

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

3D modeling of ribosomal RNA using cryo-electron microscopy density maps



Alexander Jarasch

aus

Leipzig

2011

Erklärung

Diese Dissertation wurde im Sinne von §13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 (in der Fassung der vierten Änderungssatzung vom 26. November 2004) von Herrn Prof. Dr. Roland Beckmann betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 31.03.2011

Alexander Jarasch

Dissertation eingereicht am: 31.03.2011

1. Gutachter: Prof. Dr. Roland Beckmann
2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am: 25.05.2011

Parts of this thesis have been published:

Thomas Becker, Jean-Paul Armache, **Alexander Jarasch**, Andreas Anger, Elizabeth Villa, Basma Abdel Motaal, Thorsten Mielke, Otto Berninghausen and Roland Beckmann

“Structure of the No-Go mRNA decay complex Dom34/Hbs1 bound to a stalled 80S ribosome”
NSMB, accepted

Armache JP* and **Jarasch A*** and Anger AM*, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Söding J, Westhof E, Wilson DN, Beckmann R.

“Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution.”
PNAS USA. 2010 Nov 16;107(46):19748-53. Epub 2010 Oct 27.

Armache JP* and **Jarasch A*** and Anger AM*, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Söding J, Westhof E, Wilson DN, Beckmann R.

“Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome.”

PNAS USA. 2010 Nov 16;107(46):19754-9. Epub 2010 Oct 25.

Bhushan S, Gartmann M, Halic M, Armache JP, **Jarasch A**, Mielke T, Berninghausen O, Wilson DN, Beckmann R.

“alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel.”

NSMB 2010 Mar;17(3):313-317

Becker T, Bhushan S, **Jarasch A**, Armache JP, Funes S, Jossinet F, Gumbart J, Mielke T, Berninghausen O, Schulten K, Westhof E, Gilmore R, Mandon EC, Beckmann R.

“Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome.”

Science. 2009 Dec 4;326(5958):1369-73.

Parts of this thesis have presented at the following international conferences:

Talks:

RNA 2010, 15th Annual Meeting of the RNA Society, Seattle (USA)

Alexander Jarasch and Jean-Paul Armache, Andreas Anger, Elizabeth Villa, T. Becker, F. Jossinet, J. Soeding, E. Westhof, Daniel Wilson and Roland Beckmann

“The structure of a translating eukaryotic 80S ribosome at 5.5 Å” (invited speaker)

Posters:

ISMB/ECCB 2009, Stockholm (Sweden)

Alexander Jarasch, F. Jossinet, E. Villa, T. Becker, K. Schulten, E. Westhof and Roland Beckmann

“3D modeling of ribosomal RNA based on cryo-EM maps”

RNA 2008, 13th Annual Meeting of the RNA Society, Berlin (Germany)

Alexander Jarasch, Jean-Paul Armache, F. Jossinet, T. Becker, M. Topf, J. Soeding, E. Westhof and Roland Beckmann

“Molecular model of the yeast ribosomal tunnel and exit site”

EMBO Practical Course on computational RNA biology 2008, Cargese (France)

Alexander Jarasch, F. Jossinet, T. Becker, E. Westhof and Roland Beckmann

“Modelling of ribosomal rRNA using high-resolution cryo-EM maps”

Participant:

[BC]² 2008 - Basel Computational Biology Conference, Basel (Switzerland)

Contents

I. Introduction	10
1. Ribosomes	11
1.1. Ribosome structures	11
1.1.1. Structure of the bacterial small ribosomal subunit	11
1.1.2. Structure of the <i>archaeal</i> large ribosomal subunit	12
1.1.3. Structures of bacterial 70S ribosomes	13
1.1.4. Studies of eukaryotic 80S ribosomes	14
1.1.5. RNA expansion segments and variable regions	16
1.1.6. Ribosomal proteins	18
1.2. Functional site of ribosomes	18
1.2.1. The decoding site	18
1.2.2. Peptidyl transferase center	21
1.2.3. The ribosomal tunnel and tunnel exit site	22
1.2.4. Translation factor binding site	23
1.3. Evolution of ribosomal RNA	24
2. RNA	26
2.1. Base pairing	26
2.2. Structural organization of RNA	28
2.3. RNA structural motifs	30
2.3.1. SCOR database	31
2.3.2. A-minor motif	31
2.3.3. Ribose-Zipper	32
2.3.4. Kink-turn	33
2.3.5. Tetraloop motifs	34
2.3.6. Pseudoknots	34
2.4. RNA-protein interactions	35
2.5. RNA structure prediction and modeling	36
3. Cryo-electron microscopy	38
3.1. Cryo-EM and single particle analysis	38
3.2. Limitations of cryo-EM	39
4. Aim of this study	41

II. Materials and Methods	42
5. Cryo-Electron microscopy	43
6. Templates and sequences	44
6.1. Ribosomal RNA sequences	44
6.2. Templates for the large ribosomal subunit	44
6.3. Templates for the small ribosomal subunit	45
6.4. Template for tRNA	45
6.5. Templates for structural motifs	45
7. Modeling	46
7.1. RNA homology modeling	46
7.1.1. Structure-based sequence alignment	46
7.1.2. S2S	47
7.2. <i>De novo</i> modeling of RNA	47
7.2.1. Secondary structure prediction of RNA	47
7.2.2. Tertiary structure modeling of RNA	47
7.2.3. Assemble	48
7.3. Fitting and refinement of RNA models	48
7.3.1. Assemble refinement	48
7.3.2. Molecular Dynamics Flexible Fitting (MDFF)	48
7.4. Visualization and structure analysis	50
III. Results	51
8. Workflow for modeling and fitting large RNA molecules	52
8.1. RNA homology modeling	53
8.1.1. Structure-based sequence alignment	54
8.1.2. Deducing the homology model	55
8.1.3. Homology modeling of structural motifs	55
8.2. RNA <i>de novo</i> modeling	55
8.3. Refinement and flexible fitting of RNA	57
9. Model of <i>T. aestivum</i> and <i>S. cerevisiae</i> 80S ribosomes based on high-resolution cryo-EM structures	58
9.1. Cryo-EM reconstruction	58
9.2. Near complete model of the eukaryotic 80S ribosome	60
9.2.1. Atomic model of the eukaryotic ribosomal RNA core structure	68
9.2.2. Atomic models for eukaryote-specific ribosomal RNA expansion segments	69
9.2.2.1. ES1 ^S (h6)	73
9.2.2.2. ES3 ^S (h9, h10)	73
9.2.2.3. ES5 ^S (h16, h17)	75
9.2.2.4. ES6 ^S (h21)	77

Contents

9.2.2.5.	ES7 ^S (h26)	79
9.2.2.6.	ES8 ^S (h33)	80
9.2.2.7.	ES9 ^S (h39)	81
9.2.2.8.	ES10 ^S (h41)	82
9.2.2.9.	ES12 ^S (h44)	83
9.2.2.10.	ES3 ^L / ES4 ^L (H9, H10)	84
9.2.2.11.	ES5 ^L (H15, H16-H18)	86
9.2.2.12.	ES7 ^L of <i>T. aestivum</i>	88
9.2.2.13.	ES7 ^L of <i>S. cerevisiae</i>	90
9.2.2.14.	ES9 ^L (H31, H28-H30)	92
9.2.2.15.	ES10 ^L / ES12 ^L (H38)	94
9.2.2.16.	ES15 ^L (H45)	96
9.2.2.17.	ES19 ^L (H52)	97
9.2.2.18.	ES20 ^L / ES26 ^L (H54, H55)	98
9.2.2.19.	ES24 ^L (H59)	99
9.2.2.20.	ES27 ^L of <i>T. aestivum</i>	100
9.2.2.21.	ES27 ^L of <i>S. cerevisiae</i>	103
9.2.2.22.	ES31 ^L (H79)	105
9.2.2.23.	ES39 ^L (H98)	107
9.2.2.24.	ES41 ^L (H101)	109
9.3.	Comparison between <i>S. cerevisiae</i> and <i>T. aestivum</i>	110
9.3.1.	Comparison of the cryo-EM densities	110
9.3.2.	Comparison of expansion segments	110
9.3.2.1.	Interaction of ES3 ^S and ES6 ^S	111
9.3.2.2.	Comparison between ES7 ^L in <i>T. aestivum</i> and <i>S. cerevisiae</i>	112
9.3.2.3.	Motion of ES27 ^L	113
10.	Molecular interpretation of functional sites of the eukaryotic 80S ribosome	114
10.1.	The decoding site and the PTC	114
10.1.1.	The decoding site	114
10.1.2.	The PTC	115
10.2.	Model of the eukaryotic ribosomal tunnel	116
10.2.1.	Interaction of the ribosomal tunnel with nascent chains	116
10.3.	The eukaryotic ribosomal tunnel exit site	120
10.3.1.	Interaction of the protein-conducting channel and the tunnel exit site	121
10.4.	Canonical Translation factor binding site	123
IV.	Conclusion and discussion	129
11.	Pipeline for modeling and fitting RNA	130

12. Molecular model of the 80S ribosome	134
12.1. Evaluation of RNA models for the eukaryotic ribosome	135
12.1.1. Comparison to the RNA model of the <i>mammalian</i> 80S ribosome	135
12.1.2. Comparison to the RNA model of the <i>Thermomyces lanuginosus</i> 80S ribosome	135
12.1.3. Comparison to the X-ray structure of the <i>S. cerevisiae</i> 80S ribosome at 4.5 Å	138
12.2. The ribosomal expansion segments and variable regions	140
12.2.1. Interaction between ES3 ^S and ES6 ^S	140
12.2.2. Expansion segment ES6 ^S	141
12.2.3. Expansion segment ES7 ^L	142
12.2.4. Expansion segment ES27 ^L	142
12.3. Functional sites on the 80S ribosome	143
12.3.1. The decoding site and the PTC	143
12.3.2. The eukaryotic ribosomal tunnel	144
12.3.3. The eukaryotic tunnel exit site	146
12.3.4. Canonical translation factor binding site	147
12.4. Co-evolution	147
12.5. Evolution of ribosomes	148
13. Summary	150

Abstract

Ribosomes are macromolecular protein-RNA complexes translating mRNA into protein. To date, crystal structures are available for the bacterial 30S and archaeal 50S subunits, as well as the complete bacterial 70S ribosomes. Eukaryotic ribosomes are much more complex in terms of ribosomal RNA and proteins. However, to date high-resolution crystal structures of eukaryotic ribosomes or ribosomal subunits are lacking.

In order to build reliable models for the eukaryotic rRNA, we developed an approach for large scale homology and *de novo* modeling of RNA and subsequent flexible fitting into high-resolution cryo-EM density maps.

Using this approach we built a model of the *T. aestivum* and the *S. cerevisiae* ribosome based on available cryo-EM maps at 5.5 Å and 6.1 Å resolution, respectively. The model comprises of ~98% of the eukaryotic rRNA including all 21 RNA expansion segments (ES) and structurally six variable regions. Further, we were able to localize 74/80 (92.5%) of the ribosomal proteins. The model reveals unique ES-ES and r-protein-ES interactions, providing new insight into the structure and evolution of the eukaryotic ribosome. Moreover, the model was used for analyzing functional ribosomal complexes, i.e. the characterization of different nascent polypeptide chains within the ribosomal tunnel, intermediates of protein translocation as well as mRNA quality control.

Part I.

Introduction

1. Ribosomes

In all three kingdoms of life the translation of messenger-RNA (mRNA) into polypeptides/proteins is carried out by a large ribonucleoprotein complex, called the ribosome [1]. Ribosomes consist of two subunits of unequal size and shape. The bacterial 70S ribosome consists of a small (30S) subunit and a large (50S) subunit. In contrast to the bacterial ribosome the eukaryotic counterpart sediments at 80S and consists of a small 40S subunit and a large 60S subunit. The additional mass is caused by additional rRNA expansion segments (ES; see section 1.1.5; 9.2 and 9.2.2) and eukaryote-specific ribosomal proteins (r-proteins) and r-protein extensions (see section 1.1.6 and 9.2). The molecular weight of the ribosome ranges between ~ 2.5 MDa in prokaryotes, where approximately two-thirds correspond to rRNA and one-third to r-proteins [2], and ~ 4.5 MDa in eukaryotes, where the distribution is approximately 50% rRNA and 50% r-proteins. The overall diameter of the ribosome is about ~ 250 Å [2].

1.1. Ribosome structures

The determination and interpretation of ribosome structures and structures of the ribosome in different functional complexes by X-ray crystallography and cryo electron microscopy (cryo-EM) provided extensive insights into the field of protein translation [1]. Moreover these structures contributed significantly to the knowledge about RNA folding and RNA-protein interactions.

1.1.1. Structure of the bacterial small ribosomal subunit

In 2000, two independent high-resolution crystal structures of the *Thermus thermophilus* 30S subunit were published, one at 3.3 Å resolution and another at 3.05 Å resolution [3, 4]. These structures provided a structural basis for understanding the functions of the 30S subunit, such as decoding on a molecular level.

The bacterial 30S subunit consists of the 16S rRNA molecule (~ 1500 nts) and 21 r-proteins. The shape is largely determined by the 16S rRNA which consists of 4 domains, namely the 5'-domain, the central domain, the 3'-major domain (3'-M domain) and the 3'-minor domain (3'-m domain) (Figure 1.1.1 A, B). The secondary structure contains more than 50 RNA helices connected by single stranded internal and terminal loops. In the crystal structure the internal loops were identified to be involved in long-range non-Watson-Crick base pairing which stabilizes the packing of helices.

The r-proteins are distributed asymmetric on the 30S surface. In addition to their globular domain most of the r-proteins have long non-globular C- or N-terminal extensions [5]. These protein extensions are located between different RNA elements and stabilize the tertiary structure. The 30S plays an essential role in decoding and mediates the interaction between the mRNA codons and the anticodons of the tRNA [6] (see section 1.2.1).

1. Ribosomes

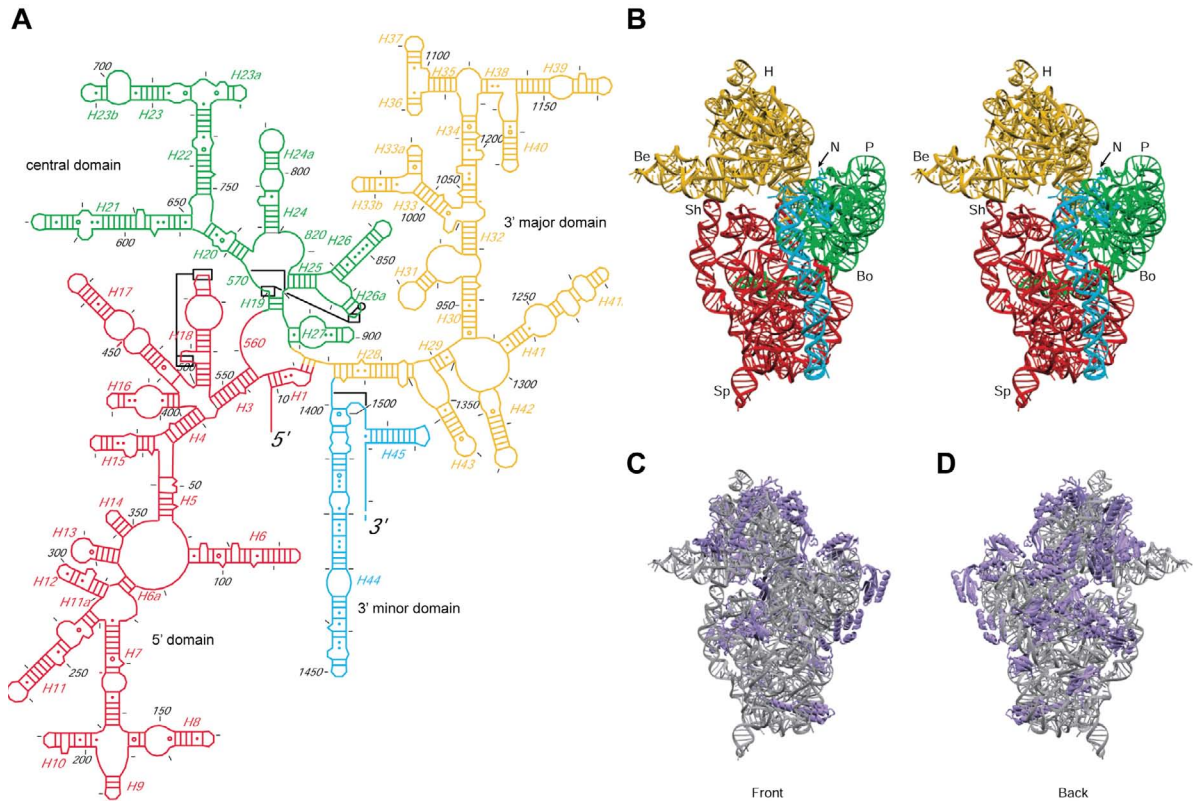


Figure 1.1.1.: **Overview of the 30S subunit.** (A) Secondary structure diagram of the 16S rRNA. The 5' domain is colored in red, the central domain in green, the 3'-major domain in yellow and the 3'-minor domain in blue. (B) Stereo view of the tertiary structure of the 16S rRNA with the same coloring for the domains. (C and D) Front and back view of the 30S. The rRNA is shown in gray and the r-proteins in blue. The figures were taken from [3].

