

1

Ribosomal Protein Synthesis

Prof. Dr. Wolfgang Wintermeyer¹, Prof. Dr. Marina V. Rodnina²

¹ Institut für Molekularbiologie, Universität Witten/Herdecke, Stockumer Straße 10, 58448 Witten, Germany; Tel.: +49-2302-669140; Fax: +49-2302-669117; E-mail: winterme@uni-wh.de

² Institut für Physikalische Biochemie, Universität Witten/Herdecke, Stockumer Straße 10, 58448 Witten, Germany; Tel.: +49-2302-669205; Fax: +49-2302-669117; E-mail: rodnina@uni-wh.de

1	Introduction	2
2	Historical Outline	3
3	The Genetic Code	4
3.1	The Four Letters of the Genetic Code	5
3.2	Mutations	5
3.3	Recoding	6
4	The Translational Apparatus	6
4.1	Transfer RNA	7
4.2	Ribosomes	8
4.3	Translation Factors	11
4.4	Messenger RNA	12
5	Protein Synthesis	12
5.1	Initiation of Translation in Bacteria	13
5.2	Initiation of Translation in Eukaryotes	13
5.3	The Elongation Cycle	15
5.3.1	Aa-tRNA Binding	15
5.3.2	Aa-tRNA Selection	15
5.3.3	Peptide Bond Formation	17
5.3.4	Translocation	18
5.3.5	Rate of Elongation	19
5.4	Termination	19

6	Regulation of Translation in Eukaryotes	20
7	Inhibitors, Antibiotics, and Toxins	22
8	Further Reading	24

Aa-tRNA	aminoacyl-tRNA
EF	bacterial elongation factor
eEF	eukaryotic elongation factor
IF	bacterial initiation factor
eIF	eukaryotic initiation factor
IRES	internal ribosome entry site
mRNA	messenger RNA
Pept-tRNA	peptidyl-tRNA
PT	peptidyl transferase
RF	bacterial release factor
eRF	eukaryotic release factor
rRNA	ribosomal RNA
tRNA	transfer RNA
UTR	untranslated region

1 Introduction

Information for the amino acid sequence of proteins is encoded in the nucleotide sequence of the DNA. Gene expression, i.e., the synthesis of a protein encoded in a gene comprises two main phases: transcription of one DNA strand into the complementary RNA copy (messenger RNA, mRNA) and translation of the mRNA into protein by polymerizing amino acids in the sequence specified by the nucleotide sequence of the mRNA. Three nucleotides out of four specify one of the 20 amino acids incorporated into protein. The substrates of protein synthesis are aminoacylated tRNAs (Aa-tRNAs) that at their 3' end carry specific amino acids in an energy-rich ester bond. Aa-tRNAs are produced by the action of aminoacyl-tRNA synthetases, which each specifically recognize one amino acid, and the tRNA that is specific for this amino acid, as it has the

anticodon triplet matching the respective codon triplet. Thus, the tRNA serves as an adaptor between the amino acid and the codon on the mRNA. Protein synthesis takes place on ribosomes that are large macromolecular assemblies consisting of several RNAs (ribosomal RNA, rRNA) and numerous proteins. The ribosome presents the mRNA such that Aa-tRNAs can bind to their respective codon and take part in peptide bond formation catalyzed by the ribosome. In their functions, ribosomes are assisted by a number of accessory enzymes (translation factors).

In prokaryotes, transcription and translation are coupled spatially and functionally, such that ribosomes initiate translation on mRNA while it is being synthesized. In eukaryotes, the two processes are separated spatially, in that transcription and processing of primary transcripts takes place in the nucleus, and mature mRNAs are exported from the nucleus to the cytosol for translation.

Translational control, i.e., the regulation of the efficiency of translation of a given mRNA, is an important means of regulation by which the cell can adapt to changes of its physiological status or to external stimuli. Compared with the control of transcription and posttranscriptional processing, translational control can elicit a more rapid response. It is particularly important in eukaryotes, and in most cases the regulated step is initiation.

