, Stp map within the loop of S12 that directly contacts the molecule (see Table 12.6, with the exception of K53 which contacts h44). Of these mutations, only position K42 directly interacts with Stp, forming a hydrogen-bond with ring I, explaining why mutation of this residue to Arg (R) or Gln (Q) confers resistance. Mutations in other ribosomal proteins, mainly S4 and S5, were found to reverse the Stp-dependent phenotype of the S12 mutations [65]. Ribosomes containing one of the S12 mutations are hyperaccurate in tRNA selection (with the exception of the K42R resistance mutant, which does not alter translation accuracy), i.e., they restrict errors. In contrast, the S4 or S5 mutants are characterized as error-prone or ribosomal ambiguity mutants (*ram*). The same interplay in controlling accuracy is also evident when the yeast *Saccharomyces cerevisiae* harbors the analogous mutations in the equivalent proteins. It follows that the accuracy balance exerted by S12 versus S4 and S5 has been conserved during 2 billion years of evolution underlining its importance for all ribosomes.

The first real insights into the action of, and resistance to, Stp were concurrent with a better understanding of the ribosomal changes associated with tRNA selection. Comparison of 30S subunit crystal structures bound with codon and anticodon, in one case cognate to the codon and in the other, near-cognate, led to the proposal that selection of the correct or cognate tRNA by the ribosome requires a transition from an open to a closed form [66] (discussed in more detail in Chap. 8.2). Stp binding stabilizes the closed form and, by doing so, explains the lower translational fidelity. Transition into the closed form involves (i) disruption of

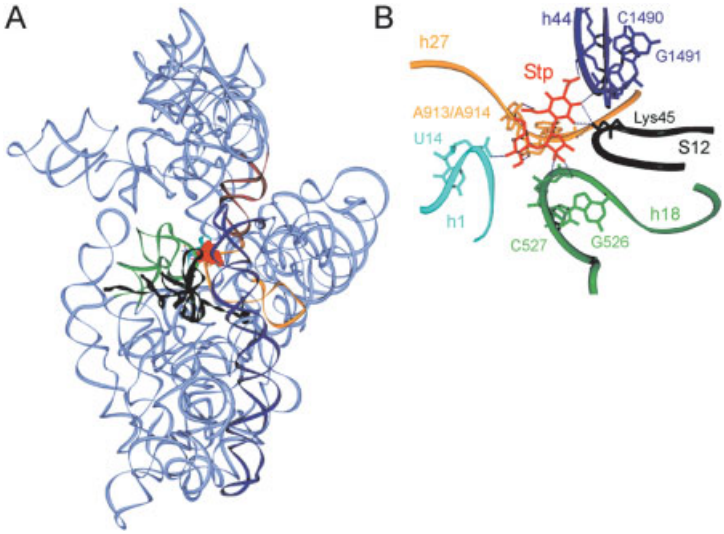


Figure 12.7 The streptomycin binding site on the 30S subunit. (A) Overview Streptomycin binding site (pdb 1FJG)[63] with streptomycin in red space-fill. The 16S rRNA is in ribbons with h1 (cyan), 530 loop or h18 (green), h27 (yellow), h28 (magenta) and h44 (dark blue) and ribosomal protein S12 (dark green) illustrated. (B) Detailed view of the streptomycin binding site. Streptomycin (red) interacts exclusively with sugar-phosphate backbone of the 16S rRNA and in doing so locks together all four of the 16S rRNA domains, namely, the 5′domain (U14 in h1), the central domain (G526 and C527 in the 530 loop), 3′ major domain (A913 and A914 in h27/h28) and 3′ minor domain (C1490 and G1491 of h44). Lysine 45 of S12 interacts with ring I of streptomycin and also the phosphate oxygen of A913. All colours are as in (A).

multiple interactions at the interface between S4 and S5 and (ii) establishment of salt-bridge interactions between S12 and either h44 or h27 of the rRNA. Consistently, (i) mutations in S4 and S5 that promote formation of the *ram* state would also lead to disruptions at this interface, suggesting that the observed error-prone phenotype stems from partially inducing the closed form and (ii) mutations in S12 that block salt-bridge formation may destabilize the closed form and thus confer resistance (or in some cases even dependence on the drug). The antagonistic effects of the S4/S5 *ram* mutants to destabilize the closed form, and the S12 streptomycin-resistant mutants to stabilize it, rationalizes the compensatory effects observed on translational fidelity.

Distinct from the open and closed forms, the ribosome has also been proposed to exist in two distinct states, the *ram* state, which as mentioned above is stabilized by Stp or established by mutations in S4 or S5, and the restrictive state as seen in S12 mutants. The *ram* state is characterized functionally by a large decoding error caused by a high affinity for aa-tRNAs, and was thought also to be achieved by base-pairing of the bases 885–887 with 912–910 in h27 [55]. If the base-pairing in the "switch region" is shifted by three nucleotides so that 912–910 now base-pairs to

888–890, the ribosome was believed to reside in a restrictive state [55]. However, this latter state has not been observed in any of the 30S crystal structures determined to date; in particular, it was not observed in either the closed or open forms, both of which had the 885–887/912–910 base-pairing combination. Furthermore, the importance of this proposed switching mechanism involving universally conserved bases is questionable, since it was shown not to play a role in yeast [56].

Aminoglycosides (2-deoxystreptamines)

Aminoglycosides can be divided into two categories based simply on whether they contain a 2-deoxystreptamine (2-DOS) group or not. Aminoglycosides that target the ribosome predominantly fall into the former class and can be further sub-classified depending on the substitution pattern of the 2-DOS ring, namely, those having sugar rings at position 4 (as for apramycin), both 4 and 5 positions (such as the neomycins (neo), paramomycins (Par) and ribostamycins) or 4,6 di-substituted (for, e.g., gentamycin, kanamycin (Kan) and tobramycin) as seen in the Appendix C1. The amino groups, which are protonated at neutral pH, make them positively charged molecules with high affinity to RNA – a property that was at one time exploited to recover RNA molecules from solution by centrifugation.

Paromomycin (Par) has been solved in complex with the complete 30S subunit [66–68] and also at higher resolution using short rRNA fragments mimicking their binding site on the ribosome [69–72]. The latter technique has also been successfully used to solve the structures of the binding site with a number of 4,6-disubstituted aminoglycosides (reviewed in Ref. [73]), namely, gentamycin C1a [71, 74] and the closely related Geneticin (also called G418 or gentamycin G; [75]) as well as tobramycin [76].

These structures show that the aminoglycoside family of antibiotics target an internal loop in h44 within a region referred to as the decoding site (A-site on 30S, see Fig. 12.8). Binding of aminoglycosides induces conformational changes in the ribosome, first evident from the altered reactivity to modifying agents of nucleotides A1408 and G1494 within this region [23]. The aminoglycoside-bound conformation has a higher affinity for the codon–anticodon complex, which increases the selection of near-cognate tRNAs [77], thereby explaining the decrease in translational fidelity in the presence of this class of antibiotics, as first observed in the 1960s (reviewed in Ref. [28]).

In detail, binding of Par within the decoding center induces the universally conserved residues A1492 and A1493 (green) to flip out of h44 (Figs. 12.8A and B), in a fashion reminiscent to that observed during aa-tRNA binding to the A-site [68]. This conformational change results from the insertion of one (ring I) of the four rings of Par into h44. By doing so, ring I mimics a nucleotide base, stacking against G1491 and forming a hydrogen-bond from the 6′-OH of ring I with the N1 position of A1408. The stability of this conformation is strengthened further by hydrogen-bonding between ring I and the backbone of the flipped out A1493.

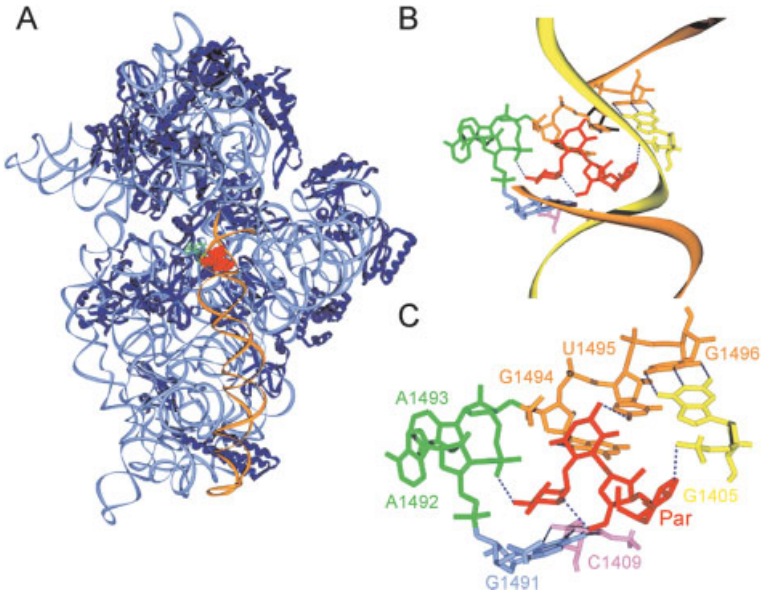


Figure 12.8 Overview of the aminoglycoside paromomycin binding site on the *T. thermophilus* 30S subunit (pdb 1ibk; [68]). (A) Ribbons representation of 16S rRNA (light blue) and ribosomal proteins (dark blue) with h44 high-lighted in yellow and the flipped out A1492 and A1493 in green. Paromomycin is shown in red spacefill representation bound at the top of helix 44. (B) Close-up view of paromomycin site binding site within h44. The different strands of h44 are coloured yellow (1400-1410) and gold (1490-1496) and paromomycin is red. Note the ribbon representation is broken on one strand between 1491-1494. (C) Close-up view of the bases associated with the paromomycin binding site. The flipped out bases of A1492 and A1493 (green), the G1491-C1409 base-pair that forms the shelf upon which ring I sits (pale blue and pink respectively) as well as U1405 (yellow) and G1494, U1495 and G1496 (orange). Hydrogen bond interactions are indicated with a dashed line, for example, in the Watson-Crick 1405-1496 base pair. Paromomycin is coloured red.

As discussed in Chap. 8.2, the formation of correct codon–anticodon interactions is monitored by the formation of A-minor interactions between A1492 and A1493 with the codon–anticodon helix. Presumably, the energy required to flip out A1492 and A1493 during decoding is compensated for by interactions established with the codon–anticodon helix, thus stabilizing this conformation [68]. In the presence of near-cognate tRNA, the prediction is that these compensatory interactions are insufficient to stabilize the flipping out of A1492 and A1493 and, thus, A-site accommodation does not occur. However, in the presence of Par, the uncompensated losses of energy are absorbed by Par that has already induced A1492 and A1493 to

flip out and has stabilized them in this open conformation. The outcome being that a near-cognate tRNA becomes fully accommodated into the A-site and thus results in mis-incorporation of an amino acid (reviewed in Ref. [78] and also discussed in Sect. 8.2.3).

Comparison of the structures for the aminoglycosides complexed with the 30S subunit and with the decoding site rRNA fragment reveal their striking similarity, especially with regard to the position in the A-site of the common neamine core (rings I and II) of the 4,5- and 4,6-disubstituted varieties. This latter point is consistent with the conclusion that the neamine core is sufficient for A-site binding [15, 79]. Furthermore, despite differences in the specific contacts between the aminoglycoside subclasses and the rRNA – resulting from both the different position of substituted ring III and also the different substitutions at each position within the rings themselves – the number of contacts remain equivalent, for example, direct hydrogen-bonds from ring III in one subclass are replaced by water bridges in the other [76]. Similarly, Par forms hydrogen-bonds with the 3′- and 4′-OH from ring I with G1491, whereas the gentamycin C class have methyl groups in this position; thus the hydrogen-bonds are replaced by hydrophobic interactions [74]. This explains the similar minimal inhibitory concentrations (MIC) between the different subclasses, for example, 2.5, 5, 10 and 40 μg ml^{-1} for tobramycin, geneticin, Par/ribostamycin and neamine, respectively. However, the affinity of aminoglycosides for the A site decreases significantly with the absence of rings III and IV, thus explaining the higher MIC of neamine with respect to the other aminoglycosides. As well as providing additional contacts with the rRNA, the presence of the additional rings III and IV contributes to a stabilization in the positioning of rings I and II.

All aminoglycosides that bind to the decoding center have a hydrogen-bond donor at the 6′ position on ring I. The 1408 position is an adenosine in all bacterial sequences whereas it is usually guanosine in eukaryotic sequences. This has implications for the specificity of aminoglycosides since mutations of A1408 to G confer high-level resistance to all of the aminoglycosides except those with a 6′ OH. Structural studies suggest that a G1408–A1493 base-pair would disrupt the binding site for these aminoglycosides and the 6′ amino group would be unable to hydrogen-bond the phosphate backbone at A1493 [80, 81]. Organisms that produce aminoglycosides protect themselves by having methylases that specifically modify either the drug (see Sect. 12.5.5) or the 16S rRNA at nucleotides G1405(N7) (Kan and gentamycin-resistant but not Par or neo) or A1408(N1) (Kan, apramycin and istamycin-resistant). Methylation of G1405 would be predicted to prevent hydrogen-bonding with ring III of Kan or gentamycin (whereas no contacts between Par and the base of G1405 are observed), whereas methylation of A1408 prevents formation of the A1408–A1493 base-pair that is essential for aminoglycoside binding. Similarly, the base-pair between C1409 and G1491 is also important for aminoglycoside binding: This base-pair forms a “shelf” in the aminoglycoside-binding pocket, upon which ring I makes stacking interactions with the base moiety of G1491. Disruption of this base-pair, for example by the mutation G1491C (or G1491U), leads to neomycin resistance in *E. coli*. Higher eukaryotes have a mispair at 1409–1491 positions,

which together with G1408, accounts for the 10-fold lower affinity of aminoglycosides for eukaryotic ribosomes. However, it should be noted that breaking the C1409-G1491 base-pair is not always sufficient in bacteria to attain a significant level of resistance to Par, nor to other members of the neomycin family. Furthermore, the G1491U mutation (in contrast with the G1491C mutation) confers resistance only in strains that contain additionally the streptomycin-resistant S12 allele. On the other hand, the 16S rRNA mutations G1491U or G1491C in strains carrying a mutant S12, influence the interaction of the ribosomes with streptomycin in a more complex way. The combination of the 16S rRNA mutations with highly restrictive S12 mutations produces a streptomycin-dependent phenotype (the strain can grow only in the presence of the drug), whereas, when combined with weak or non-restrictive streptomycin-resistant S12 mutations, the streptomycin sensitivity is partially restored. The interplay of mutations in both S12 and helix h44 can be easily envisioned if one considers the binding site of streptomycin.

One interesting observation when comparing the 4,5- and 4,6-disubstituted aminoglycosides is that there are no members of the latter family that contain a four-ring system. In light of structures for representatives of each family bound to the decoding center, the reason for this is that the 4,6-disubstitution generates a rather linear molecule which when extended by an additional ring would be difficult to accommodate in the binding site due to space restrictions [74]. Despite this restriction, the 4,6-disubstituted family of aminoglycosides are clinically preferred, i.e., the predominantly used aminoglycosides are gentamycin (Garamycin® introduced in the mid-1960 s by Schering-Plough) and two kanamycin derivatives, tobramycin (Nebcin® marketed Eli Lilly and Company) and amikacin (marketed under the name Amikin® by Bristol-Myers Squibb). This may in part relate to the fact that ring III of this family makes additional base-specific contacts with the A-site which are not present in the neomycin family [74]. Furthermore, position 1 of ring II of amikacin bears a bulky chain that blocks resistance enzymes from modifying this ring, while at the same time still allowing hydrogen-bonding with O4 of U1495. Similarly, the absence of 3′- and 4′-OH groups on tobramycin and certain gentamycins also protect the drug from inactivation since these groups are also targets for resistance enzymes.

The binding site for the aminoglycoside Hygromycin B has been determined on the 30S subunit [29]. Although Hygromycin B also binds within the decoding center, making exclusively contacts with h44, the location is slightly displaced towards the top of h44 when compared with the position of, for example, paromomycin. The binding of Hygromycin B is very sequence-specific since all interactions are with the bases of the rRNA, rather than the backbone, however, no conformational changes were observed within the rRNA upon binding of the drug [29]. Hygromycin B has been demonstrated to be a powerful translocation inhibitor [47]. Since the top of h44 is thought to move with the A- and P-site tRNAs during the translocation reaction (see Chap. 8), the inhibition of translocation probably results from the reduced flexibility arising from drug binding within this region [29].

12.3.1.3 Inhibitors of EF-Tu-mediated Reactions

There are three basic mechanisms by which antibiotics can disrupt EF-Tu function:

1. Preventing the release or dissociation of EF-Tu from the ribosome without necessarily affecting the GTPase activity, as exemplified by kirromycin and enacyloxin IIa.
2. Blocking EF-Tu from forming the ternary complex, as evident for GE2270A and pulvomycin.
3. Disrupting the capability of ribosomes to bind the aa-tRNA•EF-Tu•GTP ternary complex, as exemplified by the ribotoxin family of proteins, such as the well-known α-sarcin and ricin.

Point 1 is well reviewed in Refs. [82, 83], whereas points 2 and 3 are discussed thoroughly in a recent review by Hilgenfeld and co-workers [84].

The Kirromycins Trap EF-Tu on the Ribosome

Initial binding of the ternary complex aa-tRNA•EF-Tu•GTP to the A-site results in EF-Tu-dependent GTP hydrolysis. Cleavage of GTP causes EF-Tu to adopt the low-affinity GDP conformation that dissociates from the ribosome – analogous with the action of typical G-proteins. Kirromycin and aurodox (*N*-methyl kirromycin) are members of a large family of agents produced by the actinomycetes (see Appendix D1 for structures). The kirromycins stall the ternary complex on the ribosome, not by preventing hydrolysis of GTP to GDP by EF-Tu, but by preventing the conformational changes in EF-Tu that are associated with GTP hydrolysis. In effect, kirromycins lock EF-Tu in a high-affinity state, which prevents both release of the aa-tRNA from EF-Tu and dissociation of EF-Tu from the ribosome. This made the kirromycin-stalled ternary complex-bound ribosomes ideal for analysis using cryo-electron microscopy, which provided the first direct visualization of this state [85]. The recent more refined reconstructions reveal that kirromycin captures the ternary complex during the initial stages of A-site binding; in this situation the anticodon of the tRNA is "kinked" to enable it to undergo codon–anticodon interactions and simultaneously maintain interaction between the acceptor stem and EF-Tu [86]. Correct codon–anticodon interaction is proposed to trigger hydrolysis of GTP by EF-Tu leading to dissociation of EF-Tu from the ribosome, which in turn enables the aa-tRNA to move into the A-site on the 50S subunit, become fully "accommodated" and to partake in peptide-bond formation. The corollary being that in the kirromycin-stalled complex, the acceptor stem and aminoacyl moiety of the aa-tRNA, by remaining bound to EF-Tu, are unable to move into the PTF center and thus progression of the elongation cycle is prevented since no subsequent peptide-bond formation is possible.

Since the structures for EF-Tu in both the GTP (actually with the non-hydrolyzable GTP analog, GDPNP) and GDP form have been solved to high resolution, the conformational rearrangements that are undertaken as a results of GTP hydrolysis are well understood (reviewed in Refs. [87, 88]). Briefly, following GTP hydrolysis,

domain I (the GTP binding domain) of EF-Tu undergoes a conformational change and is rotated relative to domains II and III by up to 40 Å (cf. Figs. 12.9A and B). Comparison of these structures with the recent structure of *T. thermophilus* EF-Tu•GDP•aurodox [89] confirms the resemblance of this latter structure with the GTP conformation (cf. Figs. 12.9A and C). In this structure, aurodox is wedged between domains I and III making almost exclusively hydrophobic interactions. By binding within this region, kirromycin may lock domains I and III together thereby preventing allosteric switching.

This location is in agreement with the numerous kirromycin-resistant mutations that map to amino acids clustering at the interface between domains I and III in the

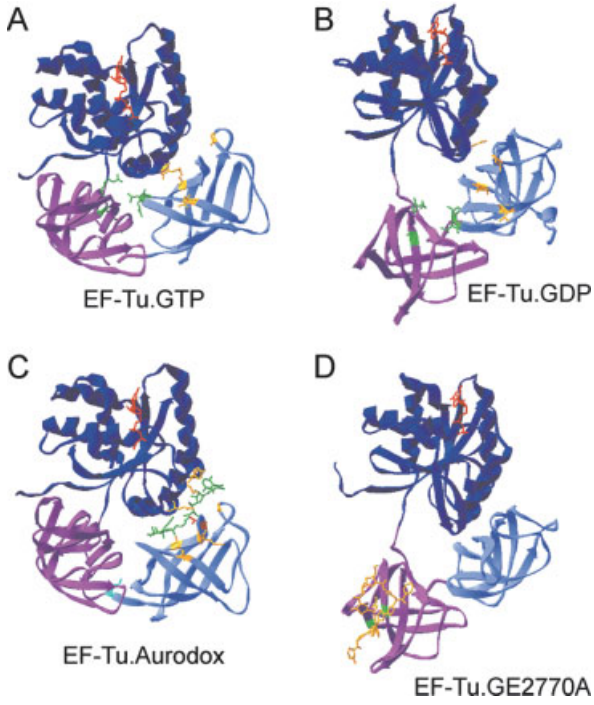


Figure 12.9 The structures of aurodox and GE2770A bound to EF-Tu. (A) EF-Tu•GDPNP (PDB1EXM; [88]), (B) EF-Tu•GDP (PDB1D2E; [87]), (C) EF-Tu•GDP•Aurodox (PDB1HA3; [89]) (D) EF-Tu•GDP•GE2770A (PDB1D8T; [94]). In A and B, the residues conferring resistance to Enacyloxin IIa and pulvomycin are coloured yellow and green, respectively. In C and D, the antibiotics aurodox and GE2770A are coloured green and yellow, respectively. In C, substitutions at residues giving rise to kirromycin resistance are coloured yellow. Position I298 and T382, which when deleted or phosphorylated (respectively) confer resistance to kirromycin are coloured cyan and red, respectively. In Figure D, sites of mutations giving rise to GE2770A are indicated in green. In Figures A-D, domains I, II and III are coloured dark blue, magenta and light blue, respectively, and the guanine nucleotide is always red.

GTP form (but not the GDP form) (see Table 12.4 and as indicated in Fig. 12.9C). Resistance is conferred through two distinct mechanisms:

1. Obstruction of the binding of kirromycin to EF-Tu•GTP by disrupting the kirromycin-binding site.
2. Promoting dissociation of kirromycin from the ternary complex following GTP hydrolysis since kirromycin has a lower affinity for EF-Tu•GDP. This protection mechanism implies that some mutants permit sufficient conformational change within EF-Tu concomitant to GTP hydrolysis despite the presence of kirromycin.

Kirromycin resistance is a recessive phenotype and therefore sensitivity to kirromycin is dominant. Moreover, many bacteria, including *E. coli*, contain two genes encoding EF-Tu (*tufA* and *tufB*). In these particular cases, resistance is dependent on an alteration in both genes. Alteration of only one of the genes will not confer resistance to kirromycin [90], since a ribosome blocked by a sensitive EF-Tu prevents translation for the subsequent ribosomes of the polysome.

Enacyloxin IIa is a linear polyenic acid similar to kirromycin (Appendix D1) isolated from *Fratueria* W-315 (i.e., not from the actinomycetes as all other EF-Tu-binding inhibitory antibiotics) and is active against Gram-positive and Gram-negative bacteria [91, 92]. It is easy to envisage enacyloxin IIa operating through a similar

Table 12.4 Mutations in EF-Tu giving rise to antibiotic resistance

Antibiotic	Mutation position[a]	EF-Tu domain	Reference
Kirromycin	L120Q,	I	240–242 and references therein.
	Q124R[b], E, K, N	I	
	Y160N, D, C	I	
	Δ298	II–III linker	
	G316D[b]	III	
	Q329H,	III	
	A375T, N, V, S	III	
	E378K	III	
Kirromycin	T382- phosphorylation	III	243
Enacyloxin IIa	Q124K,	I	93
	G316, Q329, A375	III	
GE2270A	G257A, G275A, S	II	244
Pulvomycin	R230C, R233S,	II	245, 96
	R333C, T334A	III	
	R230V-R233F	II; III	

a *E. coli* numbering is given although resistance mutants were not necessarily identified in *E. coli*.

b Kirromycin-resistant eukaryotic and archaebacterial EF-1 (the EF-Tu homolog) have substitution and deletion of the positions equivalent to Q124T and ΔG316, respectively [246].

mechanism to kirromycin, since enacyloxin is structurally similar and the mutations that give rise to enacyloxin resistance were originally identified as kirromycin-resistant mutations (indicated in Figs. 12.9A and B). This suggests that they have, at least to some extent, overlapping binding sites on EF-Tu. However, there are some differences with the kirromycins, for example, the enacyloxin IIa-bound EF-Tu•GDP complex has an even higher affinity for aa-tRNA than kirromycin and enacyloxin IIa does not enhance the GTPase activity of EF-Tu in the way kirromycin does. Furthermore, subtle substitution changes at sites in EF-Tu can affect resistance to kirromycin and enacyloxin differently, for example, A375V confers resistance to kirromycin but not enacyloxin whereas A375T confers resistance to both [93]. This raises doubt as to whether Enacyloxin IIa should simply be classified along with kirromycin or whether it should be differentiated. Structural studies with the enacyloxin-bound EF-Tu complexes should resolve this issue.

Inhibition of Ternary Complex Formation: Pulvomycin and GE2270A

Both pulvomycin and GE2270A (also referred to as MDL 62, 879) prevent the binding of the aa-tRNA to EF-Tu, i.e., they prevent ternary complex formation; however it appears likely that they operate through dissimilar mechanisms, although this remains to be verified. First, pulvomycin and GE2270A are structurally unrelated. While pulvomycin bears some resemblance to the kirromycins, GE2770A is a member of the cyclic thiazolyl peptide family (Appendix D1). In fact, GE2770A is more closely related in structural terms to the thiostreptons and micrococcins, which are also inhibitors of protein synthesis, but act by binding directly to the ribosome (see Sect. 12.3.3.1). Recently, the crystal structure of EF-Tu•GDP•GE2770A was solved to 2.35 Å resolution [94]. GE2770A was located in a cleft of domain II (Fig. 12.9D), where ionic interactions with R223 and E259 account for the strong affinity of this antibiotic. In agreement, mutations at these two residues are associated with resistance to this antibiotic (Table 12.4 and as colored green in Fig. 12.9D) – however, it should be noted that these residues are not invariant throughout the prokaryotes suggesting that some organisms may be naturally resistant (a good example being the producer of GE2770A, *Planobispora rosea*). A comparison of the aforementioned structure with that for binary complex EF-Tu•GTP and the ternary complex EF-Tu•GTP•tRNA suggested that the binding position of GE2770A would sterically clash with that of the amino-acyl moiety of the aa-tRNA. This in itself explains the inhibitory action of GE2770A to prevent ternary complex formation, but in addition, the binding position of GE2770A at the interface of domain II would prevent tight association with domain I, an interaction necessary to adopt fully the GTP conformation. Therefore, GE2770A can be supposed as having a dual action in preventing ternary complex formation.

In contrast with GE2770A, the action of pulvomycin (Appendix D1) is not so well understood. Although pulvomycin, like GE2770A, inhibits ternary complex formation, the locations of the resistance mutations are distinct from those of GE2770A (Table 12.4), suggesting that the inhibitory mechanism also differs. The location of the pulvomycin-resistant mutations is at the junction between domains II and III

(as indicated in Figs. 12.9A and B). Since pulvomycin (as GE2770A) does not compete with kirromycin for binding to EF-Tu [95], this suggests that if pulvomycin does interact with domain II, it is probably not at the I–II interface. Domain III of EF-Tu has been shown to be necessary for pulvomycin binding, leading to the speculation that the binding site of pulvomycin lies at the domain II–III interface [84]. A striking difference between pulvomycin and GE2770A relates to their resistance phenotype; pulvomycin sensitivity was found to be dominant to resistance [96], whereas in contrast GE2770A resistance was dominant over sensitivity [97, 98]. This suggests that in the case of pulvomycin an additional mechanism must be operating other than simply limiting the availability of active ternary complex.

The ribotoxins: α-sarcin and ricin A

Ribosome-inactivating proteins (RIPs) are ribotoxins produced by bacteria, fungi, and plants to damage the ribosomes of other organisms, either prokaryotic or eukaryotic. These ubiquitous proteins can be grouped, based on their method of inactivation, into either the α-sarcin-like fungal ribonucleases (RNases) or the bacterial and plant RIP family of glycosidases, for which ricin is perhaps the best known member. The action of α-sarcin or ricin A on *E. coli* ribosomes results in a direct loss in binding of both elongation factors. Although EF-Tu-dependent A-site occupation and EF-G-catalyzed translocation are blocked, all other functions of the ribosome including non-enzymatic A-site binding and spontaneous translocation remain unaffected [99]. The target of the RIPs is a 12-base loop (termed the sarcin-ricin loop (SRL)), a constituent of domain VI of large subunit ribosomal RNA, and contains the longest stretch of universally conserved nucleotides in the cell (Fig. 12.2B). The SRL is a continuous irregular helix with a bulged G2655, which distorts the backbone creating a characteristic S-shape. The helix is closed by a GAGA tetraloop, of which the two bases A2660 and G2661 are 'looped out' and accessible for possible interactions with EFs or RIPs.

The cytotoxic protein α-sarcin is produced by the *Aspergillus* species and inhibits protein biosynthesis by cleavage of the SRL. The high specificity and effectiveness of this single cut is illustrated by the observation that fragmentation of the rRNA by introducing up to 10 randomly distributed breaks does not affect protein synthesis *in vitro*, but the α-sarcin specific cleavage at the 3′ side of G2661 in 23S rRNA (or G4326 in the rat 28S rRNA) completely abolishes protein synthesis [99, 100]. The extreme conservation probably explains the observation that α-sarcin is active against ribosomes across all kingdoms. Although this is also true for glycosidase family of ribotoxins, for example, gypsophilin (from *Gypsophilia elegans*) acts on both prokaryotic and eukaryotic ribosomes [101], there are a number that effectively target only the SRL of eukaryotes, for example, pepocin is 10000-fold less effective against *E. coli* as against mammalian ribosomes [102]. RIPs from higher plants can be further divided into two distinct categories based on their structures; the type I RIPs composed of a single protein chain of ~30 kDa and the larger type II RIPs, which are composed of two unequal chains, an A chain homologous to the type I RIP and a B chain that binds to the A chain and facilitates its cellular uptake. The A

chain of the type II ribotoxin ricin depurinates (depurination involves hydrolysis of the N-glycosidic linkage between the ribose and sugar resulting in the removal of a purine base, which in this case is adenine) only eukaryotic ribosomes at A4324 of the 28S rRNA (which corresponds to *E. coli* A2660). Only the naked 23S rRNA of *E. coli,* but not the complete ribosome, is a substrate for ricin A, suggesting that the ricin-binding site on the mature *E. coli* ribosomes is not exposed in the same manner as that for eukaryotic ribosomes. Recently, this was elegantly confirmed using hybrid ribosomes constructed from *E. coli* ribosomes where the pentameric L10•(L7/L12)$_4$ complex and L11 were substituted for the rat counterparts (P0•(P1/P2) and eL12, respectively). Pepocin, which normally acts only on eukaryotic ribosomes, was demonstrated to act on the *E. coli* hybrid ribosomes but only in the presence of both P0•(P1/P2)$_2$ and eL12 [103]. Similarly, the binding of L10•(L7/L12)$_4$ complex and L11 to *E. coli* ribosomes was necessary for susceptibility to gypsophilin. This suggests that the binding of these specific ribosomal proteins dictates the conformational state of the SRL and its accessibility to the RIPs.

12.3.2 Inhibitors of Peptide-bond Formation and Nascent Chain Progression

The central enzymatic function of the ribosome is peptidyl transferase (PTF), which is the domain of the large ribosomal subunit. The recent structures of the 50S subunit, alone and in complex with various ligands and antibiotics, has led to rapid improvement in our understanding of the PTF reaction itself (see Chap. 8.3 for more details), but also the mechanism of action of the multitude of antibiotics that target this region. The PTF center of the ribosome is composed entirely of rRNA, made up mostly of residues from the PTF ring of domain V of the 23S rRNA (Fig. 12.2B). This active center and the associated region of tunnel extending from it comprise the target for the majority of large subunit binding antibiotics that have been structurally characterized to date. For this reason, the predominant interactions are made with rRNA, relegating ribosomal proteins almost exclusively to indirect role in antibiotic interactions. Despite this, numerous ribosomal proteins have been associated with antibiotic resistance (summarized in Table 12.6). A special feature of about two-thirds of the ribosomal proteins is the presence of a globular domain, usually located at the solvent surface of the ribosome, and long finger-like extensions that weave their way through the rRNA into the core of the ribosome. These protein extensions are thought to act like "glue" and provide a scaffold for the rRNA. For this reason, mutations in ribosomal proteins that confer resistance to particular antibiotics, despite making no direct contact, probably do so by altering the architecture of the rRNA and therefore the antibiotic binding site.

12.3.2.1 Puromycin and Blasticidin S mimic the CCA end of tRNAs

Puromycin (Puro) is a structural analog of the 3′-end of aminoacyl-tRNA, except that the aminoacyl residue is linked to the ribose via an amide bridge rather than an

ester bond (Fig. 12.10A). Puro binds to the A-site region of the PTF center. Following A-site binding, peptidyl transfer links the peptidyl residue covalently to the drug. The peptidyl-Puro then dissociates from the ribosome since it has a low affinity being bound at the A-site only via the 3′-terminal adenine. Furthermore, should the peptidyl-Puro arrive at the P-site, by rebinding, no further peptidyl transfer could ensue as the amide bridge cannot be cleaved by the ribosome. Thus, Puro is effectively terminating peptide chain elongation by exploiting the ribosomal PTF activity. Peptide-bond formation of Puro with peptidyl-tRNA, peptidyl-tRNA analogs or 3′-terminal fragments of these tRNAs, are important tools for studying the PTF reaction. In fact, the classical definitions of A- and P-sites are based on the inability or ability, respectively, of peptidyl-tRNA or its analog *N*-acetyl-Phe-tRNA (AcPhe-tRNA) to react with Puro.

The binding site of Puro was first visualized at high resolution in the 3.3 Å structure of the *H. marismortui* 50S subunit bound with the Yarus inhibitor [104], an analog of the peptide-bond intermediate formed by linkage of CCdA to Puro via a phosphoramide group [105](see Chap. 8.3). In this structure, the Puro component of the Yarus inhibitor occupies the A-site position: a type-I A-minor interaction between the adenosine base (termed A76 since its position mimics the terminal adenosine (A76) of a tRNA) and the G2583–U2506 base-pair as well as an additional H-bond interaction between the 2′-OH of A76 with U2585 and stacking of A76 with U2558 (Fig. 12.10B). Subsequently, similar positions for puromycin were observed as the substrate in the A- and P-sites namely CC-pcb (CC-Puro-caprioic acid-biotin) [106] and as the post-peptide-bond formation product at the P-site of the *H. marismortui* 50S (Fig. 12.10C) [107], as well as, ACC-Puro with the *D. radiodurans* 50S subunit [108]. Interestingly, CCA-Pcb bound to both A- and P-sites suggesting that it had equal affinity for both sites, but in the presence of sparsomycin, was only present in the P-site [106]. In contrast, CACCA-Leu and CACCA-LeuAc (α-amino group is blocked by an acetyl residue) show a perfect specificity for A and P-site regions of the PTF center, respectively [109].

Unlike puromycin, which mimics an aminoacylated-terminal adenosine (A76) of a tRNA, blasticidin S has a structure resembling either of the preceding cytosines (C74 or C75) of the tRNA. Specifically, blasticidin S is composed of a cytosine base and a pyranose sugar with an N-methylguanidine tail (Appendix E1). Although few studies have addressed the mechanism of blasticidin S, it has been reported to inhibit the PTF reaction of both bacterial 70S and eukaryotic 80S ribosomes [110, 111].