1.1.2. Structure of the *archaeal* large ribosomal subunit

The essentially complete high-resolution crystal structure of an *archaeal* 50S subunit from *Haloarcula marismortui* was solved in 2000 at 2.4 Å resolution [7]. This crystal structure provided the structural basis for the molecular interpretation of peptide bond formation reaction of the ribosome [8]. In this X-ray structure a new structural RNA motif was identified - the kink-turn (see section 2.3.4) [9]. Moreover the A-minor motif, a long-range RNA interaction motif, has been discovered by examining the reason for the large number of conserved adenines in the 23S rRNA molecule (see section 2.3.2) [10].

The *archaeal* 50S subunit is composed of the 23S rRNA molecule (~ 3000 nts), the 5S rRNA (~ 120 nts) and 31 r-proteins. The *archaeal* 23S rRNA consists of 6 domains, namely domain I-VI (Figure 1.1.2 A, B) having a complicated and asymmetric structure (Figure 1.1.2 C-F).

An exhaustive analysis of the large subunit r-proteins revealed a tremendous diversity of interactions with the 23S rRNA [11]. Similar to the 30S subunit the r-proteins appear on the surface of the 50S subunit and largely absent from the interface that contacts the 30S subunit. The long C- or N-terminal extensions of the r-proteins penetrate and stabilize the 3D structure of the rRNA.

1. Ribosomes

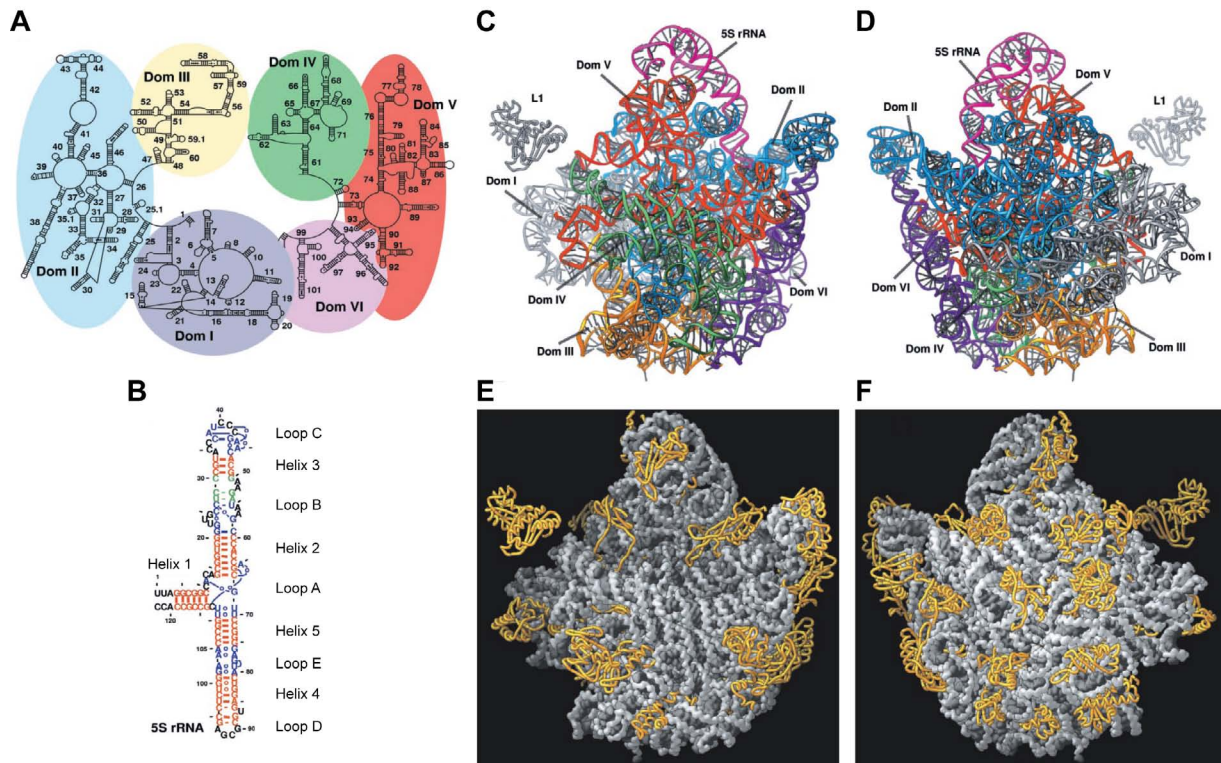


Figure 1.1.2.: **Overview of the 50S subunit.** (A) Schematic secondary structure of the 23S rRNA with coloring for the different domains. (B) Secondary structure of the 5S rRNA from *H. marismortui*. (C and D) Front and back view of the 23S rRNA 3D structure. Domains are colored as in (A). (E and F) Front and back view of the 50S subunit with rRNA shown in gray and proteins highlighted in gold. The figures were taken from [7].

The most significant result derived from the 50S X-ray structure was that the ribosome is indeed a ribozyme [2]. All components at the Peptidyl transferase center (PTC) involved in orienting the A-site α -amino group and the carbonyl carbon bound at the P-site are exclusively made of RNA (see section 1.2.2).

Another two important features of the large subunit are the ribosomal tunnel and tunnel exit site which all nascent polypeptide chains pass before emerging from the ribosome. The tunnel immediately starts below the PTC, spans 80 – 100 Å through the large subunit and ends at the tunnel exit site (see section 1.2.3).

1.1.3. Structures of bacterial 70S ribosomes

Structures of the isolated ribosomal 30S and 50S subunits in complex with oligonucleotide analogs of tRNA and mRNA-tRNA revealed insights into the mechanism of the peptidyl transferase reaction and codon-anticodon recognition, respectively. Beside the X-ray structures of the separate subunits, a crystal structure of the *Thermus thermophilus* 70S ribosome containing mRNA, the A-, P-, and E-site full-length tRNAs was resolved at 5.5 Å resolution [12]. This model provided the details of the subunit-subunit interactions and tRNA-ribosome interactions (Figure 1.1.3 A-C). In 2005 two high-resolution X-ray structures at 3.5 Å resolution from the bacterial *E. coli* 70S ribosome improved the structural knowledge of the intersubunit interactions [13]. This study also revealed details of the conformational flexibility of the ribosome at a molecular level.

1. Ribosomes

The observed swiveling of the 30S head together with the coupled ratchet-like motion of the two subunits suggested a mechanism for the movement of the mRNA and tRNA during translocation. One year later the first all-atom structures of the *T. thermophilus* 70S ribosome in complex with mRNA and tRNA were reported at 2.8 Å resolution [14].

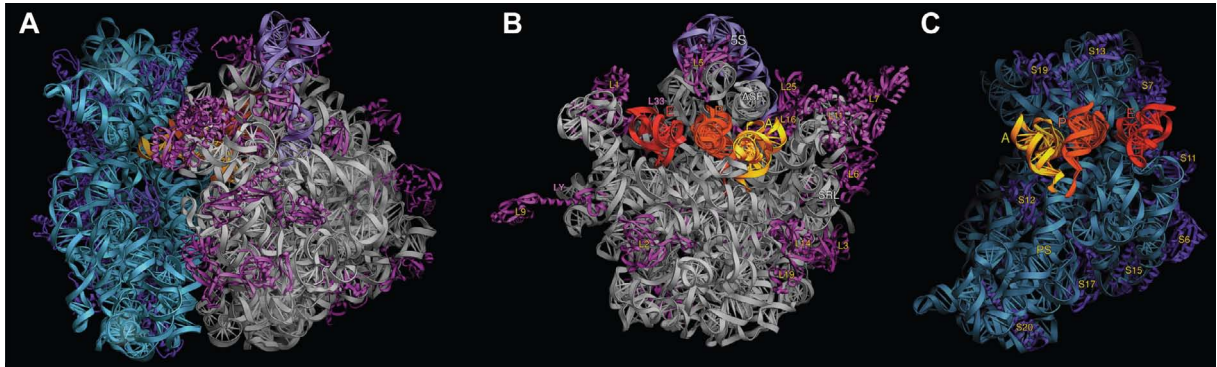


Figure 1.1.3.: **Crystal structure of the *T. thermophilus* 70S ribosome in complex with the full length tRNAs.** (A) Classical view of 70S ribosome. The 16S rRNA is colored in blue, the 50S subunit r-proteins in dark blue, the 23S rRNA in gray, the 5S rRNA in purple and the 30S subunit r-proteins in magenta. The A-site tRNA is colored in yellow, the P-site tRNA in orange and the E-site tRNA in red. (B) View on the 50S subunit from the intersubunit surface. The color code is the same as in (A). (C) View on the 30S subunit from the intersubunit surface. The color code is the same as in (A). The figures were taken from [12].

1.1.4. Studies of eukaryotic 80S ribosomes

In eukaryotes such as *T. aestivum* and *S. cerevisiae* the 40S subunit consists of an 18S rRNA molecule (~ 1800 nts) and 33 r-proteins. Eighteen of the 33 have no bacterial homologue and are therefore eukaryote-specific. The 60S subunit consists of the 5S, the 5.8S and the 25S rRNA molecules and ~ 46 proteins. The additional 28 r-proteins are specific to eukaryotes (and in part *archaea*). Most of the eukaryotic r-proteins, having a bacterial homologue, contain an additional C- or N-terminal extension.

Although the two bacterial 30S and 50S subunits as well as the 70S ribosome could be solved at high-resolution, no high-resolution crystal structure of a eukaryotic 80S ribosome has been published, yet. A number of cryo-EM structures of various 80S ribosomes have been published: i.e. an 80S ribosome from rat liver was interpreted in terms of rRNA inserts and additional r-proteins at 25 Å resolution [15]. In another structural study the eukaryote-specific rRNA expansion segments in the *S. cerevisiae* 80S ribosome a cryo-EM reconstruction have been localized and characterized at 15 Å resolution [16]. Here the atomic coordinates of the *H. marismortui* 50S and *T. thermophilus* 30S were docked into cryo-EM density map. According to the secondary structure diagrams the helical anchor points for each expansion segment were then mapped (Figure 1.1.4 A, B). For the first time, it could be shown that the expansion segments are located at the surface of the ribosome (Figure 1.1.4 C, D). In addition homology models for r-proteins, having homologs in bacteria, were generated and localized in the EM density. Moreover the conserved intersubunit bridges were mapped in the 80S ribosome and in addition four new bridges were suggested.

1. Ribosomes

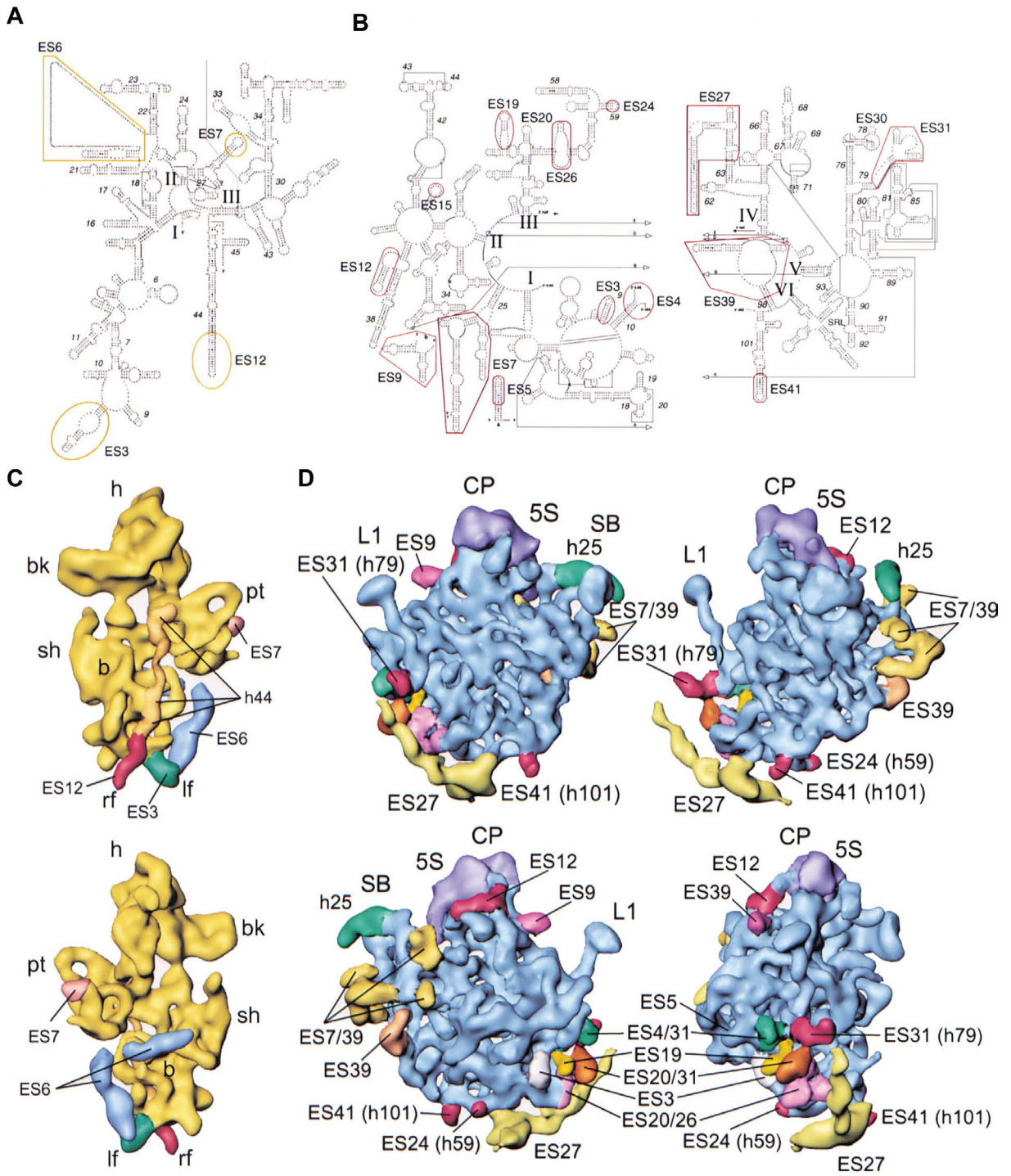


Figure 1.1.4.: **Mapping of expansion segments in *S. cerevisiae*.** (A) Mapping of ES in the 18S secondary structure. (B) Mapping of ES in the 28S secondary structure. (C) Mapping of ES in the 40S cryo-EM reconstruction. (D) Mapping of ES in the 60S cryo-EM reconstruction. The figures were taken from [16]

1. Ribosomes

Another reconstruction showed the yeast Ribosome-Nascent-chain Complex (RNC) bound to the Protein-Conducting-Channel (PCC) of yeast endoplasmatic reticulum, the Sec61 complex, at 15.4 Å resolution [17]. In this reconstruction the highly flexible RNA expansion segment *ES27^L* could be visualized in two distinct conformations.