Various components of the translational apparatus, including the ribosome and translation factors, are targeted by naturally occurring inhibitors, such as many of the classic antibiotics, that inhibit specific steps of translation. The development of new inhibitors of translation that can be used for therapeutic purposes is an important challenge for research on translation.

2

Historical Outline

Our present knowledge about ribosomal protein synthesis is the result of five decades of research. Soon after DNA had been identified as the molecule that, in its base sequence, encodes the information for the amino acid sequence of proteins, it became clear that DNA does not serve as a template for protein synthesis directly. In the early 1950s, ribosomes were identified as the site of amino acid incorporation into protein, taking place in the cytosol of the cell. Soon thereafter, protein synthesis on ribosomes was demonstrated in bacteria as well. In the late 1950s, the function of transfer RNA (tRNA) in carrying amino acids to the ribosome was elucidated, and the enzymes that attach the amino acids to tRNAs at the expense of ATP (aminoacyl-tRNA synthetases) were characterized. As ribosomes contain RNA, it was assumed for some time that it was ribosomal RNA (rRNA) that served as

a template and determined the sequence of the protein to be synthesized. It took several years until, in 1960, the template was identified as messenger RNA (mRNA) that, in eukaryotic cells, was transcribed from DNA in the nucleus and translated in the cytosol. Thus, by the end of the first decade of research on protein synthesis, the important role of different RNAs (rRNA, tRNA, mRNA) in translation had been recognized and the central function of the ribosome was established.

The genetic code, i.e., the meaning of each of the 64 trinucleotide codons of the mRNA, was deciphered in the first half of the 1960s. The first sequence of a nucleic acid, the 76-nucleotide sequence of alanine-specific tRNA, was determined in 1965, and it was soon followed by the determination of several more tRNA sequences. In these years it was also found that, apart from RNA, protein synthesis required soluble protein factors, called translation factors, in order to proceed at a significant rate. The first translation factors to be characterized were the factors that promote partial reactions of the elongation cycle of protein synthesis. The identification of factors for initiation and termination followed, and by the end of the second decade most of the components of the translation apparatus were known. However, the knowledge of the biochemical mechanisms was still in its infancy, and in subsequent years a large body of biochemical information was gathered as a result of the massive efforts of many groups.

One major development during the 1970s was the determination of the primary structures of ribosomal proteins and ribosomal RNA, such that by 1980 the complete sequence of the ribosome from *Escherichia coli* was known and secondary structure models of ribosomal RNAs were developed. Another important milestone in structure determination was reached when the struc-

ture of a tRNA molecule was solved in 1974 – the first tertiary structure of a nucleic acid at atomic resolution. At the same time, the architecture of ribosomes at low resolution was visualized by electron microscopy. The topography of functional centers, such as tRNA binding sites, decoding sites, peptidyl transferase centers, and factor binding sites, as well as of individual ribosomal proteins, was studied by using cross-linking and immuno-electron microscopy techniques. The basal biochemical mechanisms of protein synthesis were worked out, showing similarities and differences in prokaryotic and eukaryotic systems.

As a consequence of the advent, in the early 1980s, of catalytic RNA, and of the difficulty in assigning defined biochemical functions to ribosomal proteins, there was a shift from the protein paradigm, which assumed that major ribosome functions were carried out by ribosomal proteins, to the RNA paradigm, attributing both structural and functional roles to ribosomal RNA. Structural studies on the ribosome employed cross-linking and scattering techniques in order to determine interactions between different parts of ribosomal RNAs and between RNA and proteins, as well as the locations and interactions of proteins. These studies provided fairly well-defined, low-resolution structural models of the ribosome. Atomic structures of aminoacyl-tRNA synthetases and their tRNA complexes became available. Improved preparation procedures and the application of recombinant DNA technology led to an ever-increasing sophistication of mechanistic understanding of ribosome function. Many aspects of translational regulation in eukaryotes, of the initiation step in particular, were worked out and linked translation to general regulatory circuits in the cell.