The structure of blasticidin S soaked *H. marismortui* 50S subunits was reported at 3 Å resolution [112]. Two molecules of blasticidin S were identified, both bound at the PTF center, where they are positioned so as to mimic C74 and C75 of a tRNA at the P-site by making interactions with P loop residues. In the higher occupancy site, the cytosine base of blasticidin S makes Watson–Crick (WC) base-pairs with G2251, whereas in the lower affinity site the cytosine of the second molecule of blasticidin S forms WC interactions with G2252. Binding to the higher affinity site is further stabilized by stacking of the *N*-methylguanidine tail onto the base of A2439 as well

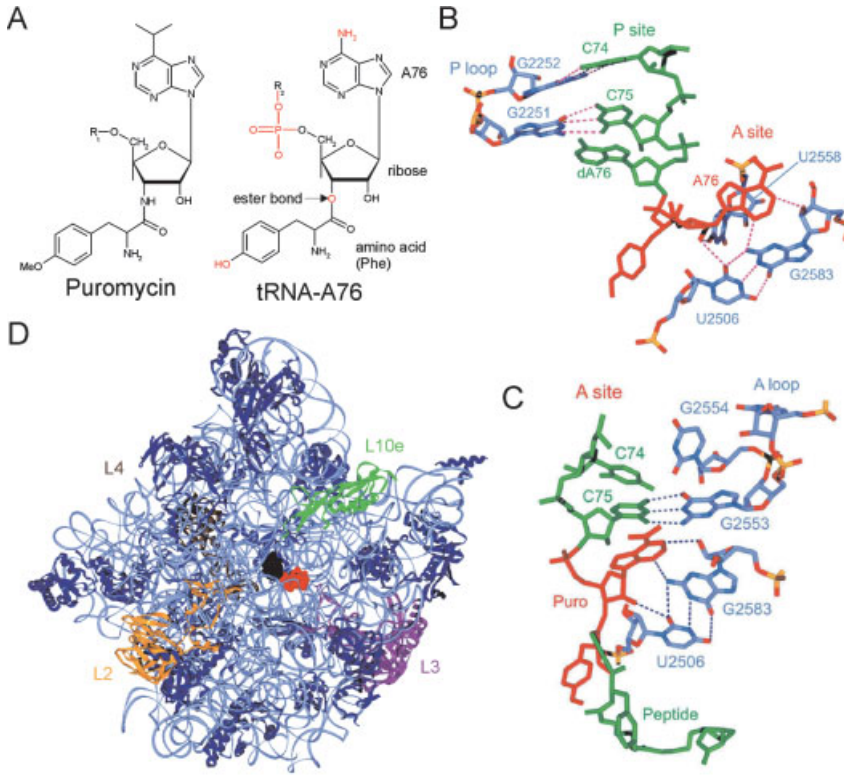


Figure 12.10 Puromycin binds at the peptidyl-transferase center of the 50S subunit. (A) Comparison of structures of puromycin with the terminal adenine (A76) aminoacylated with phenylalanine. Differences between puromycin and the physiological tRNA substrate are indicated in red on the tRNA. Puromycin bound to the *H. marismortui* 50S ribosomal subunit in the form of (B) the Yarus inhibitor (pdb 1FFZ)[104] and (C) the products following peptide bond formation (pdb 1KQS)[107]. The puromycin part of each of the respective compounds is coloured red. Selected rRNA residues of domain V of the 23S rRNA are colored light blue, including the A- and P-loop bases that participate in A and P site CCA end fixation (*E. coli* numbering). In (B) the A site C74 and C75 mimics have been omitted for clarity, likewise in (C) for the P site product. Dashes indicate hydrogen bonding and rRNA nucleotides use the following color scheme: Oxygen, red; phosphorus, yellow; nitrogen, blue; carbon, dark blue. (D) Overview of where the PTC is on the ribosome using the A site and P site products shown in red and green respectively. Ribosomal Protiens L4 (purple) L10e (green) L16 in bacteria L2 (yellow), and L3 (crimson) have long extenstions that reach towards the PTC.

as hydrogen-bonding interactions with the phosphates of both A2439 and A2600. The removal of or alterations in the tail markedly decreases the effectiveness of blasticidin S on translation [113], suggesting these interactions are important for drug binding. Residue 2439 has also been identified as being protected by blasticidin S from chemical probing [114] and mutation of the neighboring base U2438 to C confers resistance to blasticidin S in the archaeabacterium *H. halobium* [115].

12.3.2.2 **Sparsomycin Prevents A-site Binding and Stimulates P-site Binding**

Sparsomycin (Spm), a modified uracil antibiotic produced by *Streptomyces sparsogenes*, has long been known as a potent inhibitor of PTF activity in all organisms studied (reviewed in [28]). Specifically, Spm has been shown to interfere with binding of tRNA (and CCA-end fragments) to the A-site, while enhancing the affinity of peptidyl-tRNAs (especially N-acetylated aa-tRNAs and aa-ACC end fragments) for the P-site. Despite this early biochemical characterization, the binding site for Spm has remained elusive since Spm did not produce clear footprints on the rRNA [116, 117] nor could resistant mutants be isolated in *E. coli*. However, the subsequent isolation of Spm-resistant mutations was successfully achieved in several archaeal species indicating that the PTF center was the likely site of drug interaction: In *H. halobium,* mutation of C2518U and to a lesser extent C2471 and U2519 (C2499, C2452 and U2500 in *E. coli*, respectively) conferred resistance to Spm [117], as did loss in methylation (probably at the N3 position) of U2603 in *H. salinarium* (U2584 in *E. coli*) [118]. These residues are to be found in the PTF ring of domain V of the 23S rRNA (Fig. 12.2B). Consistently, Spm competes for binding with the PTF inhibitors chloramphenicol and lincomycin [119] and has been crosslinked to residue A2602 [120].

The high-resolution structures of Spm bound to the 50S subunit of both *H. marismortui* (H50S) [112, 106] and *D. radiodurans* (D50S) [108] have revealed why the footprinting techniques had been so unsuccessful: in the H50S complexes, density for Spm was only observed when a P-site substrate was included in the co-crystallization experiment [112], whereas, in the D50S, although density for Spm was observed in the absence of a P-site ligand, the sole interaction with the ribosome was through a stacking interaction between the modified uracil ring of Spm and the base of A2602. This stacking interaction is also observed in both H50S and D50S complexes, where P-site ligands are included, however, despite this similarity, the orientation of A2602 and Spm itself differed significantly.

In the D50S-Spm structure with a tRNA acceptor-stem mimic (ASM), the presence of Spm has pushed the helical region of the ASM towards the P-site, but the CCA-end of the ASM still maintains interactions with the A loop, thus placing it in the A-site. In contrast, the H50S-Spm structure with the tRNA mimic, CCA-phe-cap-biotin (CCA-pcb), the mimic is clearly bound at the P-site. Furthermore, in this latter structure, the extensive interactions made with Spm explain how the presence of a P-site substrate significantly enhances Spm binding and *vice versa:* the uracil ring of

Spm is sandwiched between the P-site substrate and A2602, the N3 position of the uracil makes a hydrogen-bond with the phosphate group of C75, the exocyclic methyl group of the uracil forms van der Waals contacts with C75 and A76 and the C3-keto group of uracil is co-ordinated through a bound Mg ion with two phosphate-oxygen atoms of the P-site substrate (Fig. 12.11). In addition, the sulphurous tail of Spm contributes to the binding affinity by inserting into the A-site, where it would be predicted to inhibit binding of A-site ligands, thus explaining the observed competition with drugs such as chloramphenicol and puromycin. Therefore, the inhibitory action of Spm is likely to be due to blocking the binding of the A-site ligand in a fashion that is dependent on the presence of a P-site ligand necessary to stabilize Spm binding. Ribosomes that are in a PRE state, i.e., having both A- and P-sites occupied, are not protected from the action of Spm, since under these conditions Spm induces translocation of the A and P substrates [121], relegating the situation to a peptidyl-tRNA at the P-site and an A-site blocked by Spm. It is worth mentioning that the base in the PTF center that Spm stacks upon, A2602, is at the center of the rotational symmetry of the PTF center, where it has been proposed to play a role in guiding the CCA ends from the A- to P-site during translocation [122].

12.3.2.3 **Antibiotic Overlap in the PTF Center: chloramphenicol, Anisomycin and the Lincosamides**

Chloramphenicol (Cam; Appendix E1) inhibits several kinds of PTF assays. However, it does not interfere with, but rather stimulates, tRNA fragment binding to the

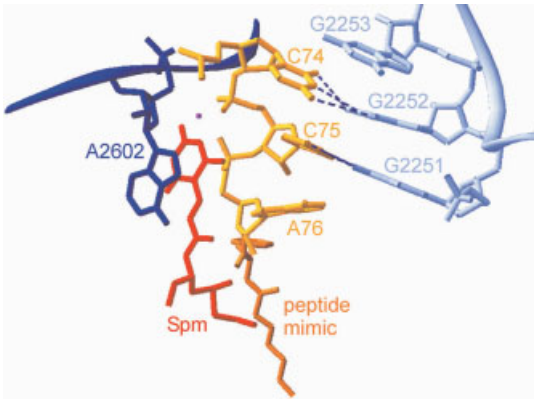


Figure 12.11 Sparsomycin binding is stabilized by interaction with the P site substrate. Spm (red) bound within the PTF center of the *H. marismortui* 50S subunit in the presence of P site substrate (CCA-pcb, yellow) [112, 106]. The uracil of Spm can be seen to make stacking interactions with the base of A2602 (blue) of the 23S rRNA. The P site substrate interacts with the residues G2251 and G2252 of the P loop (23S rRNA) by making hydrogen bond interactions (dashed blue lines) from the positions mimicking C75 and C74 of a bound tRNA.
Note additional interactions between Spm and P site substrate can be coordinated through the bound Mg ion (purple).

P-site of the PTF center [109], and competes with the binding of tRNA fragments to the A-site and with puromycin [123]. In intact bacteria, Cam "freezes" the polysome profile, suggesting that the drug inhibits the PTF reaction by disturbing the binding of the CCA 3′-end at the A-site within the catalytic center but without weakening tRNA binding *per se*. Curiously, the degree of inhibition depends on the character of the peptidyl residue or the A-site substrate, for example, AcPhe-puromycin formation is blocked, as well as Gly-Phe-puromycin formation, but the formation of AcPhe-Phe, AcPhe-Phe-puromycin or Leu-Phe-puromycin is not [124–126]. On this evidence it was argued that aromatic amino acids can displace Cam during peptide-bond formation by competing with the phenyl group of the drug.

The ribosome-binding site of Cam was determined to 3.5 Å by soaking 50S crystals of *D. radiodurans* in a solution containing 100 µM Cam and shown to involve interaction with seven nucleotides within the PTF center [127] (Fig. 12.12). Many of these interactions were indirect being mediated through two putative Mg^{2+} ions. Functionally important moieties for the antibiotic action of Cam constitute these interactions suggesting the presence of ions is of utmost importance for antibiotic binding and therefore PTF inhibition. The position of the dichloroacetamido tail of

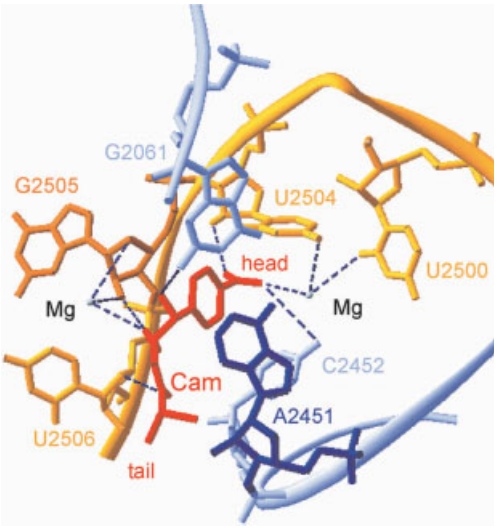


Figure 12.12 Chloramphenicol binding within the A site of the peptidyl-transferase center of the *D. radiodurans* 50S subunit (pdb 1k01) [127]. Chloramphenicol (Cam, red) interacts with 7 nucleotides within the PTF center, three through interactions with the head and four through the tail. Hydrogen bond interactions (blue dashed lines) between the p-NO_2 group of the head with U2500, U2504 and C2452 are evident as well as from the tail with G2061, G2505 and U2506. Some of these interactions are direct whereas others are stabilized through either of the two Mg ions present in the binding site.

Cam within the PTF center is located such that it extends towards, and may even displace, the CCA end of the A-site tRNA. This is in agreement with the observation that tRNAs can bind in the presence of Cam but they cannot undergo peptide-bond formation. Furthermore, the major overlap between Cam and an A-tRNA encompasses predominantly the amino acid moiety, lending credence to the idea that Cam operates predominantly by displacing the aminoacyl residue of the A-site tRNA and thus probably displaces the CCA end as an indirect result of this.

Cam protects bases A2059, *A2062*, A2070, A2451, *G2505*, *U2506* and enhances the reactivity of A2058 [128, 114]. Interactions with three of these residues (italicized) are made by Cam and A2451 is within 3.0 Å, whereas the other three (A2070, A2058 and A2059) are located over 10 Å away. Mutations that confer resistance to Cam map near to the sites of protection at positions 2057, 2451, *2452*, 2447, *2503* and *2504*, whereas the mutations *G2505C* and G2583U or C (but not G2583A) cause hypersensitivity towards Cam in an *in vitro* translation system [129]. The discrepancy between the Cam-binding site at the A-site and the protections (and some modest resistance mutations) located within the tunnel region, can be best explained by a second lower affinity binding site for Cam. Indeed, the structure of Cam bound to the *H. marismortui* 50S subunit identified a site distinct from that found in *D. radiodurans* located within the tunnel, overlapping the binding site of the macrolide class of antibiotics [112]. This is not totally unexpected since most archaea exhibit a natural resistance to Cam (probably due to sequence differences within the region of the D50S Cam site) and thus unusually high concentrations of Cam (20 mM) were necessary to bind Cam to this lower affinity site. This suggests that eubacterial ribosomes may also bind a second molecule of Cam at this lower affinity secondary site, which would correlate nicely with the previously contradictory biochemical data. It should be noted that the primary binding site is sufficient to account for the inhibitory action of Cam and also that in eukaryotes, Cam is not even taken up by the cells, although cycloheximide functions in a very similar fashion.

Cam has been crosslinked to ribosomal proteins L16 and L27 [130], extensions of both proteins approach the PTF center, although not close enough in the case of L16 for any direct contact with Cam in either the primary or secondary site. Therefore, the conclusion from reconstitution studies that L16 constitutes a component of the Cam-binding site [131], probably reflects instead alterations in the PTF center, due to the proteins absence, that perturb Cam binding. Unfortunately, the N-terminal of L27 in the D50S structure is relatively disordered and there is no counterpart in the H50S structure. The N-terminal residues of L27 have been crosslinked to the acceptor stem of a P-site-bound tRNA leading to the proposal that it may play a role in positioning the acceptor ends of the tRNA within the PTF center – this could place L27 in direct contact with the primary Cam-binding site.

In contrast with Cam, the anisomycin and lincosamide classes of antibiotics have been reported to interfere with binding of ribosomal ligands at both the A- and P-sites. Anisomycin acts exclusively against eukaryotic cells, whereas lincosamides specifically inhibit bacterial protein synthesis (reviewed in Ref. [28]). Structurally, anisomycin shares similarity with both puromycin and chloramphenicol; therefore, it

was not surprising that it was found to be bound at the A-site of the PTF center in the *H. marismortui* 50S subunit [112]. The aromatic methoxyl-phenyl ring of anisomycin inserts into a pocket created by A2451–A2452, where it makes stacking interactions with the latter. The N3 of the pyrrolidine sugar also forms a hydrogen-bond with A2452 and the hydroxyl group of this sugar, hydrogen-bonds with the O1P of 2504, and is co-ordinated by a potassium ion that is chelated between positions 2061, 2447 and 2501 of the 23S rRNA. The outcome is that anisomycin binds in a position that overlaps extensively with puromycin or an A-site-bound tRNA, except that anisomycin approaches the binding site from the opposite side. This means that it is predominantly the aromatic ring and methoxy sidechain that superimposes with the position of the amino acyl moiety (tyrosine-like moiety in the case of puromycin), whereas the pyrrolidine ring and tail have no counterpart, being located on the side opposite to the A-site ligand. What is surprising about the binding position of anisomycin is that no part of the molecule encroaches the position of the P-site. This suggests that anisomycin interferes with P-site binding indirectly, perhaps by inducing conformational changes within the PTF center. In fact, modest conformational changes are observed within the PTF center, but not with bases that directly interact with the CCA end of a P-site bound ligand.

Two commonly discussed lincosamides are lincomycin, naturally produced by several species of actinomycetes (such as *Streptomyces lincolnensis*, *espinosus* and *Actinomyces roseolus*), and clindamycin, a semi-synthetic derivative of lincomycin (Appendix E1). Although they exhibit similar affinities for the ribosome (5 and 8 μM respectively), clindamycin is generally a more effective inhibitor (probably due to better cellular uptake) and is used clinically, for example, as part of combination therapy, with pyrimethamine and folinic acid, as treatment against toxoplasmosis. The dual-site interference of lincosamides is evident from the inhibition by lincomycin of the transfer of fMet or AcPhe to Puro as well as preventing the binding of small tRNA 3′-end mimics, namely, CACCA-Leu to the A-site and CACCA-AcLeu to the P-site (see Ref. [132] and references therein). Furthermore, lincomycin has been shown to compete for binding with both erythromycin and Cam [28]. This latter point is consistent with observation that a number of strains exhibiting resistance to macrolides also protect the ribosomes from the action of lincomycin and streptogramin Bs (the so-called MLS_B resistance, reviewed in Ref. [133]).

In agreement with most of the biochemical data, the binding site determined for clindamycin spans between both A and P-sites at the PTF center, with the sugar moiety extending towards the tunnel and the prolyl moiety encroaching on the A and P-sites [127]. The majority of the interactions involve hydrogen-bonds from hydroxyl groups on the sugar moiety with nucleotides within the PTF center, specifically, to bases of A2058, A2059 and C2611 (U2590 in *D. radiodurans*), the phosphate of G2505 and the ribose of A2503 (Fig. 12.13). Bases A2058, A2059 and G2505 (as well as A2451) are strongly protected from chemical modification in the presence of clindamycin [134]. The same pattern of protection was found in the presence of lincomycin except that the protection of A2059 was absent (Fig. 12.2B). This suggests that there is some difference in the binding positions, although this is somewhat surprising since the structural differences between the two lincosamides occur at the C7

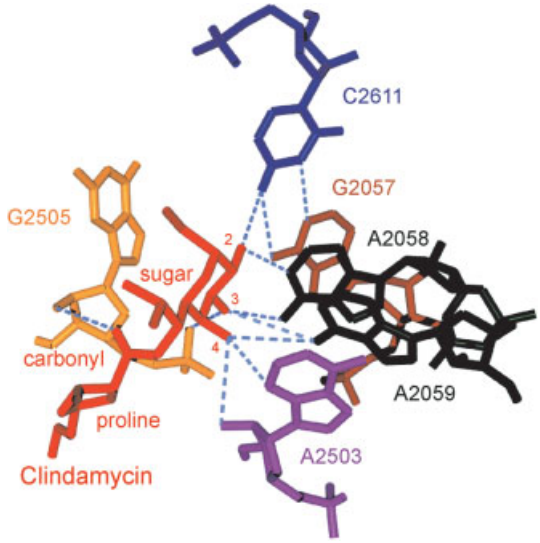


Figure 12.13 Interaction of clindamycin with the peptidyl-transferase center of the *D. radiodurans* 50S subunit (pdb1jzx; [127]). The majority of the interactions are with the OH groups of the sugar moiety of clindamycin (red) and consist of interactions between 2 OH and the base at position C2611 (blue), which is also base-paired with G2057 (brown). The bases of A2058 and A2059 (green) interact with 2, 3 and 4 OH groups. The sugar of A2503 (magenta) with 4 OH and the sugar and phosphate at position G2505 (yellow) with the carbonyl group and 3OH respectively.

position – a position that is expected to make a less significant contribution to the binding of the drug to the ribosome. The base specificity of the interaction with A2058 and A2059 correlates with the fact that mutations at these positions [135] and dimethylation of the N6 position of A2058 (which would prevent hydrogen-bonds with the 2-OH and 3-OH groups of the sugar moiety of the drug) confer resistance to clindamycin (in fact, to members of the MLS_B antibiotics in general). At the other end of the molecule, the proline group of clindamycin overlaps in position with that of phenyl group of Cam, in line with the A-site nature of clindamycin inhibition (Fig. 12.13). The 8′ carbon extending from the proline moiety of clindamycin comes within 2.5 Å of the N3 of C2452 and thus in close proximity to a P-site-bound tRNA (and A2451, thus accounting for the aforementioned protection). Thus, the binding position of clindamycin traverses both A- and P-sites and would be expected to disturb the positioning of substrates at both sites.

12.3.2.4 **Blocking the Progression of the Nascent Chain by the Macrolide Antibiotics**

Macrolides represent a large class of therapeutically useful antibiotics that have been extensively studied since the 1950s when the first member, erythromycin, was discovered and introduced clinically. Since then macrolide inhibition of ribosome

function has been the topic of many reviews (Refs. [7, 28, 111, 136–138], to name but a few). Macrolides are polyketide compounds synthesised by the actinomycetes and can be classified structurally into groups in a variety of different ways, the easiest being on the basis of the size of their lactone ring, which can vary significantly from as small as 8–12 ring members (methylmycin) to as large as 20 (rapamycin), but here we will consider mainly those of the between 14 and 16: for example, those with 14 (erythromycin, cethromycin, telithromycin and troleandomycin), 15 (azithromycin) or 16 (tylosin, spiramycin, and carbomycin A) atoms. Alternatively (or additionally), macrolides can be distinguished based on the number, position, size and type of sugar sidechains extending from the lactone ring, for example, erythromycin and azithromycin have single C3-cladinose and C5-desosamine sugar moieties, while the larger macrolide spiramycin has C5-mycaminose–mycarose disaccharide and tylosin, an additional C14-mycinose (see Fig. 12.14D and Appendices E2 and E3 for more details). Classification can also be applied using an evolutionary viewpoint since numerous attempts have been made, and are constantly being pursued, to develop more potent macrolide inhibitors. This has led to the discovery of the second-generation erythromycin derivatives, roxithromycin and clarithromycin, which exhibited a broader spectrum of activity. The emergence of bacterial strains resistant to both first- and second-generation macrolides has resulted in the recent introduction of the third-generation ketolide antibiotics.

Early studies suggested that macrolides have a single binding site on the 50S subunit (K_D in the range of 10–100 nM). The binding site was found to be vacant on free or initiating ribosomes, but unavailable in actively elongating ribosomes. The observation that most macrolides had no effect on the ribosomal PTF activity, coupled with the observed accumulation of short oligo-peptidyl-tRNAs in the presence of certain macrolides, led to the suggestion that the action of macrolide inhibition was to block the path of the nascent chain through the exit tunnel. In fact, a distinction can also be made functionally between particular macrolides: macrolides with extensive sidechains extending from position C5 have been shown to inhibit PTF activity, such as carbomycin (100% inhibition), spiramycin (85%) and tylosin (~60%), whereas those with shorter sidechains, such as erythromycin, do not [139].

Two regions of the 23S rRNA, the central loop of domain V and H35 of domain II (Fig. 12.2B), as well as two ribosomal proteins, L4 and L22, have been implicated with macrolide activity. Specifically, footprinting studies demonstrated that erythromycin protected A2058 and A2059 in domain V from chemical modification [128] and enhanced the reactivity of A752 in H35 of domain II [140, 141]. Methylation of A2058, as well as mutations at this and neighboring positions (2057, 2059, 2062 and 2611) confer resistance to macrolides, as do mutations in ribosomal proteins L4 and L22 (see Table 12.6).

The past 2 years has seen a plethora of publications reporting the structures of the 50S subunit in complex with many antibiotics, especially within the macrolide family (see Table 12.1): from the Yonath and Franceschi groups, the D50S subunit has been solved in complex with erythromycin, the second-generation derivatives clarithromycin, roxithromycin [127] and troleandomycin [142], the third-generation

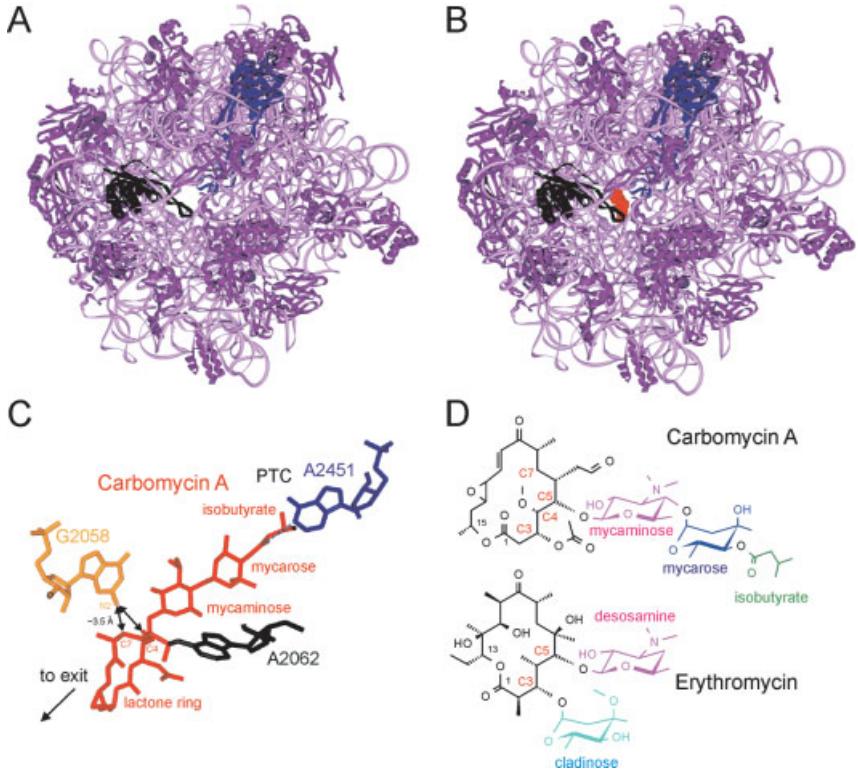


Figure 12.14 Macrolides bind within the tunnel of the 50S subunit. View from the base of the 50S subunit looking up the tunnel towards the PTC in the absence (A) and presence (B) of the macrolide carbomycin (red spacefill in (B)). The rRNA (pink) and ribosomal proteins (magenta) are in ribbons, with ribosomal protein L4 (blue) and L22 (green), whose long extensions reach into the tunnel of the 50S subunit, are highlighted. (C) The C5-sidechain of carbomycin (red) reaches into the PTC and approaches A2451 (very close to the site of peptide bond formation as discussed in Chapter 8.3). Also indicated are 23S rRNA residues A2062 (green), the N6 of which forms a covalent bond with the acetaldehyde group at the C6 position of the lactone ring of carbomycin, and G2058 (yellow), the N2 of which is in van der Waals distance with the C4 and C7 positions of the lactone ring preventing hydrogen bond foramation with the mycaminose C5 sugar position. (D) Comparison of chemical structures of carbomycin A and erythromycin, illustrating the sugar sidechains at the C3 and C5 position and the aldehyde of carbomycin at the C6 position.

ketolide antibiotics, cethromycin (more commonly known as ABT-773) [143], telithromycin (HMR-3647) [144], as well as the azalide azithromycin [143]. Steitz and coworkers [145] have reported H50S complexes with four different macrolides: azithromycin, spiramycin, carbomycin A and tylosin. This wealth of structural information enables us to attain a new level of understanding into the action of these antibiotics and reinterpret the accumulated treasure trove of biochemical and genetic data.

Consistent with the idea that the macrolide class of antibiotics inhibit progression of the nascent chain, the binding site of all the macrolide antibiotics determined to date was found to be within a distinct region of the tunnel, located adjacent to the PTF center. It is immediately apparent that in this position the tunnel lumen is significantly narrowed, being reduced in diameter from ~15 Å to almost 6–8 Å, such that the nascent chain cannot pass by (Figs. 12.14A and B). The binding site and interactions between different macrolides and ribosomal components exhibit extensive similarities, even when comparisons between the H50S and D50S structures are made (although minor but significant differences do exist and will be discussed later).

The general location is in agreement between each study to the extent that the common features of the binding site constitute primarily interactions of the lactone ring, as well as the sugar moiety attached to position C5, with the central portion of domain V. Residues in domain V that had been previously implicated in macrolide-binding make extensive interactions with the C5 sugars, including hydrogen-bonding and hydrophobic interactions with G2057, A2058, A2059 and G2505. They also provide an explanation as to why methylation, especially dimethylation at the N6 position of A2058, as well as the mutation A2058G, confer resistance to macrolides: modification of the N6 position would result in steric clashes between the methyl groups and the first C5 sugar (desosamine/mycaminose position), whereas replacing adenine with guanine at this position removes the hydrogen-bonding potential with the 2′-OH of this sugar. The importance that the sugar moieties make to the overall binding affinity of the macrolides is reflected by their sizeable contribution (between one-half to two-thirds) to the interaction surface with the ribosome [145]. Macrolides with an additional mycarose sugar attached to the mycaminose, such as carbomycin, spiramycin and tylosin, make an extra hydrogen-bond from the hydroxyl at position 3 with the ribose of G2505. Surprisingly, these larger macrolides also have a C6-ethyl-aldehyde, which forms a covalent carbinolamine bond with the N6 of base A2062 [145]. Tylosin and spiramyin have two further sugar sidechains that make additional contacts: tylosin has a mycinose sugar at the C14 position, which interacts with domain II establishing a hydrogen-bond between the N6 of A748 and 2′-OH of the sugar as well as contacting ribosomal protein L22. Methylation of A748, which confers resistance to tylosin would be predicted to disrupt this hydrogen-bond. In contrast, spiramycin contains a forosamine sugar, glycosidically linked to the C9 position that, despite being poorly resolved, clearly approaches L4 [145].

The orientation of the macrolides within the tunnel is such that the C5 position faces towards the PTF center. Thus, macrolides that bear longer C5 extensions restrict the length of the oligopeptide more than macrolides with shorter sidechains. Macrolides with C5 monosaccharides, such as erythromycin, have been observed to prevent progression of the nascent chain beyond a length of up to 6–8 amino acids [146], whereas macrolides with C5 disaccharides permit only dipeptide formation. The extreme case is carbomycin A, which has an additional isobutyrate extension on the C5 amino sugar enabling it to reach into the A-site of the PTF center,

where contacts with bases A2451 and A2452 are made (Fig. 12.14C). This explains why this drug is such a strong inhibitor of PTF activity and confirms the carbomycin/tylosin-specific protection of U2506 from chemical modification [139], the backbone of which is within 4–5 Å of the isobutyrate extension.

The removal of the desosamine sugar at the C5 position totally abolishes the effectiveness of these compounds, for example, although rapamycin still binds to the ribosome its exhibits little or no inhibitory activity for bacterial translation. In contrast, the importance of the cladinose sugar attached to the C3 position of the lactone ring seems to be less clear. The simplest ketolide (RU 56006), a derivative of erythromycin where a ketone group replaces the C3 cladinose (the basis for the name ketolide), reduces drug binding 70-fold (a change in K_D from 0.014 μM (erythromycin) to 0.980 μM; [147, 140]), suggesting this position influences the binding affinity of the drug. In the D50S-azithromycin structure, a hydrogen-bond from the cladinose is formed with the N4 of U2586 [143], however, this is not observed in the D50S-erythromycin or H50S-azithromycin structures, where the cladinose sugar seems to make a comparatively low contribution to binding and, although G2505 comes within 3–6 Å of the single hydroxyl group (4′-OH) of the cladinose sugar, it cannot make hydrogen-bond interactions. This latter point is at least consistent with the fact that this hydroxyl group has been shown to be dispensable for anti-microbial activity [148]. In contrast, the broader spectrum of activity of the ketolides, despite the absence of the C3-cladinose, seems to be related to the presence of their additional sidechains and modifications, for example, the ketolides ABT-773 and telithromycin have a cyclic carbamate inserted at the C11–C12 position of the lactone ring and an alkyl–aryl sidechain attached to the C6–O position or to the carbamate, respectively. In the D50S-ABT-773 structure, the N2 of the carbamate forms a hydrogen-bond with O4 of U2609, whereas the N3 of the quinolylallyl can hydrogen-bond with the O2′ of U790 of domain II of the 23S rRNA. Since telithromycin has the alkyl–aryl side chain attached to the N2 of the carbamate, the analogous hydrogen-bond with U2609 cannot be formed; however, the sidechain itself also makes contacts with domain II of the rRNA. The position of attachment and the longer aryl-linker allows it to penetrate deeper into the tunnel, binding in a cleft composed of positions 789–791 and A764 of domain II. These additional interactions might explain why the ketolides have ~10-fold higher affinity for the ribosome than macrolides [140]. Significant interaction between ketolides and domain II is supported by the strong protection of A752 from chemical modification by ketolides, such as telithromycin, as well as mutations at position U754A in H35 that confer resistance to telithromycin [147, 145, 141]. However, these residues are not directly involved in ketolide binding; thus their influence must be exerted through conformational changes in the loop connecting H35a and H32 [143].

Despite the observation that mutations in L4 and L22 confer resistance to erythromycin, no direct interactions between erythromycin and any part of these proteins was evident, suggesting that resistance is conferred indirectly by inducing conformational changes within the rRNA. This is easy to envisage since the β-hairpin structures extending from L4 and L22 interact intimately with the rRNA associated

with the macrolide-binding site. This observation was not altogether surprising, since it had been reported that these mutations perturbed the 23S rRNA in the vicinity of the macrolide-binding site [149]. Although interaction between sidechains of other macrolide antibiotics and these ribosomal proteins has been observed, for example, the C9–forasamine sugar of spiramycin and L4, the C14 mycinose sugar of tylosin with L22 [145] and the second (see following) azithromycin-binding site with both L4 and L22 [143], the sites of interaction are distinct from those where the mutations arise. Perhaps the most intriguing interaction is made by troleandomycin, the binding position of which is displaced deeper in the tunnel, where it may transiently interact with L22 to induce the β-hairpin of L22, which normally lies flat against the tunnel wall, to swing out across the tunnel [142]. The implication is that this conformational change provides insight into a ribosomal gating mechanism, whereby specific sequences within the nascent chain are recognized within the tunnel to regulate their translation (reviewed in Ref. [150]).

Since the azalide azithromycin has been solved in complex with both the D50S [143] and H50S subunits [145], this provides us with the opportunity to compare directly their interactions. An immediately apparent difference is that two binding sites were found in the D50S structure, one that is equivalent to the position determined for all the other macrolide antibiotics and a second site located deeper in the tunnel, directly following the first. This second binding site is likely to be specific for *D. radiodurans*, since the interactions involve residues specific to this species; however, it nicely illustrates how small sequence deviations can subtly alter the rRNA architecture, which in turn can markedly influence the potential for antibiotic interaction. This is also evident when making a comparison between the first binding site of the D50S structure with the H50S-azithromycin-binding position: comparing the conformation of the macrolides themselves reveals no significant difference in the orientation of the C3 cladinose sugar with respect to one another. However, the orientation of the lactone ring is significantly displaced, particularly on the C9–C14 side, which faces the solvent. This may in part result from the additional azithromycin-binding site in the D50S structure, since there is direct interaction between the two molecules, via a hydrogen-bond between the desosamine sugar and the O1 in the lactone ring of the first binding site [143]. On the whole, it seems that small but distinct changes within the 23S rRNA seem to be responsible for altered lactone ring orientation; however, at a few positions more significant differences are observed, for example, the orientation of bases 2586 and 2609 are sufficiently different to preclude even similar binding modes for the lactone ring. It is interesting to note that although both *H. marismortui* and *D. radiodurans* 23S rRNAs have pyrimidines at these positions they differ in whether they have cytosine or uracil. Furthermore, the exchange of purines is seen at position 2058, which is adenine in all bacterial species (D50S) and guanine in most archaea (H50S) and eukaryotes. Replacing adenine with guanine at 2058 eliminates the potential for H-bonding with the C5-sugar and may disturb the orientation of the antibiotic, thus explaining the natural resistance of *H. marismortui* to this class of antibiotics. That the A2058G substitution in *E. coli* ribosomes reduced the binding of erythromycin by almost four orders of magnitude

[147] could explain why even at high antibiotic concentrations (1–10 mM) no binding of erythromycin was observed when soaking the H50S crystals [145], whereas more physiological concentrations (10–100 μM) were sufficient for the D50S complexes [143, 127]. However, when comparing the position of the lactone ring of the azithromycin structures (and the larger 16-membered ring macrolides) with those of the smaller 14-membered ones, even more marked differences are observable. In the latter structures, the lactone ring adopts a more compact "folded-in" structure, whereas the former exhibited an extended or "folded-out" conformation [145]. The upshot being that the lactone ring of the smaller macrolides is almost perpendicular to that observed for the larger ones and in fact results in a larger constriction of the tunnel [143]. It would be interesting to see whether this is a general feature of the 14-membered ring macrolides or whether the conformation depends on species-specific interactions: the structures of the same antibiotics in complex with different species will go some way to answering these questions.