In 2007 a cryo-EM reconstruction at 13.5 Å from *Thermomyces lanuginosus* has been used to compare expansion segments of fungal 80S ribosomes in structure and conformation [18].

During the writing of this manuscript a crystal structure of the *S. cerevisiae* 80S ribosome at 4.5 Å resolution [19] and a crystal structure of the *Tetrahymena thermophila* 40S ribosomal subunit in complex with eukaryotic initiation factor 1 (eIF1) at a resolution of 3.9 Å resolution [20] were published and will be discussed in section 12.1.3.

Models for 80S ribosomes A cryo-EM reconstruction at 8.7 Å of a canine ribosome was used for a model of the eukaryotic 80S ribosome including models for ~ 50% of the RNA expansion segments, which were modeled as A-form helices and fitted into the EM density [21]. Recently, the cryo-EM reconstruction from *Thermomyces lanuginosus* was refined to 8.9 Å resolution. On this basis models for all expansion segments except for *ES10^L*, *ES27^L* and the tip of *ES15^L* and models of r-proteins, for which a structural homolog could be identified, were generated [22]. Moreover, in this study the eukaryote-specific r-protein S19e was localized and modeled.

1.1.5. RNA expansion segments and variable regions

Compared to the 16S and 23S rRNAs molecules, the eukaryotic 18S and 28S rRNAs counterparts contain additional RNA segments with variable size. These RNA segments of variable size interrupt the universal rRNA core secondary structure and have been defined as *variable regions* or *expansion segments* [23].

Variable regions Variable regions (VR) are defined as regions within the rRNA that are not evolutionarily conserved [23, 24]. They are used to indicate regions of rRNA having either an increase or a decrease in size between organisms.

Expansion segments Expansion segments (ES) are defined as a subset of variable regions and interruptions with expanded size [23]. It is known that expansion segments interrupt the universal core rRNA at the same sites along different species [24]. The identity of expansion segment positions in different organisms suggests a common evolutionary origin. Despite their conserved location the sequence, the length and therefore the structure of the RNA expansion segments vary significantly between the different species [25]. Most expansion segments are highly flexible which makes it complicated to visualize them.

1. Ribosomes

Primary structure The nucleotide sequence of ES is highly variable and repetitive. It has been observed that expansion segments have a bias for certain tri- and tetranucleotide motifs that are shuffled within the sequence. Different ES in human 28S rRNA share the same trinucleotide motif repeats, suggesting that they may have a common evolutionary origin [26].

Due to the dramatic sequence and length variations, the construction of sequence alignments between expansion segments of different organisms is very difficult. Moreover, homologous nucleotides can not be easily deduced.

Secondary and tertiary structure Expansion segments have a defined secondary structure and have been visualized by cryo-EM [16]. Due to the sequence variation and the high repetitive motifs, secondary structure prediction of ES via the minimum free energy results in several secondary structure models with nearly equal free energy.

Suggestions for secondary structure predictions for some ES which are based on comparative sequence analysis are available on the CRW database¹ [27]. However, comparative sequence analysis is insufficient for predicting the secondary structure because of the species-specific variability in length and sequence. Thus, experimental data is required to identify the most probable secondary structure. A possible solution is to perform a secondary structure prediction by using the structural constraints provided by cryo-EM. For example, a secondary structure model for the interacting expansion segments $ES3^S$ and $ES6^S$ of the 18S rRNA was proposed by combining sequence with cryo-EM data [28]. $ES3^S$ and $ES6^S$ were analyzed from more than 2900 discrete eukaryotic species for a possible sequence complementarity between the two ES. In combination with the two adjacent densities in the cryo-EM map this indicated a possible interaction of $ES3^S$ and $ES6^S$ by forming an intermolecular helix consisting of seven to nine continuous base pairs. Another approach is to combine experimental data from structure specific RNA modifying chemical reagents and enzymes with computational methods as shown for $ES15^L$ and $ES39^L$ [29, 30]. As mentioned in section 1.1.4, there are only a few 3D models for RNA expansion segments present, which are based on low-resolution cryo-EM data.

Function of expansion segments It is still an open question whether expansion segments have a functional role. It seems very unlikely that ES were maintained in eukaryotes without playing a role in ribosome assembly or in translation. However, it has been suggested that they are tolerated because they do not disturb the essential ribosomal components [31]. On the one hand, a 119-nt insertion is tolerated in an ES near the 3' end of the 26S rRNA in *Tetrahymena thermophila* [32]. On the other hand, a deletion or substitution of $ES27^L$ in the *T. thermophila* 26S rRNA with an unrelated sequence of similar length is lethal, suggesting an essential function of this region [33]. Deletion of small portions of $ES7^L$ strongly reduces cellular growth and larger deletions are lethal [34]. Some of the deletions in $ES7^L$ cause accumulation of 27S-pre-rRNA, suggesting an important role in the 60S assembly [34].

¹<http://www.rna.cccb.utexas.edu/>

1.1.6. Ribosomal proteins

Ribosomal proteins (r-proteins) are mainly found on the outer surface of the ribosome. Their globular domains are found on the subunits' exterior while their long extended tails are nested in the gaps and cracks between the rRNA [7, 11]. The r-proteins do not only stabilize the RNA 3D structure but are also necessary in the process of ribosome assembly and therefore influence the RNA folding [11]. Here the tails co-assemble with the RNA and therefore ensure the correct folding and assembly of the subunits [5, 11, 35]. Compared to DNA-binding proteins, which mostly recognize the base sequence, ribosomal proteins preferentially bind the minor groove surface of rRNA helices and recognize higher-order structural motifs [35].

Ribosomal proteins neither directly participate in peptide bond formation nor in the decoding process. Thus, their role is rather to improve the accuracy and efficiency of mechanisms that are based on RNA [35]. In *E. coli* it has been observed that at least one-third of the r-proteins can be deleted singly without being lethal [36].

Ribosomal proteins in eukaryotic ribosomes Eukaryotic ribosomes have an additional set of 20-30 r-proteins compared to their bacterial counterparts. For example the *S. cerevisiae* 80S ribosome contains additional 21 r-proteins (9 on the 40S subunit, 12 on the 60S subunit) compared to the *E. coli* 70S ribosome. Besides the eukaryote-specific r-proteins most of the r-proteins having a bacterial counterpart are extended by either a C- or an N-terminal extension, which contribute to the additional protein mass, too.

Evolution of ribosomal proteins With respect to the r-proteins a comprehensive analysis of 66 complete genomes provided insights into distribution and conservation of r-proteins between and within the three domains of life [37]. Regarding the inter-domain distribution, 34 r-proteins (15 in the SSU, 19 in the LSU) are conserved in all three domains. This is in agreement with the structural comparison between the *bacterial* and the *eukaryotic* ribosome [16] which shows the remarkable degree of conservation of the ribosomal core and the global shape of the ribosome. In contrast to *bacteria*, ribosomes from *archaea* and eukaryotes consist of additional 33 r-proteins (13 in the SSU, 20 in the LSU) [37].

1.2. Functional site of ribosomes

During protein synthesis the mRNA is translated into protein via the ribosome. This process can be subdivided in four steps: initiation, elongation, termination and ribosome recycling. All steps are highly dynamic and require several translation factors for ribosome activity. These factors transiently bind the ribosome at specific sites on the two subunits.

1.2.1. The decoding site

Accurate decoding is crucial in protein synthesis and occurs at the A-site on the small ribosomal subunit (Figure 1.2.1 A, B). In this step the anticodon of the correct (cognate) aminoacyl-tRNA (aa-tRNA) is recognized by the mRNA codon. The aa-tRNA is delivered to the ribosomal A-site with elongation factor EF-Tu. The tRNA anticodon is associated reversibly with the codon

1. Ribosomes

of mRNA while the aminoacyl acceptor stem of the tRNA is still bound to EF-Tu. Cognate anticodon-codon binding stimulates a conformational change of EF-Tu and leads to a GTP hydrolysis. This leads to a dissociation of EF-Tu and the aminoacyl moiety is positioned in the A-site of the large subunit [6].

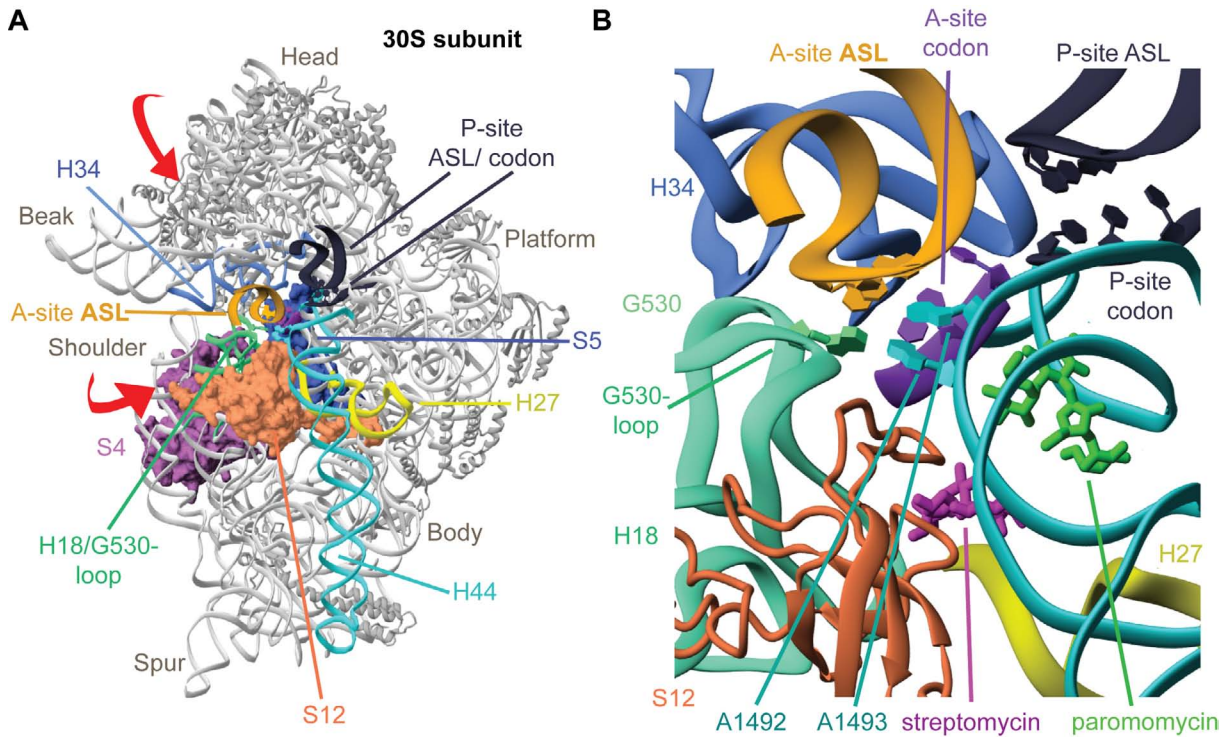


Figure 1.2.1.: **The decoding site on the small ribosomal subunit.** (A) Overview of the 30S subunit structure in complex with A-site tRNA anticodon stem-loop (ASL, gold). Red arrows indicate the movement of domains during the transition to the closed 30S conformation. The 16 rRNA domains are colored in gray, h44 cyan, h34 blue, h27 yellow, h18/G530-loop mint, P-site ASL black, S4 purple and S12 orange. (B) Close-up of selected 30S elements around the decoding center. The color code as in (A). The A-site codon is colored purple, P-site codon in black, antibiotics streptomycin in pink and paromomycin in green. The two adenine A1492 and A1493 and G530 are shown in the A-minor conformation. The figures were adapted from [6].

Crystal structures of the 30S subunit in complex with mRNA and cognate tRNA anticodon stem loops (ASL) in the A site for both in the presence and absence of the antibiotic paromomycin revealed how the ribosome discriminates geometrically between cognate and non-cognate tRNAs [38]. Binding the cognate tRNA leads to a domain rearrangement of the 30S subunit. Further the highly conserved and essential bases G530, A1492 and A1493 change their conformation and closely interact with the first two base pairs via the minor groove between the codon and anticodon (Figure 1.2.2 A-C). Thus these A-minor interactions (A1492, A1493) serve as sensing mechanism for cognate tRNA and discriminate against non-cognate tRNAs.

Another two crystal structures of the 30S subunit in complex with mRNA for both cognate and near-cognate tRNA ASL explained selection of tRNAs and the induced rearrangement of the 30S upon cognate tRNA binding [39]. Binding a near-cognate tRNA leads to a non-canonical codon-anticodon base pair resulting in a different minor groove geometry of the base pair and unsatisfied hydrogen-bonding (Figure 1.2.2 D-F). Geometry is the only criterion for discrimination

1. Ribosomes

since non-Watson-Crick base pairs form stable hydrogen bonds often with the same number as Watson-Crick base pairs. Thus, the importance of Watson-Crick base pairing is the relative orientation of the two bases and hence the geometry of the minor groove. Recently the crystal structure of a 70S ribosome bound to EF-Tu and aa-tRNA provided insights to the complete decoding pathway [40]. This structure provided information of the tRNA distortion, which allows the aa-tRNA to interact simultaneously with the decoding center of the small subunit, and EF-Tu at the factor binding site.

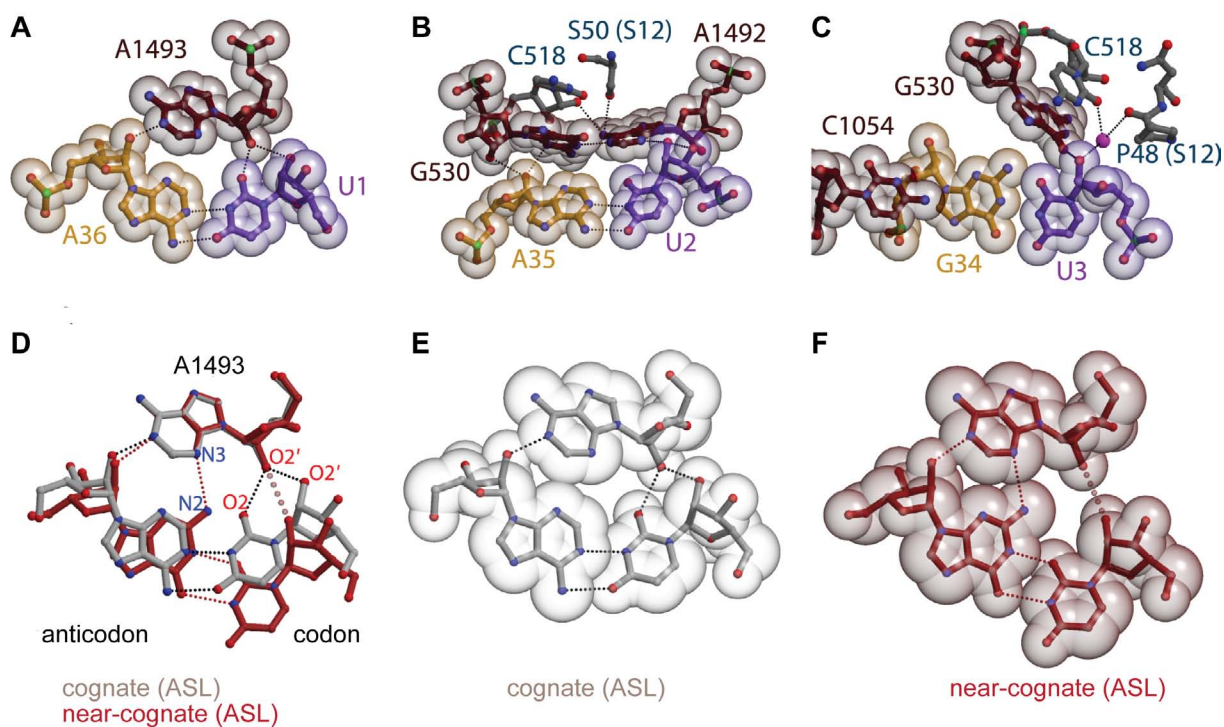


Figure 1.2.2.: **Interactions of the ribosome with the codon-anticodon base pairs.** (A-C) Minor groove interactions of the three positions of the codon-anticodon base pairs. (A) The first position A1493 binds to the minor groove of the A36-U1 base pair. (B) The second position A1492 binds to the minor groove of the A35-U2 base pair. Nucleotide G530 participates at the interaction. (C) The third (wobble) position, showing the G34-U3 base pair. Nucleotide C1054 stacks against G36 of the ASL. U3 interacts with G530, and indirectly through a magnesium ion (magenta) with C518 and residue Pro48 from protein S12 (gray). (D-F) Comparison between cognate and near-cognate codon-anticodon base pairing at the first position. (D) Superposition of the cognate (gray) and near-cognate (red) first base pair. Hydrogen bonding is shown with dashed lines. (E) Cognate A-U codon-anticodon base pair with A-minor interaction of A1493. (F) Near-cognate G-U codon-anticodon base pair with the loss of one hydrogen bonding of the A-minor interaction of A1493. Figures A-C were taken from [38], figures D-F from [39].