The last decade was dominated by structure, although major steps in unraveling

detailed kinetic mechanisms of ribosome function were made in parallel. Atomic structures of translation factors provided hints as to their function on the ribosome. Cryo-electron microscopy provided three-dimensional models of ribosomes and ribosome complexes with various ligands of ever-increasing detail. The structural work culminated in the recent determination, at atomic resolution, of the structure of ribosomal subunits and, at lower resolution, of whole ribosomes with tRNAs bound to their respective sites. These structures, together with advanced enzymology, form the basis for future work toward molecular mechanisms of translation. It has become clear by now that the ribosome has a dynamic structure that changes in response to substrate binding and the action of translation factors and takes an active part in all steps of translation. As to the enzymology of the ribosome, the structures show that the ribosome is a ribozyme whose active sites consist of RNA and whose assembly and functions are assisted by ribosomal proteins.

3 The Genetic Code

The genetic code provides the nucleic-acid alphabet that is used to encode the sequence of proteins in their respective genes. The structure of the code and the mechanism of decoding are determined by the fact that there are 4 nucleic acid bases and at least 20 amino acids to be coded for. Alterations of the coding sequence by mutations frequently lead to changes of the amino acid sequence of the encoded protein. Recoding events during translation can lead to changes in the amino acid incorporated in response to certain codons or to changes in the frame in which a particular mRNA is translated.

3.1

The Four Letters of the Genetic Code

The bases of the mRNA (guanine, G; adenine, A; cytosine, C; uracil, U) are the four letters of the genetic code, of which three each form one word that specifies an amino acid. From the total of 64 (4^3) code words, 61 stand for the 20 standard amino acids to be incorporated into proteins during translation (sense codons); the remaining three serve as termination signals that specify the end of the coding sequence (nonsense codons) (Table 1). Because the number of codons is much larger than the number of amino acids to be coded for (degenerated code), most amino acids (except methionine and tryptophan) are encoded by two or more codons. One codon, AUG, stands for methionine; in a particular mRNA context, it specifies the start of the coding sequence, implying that the first amino acid of newly synthesized proteins is methionine. The codon triplets of the mRNA are decoded by the formation of

three complementary base pairs with the anticodons of tRNAs. Elongator tRNAs decode internal codons. Initiation codons are decoded by a particular methionine-specific tRNA, initiator tRNA, which differs from elongator tRNA^{Met} used for decoding internal methionine codons. The three termination codons, UAA, UAG, UGA, stand for “Stop”; these codons are recognized by proteins, the termination (or release) factors, rather than by tRNAs. In a particular mRNA context, UGA also may be decoded as selenocysteine (see Section 3.3), which is therefore the 21st amino acid incorporated into protein during ribosomal protein synthesis. The genetic code is the same in all organisms. Slight deviations, i.e., a few codons with different meanings, were found in mitochondria and ciliates.

3.2

Mutations

Deletions or insertions of bases in DNA/RNA sequences coding for protein nearly

Tab. 1 Genetic code

1st Base	2nd Base				3rd Base				
	U	C	A	G					
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

always (except when three bases are deleted or inserted) lead to a totally different amino acid sequence, as downstream of the mutated position the reading frame is shifted (frameshift mutation).

Single amino acid exchanges are the result of point mutations in which one base in a codon is replaced with another. Base exchanges in the first and second codon positions frequently result in an exchange of the amino acid incorporated at the mutated codon. Base exchanges that lead to the incorporation of chemically related amino acids are referred to as neutral mutations. There is no change in the amino acid sequence when a base exchange, mostly in the third codon position, creates a codon that specifies the same amino acid (silent mutation).

Single-base substitutions that change a sense codon coding for an amino acid into a termination codon (nonsense mutation), for instance UGG (tryptophan) to UGA (stop), are deleterious, as at such mutated positions the synthesis of the protein is terminated prematurely and full-length, functional protein is not made. Truncated proteins formed as a consequence of nonsense mutations may impair cell functions. In eukaryotic cells, mRNAs containing premature stop codons are subject to rapid, preferential nucleolytic degradation (nonsense-mediated decay).