12.3.2.5 **Streptogramins**

Streptogramins are produced as a mixture of two chemically unrelated compounds, type A and type B (Appendix E3), which act synergistically *in vivo* and *in vitro*, such that the binding of one class stimulates the binding of the other. The net result is that significantly lower concentrations of both antibiotics are needed to obtain the same level of inhibition compared with the use of each compound separately, for example, a 20-fold lower concentration of virginiamycin M and pristinamycin I_A (streptogramin A (S_A) and B (S_B), respectively) was needed when used in combination, than when used alone [43]. However, most importantly, the combination of some streptogramin A and Bs can convert a bacteriostatic effect into bactericidal lethality. This, coupled with the observation that bacteria resistant to the MLS_B class of antibiotics are still sensitive to streptogramin A inhibition, has led to the use of a streptogramin A and B mixture (in ratio of 3:7) of dalfopristin and quinupristin as a new antimicrobial agent [151] called Synercid®, marketed by the company Rhone-Poulenc Rorer.

In general, S_A antibiotics have been reported to interfere with the puromycin reaction and with P-site binding, suggesting that S_A act at both A- and P-sites. In contrast, antibiotics of the B type (S_B) do not tend to inhibit the puromycin reaction, but rather stimulate the binding of 3′-terminal fragments of aminoacyl-tRNA or N-blocked aminoacyl-tRNA to the ribosome. Binding of virginiamycin S (S_B) is inhibited by erythromycin and stimulated by virginiamycin M, the corresponding S_A. Virginiamycin S has been crosslinked to ribosomal proteins L18 and L22 and protects nucleotides A2062 and G2505 within the central loop of domain V of 23S rRNA from chemical modification. These bases are also protected by vernamycin B, another member of the S_B group, which additionally shows strong protection of A752 as well as other weaker affects. Virginiamycin M (S_A) protects A2037, A2042, G2049 and C2050 near the PTF ring (Fig. 12.2B), suggesting that binding of S_A to the ribosome induces conformational change within the PTF center [152]. In this

regard, it is interesting to note that suppression of bacterial growth persists for a prolonged period subsequent to the removal of S_A drugs [153, 154]. The implication here is that S_A binding induces a conformational change within the PTF center that is *slowly* reversible.

Conformational change within the ribosome seems to be also important for the synergistic action of S_A and S_B. The mutation A2058U in 23S rRNA causes resistance to MLS_B antibiotics and prevents binding of virginiamycin S, whereas the inhibitory action of the A-type streptogramin virginiamycin M remains unaffected. In the presence of virginiamycin M (S_A), however, the binding of virginiamycin S (S_B) to mutant ribosomes occurs with the same affinity as with wild-type ribosomes. Indeed, in some cases, following the removal of the S_A compound, the affinity of the S_B antibiotic for the ribosome is still enhanced, providing a strong argument for conformational changes in the PTF center induced by S_A binding.

A crystal structure of virginiamycin M in complex with the *H. marismortui* 50S subunit shows that this S_A does indeed bind in a position that overlaps both A- and P-sites [112], consistent with the earlier biochemical information. Specifically, the majority of the 20-membered lactone ring overlaps the P-site, whereas only the oxazole ring is inserted into the hydrophobic A-site crevice [112]. The conjugated amide group of virginiamycin M occupies the position that A2602 has in the native H50S structure, whereas the base of A2602 is rotated by 90° to end up in a position perpendicular to the tunnel wall. However, similar conformations for A2602 have also been observed in the presence of tRNA mimics, the binding sites of which do not overlap with the 'native' position of A2602. This suggests that overlap in position of the conjugated amide group of virginiamycin M with the base of A2602 is not *per se* necessary for the change in the position of this base.

Other than the rotation of A2602, little other change is observed within the PTF center, which is surprising since this does not reflect the significant conformational changes predicted from biochemical experiments. Furthermore, in these co-crystallization experiments, virginiamycin S (S_B) was also included, yet only poorly resolved electron density was observable. Despite this, the highest additional electron density was located along the opposite face of the rotated A2602, thus sandwiching between both streptogramins, possibly hinting at the basis for co-operative nature of S_A and S_B binding. It is tempting to speculate that because *H. marismortui* is naturally resistant to the MLS_B class of antibiotics (due mainly to guanine at position 2058), this may provide a reason for the weak binding of the S_B antibiotic that was observed. However, it does not explain the absence of any observable rearrangement within the PTF center, nor that in the presence of S_A, the A2058G mutations, or methylation at this position, hardly affect S_B affinity (although whether this also applies to archaea-bacteria remains to be seen).

Recently, the structure of D50S in complex with both dalfopristin (S_A) and quinupristin (S_B) was determined [274]. The binding site of dalfopristin was consistent with that of virginiamycin M (S_A) in the H50S structure, and quinupristin was shown to occupy a position similar to that of the macrolide antibiotics. Analysis of this structure suggested that the synergistic binding of S_A and S_B antibiotics probably results

from their direct interaction on the ribosome, with each other, as well as indirect contact through residue A2602 [274]. In contrast to the H50S structure, significant conformational changes were observed in the PTF centre, particularly, the universally conserved residue U2585, which is displaced by dalfopristin binding and rotates 180° to establish hydrogen bonds with C2606 and G2588 of the 23S rRNA. This stable non-productive conformation provides an explanation for the bacteriocidal properties and the prolonged inhibitory effects of the S_A antibiotics.

12.3.2.6 New Classes of Translation Inhibitors; the Oxazolidinones and Novel Ribosome Inhibitors

The oxazolidinone class of antibiotics represent the first new class of drug to enter the antibiotic market in over 20 years, with the most well-characterized member being linozolid (Lin), marketed by Pharmacia under the name Zyvox® (PNU-100766; Appendix E1). The oxazolidinones act against a wide spectrum of Gram-positive and anaerobic bacteria and, of special importance, also exhibit activity against multi-drug-resistant bacteria.

There has been some confusion as to the exact binding site for this class of antibiotics. *In vitro* experiments detected crosslinks between oxazolidones and rRNA from both the small (residue A864) and the large subunit (all residues detected were within domain V near to binding site of the E-site tRNA and ribosomal protein L1) [155]. However, these positions are at odds with the multitude of oxazolidinone-resistant mutants, all of which map within the vicinity of the PTF center, i.e., far from L1 and the E-tRNA-binding site (Fig. 12.2B). This discrepancy between the biochemical and genetic data has been attributed to the tendency of the oxazolidinones to bind non-specifically *in vitro* [137]: The idea that oxazolidinones may bind a particular conformation of the ribosome led to the use of *in vivo* crosslinking experiments with an ^{125}I-labelled oxazolidinone to address this issue. Residue A2602 of the 23S rRNA and the N-terminal of ribosomal protein L27 were identified as being in proximity to the drug-binding site [156]. The specificity of the crosslink was demonstrated by addition of the non-labelled oxazolidinone eperezolid, which abolished the crosslink, or its inactive enatiomer (PNU-107112(R)), which exhibited no effect on the crosslink efficiency. Furthermore, a marked reduction in crosslinking was observed in two strains of *Staphylococcus aureus* bearing mutations (G2447U and G2576U) conferring resistance to oxazolidinones. The close proximity of A2602 to these various mutation sites not only provides support for the PTF center as the oxazolidinone-binding site but may provide some general information as to the orientation of the drug, namely that the side of ring C to which the azido group is attached should extend towards A2602, whereas the pharmacophoric oxazolidinone ring A (and attached N-acetylaminomethyl side chain) angles toward the PTF center, where the resistance mutations are located. Crosslinking to the N-terminal region of L27 is also consistent with the PTF center location, since crosslinks to L27 were also observed from other antibiotics that bind to the PTF center as well as 3′-azido-labeled aa-tRNAs (see Ref. [156] and references cited therein). Unfortunately, the

N-terminal region of L27 is disordered in the *D. radiodurans* 50S crystal structure so it remains to be seen how close this flexible extension comes to the PTC [39].

It should be noted that there seems to be a unique species-specific pattern of resistance to oxazolidinones, since there is little overlap in the reported mutation sites (Table 12.5; [137, 157]). This is probably due to slight differences within the PTC of the ribosomes, which may affect the drug-binding site and/or could simply reflect variation in the tolerance to mutations at specific sites in the PTC of the respective ribosomes. In this respect, it is interesting that the G2447U mutation in *Mycobacterium smegmatis* is lethal in *E. coli* (see Ref. [157] and references therein).

The specific binding of oxazolidinones with the 50S subunit is supported by the fact that ^{14}C-labeled eperezolid (PNU-100592(S)) bound only the 50S subunit (K_d = 20 μM), a result which was recently supported by NMR studies [158]. That the

Table 12.5 23S rRNA mutations of different organisms conferring resistance to the oxazolidinone linezolid

Organism	Mutation position[a]	Mutation creation[b]	Selected references
H. halobium	U2500C$^{>200}$	spon.	160
	U2504C^{180}	spon.	
	C2452U^{160}	spon. + engin	
	A2453G^{130}	spon.	
	A2062C^{80}	cross.	
	C2499U^{30}	cross.	
	A2453C^{15}	cross.	
E. coli JM109	G2032A	spon.	247
E. coli HN818	G2032A^{80}	spon. + engin.	247
	G2032C^{60}	engin.	
	G2032U^{40}	engin.	
	G2447U$_{\sim 20}$[c]	engin.	
M. smegmatis (rrn-)	G2447U	spon. + engin.	157
Enterococcus sp.	G2505A G2576U	spon.	248
Staphylococcus sp. / Streptococcus sp.	G2447U G2576U	spon.	275

a *E. coli* numbering is given; the minimal inhibitory concentration (MIC) is given where known (μM) as the superscripted number following the base substitution.

b Spon. and Engin. indicate whether the resistance was discovered through selecting for spontaneous-resistant mtutants or whether the mutation was engineered and then the level of resistance being determined. Lin mutants originally discovered as having resistance to another antibiotic are indicated by +.

c It should be noted that resistance was conferred in *trans* by overexpressing the 23S rRNA bearing this mutation since attempts to generate this mutation were unsuccessful suggesting it is lethal in *E. coli* [250].

binding position locates to the PTC was indicated biochemically by the competition for binding between eperezolid and chloramphenicol, lincomycin and clindamycin, but not puromycin [159]. Consistently, neither eperezolid, nor Lin, has been shown to inhibit the puromycin reaction [159, 160], although certain derivatives of Lin (PNU-176798 and DuP 791) have [157, 161, 162]. Whether this reflects differences in the drugs or the biochemical systems remains to be resolved. In any case, there is mounting evidence suggesting that the oxazolidinones exert their influence through the positioning or accommodation of initiator fMet-tRNA on the ribosome, leading to the proposal that the site of action of oxazolidinones may in fact be the initiation phase of translation [162–165]. In support of this, Colca et al. [156] also detected an oxazolidinone specific crosslink to tRNA; unfortunately, however, whether the species was the initiator-tRNA could not be determined. If this were true, then one could envisage that the PTF activity *per se* is not inhibited by the oxazolidinones, instead this inhibition is by incorrect positioning of the P-site substrate. Additionally, the oxazolidones (as well as chloramphenicol) have been shown *in vivo* to induce translational inaccuracy in the form of increased frameshifting and nonsense suppresion, which was proposed to result from perturbation of tRNA–ribosome interactions [166]. Moreover, Matassova et al. [155] detected no inhibition on any steps leading to dipeptide formation, but observed inhibition during the elongation phase, prompting the suggestion that the translocation reaction may be the site of action. In this regard, it is interesting that the residue crosslinked by the oxazolidinones, A2602, has been proposed to play a role in guiding the translocation of the CCA–end of the post-peptide-bond formation peptidyl-tRNA from the A- to P-site [108]. However, the azido crosslinking moiety that was attached to the ring C of Linezolid adds significant length to the drug making it unlikely that the drug actually comes into contact with A2602. Further work will be required to reconcile the conflicting data and determine the exact mechanism of inhibition of this clinically important class of antibiotics.

The search for novel ribosome inhibitors (NRI) has led to the recent discovery by Abbott laboratories of a series of translation inhibitors [167], structurally related to the antibacterial quinolones (Appendix E1). The quinolones are well known as DNA gyrase and topoisomerase IV inhibitors; therefore, it was surprising that small modifications in structure were sufficient to completely alter the mechanism of action of these compounds turning them into translation inhibitors. Moreoever, the specificity of action of the NRI compounds was such that both Gram-positive and Gram-negative bacteria were inhibited, including a number of common respiratory pathogens, while human cell lines remained unaffected [167]. Perhaps the most exciting observation is that the NRI compounds appear to inhibit translation using a new mechanism since NRI-resistant strains show no cross-resistance with other translation inhibitors, such as macrolides, tetracyclines, aminoglycosides or oxazolidinones. Defining the binding site of these antibiotics on the ribosome is certain to provide insights into their mechanism of inhibiton.

12.3.3 Translocation Inhibitors

12.3.3.1 Thiostrepton and Micrococcin

Thiostrepton, a modified peptide antibiotic (Appendix F1), has several effects on a number of partial reactions during translation. One prominent effect is the inhibition of EF-G-dependent uncoupled GTPase, i.e., a strong EF-G-dependent GTPase activity in the presence of idle ribosomes (those not active in protein synthesis, i.e., non-polysomal). However, thiostrepton blocks spontaneous (EF-G-independent) translocation as well. Furthermore, the antibiotic inhibits A-site occupation of the e-type and to a lesser extent the A-site binding of the i-type. Therefore, the prevalent effect of thiostrepton is probably an inhibition of the transition between PRE and POST states, regardless of the direction [47].

Thiostrepton binds to the 50S subunit with a remarkably high association constant ($K_a > 10^9\ M^{-1}$) [168], which is 2–4 orders of magnitude higher than the association constant for many other antibiotics and six orders of magnitude higher than that for puromycin. Thiostrepton also binds to naked 23S rRNA and this affinity is enhanced in the presence of L11; however, L11 in isolation does not bind the drug. L11 binds a fragment of 23S rRNA (nucleotides 1052–1112), as do thiostrepton and micrococcin, an antibiotic structurally related to thiostrepton (see Appendix F1), providing strong evidence that both drugs bind within this region. The producer of thiostrepton, *Streptomyces azureus*, protects its own rRNA by introducing a methyl group onto the ribose moiety of A1067, thereby causing resistance to thiostrepton (and micrococcin). This small 2′-O-methyl group decreases the binding affinity of thiostrepton by at least six orders of magnitude [168].

The 23S rRNA/L11 complex, together with the pentameric complex L10•$(L12)_4$, has been implicated in elongation factor-dependent GTP hydrolysis. An EF-G crosslink within the 70S ribosomes maps to a fragment of 23S rRNA containing A1067, and bases A1067 and A1069 are protected by EF-G in addition to those protected in the sarcin-ricin loop [169]. Although EF-Tu shows no direct protection of bases in domain II of 23S rRNA [169], a mutation of A1067 affects EF-Tu function, and residues G1041, G1068 and G1071 are shielded by aminoacyl-tRNA-bound enzymatically to the A-site with or without kirromycin [170]. Chemical and enzymatic probing studies show that thiostrepton protects 13 nucleotides between A1067 and A1098 [171]. The same set of protections were found in the presence of micrococcin with one important exception: namely that the N-1 position of A1067 was protected against dimethylsulfhate (DMS) by thiostrepton but exhibits enhanced reactivity upon micrococcin binding. Since A1067 is also strongly shielded by EF-G, this observation may well be related to the contrasting effects that thiostrepton and micrococcin have on EF-G-dependent GTPase, i.e., thiostrepton is inhibitory whereas micrococcin is stimulatory (reviewed in Ref. [64]).

The protection pattern of thiostrepton encompasses the loops of two helices of the 23S rRNA, H43 and H44. The importance of both loop regions for drug binding is

supported by the fact that methylation of A1067 confers thiostrepton resistance [168]. Similarly, mutations of the residue A1067 to U or C and, with a weaker effect, to G, result in thiostrepton resistance *in vitro*. The mutations at both A1067 and A1095 strongly reduce the binding of micrococcin and thiostrepton to the ribosome, whereas L11 binding is not affected by these mutations. Note that methylation of the ribose at position 1067 is a much smaller change when compared with a transversion at this position, and yet the effect of methylation on thiostrepton binding is much more drastic.

The L11-binding region of *E. coli* 23S rRNA can be replaced by the homologous stretch of residues from the yeast *S. cerevisiae*, with the result that these engineered ribosomes are also thiostrepton-sensitive [172]. Potentially, this raises a paradox, since yeast ribosomes are naturally resistant to thiostrepton. However, the naked yeast 26S rRNA is able to bind thiostrepton and thus the potential thiostrepton-binding site may simply be masked by proteins in the intact yeast ribosome. The exchange of the L11 binding site between *E. coli* and yeast is equivalent to the simultaneous introduction of 20 mutations between positions 1056 and 1103. Despite these differences in primary sequence, the conformation of this region is likely to be evolutionary highly conserved, since it has been demonstrated that L11 can recognize the corresponding region in the rat 28S rRNA and *vice versa* [173]. Furthermore, *E. coli* L11 and thiostrepton bind co-operatively to the *E. coli* 23S rRNA region and to the rat 28S rRNA region, if position 1878 (the equivalent of 1067 in *E. coli*) is mutated from G to A. No such co-operativity was seen with rat L12 (the *E. coli* L11 homolog). The lack of co-operativity in eukaryotic ribosomes may also play a role in thiostrepton resistance [173].

Protein L11 consists of two domains: the C-terminal part is responsible for the tight interaction with the rRNA-binding region for L11, and the N-terminal part is required for the co-operative binding of thiostrepton. Mutations in the N-terminal region (specifically, substitutions of P18S/T and P22S/T in *E. coli* L11) confer resistance to thiostrepton [174, 175], although not by affecting interaction of thiostrepton with the rRNA, but perhaps by allowing L11 the freedom to move despite the presence of thiostrepton. Interestingly, in bacterial and archaeal ribosomes, on which thiostrepton is active, P18/22 in L11 are conserved, whereas in the equivalent position in eukaryotic L11 the prolines are not conserved, consistent with their natural resistance. The crystal structure of L11 in complex with its rRNA-binding site has been solved to 2.57–2.8 Å resolution [176, 177]. It has been proposed that the sugar of A1067 and the base A1095 interact directly with thiostrepton by forming a binding pocket in conjunction with the prolyl residues of the N-terminal domain of L11 (Fig. 12.15). Unfortunately, the L11 region (as well as the pentameric L10•$(L7/L12)_4$) are highly disordered in all the available high-resolution crystal structures of the ribosomal subunits [38, 39], making determination of thiostrepton-bound 50S subunit structures more complicated.

There is some controversy over the exact step of thiostrepton action. In contrast with numerous previous results (reviewed in Ref. [28]), a recent publication purported that thiostrepton allowed binding of EF-G to the ribosome, a single translocation reaction and hydrolysis of GTP, but prevented release of EF-G.GDP [178]. These

thiostrepton-stabilized EF-G–ribosome complexes were analyzed by cryo-EM [179]. The conclusions drawn from this analysis, interpreted on the basis that the thiostrepton-stabilized EF-G was the GDP conformer, was at odds with the interpretation of the fusidic acid complexes (see Ref. [180]). In direct contradiction with this model, recent evidence was presented demonstrating that thiostrepton is, in agreement with earlier works, a strong inhibitor of EF-G-dependent GTP hydrolysis and acts by destabilizing the interaction between EF-G and the ribosome [181]. Numerous attempts were made to reconcile the two disparate results but even repeating the experiments under the conditions used by Rodnina and coworkers [178], the conclusion remained the same, namely that thiostrepton strongly inhibited GTP hydrolysis and association of EF-G with the ribosome. Additionally, thiostrepton was shown to stimulate IF2-dependent GTP hydrolysis, whereas micrococcin stimulated the GTP hydrolysis of both factors. These results can be interpreted best if thiostrepton (and micrococcin) are thought to stabilize the L11 region in a particular conformation that prevents (or reduces) stable binding of EF-G and EF-G-dependent GTP hydrolysis. In the case of micrococcin the weakened interaction of EF-G with the ribosome, although sufficient to stimulate GTP hydrolysis, results in rapid dissociation

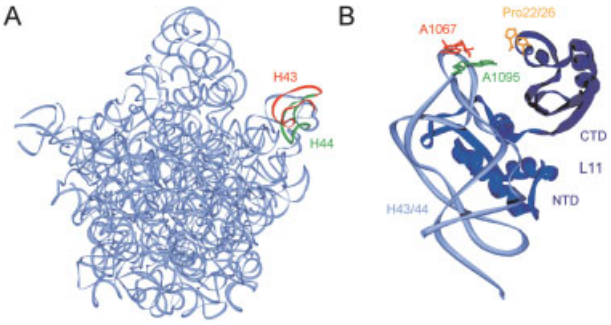


Figure 12.15 Putative thiostrepton binding site on the 50S subunit. (A) Overview of the L11-binding site on the *D. radiodurans* 50S subunit. Only 23S and 5S rRNA are shown in ribbons representation (pdb file 1kpj) with the L11-binding site, H43 (red) and H44 (green) indicated. Note this region is relatively disordered in both D50S and H50S structures with the L11 protein being totally disordered in the H50S structure (pdb 1jj2) but the CTD being present in the D50S structure (pdb 1kpj). (B) The putative thiostrepton binding site illustrated using the isolated L11-rRNA structure (pdb1mms; [177]). Ribbons representation of a fragment of RNA (light blue) that mimics the L11 binding sites. Ribosomal protein L11 is shown with the C-terminal domain (CTD, blue), which binds to the rRNA fragment and the free N-terminal domain (NTD, dark blue) indicated. Mutations in H43 (base A1067 shown in red) and H44 (with base A1095 in green) of the 23S rRNA and in L11 (Prolines at position 22 and 26 in yellow) that confer resistance to thiostrepton are indicated.

of EF-G•GDP from the ribosome, thus resulting in increased turnover and therefore accounting for the observed stimulation of GTP hydrolysis [181].

12.3.3.2 **Viomycin Blocks Coupled GTPase Activity**

Viomycin is a cyclic peptide antibiotic composed of an unusual assortment of amino acids; the non-coded tuberactidine and ureidodehydroalanine in addition to β-lysine, serine and diaminopropionic acid (as seen in Appendix F1). Viomycin (sometimes referred to as tuberactinomycin B) belongs to the tuberactinomycin family of antibiotics and has seen limited use against tuberulosis. Viomycin exhibits inhibitory characteristics similar to the aminogylcoside family of antibiotics, in particular to hygromycin B, since viomycin acts both to induce misreading as well as inhibit EF-G-dependent translocation *in vitro* [182], but not GTP hydrolysis [178]. In fact, viomycin has been shown to compete with aminoglycosides for binding to the 30S subunit and *vice versa* [183]. However, the binding site of viomycin, although still uncertain, is probably very different to that of hygromycin B, since it seems to encompass components from both the small and large subunits: Resistance to viomycin results from ribosomes that have alterations in either rRNA from the small or large subunit, although the exact location of these mutations has not been determined [184]. Protection from chemical probing by viomycin identified bases 912–915 and A1408 in the 16S rRNA and in addition bases U913 and G914 in the 23S rRNA (Ref. [128] and unpublished data cited in Ref. [185]; see Fig. 12.2). Furthermore, conformational changes in the ribosome, measured using a toeprinting assay, demonstrated that viomycin induced an effect only with 70S ribosomes and not with 30S subunits, in contrast with all the aminoglycosides tested which affected both [185]. This suggests that viomycin may not even bind the 30S subunit in the absence of the 50S and further emphasizes the differences between viomycin and the aminoglycosides. An *E. coli* mutant strain lacking N1 methylation of G745 in H35 of domain II of the 23S rRNA also exhibits 4-fold increased resistance to viomycin as well as having a markedly reduced growth rate and a large reduction in the number of 70S ribosomes [186]. Furthermore, viomycin protected G914 from kethoxal probing more efficiently in strains lacking the G745 methylation than the wild type, suggesting that the relatively moderate increase in the level of resistance conferred by the methylation probably does not reflect a specific resistance mechanism but results from conformational changes in the 50S subunit that affect the viomycin-binding site [186]. Collectively, these data suggest that viomycin binds at the interface of the ribosomal subunits; however the presence of two binding sites, one on each subunit, cannot be excluded.

Recently, viomycin has been shown to prevent the EF-G (and RRF)-mediated release of tRNA from the ribosome during ribosome recycling [187]; however, not as effectively as thiostrepton and, unlike thiostrepton, viomycin did not inhibit the release of the RRF [188]. This suggests that viomycin and thiostrepton operate using different mechanisms, i.e., maybe viomycin allows EF-G binding and GTP hydrolysis, whereas thiostrepton severely reduces the binding of EF-G (as mentioned in Sect. 12.3.3.1).

Another feature that viomycin has in common with aminoglycosides relates to competitive inhibition of several ribozymes and also of group I intron splicing (Ref. [189]; reviewed in Ref. [190]). This finding led to the provocative proposal that low-molecular-weight molecules, such as small RNAs or antibiotic progenitors, may have been present in the primordial soup and co-evolved with the modern ribosome, originally serving a regulator function [191]. Indeed, the recent findings that Ede and Pct work antagonistically to influence initiation complex formation certainly demonstrates the potential existing for such small molecule regulation (see Sects. 12.2.2 and 12.2.3).

12.3.3.3 **Spectinomycin Interferes with EF-G Binding**

Spectinomycin is sometimes listed under "aminoglycosides", but it has in fact nothing in common with this group of antibiotics, neither structurally (Appendix F1) nor functionally. The drug inhibits translocation by interfering with the binding of EF-G to the ribosome, probably by preventing the conformational changes in the ribosome associated with EF-G binding. Consistent with this idea, the binding position of spectinomycin has been located to the elbow of h34 and h35 within the head of the *T. thermophilus* 30S subunit (Fig. 12.16) [67]. Helix 34 has a putative role in trans-location of the tRNAs from the A to P-site, probably requiring a rearrangement in its interaction with neighboring helices 35 and 38. Spectinomycin makes interactions

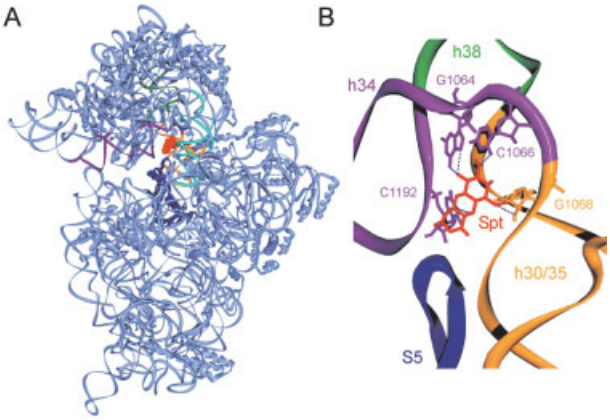


Figure 12.16 Spectinomycin binding site on the 30S subunit. (A) Overview of the spectinomycin binding site on the 30S subunit (pdb 1fjg [67]). Ribbons representation of 16S rRNA (light blue), with h34 (purple), h30/35 (yellow), h38 (green) and h28 (cyan) highlighted, as well as ribosomal protein S5 (dark blue). (B) Close-up view of the spectinomycin binding site at the elbow junction of h34 and h30, where hydrogen bond interactions (dashed blue line) between spectinomycin (Spt, red) and G1068 (h30) and C1066, G1064 and C1192 (counter clockwise) are shown. Other colours as in (A)

with three bases of h34 (G1064, C1066, and C1192) and with the phosphate of G1068 in h35, which together may stabilise this region, preventing the conformational changes necessary for translocation. This is supported by cryo-EM reconstructions of functional states that reveal movements within the head region during translocation [192].

The determined spectinomycin-binding site is in agreement with spectinomycin-resistant mutations that map within h34, specifically nucleotides G1064, C1192 and C1066 [193–196]. Furthermore, spectinomycin protects C1063 and G1064 from chemical probing [23] and overexpression of an RNA fragment resembling helix 34 confers resistance to spectinomycin [197]. The fragment is proposed to sequester the drug thereby permitting the intact ribosome to function and consequently display a spectinomycin-resistant phenotype. Surprisingly, mutations in S5 also give rise to spectinomycin resistance [198], although S5 is not required for spectinomycin binding. These mutations map to a loop in S5 that does not make direct interaction with spectinomycin; instead it stabilizes the pseudoknot region that is connected to h34 through a network of interactions. This led to the proposal that mutations in this loop, by disrupting this network of interactions, allow the head to move freely during translocation, even when spectinomycin is bound [67].

12.3.3.4 **Fusidic Acid is the Counterpart of Kirromycin**

Fusidic acid is a steroidal antibiotic (Appendix F1), which has been extensively studied since the 1960s and whose mode of action was well characterized by the mid-1970s (reviewed in Ref. [28]). Fusidic acid allows translocation and GTP hydrolysis, but prevents the associated conformational changes in EF-G, thus stabilizing the EF-G·GDP complex on the ribosome in an analogous fashion to how kirromycin prevents dissociation of EF-Tu. It is used clinically, primarily against *S. aureus*, from which most resistance mutations have been selected [199, 200]. Most resistance mutations are clustered within three distinct regions that map to the *fusA* gene, which encodes EF-G (Fig. 12.17A) [200]. According to the crystal structures of EF-G·GDP and nucleotide-free EF-G, the three mutational regions are confined to a central region of EF-G, localizing to a three-way junction between the G domain and domains III and V. It is possible that fusidic acid binds to the interface of these three domains and by restricting their movement prevents EF-G from adopting the GDP conformation (reviewed in Ref. [201]). Therefore, mutations within this region probably either facilitate the conformational changes in EF-G required for dissociation from the ribosome despite the presence of the drug, whereas others may simply prevent the drug from binding to EF-G [200].

The high occupancy and stability of the fusidic acid-bound EF-G ribosome complex has led to medium resolution cryo-EM reconstructions revealing the binding and interaction of EF-G with the ribosome [202–204]. Fusidic acid does not work on eukaryotes, but sordarin is thought to act in an analogous fashion on yeast EF2 (the homolog of EF-G). Indeed, this antifungal was used to construct stable EF2-80S yeast ribosome complexes for cryo-EM analysis (at 17.5 Å [205] and more recently at

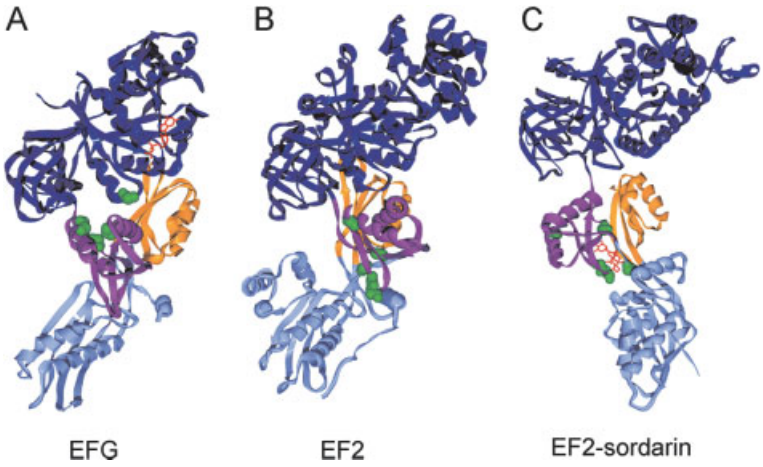


Figure 12.17 Comparison of bacterial EFG, yeast EF2 and EF2-sordarin. (A) EF-G.GDP (pdb1fmn; [200]), (B) EF2, apo form (pdb1nov; [206]), (C) EF2 bound with Sordarin (pdb1nou; [206]). Domains I, II and G domain are coloured dark blue, domain III, purple, domain IV, light blue, domain V, yellow. Residues in EF-G that confer resistance to fusidic acid (A, F90L, P405L/Q, L431Q, A435N, P436Q, H458Y) or in EF2, to sordarin (B and C; Q490E; Y512S, S523E, E524P, A562P) are coloured in green. The GDP and sordarin molecule are coloured red in A and C, respectively.

12 Å; [276]. The crystal structure of EF2 in the apo-form and in complex with sordarin were solved to 2.85 and 2.12 Å resolution, respectively [206] (Figs. 12.17B, and C). Sordarin forms hydrogen-bonds to residues Gln490, Ala562 and Phe798, thus linking domains III, IV and V. Since no contact with the G domain is made, it would seem that the binding site for sordarin and fusidic acid may differ, although the mechanism of inhibition is probably similar, i.e., preventing the conformational changes necessary to form the low-affinity state that allows EF dissociation from the ribosome. Interestingly, sordarin locates within a very enclosed binding pocket, where extensive van der Waals interactions are made, such that almost two-thirds of the surface area of the drug is buried [206]. This binding site is not present in the apo form of EF2 suggesting that sordarin binds with an induced fit.

There is good agreement between the structurally determined sordarin-binding site and mutations in EF2 that give rise to sordarin resistance: substitutions in domains III (Q490E) and IV (Y512S, S523E, E524P, A562P) as well as a deletion in domain V (ΔG790) have been reported to confer resistance in *S. cerevisiae* [207, 208]. Many of these substitutions are naturally occurring in plant and mammalian EF2s explaining why sordarin is a fungal-specific translation inhibitor. Additionally, sordarin resistance results from mutations of the large-subunit ribosomal protein L10e (see Table 12.6; [209]).

Table 12.6 Antibiotic resistance resulting from mutations in ribosomal proteins

Antibiotic	Ribosomal component	Mutation position[a]	Selected reference[a]
Cycloheximide	L41	Q56P	251, 252
Evernimicin	L16	See Table 12.2	See Table 12.2
Erythromycin	L4	K63E	253, 254
Erythromycin	L22	Δ_{82}MKR$_{84}$	253, 254
Quinupristin-dalfopristin and erythromycin	L22	Δ_{79}GP$_{80}$ Insertions, e.g. $_{84}$GPTLK	255
Emetine	S14	R149C R149C-R150H	256
Gentamycin	L6		257
Streptomycin	S12	P41L/S, 42Q/R, K43E, K53, R85C/H, K87R/E, P90L, G91D	258-265
Streptomycin	S4		266
Spectinomycin	S5		267, 198
Neamine	S17	H32	268, 269
Sordarin	L10e	Q137P, K T143I, A ΔS134	209
Thiostrepton	L11	P18S, T P22S, T	174, 175

a Selected examples only of amino acid mutations and corresponding references (by no means complete) presented using equivalent *E. coli* numbering except for cycloheximide resistance in eukaryotes due to L41 and sordarin resistance in some yeast due to L10e.

12.4 Inhibitors of Termination, Recycling and *trans*-Translation

The final step of protein synthesis is signalled by the presence of a stop codon in the A-site. In *E. coli,* this termination signal is recognized by the decoding termination release factors, RF1 and RF2, which mediate the release of the nascent polypeptide from the ribosome (see Chap. 9). Subsequently, RF3 binds to recycle the decoding RFs from the ribosome in a GTP-dependent manner before also dissociating (reviewed in Ref. [210]). The disassembly of the post-termination complex is accomplished by the tandem effort of two factors, the ribosome-recycling factor RRF and EF-G (reviewed in Ref. [211]). Entry into the termination of protein synthesis can also be initiated through the co-called trans-translation pathway. This pathway is evoked most often when ribosomes become stalled because the mRNA that they are translating is truncated, i.e., the mRNA does not finish with a stop codon in the A-site; instead the A-site is empty. In this specific situation, a specific transfer-messenger RNA (tmRNA) enters the ribosomal A-site to continue translation and, in doing so, frees the ribosome and tags the truncated protein for degradation (see Chap. 10).