1.2.2. Peptidyl transferase center

The catalytic active site of the ribosome is the Peptidyl Transferase Center (PTC), which is located at the center of the large subunit (Figure 1.2.3 A, B). The PTC is entirely composed of RNA [1, 8] and the participating nucleotides of the 23S domain V are > 95% conserved in all three kingdoms of life. The peptide bond is formed when the nucleophilic α -amino group of the aminoacyl-tRNA bound to the A-site attacks the carbonyl carbon of the ester bond linking the peptide moiety of the peptidyl-tRNA bound to the P-site (Figure 1.2.4 A). Although peptide bond formation is catalyzed by RNA residues only, r-proteins L27 and L16 seem to aid the reaction by interacting with the A- and P-site tRNAs. The PTC is one of the major targets for antibiotics in the cell [41]. Some antibiotics inhibit ribosome function when binding to the PTC.

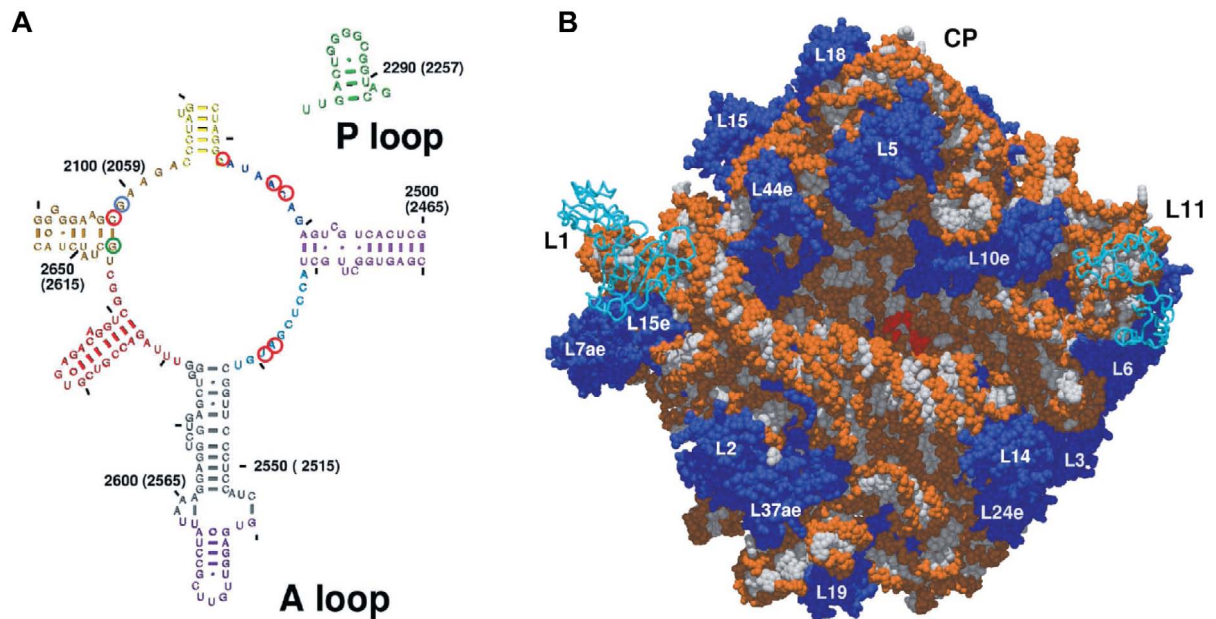


Figure 1.2.3.: **The peptidyl-transferase-center (PTC) in *H. marismortui*.** (A) Secondary structure diagram of the PTC (*E. coli* numbering). (B) Space filling model of the 23S rRNA and r-proteins viewed on the active site. The r-proteins are shown in blue, the 23S rRNA in white (bases) and orange (backbone). The active site is shown in red. The central protuberance is labeled as 'CP'. The figures were taken from [8].

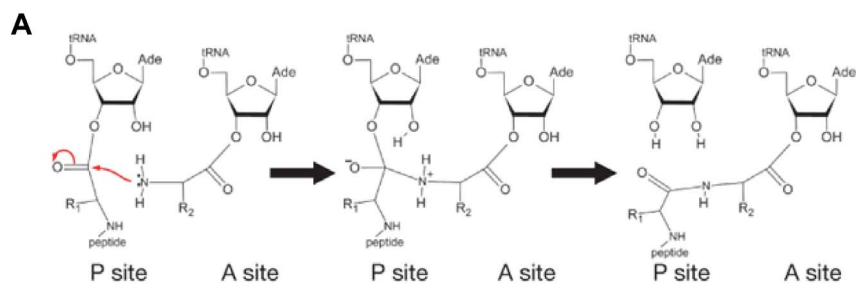


Figure 1.2.4.: **Peptide bond formation.** (A) Schematic drawing of the reaction. The figure was taken from [1].

1.2.3. The ribosomal tunnel and tunnel exit site

The ribosomal tunnel A universal feature of all ribosomes is the ribosomal tunnel. The tunnel connects the PTC with the tunnel exit site (Figure 1.2.5 A) [8]. Its length of 80 – 100 Å is consistent with the length of the nascent polypeptide being protected from proteolytic cleavage [42]. The tunnel is straight and consists of rRNA only, except for a small bended stretch 20 – 30 Å away from the PTC. This constriction site is formed by portions of the r-proteins L4 and L22. The diameter of the tunnel varies from 10 Å at the constriction, to ~ 20 Å at the lower region. The tunnel has an electronegative potential, which is consistent with the dominating RNA environment [43, 44]. For many years the tunnel was assumed to be like a “Teflon wall” and all nascent chains pass through without interacting with the tunnel wall [8]. However, there is evidence that at least for some nascent chains (NC), the tunnel plays an important role. Some leader peptides induce translational stalling in response to the presence or absence of an effector molecule resulting in translation regulation of a downstream gene [45]. There are three well-characterized bacterial leader peptides, namely SecM, TnaC and ErmC, for which mutations in the leader peptide, or in the ribosomal tunnel components can relieve the translational arrest [46, 47, 48]. These data imply direct interactions between specific amino acids within the leader peptide and distinct locations of the ribosomal tunnel. Also for eukaryotes leader peptides causing translational stalling have been described, i.e. the arginine attenuator peptide (AAP), which stalls in the presence of arginine; or the human cytomegalovirus (hCMV) gp48 upstream open reading frame 2 (uORF2) [49].

In principle, the diameter of the ribosomal tunnel would allow the accommodation of α -helical secondary structure formation of the nascent chain. However, folding of small protein domains such as an IgG domain (~ 17 kDa) would be precluded [50, 51]. Fluorescence resonance energy transfer (FRET) studies have indicated a compacting of a transmembrane sequence as it travels through the tunnel, which is consistent with α -helix formation in the tunnel [52]. Independent biochemical analyses provided support for the potential of the NC to adopt compacted or helical conformations in the tunnel and identified specific regions of the tunnel that promote compaction [53, 54, 55, 56]. However, none of these complexes have been visualized, yet.

The ribosomal tunnel exit The ribosomal tunnel exit is clearly an important component of the large subunit and has been well characterized in *bacteria* [50]. In contrast to the tunnel, the tunnel exit site is mainly composed of r-proteins (Figure 1.2.5 B). The tunnel exit site consists of rRNA helices H6, H7, H24, H47 and H50 and the r-proteins L19, L22, L23, L24, L29 and L31e. The tunnel exit site mediates the binding of the ribosomes to factors involved in post- and co-translational protein folding, targeting and translocation. There are two universal adapter sites being characterized to bind to the factors (Figure 1.2.5 C, D). The first universal adapter site involving r-proteins L23, L29 and H7 (Figure 1.2.5 C, D) was shown to bind to SRP54 [57, 58] and Sec61/SecYEG [59, 60] as well as YidC/Oxa1 [61] and trigger factor [62] in *bacteria* (Figure 1.2.5 D). The second universal adapter site is located on the opposite site of the tunnel exit and involves r-proteins L22 and L31e (Figure 1.2.5 C, D). It was shown that the second adapter site binds the putative chaperones RAC (ribosome associated complex) [63], NAC (nascent-polypeptide associated complex) [64], as well as the SRP-Receptor (SR) [65] (Figure 1.2.5 D).

1. Ribosomes

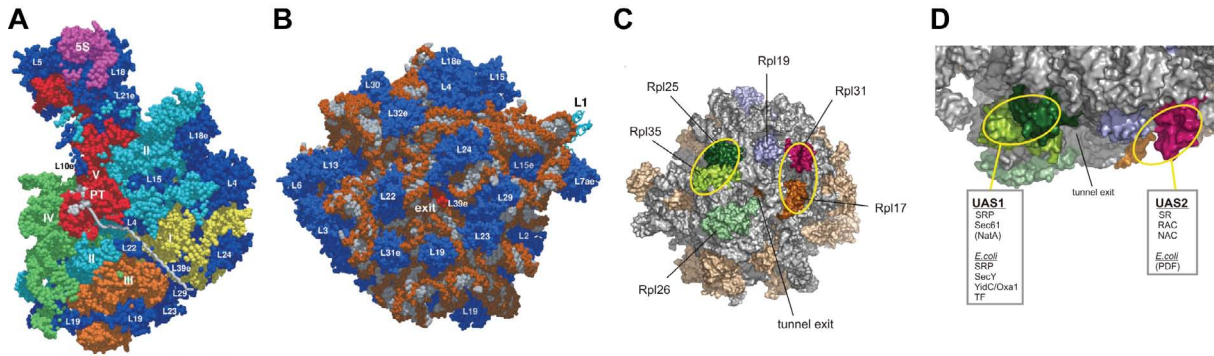


Figure 1.2.5.: **Structure of the ribosomal tunnel and tunnel exit site.** (A) Cut through the *H. marismortui* 50S subunit along the ribosomal tunnel. The space filling representation of the 23S rRNA is colored according to the domains (yellow, cyan, orange, green and red; the 5S in magenta). The r-proteins are shown in blue. The Peptidyl-transferase-center is part of domain V and termed as 'PT'. A model for a polypeptide chain is shown along the ribosomal tunnel in gray. (B) The space filling representation of the 50S subunit at the bottom-view. The r-proteins are shown in blue, the 23S rRNA in white (bases) and orange (backbone). The model of the nascent chain is shown in red. The tunnel exit site is termed as 'exit'. (C) Molecular model of the *S. cerevisiae* tunnel exit site of the large subunit showing the two universal adapter sites. The rRNA is shown in gray, the r-proteins at the exit site are colored differently. Yellow circles highlight the two universal adapter sites. (D) Detailed view on the tunnel exit. The colors are the same as in (C). Figures (A) and (B) were taken from [8]. Figures (C) and (D) were taken from [64].

1.2.4. Translation factor binding site

During translation, protein translation factors play a crucial role. Translation factors transiently bind to the ribosome and thereby promote different steps in protein synthesis. Crystal structures of the bacterial ribosome in complex with elongation factors EF-Tu [40] and EF-G [66], respectively, as well as cryo-EM reconstructions of eukaryotic ribosomes bound to eEF2 [67] show a common binding site located adjacent to the decoding center and shared by the two subunits. Moreover, structural studies of prokaryotic ribosome bound to release factors RF1 [68, 69], RF2 [70, 71, 72, 73] and RF3 [74] as well as prokaryotic ribosome bound to recycling factors RRF [75, 76, 77] suggest a similar interactions of RFs and RRFs with ribosomes and thus provide evidence for a canonical translation factor binding site. Most likely, eukaryotic release factors (eRF) and their homologs involved in mRNA quality control, like the eRF1 homolog Dom34 or the eRF3 homolog Hbs1 and Ski7, are factors binding to the canonical translation factor binding site. All the translocational GTPases (EF-Tu/eEF1 α , EF-G/eEF2, RF3, eRF3 and eRF3 homologs) contact the GTPase-associated center (GAC), which is located near the stalk base (H43 and H44) on the large subunit. The main contacts on the large subunit are made by the highly conserved Sarcin-Ricin-Loop (SRL; H95) as well as r-protein L11 (rpL12). The small subunit interacting partners are the RNA helices h5, the junction between h8 and h14, h15 as well as r-protein S12 (rpS23).

1.3. Evolution of ribosomal RNA

It is widely accepted that the ribosome evolved from a so-called “RNA-world” when proteins did not exist and chemical reactions were catalyzed by RNA [78, 79]. Although the ribosome contains several proteins, the two essential functions, such as decoding and peptide bond formation, are performed by RNA only [6, 8, 1]. Because all ribosomes share a common RNA core structure, it must be formed before splitting of the lineages [80]. Consequently, the primary sequence is not sufficient to study evolutionary aspects, but rather the analysis of the tertiary structure.

Recently, a detailed analysis of the ribosomal large subunit revealed a concerted and modular scheme of early ribosome evolution [81]. In this study, the A-minor interactions (see section 2.3.2) were analyzed in the whole 23S rRNA. Interestingly, in the A-minor interactions that domain V forms with other parts of the 23S rRNA, the acceptor double helix almost exclusively belongs to domain V, while the adenine usually belongs to the rest of the molecule (Figure 1.3.1 A). Thus, it is suggested that this observation reflects the order in which different RNA elements were added to the 23S rRNA as it evolved. In their model they systematically dismantled the large subunit structure through elimination of those RNA elements that are considered to be added recently. Hereby, they suggested that an element could not be a recent addition if its removal compromised the integrity of the remaining parts of the ribosome. If a removed element forms an A-minor interaction with the remaining part of the ribosome, it must contain the single stranded adenine and not the acceptor helix. The systematic dismantling revealed a hierarchical architecture of the 23S rRNA. The removal of 12 generations of acquired elements corresponded to 93% of the original 23S rRNA (Figure 1.3.1 B-D). The remaining part corresponds only to domain V, the central internal loop of which forms the PTC (Figure 1.3.1 B, C). Recently, it was observed that this remaining region consists of two consecutive parts and having almost the same secondary and tertiary structures [82]. These parts are arranged symmetrically to each other and form the binding sites for CCA-3'-end of the tRNAs in the A-site and P-site, respectively. It was suggested that the two parts originate from a duplication of the small RNA fragment since both of them are so similar. From this point of view the evolution of ribosomes started with an initial 110 nts RNA fragment which was able to bind CCA-3'-ends which would later become tRNA molecules. By duplicating the same fragment, the resulting molecules could bind two CCA-3'-ends simultaneously. Within this spatial arrangement of the two adjacent CCA-3'-ends associated with the RNA parts, the first peptide bond could have been formed. Most likely this dimer was already able to synthesize random oligopeptides. This view was supported by *in vitro*-selected RNA molecules resembling the PTC, which could perform a transpeptidation [83] and thus, showing that the transpeptidation reaction only requires RNA. Assuming the pre-existing RNA-world, the ribosomal proteins joined the RNA part in a later stage of evolution to stabilize the inefficiently packed RNAs.