3.3

Recoding

Recoding is a programmed event (not an error) during translation that leads to a protein that differs in sequence from the sequence predicted from the mRNA on the basis of the standard genetic code and that must take place to obtain functional protein. This can be redefinition of a single codon or a shift of the reading frame (mostly by one

base in +1 or -1 direction). Recoding events depend on the presence of particular structural elements in the mRNA which constitute recoding signals of varying efficiency. Frameshift signals are frequently provided by stable hairpins or pseudoknots in the mRNA that lead to ribosome pausing on the sequence on which the frameshift takes place. The nature of the shift sequence is such that it provides sufficiently stable base pairing between codon and anticodon after the anticodon has shifted by one base.

One important example of codon redefinition is the incorporation of selenocysteine at UGA codons, which normally function as termination codons. Selenocysteine is found in the active site of a number of oxidoreductases, such as glutathione peroxidase or deiodinases for thyroxine and triiodothyronine. It is formed from serine by enzymatic transformation of a particular tRNA charged with serine to form Sec-tRNA^{Sec} (Sec, selenocysteine). With the help of a specialized translation factor (SelB in bacteria), Sec-tRNA^{Sec} binds to the ribosome and recognizes UGA when a particular structural element of the mRNA, the selenocysteine insertion sequence, which recruits SelB, is present either within the coding sequence downstream of the UGA (bacteria) or further away in the 3'-untranslated region (3'-UTR) of the mRNA (eukaryotes).

4

The Translational Apparatus

Translation of the nucleotide sequence of mRNAs into protein requires a complex apparatus that consists of a large variety of proteins and RNAs. Transfer RNA (tRNA) functions as an adapter between the codon of the mRNA and the respective amino acid specified by the codon. tRNAs are aminoacylated by aminoacyl-tRNA synthetases.

Decoding and peptide synthesis take place on ribosomes, large ribonucleoprotein complexes consisting of three or four RNAs and 50 to 80 proteins, depending on organism and organelle. Ribosomes are assembled in the nucleus and carry out protein synthesis in the cytosol. All phases of protein synthesis, i.e., initiation, elongation, and termination, require the action of translation factors that interact with the ribosome at defined stages of translation.

4.1

Transfer RNA

tRNAs are small RNAs (75–85 nucleotides) that have very similar secondary and tertiary structure (Figure 1) and contain a large number of modified nucleosides. The functional centers of the tRNA molecule are the anticodon and the 3' terminus to which the amino acid is attached by an energy-rich ester bond.

Decoding of mRNA codons on the ribosome entails the formation of base pairs

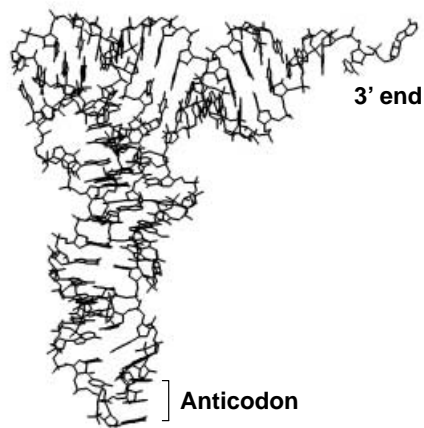


Fig. 1 Tertiary structure of yeast tRNA^{Phe}: The three anticodon bases recognize the codon triplet on the ribosome, and during aminoacylation the amino acid is attached to the 3'-terminal adenosine.

between three bases of the codon and complementary bases of the anticodon. Most amino acids are specified by more than one codon, in some cases by up to six (Table 1). In such cases, there are two or more tRNAs that are charged with the same amino acid (isoacceptors) and have different sequences, including different anticodons.