12.4.1 Termination

As yet there are no known specific inhibitors of the termination process. A number of antibiotics that effect termination are better known as inhibitors of the elongation cycle. Members of aminoglycoside family of antibiotics, including streptomycin, have been shown to specifically inhibit RF-dependent peptidyl-tRNA hydrolysis activity without preventing the binding of the termination factors to the ribosome [212]. Comparison of the solution structure of the termination factor RF2 [213], with that of the ribosome-bound form [214, 215] suggests that, upon ribosome binding, RF2 undergoes significant conformational changes. This rearrangement is probably initiated by recognition of the stop codon in the A-site and results in the movement of domain III of RF2 into the peptidyl-transferase center to mediate peptidyl-tRNA hydrolysis. Therefore, the aminoglycoside family of antibiotics could perturb the orientation of the RF on the ribosome, preventing the correct positioning of domain III necessary to activate peptide release. Sparsomycin, which binds in the PTF center and inhibits PTF activity (see Sect. 12.3.2.2) also prevents peptide release without affecting the binding of the RFs to the ribosome [212, 216]. Interestingly, a number of antibiotics have been identified that differentially effect peptide-bond formation and peptide release, being notably less effective against the latter, especially the RF-independent release of the peptide [216], the most significant difference being seen for the lincosamides, lincomycin and clindamycin [216], all of which bind directly in the PTF [217].

In contrast, thiostrepton has been shown to inhibit the binding of both RF1 and RF2 to the ribosome [217]. This can be understood in light of the recent cryo-EM reconstructions of RF2 on the ribosome where domain I of RF2 makes extensive contact with L11 and the GTPase-associated region, where thiostrepton also binds [214, 215].

The situation with spectinomycin however is not so clear. At low concentrations (1 μM) the *in vitro* peptide release activity of RF1 and RF2 at UGA and UAA codons, respectively, is enhanced, whereas at high concentrations (100 μM) RF2 binding is inhibited [212]. Furthermore, ribosomes carrying spectinomycin-resistant mutations C1192A or C1192U inhibited ribosome binding of both RF1 and RF2 *in vitro* but RF2 was affected more severely [212]. This suggests that binding of the RFs to the ribosome may require conformational changes within the ribosome, which are prevented by the binding of a rigid spectinomycin molecule to the head of the 30s subunit.

The challenge for the future will be to develop antibiotics that are specific for the termination phase, perhaps by taking advantage of the fact that the termination factors between prokaryotes and eukaryotes are structurally unrelated [213].

12.4.2 Recycling

It seems that the majority of antibiotics that inhibit translocation; thiostrepton, viomycin, fusidic acid and aminogylcosides; also effectively inhibit RRF-dependent

ribosome recycling [187]. This is not totally unexpected since this RRF-mediated reaction is strictly dependent on EF-G. Kaji and co-workers are working on the hypothesis that during RRF-mediated ribosome recycling the role of EF-G is analogous to that during elongation, i.e., EF-G actually translocates the RRF molecule, whose structure incidentally closely resembles that of a tRNA, from the A to the P-site (see Ref. [218]). However, this model has not yet been proven. In fact, recent evidence from chemical probing experiments suggest that RRF can bind to the ribosome in a completely different orientation to that of a tRNA (reviewed in Ref. [219]).

Since no counterpart to RRF has been identified in eukaryotes, this makes ribosome recycling an attractive target for antibiotic inhibition. The goal would be to find or develop antibiotics that can specifically inhibit RRF function without affecting other steps of proteins synthesis, instead of the current situation where all antibiotics targeting this phase operate through inhibition of EF-G function. To date, no such antibiotics have been found, but with an optimistic prospect for an RRF–ribosome complex crystal structure in the near future, designing of such antibiotics may soon be within reach.

12.4.3
Trans-translation

The importance of tmRNA was questioned when deletion strains (*ΔssrA*) were generated that had growth rates similar to the wild type strain. However, under certain stress conditions, even the *ΔssrA* strain was non-viable. One of these stress conditions is the presence of protein synthesis inhibitors. Unfortunately, the *ΔssrA* strain was generated by insertion of a kanamycin resistance cassette into the tmRNA gene, making the study of aminoglycoside effects on this strain uninformative. Aminoglycosides have been shown to increase the amount of ssrA-mediated tagging [220], suggesting the effect on the *ΔssrA* strain would be severe, especially since it was not viable in the presence of a number of antibiotics that bind to the PTF center or tunnel, such as chloramphenicol, lincomycin, spiramycin, tylosin, erythromycin (as well as spectinomycin which binds to the 16S rRNA) at drug concentrations that had no significant effect on the wild type strain [221]. Such a result would not be at all surprising in the presence of aminoglycosides: since they induce misreading which can lead to ribosomal stalling, the presence of tmRNA under these conditions would be necessary to rescue the stalled ribosomes. Although more unexpected for antibiotics that target the 50S subunit, it has been recently demonstrated for chloramphenicol and the oxazolidinone linezolid that they induce frameshifting and nonsense suppression [166], probably due to perturbation in the binding of the A- and/or P-site tRNAs.

12.5
Mechanisms Causing Drug Resistance

All antibiotics discussed here (at least the *source* compounds) are typically produced in *microorganisms* as secondary metabolites. In some cases, we have already discussed

how the 'producer' survives, and many of the known resistance mechanisms are strategies that are applied by the antibiotic-producing organisms themselves and can also be found in drug-resistant microorganisms that have become a major problem in the treatment of infections. We distinguish between six different mechanisms of resistance, four of which are non-ribosome-related and are discussed briefly, whereas the remaining two are directly related to the ribosome and are therefore discussed in more detail.

12.5.1 Modification of the Antibiotic

Modification of the antibiotic can inactivate the drug or alternatively prevent its cellular uptake. Common mechanisms of inactivation include adenylation, acetylation or phosphorylation, as seen for the (i) aminogylcosides, which are modified primarily on the rings I (3′- and 4′-OH are phosphorylated and adenylated) and II (amino group at position 3 is acetylated) [222], (ii) chloramphenicol, which is acetylated by chloramphenicol-acetyltransferase, and (iii) viomycin, which is phosphorylated by a viomycin-phosphotransferase. Capreomycin (a derivative of viomycin) is both acetylated and phosphorylated in the *Streptomyces* species that produce it [223]. Alternatively, cleavage or even degradation of the drug occurs, for example, penicillin cleavage by β-lactamase. Usually, these enzymes reside in the periplasmatic space and inhibit uptake of the drugs through the cellular membrane via direct modification of the antibiotic.

12.5.2 Blockage of Transport (without Modification of the Drug)

Modification of certain membrane components can prevent drug accumulation in the cell either by passively preventing their uptake (fusidic acid, tetracycline) or actively by promoting their active efflux from the cell (tetracycline).

12.5.3 Overproduction of the Inhibited Substrate (Target Dilution)

Overproduction of the target can, in principle, sequester the antibiotic, enabling the remaining active molecules to maintain growth. This principle is observed with trimethoprim, a drug that inhibits folic acid metabolism. In this case, a strong overproduction of dihydrofolate reductase was identified as the reason for resistance. As previously mentioned, exogenous expression of an rRNA fragment containing the binding site of spectinomycin also confers resistance – most probably by a "titration" of the antibiotic.

12.5.4
Bypassing or Replacement of the Inhibited Reaction

In principle, an inhibited reaction can by bypassed or substituted by another reaction. An example is seen in a resistance mechanism against penicillin. Penicillin works by cleaving the peptidoglycans within the cell wall. Resistance can arise when peptidoglycan is substituted for another component within the cell wall; the biosynthesis of peptidoglycan is inhibited at a late step by the drug. It is clear that neither one of the elongation steps nor the elongation cycle itself can be substituted or bypassed, since the ribosome is the sole protein synthesizing enzyme in all cells.

12.5.5
Alteration of the Target Site

Many examples exist where the target site of the drugs has been altered, of which a number have been mentioned already. In *E. coli* substitutions of C912U in 16S rRNA confer resistance to streptomycin, whereas in the 23S rRNA, substitution of A1067U and any change at A2058 confer resistance to thiostrepton and erythromycin, respectively. Mutations in the A-site that confer resistance to the aminoglycosides antibiotics have been reviewed in Westhof and colleagues [224]. However, for most bacterial species it is prohibitively difficult to acquire such mutations, since the rRNA is usually encoded in multiple operons, and thus mutations must be introduced into all rRNA operons (rrn) to confer resistance. This is particularly true for *E. coli*, which has seven copies of the rrn operon. Single mutations in the rRNA operon will result in a mixture of resistant and sensitive ribosomes leaving the bacteria still susceptible to drugs that exert a dominant inhibition effect. This occurs because on bacterial polysomes, blockage of one sensitive translating ribosome on a distinct mRNA will indirectly block following ribosomes on the same mRNA (whether the following ones are resistant or susceptible). For this reason rrn operons that confer resistance to protein synthesis inhibitors are typically recessive or weakly co-dominant.

Mutations conferring resistance to these inhibitors usually arise more frequently in bacteria that harbour one (as in *H. halobium*) or two copies of the rrn operon (e.g. *Helicobacter pylori*). Species such as these have been used successfully to screen for spontaneous resistant strains when exposed to an antibiotic of interest, for example pactamycin- and evernmicin-resistant *H. halobium* strains [35, 225]. Alternatively, Squires and coworkers have constructed an *E. coli* strain with all seven of the rrn operons deleted from the genome and having a single copy expressed from an exogenous plasmid [226, 227], which has, for example, been used to select for kasugamycin-resistant mutants [14].

However, resistance can also be mediated via ribosomal proteins where all ribosomes of a cell become resistant because in bacteria ribosomal proteins are normally encoded by unique genes. A number of such mutations that give rise to antimicrobial agents exist and are summarized in Table 12.6.

Producers of antibiotics often modify the rRNA post-transcriptionally to protect themselves. This mechanism is independent of the copy number of rRNA operons. Examples include the ribose methylation of A1067 (16S rRNA, thiostrepton-resistant) and base methylation of A2058 (erythromycin-resistant). Conversely, the undermethylation of the two 2,6,dimethyl-adenines 1518 and 1519 (16S rRNA) due to a mutation in the corresponding methylase gene (*ksgA*) renders cells kasugamycin-resistant.

12.5.6
Active Protection of the Target by a Third Component

The sixth resistance mechanism involves the involvement of protein factors termed ribosome protection proteins (RPP), which by interacting with the ribosome confer resistance to tetracycline (Tet) (reviewed in more detail in Ref. [57, 228–230]). The best studied of these are Tet(O) and Tet(M) from *Campylobacter* and *Streptococcus*. These RPPs probably originated from the natural producer of oxytetracycline, *Streptomyces rimosus*, which harbors *otr*A, a RPP-like determinant, derived about 30 million years ago from the elongation factor EF-G [57]. Spread of these factors throughout the eubacteria by lateral gene transfer events was most likely facilitated by their location on mobile genetic elements.

Not surprisingly, the RPPs display significant sequence similarity to the ribosomal elongation factors, EF-G and EF-Tu, and have been shown to have GTPase activity although they cannot substitute for the elongation factors *in vivo* or *in vitro* [231]. Instead, these RPPs confer resistance to Tet by binding to the Tet-inhibited ribosome, for example, addition of purified Tet(O) to a Tet-inhibited poly(Phe) *in vitro* system could restore activity with Tet(O) shifting the IC_{50} in this system from 100 to more than 500 µM Tet [232]. This was shown to be due to the fact that Tet(O)/(M) can dislodge tetracycline from the ribosome, a function that is dependent on the presence of GTP. The interaction of Tet(O) with the ribosome has been studied using cryo-EM, revealing that Tet(O) has an overall shape similar to that of EF-G, and as expected binds at a common site [233]. Comparison of the EF-G and Tet(O) ribosomal contacts indicates that they differ primarily in the vicinity of domain IV, where EF-G contacts H69 and Tet(O) interacts with h18/34 of the 16S rRNA. This is significant since domain IV in EF-G has been implicated as an important determinant for promoting translocation of the tRNAs (see Chap. 8.4) and is therefore consistent with the lack of translocation activity for Tet(O). Similarly the interaction of domain IV of Tet(O) with h34 of the 30S subunit is consistent with its role in Tet release as h34 is a component of the primary tetracycline-binding site (see Sect. 12.3.1.1). Recent chemical probing experiments are in agreement with this proposal, since Tet(O) interactions with the 30S subunit were localized to h34 (C1214 is protected) and h44 (A1408 is enhanced) [54]. These are components of the decoding center located near to the primary tetracycline-binding site. The protection C1214 probably results from interaction of this base with Tet(O), a conclusion supported by the fact that the Tet(O)-binding site observed by cryo-EM approaches

C1214. However, since Tet(O) does not approach A1408, the enhancement of A1408 is indicative of a conformational change. Connell and coworkers [54] suggest that Tet(O) binding to the ribosome induces long-range conformational changes, possibly through S12 and h44 to allosterically release Tet from the primary binding site. This conformational change persists after Tet(O), via GTP hydrolysis, dissociates from the ribosome, so that the conformation unfavorable for Tet binding continues to provide an advantage for binding of the ternary complex to the A-site, thus relieving the Tet-induced inhibition [54].

12.6 Future Perspectives

High-resolution structures for each of ribosomal subunits have revolutionized our ability to design biochemical and genetic experiments aimed at understanding ribosome structure and function and also to interpret the results better. Furthermore, it has enabled us to reinterpret the huge wealth of data relating to the ribosome, where the interaction of antibiotics with the ribosome is no exception. The current state of the art encompasses understanding these interactions by direct determination of these antibiotic–ribosome complexes at high resolution. Already representatives for the majority of antibiotic families that inhibit the ribosome directly have been characterized and those that have not, are for sure being completed as this review goes to press. So what does the future hold? With the ever-increasing emergence of antibiotic-resistant bacteria, the search for novel and more potent antibiotics continues to be the challenge for the future. Using the available structural information for rational design of these new and improved antibiotics is the path ahead. Such approaches may include chemically linking known antibiotics to create so-called "hybrid antibiotics" – the idea being that simultaneous mutations that confer resistance to each antibiotic would be required to confer resistance to the hybrid antibiotic. For example this has been reported for CP-544372, a hybrid between a macrolide, linked through a long anchor group at the 4"-position of a cladinose sugar, to chloramphenicol [234]. By analyzing the currently available information pertaining to the interactions of particular antibiotics with the ribosome, another approach is to alter the antibiotic, for example, by chemically modifying the sidechains, in such a way that either they establish new or additional interactions, such as hydrogen-bonds, hopefully increasing their affinity for the ribosome and/or making them more effective, even against, currently resistant strains. This is the exact goal of start-up companies, such as Rib-X and RiboTargets, where specialized software (Analog and Ribodock®, respectively) is being used iteratively to develop new antibiotic agents. Recently, RiboTargets presented evidence as to the power of this approach by the design of a new aminoglycoside antibiotic, which they demonstrated by crystallography could bind to a small RNA mimic of the ribosomal decoding site [235].

References

1 E. F. Gale, E. Cundliffe, P. E. Reynolds et al.: in *The Molecular Basis of Antibiotic Action* Wiley, Bristol, UK 1972, pp. 278–379.
2 P. Moore, T. Steitz, *Annu. Rev. Biochem.* **2003,** *72,* 813–850.
3 P. B. Moore, T. A. Steitz, *Nature* **2002,** *418,* 229–235.
4 V. Ramakrishnan, *Cell* **2002,** *108,* 557–572.
5 D. N. Wilson, G. Blaha, S. R. Connell et al., *Curr. Protein Pept. Sci.* **2002,** *3,* 1–53.
6 A. Yonath, *Annu. Rev. Biophys. Biomol. Struct.* **2002,** *31,* 257–273.
7 T. Auerbach, A. Bashan, J. Harms et al., *Curr. Drug Targets Infect. Disord.* **2002,** *2,* 169–186.
8 J. M. Harms, H. Bartels, F. Schlunzen et al., *J. Cell Sci.* **2003,** *116,* 1391–1393.
9 C. O. Gualerzi, C. L. Pon, *Biochemistry* **1990,** *29,* 5881–5889.
10 M. Kozak, D. Nathans, *J. Mol. Biol.* **1972,** *70,* 41–55.
11 T. L. Helser, J. E. Davies, J. E. Dahlberg, *Nat. New Biol.* **1971,** *233,* 12–14.
12 T. L. Helser, J. E. Davies, J. E. Dahlberg, *Nat. New Biol.* **1972,** *235,* 6–9.
13 P. F. Sparling, *Science* **1970,** *167,* 56–58.
14 A. Vila-Sanjurjo, C. L. Squires, A. E. Dahlberg, *J. Mol. Biol.* **1999,** *293,* 1–8.
15 J. Woodcock, D. Moazed, M. Cannon et al., *EMBO J.* **1991,** *10,* 3099–3103.
16 K. Chin, C. S. Shean, M. E. Gottesman, *J. Bacteriol.* **1993,** *175,* 7471–7473.
17 I. Moll, U. Bläsi, *Biochem. Biophys. Res. Commun.* **2002,** *297,* 1021–1026.
18 B. Poldermans, N. Goosen, P. H. Van Knippenberg, *J. Biol. Chem.* **1979,** *254,* 9085–9089.
19 J. Tan, U. Jakob, J. C. Bardwell, *J. Bacteriol.* **2002,** *184,* 2692–2698.
20 K. Inoue, J. Alsina, J. Chen et al., *Mol. Microbiol.* **2003,** *48,* 1005–1016.
21 bläsi, p. c..,
22 G. Dinos, D. Wilson, Y. Tereoka et al., *Mol. Cell* **2004.**
23 D. Moazed, H. F. Noller, *Nature* **1987,** *327,* 389–394.
24 M. Pioletti, F. Schlunzen, J. Harms et al., *EMBO J.* **2001,** *20,* 1829–1839.
25 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001,** *292,* 883–896.
26 F. Robert, L. Brakier-Gingras, *J. Biol. Chem.* **2003,** *278,* 44913–44920.
27 B. K. Bhuyan, *Biochem. Pharmacol.* **1967,** *16,* 1411–1420.
28 E. F. Gale, E. Cundliffe, P. E. Reynolds et al.: in *The Molecular Basis of Antibiotic Action* Wiley, Bristol, UK 1981, pp. 278–379.
29 D. E. Brodersen, W. M. Clemons, A. P. Carter et al., *Cell* **2000,** *103,* 1143–1154.
30 F. M. Aarestrup, L. B. Jensen, *Antimicrob. Agents Chemother.* **2000,** *44,* 3425–3427.
31 P. V. Adrian, W. Zhao, T. A. Black et al., *Antimicrob. Agents Chemother.* **2000,** *44,* 732–738.
32 P. M. McNicholas, P. A. Mann, D. J. Najarian et al., *Antimicrob. Agents Chemother.* **2001,** *45,* 79–83.
33 M. Zarazaga, C. Tenorio, R. Del Campo et al., *Antimicrob. Agents Chemother.* **2002,** *46,* 3657–3659.
34 P. V. Adrian, C. Mendrick, D. Loebenberg et al., *Antimicrob. Agents Chemother.* **2000,** *44,* 3101–3106.
35 L. Belova, T. Tenson, L. Q. Xiong et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 3726–3731.
36 C. B. Kofoed, B. Vester, *Antimicrob. Agents Chemother.* **2002,** *46,* 3339–3342.
37 P. A. Mann, L. Xiong, A. S. Mankin et al., *Mol. Microbiol.* **2001,** *41,* 1349–1356.
38 N. Ban, P. Nissen, J. Hansen et al., *Science* **2000,** *289,* 905–920.
39 J. Harms, F. Schluenzen, R. Zarivach et al., *Cell* **2001,** *107,* 679–688.
40 P. M. McNicholas, D. J. Najarian, P. A. Mann et al., *Antimicrob. Agents Chemother.* **2000,** *44,* 1121–1126.

41 A. La Teana, C. O. Gualerzi, A. E. Dahlberg, *RNA* **2001**, *7*, 1173–1179.
42 W. S. Champney, C. L. Tober, *Curr. Microbiol.* **2000**, *41*, 126–135.
43 W. S. Champney, *Curr. Drug Targets Infect. Disord.* **2001**, *1*, 19–36.
44 R. Mehta, W. S. Champney, *Antimicrob. Agents Chemother.* **2002**, *46*, 1546–1549.
45 H. S. Chittum, W. S. Champney, *Curr. Microbiol.* **1995**, *30*, 273–279.
46 J. Usary, W. S. Champney, *Mol. Microbiol.* **2001**, *40*, 951–962.
47 T. P. Hausner, U. Geigenmüller, K. H. Nierhaus, *J. Biol. Chem.* **1988**, *263*, 13103–13111.
48 J. P. White, C. R. Cantor, *J. Mol. Biol.* **1971**, *58*, 397–400.
49 W. Rogalski: in *The Tetracyclines (Handbook of Experimental Pharmacology)*, eds J. J. Hlavka and J. H. Boothe, Springer, Berlin 1985, pp. 179–316.
50 B. Rasmussen, H. F. Noller, G. Daubresse et al., *Antimicrob. Agents Chemother.* **1991**, *35*, 2306–2311.
51 M. M. Gerrits, M. R. de Zoete, N. L. Arents et al., *Antimicrob. Agents Chemother.* **2002**, *46*, 2996–3000.
52 J. I. Ross, E. A. Eady, J. H. Cove et al., *Antimicrob. Agents Chemother.* **1998**, *42*, 1702–1705.
53 C. A. Trieber, D. E. Taylor, *J. Bacteriol.* **2002**, *184*, 2131–2140.
54 S. R. Connell, C. A. Trieber, U. Stelzl et al., *Mol. Microbiol.* **2002**, *45*, 1463–1472.
55 J. S. Lodmell, A. E. Dahlberg, *Science* **1997**, *277*, 1262–1267.
56 I. V. Velichutina, J. Y. Hong, A. D. Mesecar et al., *J. Mol. Biol.* **2001**, *305*, 715–727.
57 S. R. Connell, D. M. Tracz, K. H. Nierhaus et al., *Antimicrob. Agents Chemother.* **2003**, *47*, 3675–3681.
58 I. Chopra, *Curr. Opin. Pharmacol.* **2001**, *1*, 464–469.
59 J. Bergeron, M. Ammirati, D. Danley et al., *Antimicrob. Agents Chemother.* **1996**, *40*, 2226–2228.
60 B. A. Rasmussen, Y. Gluzman, F. P. Tally, *Antimicrob. Agents Chemother.* **1994**, *38*, 1658–1660.
61 M. Tuckman, P. J. Petersen, S. J. Projan, *Microb. Drug Resist.* **2000**, *6*, 277–282.
62 C. G. Kurland, F. Jørgensen, A. Richter et al.: in *The Ribosome Structure, Function, and Evolution*, eds A. Dahlberg, W. E. Hill, R. A. Garrett et al., ASM Press, Washington, DC 1990, pp. 513–526.
63 A. P. Carter, W. M. Clemons, Jr., D. E. Brodersen et al., *Science* **2001**, *291*, 498–501.
64 C. M. T. Spahn, C. D. Presscott, *J. Mol. Med.* **1996**, *74*, 423–439.
65 C. G. Kurland, D. Hughes, M. Ehrenberg: in *Escherichia coli and Salmonella, Cellular and Molecular Biology*, ed. F. C. Neidhardt, ASM Press, Washington, DC 1996, pp. 979–1004,
66 J. M. Ogle, F. V. Murphy, M. J. Tarry et al., *Cell* **2002**, *111*, 721–732.
67 A. P. Carter, W. M. Clemons, D. E. Brodersen et al., *Nature* **2000**, *407*, 340–348.
68 J. M. Ogle, D. E. Brodersen, W. M. Clemons Jr et al., *Science* **2001**, *292*, 897–902.
69 D. Fourmy, M. I. Recht, S. C. Blanchard et al., *Science* **1996**, *274*, 1367–1371.
70 D. Fourmy, S. Yoshizawa, J. D. Puglisi, *J. Mol. Biol.* **1998**, *277*, 333–345.
71 S. R. Lynch, R. L. Gonzalez Jr., J. D. Puglisi et al., *Structure* **2003**, *11*, 43–53.
72 Q. Vicens, E. Westhof, *Structure (Cambridge)* **2001**, *9*, 647–658.
73 Q. Vicens, E. Westhof, *Biopolymers* **2003**, *70*, 42–57.
74 S. Yoshizawa, D. Fourmy, J. D. Puglisi, *EMBO J.* **1998**, *17*, 6437–6448.
75 Q. Vicens, E. Westhof, *J. Mol. Biol.* **2003**, *326*, 1175–1188.
76 Q. Vicens, E. Westhof, *Chem. Biol.* **2002**, *9*, 747–755.
77 R. Karimi, M. Ehrenberg, *Eur. J. Biochem.* **1994**, *226*, 355–360.
78 J. Ogle, A. Carter, V. Ramakrishnan, *TIBS* **2003**, *28*, 259–266.

79 D. Fourmy, M. I. Recht, J. D. Puglisi, *J. Mol. Biol.* **1998,** *277,* 347–362.
80 S. R. Lynch, J. D. Puglisi, *J. Mol. Biol.* **2001,** *306,* 1037–1058.
81 M. I. Recht, S. Douthwaite, J. D. Puglisi, *EMBO J.* **1999,** *18,* 3133–3138.
82 R. Kao, J. Davies: in *The Ribosome Structure, Function, Antibiotics, and Cellular Interactions,* eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, pp. 451–460.
83 I. G. Wool, A. Gluck, Y. Endo, *Trends Biochem. Sci.* **1992,** *17,* 266–269.
84 T. Hogg, J. R. Mesters, R. Hilgenfeld, *Curr. Protein Pept. Sci.* **2002,** *3,* 121–131.
85 H. Stark, M. V. Rodnina, J. Rinkeappel et al., *Nature* **1997,** *389,* 403–406.
86 M. Valle, J. Sengupta, N. K. Swami et al., *EMBO J.* **2002,** *21,* 3557–3567.
87 G. R. Andersen, S. Thirup, L. L. Spremulli et al., *J. Mol. Biol.* **2000,** *297,* 421–436.
88 R. Hilgenfeld, J. Mesters, T. Hogg: in *The Ribosome Structure, Function, Antibiotics, and Cellular Interactions,* eds R. A. Garrett, S. R. Douthwaite, A. Liljas, et al., ASM Press, Washington, DC 2000, pp. 347–357.
89 L. Vogeley, G. J. Palm, J. R. Mesters et al., *J. Biol. Chem.* **2001,** *276,* 17149–17155.
90 E. Fischer, H. Wolf, K. Hantke et al., *Proc. Natl. Acad. Sci. USA* **1977,** *74,* 4341–4345.
91 T. Watanabe, K. Izaki, H. Takahashi, *J. Antibiot. (Tokyo)* **1982,** *35,* 1141–1147.
92 T. Watanabe, T. Sugiyama, M. Takahashi et al., *J. Antibiot. (Tokyo)* **1992,** *45,* 470–475.
93 A. M. Zuurmond, L. N. Olsthoorn-Tieleman, J. Martien de Graaf et al., *J. Mol. Biol.* **1999,** *294,* 627–637.
94 S. E. Heffron, F. Jurnak, *Biochemistry* **2000,** *39,* 37–45.
95 A. Pingoud, W. Block, C. Urbanke et al., *Eur. J. Biochem.* **1982,** *123,* 261–265.
96 L. A. H. Zeef, L. Bosch, P. H. Anborgh et al., *EMBO J.* **1994,** *13,* 5113–5120.
97 P. Landini, M. Bandera, B. P. Goldstein et al., *Biochem. J.* **1992,** *283,* 649–652.
98 V. G. Mohrle, L. N. Tieleman, B. Kraal, *Biochem. Biophys. Res. Commun.* **1997,** *230,* 320–326.
99 T. P. Hausner, J. Atmadja, K. H. Nierhaus, *Biochimie* **1987,** *69,* 911–923.
100 I. G. Wool, *TIBS* **1984,** *9,* 14–17.
101 S. Yoshinari, S. Koresawa, S. Yokota et al., *Biosci. Biotechnol. Biochem.* **1997,** *61,* 324–331.
102 S. Yoshinari, S. Yokota, H. Sawamoto et al., *Eur. J. Biochem.* **1996,** *242,* 585–591.
103 T. Uchiumi, S. Honma, Y. Endo et al., *J. Biol. Chem.* **2002,** *277,* 41401–41409.
104 P. Nissen, J. Hansen, N. Ban et al., *Science* **2000,** *289,* 920–930.
105 M. Welch, J. Chastang, M. Yarus, *Biochemistry* **1995,** *34,* 385–390.
106 J. L. Hansen, T. M. Schmeing, P. B. Moore et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 11670–11675.
107 T. M. Schmeing, A. C. Seila, J. L. Hansen et al., *Nat. Struct. Biol.* **2002,** *9,* 225–230.
108 A. Bashan, I. Agmon, R. Zarivach et al., *Mol. Cell* **2003,** *11,* 91–102.
109 B. Ulbrich, G. Mertens, K. H. Nierhaus, *Arch. Biochem. Biophys.* **1978,** *190,* 149–154.
110 R. Suhadolnik: *Nucleoside Antibiotics* Wiley, New York 1970.
111 D. Vasquez: *Inhibitors of Protein Synthesis* Springer, Berlin, Heidelberg, New York 1979.
112 J. L. Hansen, P. B. Moore, T. A. Steitz, *J. Mol. Biol.* **2003,** *330,* 1061–1075.
113 F. Lichtenthaler, G. Trummlitz, *FEBS Lett.* **1974,** *38,* 237–242.
114 C. Rodriguez-Fonseca, R. Amils, R. A. Garrett, *J. Mol. Biol.* **1995,** *247,* 224–235.
115 B. Porse, C. Rodriguez-Fonseca, I. Leviev et al., *Biochem. Cell Biol.* **1995,** *73,* 877–885.
116 D. Moazed, H. F. Noller, *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 3725–3728.

117 G. T. Tan, A. Deblasio, A. S. Mankin, *J. Mol. Biol.* **1996**, *261*, 222–230.
118 E. Lazaro, C. Rodriguezfonseca, B. Porse et al., *J. Mol. Biol.* **1996**, *261*, 231–238.
119 E. Lazaro, L. A. G. M. Vandenbroek, A. S. Felix, et al., *Biochemistry* **1991**, *30*, 9642–9648.
120 B. T. Porse, S. V. Kirillov, M. J. Awayez et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9003–9008.
121 K. Fredrick, H. F. Noller, *Science* **2003**, *300*, 1159–1162.
122 I. Agmon, T. Auerbach, D. Baram et al., *Eur. J. Biochem.* **2003**, *270*, 2543–2556.
123 M. L. Celma, R. E. Monro, D. Vazquez, *FEBS Lett.* **1971**, *13*, 247–251.
124 S. Pestka, *Biochem. Biophys. Res. Commun.* **1969**, *36*, 589–595.
125 S. Pestka, *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 709–714.
126 H. J. Rheinberger, K. H. Nierhaus, *Eur. J. Biochem.* **1990**, *193*, 643–650.
127 F. Schlünzen, R. Zarivach, J. Harms et al., *Nature* **2001**, *413*, 814–821.
128 D. Moazed, H. F. Noller, *Biochimie* **1987**, *69*, 879–884.
129 U. Saarma, J. Remme, *Nucleic Acids Res.* **1992**, *20*, 3147–3152.
130 B. S. Cooperman, C. J. Weitzmann, C. L. Fernandez: in *The Ribosome: Structure, Function and Evolution*, eds W. E. Hill, A. Dahlberg, R. A. Garrett et al., ASM Press, Washington, DC 1990, pp. 491–501.
131 D. Nierhaus, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 2224–2228.
132 S. Kallia-Raftopoulos, D. L. Kalpaxis, *Mol. Pharmacol.* **1999**, *56*, 1042–1046.
133 J. Poehlsgaard, S. Douthwaite, *Curr. Opin. Investig. Drugs* **2003**, *4*, 140–148.
134 S. Douthwaite, *Nucleic Acids Res.* **1992**, *20*, 4717–4720.
135 J. I. Ross, E. A. Eady, J. H. Cove et al., *Antimicrob. Agents Chemother.* **1997**, *41*, 1162–1165.
136 S. Jenni, N. Ban, *Curr. Opin. Struct. Biol.* **2003**, *13*, 212–219.
137 A. S. Mankin, *Mol. Biol.* **2001**, *35*, 509–520.
138 H. Takashima, *Curr. Top. Med. Chem.* **2003**, *3*, 991–999.
139 S. M. Poulsen, C. Kofoed, B. Vester, *J. Mol. Biol.* **2000**, *304*, 471–481.
140 L. H. Hansen, P. Mauvais, S. Douthwaite, *Mol. Microbiol.* **1999**, *31*, 623–631.
141 L. Xiong, S. Shah, P. Mauvais et al., *Mol. Microbiol.* **1999**, *31*, 633–639.
142 R. Berisio, F. Schluenzen, J. Harms et al., *Nat. Struct. Biol.* **2003**, *31*, 31.
143 F. Schlunzen, J. M. Harms, F. Franceschi et al., *Structure (Cambridge)*, **2003**, *11*, 329–338.
144 R. Berisio, J. Harms, F. Schluenzen et al., *J. Bactriol* **2003**, *185*, 4276–4279.
145 J. L. Hansen, J. A. Ippolito, N. Ban et al., *Mol. Cell* **2002**, *10*, 117–128.
146 T. Tenson, M. Lovmar, M. Ehrenberg, *J. Mol. Biol.* **2003**, *330*, 1005–1014.
147 S. Douthwaite, L. H. Hansen, P. Mauvais, *Mol. Microbiol.* **2000**, *36*, 183–193.
148 J. C. Mao, M. Putterman, *J. Mol. Biol.* **1969**, *44*, 347–361.
149 S. T. Gregory, A. E. Dahlberg, *J. Mol. Biol.* **1999**, *289*, 827–834.
150 T. Tenson, M. Ehrenberg, *Cell* **2002**, *108*, 591–594.
151 H. M. Lamb, D. P. Figgitt, D. Faulds, *Drugs* **1999**, *58*, 1061–1097.
152 B. Porse, R. Garrett, *J. Mol. Biol.* **1999**, *286*, 375–387.
153 E. Nyssen, M. Di Giambattista, C. Cocito, *Biochim. Biophys. Acta* **1989**, *1009*, 39–46.
154 R. Parfait, C. Cocito, *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 5492–5496.
155 N. B. Matassova, M. V. Rodnina, R. Endermann et al., *RNA* **1999**, *5*, 939–946.
156 J. R. Colca, W. G. McDonald, D. J. Waldon et al., *J. Biol. Chem.* **2003**, *10*, 10.
157 P. Sander, L. Belova, Y. G. Kidan et al., *Mol. Microbiol.* **2002**, *46*, 1295–1304.
158 C. C. Zhou, S. M. Swaney, D. L. Shinabarger et al., *Antimicrob. Agents Chemother.* **2002**, *46*, 625–629.
159 A. H. Lin, R. W. Murray, T. J. Vidmar et al., *Antimicrob. Agents Chemother.* **1997**, *41*, 2127–2131.