1. Ribosomes

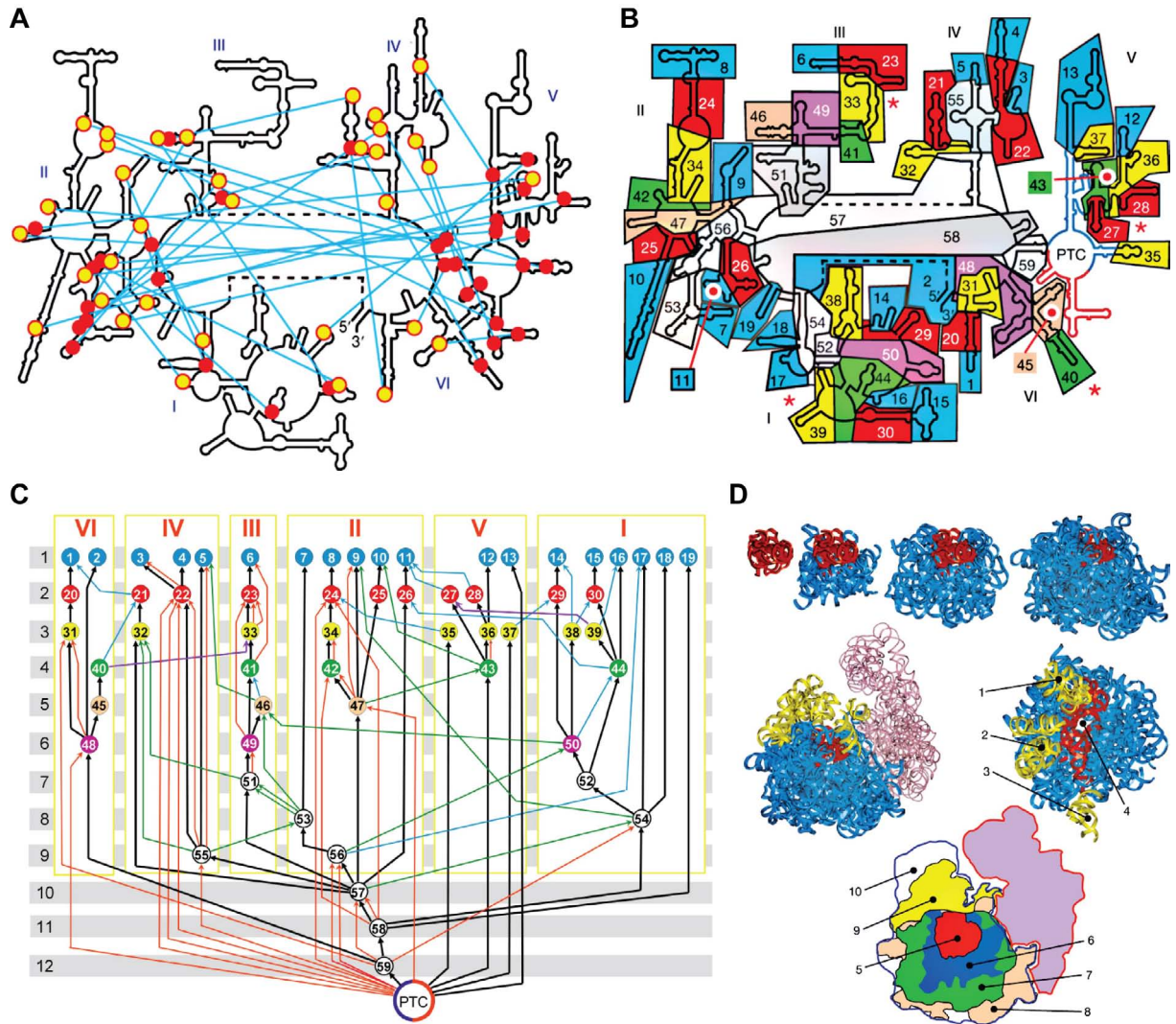


Figure 1.3.1.: **Evolution of ribosomal RNA.** (A) Location of the inter-domain A-minor interactions in the schematic secondary structure of the *E. coli* 23S rRNA. Each A-minor interaction is shown by a cyan line connecting the acceptor double helix (red circle) and the corresponding adenine (yellow circle). Unlike other domains, domain V almost exclusively forms these interactions using double helices and not adenines. (B) Location of the elements in the schematic *E. coli* 23S rRNA secondary structure. Each element has the same color as in (C). The asterisk indicate the non local pseudoknots. (C) Network of the dependencies between the different elements. The number of levels are shown on the left. (D) The aggrandizement of the 23S rRNA structure during its evolution. The proto-ribosome is shown in red with gradually added elements (0 to 59 elements from left to right). The protuberances are colored in yellow and 16S rRNA in purple. The figure was adopted from [81]

2. RNA

Like proteins, several RNAs play an essential role in the cell by catalyzing reactions, sensing gene expression and protein synthesis. In order to perform these functions RNA molecules fold into complex macromolecular structures. The underlying principle for folding is the formation of hydrogen bonds between the RNA bases to form base pairs and stacking. Base pairs and single stranded RNA nucleotides then organize in different structural units (see section 2.2) or form into complex structural motifs (see section 2.3).

2.1. Base pairing

In structured RNAs the majority of nucleotides is base paired in which most of them engage in Watson-Crick base pairing (WC). Besides the classical Watson-Crick base pairing a large fraction of nucleotides assembles in non-Watson-Crick base pairs (non-WC) mediating the long-range intermolecular RNA-RNA interactions. Base pairing is very diverse and almost all combinations of base pairing are observed [84, 85, 86]. In order to characterize base pairs, definitions for an unambiguous and descriptive nomenclature was proposed [85]:

1. The hydrogen bonding between bases is distributed on three edges (Figure 2.1.1 A).
 - a) the **Watson-Crick edge** which presents the canonical Watson-Crick sites
 - b) the **Hoogsteen edge** (purine) or 'C-H' edge (pyrimidine)
 - c) the **Sugar edge** (or shallow groove edge). Base pairing involving the shallow groove edge are also called '*sheared*'. The hydrogen bond is formed with the 2'-OH group (O2') of the sugar moiety which is acting either as donor or acceptor.
2. The orientation of the sugars can either be on the same side or on the opposite sides of a line median to the hydrogen bonds.
 - a) If the sugars are on the **same** side the conformation is called *cis* (Figure 2.1.1 B top).
 - b) If the sugars are on **opposite** sides the conformation is called *trans* (Figure 2.1.1 B bottom).

2. RNA

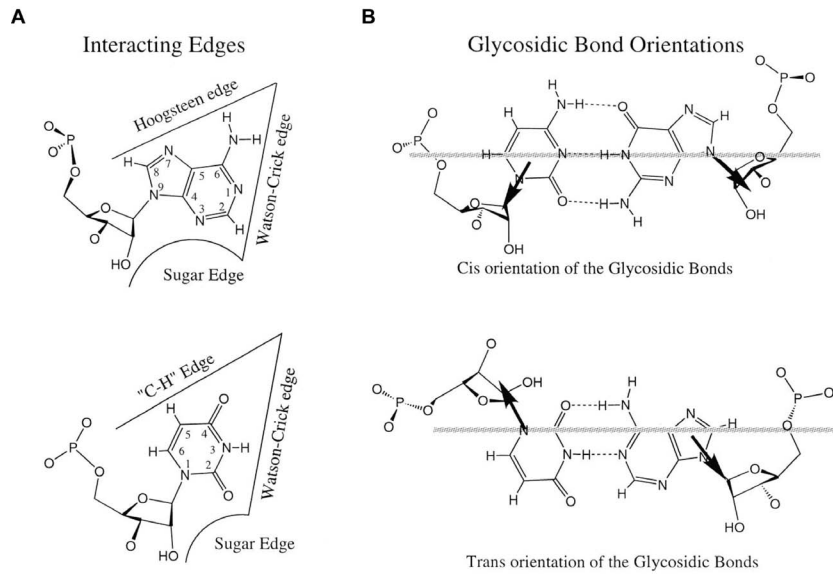


Figure 2.1.1.: **Characterization of RNA base edges and orientation of glycosidic bond.** (A) Definition of the different interacting edges of RNA bases. (Top) Interacting edges of a purine base: The Hoogsteen edge, the Watson-Crick edge, and the Sugar edge. (Bottom) Interacting edges of a pyrimidine edge: the 'C-H' edge, the Watson-Crick edge, and the Sugar edge. (B) *cis* and *trans* orientation of the glycosidic bonds. (Top) *C* – *G* Watson-Crick base pair with *cis* orientation of the glycosidic bond (cWW). (Bottom) *U* – *A* Watson-Crick base pair with *trans* orientation of the glycosidic bond (tWW). The pictures were taken from [85].

Watson-Crick base pairing Standard (canonical) *cis*-Watson-Crick base pairing (cWW) includes the two classical RNA base pairs *A* – *U* and *G* = *C* and the wobble base pair *G*^o*U*. Upon hydrogen bonding the bases interact with their Watson-Crick edge and the orientation of the glycosidic bond is in the *cis* conformation. Stacking of cWW base pairs leads to the RNA A-form double helix (see section 2.2). Note that all combinations of bases are able to form the cWW base pair, however, adapting a different geometry of the base pair.

Besides the cWW base pairing the *trans* orientation also occurs with Watson-Crick-Watson-Crick base pairing.

Non-Watson-Crick base pairing All other base pairs interacting with any combination of two edges (each for one nucleotide) are termed as non-Watson-Crick base pairs. An exhaustive analysis of all the observed edge-to-edge interactions was performed and revealed twelve basic geometry families (including Watson-Crick base pairing) [85]. The twelve families including their proposed symbols for designating each base pair family are shown in Figure 2.1.2.

2. RNA













No.	GLYCOSIDIC BOND ORIENTATION	INTERACTING EDGES	SYMBOL	DEFAULT LOCAL STRAND ORIENTATION
1	<i>Cis</i>	Watson-Crick / Watson-Crick		Anti-parallel
2	<i>Trans</i>	Watson-Crick / Watson-Crick		Parallel
3	<i>Cis</i>	Watson-Crick / Hoogsteen		Parallel
4	<i>Trans</i>	Watson-Crick / Hoogsteen		Anti-parallel
5	<i>Cis</i>	Watson-Crick / Sugar Edge		Anti-parallel
6	<i>Trans</i>	Watson-Crick / Sugar Edge		Parallel
7	<i>Cis</i>	Hoogsteen / Hoogsteen		Anti-parallel
8	<i>Trans</i>	Hoogsteen / Hoogsteen		Parallel
9	<i>Cis</i>	Hoogsteen / Sugar Edge		Parallel
10	<i>Trans</i>	Hoogsteen / Sugar Edge		Anti-parallel
11	<i>Cis</i>	Sugar Edge / Sugar Edge		Anti-parallel
12	<i>Trans</i>	Sugar Edge / Sugar Edge		Parallel

Figure 2.1.2.: **The 12 geometric families.** The 12 families of edge-to-edge base pairs defined by the relative orientation of the glycosidic bonds of the interacting bases. The figure was taken from [85].

Isostericity Like in proteins the three-dimensional structure of homologous RNAs is more conserved than the primary sequence. Point mutations in structurally essential parts of the RNA must be compensated by other mutations. In general, base pairs belonging to the same geometric family share a similar orientation of the glycosidic bonds which implies a similar orientation of the strands in the three-dimensional structure. However, similar base pairs belonging to the same geometric family differ in their C1'-C1' distances significantly. So a random compensatory base mutation (keeping the same edge-to-edge interaction and orientation of the glycosidic bonds) would lead to a distorted three-dimensional structure and maybe to a loss of function.

In order to define a valid compensatory base pair substitution the term of *isostericity* was defined as follows: “When two base pairs display nearly the same C1'-C1' distance and have their glycosidic bonds oriented in the same way, they can replace each other without drastically changing the three-dimensional path and relative geometric orientations of the phosphate-sugar backbones” [85]. For each of the twelve geometric families an *isostericity matrix* was constructed. The isostericity matrix for each family holds the information which base pairs can be substituted by each other while maintaining isostericity.

2.2. Structural organization of RNA

A-form helix The most important structural element in RNA is the so-called A-form double helix [87, 88]. It consists of two or more consecutive cis-Watson-Crick base pairs stacking on top of each other to form a stem-like structure. The cis-Watson-Crick base pairs $A - U/U - A$, $G = C/C = G$ and the wobble base pair G^*U/U^*G are accepted equivalently and can be substituted by each other without disturbing the A-form. In the A-form the major groove is deep and narrow while the minor groove is wide. Although RNA is primarily composed of A-form helices other helical structures of RNA like the left-handed Z-RNA duplex [89] and the L-conformation Spiegelmer RNA duplex [90] have been determined.

2. RNA

Terminal loops A single stranded RNA loop linking the 3'- and 5'-ends of a helix is called terminal loop (or hairpin loop). The majority of terminal loops are tetraloops and pentaloops which fold into determined structural motifs (see section 2.3.5), however, diloops up to tetradecaloops exist [91].

Internal loops Internal loops are formed by two single stranded pieces of RNA that interrupt a pair of A-form helices. The single stranded regions are either of the same length (symmetric) or vary in length (asymmetric) and are usually involved in non-Watson-Crick base pairing. Symmetric internal loops extend the double helical structure by formation of non-Watson-Crick base leading to an alteration of the helical pitch. Asymmetric internal loops have been identified to fold into important tertiary structure formations, called structural motifs (see section 2.3) [92].

Metal ions Due to its phosphate groups the charge of RNA is highly negative. Folding of RNA into a compact structure is very inefficient. Therefore positively charged counter ions promote folding by reducing the repulsion of RNA phosphates [93]. In double-helical base pairing and stacking the relatively hydrophobic nucleotide base is shielded from water while the highly polar and negatively charged phosphates form electrostatic contacts with water and solvent cations [94]. Although Magnesium ions (Mg^{2+}) are considered to be responsible for the stabilization of RNA tertiary structures, sodium and potassium have also been observed to bind and stabilize structured RNA [95]. Besides stabilizing RNA folding it has been suggested that Mg^{2+} directly participates in the catalytic mechanism of ribozymes [96].

Coaxial stacking In 1974, the first medium size RNA crystal structure, a yeast phenylalanine tRNA, was solved [97]. Here the most striking observation was the L-shaped structure of the tRNA formed from its four-helix cloverleaf secondary structure. Each arm of the L-shape was formed by two coaxially stacked helices. On the one hand the acceptor stem coaxially stacks on the T-stem and on the other hand the anticodon stem coaxially stacks on the D-stem [98]. In general coaxial stacking is formed by two separate helices folding into a long pseudo-continuous helix [97, 99]. Coaxial stacking stabilizes the double-helical structure and is a dominating feature in large structural RNAs [3, 7, 99, 97]. It has been suggested that the hydrophobic effect as well as the enhanced van der Waals interactions are the driving force [100]. As seen in the group I ribozyme coaxial stacking can be extended by stacking non-Watson-Crick base pairs and base stacking of innerhelical loops bridging the coaxial helices [99].

2.3. RNA structural motifs

RNA structural motifs are defined as “directed and ordered stacked arrays of non-Watson-Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands” [101]. Moreover, “a given RNA motif comprises all sequences that form the same directed and ordered array of isosteric non-Watson-Crick base pairs and fold into similar, if not essentially identical, three-dimensional structures” [101]. RNA structural motifs play an important role in RNA folding and are often essential for the biochemical function. In order to understand the complexity of RNA folding and tertiary interactions it is necessary to analyze the solved RNA structures. Therefore several high-resolution RNA crystal structures (ssRNA, ribozymes, riboswitches, tRNA, protein/RNA complexes and rRNA) were analyzed and annotated for identifying RNA structural motifs [101, 102, 103]. The data set was scanned for the seven major structural motifs: coaxial helix, A-minor, ribose-zipper, pseudoknot, kissing hairpin, tRNA D-loop/T-loop and tetra-loop tetra-loop receptor as defined in the literature and SCOR database (see section 2.3.1). Interestingly, the motifs are unevenly distributed (Figure 2.3.1).