In the first and second codon position, base pairing with the anticodon strictly follows the Watson-Crick complementarity rules, i.e., G pairs with C and A pairs with U. Base pairing at the third codon position is less restrictive because also “wobble” pairs (G:U, U:G, or I:U,C,A; I stands for inosine) are “allowed”, i.e., energetically favorable. As a consequence, there are tRNAs that can pair with two or more codons differing in the third base, which reduces the total number of different tRNAs required to decode all 61 sense codons. Most organisms have around 50 different tRNAs, and a set of 22 tRNAs suffices for protein synthesis in mitochondria of mammalian cells.

Attachment of the amino acid to the 3' end of the tRNA (aminoacylation) is catalyzed by aminoacyl-tRNA (Aa-tRNA) synthetases (amino acid-tRNA ligases). Aa-tRNA synthetases are specific for their respective amino acid and tRNA (or group of isoaccepting tRNAs). Thus, every organism contains 20 different Aa-tRNA synthetases, one for every amino acid incorporated. (An interesting exception is an Aa-tRNA synthetase with specificity for two amino acids [Pro, Cys] found in a thermophilic archaeon.)

Aminoacylation takes place in two steps that are both catalyzed by the same Aa-tRNA synthetase:

- 1) Amino acid + ATP = Aa-AMP + PP_i
- 2) Aa-AMP + tRNA = Aa-tRNA + AMP

The first step is the activation of the amino acid with ATP by formation of a mixed

anhydride formed between the carboxyl group of the amino acid and the phosphoryl group of AMP. The second reaction is the transfer of the aminoacyl residue from Aa-AMP to tRNA to form an energy-rich ester bond with a hydroxyl group of the 3'-terminal ribose.

Aa-tRNA synthetases vary in size and subunit composition; there are monomeric enzymes with molecular masses around 50 kDa and ($\alpha\beta$)₂ heterotetramers with over 200 kDa. Two classes of Aa-tRNA synthetases can be distinguished that differ in the structure of the catalytic center and transfer the amino acid to either the 2' or the 3' hydroxyl group of the terminal ribose of the tRNA. By spontaneous migration of the aminoacyl residue, the 2' derivative isomerizes to the 3' Aa-tRNA, which is the substrate for subsequent steps of translation.

The aminoacylation reaction is highly accurate, i.e., the frequency by which an incorrect amino acid is attached to a tRNA is low (about 10^{-4}), matching the overall error frequency of translation. Thus, Aa-tRNA synthetases must distinguish correct and incorrect amino acids and tRNAs with high accuracy. Most amino acids can be discriminated with sufficient accuracy in a single binding step on the basis of structure, size,

or chemical character. In the few cases where these criteria do not suffice for high-accuracy discrimination (isosteric or chemically similar amino acids, such as threonine and isoleucine or isoleucine and valine, respectively), the required low error level is attained by hydrolytically discarding incorrectly formed products (editing). Editing synthetases have a second catalytic site where incorrectly formed Aa-AMP or Aa-tRNA is hydrolyzed.

The recognition of the tRNA substrate is brought about by specific interactions between the Aa-tRNA synthetase and structural elements at various positions of the tRNA molecule (identity elements), frequently including residues in the acceptor arm and the anticodon. Isoaccepting tRNAs that are aminoacylated by the same synthetase possess the same pattern of identity elements.

4.2

Ribosomes

Ribosomes are large (diameter, 20 nm) ribonucleoprotein particles (RNPs) that consist of two subunits of different sizes. They contain several molecules of rRNA and many proteins that are mostly small (10–20 kDa) and basic (Table 2). Ribosomes from bacteria and eukaryotes have similar archi-

Tab. 2 Composition of ribosomes

Organism	Ribosome ^a	Small subunit ^a	Large subunit ^a
Eukaryotes (Mammalia)			
Size	80S (4.2 MDa)	40S (1.4 MDa)	60S (2.8 MDa)
rRNAs		18S rRNA (1874 Nt)	28S rRNA (4718 Nt) 5.8S rRNA (160 Nt) 5S rRNA (120 Nt)
Proteins		33 proteins	49 proteins
Prokaryotes (<i>E. coli</i>)			
Size	70S (2.5 MDa)	30S (0.9 MDa)	50S (1.6 MDa)
rRNAs		16S rRNA (1542 Nt)	23S rRNA (2904 Nt) 5S rRNA (120 Nt)
Proteins		21 proteins	34 proteins

^a Abbreviations: S = sedimentation constant; MDa = megadalton; Nt = nucleotide.

ture, although eukaryotic ribosomes are larger and have higher molecular mass because they have larger rRNAs and more proteins. Mitochondrial ribosomes resemble bacterial ribosomes but contain smaller rRNAs and many more proteins.