160 P. Kloss, L. Xiong, D. L. Shinabarger et al., *J. Mol. Biol.* **1999,** *294,* 93–101.
161 H. Aoki, L. Ke, S. M. Poppe et al., *Antimicrob. Agents Chemother.* **2002,** *46,* 1080–1085.
162 H. Burghardt, K. L. Schimz, M. Muller, *FEBS Lett.* **1998,** *425,* 40–44.
163 D. C. Eustice, P. A. Feldman, I. Zajac et al., *Antimicrob. Agents Chemother.* **1988,** *32,* 1218–1222.
164 D. L. Shinabarger, K. R. Marotti, R. W. Murray et al., *Antimicrob. Agents Chemother.* **1997,** *41,* 2132–2136.
165 S. M. Swaney, H. Aoki, M. C. Ganoza et al., *Antimicrob. Agents Chemother.* **1998,** *42,* 3251–3255.
166 J. Thompson, M. O'Connor, J. A. Mills et al., *J. Mol. Biol.* **2002,** *322,* 273–279.
167 P. Dandliker, S. Pratt, A. Nilius et al., *Antimicrob. Agents Chemother.* **2003,** *47,* 3831–3839.
168 E. Cundliffe: in *The Ribosome: Structure, Function and Evolution,* eds W. E. Hill, A. Dahlberg, R. A. Garrett et al., ASM Press, Washington, DC 1990, pp. 479–490.
169 D. Moazed, J. M. Robertson, H. F. Noller, *Nature* **1988,** *334,* 362–364.
170 D. Moazed, H. F. Noller, *Cell* **1989,** *57,* 585–597.
171 J. Egebjerg, S. Douthwaite, R. A. Garrett, *EMBO J.* **1989,** *8,* 607–611.
172 J. Thompson, W. Musters, E. Cundliffe et al., *EMBO J.* **1993,** *12,* 1499–1504.
173 T. Uchiumi, A. Wada, R. Kominami, *J. Biol. Chem.* **1995,** *270,* 29889–29893.
174 A. S. Mankin, I. Leviev, R. A. Garrett, *J. Mol. Biol.* **1994,** *244,* 151–157.
175 B. T. Porse, I. Leviev, A. S. Mankin et al., *J. Mol. Biol.* **1998,** *276,* 391–404.
176 G. L. Conn, D. E. Draper, E. E. Lattman et al., *Science* **1999,** *284,* 1171–1174.
177 B. T. Wimberly, R. Guymon, J. P. McCutcheon et al., *Cell* **1999,** *97,* 491–502.
178 M. V. Rodnina, A. Savelsbergh, N. B. Matassova et al., *Proc. Natl. Acad. Sci. USA* **1999,** *96,* 9586–9590.
179 H. Stark, M. V. Rodnina, H. J. Wieden et al., *Cell* **2000,** *100,* 301–309.
180 J. Frank, R. Agrawal, *Cold Spring Harb. Symp. Quant. Biol.* **2001,** *66,* 67–75.
181 D. M. Cameron, J. Thompson, P. E. March et al., *J. Mol. Biol.* **2002,** *319,* 27–35.
182 P. Wurmbach, K. H. Nierhaus, *Eur. J. Biochem.* **1983,** *130,* 9–12.
183 M. Misumi, T. Nishimura, T. Komai et al., *Biochem. Biophys. Res. Commun.* **1978,** *84,* 358–365.
184 T. Yamada, Y. Mizugichi, K. H. Nierhaus et al., *Nature* **1978,** *275,* 460–461.
185 O. Jerinic, S. Joseph, *J. Mol. Biol.* **2000,** *304,* 707–713.
186 C. Gustafsson, B. C. Persson, *J. Bacteriol.* **1998,** *180,* 359–365.
187 G. Hirokawa, M. C. Kiel, A. Muto et al., *EMBO J.* **2002,** *21,* 2272–2281.
188 M. Kiel, V. Raj, H. Kaji et al., *J. Biol. Chem.* **2003,** *278,* 48041–48050.
189 U. von Ahsen, J. Davies, R. Schroeder, *Nature* **1991,** *353,* 368–370.
190 R. Schroeder, C. Waldsich, H. Wank, *EMBO J.* **2000,** *19,* 1–9.
191 J. Davies, *Mol. Microbiol.* **1990,** *4,* 1227–1232.
192 J. Frank, R. K. Agrawal, *Nature* **2000,** *406,* 318–322.
193 M. F. Brink, G. Brink, M. P. Verbeet et al., *Nucleic Acids Res.* **1994,** *22,* 325–331.
194 H. Fromm, M. Efelman, D. Aviv et al., *EMBO J.* **1987,** *6,* 3233–3237.
195 U. Johanson, D. Hughes, *Nucleic Acids Res.* **1995,** *23,* 464–466.
196 Z. Svab, P. Maliga, *Mol. Gen. Genet.* **1991,** *228,* 316–319.
197 G. Thom, C. D. Prescott, *Bioorgan. Med. Chem.* **1997,** *5,* 1081–1086.
198 G. Funatsu, E. Schilitz, H. G. Wittmann, *Molec. Gen. Genet.* **1971,** *114,* 106–111.
199 S. Besier, A. Ludwig, V. Brade et al., *Mol. Microbiol.* **2003,** *47,* 463–469.

200 M. Laurberg, O. Kristensen, K. Martemyanov et al., *J. Mol. Biol.* **2000**, *303*, 593–603.

201 A. Liljas, O. Kristensen, M. Laurberg et al.: in *The Ribosome. Structure, Function, Antibiotics, and Cellular Interactions*, eds R. A. Garrett, S. R. Douthwaite, A. Liljas et al., ASM Press, Washington, DC 2000, pp. 359–365.

202 R. Agrawal, P. Penczek, R. Grassucci et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6134–6138.

203 R. K. Agrawal, A. B. Heagle, P. Penczek et al., *Nat. Struct. Biol.* **1999**, *6*, 643–647.

204 R. K. Agrawal, J. Linde, J. Sengupta et al., *J. Mol. Biol.* **2001**, *311*, 777–787.

205 M. G. Gomez-Lorenzo, C. M. T. Spahn, R. K. Agrawal et al., *EMBO J.* **2000**, *19*, 2710–2718.

206 R. Jorgensen, P. A. Ortiz, A. Carr-Schmid et al., *Nat. Struct. Biol.* **2003**, *10*, 379–385.

207 M. C. Justice, M. J. Hsu, B. Tse et al., *J. Biol. Chem.* **1998**, *273*, 3148–3151.

208 M. Shastry, J. Nielsen, T. Ku et al., *Microbiology* **2001**, *147*, 383–390.

209 M. C. Justice, T. Ku, M. J. Hsu et al., *J. Biol. Chem.* **1999**, *274*, 4869–4875.

210 L. Kisselev, M. Ehrenberg, L. Frolova, *EMBO J.* **2003**, *22*, 175–182.

211 A. Kaji, M. Kiel, G. Hirokawa et al., *Cold Spring Harb. Symp. Quant. Biol.* **2001**, *66*, 515–529.

212 C. M. Brown, K. K. Mccaughan, W. P. Tate, *Nucleic Acids Res.* **1993**, *21*, 2109–2115.

213 B. Vestergaard, L. Van, G. Andersen et al., *Mol. Cell* **2001**, *8*, 1375–1382.

214 B. P. Klaholz, T. Pape, A. V. Zavialov et al., *Nature* **2003**, *421*, 90–94.

215 U. B. Rawat, A. V. Zavialov, J. Sengupta et al., *Nature* **2003**, *421*, 87–90.

216 N. Polacek, M. J. Gomez, K. Ito et al., *Mol. Cell* **2003**, *11*, 103–112.

217 N. Brot, W. P. Tate, C. T. Caskey et al., *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 89–92.

218 M. Selmer, S. Al-Karadaghi, G. Hirakawa et al., *Science* **1999**, *286*, 2349–2352.

219 D. E. Brodersen, V. Ramakrishnan, *Nat. Struct. Biol.* **2003**, *10*, 78–80.

220 T. Abo, K. Ueda, T. Sunohara et al., *Genes Cells* **2002**, *7*, 629–638.

221 J. de la Cruz, A. Vioque, *RNA* **2001**, *7*, 1708–1716.

222 K. J. Shaw, P. N. Rather, R. S. Hare et al., *Microbiol. Rev.* **1993**, *57*, 138–163.

223 R. H. Skinner, E. Cundliffe, *J. Gen. Microbiol.* **1980**, *120*, 95–104.

224 P. Pfister, S. Hobbie, Q. Vicens et al., *Chembiochem.* **2003**, *4*, 1078–1088.

225 A. S. Mankin, *J. Mol. Biol.* **1997**, *274*, 8–15.

226 T. Asai, C. Condon, J. Voulgaris et al., *J. Bacteriol.* **1999**, *181*, 3803–3809.

227 T. Asai, D. Zaporojets, C. Squires et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1971–1976.

228 I. Chopra, M. Roberts, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 232–260.

229 M. C. Roberts, *FEMS Microbiol. Rev.* **1996**, *19*, 1–24.

230 M. C. Roberts, *Mol. Biotechnol.* **2002**, *20*, 261–283.

231 V. Burdett, *J. Bacteriol.* **1996**, *178*, 3246–3251.

232 C. A. Trieber, N. Burkhardt, K. H. Nierhaus et al., *Biol. Chem.* **1998**, *379*, 847–855.

233 C. M. Spahn, G. Blaha, R. K. Agrawal et al., *Mol. Cell.* **2001**, *7*, 1037–1045.

234 Y. J. Wu, W. G. Su, *Curr. Med. Chem.* **2001**, *8*, 1727–1758.

235 R. J. Russell, J. B. Murray, G. Lentzen et al., *J. Am. Chem. Soc.* **2003**, *125*, 3410–3411.

236 D. Fourmy, M. I. Recht, S. C. Blanchard et al., *Science* **1996**, *274*, 1367–1371.

237 R. A. Sayle, E. J. Milner-White, *Trends Biochem. Sci.* **1995**, *20*, 374.

238 N. Guex, M. C. Peitsch, *Electrophoresis* **1997**, *18*, 2714–2723.

239 W. Humphrey, A. Dalke, K. Schulten, *J. Molec. Graphics* **1996**, *14*, 33–38.

240 F. Abdulkarim, L. Liljas, D. Hughes, *FEBS Lett.* **1994**, *352*, 118–122.

241 J. R. Mesters, L. A. H. Zeef, R. Hilgenfeld et al., *EMBO J.* **1994**, *13*, 4877–4885.

242 L. A. H. Zeef, L. Bosch, *Mol. Gen. Genet.* **1993**, *238*, 252–260.

243 C. Alexander, N. Bilgin, C. Lindschau et al., *J. Biol. Chem.* **1995,** *270,* 14541–14547.

244 M. Sosio, G. Amati, C. Cappellano et al., *Mol. Microbiol.* **1996,** *22,* 43–51.

245 K. Boon, I. Krab, A. Parmeggiani et al., *Eur. J. Biochem.* **1995,** *227,* 816–822.

246 P. Londei, J. L. Sanz, S. Altamura et al., *J. Bacteriol.* **1986,** *167,* 265–2671.

247 L. Xiong, P. Kloss, S. Douthwaite et al., *J. Bacteriol.* **2000,** *182,* 5325–5331.

248 J. Prystowsky, F. Siddiqui, J. Chosay et al., *Antimicrob. Agents Chemother.* **2001,** *45,* 2154–2156.

249 S. Tsiodras, H. S. Gold, G. Sakoulas et al., *Lancet* **2001,** *358,* 207–208.

250 J. Thompson, D. F. Kim, M. O'Connor et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 9002–9007.

251 L. Del Pozo, D. Abarca, J. Hoenicka et al., *Eur. J. Biochem.* **1993,** *213,* 849–587.

252 S. Kawai, S. Murao, M. Mochizuki et al., *J. Bacteriol.* **1992,** *174,* 254–262.

253 H. S. Chittum, W. S. Champney, *J. Bacteriol.* **1994,** *176,* 6192–6198.

254 H. G. Wittmann, G. Stoffler, D. Apirion et al., *Mol. Gen. Genet.* **1973,** *127,* 175–189.

255 B. Malbruny, A. Canu, B. Bozdogan et al., *Antimicrob. Agents Chemother.* **2002,** *46,* 2200–2207.

256 D. D. Rhoads, D. J. Roufa, *Mol. Cell. Biol.* **1985,** *5,* 1655–1659.

257 C. Davies, D. E. Bussiere, B. L. Golden et al., *J. Mol. Biol.* **1998,** *279,* 873–888.

258 J. Bjorkman, P. Samuelsson, D. Andersson et al., *Mol. Microbiol.* **1999,** *31,* 53–58.

259 K. Bohman, T. Ruusala, P. C. Jelene et al., *Mol. Gen. Genet.* **1984,** *198,* 90–99.

260 G. Funatsu, K. Nierhaus, H. G. Wittmann, *Biochim. Biophys. Acta* **1972,** *287,* 282–291.

261 G. Funatsu, H. G. Wittmann, *J. Mol. Biol.* **1972,** *68,* 547–550.

262 S. Galili, H. Fromm, D. Aviv et al., *Mol. Gen. Genet.* **1989,** *218,* 289–292.

263 S. Gregory, J. Cate, A. Dahlberg, *J. Mol. Biol.* **2001,** *309,* 333–338.

264 A. R. Timms, H. Steingrimsdottir, A. R. Lehmann et al., *Mol. Gen. Genet.* **1992,** *232,* 89–96.

265 U. van Acken, *Mol. Gen. Genet.* **1975,** *140,* 61–68.

266 G. Funatsu, W. Puls, E. Schiltz et al., *Molec. Gen. Genet.* **1972,** *115,* 131–139.

267 A. Bollen, T. Helser, T. Yamada et al., *Cold Spring Harb. Symp. Quant. Biol.* **1969,** *34,* 95–100.

268 A. Bollen, T. Cabezon, M. de Wilde et al., *J. Mol. Biol.* **1975,** *99,* 795–806.

269 B. L. Golden, D. W. Hoffman, V. Ramakrishnan et al., *Biochemistry* **1993,** *32,* 12812–12820.

270 J. J. Cannone, S. Subramanian, M. N. Schnare et al., *BioMed Central Bioinformatics* **2002,** *3.*

271 B. T. Wimberly, D. E. Brodersen, W. M. Clemons et al., *Nature* **2000,** *407,* 327–339.

272 M.A. Schaefer, A.O. Tastan, S. Patzke et al., *J. Biol. Chem.* **2002,** *277,* 19095–19105.

273 J.E. Wilson, T.V. Pestova, C.U. Hellen et al., *Cell* **2000,** *102,* 511–520.

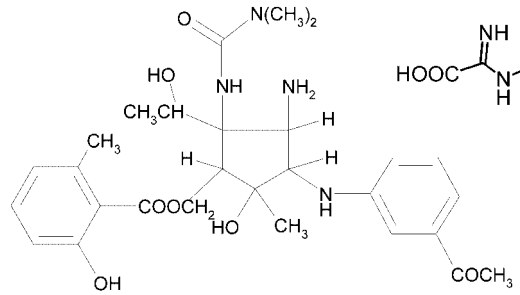
274 J.M. Harms, F. Schluenzen, P. Fucini et al. *BMC Biol.* **2004,** Apr 1.

275 S.M. Swaney, D.L. Shinabarger, R.D. Schaadt et al., *Abtracts of the 38th Interscience conference on Antimicrobial Agents and Chemotherapy* **1998** Abstract C104.

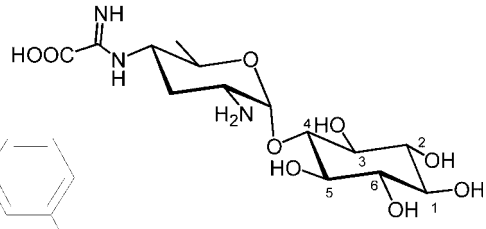
276 C.M. Spahn, M.G. Gomez-Lorenzo, R.A. Grassucci et al., *EMBO J.,* **2004,** *23,* 1008–1019.

Appendix A1

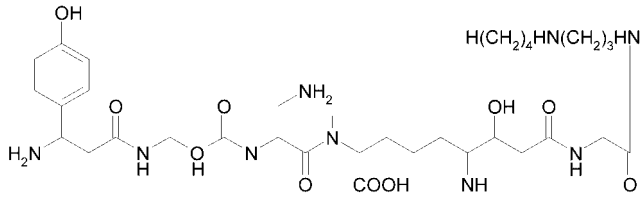
Pactamycin



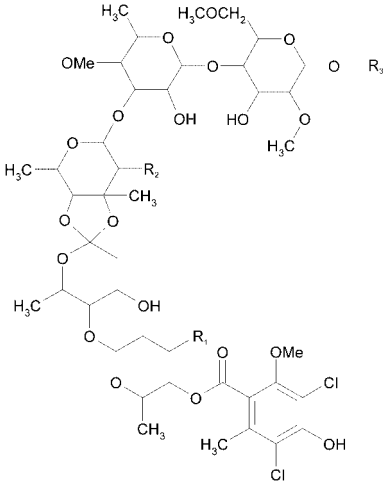
Kasugamycin



Edeine



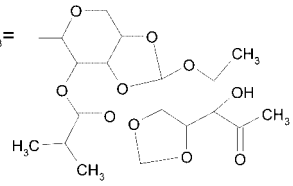
Evernimicin and Avilamycin



Avilamycin

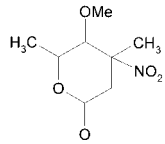
R_1= OH, R_2= H

R_3=



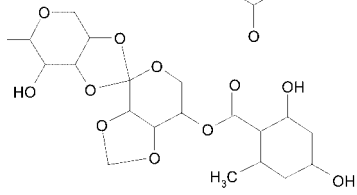
Evernimicin

R_1=



R_2= OH

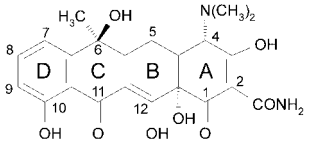
R_3=



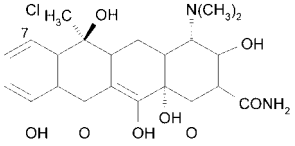
Appendix B1

Tetracyclines

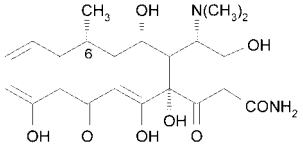
Tetracycline (1953)



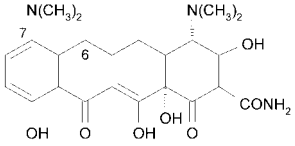
7-Chlortetracycline (1948)



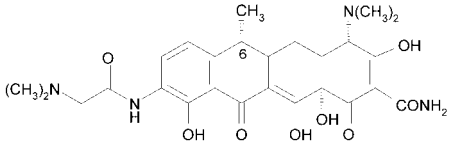
Doxycycline
(6-deoxy-tetracycline) (1967)



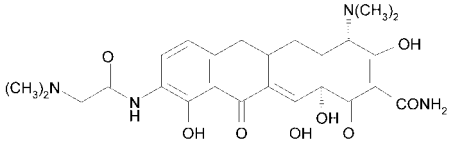
Minocycline
(7-dimethylamino-6-demethyl-6-deoxy-tetracycline) (1972)



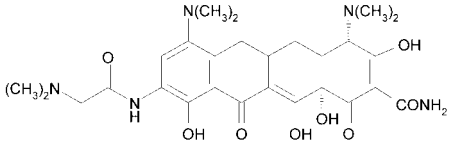
Glycylcyclines



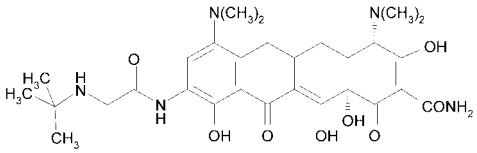
DMG-DOT
(DMG-doxycycline)
(9-(*N,N*-dimethylglycylamido)-6-deoxytetracycline)
(1993)



DMG-DMDOT
(9-(*N,N*-dimethylglycylamido)-6-demethyl-6-deoxytetracycline)
(1993)



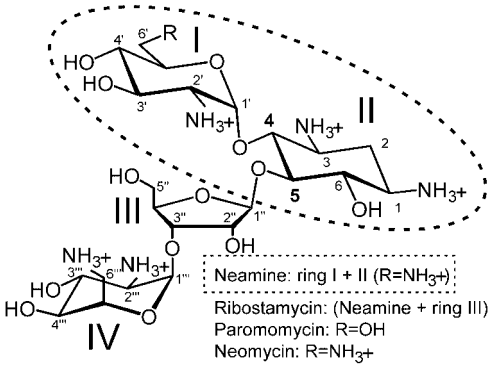
DMG-MINO
(9-(*N,N*-dimethylglycylamido)-minocycline)
(1993)



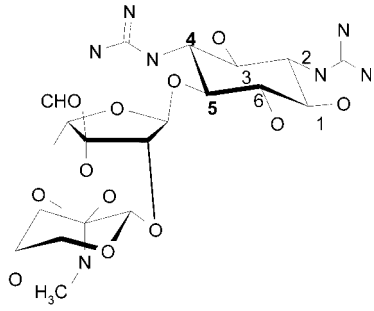
GAR-936
(tigecycline) or
(TBG-MINO) or
(9-(t-butylglycylamido)-minocycline)
(1993)

Appendix C1

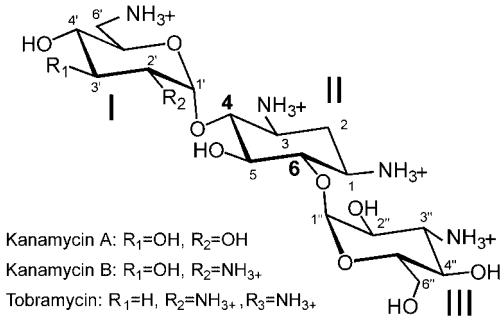
Aminoglycosides: 4,5-2-DOS

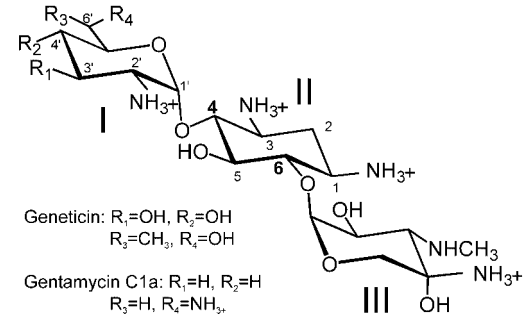


Streptomycin



Aminoglycosides: 4,6-2-DOS

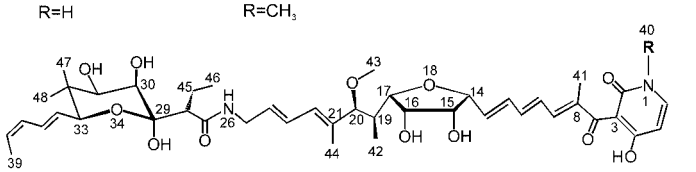




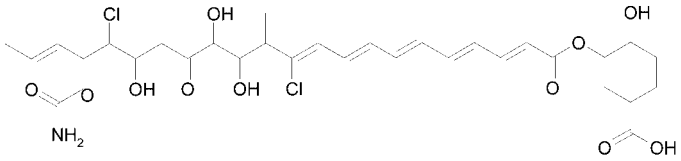
Appendix D1

Kirromycin
R=H

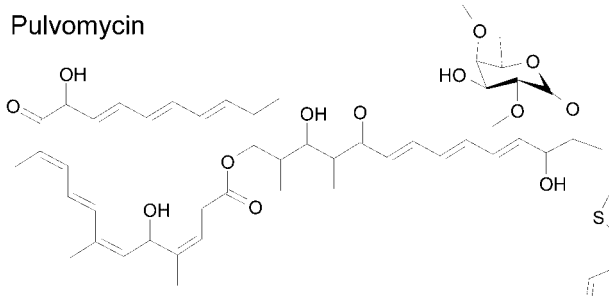
Aurodox (N-methyl-Kirromycin)
R=CH_3



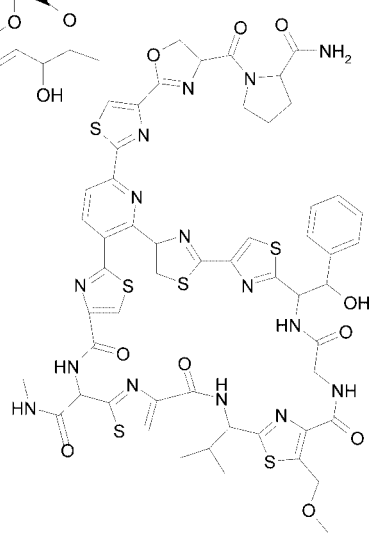
Enacyloxin IIa



Pulvomycin



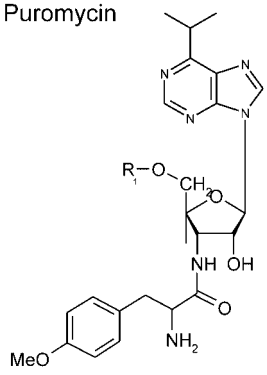
GE2770A
(MDL 62,879)



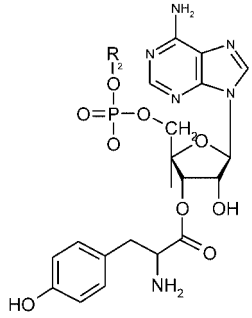
Appendix E1

Peptidyl-transferase inhibitors

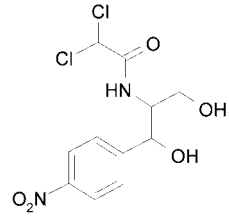
Puromycin



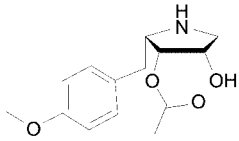
tRNA-A76



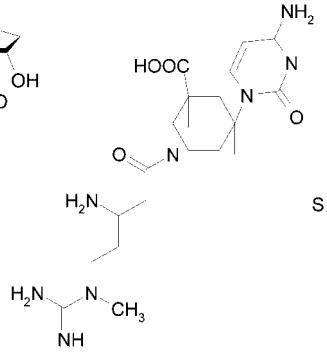
Chloramphenicol



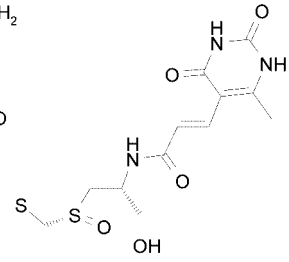
Anisomycin



Blasticidin S

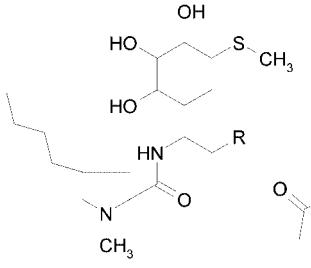


Sparsomycin



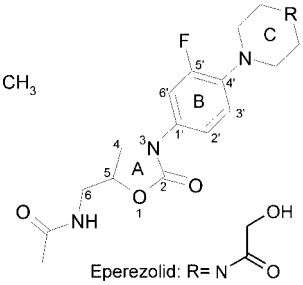
Lincosamides

Lincomycin: R= OH
Clindamycin: R= Cl



Oxazolidinones

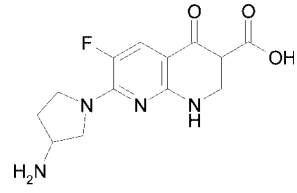
Linezolid: R= O



Eperezolid: R= N

NRI

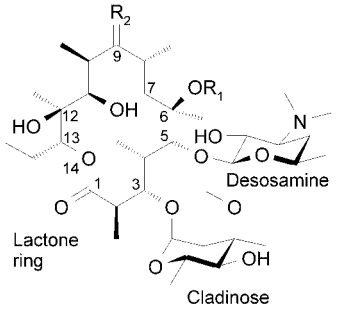
A-72310



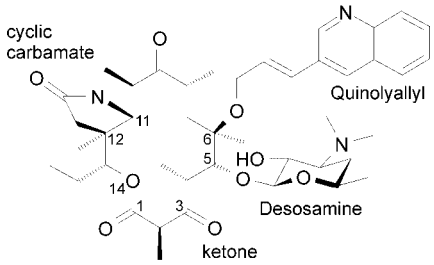
Appendix E2

Macrolides: 14 and 15 membered

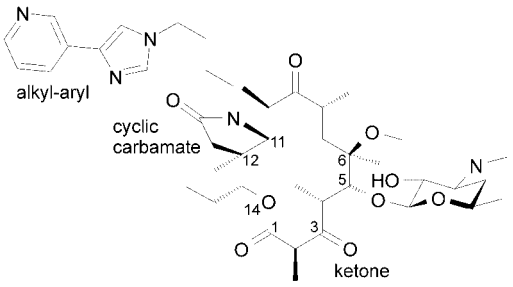
Erythromycin: R_1= H, R_2= O
Clarithromycin: R_1= CH_3, R_2= O
Roxithromycin: R_1= H, R_2 = $NOCH_2O(CH_2)_2OCH_3$



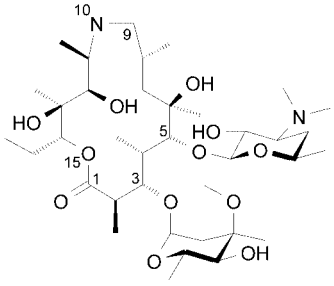
Cethromycin (ABT-773)



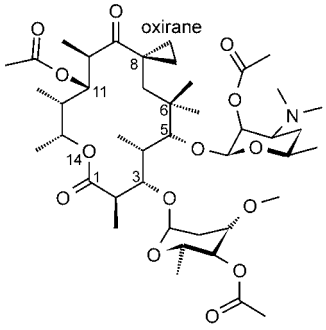
Telithromycin (HMR-3647)



Azithromycin



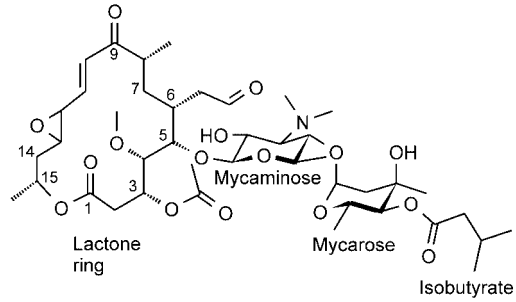
Troleandomycin



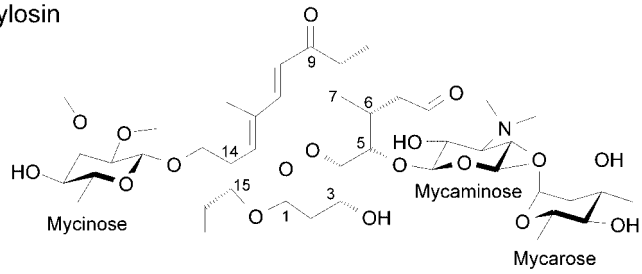
Appendix E3

Macrolides: 16 membered

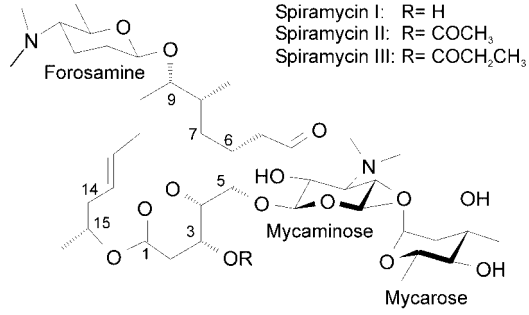
Carbomycin A



Tylosin

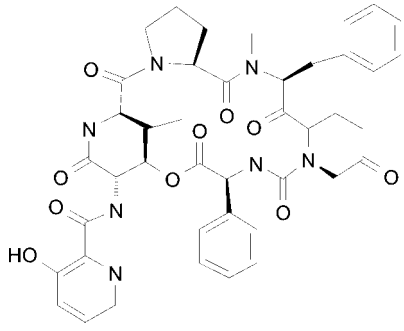


Spiramycin

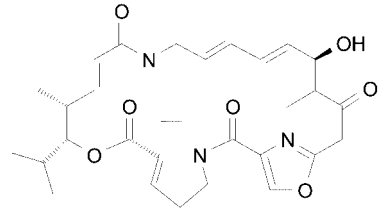


Streptogramins

Virginiamycin S (S_A)



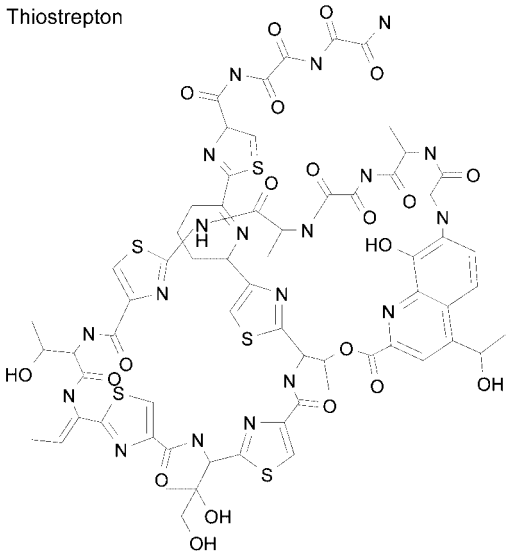
Virginiamycin M (S_B)



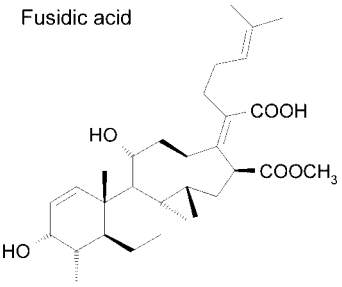
Appendix F1

Translocation inhibitors

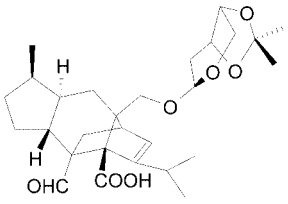
Thiostrepton



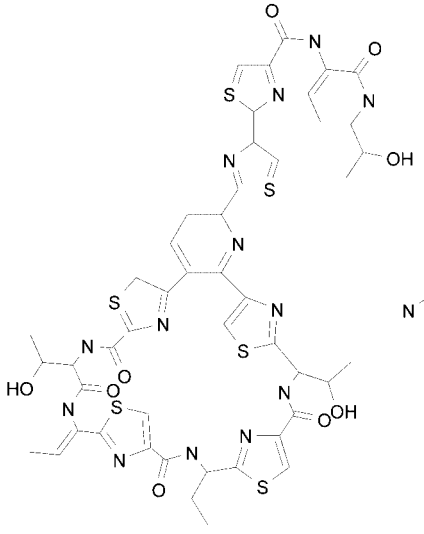
Fusidic acid



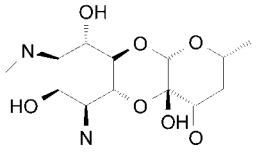
Sordarin (GM193663)



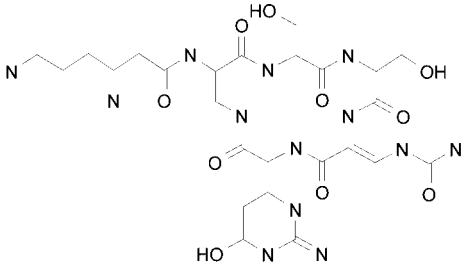
Micrococcin



Spectinomycin



Viomycin



13
The Work of Chaperones

Jean–Hervé Alix

13.1
From The Levinthal Paradox To The Anfinsen Cage

The classic experiment of Anfinsen showing that the unfolded ribonuclease folds spontaneously *in vitro* established the thermodynamic hypothesis of protein stability, i.e., that a protein's primary sequence dictates three essential and partially overlapping features of proteins, viz. the assembly pathway, the structure and function, without need of any further genetic information [1]. However, how the correct folding of a protein is selected among an astronomically large number (10^{16}) of possible conformations to give the native active state was enigmatic for a long time, a problem known as the Levinthal paradox. This was particularly true for large proteins, but it is now clear that folding pathways *guide* the protein, along energy landscapes, towards the unique (lowest energy) native conformation, through a series of partially folded intermediate states known as molten globules [2]. In other words, it is possible to arrive at the native state of a protein after having searched through *only* a minute fraction of the total number of conformations [3, 4]. The transition from the molten globule state (which contains elements of secondary structure such as α-helices and β-sheets, but lack well-defined, unique tertiary interactions) to the native state is often the rate-limiting step (Fig. 13.1).