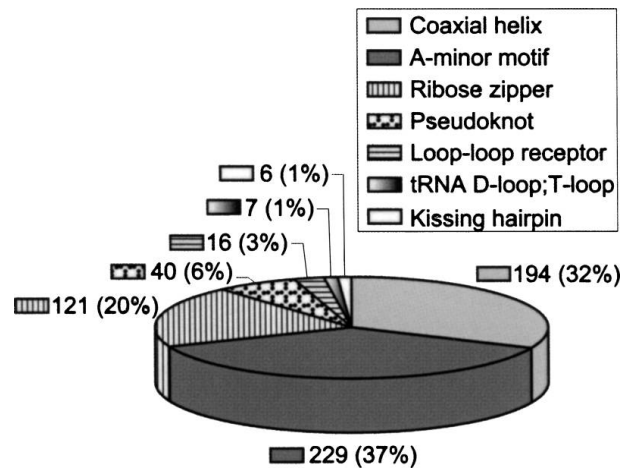


Figure 2.3.1.: **Distribution of RNA structural motifs.** The distribution of RNA structural motifs in a non-redundant data set of 54 high-resolution crystal structures. More than two thirds correspond to the two most abundant motifs: the A-minor motif (37%) and coaxial stacking (32%). The remaining one third is shared by the other five motifs: ribose zipper (20%), pseudoknot (6%), loop-loop receptor (3%), tRNA D-loop/T-loop (1%) and kissing hairpin (1%). The figure was taken from [102].

In most RNA structures such as the ribosome a variety of motifs are used to fold into a compact structure. Statistics of RNA structural motifs are also available on the SCOR database. In the following the structural motifs of RNA occurring in ribosomal RNA (rRNA) will be introduced. These structural motifs are essential for correct assembly and folding of rRNA.

2.3.1. SCOR database

The Structural Classification of RNA (SCOR) database¹ was developed to provide a comprehensive perspective and understanding of RNA structural motifs, function, tertiary interactions and their relationships [91, 104, 105]. The SCOR database provides a survey of three-dimensional RNA motifs extracted from NMR and X-ray RNA structures. The structures are retrieved from the Protein Data Bank² (PDB) and Nucleic Acid Database³ (NDB). The SCOR database includes structural motifs (i.e. kink-turns, S-turns, GNRA-tetraloops) and provides sequence information, descriptions and literature.

2.3.2. A-minor motif

The analysis of the large ribosomal subunit from *H. marismortui* revealed the existence of a significant structural motif called the **A-minor motif**. The interaction involves a single stranded adenine residue, which is inserted into the minor groove of another helix [10]. The adenine interacts primarily with one of the two bases but also forms a hydrogen bond to the other. There are four types of A-minor interactions (Figure 2.3.2 A-D).

Type 0 In type 0 the N3 atom of the adenine (or other) residue forms a hydrogen bond with the O2' of the receptor helix (Figure 2.3.2 A). The A-minor type 0 interaction is observed upon helix packing.

Type I Type I the O2' and N3 atoms of the adenine are inserted in the minor groove of the receptor helix. In type I the adenine is invariant by forming a cis-Sugar-Sugar edge (cSS) and trans-Watson-Crick-Sugar edge (tWS) interaction (Figure 2.3.2 B).

Type II The O2' of the adenine is placed outside of the O2' of the residue from the receptor helix while the N3 of the adenine forms a hydrogen bond in the minor groove. The adenine is invariant and forms a cSS interaction (Figure 2.3.2 C), however, an example of a G-minor type II has been observed in the structure of the S-adenosylmethionine riboswitch (PDB: 2GIS)⁴.

Type III Both the N3 and the O2' of the adenine are outside the nearest O2' atom of the receptor helix. The interaction is weaker than the cSS interaction (Figure 2.3.2 D) and is not an interaction in the minor groove.

¹<http://scor.berkeley.edu/scor.html>

²<http://www.pdb.org/pdb/home/home.do>

³<http://ndbserver.rutgers.edu/>

⁴<http://www.pdb.org/pdb/explore/explore.do?structureId=2GIS>

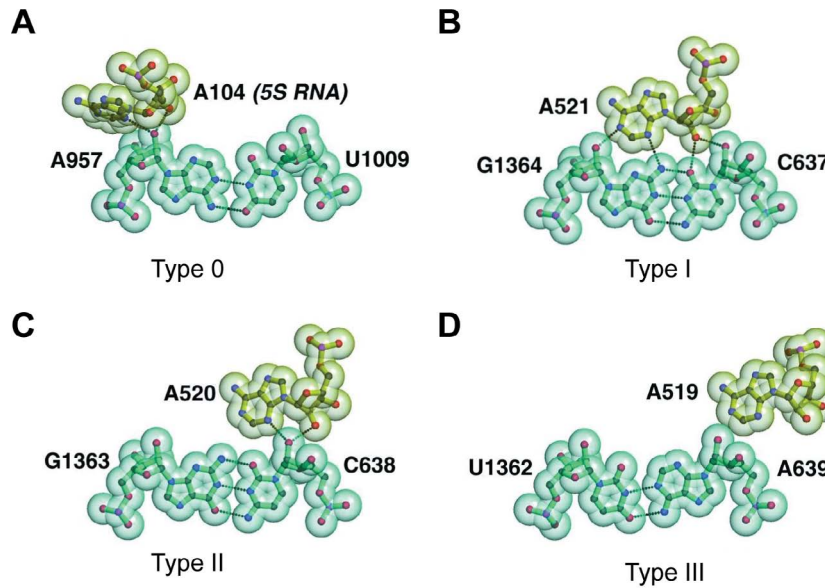


Figure 2.3.2.: **The A-minor interaction.** Examples of the four different A-minor interactions observed in the *H. marismortui* 50S subunit. The pictures were taken from [10].

In the 16S rRNA of *E. coli* the proportions of single stranded adenines is 62% compared to G, C and U with 31%, 29%, and 33%, respectively [106]. For the 23S rRNA from *H. marismortui* 186 of the 721 adenines are involved in an A-minor interaction [107]. It should be mentioned that the A-minor motif is not limited to adenines. Several C-minor and U-minor motif interactions have been observed [85].

A-platforms A recurring motif is the A-platform. Here adenines are positioned side by side in a pseudo-base pair within a helix. In this way the minor groove opens for tertiary interactions.

Function A-minor interactions are more abundant than tertiary base pairs and play a significant role in stabilizing the RNA tertiary structure.

2.3.3. Ribose-Zipper

The ribose-zipper motif is yet another way to pack minor grooves of ribosomal RNA helices against each other. The motif is formed by hydrogen bonding between at least two consecutive backbone ribose 2'-OH from two distant helices [99]. The two chains interact in an antiparallel manner. There are a total of 66 ribose-zippers in the 16S (*T. thermophilus*) and 23S (*H. marismortui*) ribosomal RNA. Moreover, 43 ribose-zippers have been identified to interact with ribosomal proteins [108]. The extended C- and N-terminal extension form hydrogen bonds between arginine and lysine residues and the ribose-zipper and neutralize the negative charge of the RNA backbone. The r-proteins S4, S7 and S20 are known to be involved in ribosome assembly. Binding of these r-proteins to ribose-zippers suggest an essential role of these motifs for ribosome assembly and compaction of the rRNA.

2.3.4. Kink-turn

Kink-turns (k-turns) are asymmetric internal loops that result in sharp bends (kinks) of the backbone in RNA helices. The kink occurs on the minor groove side and so brings the two minor grooves of the adjoining helices together [9, 103]. The canonical strand (C-strand) consists only of Watson-Crick base pairs while the other so-called non-WC stem (NC-stem) is composed of non-Watson-Crick base pairs. Usually, the two conserved adenines that are involved in a tandem-sheared base pair (tHS) with a guanosine. The adenines face their WC-edge and sugar-edge towards the minor groove and so stabilize the C-stem via two A-minor interactions (Figure 2.3.3 A, B). Besides minor groove k-turns there exist k-turns kinking major rather than the minor grooves of the flanking helices together (Figure 2.3.3 C, D) [109].

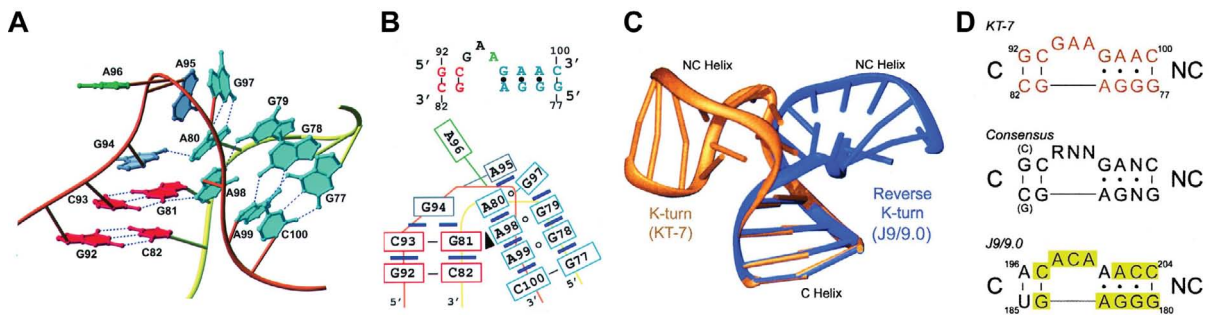


Figure 2.3.3.: **The kink-turn motif.** (A) Structure of the k-turn KT-7 with backbone of the kinked strand in orange and the unkinked strand in yellow. (B) Secondary structure with C-stem in red and the NC-stem in blue (top). Schematic representation of the relative base stacking. The black arrow indicates the A-minor interaction. (C) Structure KT-7 (orange) and reverse kink-turn (blue). (D) Secondary structure of the KT-7 (top) the consensus structure (middle) and reverse kink-turn (bottom). The C-stem and NC-stem are labeled. Figures (A and B) were adapted from [9]. Figures (C and D) were adapted from [109].

Kink-turns are recurrently observed in ribosomes. One k-turn is present in the 16S rRNA of *T. thermophilus* and eight k-turns motifs occur in the 23S rRNA of *H. marismortui* (six are shown in Figure 2.3.4 A, B) [103]. Interestingly all kink-turns appear on the surface of the ribosomal particle. The kink-turn on the small subunit is located at the intersubunit bridge B7.

Function Kink-turns function as an important RNA recognition motif for the ribosomal proteins, however, the way how the r-proteins interact with the different kink-turns varies. Besides recognition of r-proteins k-turns also functional mediate RNA backfolding of helices to stabilize RNA tertiary structure.

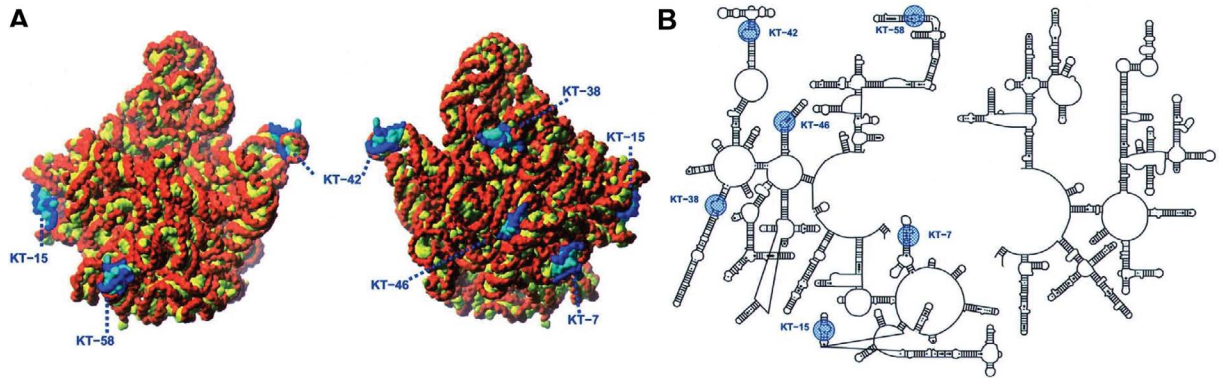


Figure 2.3.4.: **Occurrence of the kink-turn motif in the large ribosomal subunit.** (A) Location of the k-turns in the 50S ribosomal subunit from *H. marismortui*. The ribosome is shown in the crown view (left) and 180° rotated (right). The k-turns are indicated in blue. (B) Secondary structure scheme of the 23S rRNA with k-turns highlighted in blue. The figures were adapted from [9].

2.3.5. Tetraloop motifs

The majority of terminal loops in ribosomal RNA (rRNA) and other structural RNAs consist of four nucleotides, so-called tetraloops. Most of the tetraloops conform to one of the three sequence motifs: GNRA, UNCG or CUUG [110]. These motifs are favored because of their structural stability [110, 111]. The most frequently occurring tetraloop is the GNRA-tetraloop, which has been shown to be involved in tertiary interactions of self-splicing introns, RNase P RNA and ribosomes [112, 113, 114, 110]. In one type of the GNRA-tetraloop interaction the last two purine nucleotides face their Watson-Crick edges towards the minor groove of another RNA helix [112]. Another interaction of GAAA-tetraloop with an 11 nt RNA motif (CCUAAG...UAUGG) was observed in the internal ribosome entry site (IRES) [115]. The second type of the interaction was solved in the X-ray structure of P4-P6 ribozyme domain in *T. thermophila* [99]. Besides of RNA-RNA interaction the GNRA tetraloop is known to be recognized by proteins. The highly conserved Sarcin-Ricin-Loop (SRL) is capped by a GAGA loop and binds to EF-G [116]. Another example is the GAAA tetraloop of h26 of the 16S rRNA which is capped by r-protein S15 and is so protected from chemical modifications [117].

2.3.6. Pseudoknots

Large RNA molecules having extensive tertiary structure often contain pseudoknots. Although several distinct folding topologies of pseudoknots exist, the best characterized is the H type. In the H-type fold, the bases in the loop of a hairpin form intramolecular pairs with bases outside of the stem. In most cases the pseudoknot structure forms coaxial stacked helices. The tight packing of the stems and single stranded regions in the pseudoknot lead to a highly compact structure which is often stabilized by metal ions and tertiary hydrogen bonding. The H-type pseudoknot occurs in the human telomerase (hTR) (see Figure 2.3.5) and is critical for activity [118].

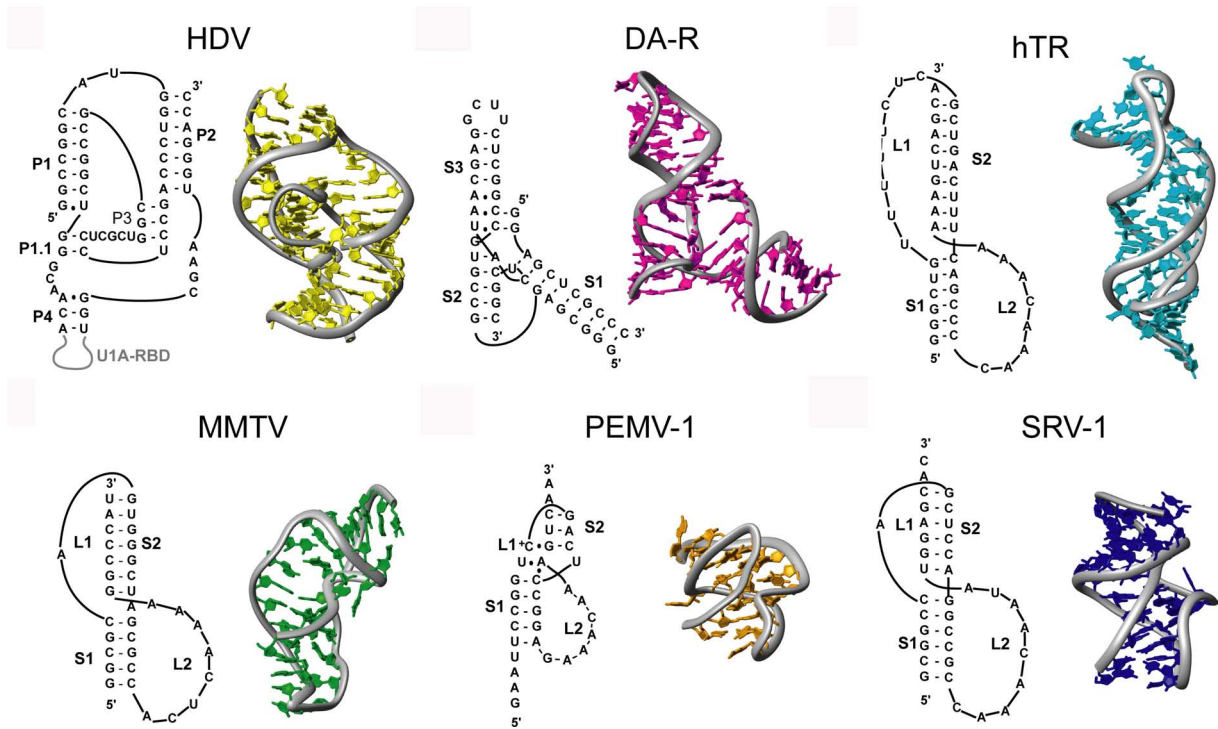


Figure 2.3.5.: **Sequences and structure of RNA pseudoknots.** Pseudoknots from the Hepatitis Delta Virus (HDV), the Diels-Alder ribozyme (DA-R), the Human telomerase (hTR), the mouse mammary tumor virus (MMTV), the Pea enation mosaic virus RNA1 (PEMV-1), and the Simian retrovirus 1 (SRV-1). The figure was taken from [119].