The structures of 30S and 50S ribosomal subunits are known at atomic resolution. In both subunits, the structure is determined by the tertiary structure of the RNA to which the ribosomal proteins are attached. The three domains of 16S rRNA are represented in the body, the platform, and the head domain of the 30S particle (Figure 2A). The decoding center is built of parts of helices 44,

18, and 34 of 16S rRNA, and the only protein that is close to the decoding center is S12, which influences fidelity. 16S rRNA forms contacts with the codon–anticodon complex that are essential for the accuracy of decoding. Monitoring the quality of codon–anticodon interaction is performed by helix 44 through A1493 and A1492 that contact the 1st and 2nd position, respectively, in the minor groove of the codon–anticodon duplex; the 3rd position is monitored less stringently by a contact from G530 (Figure 2B).

Major landmarks of the 50S subunit (Figure 3A) are the L1 region; the central

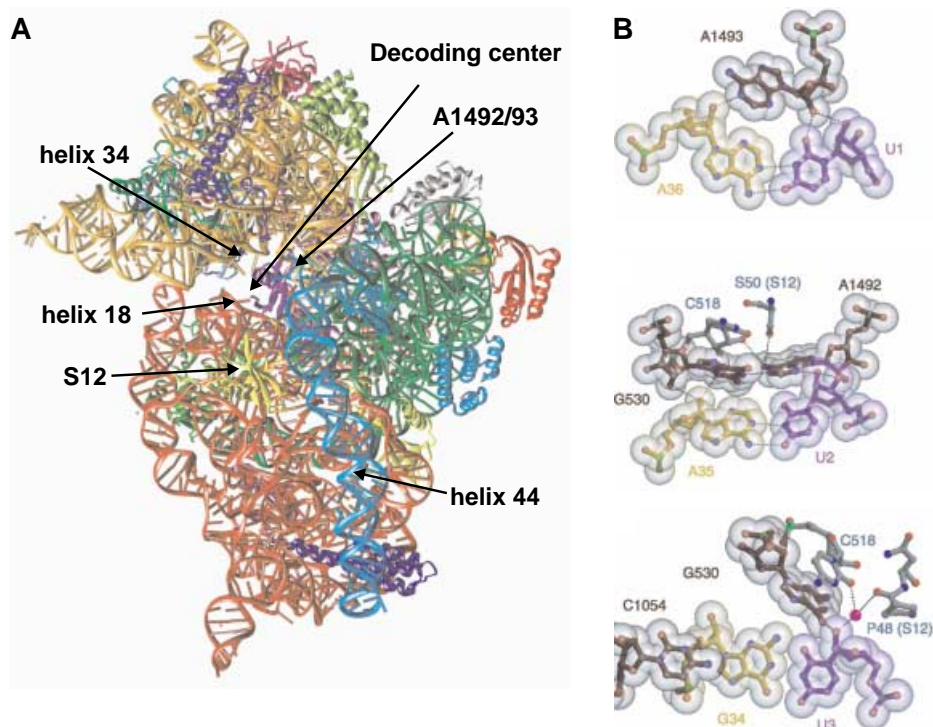


Fig. 2 Atomic structure of the 30S ribosomal subunit of ribosomes: (A) Overview of the 30S structure from *Thermus thermophilus* (ribbon representation). The three major domains of 16S rRNA are colored differently, and proteins are shown in various colors. (B) Recognition of codon–anticodon base pairs by ribosomal residues of the decoding center. Shown are the three anticodon bases (G34A35A36) bound to a codon triplet (U1U2U3) in the 30S crystal. Reprinted from *Science* 292 (2001) 897, with permission, Copyright 2001 American Association for the Advancement of Science.