In vitro, self-folding gives low yields and slow rates, particularly at temperatures above 15–20°C and at high protein concentrations, i.e., under situations of incredibly high macromolecular crowding as is the case in living cells [4b]. Also, unfavorable side-reactions, such as misfolding or aggregation of partially folded intermediates (Fig. 13.2), often compete with the correct assembly pathway [5]. Therefore, self-assembly is not the predominant form of protein assembly *in vivo*, and proteins will be assisted by a particular class of proteins, sequestering them during folding in a safe environment protected from aggregation, sometimes referred to as the Anfinsen cage [6] or box of infinite dilution. These folding or assembly helpers have been termed molecular chaperones. More than just playing the role of a passive cage, chaperones also decrease the roughness of the energy lanscape of the substrate protein. After assisting the correct folding and assembly of other proteins, they are not themselves components of the final functional structures, nor do they cause covalent modifications of the target protein or protein complex, for example, the *Escherichia coli* GroEL/GroES chaperonin, which participates in bacteriophage

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

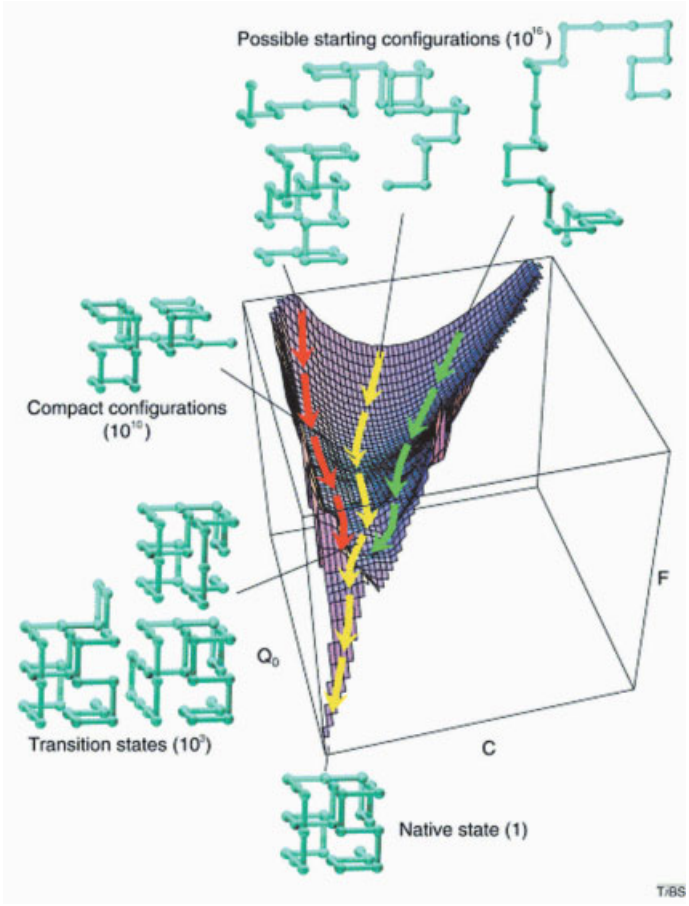


Figure 13.1 Free-energy landscape of the fold-ing of a protein. The free energy (F) surface or free energy landscape of a protein is represen-ted as a function of the number of native contacts (Q) and the total number of (native and nonnative) contacts (C). Native and non-native contacts refer to contacts bringing or not towards the native state, respectively. The surface shown in the figure illustrates that at the beginning of the folding reaction there are many conformations of similar free energy, so that the accessible surface is very broad. As folding progresses, the energy of the system decreases with the formation of native contacts that are generally more stabilizing than the non-native ones. Thus, the entropy decreases as the native state is approached, and the free-energy surface has a funnel-like shape that guides the system towards the unique (lowest energy) native conformation. Among the intermediate folding species are the molten globules, which are generally close to the native state. The yellow trajectory shows the average folding pathway, and the other two trajectories (green and red) show a range of two standard deviations around the average and are thus expected to include 95% of the trajectories. Reprinted with per-mission from Dinner et al. [4].

Figure 13.2 Formation of a domain-swapped aggregate in the process of protein folding. A newly synthesized protein molecule in the process of folding seeks a thermodynamically stable structure. Among possible low-energy structures that it may seek are a monomer, a domain-swapped dimer, and a domain-swapped aggregate. Reprinted with permission from Eisenberg [298].

λ head assembly, is not found in the assembled head structures. Molecular chaperones also act as refolders of misfolded or aggregated substrates, probably through substrate unfolding (either local or global) [7, 8].

Chaperones are ubiquitous, universal, highly conserved throughout evolution in both structural and functional properties: HSP70 is the most conserved protein known to date that is found in all biota, i.e., eubacteria, eukaria and some of the archaea [9]. This high conservation across the phylogenetic domains [10, 11] has provided support for a phylogenetic classification of all living cells [12–14] that intriguingly differs from that based on comparative analyses of 16S rRNA sequences: the HSP70-based phylogenies predict a specific evolutionary relationship between the archea and Gram-positive bacteria on the one hand, and between the Gram-negative bacteria and eukaryotes on the other [15].

The myriad of functions of the molecular chaperones [16–18] can be summarized as follows:

1. *De novo* protein folding, i.e., co- or post-translational folding of ribosome-bound nascent polypeptide chains.
2. Refolding and prevention and reversion of aggregation of misfolded or denatured proteins.
3. Protein translocation across membranes.
4. Post-translational quality control, to detect and eliminate, in co-operation with proteases, the proteins irreversibly unfolded.
5. Assembly and dissasembly of protein and nucleoprotein complexes.
6. Modulation of the heat-shock response.

In summary, two assignments can be made for the molecular chaperones: housekeeping and defence against stress functions. It is therefore not surprising that, to fulfill all their roles, the chaperones function with a cohort of accessory factors [19a].

13.2 The Folding Machines

13.2.1 The Trigger Factor (TF)

The trigger factor (TF) is an ATP-independent chaperone. It displays both chaperone and peptidyl-prolyl-*cis-trans*-isomerase (PPIase) activities [19b]. It is not a heatshock protein, but is induced upon cold shock and enhances *E. coli* viability at low temperatures [20a]. The lack of TF (deletion of the gene *tig*) has almost no effect on *E. coli* growth, but a strain with deletions of both *tig* and *dnaK* can survive only at low temperature [20b, 20c].

The trigger factor interacts with the ribosome [21–23] and is, along the chaperone pathway, the first that affects the folding of newly formed protein chains, scanning for prolyl bonds that need catalysis of isomerization. However, the binding of TF to peptides is not dependent on the presence of proline residues, and it is not known whether PPIase activity is required for the TF chaperoning of nascent chains. Binding of TF to the ribosome is important for creating a high local concentration of substrates [24, 25].

The trigger factor is composed of three domains: an N-terminal domain (NTD), which mediates association with the large ribosomal subunit, a central substrate binding and PPIase domain, and a C-terminal domain (CTD). It is monomeric in its ribosome-associated state, but uncomplexed TF is in monomer–dimer equilibrium, with two-thirds existing in a dimeric state [26].

13.2.2 The DnaK/DnaJ/GrpE System

The control of protein folding by DnaK is coupled to its ATPase activity, and these two activities correspond to two functional domains: the 44 kDa NTD (residues 1–385) which binds and hydrolyses ATP; a CTD (residues 390–638) which consists of (i) a 18 kDa β-sandwich subdomain which binds and releases polypeptide targets (substrate-binding domain) [27] and (ii) another subdomain (10 kDa) composed of five α-helices (residues 537–638) [28a] that acts like a lid [286] over the β-sandwich subdomain to encapsulate the bound peptide in the ADP-bound state; the bound peptide contacts the β-sandwich but not the lid.

The DnaK-binding motif in substrates consists of a core of up to five large hydrophobic or aromatic residues, and flanking regions enriched in positively charged residues which are of decreasing importance with increasing distance from the core [29]. As both denatured proteins and folding intermediates display hydrophobic surfaces, DnaK and the various eukaryotic HSP70s stabilize non-native polypeptides through the binding and release of these extended hydrophobic peptide segments that are normally buried in the fully folded form, but are exposed during protein synthesis, protein translocation and protein degradation. However, *E. coli* DnaK also binds some native proteins such as λP, λO, λcIII, RepA, heat-shock

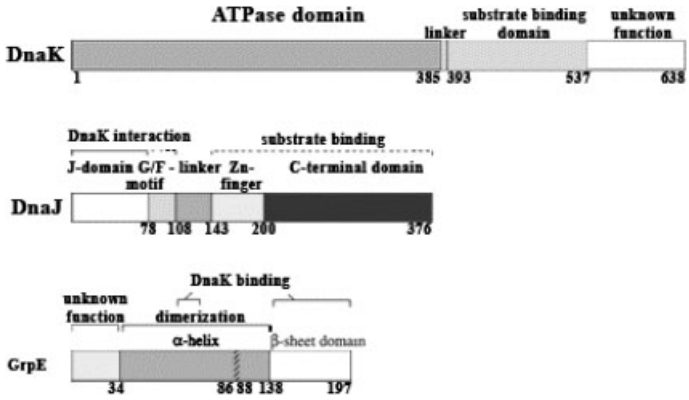


Figure 13.3 Domain organization of the *E. coli* chaperones DnaK, DnaJ and GrpE. Residue numbers define the approximate individual domain borders. Residues 386–392 of DnaK constitute a linker between the ATPase and the substrate-binding domain. Residues 86–88 of GrpE constitute a break of the long N-terminal α helix in the GrpE monomer that interacts with DnaK. Reprinted with permission from Bukau and Horwich [299].

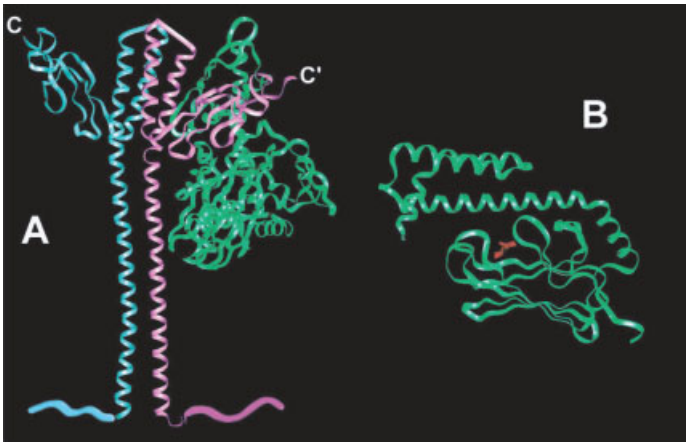


Figure 13.4 Structures of the *E. coli* chaperones GrpE and DnaK. (A) Structure of the GrpE homodimer complexed to the ATPase domain of DnaK. The proximal and distal GrpE monomers are shown in purple and light blue, respectively. The ATPase domain is shown in green. In total, there are six contact areas between DnaK and GrpE. (B) Structure of the C-terminal substrate binding and lid domains of DnaK. DnaK is shown in green and the bound peptide in red. Reprinted with permission from Chesnokova et al. [300].

transcription factor σ^{32} [30], the tumor suppressor protein p53 [31]. BIP, a resident endoplasmic reticulum HSP70, associates with immunoglobulin heavy and light chains [32] and human immunodeficiency virus envelope glycoprotein gp160 [33].

There is a mutual stimulation of ATPase activity and substrate release [34–40]. Interestingly, interdomain coupling occurs even when the lid is deleted, but potassium ions are indispensable for the mutual functional control of the two domains [41a]. The ATP-bound form of DnaK is characterized by high on- and off-rates of substrate interactions (fast peptide binding and release) and a low affinity, whereas the ADP-bound form is characterized by low on- and off-rates (slowly binding and release) and high affinity for substrates. In other words, DnaK in its ADP state captures substrates, and ATP inhibits the capture [40]. The ATPase activity thus constitutes a switch regulating the velocity and stability of substrate binding by DnaK.

An interesting difference between proteins of the DnaK/HSP70 family and the GTPases is that the latter are activated to bind proteins when they contain bound nucleotide *tri*phosphate (GTP) [41b], whereas HSP70 forms stable complexes with protein substrates when nucleotide *di*phosphate (ADP) is bound.

The ATPase activity of DnaK is itself under tight control of two cofactors: DnaJ, which markedly stimulates the ATPase activity [42, 43], and GrpE, which facilitates the ADP/ATP exchange since DnaK binds ADP more tightly than ATP [44]. The stimulation of the ATPase activity requires the conserved J domain [45] of DnaJ (residues 2–78 in *E. coli*; [46]). DnaJ tightly couples ATP hydrolysis with binding of protein substrate by DnaK [47] through a mechanism that involves communication between the ATPase and the substrate-binding domains of DnaK. But DnaJ itself is also capable of associating with unfolded substrates and preventing aggregation, having most binding motifs in common with DnaK [48]. This qualifies DnaJ as a chaperone in its own right and as a targeting partner for DnaK [42].

The DnaK reaction cycle in protein folding is therefore the following: DnaJ acts on the ATP-bound DnaK (rapid substrate binding and release, low affinity, so-called T state) to stimulate the hydrolysis of ATP, resulting in the ADP-bound form of DnaK, which binds the substrate tightly (so-called R state). DnaK probably transduces free energy from ATP binding and hydrolysis to produce a conformational change in the substrate protein that increases the probability of proper folding. Substrate ejection then requires the dissociation of ADP, which is catalyzed by GrpE, and which occurs in concert with binding of a new ATP molecule. Therefore, under *in vivo* conditions with an estimated chaperone ratio of DnaK/DnaJ/GrpE = 10/1/3, both DnaJ and GrpE appear to control the chaperone cycle by transient interactions with DnaK [49a].

In mammalian cells, a network of co-operating and competing chaperone cofactors, such as the DnaJ-like HSP40s, HAP46/BAG-1[49b] (Snl 1 in yeast; [50]), Hip, Hop (Sti 1 in yeast; [51]), and CHIP [52] modulates the chaperone activity of the heat-shock cognate protein HSC70 [53] (see Fig. 13.12). For a synopsis of the *E. coli*, yeast and mammalian HSP40s proteins see Table IV in Ref. [42]. The large tumor antigen (T antigen) of simian virus 40 (SV40) is also a DnaJ molecular chaperone [54]).

Concerning the role of DnaK in assisting folding of newly synthesized cytosolic proteins, it had been thought for long time that DnaK plays a critical role in *de novo* folding (see Sect. 13.3.1). Indeed, DnaK is present in the *E. coli* cytosol at ~50 μM, roughly equivalent to the concentration of ribosomes. However, the fraction of newly translated proteins that is recovered in a complex with DnaK is only 5–10% of the total soluble *E. coli* proteins at 30°C, preferentially in the size range of 30–75 kDa [55]. Also a study with a mutant lacking the function of DnaK (*ΔdnaK*) indicates that, at 30°C at least, DnaK is not essential for bacterial viability, nor for *de novo* folding of the majority of *E. coli* proteins [56]. However, DnaK is essential for *E. coli* growth at 37°C and above, and a large set of thermolabile *E. coli* proteins are substrates of DnaK and of two other chaperones, ClpB [57] and trigger factor [58] during heat stress, both *in vivo* and in cell extracts.

In eukaryotes, the so-called heat-shock "cognate" proteins HSC70s are non-inducible but constitutively expressed homologs of the HSP70s. Interestingly, the HSP70 genes (but not the HSC70 genes) are intronless. Although unusual (other genes that lack intervening sequences include histones and α-interferon), this feature is perhaps significant for genes that are rapidly activated at the transcriptional level. Two DnaK homologs have been found in *E. coli*: Hsc66 (=HscA) which seems to be specialized for the assembly of iron–sulfur cluster proteins [59], and Hsc62 (=HscC) [60], the function of which is less clear: it forms a complex with the transcription factor σ^{70}, and may function as its negative modulator [61], but it may also be involved in the repair of damage induced by radiation and cadmium since mutations in this protein have been identified that are hypersensitive to ultraviolet light and cadmium. Triple knock-outs of all *E. coli* genes encoding HSP70 proteins *(ΔdnaK ΔhscA ΔhscC)* are viable, indicating that HSP70 proteins are not strictly essential for viability. There are also four additional DnaJ proteins in *E. coli*: (i) CbpA, (ii) DjlA, both of which interact with DnaK, (iii) Hsc20 (=HscB), which interacts with HscA, and (iv) DjlC, which seems to be the appropriate DnaJ cochaperone for HscC. The unique common feature of all the DnaJ homologs is a short sequence of about 75 amino acid residues called the J domain, which is essential for the interaction with an HSP70 chaperone partner and for the stimulation of its ATPase activity [62].

13.2.3 The GroEL/GroES System

Early genetic studies identified the *E. coli groES* and *groEL* genes because mutations in them blocked the growth of bacteriophages λ, T4 and T5. GroEL/GroES (HSP60/HSP10) are also known as Group I chaperonins, to make a distinction with Group II chaperonins of archea and eukarya (also called TRiC/CCT) [63]) (see Fig. 13.7): the latter have a few substrates including the cytoskeletal proteins actin and tubulin [64–66], Cdc20 [67a], the Von Hippel-Landau tumor suppressor complex [67b] and the WD-repeat proteins [67c]. GroEL/ES are large cylindrical complexes that promote protein folding in the sequestered environment of their central cavity. Group I chaperonins are present not only in all eubacteria [67d], but also in mitochondria and chloroplasts [68a, 68b].

GroEL/GroES is the only chaperone system in *E. coli* cytoplasm essential under all growth conditions [69]. GroEL is organized in two stacked rings, each composed of seven 60 kDa subunits [70, 71] (see Fig. 13.5). However, mutants of GroEL that are fully functional as single rings have been recently isolated [72].

The co-chaperonin GroES, a single ring of seven 10 kDa subunits, forms the lid on a folding cage (Group II chaperonins have not such a detachable GroES-like co-chaperonin [73], see Fig. 13.7). Non-native proteins are encapsulated in this cage, or passive box, which, when capped by GroES in the presence of ATP, creates an environment of infinite dilution inside its central cavity where individual polypeptide chains are free to fold without risk of aggregation. Then the dissociation of GroES allows release of the trapped protein from the cavity [74] (see Fig. 13.6). However it remains a matter of debate whether the chaperonin cage plays only a passive role in protecting the protein substrate from aggregation, or an active role in accelerating folding rates [75].

Typical GroEL substrates consist of two or more domains with $\alpha\beta$-folds, which contain α-helices and buried β-sheets with extensive hydrophobic surfaces. Binding takes place between the hydrophobic residues in the apical domain of GroEL and the hydrophobic faces exposed by the β-sheets in the $\alpha\beta$-domains of protein

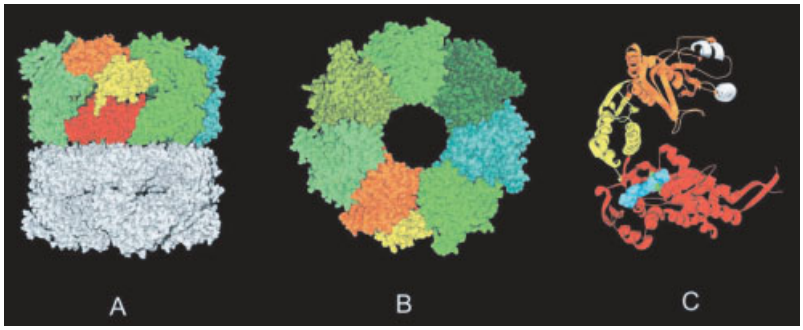


Figure 13.5 Structure of the GroEL chaperonin from *E. coli*. (A) Side-view of the GroEL tetrade-camer. Subunits comprising the top ring are shown in colour, subunits of the bottom ring are shown in grey. Each subunit can be dissected into three domains: apical (orange), intermediate (yellow) and equatorial (red). (B) Top view of the GroEL tetradecamer. The seven subunits of the ring are shown in shades of green. For one subunit, the apical and the intermediate domains are highlighted in orange and yellow, respectively. (C) Ribbon representation of a GroEL subunit. The equatorial domain (red) consists almost exclusively of α helices and contains the nucle-otide-binding site, which is occupied by ATPγS (blue). The intermediate domain (yellow) serves as a molecular hinge that connects the equatorial domain with the apical domain (orange). Binding of GroES and polypeptides occurs in a hydrophobic groove formed by the two helices (white) facing the central cavity. Reprinted with permission from Walter [71].

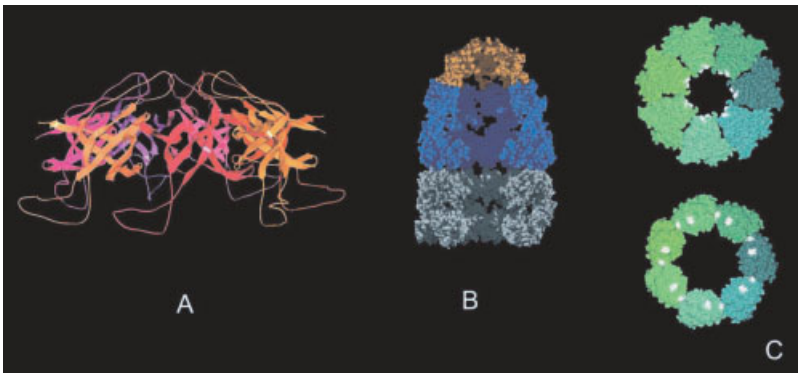


Figure 13.6 Structure of the GroELS chaperonin from *E. coli*. (A) Sideview of the GroES heptamer. The individual subunits (in shades of red) consist mainly of β-sheets and form a dome. The flexible extensions on the bottom are the so-called mobile loops that mediate binding to GroEL. (B) Cross-section of a GroE "bullet". Each GroEL ring encloses a cavity that serves as a folding compartment for a polypeptide substrate. Binding of GroES (orange) to the top GroEL ring (blue) blocks the access to the upper cavity and concomitanly induces an *en bloc* movement of the apical domains. (C) Changes in the GroEL structure upon binding of GroES. In this top view, the seven subunits comprising one ring of GroEL are shown in shades of green and blue. The hydrophobic residues in the apical domains important for binding of polypeptide and GroES are shown in white. In the absence of GroES (top panel), these residues coat the inside of the central cavity and account for the high affinity for unfolded polypeptides. Upon binding of GroES (lower panel) the apical domains rotate outwards by 90°. The hydrophobic patches become buried in the subunit interfaces, rendering the inner surface of the cavity mainly hydrophilic and causing the release of a bound polypeptide. Concomitantly, the diameter of the cavity increases from 45 to 80 Å. Reprinted with permission from Walter [71].

substrates [76]. Binding stimulates ATP binding and hydrolysis, causing a conformational change in the box. The transmission of an allosteric signal between the two rings of the GroEL complex is a key aspect of the reaction mechanism [77, 78].

Only 10–15% of all newly synthesized polypeptides transit GroEL post-translationally [79, 76], which agrees with the intracellular concentrations of 2.6 and 5.1 μM reported for GroEL and GroES, respectively, since this is sufficient to facilitate the folding of no more than 5% of all of the proteins within the *E. coli* cell [80]. GroEL is absolutely essential for the correct folding of *E. coli* dihydropicolinate synthase, the first enzyme in the diaminopimelic acid synthesis pathway, and therefore also for cell-wall synthesis [81].

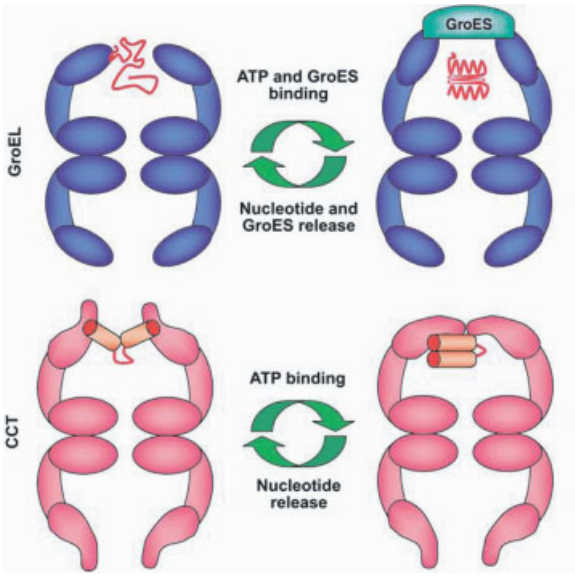


Figure 13.7 Schematic model of the folding mechanism of GroEL and CCT. Both chaperonins cycle between an open, substrate-receptive conformation and a closed conformation. In GroEL, substrate recognition and binding is performed in the open conformation by a hydrophobic region in the apical domain of the GroEL subunits. The closed conformation is generated upon ATP binding in the presence of the cochaperonin GroES, and the unfolded polypeptide is liberated in the GroEL cavity where folding takes place. The polypeptide is then liberated from the GroEL cavity after GroES release, which is induced upon ATP hydrolysis. In CCT, built up by two superimposed rings, each ring is constituted by eight different, albeit homologous subunits (30% identity). The apical domains of specific subunits recognize the substrate. The sealing of the CCT cavity, carried out by the movements of the apical domains induced upon ATP binding, is performed by the helical protrusions present at the tip of the apical domains, and the substrate is not liberated in the CCT cavity but remains bound to the apical domains and forced to acquire a more compact, native conformation. Reprinted with permission from Valpuesta et al. [66].

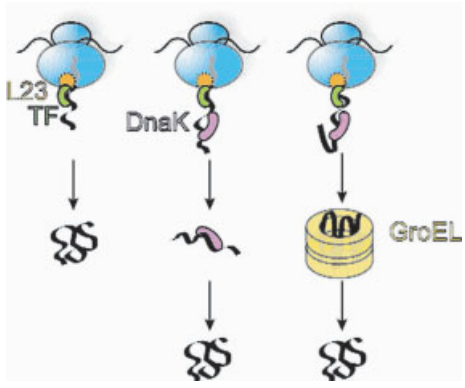


Figure 13.8 Three bacterial chaperones participating in *de novo* protein folding. The Trigger Factor interacts with emerging nascent chains via its interaction with ribosomal protein L23 at the ribosomal exit site. Some polypeptides then interact with DnaK and GroEL, which assist the folding of selected subsets of cellular proteins. Reprinted with permission from Albanese and Frydman [142].

13.2.4 Other Chaperones

13.2.4.1 HSP90

HSP90 is one of the most abundant chaperones in the cell, and is already at high levels prior to cellular stress. HSP90 is a constitutive homodimer and has only recently been recognized as an ATP-dependent chaperone [82] (see Fig. 13.9).

HtpG, the bacterial HSP90 homolog is non-essential in *E. coli* [83], but is essential for the thermal stress management in cyanobacteria [84]. In eukaryotes, HSP90-null mutants are lethal. In yeast, HSP90 does not act generally in nascent protein folding [85], but some substrates (i.e., the α-complemented β-galactosidase) depend on HSP90 for folding, assembly and/or stabilization. In higher eukaryotes, most HSP90 substrates are signal transduction proteins, such as steroid hormone receptors and signaling kinases. Interaction of the glucocorticoid receptor with HSP90 is essential for its activity. HSP90 operates as part of a multichaperone family which includes HSP70 and several cochaperones such as p23 [85b], Hop, CHIP, Cdc37, Aha1 (see [86] for a Table of the HSP90 interacting proteins). Ansamycin drugs which specifically target HSP90, the most representative being geldanamycin, makes HSP90, involved in many growth-regulatory pathways, an attractive target for cancer therapeutics [87a, 87b]. For another astonishing property of HSP90, see Sect. 13.7.1.

13.2.4.2 Clp/HSP100 Family

ClpB/HSP104 and ClpA, ClpX and ClpY (=HslU) are ATP-dependent chaperones associated with disassembly and degradation of protein complexes. They are members of the AAA+ family of proteins, which are ATPases Associated with a variety of cellular Activities [88–90].

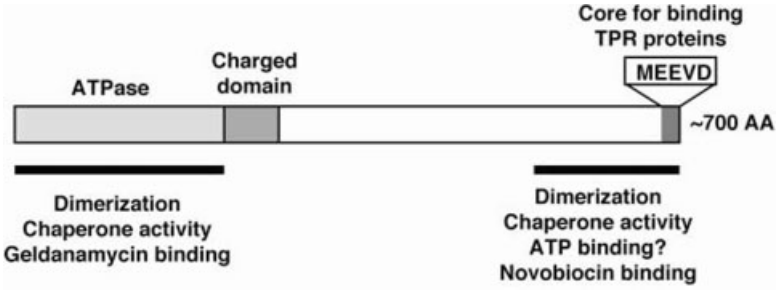


Figure 13.9 Schematic representation of the domain structure of HSP90. ATP/ADP and the HSP90 inhibitors geldanamycin and radicicol bind to the same pocket inside the N-terminal domain. Following the N-terminal ATPase domain, all eukaryotic HSP90 proteins have a charged domain of varying size. They end at the very C-terminus with the pentapeptide MEEVD, which constitutes the core of the HSP90 interaction surface for the tetratricopeptide repeats (TPR) of HSP90 co-chaperones. Neither the charged domain nor the pentapeptide is required for viability in *S. cerevisiae*. Reprinted with permission from Picard [86].

The common physical feature of *E. coli* ClpA, ClpX and ClpY is a ring structure formed by their ATPase subunits, which arrange into ring-shaped hexameric or heptameric complexes enclosing a central cavity. They act as regulatory subunits of proteases, for example, the homohexameric ClpA, or ClpX, associate with each side of the heptameric double-ring protease ClpP, delivering recognized substrates to it for degradation [91]. This structure (ClpA6, ClpP7×2, ClpA6) resembles that of the eukaryotic 20S core proteasome (=28 subunits) arranged in four homoheptameric rings, as $\alpha 7$, $\beta 7$, $\beta 7$ and $\alpha 7$ [92]. Similarly, in the *E. coli* HslUV complex, each HslU (=ClpY) hexamer binds to opposite ends of the dodecamer HslV (=CplQ) protease component [93]. In the absence of association with the protease partner ClpP, ClpA or ClpX can mediate disassembly of oligomeric substrate proteins, such as the ClpX-mediated disassembly of the Mu transposase tetramer [94] and the ClpA-mediated remodeling of bacteriophage P1 RepA dimers (inactive) into active monomers [95] (see Fig. 13.13) or the *in vitro* unfoldase activity of ClpA [96]. Similarly, without its protease partner HslV (=ClpQ), the HslU (=ClpY) ATPase acts as a molecular chaperone to prevent aggregation of SulA, an inhibitor of cell division in *E. coli* [97].

Unlike ClpA, ClpX and HslU, the protein ClpB is unique among the HSP100 proteins of *E. coli* since it does not interact with a proteolytic partner. *E. coli* cells deleted of *ClpB* show a higher rate of death above 50°C, indicating that ClpB is essential for cell survival at high temperatures, like HSP104 [98] in yeast. ClpB, which has two ATPase domains [99] but whose physiological oligomeric state is unclear at the moment (most probably an hexamer)[100a–c] , does not assist itself in protein folding, but disaggregates preformed protein aggregates before transferring them to HSP70 (see Sect. 13.3.2).

13.2.4.3 **DegP**

E. coli DegP (=HtrA) is a periplasmic heat-shock protein, which possesses both chaperone and protease activities. The chaperone function dominates at low temperatures and the protease function, at elevated temperatures [101]. DegP becomes necessary for *E. coli* growth at temperatures above 39°C for the removal of the misfolded outer-membrane proteins [102]. The proteolytic sites are located in the central cavity of the DegP hexamer formed by the staggered association of trimeric rings [103]. This unique structural organization indicates a new type of protease chaperone machine.

13.2.4.4 **Periplasmic Chaperones**

Aside from DegP (see Sect. 13.2.4.3) and pili chaperones (see Sect. 13.2.4.5), the *E. coli* periplasm contains many other chaperones involved in the folding and targeting specifically of outer-membrane proteins [104] with the exception of TolC [105]. Some of these chaperones have redundant functions in protein folding [106], perhaps because of the particular conditions (absence of ATP and a relatively more

oxidizing state than in the cytoplasm) prevailing in the periplasm [107]. Misfolding of proteins in the cell envelope of *E. coli* induces a cascade of particular signaling pathways mediated by the Rse proteins and the transcription factor σ^E [108].

13.2.4.5 **Pili Chaperones**

Some periplasmic chaperones such as PapD and FimC are specialized in the assembly of bacterial pili (the adhesive structure that enables bacteria to bind to host cells). The priming action of the chaperone drives subunit assembly into the fiber [109]. The integration of the next pilin, the building element of the *growing* pilus, requires the removal of the chaperone sitting at the growing end. Each pilin is also complexed with a chaperone which promotes a partially unfolded state of the pilin that is required for assembly (see Fig. 13.10): in other words, the chaperone provides the missing information for folding of the pilin [110] and this case of structural complementation represents therefore an interesting extension of the Anfinsen dogma described in Sect. 13.1.

13.2.4.6 **Small HSPs**

The α-crystallins [111, 112] are the most representative members of the small heat-shock protein family (sHSPs) which form dynamic oligomeric structures [113]. Their chaperone activity, required to prevent aggregation of intermediate filaments [114], is of potential importance in human disease [115]: for example, a mutation in the αB-crystallin chaperone gene causes a desmin-related myopathy [116a–d]. There is another example of a mutation in a gene encoding a putative chaperonin, which causes the McKusick–Kaufman syndrome [117a], and of mutations in the sHSP22 [117b] and sHSP27[117c] which cause neurodegenerative disorders.

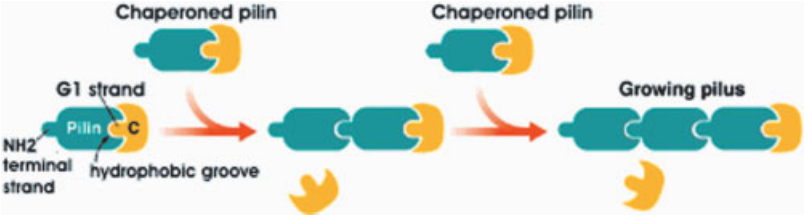


Figure 13.10 Benefits of a chaperone for a growing pilus. The pilus is the adhesive structure that enables bacteria to bind to host cells. A chaperoned pilin protein is added to a growing pilus in the following way: the chaperone carries the pilin subunit to the large pore subunit (called the usher) where the pilin subunit is released by the chaperone and becomes attached to the end of the gowing pilus rod. Therefore, the pilin can be complemented by either a different molecule (the chaperone) or the same molecule (another pilin). Reprinted with permission from Eisenberg [298].

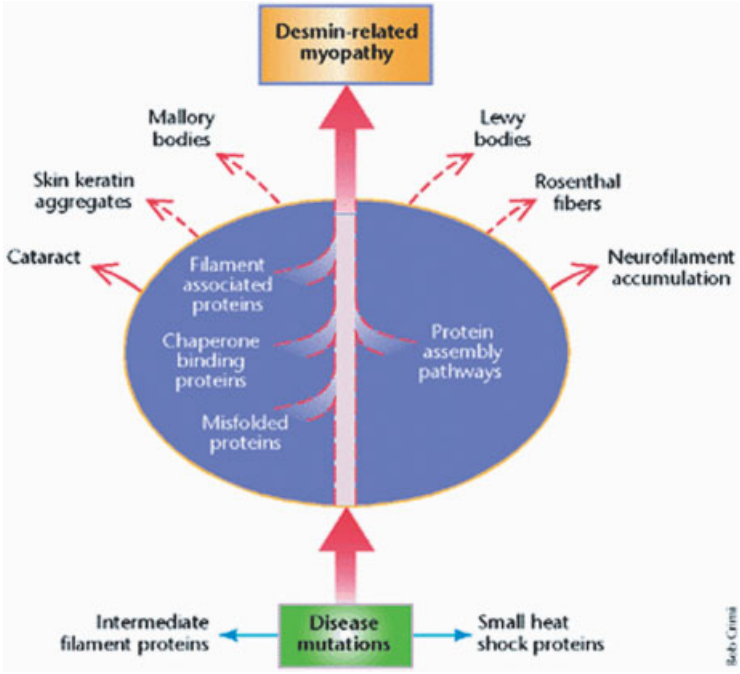


Figure 13.11 Small heat-shock protein chaperones for the intermediate filaments. Mutations in either small heat-shock proteins, such as α B-crystallin, or intermediate filament proteins lead to collapse and aggregation of cellular intermediate filament networks resulting in skeletal muscle and cardiac myopathy. Reprinted with permission from Quinlan and Van Den Ijssel [114].

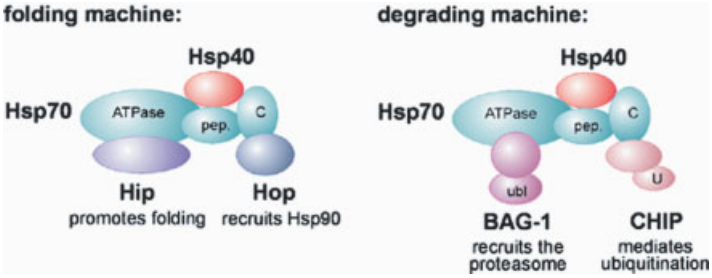


Figure 13.12 HSP70 as a folding machine or a degrading machine. Binding of distinct co-chaperones to the N-terminal ATPase domain of HSP70 (ATPase) and to its C-terminus (C) gives rise to chaperones machines involved in protein folding or in protein degradation. The co-chaperones Hip and BAG-1 compete for binding to the ATPase domain, whereas Hop and CHIP associate with the C-terminus in a competitive manner. (pep = peptide-binding domain of HSP70; ubl = ubiquitin-like domain of BAG-1; U = U-box of CHIP). Reprinted with permission from Höhfeld [195].