Several biological processes rely on RNAs containing a pseudoknot. For example the bacterial 16S and eukaryotic 18S rRNA contain a pseudoknot at the beginning of the 5'-domain (Figure 9.2.3; 9.2.4). Viruses like the turnip yellow mosaic virus (TYMV) use a pseudoknot structure at the 3' termini to form a tRNA-like motif to infiltrate the host cell transfer RNA-specific proteins [120]. Hepatitis C viruses (HCV) internal ribosome entry sites (IRES) includes a 5'-terminal pseudoknot structure which is essential for viral protein translation initiation [121]. The pseudoknot region of RNase P is one of the most conserved elements in all of evolution.

2.4. RNA-protein interactions

RNA-protein interactions are abundant and diverse in cellular processes like transcription, splicing and translation. A key question in structural biology is how proteins recognize RNA (or DNA) molecules. The structures of the 50S and 30S ribosomal subunits provided a tremendous amount of structural information [3, 7]. In 2007, a data set of 89 non-homologous RNA-protein complexes has been analyzed [122]. The analysis revealed a number of different binding sites and binding motifs. Van der Waals interactions are much more prevalent than hydrogen bonds. Moreover RNA-protein interactions via the RNA backbone occur more frequently than specific contacts with a nucleotide base. As expected arginine and lysine residues, both positively charged, have a clear propensity to bind to the negatively charged phosphate groups of the backbone. The non polar residue tryptophan has a propensity to stack with double ringed nucleotide bases. Surprisingly, glycine residues seem to have a preference to bind guanosine.

RNA-protein interactions in ribosomes In ribosomes, RNA-protein interactions play a significant role in stabilizing the whole complex and so maintaining the function of the ribosome. Ribosomal proteins interact with rRNA in four different types: i) recognition of base functional groups that are exposed in the minor groove of an RNA helix, ii) recognition of base groups via the major groove of a distorted RNA helix, iii) recognition of bulged or flipped out nucleotide bases and iv) insertion of amino acid residues into hydrophobic binding pockets in RNA [11]. Minor groove recognition is the most common type of sequence specific RNA-protein interaction, due to the accessibility of the minor groove in A-form helices. Interactions of proteins via the major groove is only observed when the major groove is widened up due to non-Watson-Crick base pairs. The arginine and lysine residues in the buried N- and C-terminal extensions of r-proteins interact with the negatively charged phosphates of the RNA backbone.

The interaction of r-proteins with RNA also plays an essential role during co-translational assembly of the ribosomal subunits. The assembly is coupled to transcription and assures each domain to fold before transcribing the next [123]. For example, cooperative binding of r-proteins S4, S17 and S20 to the transcribed 5' domain of 16S rRNA ensures the formation of a stable structure of the 30S body before transcribing and folding of the central domain.

2.5. RNA structure prediction and modeling

Studying the 3D structure and dynamics of RNA is important to understand its function in the cell. Experimental techniques such as X-ray crystallography, NMR spectroscopy and cryo-electron microscopy are used to derive the 3D coordinates. Unfortunately all techniques are time consuming and expensive. Moreover, the complexity and flexibility of RNA molecules makes it more complicated to determine the 3D structures. Thus, the gap between known RNA 3D structures and RNA sequences increased dramatically. This encourages the approach to predict the 3D coordinates of RNA molecules with computational methods. Although significant progress on RNA secondary structure prediction has been made in the past several years, the prediction of an RNA 3D structure from its sequence has to follow. Thus, the determination of RNA 3D structures has to combine experimental data (observations) with computational approaches (algorithms) [124].

Prediction of RNA 3D structures Much progress has been made in the field of automated RNA structure prediction and the prediction of small and simple RNA structures is feasible [125]. However, large RNA complexes with several helices and/or pseudoknots can not be predicted accurately. At present, the automated 3D prediction tools reported by Baker [125] and Major [126] are either limited to 50-nt sequences or still require improvement in non-helical regions and tertiary (long-range) interactions. Especially the tertiary interactions, that are formed by non-Watson-Crick base pairs, are important for stabilizing the entire 3D fold. Predicting those interactions require an intensive analysis and annotation of structural motifs. To assist RNA 3D

2. RNA

structure prediction, several tools have been developed and applied. Extensive 3D modeling tools developed by Westhof [127] and Harvey [128, 129, 130] require manual application and expert knowledge [102]. ERNA-3D⁵⁶ can automatically generate coordinates for A-form helices given a secondary structure input. In 2005, ERNA-3D was used to build a 3D structure of the transfer-messenger RNA (tmRNA) of *E. coli* [131]. RNA2D3D⁷ can generate a first-order approximation of the actual 3D model but needs interactive editing of nucleotides to remove structural clashes and manipulates tertiary interactions [132]. RNA2D3D was used to build the pseudoknot domain of wild-type human telomerase RNA [133].

Another approach to predict RNA 3D structures is to use the information of homologous structures that have been solved already by experimental methods. MC-Sym⁸ builds RNA 3D structures by using coordinates and relationships between residues that are extracted from known 3D structures. Structural constraints derived from crosslinking-assays can be applied to the building procedure [126]. The model is then refined using molecular mechanics. MC-Sym was used to build several *de novo* models including the structure of the hairpin ribozyme catalytic core [134]. The software package S2S 1.0 (sequence to structure) has been designed to construct and manipulate multiple sequence alignments based on a known RNA 3D structure [135]. The manipulations are constrained by the rules of RNA structure folding. Besides studying the sequence and secondary structure similarities to infer similarities in the tertiary structure the graphical tool Assemble [136] was developed based on the functionality of MANIP⁹ [127].

Limitations of automated predictions Computational methods predicting an RNA 3D structure automatically from its sequence are still very limited. Especially large RNA molecules like those of ribosomes can not be predicted by any automated method. Besides the limitation of the input size, the problem becomes even more complex if one considers that the predicted structure is not necessarily the functional active conformation. Moreover, RNA folding is highly dependent on its environment such as water, ions or proteins. Thus, another issue is the limitation of accurate force-field and energy terms used in molecular mechanics or molecular dynamics.

None of the above mentioned tools is able to include experimental data such as X-ray, NMR or cryo-EM data [124]. A combination of experimental data and computer-based methods is necessary to derive more accurate atomic detail of the RNA 3D structures.

⁵http://www.molgen.mpg.de/~ag_ribo/ag_brimacombe/ERNA3D/ERNA-3D.html

⁶<http://www.erna-3d.de/>

⁷<http://www-lmmb.ncifcrf.gov/~bshapiro/rna2d3d/rna2d3d.html>

⁸<http://www.major.irc.ca/MC-Sym/>

⁹http://www-ibmc.u-strasbg.fr/arn/Westhof/them4_West.html

3. Cryo-electron microscopy

Cryo-electron microscopy (Cryo-EM) is a method for obtaining 3D reconstructions of macromolecules from 2D images of a transmission electron microscope. In the last years, cryo-EM has been established as a technique to get structures of biological complexes at subnanometer resolution allowing the visualization of protein and RNA secondary structure. This is mainly due to significant improvements at the level the microscope itself, new algorithms and software as well as the dramatically increased computational power [137]. In contrast to X-ray crystallography (X-ray) and nuclear magnetic resonance (NMR) spectroscopy, cryo-EM is generally not limited by the size and the functional state of the macromolecular complex [138]. Complexes of 150 kDa size up to multi-component complexes like ribosomes can be visualized. In contrast to X-ray crystallography, cryo-EM samples are recorded under native conditions using vitreous ice. Moreover, tiny amounts of a homogeneous sample (μg range; in contrast to mg range in X-ray and NMR) are sufficient to study the complex.

3.1. Cryo-EM and single particle analysis

The general assumption of this approach is that every image is taken from the same homogeneous sample (Figure 3.1.1 A). The sample (including its buffer) is applied on an EM grid, which is covered by a holey carbon (Figure 3.1.1 B). In order to get the sample over the holes, an additional thin layer of carbon is provided and made hydrophilic by glow discharging (Figure 3.1.1 B middle). The EM grid is then blotted and quickly frozen in liquid ethane. As a result a thin layer of vitreous ice is formed embedding the particles (Figure 3.1.1 B bottom). In an ideal case the particles have a random distribution of orientations in the ice layer. The vitrified EM grid is then inserted into a microscope and kept under liquid nitrogen temperatures (Figure 3.1.1 C). A data set of 2D images is collected under low dose conditions using $\sim 20e^- - 25e^-/\text{\AA}^2$ and is recorded either on CCD or film. In case of film the recorded images have to be scanned. The single particles on the images (micrographs) have to be identified, boxed out, centered and normalized.

3. Cryo-electron microscopy

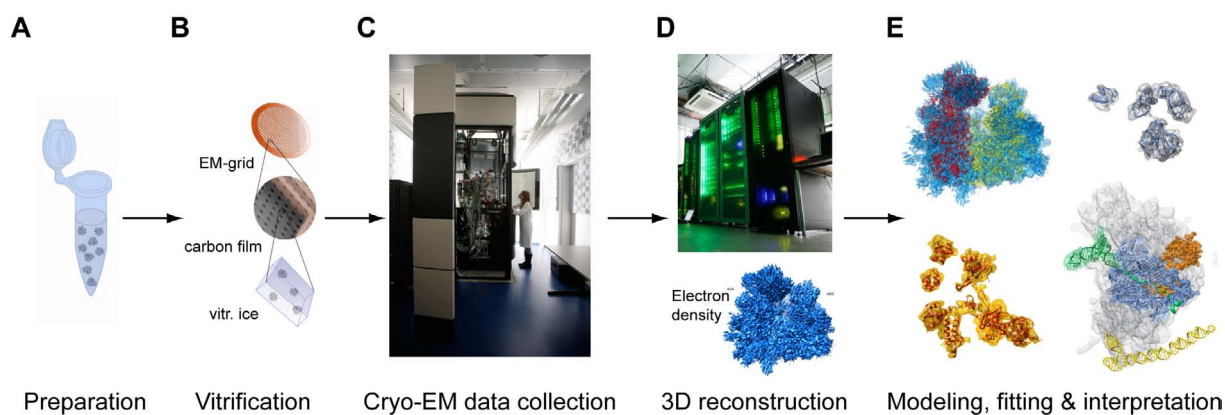


Figure 3.1.1.: **Workflow of Cryo-EM.** (A) Homogeneous sample in buffer. (B) EM grid (top), Carbon film with holes (middle) and Vitrified ice layer (bottom). (C) Electron microscope. (D) Computer cluster for data processing. (E) Molecular models with electron densities for interpretation.

The next step is the alignment of the particles, i.e. determination of the particle orientation. In case of ribosomes, the projection matching technique is applied. Here, each particle is aligned against reference projections with known Eulerian angles and scored for similarity (cross-correlation). The particle is then classified to the reference's class with the highest similarity score. The result of the alignment is the assignment of the three Eulerian angles and the x-, y-shifts to each particle. In the following step the angular information of each particle is used for calculating the backprojection of the 2D images into a 3D reconstruction. The alignment and backprojection as well as a refinement is applied iteratively for several rounds in order to improve the accuracy of parameters resulting in improved resolution of the reconstruction. The computations, especially the alignment, is highly intensive and requires several hundreds to thousands of CPU-hours on a cluster (Figure 3.1.1 D).

The resulting 3D reconstruction is then used for fitting and docking of atomic structures. This step involves the inspection and interpretation of the structure (Figure 3.1.1 E). The interpretation is the crucial step and biological information or function can be concluded.

3.2. Limitations of cryo-EM

Cryo-EM with single particle analysis has become a successful approach in structural biology despite not providing atomic resolution. Some important limitations are necessary to be mentioned.

Limitation of resolution In the recent years, a number high resolution cryo-EM structures have been reported. Near-atomic resolution of highly symmetric particles like the 14-fold symmetric chaperone GroEL complex [139] and the rotavirus particle [140] could be visualized at a resolution of 4 Å and 3.8 Å, respectively. As an example for asymmetric particles the ribosomes could be visualized at 6.1 Å [141]. Theoretically a better imaging is possible [142], however, in most

3. Cryo-electron microscopy

applications the resolution is the limiting factor. Achieving near-atomic resolution (below 4 Å) in single particle analysis requires millions of particles to be averaged. One issue is the quality of the EM images where the signal-to-noise ratio decreases dramatically at resolutions beyond ~ 10 Å. Another issue arises from conformational and heterogeneity of the sample.

Conformational heterogeneity and flexibility Cryo-EM reconstructions of heterogeneous data sets originating from 2D projections of different conformations or flexible parts of the complex result in weak or even loss of density due to the averaging of the picture information [143, 144].

To tackle this problem approaches of supervised and unsupervised sorting have been established [145]. Supervised sorting, also called classification, requires knowledge about the 3D object. In the refinement process the data set is sorted, or classified, against two references (i.e. two references of different conformations). Each of the 2D projections (experimental images) is matched against the reference projections and classified according to the class with higher similarity (cross-correlation). Classification can lead to overfitting or bias due to the given reference. In order to overcome reference bias, the references are only offered in the first round of refinement, followed by several rounds of independent refinement based on the 3D reconstructions of the first round.

4. Aim of this study

The main focus of this work was to construct a molecular model for the eukaryotic ribosomal RNA including the RNA expansion segments and structural variable regions. To be able to model RNA molecules that large, the first aim was to develop a new approach for large scale RNA modeling, which could also include empirical data such as cryo-EM density maps. This approach should include methods for highly accurate RNA alignments, for homology modeling of conserved regions. Regions having no structural template should be modeled *de novo* and fitted into cryo-EM density maps. The final 80S model, including rRNA and r-proteins, should be used for an all-atom molecular dynamics flexible fitting simulation into the cryo-EM map. Together with the eukaryotic ribosomal proteins this thesis should provide a structural inventory for all expansion segments and their interaction partners.

Further the model should be used to analyze functionally important sites of the ribosome by studying intermediates of co-translational translocation (RNC-PCC complexes), no-go mRNA decay (RNC-Dom34/Hbs1 complex) and protein folding (helix formation in the ribosomal tunnel).

Part II.

Materials and Methods

5. Cryo-Electron microscopy

Sample preparation and data collection The *Triticum aestivum* Ribosome-Nascent-chain-Complexes (RNCs) were prepared from a home-made wheat germ *in vitro* translation system [146]. Further the RNCs were purified according to [147]. The samples were applied on carbon-coated holey grids. The micrograph images were recorded on a Polara field emission gun electron microscope at 300 kV (FEI). The defocus range was between 1.0 and 4.5 μm . The micrographs were then scanned on a Heidelberg Primescan D8200 drum scanner. The pixel size was 1.24 \AA .