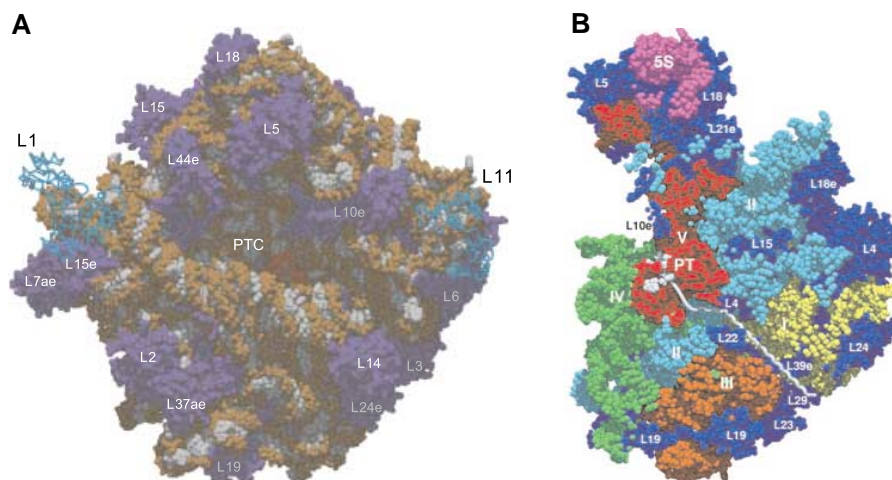


Fig. 3 Atomic structure of the 50S ribosomal subunit: (A) Subunit interface view of 50S subunits from *Haloarcula marismortui*. The space-filling model shows RNA in gray/brown and proteins in blue. The peptidyl transferase center (PTC) is indicated by an inhibitor (red) mimicking the tetrahedral intermediate of the PT reaction. (B) Side view cut to show the peptide exit tunnel. Domains I-V of 23S rRNA are shown in different colors, proteins in blue. The tunnel is indicated by a modeled peptide extending from the PT center to the exit. Reprinted from *Science* 289 (2000) 920, with permission, Copyright 2000 American Association for the Advancement of Science.

protuberance that contains 5S rRNA; and the stalk region where proteins L11, L10, and L7/12 (not seen in the structure) form an important site for elongation factor binding. The peptidyl transferase center, as localized by an analogue mimicking the tetrahedral intermediate of the peptidyl transferase reaction, is made up of RNA only, suggesting that the ribosome is a ribozyme. The exit tunnel for the growing peptide extends from the peptidyl transferase center through the body to the back of the subunit where the peptide emerges (Figure 3B).

The structure of 70S ribosomes with three tRNA molecules bound in the binding sites for Aa-tRNA (A site), Pept-tRNA (P site), and the exit site for deacylated tRNA (E site) was determined by X-ray crystallography at somewhat lower resolution (Figure 4). The arrangement of the tRNAs in A and P sites with their anticodons and CCA ends, respectively, coming close together is clearly

seen in the structure. There are a number of connections between the subunits, most of them made up of RNA. One prominent connection (bridge 2a) is formed of helix 69 of 23S rRNA and connects to the 30S decoding site (helix 44), providing the basis for transmitting conformational signals between subunits. Some bridges contact the A-site tRNA, e.g., helix 38 (bridge 1a, or “A-site finger”) from above and helix 69 from below; the latter also contacts the P-site tRNA. These interactions stabilize the tRNAs in their binding positions and presumably have to be released to allow tRNA movement during translocation.

Structural and functional analyses suggest that rRNA is involved in the major functions of the ribosome, i.e., decoding, peptide bond formation, and translocation. This is consistent with the fact that the respective functional centers of rRNA are highly conserved. However, ribosomal proteins