In *E. coli*, the genes encoding HSP15, a ribosome-associated heat-shock protein [118], and HSP33, a redox-regulated molecular chaperone [119a, 119b] form part of a heat-shock-regulated multigene operon. HSP31, the *yedU* (*hchA*) gene product, alleviates protein misfolding by interacting with early unfolding intermediates [120a–f]. The 16 kDa IbpA and IbpB (for inclusion-body associated heat-shock proteins) are dispensable in *E. coli*, but cooperate with ClpB and DnaK/DnaJ/GrpE (and not with GroEL/GroES) *in vivo* at high temperatures [121, 122] and *in vitro* [123]. A novel sHSP named AgsA, which supresses protein aggregation, has been recently discovered in *Salmonella enterica* [124].

In plant cells, sHSPs, which are the predominant proteins synthesized under stress conditions, exhibit chaperone activity in cooperation with HSP70 [125, 126a, 126b]. sHSPs which are incorporated into protein aggregates help to the disaggregation reaction mediated by ClpB/DnaK, highlighting a role of sHSPs in cellular protein quality control [127] (see section 13.3.2).

13.2.4.7 **Endoplasmic Reticulum (ER) Chaperones**

Several ER chaperones, including calnexin and calreticulin, co-operate to ensure the correct biogenesis of glycosylated proteins [128, 129a, 129b].

13.2.4.8 **Intramolecular Chaperones**

The prosequences of some proteases act as intramolecular chaperones for the correct folding of the polypeptide chains to which they are covalently but transiently attached: this is the case for the 77-residue propeptide of subtilisin [130]. Similar folding mechanisms are used by several prokaryotic and eukaryotic proteins, including prohormone convertases. In the *Bordetella pertussis* autotransporter BrkA, which contributes to adherence of the bacterium to host cells, a conserved region acts as an intramolecular chaperone to effect folding of the 73 kDa domain (referred to as the passenger domain) and ferry it to the bacterial surface [131].

13.3
Chaperone Networks

13.3.1
De novo Protein Folding

It is important to know whether chaperones act independently, or are organized in a universal chaperone pathway, through which proteins, especially newly translated polypeptides, will be channelled while they are still ribosome-bound. Alternatively, unfolded polypeptides following translation will be discharged into the soluble cytosolic medium and then cycle between the endogenous chaperones (kinetic partitioning) [132]. To address that question, a mutant form of GroEL called “trap GroEL”, which irreversibly captures unfolded polypeptides, when overexpressed in growing yeast and mammalian cells, did not interrupt the productive folding pathway, indicating a high level of organization in folding reactions [133, 65, 134], i.e., the existence of

an integrated folding compartment, directly coupled with the translation machinery [135]. Residing on the yeast ribosome, the HSP70 Ssb [136, 137] forms a chaperone triad with the HSP70 homolog Ssz1 and zuotin, a DnaJ-related HSP40, to act on nascent chains emerging from the ribosome [138]. During *in vitro* translation, a nascent polypeptide-associated complex (NAC) has been shown to form a protective environment for regions of nascent chains just emerging from the ribosomal tunnel [139, 140a]. In the endoplasmic reticulum too, a subset of chaperones and folding enzymes form multiprotein complexes to bind nascent proteins [140b].

In the case of prokaryotes, such a rigid sequential pathway is less obvious [140c], but nascent chains emerging at the peptide exit tunnel of the ribosome are awaited by a "welcoming committee" consisting of three chaperones [141]: the sequence of interactions of newly synthesized proteins with chaperones is proposed to be trigger factor (TF), then the HSP70 system, and then the GroEL/ES system [142] (see Fig. 13.8). TF interacts with relatively short emerging nascent chains, via its interaction with ribosomal protein L23 at the ribosomal exit site [21, 22], and may then migrate transiently with the nascent chain, followed by rapid dissociation and rebinding to ribosomes. DnaK, which has partially overlapping functions with TF [58, 143, 55], binds to longer chains and allows the larger polypeptides to fold. Although complementary, the mechanisms of DnaK and TF in protein folding are distinct from each other [144–146]. Indeed, DnaK is not recruited to translating ribosomes in the absence of TF [147].

Finally, GroEL functions post-translationally to assist folding of a subset of cytosolic proteins [76]. DnaK and GroEL have been shown to co-operate in many folding pathways [148–152, 20c] and in ribosome biogenesis at high temperature [153]. However, DnaK and GroEL do not obligatorily act in succession by promoting earlier and later folding steps, respectively: rather, they can form a lateral network, for example for a large protein that does not fit the GroEL/GroES cavity [154]. ClpB and HtpG also participate in *de novo* folding in mildly stressed *E. coli* cells [155], and in the eubacteruim *Thermus thermophilus*, a protein, DafA, cycles between the DnaK chaperone system and translational machinery, although its role is unknown at the moment [301].

However, the ribosome itself could be the first player in facilitating cotranslational and sequential *domain folding*, in the case of eucaryotic proteins composed of multiple domains [156]. Indeed, co-translational folding of a virus capsid protein has been observed without release from the ribosome and without the assistance of HSP70: folding is more favorable kinetically if it occurs at the same time as protein synthesis, but it remains to be seen whether this protein is exceptional in its capacity to fold so efficiently while still bound to ribosomes [157]. The ribosome may play a key role [158], either because the vectorial aspect of protein biosynthesis is critical in the process of protein folding, or in recruiting the folding machinery [159], or in controlling the subsequent fate of the translated polypeptide, through its interaction with the peptide exit tunnel as shown in *E. coli* [160–162] or even in actively refolding unfolded proteins, as shown *in vitro* [163].

13.3.2 Protein Disaggregation

During prolonged heat shock of *Saccharomyces cerevisiae* cells, the high-molecular-weight aggregated proteins that form are rapidly eliminated in wild-type cells, while they persist in cells lacking the stress tolerance factor HSP104. Refolding from the aggregated state requires not only HSP104, but also the HSP70 homolog Ssa and the DnaJ homolog YDJ1 [164]. A close functional relationship between HSP104 and HSP70 has also been shown *in vivo* [165], and in their ability to restore mRNA splicing after heat inactivation [166]. In maintenance of mitochondrial function, HSP78, the mitochondrial ClpB homolog, co-operates with matrix HSP70 [167, 168]. However, HSP104 and HSP70 have antagonistic interactions in yeast prion curing [169] (see Sect. 13.7.2).

In Thermus thermophilus [170, 171] as well as *E. coli* [172–176, 57, 177], a multichaperone system composed of ClpB and DnaK/DnaJ/GrpE has been shown to prevent and revert protein aggregation both *in vivo* and in cell extracts. There is a recent indication that ClpB acts prior to DnaK on protein substrates [178]. Therefore, the sequential mechanism could be the following: (i) ClpB exposes new DnaK-binding sites on the surface of the aggregates; (ii) DnaK binds the aggregate surfaces and melts the incorrect hydrophobic associations; and (iii) DnaK and GroEL complete refolding of solubilized proteins [179]. Small HSPs, which are incorporated into protein aggregates, help the disaggregation reaction mediated by ClpB and the DnaK system [123, 127].

The interactions of chaperones with disease-causing misfolded proteins [180a, 180b] and toxic aggregates, such as polyglutamine-rich proteins [181a, 181b], neurodegenerative amyloid plaques and prions is therefore a promising field of research [182–186].

ClpB also cooperates with HtpG in facilitating *de novo* protein folding in mildly stressed *E. coli* cells [155], and with DnaK/DnaJ/GrpE in the activation of the plasmid RK2 replication initiation protein TrfA by converting inactive dimers to an active monomer form [187].

13.3.3 Posttranslational Quality Control

Proteins that cannot fold properly because of mutations, errors in translation or stress damage are degraded by the ubiquitin–proteasome system in eukaryotes [188], or by specific proteases (FtsH, Lon, DegP, HslUV, ClpAP and ClpXP) in bacteria. Co-operation between chaperones and proteases, both of which recognize hydrophobic regions exposed on unfolded polypeptides, ensure post-translational quality control, i.e., whether a damaged protein will be destroyed before it folds properly [189]. The kinetics of partitioning of non-native proteins between chaperones and proteases, in other words, the competition between these two sets of proteins has been proposed to underlie the basis for protein quality control [190, 191].

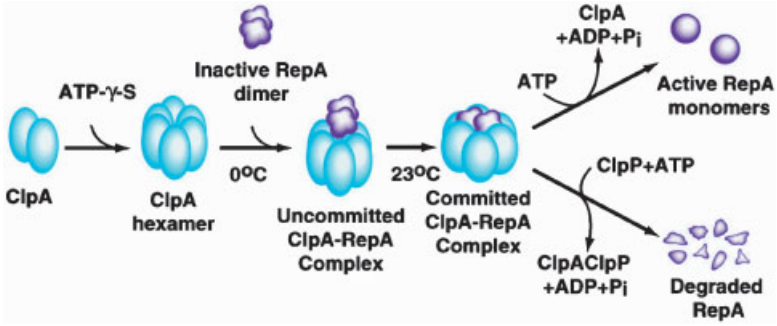


Figure 13.13 Dual function of the *E. coli* ClpA chaperone in RepA monomerization and in the ClpAP protease activity. For remodeling the bacteriophage P1 RepA dimers into monomers, the initial step is the self-assembly of ClpA and its association with inactive RepA dimers. Upon ATP binding, stable complexes containing 1 mol of RepA dimer per mol of ClpA hexamer are formed and are committed to activating RepA. Finally, active RepA monomer is released upon ATP hydrolysis. The alternate fate for RepA is the degradation by ClpAP protease. Reprinted with permission from Pak and Wickner [95].

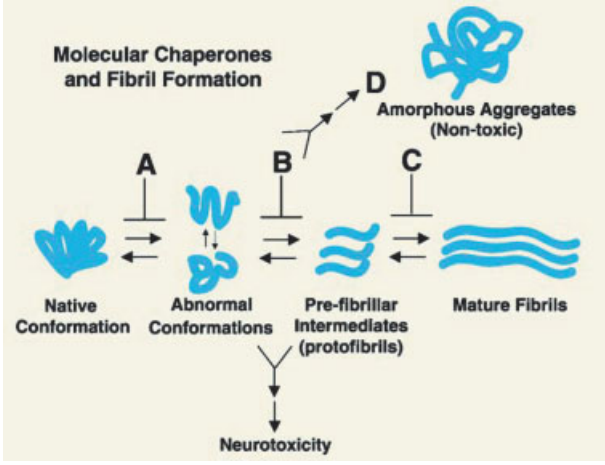


Figure 13.14 A model for molecular chaperone suppression of neurotoxicity. Amyloid is an ordered fibrillar structure arising from partial unfolding and exposure of hydrophobic surfaces that are normally buried in the core of a folded protein. Amyloid is also formed by proteins containing an expanded series of glutamine repeats (polyQ). Molecular chaperones may suppress neurotoxicity of amyloid-forming proteins by preventing their conversion between native conformations and toxic conformations (A), or the formation of pre-fibrillar intermediates (B), or the conversion between pre-fibrillar intermediates and mature fibrils (C), or by facilitating the conversion of toxic intermediates into nontoxic amorphous aggregates (D), readily degraded by the proteolytic machinery. Reprinted with permission from Muchowski [186].

In *E. coli*, a synergistic action of two proteases, Lon and ClpXP, and the DnaK system is revealed at 42°C, as proteases became essential for survival at low DnaK levels [192]. Recent results suggest that DnaK operates as a functional coupling between the posttranslational quality control of proteins and the late stages of *E. coli* ribosome biogenesis [193] (see section 13.7.3 and Fig. 13.15).

In eukaryotes, misfolded proteins are targeted for proteosomal degradation through polyubiquitination by the ubiquitin ligase CHIP, which is also a co-chaperone of HSP70 and HSP90, therefore linking chaperones to the degradation machinery [194, 195]. The fact that an increased ubiquitin-dependent degradation can replace the requirement for HSPs [196] is also indicative of the protein quality control exerted by the chaperones, and furthermore suggests that heat-stressed cells do not die because of the loss of protein activity, but because of the inherent toxicity of denatured or aggregated proteins.

Another quality control mechanism in *E. coli*, triggered when protein synthesis on the ribosome stalls for any of a variety of reasons, is the tmRNA-mediated tagging of incomplete nascent proteins, which targets them for degradation by specific proteases including ClpXP and ClpAP [197]. When normal protein synthesis is compromised, a ribosomal A-site mRNA cleavage also contributes to ribosome rescue and protein quality control [198].

13.4 Chaperones and Stress

13.4.1 The Heat-shock Response and its Regulation

Cells respond to stress by transcriptional activation of heat-shock genes, most of which encode molecular chaperones. Instead of increasing the concentration of a transcriptional activator, the heat-shock response in eukaryotes is mediated by activation of a pre-existing pool of transcription factors, the family of heat-shock factors (HSF1–4). In unstressed cells, HSFs are present in both the cytoplasm and nucleus as inert monomers that have no DNA-binding activity. In response to heat shock, HSFs assemble into trimers, which accumulate within the nucleus, bind to the heat-shock elements (HSE) located in the heat-shock-responsive gene promoters and become phosphorylated. This in turn leads to increased levels of HSP70. The attenuation of the heat shock occurs through conversion of the active trimeric form of HSF to the non-DNA-binding monomer. This negative regulation of HSF is mediated by direct binding of HSP70 and by another negative regulator, the heat-shock factor binding protein 1 (HSBP1), which interacts with both the trimeric state of HSF and HSP70, leading finally to the dissociation of HSF trimers [199].

It has been observed that HSP70/HSC70, which shuttle between nucleus and cytoplasm, localize to the nucleus and to the nucleolus upon exposure to heat stress, where they bind to incompletely folded proteins in the pre-ribosome assembly unit to protect them from irreversible denaturation [200]. Starvation also promotes nuclear

accumulation of the HSP70 Ssa4 in yeast cells [201]. However, neither HSC70 nor HSP70 is involved in nucleolar transcription in the amphibian oocyte [202].

In Gram-negative bacteria such as *E. coli*, the transcription factor σ^{32}, product of the *rpoH* gene [203], associates with RNA polymerase to direct transcription of heat-shock genes. Under normal conditions, DnaK and DnaJ sequester σ^{32} through direct binding, inhibiting its association with RNA polymerase and promoting its degradation (σ^{32} is an extremely unstable protein with a half-life of less than 1 min) [204]. Following heat shock, Dnak/DnaJ is titrated by misfolded proteins, allowing σ^{32} to recruit RNA polymerase to the heat-shock promoters. Then, as more non-native proteins are removed, DnaK becomes once again available to bind σ^{32}, thereby turning-off the heat-shock response [205, 206]. But heat-shock regulatory mechanisms exhibit great diversity among bacteria and are quite different in Gram-positive, proteobacteria and cyanobacteria, where both σ^{32} and another mechanism (the HrcA/CIRCE repressor-operator system) monitor the levels of unfolded protein in the cell to determine the need for the expression of major cellular chaperones [207].

13.4.2 Thermotolerance

Thermophilic organisms from the three phylogenetic domains (eubacteria, archea and eukarya) acquire thermotolerance, i.e., an enhanced survival at lethal temperatures, after a brief exposure to near-lethal temperatures [208]. This led to the hypothesis that thermotolerance depends on one or more of the HSPs synthesized after heat shock [209]. Indeed, HSP104 expression plays a central role in thermotolerance in yeast [98]. However, uncoupling thermotolerance from the induction of the HSPs has been observed [210, 211], indicating that other processes outside the heat-shock response are essential to the development of thermotolerance. For example, constitutive expression of HSP70 increases heat resistance [212, 213], but HSP70 alone is not able to confer the degree of resistance to heat killing seen with heat-induced thermotolerance [214].

In *E. coli,* DnaK plays an essential role in protection against protein oxidative damage and starvation-induced thermotolerance [215–217], probably in co-ordinating the sigma factors σ^{32} and σ^{S} levels [218, 219].

13.4.3 Who Detects Stress?

Temperature controls the expression of many bacterial genes; for instance, the transcription of genes encoding virulence factors which are expressed at 37°C, the host temperature, but are turned off at 30°C or below. Changes in DNA supercoiling, mRNA conformation, protein conformation, and chaperone-mediated capture of regulators have been implicated in thermosensing [220]. Highly conserved RNA sequences within 3′ untranslated regions have been postulated as sensors of environmental stress [221]. An mRNA-based thermosensor controls expression of rhizobial

heat-shock genes [222], of the virulence genes in *Listeria monocytogenes* [223] and of the primary step of the heat-shock response in *E. coli*, i.e. the translational induction of σ^{32} synthesis [224]. Temperature control of the *E. coli* transcription factor RpoS (σ^{S}) also depends on the synthesis and stability of the untranslated RNA DsrA [225] through temperature sensing by the *dsrA* promoter [226].

Many cases of temperature sensing by a protein are also known. The virulence plasmid-encoded protein TlpA in *Salmonella typhimurium* [227], the RheA repressor in *Streptomyces albus* [228a], the HrcA repressor in *Bacillus thermoglucosidasius* [228b], the small chaperone HSP26 in *Saccharomyces cerevisiae* [229], the DegP protease/chaperone in *E. coli* [101] (see Sect. 13.2.4.3) and DnaK itself [230, 231] serve as a cellular thermometer. The co-chaperone GrpE may act as a thermosensor, and when the temperature increases its reduced ADP/ATP exchange factor activity increases the time in which DnaK binds its substrates [232] and in this way adapts the DnaK/DnaJ/GrpE system to heat-shock conditions [233]. In eukaryotes also, HSP70 functions as a sensor in a Bag1/HSP70-mediated stress signalling pathway [234], and the regulatory domain of human heat shock factor HSF is sufficient to sense heat stress [235].

13.5 Assembly and Disassembly of Macromolecular Complexes

The first chaperone to be described was nucleoplasmin, a pentameric nuclear protein that mediates the formation of nucleosomes during early development [236]. Then it was shown that the *E. coli* GroEL/GroES chaperonin acts at an early stage in the head assembly pathways of bacteriophages λ [237], T4, T5 and μ [238]. GroEL/GroES also promotes assembly of the plastid ribulose-1,5-bisphosphate carboxylase (rubisco) [239], and of the molybdenum–iron protein of nitrogenase [240]. Moreover, it mediates iterative annealing of non-productive assembly intermediates at the quaternary structure level [241a]. Some histone chaperones assist chromatin assembly [241b] and small HSPs (see section 13.2.4.6), control the polymerisation of microtubules [184], microfilaments and intermediate filaments, i.e. the formation of the cytoskeleton [241c].

There are many examples in which DnaK or HSP70 plays a role in protein quaternary structure changes: disentanglement of the λDNA–λO–λP–DnaB preprimosomal complex found at the origin of λDNA replication [242], multimerization of the C protein, a positive regulator of bacteriophage μ late transcription [243], monomerization of the replication initiation protein RepE of mini-F plasmid, mostly present as inactive RepE dimers [244]; monomerization of the replication initiation protein RepA of P1 plasmid, mostly present as inactive RepA dimers [245], monomerization of the replication initiation protein TrfA of RK2 plasmid, mostly present as inactive dimers [187] and disassembly of clathrin-coated vesicles in cooperation with the cochaperone auxilin [246].

DnaK has also been implicated in the assembly of polyomavirus capsids [247], of bacterial ribosomes [193, 248] (see Fig. 13.15) and in the α-complementation

of β-galactosidase in *E. coli* [249] (the same is true for HSP90 in yeast) [250]. *In vitro*, denatured β-galactosidase forms active tetramers upon addition of HSP70 [251].

HSP70, which is associated with a distinct cytoplasmic aggregate during lymphocyte activation [252] and in *E. coli* with the insoluble (aggregated) proteins under conditions in which proteins tend to aggregate (severe heat shock) [56], dissociates hydrophobic protein aggregates, for example DnaA [253], alone or in cooperation with ClpB [179].

Finally, HSP70 participates, although transiently, in chaperone-mediated telomerase assembly (in contrast to HSP90 and its cochaperone p23 which remain associated with the active telomerase) [254]. It also works together with HSP90 in the formation of a ribonucleoprotein complex necessary for hepadnavirus assembly [255], in the biogenesis of the heme-regulated kinase of the α subunit of eukaryotic translation initiation factor 2 [256], and in the assembly and disassembly of steroid hormone receptors [257], whereas it blocks the assembly of a functional apoptosome (anti-apoptotic effect of HSP70) [258, 259].

Two other chaperones specifically suited for disaggregating proteins are members of the Clp family, ClpX, which disassembles the Mu transposase tetramer [94], and ClpA, which remodels bacteriophage P1 RepA dimers into monomers [95] (see Fig. 13.13), when they do not function with their collaborating protease ClpP (see ClpXP and ClpAP in Sect. 13.2.4.2).

Some proteins have their 'private chaperone', such as the LipA lipase of *Pseudomonas cepacia* [260] and myosin [261] : the protein UNC-45 functions both as a molecular chaperone and as an HSP90 co-chaperone for assembly of myosin into motile cellular structures essential for cell division, cell motility, and muscle contraction [262]. CcmE is a heme chaperone that binds heme transiently in the periplasm of *E. coli* and delivers it to newly synthesized and exported c-type cytochromes [263].

13.6 Protein Translocation Across Membranes

Some molecular chaperones are necessary for postranslational protein secretion in all three kingdoms of life (bacteria, archaea and eukaryotes) [264].

In yeast, HSP70 is involved in the routing of proteins to mitochondria, as it assists mitochondrial precursors to achieve "import-competence", probably by stabilizing translocation-competent conformations until the outer mitochondrial membrane is contacted [265]. In mammals, both the chaperones HSP90 and HSP70 bind to the newly synthesized preproteins in the cytosol to target them to the import receptor Tom70 at the outer mitochondrial membrane [266].

Also, in the bacterial type III protein secretion system (used by many bacterial pathogens to deliver virulence effector proteins directly into the host-cell cytosol), specific chaperones are required to maintain their substrates in a secretion-competent state [267].

In Gram-negative bacteria such as *E. coli*, an export-specific molecular chaperone, SecB, keeps preproteins destined to be posttranslationally translocated in a partially unfolded conformation, and pilots them to a membrane-associated receptor, SecA, which provides the link to the translocon complex. SecB (a non-essential protein)

differs from other chaperones in that it is a homotetramer and functions independently of nucleotide triphosphate hydrolysis [139]. Discrimination between the SecA/SecB-dependent targeting and the signal recognition particle (SRP)–dependent targeting to the (same) translocon complex (SecYE) involves the ribosome-associated chaperone TF (see Sect. 13.2.1), which interacts with sequences near the N-terminus of the mature regions of presecretory proteins and therefore occludes SRP binding [268a, 268b]. But other data argue against a role for TF in pathway discrimination [269]. Nevertheless, TF has a unique ability to sequester nascent polypeptides for a relatively prolonged period, and in this way retards protein export [144]. Other chaperones not specific for secretory proteins such as DnaK also maintain the transport competence of presecretory proteins [270]. The functional redundancy of chaperones in the protein export pathway may explain why SecB is not essential for cell viability. In Gram-negative bacteria, distinct molecular chaperones of the periplasm, such as Skp, SurA and PpiD, are involved in the biogenesis of outer membrane proteins [104].

13.7 New Horizons in Chaperone Research

13.7.1 HSP90 and the Pandora's Box of Hidden Mutations

Morphological mutations are stored without expressing phenotypes because they are "buffered" by the molecular chaperone HSP90 in *Drosophila melanogaster* [271] and in *Arabidopsis thaliana* [272]. HSP90 functions as a chaperone for mutated abnormal proteins to enable them to "behave" functionally like normal proteins. Decreasing the levels of HSP90, whether by mutation or by an HSP90 inhibitor, uncovers a Pandora's box of developmental abnormalities [257]. HSP90 therefore appears as a capacitor of morphological evolution and phenotypic variation. Interestingly, the yeast prion ψ+ also provides a mechanism, although totally different, for genetic variation and phenotypic diversity [273]. Also in bacteria, GroEL buffers against deleterious mutations [274].

13.7.2 Chaperones and Prions

By analogy with the mode of transmission of the abnormal prion protein, the causative agent of neurodegenerative diseases [275], two self-propagating proteins, ψ+ and [URE3], abnormal isoforms of SUP35 and URE2 respectively, have been characterized and named prions, in the yeast *S. cerevisiae* [276]. The importance of yeast chaperones HSP70 and HSP104 in the propagation and maintenance of ψ+ [277], their antagonistic interactions in prion curing [169] is a promising research topic. The conformational conversion of the normal prion protein into its abnormal infectious isoform could therefore involve a chaperone-like factor [278].

13.7.3
Chaperones and Ribosome Biogenesis

DnaK is necessary for the late stages of ribosome assembly in *E. coli* at high temperature (above 37°C), but not at 30°C. Several observations [279] suggest that an extra-ribosomal factor, naturally thermosensitive and therefore DnaK-dependent *only* at high temperature, would mediate ribosome assembly, rather than DnaK itself. The benefit for the bacterial cell would be to have, through DnaK, a regulatory link between the late stages of ribosome biogenesis and the protein quality control. The titration of DnaK by the thermodenatured proteins during thermal stress would both mediate the heat-shock response (see Sect. 13.4.1) and prevent the late steps of ribosome biogenesis, providing an additional control level of ribosome biogenesis (for the other well-known control mechanisms, see Fig. 13.15). In other words, DnaK would play a role of thermometer in the late stages of ribosome biogenesis [193]. Such a link between the function of mitochondrial chaperones (HSP70–HSP78) and the naturally thermosensitive mitochondrial DNA synthesis has been shown in the yeast *S. cerevisiae* [280].

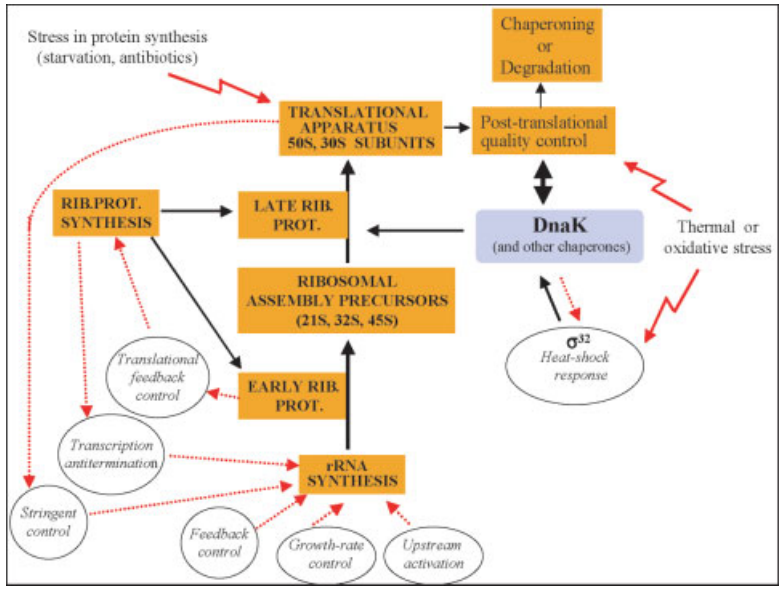


Figure 13.15 An additional level of control of ribosome bio-genesis in *E. coli*: the DnaK-assisted late steps of ribosome assembly at high temperature. Biosynthetic pathways are symbolized by rectangles and full arrows; control mechanisms by circles and dotted arrows. Ribosomal RNA synthesis is known to be regulated by different mechanisms [302], such as transcription antitermination [303], stringent control, feedback control, growthrate control and upstream activation. Ribosomal protein synthesis is translationally auto-controlled [304].
A possible mechanism for the DnaK-assisted ribosome bio-genesis at high temperature is described in Sect. 13.7.3.

13.7.4 RNA Chaperones

There is a growing body of evidence for RNA recognition and binding by molecular chaperones [281]. The chaperonin of the archeon *Sulfolobus solfataricus* is an RNA-binding protein that participates in ribosomal RNA processing [282]. The existence of RNA chaperones that resolve misfolded RNA structures *in vivo* has been shown in many cases [283, 284].

13.7.5 Chemical Chaperones

Some low-molecular-weight compounds are able to stabilize the conformation of proteins that are defective in patients of inherited diseases. Therefore, "chemical-chaperone" is a new concept in drug research [285]. For example, some chemical chaperones increase the activity of N370S β-glucosidase, the most common mutation causing Gaucher disease [286]. A designed peptide is able to rescue mutants of the tumor suppressor p53 in cancer cells (chaperoning strategy) [287].

The osmolyte trimethylamine-oxide stimulates the *in vitro* reconstitution of functional 50S ribosomes [288] and decreases the *E. coli* homoserine *trans*-succinylase aggregation at high temperature [289]. Some hydrogel nanoparticles (nanogels) assist protein refolding in a manner similar to the mechanism of molecular chaperones, namely by catching and releasing proteins [290].

13.7.6 Medical implications

The therapeutic applications of heat shock/stress proteins and chaperone inducers for medicine constitute a promising field [291a, 291b], particularly as far as neurodegenerative disorders are concerned [186, 292]. Two HSP70 family members are expressed in atherosclerotic lesions [293]. HSP70s have also been implicated in many pathways in immunology (presentation of the antigen [294a, 294b] and in molecular cancerology [295a, 31]. For example, when secreted by viable immunocompetent cells, HSP70 in the extracellular milieu acts as a powerful cytokine [295b].

13.7.7 Chaperoning the chaperones

But, and this is an unavoidable final question, who chaperones the chaperones? Is it a self-assembly process? [296, 297a]. There is no doubt that the study of molecular chaperones will continue to surprise us. A recent example [297b] shows that it is even possible to mutationally improve the efficiency of a chaperonin, as soon as an efficient method of selection can be devised!

Table 13.1 Major chaperone families

In prokaryotes	In eukaryotes	Function
DnaK	HSP70, HSC70	*De novo* protein folding.
Co-chaperones: DnaJ and GrpE (DnaK homologs =HscA and HscC)	Co-chaperones : HSP40, BAG-1, HIP, HOP, CHIP	Protein refolding; prevention and reversion of aggregation
		Protein quality contol, in cooperation with proteases
GroEL Co-chaperonin : GroES	HSP60 (mitochondria) =Cpn60 (chloroplasts) co-chaperonin: HSP10 = Cpn10 (same localization)	Protein translocation accross membranes Oligomerization; assemblies and dissasem-blies.
Trigger factor		Modulation of the heat-shock response
	TCP-1, CCT/TriC NAC GimC = prefoldin	Protein folding (in particular actin and tubulin)
HtpG	HSP9O	Signal transduction proteins: steroid hormone receptors, signaling kinases.
ClpB	HSP100 / HSP104	Thermotolerance; protein desaggregation
ClpA, ClpX (without ClpP)		Disassembly of quaternary structures
ClpA, ClpX (with ClpP) HslU (with HslV = ClpQ) = ClpY (without HslV = ClpQ)	Proteasome	Proteolysis
		Prevention of aggregation
IbpA, IbpB	HSP26, α-crystallin and other small HSPs	Protein folding and thermoprotection
SecB		Protein export via the general secretory pathway
	Nucleoplasmin	Assembly of nucleosomes
	Calnexin Calreticulin	Folding of glycosylated proteins in the ER

References

1 C. B. Anfinsen, *Science* **1973,** *181,* 223–230.
2 C. M. Dobson, *Curr. Biol.* **1994**, *4,* 636–640.
3 V. Daggett, A. R. Fersht, *Trends Biochem. Sci.* **2003,** *28,* 18–25.
4a A. R. Dinner, A. Sali, L. J. Smith et al., *Trends Biochem. Sci.* **2000,** *25,* 331–339.
4b S. B. Zimmerman, S. O. Trach, *J. Mol. Biol.* **1991**, *222,* 599–620.
5 R. A. Broglia, G. Tiana, S. Pasquali et al., *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 12930–12933.
6 R. J. Ellis, *Curr. Biol.* **2001,** *11,* R1038–1040.
7 M. Shtilerman, G. H. Lorimer, S. W. Englander, *Science* **1999,** *284,* 822–825.
8 S. V. Slepenkov, S. N. Witt, *Mol. Microbiol.* **2002,** *45,* 1197–1206.
9 S. Gribaldo, V. Lumia, R. Creti et al., *J. Bacteriol.* **1999,** *181,* 434–443.
10 W. R. Boorstein, T. Ziegelhoffer, E. A. Craig, *J. Mol. Evol.* **1994,** *38,* 1–17.

11 S. A. Rensing, U. G. Maier, *J. Mol. Evol.* **1994,** *39,* 80–86.
12 C. Borchiellini, N. Boury-Esnault, J. Vacelet et al., *Mol. Biol. Evol.* **1998,** *15,* 647–655.
13 R. S. Gupta, K. Bustard, M. Falah et al., *J. Bacteriol.* **1997,** *179,* 345–357.
14 E. A. Snell, R. F. Furlong, P. W. Holland, *Curr. Biol.* **2001,** *11,* 967–970.
15 R. S. Gupta, B. Singh, *Curr. Biol.* **1994,** *4,* 1104–1114.
16 F. U. Hartl, M. Hayer-Hartl, *Science* **2002,** *295,* 1852–1858.
17 W. A. Houry, Curr. *Protein Pept. Sci.* **2001,** *2,* 227–244.
18 P. C. Stirling, V. F. Lundin, M. R. Leroux, *EMBO Rep.* **2003,** *4,* 565–570.
19a J. C. Young, J. M. Barral, F. Ulrich Hartl, *Trends Biochem. Sci.* **2003,** *28,* 541–547.
19b W. R. Lyon, M. G. Caparon, *J. Bacteriol.* **2003,** *185,* 3661–3667.
20a O. Kandror, A. L. Goldberg, *Proc. Natl. Acad. Sci. USA* **1997,** *94,* 4978–4981.
20b P. Genevaux, F. Keppel, F. Schwager et al., *EMBO reports* **2004,** *5,* 195–200.
20c S. Vorderwülbecke, G. Kramer, F. Merz et al., *FEBS letters* **2004,** *559,* 181–187.
21 G. Blaha, D. N. Wilson, G. Stoller et al., *J. Mol. Biol.* **2003,** *326,* 887–897.
22 G. Kramer, T. Rauch, W. Rist et al., *Nature* **2002,** *419,* 171–174.
23 R. Maier, B. Eckert, C. Scholz et al., *J. Mol. Biol.* **2003,** *326,* 585–592.
24 R. Maier, C. Scholz, F. X. Schmid, *J. Mol. Biol.* **2001,** *314,* 1181–1190.
25 H. Patzelt, S. Rudiger, D. Brehmer et al., *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 14244–14249.
26 H. Patzelt, G. Kramer, T. Rauch et al., *Biol. Chem.* **2002,** *383,* 1611–1619.
27 R. C. Morshauser, W. Hu, H. Wang et al., *J. Mol. Biol.* **1999,** *289,* 1387–1403.
28a C. C. Chou, F. Forouhar, Y. H. Yeh et al., *J. Biol. Chem.* **2003,** *278,* 30311–30316.
28b F. Moro, V. Fernandez-Saiz, A. Muga, *J. Biol. Chem.* **2004,** *279,* 19600–19606.
29 S. Rüdiger, A. Buchberger, B. Bukau, *Nat. Struct. Biol.* **1997,** *4,* 342–349.
30 M. J. Gething, S. Blond-Elguindi, J. Buchner et al., *Cold Spring Harb. Symp. Quant. Biol.* **1995,** *60,* 417–428.
31 M. Zylicz, F. W. King, A. Wawrzynow, *EMBO J.* **2001,** *20,* 4634–4638.
32 Y. K. Lee, J. W. Brewer, R. Hellman et al., *Mol. Biol. Cell* **1999,** *10,* 2209–2219.
33 G. Knarr, S. Modrow, A. Todd et al., *J. Biol. Chem.* **1999,** *274,* 29850–29857.
34 A. Buchberger, H. Theyssen, H. Schroder et al., *J. Biol. Chem.* **1995,** *270,* 16903–16910.
35 A. Buchberger, A. Valencia, R. McMacken et al., *EMBO J.* **1994,** *13,* 1687–1695.
36 K. L. Fung, L. Hilgenberg, N. M. Wang et al., *J. Biol. Chem.* **1996,** *271,* 21559–21565.
37 P. Lopez-Buesa, C. Pfund, E. A. Craig, *Proc. Natl. Acad. Sci. USA* **1998,** *95,* 15253–15258.
38 M. Mayer, J. Reinstein, J. Buchner, *J. Mol. Biol.* **2003,** *330,* 137–144.
39 S. Sadis, L. E. Hightower, *Biochemistry* **1992,** *31,* 9406–9412.
40 S. Takeda, D. B. McKay, *Biochemistry* **1996,** *35,* 4636–4644.
41a B. Feifel, E. Sandmeier, H. J. Schönfeld et al., *Eur. J. Biochem.* **1996,** *237,* 318–321.
41b C.E. Caldon, P.Yoong, P.E. March, *Mol.Microbiol.* **2001,** *41,* 289–297.
42 W. Han, P. Christen, *J. Biol. Chem.* **2003,** *278,* 19038–19043.
43 R. Russell, A. Wali Karzai, A. F. Mehl et al., *Biochemistry* **1999,** *38,* 4165–4176.
44 D. Brehmer, S. Rudiger, C. S. Gassler et al., *Nat. Struct. Biol.* **2001,** *8,* 427–432.
45 W. L. Kelley, *Curr. Biol.* **1999,** *9,* R305–308.
46 P. Genevaux, F. Schwager, C. Georgopoulos et al., *Genetics* **2002,** *162,* 1045–1053.

47 T. Laufen, M. P. Mayer, C. Beisel et al., *Proc. Natl. Acad. Sci. USA* **1999,** *96,* 5452–5457.

48 S. Rüdiger, J. Schneider-Mergener, B. Bukau, *EMBO J.* **2001,** *20,* 1042–1050.

49a E. V. Pierpaoli, E. Sandmeier, H. J. Schönfeld et al., *J. Biol. Chem.* **1998,** *273,* 6643–6649.

49b U. Gehring, *EMBO reports* **2004,** *5,* 148–153.

50 H. Sondermann, A. K. Ho, L. L. Listenberger et al., *J. Biol. Chem.* **2002,** *277,* 33220–33227.

51 H. Wegele, M. Haslbeck, J. Reinstein et al., *J. Biol. Chem.* **2003,** *278,* 25970–25976.

52 H. H. Kampinga, B. Kanon, F. A. Salomons et al., *Mol. Cell. Biol.* **2003,** *23,* 4948–4958.

53 J. Demand, J. Luders, J. Hohfeld, *Mol. Cell. Biol.* **1998,** *18,* 2023–2028.

54 C. S. Sullivan, J. M. Pipas, Microbiol. *Mol. Biol. Rev.* **2002,** *66,* 179–202.

55 S. A. Teter, W. A. Houry, D. Ang et al., *Cell* **1999,** *97,* 755–765.

56 T. Hesterkamp, B. Bukau, *EMBO J.* **1998,** *17,* 4818–4828.

57 A. Mogk, T. Tomoyasu, P. Goloubinoff et al., *EMBO J.* **1999,** *18,* 6934–6949.

58 E. Deuerling, H. Patzelt, S. Vorderwulbecke et al., *Mol. Microbiol.* **2003,** *47,* 1317–1328.

59 J. J. Silberg, K. G. Hoff, T. L. Tapley et al., *J. Biol. Chem.* **2001,** *276,* 1696–1700.

60 K. Yoshimune, T. Yoshimura, T. Nakayama et al., *Biochem. Biophys. Res. Commun.* **2002,** *293,* 1389–1395.

61 M. Arifuzzaman, T. Oshima, S. Nakade et al., *Genes Cells* **2002,** *7,* 553–566.

62 C. J. Kluck, H. Patzelt, P. Genevaux et al., *J. Biol. Chem.* **2002,** *277,* 41060–41069.

63 J. M. Archibald, J. M. Logsdon, Jr., W. F. Doolittle et al., *Mol. Biol. Evol.* **2000,** *17,* 1456–1466.

64 A. Y. Dunn, M. W. Melville, J. Frydman, *J. Struct. Biol.* **2001,** *135,* 176–184.

65 K. Siegers, T. Waldmann, M. R. Leroux et al., *EMBO. J.* **1999,** *18,* 75–84.

66 J. M. Valpuesta, J. Martin-Benito, P. Gomez-Puertas et al., *FEBS Lett.* **2002,** *529,* 11–16.

67a A. Camasses, A. Bogdanova, A. Shevchenko et al., *Mol. Cell* **2003,** *12,* 87–100.

67b M.W. Melville, A.J. McClellan, A.S. Meyer et al., *Mol.Cell.Biol.* **2003,** *23,* 3141–3151.

67c E.A. Craig, *Current Biology* **2003,** *13,* R904–R905.

67d M. Ferrer, T.N. Chernikova, M.M. Yakimov et al., *Nature Biotechnology* **2003,** *21,* 1266–1267.

68a G. Levy-Rimler, R. E. Bell, N. Ben-Tal et al., *FEBS Lett.* **2002,** *529,* 1–5.

68b J.J. Hansen, A. Durr, I. Cournu-Rebeix et al., *Am.J.Hum.Genet.* **2002,** *70,* 1328–1332.

69 O. Fayet, T. Ziegelhoffer, C. Georgopoulos, *J. Bacteriol.* **1989,** *171,* 1379–1385.

70 W. A. Houry, *Biochem. Cell. Biol.* **2001,** *79,* 569–577.

71 S. Walter, *Cell. Mol. Life Sci.* **2002,** *59,* 1589–1597.

72 Z. Sun, D. J. Scott, P. A. Lund, *J. Mol. Biol.* **2003,** *332,* 715–728.

73 T. Yoshida, R. Kawaguchi, H. Taguchi et al., *J. Mol. Biol.* **2002,** *315,* 73–85.

74 H. Taguchi, T. Ueno, H. Tadakuma et al., *Nat. Biotechnol.* **2001,** *19,* 861–865.

75 A. Baumketner, A. Jewett, J. E. Shea, *J. Mol. Biol.* **2003,** *332,* 701–713.

76 W. A. Houry, D. Frishman, C. Eckerskorn et al., *Nature* **1999,** *402,* 147–154.

77 A. Horovitz, Y. Fridmann, G. Kafri et al., *J. Struct. Biol.* **2001,** *135,* 104–114.

78 J. Wang, L. Chen, *J. Mol. Biol.* **2003,** *334,* 489–499.

79 K. L. Ewalt, J. P. Hendrick, W. A. Houry et al., *Cell* **1997,** *90,* 491–500.

80 G. H. Lorimer, *FASEB J.* **1996,** *10,* 5–9.

81 N. McLennan, M. Masters, *Nature* **1998,** *392,* 139.

82 J. C. Young, I. Moarefi, F. U. Hartl, *J. Cell Biol.* **2001**, *154*, 267–273.

83 J. C. Bardwell, E. A. Craig, *J. Bacteriol.* **1988**, *170*, 2977–2983.

84 N. Tanaka, H. Nakamoto, *FEBS Lett.* **1999**, *458*, 117–123.

85a D. F. Nathan, M. H. Vos, S. Lindquist, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12949–12956.

85b S. Zhu, J. Tytgat, *FASEB. J.* **2004**, *18*, 940–947.

86 D. Picard, *Cell. Mol. Life Sci.* **2002**, *59*, 1640–1648.

87a A. J. Caplan, S. Jackson, D. Smith, *EMBO Rep.* **2003**, *4*, 126–130.

87b A. Kamal, L. Thao, J. Sensintaffar et al., *Nature* **2003**, *425*, 407–410.

88 D. A. Dougan, A. Mogk, K. Zeth et al., *FEBS Lett.* **2002**, *529*, 6–10.

89 J. A. Kenniston, T. A. Baker, J. M. FeRNAndez et al., *Cell* **2003**, *114*, 511–520.

90 A. F. Neuwald, L. Aravind, J. L. Spouge et al., *Genome Res.* **1999**, *9*, 27–43.

91 J. M. Flynn, S. B. Neher, Y. I. Kim et al., *Mol. Cell* **2003**, *11*, 671–683.

92 A. L. Horwich, E. U. Weber-Ban, D. Finley, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11033–11040.

93 A. R. Kwon, B. M. Kessler, H. S. Overkleeft et al., *J. Mol. Biol.* **2003**, *330*, 185–195.

94 I. Levchenko, L. Luo, T. A. Baker, *Genes Dev.* **1995**, *9*, 2399–2408.

95 M. Pak, S. Wickner, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4901–4906.

96 E. U. Weber-Ban, B. G. Reid, A. D. Miranker et al., *Nature* **1999**, *401*, 90–93.

97 I. S. Seong, J. Y. Oh, J. W. Lee et al., *FEBS Lett.* **2000**, *477*, 224–229.

98 S. Lindquist, G. Kim, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5301–5306.

99 D. A. Hattendorf, S. L. Lindquist, *EMBO J.* **2002**, *21*, 12–21.

100a A. Mogk, C. Schlieker, C. Strub et al., *J. Biol. Chem.* **2003**, *278*, 17615–17624.

100b S. Lee, M.E. Sowa, Y. Watanabe et al., *Cell* **2003**, *115*, 229–240.

100c V. Akoev, E. P. Goel, M. E. Barnett et al., *Protein Science* **2004**, *13*, 567–574.

101 C. Spiess, A. Beil, M. Ehrmann, *Cell* **1999**, *97*, 339–347.

102 M. CastilloKeller, R. Misra, *J. Bacteriol.* **2003**, *185*, 148–154.

103 T. Krojer, M. Garrido-Franco, R. Huber et al., *Nature* **2002**, *416*, 455–459.

104 M. Müller, H. G. Koch, K. Beck et al., *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *66*, 107–157.

105 J. Werner, A. M. Augustus, R. Misra, *J. Bacteriol.* **2003**, *185*, 6540–6547.

106 A. E. Rizzitello, J. R. Harper, T. J. Silhavy, *J. Bacteriol.* **2001**, *183*, 6794–6800.

107 D. Missiakas, S. Raina, *J. Bacteriol.* **1997**, *179*, 2465–2471.

108 S. E. Ades, I. L. Grigorova, C. A. Gross, *J. Bacteriol.* **2003**, *185*, 2512–2519.

109 M. M. Barnhart, F. G. Sauer, J. S. Pinkner et al., *J. Bacteriol.* **2003**, *185*, 2723–2730.

110 M. M. Barnhart, J. S. Pinkner, G. E. Soto et al., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7709–7714.

111 J. Graw, *Biol. Chem.* **1997**, *378*, 1331–1348.

112 F. Narberhaus, *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 64–93.

113 M. Haslbeck, *Cell. Mol. Life Sci.* **2002**, *59*, 1649–1657.

114 R. Quinlan, P. Van Den Ijssel, *Nat. Med.* **1999**, *5*, 25–26.

115 J. I. Clark, P. J. Muchowski, *Curr. Opin. Struct. Biol.* **2000**, *10*, 52–59.

116a P. Vicart, A. Caron, P. Guicheney et al., *Nat. Genet.* **1998**, *20*, 92–95.

116b M.P. Bova, O. Yaron, Q. Huang et al., *Proc.Natl.Acad.Sci.USA* **1999**, *96*, 6137–6142.

116c M.D. Perng, P.J. Muchowski, P.van den Ijssel et al., *J.Biol.Chem.* **1999**, *274*, 33235–33243.

116d A.T. Chavez Zobel, A. Loranger, N.Marceau et al., *Human Molecular Genetics* **2003**, *12*, 1609–1620.

117a D. L. Stone, A. Slavotinek, G. G. Bouffard et al., *Nat. Genet.* **2000**, *25*, 79–82.

117b J. Irobi, K. Van Impe, P. Seeman et al., *Nature Genet 2004*, advance online publication, 2 May 2004.

117c O.V. Evgrafov, I. Mersiyanova, J. Irobi et al., *Nature Genet. 2004,* advance online publication, 2 May 2004.

118 P. Korber, J. M. Stahl, K. H. Nierhaus et al., *EMBO J.* **2000,** *19,* 741–748.

119a P. C. Graf, U. Jakob, *Cell. Mol. Life Sci.* **2002,** *59,* 1624–1631.

119b J.H. Hoffmann, K. Linke, P.C. Graf et al., *EMBO J.* **2004,** *23,* 160–168.

120a P. M. Quigley, K. Korotkov, F. Baneyx et al., *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 3137–3142.

120b M.S.R. Sastry, K. Korotkov, Y. Brodsky et al., *J.Biol.Chem.* **2002,** *277,* 46026–46034.

120c A. Malki, R. Kern, J. Abdallah et al., *Biochem.Biophys.Res. Commun.* **2003,** *301,* 430–436.

120d M. Mujacic, M.W. Bader, F. Baneyx, *Mol. Microbiol.* **2004,** *51,* 849–859.

120e P. M. Quigley, K. Korotkov, F. Baneyx et al., *Protein Science* **2004,** *13,* 269–277.

120f M. S. R. Sastry, P. M. Quigley, W. G. J. Hol et al., *Proc. Natl. Acad. Sci. USA* **2004,** *101,* 8587–8592.

121 M. Kitagawa, M. Miyakawa, Y. Matsumura et al., *Eur. J. Biochem.* **2002,** *269,* 2907–2917.

122 D. Kuczynska-Wisnik, S. Kedzierska, E. Matuszewska et al., *Microbiology* **2002,** *148,* 1757–1765.

123 A. Mogk, E. Deuerling, S. Vorderwulbecke et al., *Mol. Microbiol.* **2003,** *50,* 585–595.

124 T. Tomoyasu, A. Takaya, T. Sasaki et al., *J. Bacteriol.* **2003,** *185,* 6331–6339.

125 G. J. Lee, E. Vierling, *Plant Physiol.* **2000,** *122,* 189–197.

126a P. Smykal, I. Hrdy, P. M. Pechan, *Eur. J. Biochem.* **2000,** *267,* 2195–2207.

126b W. Wang, B. Vinocur, O. Shoseyov et al., *Trends in Plant Science* **2004,** *9,* 244–252.

127 A. Mogk, C. Schlieker, K. L. Friedrich et al., *J. Biol. Chem.* **2003,** *278,* 31033–31042.

128 M. R. Leach, M. F. Cohen-Doyle, D. Y. Thomas et al., *J. Biol. Chem.* **2002,** *277,* 29686–29697.

129a A. J. Parodi, *Annu. Rev. Biochem.* **2000,** *69,* 69–93.

129b M. Molinari, K.K. Eriksson, V. Calanca et al., *Molecular Cell* **2004,** *13,* 125–135.

130 Y. Yabuta, H. Takagi, M. Inouye et al., *J. Biol. Chem.* **2001,** *276,* 44427–44434.

131 D. C. Oliver, G. Huang, E. Nodel et al., *Mol. Microbiol.* **2003,** *47,* 1367–1383.

132 R. J. Ellis, *Curr. Biol.* **1999,** *9,* R137–139.

133 N. Heyrovska, J. Frydman, J. Hohfeld et al., *Biol. Chem.* **1998,** *379,* 301–309.

134 V. Thulasiraman, C. F. Yang, J. Frydman, *EMBO. J.* **1999,** *18,* 85–95.

135 J. Frydman, *Annu. Rev. Biochem.* **2001,** *70,* 603–647.

136 C. Pfund, P. Huang, N. Lopez-Hoyo et al., *Mol. Biol. Cell* **2001,** *12,* 3773–3782.

137 C. Pfund, N. Lopez-Hoyo, T. Ziegelhoffer et al., *EMBO J.* **1998,** *17,* 3981–3989.

138 M. Gautschi, A. Mun, S. Ross et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 4209–4214.

139 L. L. Randall, S. J. Hardy, *Cell. Mol. Life Sci.* **2002,** *59,* 1617–1623.

140a S. Wang, H. Sakai, M. Wiedmann, *J. Cell. Biol.* **1995,** *130,* 519–528.

140b L. Meunier, Y.K. Usherwood, K.T. Chung et al., *Mol. Biol. Cell* **2002,** *13,* 4456–4469.

140c V.R. Agashe, S. Guha, H.C. Chang et al., *Cell* **2004,** *117,* 199–209.

141 B. Bukau, E. Deuerling, C. Pfund et al., *Cell* **2000,** *101,* 119–122.

142 V. Albanèse, J. Frydman, *Nat. Struct. Biol.* **2002,** *9,* 716–718.

143 E. Deuerling, A. Schulze-Specking, T. Tomoyasu et al., *Nature* **1999,** *400,* 693–696.

144 H. C. Lee, H. D. Bernstein, *J. Biol. Chem.* **2002,** *277,* 43527–43535.

145 E. Schaffitzel, S. Rudiger, B. Bukau et al., *Biol. Chem.* **2001,** *382,* 1235–1243.

146 C. Schiene-Fischer, J. Habazettl, T. Tradler et al., *Biol. Chem.* **2002,** *383,* 1865–1873.

147 G. Kramer, V. Ramachandiran, P. M. Horowitz et al., *Arch. Biochem. Biophys.* **2002,** *403,* 63–70.

148 H. M. Dionisi, S. K. Checa, A. R. Krapp et al., *Eur. J. Biochem.* **1998**, *251*, 724–728.

149 G. A. Gaitanaris, A. Vysokanov, S. C. Hung et al., *Mol. Microbiol.* **1994**, *14*, 861–869.

150 A. Gragerov, E. Nudler, N. Komissarova et al., *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10341–10344.

151 K. Nishihara, M. Kanemori, M. Kitagawa et al., *Appl. Environ. Microbiol.* **1998**, *64*, 1694–1699.

152 M. A. Petit, W. Bedale, J. Osipiuk et al., *J. Biol. Chem.* **1994**, *269*, 23824–23829.

153 A. El Hage, M. Sbai, J. H. Alix, *Mol. Gen. Genet.* **2001**, *264*, 796–808.

154 A. Buchberger, H. Schröder, T. Hesterkamp et al., *J. Mol. Biol.* **1996**, *261*, 328–333.

155 J. G. Thomas, F. Baneyx, *Mol. Microbiol.* **2000**, *36*, 1360–1370.

156 W. J. Netzer, F. U. Hartl, *Nature* **1997**, *388*, 343–349.

157 A. V. Nicola, W. Chen, A. Helenius, *Nat. Cell. Biol.* **1999**, *1*, 341–345.

158 B. Hardesty, G. Kramer, *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *66*, 41–66.

159 E. A. Craig, H. C. Eisenman, H. A. Hundley, *Curr. Opin. Microbiol.* **2003**, *6*, 157–162.

160 M. E. Butkus, L. B. Prundeanu, D. B. Oliver, *J. Bacteriol.* **2003**, *185*, 6719–6722.

161 F. Gong, C., Y., *Science* **2002**, *297*, 1864–1867.

162 H. Nakatogawa, K. Ito, *Cell* **2002**, *108*, 629–636.

163 J. Ghosh, A. Basu, S. Pal et al., *Mol. Microbiol.* **2003**, *48*, 1679–1692.

164 J. R. Glover, S. Lindquist, *Cell* **1998**, *94*, 73–82.

165 Y. Sanchez, D. A. Parsell, J. Taulien et al., *J. Bacteriol.* **1993**, *175*, 6484–6491.

166 J. L. Vogel, D. A. Parsell, S. Lindquist, *Curr. Biol.* **1995**, *5*, 306–317.

167 M. Moczko, B. Schonfisch, W. Voos et al., *J. Mol. Biol.* **1995**, *254*, 538–543.

168 M. Schmitt, W. Neupert, T. Langer, *EMBO J.* **1995**, *14*, 3434–3444.

169 G. P. Newnam, R. D. Wegrzyn, S. L. Lindquist et al., *Mol. Cell. Biol.* **1999**, *19*, 1325–1333.

170 K. Motohashi, Y. Watanabe, M. Yohda et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 7184–7189.

171 S. Schlee, J. Reinstein, *Cell. Mol. Life Sci.* **2002**, *59*, 1598–1606.

172 S. Diamant, A. P. Ben-Zvi, B. Bukau et al., *J. Biol. Chem.* **2000**, *275*, 21107–21113.

173 S. Diamant, D. Rosenthal, A. Azem et al., *Mol. Microbiol.* **2003**, *49*, 401–410.

174 P. Goloubinoff, A. Mogk, A. P. Zvi et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13732–13737.

175 S. Kedzierska, E. Matuszewska, *FEMS Microbiol. Lett.* **2001**, *204*, 355–360.

176 Z. Liu, V. Tek, V. Akoev et al., *J. Mol. Biol.* **2002**, *321*, 111–120.

177 M. Zolkiewski, *J. Biol. Chem.* **1999**, *274*, 28083–28086.

178 J. Weibezahn, C. Schlieker, B. Bukau et al., *J. Biol. Chem.* **2003**, *278*, 32608–32617.

179 A. P. Ben-Zvi, P. Goloubinoff, *J. Struct. Biol.* **2001**, *135*, 84–93.

180a R. R. Kopito, D. Ron, *Nat. Cell. Biol.* **2000**, *2*, E207–E209.

180b J. N. Buxbaum, *Trends Biochem. Sci.* **2003**, *28*, 585–592.

181a S. Iuchi, G. Hoffner, P. Verbeke et al., *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2409–2414.

181b A. Wyttenbach, O. Sauvageot, J. Carmichael et al., *Human Molecular Genetics* **2002**, *11*, 1137–1151.

182 P. K. Auluck, H. Y. Chan, J. Q. Trojanowski et al., *Science* **2002**, *295*, 865–868.

183 Y. Chai, S. L. Koppenhafer, N. M. Bonini et al., *J. Neurosci.* **1999**, *19*, 10338–10347.

184 F. Dou, W. J. Netzer, K. Tanemura et al., *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 721–726.

185 S. Krobitsch, S. Lindquist, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1589–1594.

186 P. J. Muchowski, *Neuron* **2002**, *35*, 9–12.

187 I. Konieczny, K. Liberek, *J. Biol. Chem.* **2002,** *277,* 18483–18488.
188 U. Schubert, L. C. Anton, J. Gibbs et al., *Nature* **2000,** *404,* 770–774.
189 R. Hengge, B. Bukau, *Mol. Microbiol.* **2003,** *49,* 1451–1462.
190 D. A. Dougan, A. Mogk, B. Bukau, *Cell. Mol. Life Sci.* **2002,** *59,* 1607–1616.
191 S. Wickner, M. R. Maurizi, S. Gottesman, *Science* **1999,** *286,* 1888–1893.
192 T. Tomoyasu, A. Mogk, H. Langen et al., *Mol. Microbiol.* **2001,** *40,* 397–413.
193 A. El Hage, J. H. Alix, *Mol. Microbiol.* **2004,** *51,* 189–201.
194 D. M. Cyr, J. Höhfeld, C. Patterson, *Trends Biochem. Sci.* **2002,** *27,* 368–375.
195 J. Höhfeld, D. M. Cyr, C. Patterson, *EMBO Rep.* **2001,** *2,* 885–890.
196 S. Friant, K. D. Meier, H. Riezman, *EMBO J.* **2003,** *22,* 3783–3791.
197 M. Valle, R. Gillet, S. Kaur et al., *Science* **2003,** *300,* 127–130.
198 C. S. Hayes, R. T. Sauer, *Mol. Cell* **2003,** *12,* 903–911.
199 R. I. Morimoto, *Genes Dev.* **1998,** *12,* 3788–3796.
200 K. L. Milarski, R. I. Morimoto, *J. Cell. Biol.* **1989,** *109,* 1947–1962.
201 Z. S. Chughtai, R. Rassadi, N. Matusiewicz et al., *J. Biol. Chem.* **2001,** *276,* 20261–20266.
202 C. Delelis, C. N. Angelier, M. Penrad-Mobayed, *Exp. Cell Res.* **2000,** *260,* 222–232.
203 F. Narberhaus, S. Balsiger, *J. Bacteriol.* **2003,** *185,* 2731–2738.
204 T. Tatsuta, D. M. Joob, R. Calendar et al., *FEBS Lett.* **2000,** *478,* 271–275.
205 F. Arsene, T. Tomoyasu, B. Bukau, *Int. J. Food Microbiol.* **2000,** *55,* 3–9.
206 M. T. Morita, M. Kanemori, H. Yanagi et al., *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 5860–5865.
207 A. Chastanet, J. Fert, T. Msadek, *Mol. Microbiol.* **2003,** *47,* 1061–1073.
208 J. D. Trent, M. Gabrielsen, B. Jensen et al., *J. Bacteriol.* **1994,** *176,* 6148–6152.
209 K. A. Ulmasov, S. Shammakov, K. Karaev et al., *Proc. Natl. Acad. Sci. USA* **1992,** *89,* 1666–1670.
210 B. J. Smith, M. P. Yaffe, *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 11091–11094.
211 R. A. VanBogelen, M. A. Acton, F. C. Neidhardt, *Genes Dev.* **1987,** *1,* 525–531.
212 G. C. Li, L. Li, R. Y. Liu et al., *Proc. Natl. Acad. Sci. USA* **1992,** *89,* 2036–2040.
213 L. Li, G. Shen, G. C. Li, *Exp. Cell Res.* **1995,** *217,* 460–468.
214 E. A. Nollen, J. F. Brunsting, H. Roelofsen et al., *Mol. Cell. Biol.* **1999,** *19,* 2069–2079.
215 J. M. Delaney, *J. Gen. Microbiol.* **1990,** *136 (Pt 10),* 2113–2118.
216 P. Echave, M. A. Esparza-Cerón, E. Cabiscol et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 4626–4631.
217 D. Rockabrand, T. Arthur, G. Korinek et al., *J. Bacteriol.* **1995,** *177,* 3695–3703.
218 A. Muffler, M. Barth, C. Marschall et al., *J. Bacteriol.* **1997,** *179,* 445–452.
219 D. Rockabrand, K. Livers, T. Austin et al., *J. Bacteriol.* **1998,** *180,* 846–854.
220 R. Hurme, M. Rhen, *Mol. Microbiol.* **1998,** *30,* 1–6.
221 A. Spicher, O. M. Guicherit, L. Duret et al., *Mol. Cell. Biol.* **1998,** *18,* 7371–7382.
222 A. Nocker, T. Hausherr, S. Balsiger et al., *Nucleic Acids Res.* **2001,** *29,* 4800–4807.
223 J. Johansson, P. Mandin, A. Renzoni et al., *Cell* **2002,** *110,* 551–561.
224 G. Storz, *Genes Dev.* **1999,** *13,* 633–636.
225 F. Repoila, S. Gottesman, *J. Bacteriol.* **2001,** *183,* 4012–4023.
226 F. Repoila, S. Gottesman, *J. Bacteriol.* **2003,** *185,* 6609–6614.
227 R. Hurme, K. D. Berndt, S. J. Normark et al., *Cell* **1997,** *90,* 55–64.
228a P. Servant, C. Grandvalet, P. Mazodier, *Proc. Natl. Acad. Sci. USA* **2000,** *97,* 3538–3543.

228b M. Hitomi, H. Nishimura, Y. Tsujimoto et al., *J.Bacteriol.* **2003,** *185,* 381–385.

229 M. Haslbeck, S. Walke, T. Stromer et al., *EMBO J.* **1999,** *18,* 6744–6751.

230 E. A. Craig, C. A. Gross, *Trends Biochem. Sci.* **1991,** *16,* 135–140.

231 J. S. McCarty, G. C. Walker, *Proc. Natl. Acad. Sci. USA* **1991,** *88,* 9513–9517.

232 A. D. Gelinas, J. Toth, K. A. Bethoney et al., *Biochemistry* **2003,** *42,* 9050–9059.

233 J. P. Grimshaw, I. Jelesarov, R. K. Siegenthaler et al., *J. Biol. Chem.* **2003,** *278,* 19048–19053.

234 J. Song, M. Takeda, R. I. Morimoto, *Nat. Cell Biol.* **2001,** *3,* 276–282.

235 E. M. Newton, U. Knauf, M. Green et al., *Mol. Cell. Biol.* **1996,** *16,* 839–846.

236 S. Banuelos, A. Hierro, J. M. Arizmendi et al., *J. Mol. Biol.* **2003,** *334,* 585–593.

237 H. Murialdo, *Virology* **1979,** *96,* 341–367.

238 R. Grimaud, A. Toussaint, *J. Bacteriol.* **1998,** *180,* 1148–1153.

239 P. Goloubinoff, A. A. Gatenby, G. H. Lorimer, *Nature* **1989,** *337,* 44–47.

240 M. W. Ribbe, B. K. Burgess, *Proc. Natl. Acad. Sci. USA* **2001,** *98,* 5521–5525.

241a R. M. Wynn, J. L. Song, D. T. Chuang, *J. Biol. Chem.* **2000,** *275,* 2786–2794.

241b J.K. Tyler, *Eur.J.Biochem.* **2002,** *269,* 2268–2274.

241c P. Liang, T.H. MacRae, *J.Cell Science* **1997,** *110,* 1431–1440.

242 M. Zyclicz, et al. :Role of Chaperones in Replication of Bacteriophage λ DNA, in *Molecular Chaperones and Folding Catalysts: Regulation, Cellular Functions and Mechanisms,* ed. B. Bukau, Harwood Academic Publishers Amsterdam 2000, 295–311.

243 O. Sand, L. Desmet, A. Toussaint et al., *Mol. Microbiol.* **1995,** *15,* 977–984.

244 F. Matsunaga, M. Ishiai, G. Kobayashi et al., *J. Mol. Biol.* **1997,** *274,* 27–38.

245 S. Y. Kim, S. Sharma, J. R. Hoskins et al., *J. Biol. Chem.* **2002,** *277,* 44778–44783.

246 S. L. Newmyer, A. Christensen, S. Sever, *Dev. Cell* **2003,** *4,* 929–940.

247 L. R. Chromy, J. M. Pipas, R. L. Garcea, *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 10477–10482.

248 J. A. Maki, D. R. Southworth, G. M. Culver, *RNA* **2003,** *9,* 1418–1421.

249 N. Lopes Ferreira, J. H. Alix, *J. Bacteriol.* **2002,** *184,* 7047–7054.

250 T. Abbas-Terki, D. Picard, *Eur. J. Biochem.* **1999,** *266,* 517–523.

251 B. C. Freeman, R. I. Morimoto, *EMBO J.* **1996,** *15,* 2969–2979.

252 Y. P. Di, E. Repasky, A. Laszlo et al., *J. Cell Physiol.* **1995,** *165,* 228–238.

253 D. S. Hwang, E. Crooke, A. Kornberg, *J. Biol. Chem.* **1990,** *265,* 19244–19248.

254 H. L. Forsythe, J. L. Jarvis, J. W. Turner et al., *J. Biol. Chem.* **2001,** *276,* 15571–15574.

255 J. Hu, D. O. Toft, C. Seeger, *EMBO J.* **1997,** *16,* 59–68.

256 S. Uma, V. Thulasiraman, R. L. Matts, *Mol. Cell. Biol.* **1999,** *19,* 5861–5871.

257 R. I. Morimoto, *Cell* **2002,** *110,* 281–284.

258 H. M. Beere, B. B. Wolf, K. Cain et al., *Nat. Cell. Biol.* **2000,** *2,* 469–475.

259 A. Saleh, S. M. Srinivasula, L. Balkir et al., *Nat. Cell. Biol.* **2000,** *2,* 476–483.

260 A. H. Hobson, C. M. Buckley, J. L. Aamand et al., *Proc. Natl. Acad. Sci. USA* **1993,** *90,* 5682–5686.

261 J. M. Barral, A. H. Hutagalung, A. Brinker et al., *Science* **2002,** *295,* 669–671.

262 M. G. Price, M. L. Landsverk, J. M. Barral et al., *J. Cell Sci.* **2002,** *115,* 4013–4023.

263 L. Thöny-Meyer, *Biochemistry* **2003,** *42,* 13099–13105.

264 C. A. Kumamoto, *Mol. Microbiol.* **1991,** *5,* 19–22.

265 T. Beddoe, T. Lithgow, *Biochim. Biophys. Acta* **2002,** *1592,* 35–39.

266 J. C. Young, N. J. Hoogenraad, F. U. Hartl, *Cell* **2003,** *112,* 41–50.

267 C. E. Stebbins, J. E. Galan, *Nature* **2001,** *414,* 77–81.

268a K. Beck, L. F. Wu, J. Brunner et al., *EMBO J.* **2000,** *19,* 134–143.

268b I. Buskiewicz, E. Deuerling, S. Q. Gu et al., *Proc. Natl. Acad. Sci. USA* **2004**, *101,* 7902–7906.

269 C. W. Bowers, F. Lau, T. J. Silhavy, *J. Bacteriol.* **2003,** *185,* 5697–5705.

270 H. Y. Qi, J. B. Hyndman, H. D. Bernstein, *J. Biol. Chem.* **2002,** *277,* 51077–51083.

271 S. L. Rutherford, S. Lindquist, *Nature* **1998,** *396,* 336–342.

272 C. Queitsch, T. A. Sangster, S. Lindquist, *Nature* **2002,** *417,* 618–624.

273 H. L. True, S. L. Lindquist, *Nature* **2000,** *407,* 477–483.

274 M. A. Fares, M. X. Ruiz-Gonzalez, A. Moya et al., *Nature* **2002,** *417,* 398.

275 D. Dormont, *FEBS Lett.* **2002,** *529,* 17–21.

276 E. FeRNAndez-Bellot, C. Cullin, *Cell. Mol. Life Sci.* **2001,** *58,* 1857–1878.

277 S. Narayanan, B. Bosl, S. Walter et al., *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 9286–9291.

278 J. Stöckel, F. U. Hartl, *J. Mol. Biol.* **2001,** *313,* 861–872.

279 J. H. Alix, K. H. Nierhaus, *RNA* **2003,** *9,* 787–793.

280 A. Germaniuk, K. Liberek, J. Marszalek, *J. Biol. Chem.* **2002,** *277,* 27801–27808.

281 T. Henics, *Cell Biol Int,* **2003,** *27,* 1–6.

282 D. Ruggero, A. Ciammaruconi, P. Londei, *EMBO J.* **1998,** *17,* 3471–3477.

283 G. Cristofari, J. L. Darlix, *Prog. Nucleic Acid Res. Mol. Biol.* **2002,** *72,* 223–268.

284 J. R. Lorsch, *Cell* **2002,** *109,* 797–800.

285 T. Kolter, M. Wendeler, *Chembiochem* **2003,** *4,* 260–264.

286 A. R. Sawkar, W. C. Cheng, E. Beutler et al., *Proc. Natl. Acad. Sci. USA* **2002,** *99,* 15428–15433.

287 N. Issaeva, A. Friedler, P. Bozko et al., *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 13303–13307.

288 K. Semrad, R. Green, *RNA* **2002,** *8,* 401–411.

289 E. Gur, D. Biran, E. Gazit et al., *Mol. Microbiol.* **2002,** *46,* 1391–1397.

290 Y. Nomura, M. Ikeda, N. Yamaguchi et al., *FEBS Lett.* **2003,** *553,* 271–276.

291a K. Rokutan, T. Hirakawa, S. Teshima et al., *J. Med. Invest.* **1998,** *44,* 137–147.

291b D. F. Smith, L. Whitesell, E. Katsanis, *Pharmacological Reviews* **1998**, *50,* 493–513.

292 F. R. Sharp, S. M. Massa, R. A. Swanson, *Trends Neurosci.* **1999,** *22,* 97–99.

293 Z. Han, Q. A. Truong, S. Park et al., *Proc. Natl. Acad. Sci. USA* **2003,** *100,* 1256–1261.

294a R. P. Wallin, A. Lundqvist, S. H. More et al., *Trends Immunol.* **2002,** *23,* 130–135.

294b L. Zihai, A. Menoret, P. Srivastava, *Curr. Opin. Immunol.* **2002,** *14,* 45–51

295a S. Xanthoudakis, D. W. Nicholson, *Nat. Cell. Biol.* **2000,** *2,* E163–165.

295b A. Asea et al., *Nat. Med.* **2000,** *6,* 435–442.

296 M. Y. Cheng, F. U. Hartl, A. L. Horwich, *Nature* **1990,** *348,* 455–458.

297a N. M. Lissin, S. Venyaminov, A. S. Girshovich, *Nature* **1990,** *348,* 339–342.

297b J.D. Wang, C. Herman, K.A. Tipton, et al., *Cell* **2002,** *111,* 1027–1039.

298 D. Eisenberg, *Science* **1999,** *285,* 1021–1022.

299 B. Bukau, A. L. Horwich, *Cell* **1998,** *92,* 351–366.

300 L. S. Chesnokova, S. V. Slepenkov, II Protasevich et al., *Biochemistry* **2003,** *42,* 9028–9040.

301 G. L. Dumitru, Y. Groemping, D. Klostermeier et al., *J. Mol. Biol.* **2004**, *339,* 1179–1189.

302 D. Schneider, W. Ross, R. Gourse, *Curr. Opin. Microbiol.* **2003,** *6,* 151–156.

303 M. Torres, C. Condon, J. M. Balada et al., *EMBO J.* **2001,** *20,* 3811–3820.

304 J. M. Zengel, L. Lindahl, *Prog. Nucl. Acid Res. Mol. Biol.* **1994, 47, 331–370.**

Index

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2