Data processing The contrast transfer function (CTF) of the data was determined using CTFFIND¹ [148]. The particles on the micrographs were picked automatically with SIGNATURE² [149]. Visual inspection and manual sorting against false positive particles with WEB³ [150] resulted in a data set of 2.108.230 *T. aestivum* RNC particles. The data set was further processed using SPIDER⁴ [150]. A supervised sorting [144] was applied to classify the data set into subsets of un-programmed (without P-tRNA) and programmed (with P-tRNA) particles. The resulting subset of programmed ribosomes contained 1.362.920 particles. The data processing was mainly done by Jean-Paul Armache.

Data formats The cryo-EM maps were converted from SPIDER format into BRIX format with SPIDER [150, 151]. The following EM densities file formats were used for modeling and converted from BRIX with mapman⁵ [152]:

Tool	EM density data format
Chimera	BRIX, MRC, CCP4
Coot	CCP4
Assemble	X-PLOR, MRC
PyMol	CCP4
VMD	BRIX, CCP4, MRC

Isolated densities For each model ES and VR the corresponding EM density was isolated using the “colorzone”-tool⁶ of Chimera [153]. The maps were splitted using radii between 4.5 \AA - 6.5 \AA around the model. Parts of the resulting map were zeroed out interactively using the “Volume eraser”-tool⁷ of Chimera. The resulting maps were stored in BRIX format.

¹<http://emlab.rose2.brandeis.edu/software>

²<http://emlab.rose2.brandeis.edu/software>

³http://www.wadsworth.org/spider_doc/web/docs/web.html

⁴http://www.wadsworth.org/spider_doc/spider/docs/spider.html

⁵http://www.csb.yale.edu/userguides/datamanip/uppsala/manuals/mapman_man.html

⁶<http://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/colorzone/colorzone.html>

⁷<http://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/voleraser/voleraser.html>

6. Templates and sequences

6.1. Ribosomal RNA sequences

The rRNA sequences of the *S. cerevisiae* 5S, 5.8S, 18S and 25S were taken from GenBank Accession number (Acc.) U53879¹. The rRNA sequence for the *T. aestivum* 5S (Acc. X06094²), 5.8S (Acc. FM998894³), 18S (Acc. AY049040⁴) and 28S (Acc. AY049041⁵) rRNAs were available, with the exception of five and four nucleotides at the 5' end of the 18S and 28S, respectively, and 65 nucleotides (487 – 551) in the 28S corresponding to *ES7^L*, which were filled with the corresponding sequences of *O. sativa* (Acc. M11585⁶). In summary, only 74 of the 5485 rRNA nucleotides in the *T. aestivum* 80S were from *O. sativa*, and the remaining 5411 (98.7%) were from *T. aestivum*. Moreover, sequence alignments between the available *T. aestivum* and *O. sativa* rRNAs indicated a 98% sequence identity without consideration of isosteric base pairing, indicating the suitability of using the *O. sativa* sequence for filling the missing 74 (1.3%) nucleotides in the *T. aestivum* model.

6.2. Templates for the large ribosomal subunit

The crystal structures of the 50S subunit and the 5S rRNA of *H. marismortui* (PDB 1FFK⁷ [7], PDB 1VQ8⁸ [154]) were used as a template for the structure-based sequence alignment between the *H. marismortui* 23S rRNA, the *S. cerevisiae* 28S rRNA and *T. aestivum* 28S rRNA. The structure-based sequence alignment between the *H. marismortui* 5S rRNA template and the 5S sequences of *S. cerevisiae* and *T. aestivum* was done separately. For regions like H5-H6-H7 and the stalk base the X-ray structures of *E. coli* (PDB 2QAM⁹ [155]) and of *T. thermophilus* (PDB 2J01¹⁰ [14]) were used.

¹<http://www.ncbi.nlm.nih.gov/nuccore/1262303>

²<http://www.ncbi.nlm.nih.gov/nuccore/21661>

³<http://www.ncbi.nlm.nih.gov/nuccore/FM998894>

⁴<http://www.ncbi.nlm.nih.gov/nuccore/AY049040>

⁵<http://www.ncbi.nlm.nih.gov/nuccore/AY049041>

⁶<http://www.ncbi.nlm.nih.gov/nuccore/169818>

⁷<http://www.pdb.org/pdb/explore/explore.do?structureId=1FFK>

⁸<http://www.pdb.org/pdb/explore/explore.do?structureId=1VQ8>

⁹<http://www.pdb.org/pdb/explore/explore.do?structureId=2QAM>

¹⁰<http://www.pdb.org/pdb/explore/explore.do?structureId=2J01>

6.3. Templates for the small ribosomal subunit

The crystal structure of the 30S of *T. thermophilus* (PDB 1J5E¹¹ [3]) was used as a template for the structure-based sequence alignment between the *T. thermophilus* 16S rRNA, the *S. cerevisiae* 18S rRNA and *T. aestivum* 18S rRNA. For h6 and h33 the X-ray structure of *E. coli* (PDB 2QAL¹² [155]) and of *T. thermophilus* (PDB 2J00¹³ [14]) were used.

6.4. Template for tRNA

The crystal structure of the peptidyl-tRNA of *T. thermophilus* (PDB 2J00¹⁴ [14]) and the crystal structure of aspartic acid and phenylalanine tRNA of *S. cerevisiae* (PDB 3TRA¹⁵ [156]) were used as a template for the structure-based sequence alignment between the *T. thermophilus* tRNA and the tRNA sequences of *S. cerevisiae* and *T. aestivum*.

6.5. Templates for structural motifs

The crystal structures of the following RNA loop motifs were extracted from the PDB¹⁶ and incorporated into Assemble [136]:

Loop class	Motif sequence and PDB identifier
Triloop	AUA (PDB 1ESH) [157], UCU (PDB 1GID) [99]
Tetraloops	AGNN (PDB 1AFX) [158], CAUU (PDB 1BIV) [159], CUYG (PDB 1NJP) [160], GNRA (PDB 10W9) [161], UAUU (PDB 2QAM) [155], UCAC (PDB 1JJ2) [9], UGNN (PDB 1AFX) [158] UNCG (1DK1) [162]
Pentaloop	(PDB 1EHZ) [163]
Hexaloop	AACCAUC (PDB 1HS1) [164]
Heptaloop	GAUGGUU (PDB 2BJ2) [165]
Octaloop	UUCAUAG (PDB 1J5A) [166]
Nonaloop	CUAGUAACA (PDB 1JJ2) [9]
kink-turn	GAA/AAUGU(PDB 3CC2) [167]
innerhelical motif	AAG/AGG (PDB 2QAL) [155], AA/AG (PDB 2QAL) [155]

Table 6.1.: **Structural motifs for RNA loops.** The sequence of the template loop as well as the structural template PDB file are shown. These loop motifs were integrated into Assemble.

¹¹<http://www.rcsb.org/pdb/explore/explore.do?structureId=1J5E>

¹²<http://www.rcsb.org/pdb/explore/explore.do?structureId=2QAL>

¹³<http://www.pdb.org/pdb/explore.do?structureId=2J00>

¹⁴<http://www.pdb.org/pdb/explore.do?structureId=2J00>

¹⁵<http://www.rcsb.org/pdb/explore/explore.do?structureId=3TRA>

¹⁶<http://www.rcsb.org/pdb/home/home.do>

7. Modeling

The modeling workflow was developed in collaboration with Prof. Dr. Eric Westhof and Dr. Fabrice Jossinet (University of Straßburg, France). Modeling the ribosomal RNA was splitted into RNA homology modeling and RNA *de novo* modeling. Parts that could be aligned to a structural template were modeled by homology according to their aligned template. The remaining non-aligned parts were modeled *de novo* in the cryo-EM map. Finally, the constructed models were refined and flexibly fitted into the cryo-EM map.

7.1. RNA homology modeling

In order to build reliable RNA homology models of the ribosomal core structure, a new approach of RNA homology modeling was developed. Here a structure-based sequence alignment between a structural template of the bacterial/archaeal ribosomal subunits and the eukaryotic rRNA sequence were generated manually (see section 8.1.1). Aligning only base pairs that fulfill the isostericity criterion ensured very accurate RNA homology models with correct hydrogen bonding.

7.1.1. Structure-based sequence alignment

The structure-based sequence alignments of both, the 18S and the 28S rRNA, were done using the structural templates (see section 6.2 and section 6.3). These multiple sequence alignment (MSA) for the 5S and 28S were constructed manually between *H. marismortui*, *T. aestivum* and *S. cerevisiae* and for the 18S between *T. thermophilus*, *T. aestivum* and *S. cerevisiae*, respectively. Here, base pairs were only aligned if the base pair substitution was isosteric to the one in the template (Figure 8.0.1 blue aligned parts). Nucleotides not involved in any base pairing were aligned. The resulting alignments of the 5S, 18S and 28S rRNA were then used to deduce 3D homology models. The aligned regions correspond to the common ribosomal core structure. The core structures were generated using S2S [135] (see section 7.1.2) and Assemble [136] (see section 7.2.3). The non-aligned regions correspond to the expansion segments (ES) and variable regions (VR) and were modeled as described in section 7.2. The stalk base (H43-H44), the L1 stalk (H76-H78) and the region of H5-H6-H7 were aligned separately using the *E. coli* X-ray structure (PDB 2QAM¹ [155] and PDB 3FIK² [168] [13]).

¹<http://www.pdb.org/pdb/explore/explore.do?structureId=2qam>

²<http://www.pdb.org/pdb/explore/explore.do?structureId=3fik>

7.1.2. S2S

The Sequence-to-Structure tool (S2S)³ [135] was used for constructing alignments. The construction of all alignments was done manually. In order to generate homology models for the ribosomal core, the tool was further developed and extended in collaboration with Dr. Fabrice Jossinet and Prof. Dr. Eric Westhof. The resulting core models for *S. cerevisiae* and *T. aestivum* were deduced automatically from the alignments and stored as Assemble objects. The Assemble objects were then loaded into Assemble (see section 7.2.3) for refinement and exported as file in PDB-format⁴.

7.2. *De novo* modeling of RNA

The *de novo* modeling was done for all the ribosomal expansion segments and variable regions. The unfolded rRNA sequences were loaded into secondary structure prediction services. According to the secondary structure prediction the expansion segments were modeled into the cryo-EM density semi-automatically using Assemble (see section 7.2.3) [136]. The initial models were refined in Assemble and subsequently fitted flexibly using MDFF (see section 7.3.2) [168].

7.2.1. Secondary structure prediction of RNA

Besides the secondary structure predictions from CRW database⁵ [27] the primary sequences of the expansion segments were used as an input for RNA secondary structure prediction tools RNAfold⁶ [169] and RNASHapes⁷ [170]. The secondary structure predictions with RNASHapes were done using shape types 1-5 and setting the maximum loop length to 20. For visual inspection and comparison between the predictions the structure graphs were generated.

7.2.2. Tertiary structure modeling of RNA

The ribosomal core model was used as an anchor point for modeling the expansion segments. The RNA molecules were loaded in FASTA format⁸ and secondary structure was automatically generated with RNAfold. In some cases the RNAfold predictions were modified according to the RNASHapes predictions. The helical regions were generated automatically as an A-form helix and were roughly fitted into the cryo-EM map. Then single stranded regions were placed and connected to the helical parts. Structural motif for loops, inner helical non-Watson-Crick base pairing and kink-turns were applied manually in Assemble. The models were refined using the geometric refinement tool and including all Watson-Crick and non-Watson-Crick base pairs. A subsequent all-atom refinement and flexible fitting is described in section 7.3.

³<http://www.bioinformatics.org/s2s>

⁴<http://www.wwpdb.org/docs.html>

⁵<http://www.rna.cccb.utexas.edu/>

⁶<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>

⁷<http://bibiserv.techfak.uni-bielefeld.de/rnashapes/>

⁸<http://www.ncbi.nlm.nih.gov/blast/fasta.shtml>

7.2.3. Assemble

The Assemble tool⁹ [136] was used to automatically generate the coordinates for the ribosomal core models of the 5S, 5.8S, 18S and 28S rRNA from the alignment files. The core models were refined with the embedded refinement tool for 20 iterations.

All *de novo* models for ES and VR were build in Assemble semi-automatically into the isolated ES/VR cryo-EM densities. Therefore, Assemble was adapted to use MRC or X-PLOR format for electron densities. The embedded RNAview algorithm¹⁰ [171] was used to analyze all kinds of base pairing. All expansion segments and variable regions were refined with the embedded refinement procedure for 20 iterations. Subsequently, the coordinates were exported as file in PDB-format¹¹

The PDB files for the structural motifs (see section 6.5) were parsed in Assemble. The extracted motifs were stored in the local Assemble motif-database.

7.3. Fitting and refinement of RNA models

A preliminary rigid body fitting of the RNA templates and the RNA models was done in the absence of proteins using the 'fit-in-map'-command¹² of Chimera [153]. The fit was performed into low-pass filtered electron densities. The core models were further fitted flexibly using MDFF (see section 7.3.2). Then all RNA segments were merged using VMD [172]. Single nucleotide insertion in the core structure were then analyzed using density maps.

In a subsequent process the r-proteins were combined with the rRNA models. Steric clashes between different RNA segments as well as RNA and proteins were solved using IMD (see section 7.3.2 and section 7.3.2). The entire 80S model was refined using MDFF (see section 7.3.2).

7.3.1. Assemble refinement

The initial RNA models for all ES and VR were refined using the internal geometric refinement tool. All ideal base pairs were identified automatically with RNAview¹³ [171]. In case of non-ideal base pair distances the interaction type was annotated manually in the 2D viewer of Assemble. Single hydrogen bond interactions were ignored and deleted in the 2D viewer. The models were then refined geometrically for 20 iterations.

7.3.2. Molecular Dynamics Flexible Fitting (MDFF)

MDFF¹⁴ is a method to flexibly fit molecular models into electron densities. Within the fitting process the stereochemistry of the models are preserved for RNA and proteins. MDFF uses RNA restraints to preserve RNA secondary structure and tertiary structure interactions because all simulations were performed without water and ions. Therefore, the exact hydrogen bonding of all the 96 observed base pair interactions were annotated manually (according to

⁹<http://www.bioinformatics.org/assemble/>

¹⁰<http://ndbserver.rutgers.edu/services/download/>

¹¹<http://www.wwpdb.org/docs.html>

¹²<http://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/fitmaps/fitmaps.html>

¹³<http://ndbserver.rutgers.edu/services/download/>

¹⁴<http://www.ks.uiuc.edu/Research/vmd/>

[173, 85]). RNAView was used to identify all the base pair interactions in the whole complex. All Watson-Crick and non-Watson-Crick base pair interactions, except the 'sugar edge/sugar edge' interactions, were restraint in the simulations. The simulations were carried out using NAMD 2.6¹⁵ [174], VMD 2.8.7a44¹⁶ [172], CHARMM¹⁷ [175], RNAView¹⁸ [171] and IMD¹⁹ [176]. All molecular dynamics (MD) simulations and molecular dynamics flexible fitting (MDFF) were performed in collaboration with Dr. Elizabeth Villa.

File preparation

The RNA models, which have been previously refined in Assemble, were used as an input for MDFF. For each RNA model the function 'autopsf'²⁰ in VMD was used to generate a 'psf file' and an extended pdb file including hydrogens. The models were then renumbered according to the right sequence numbers using the script 'renumber_offset.tcl'. Then all the ES and VR were merged with the RNA core models. Subsequently, the RNA models were merged with the proteins.

Interactive Molecular dynamics (IMD)

IMD²¹ was used to solve structural clashing residues. Two residues (either RNA-protein or RNA-RNA) are clashing if two atoms (from each residues) are closer than 2 Å. The script 'fix_clashes.tcl' was used for an exhaustive identification of RNA-RNA clashes and RNA-protein clashes. All clashes were solved manually. A local IMD simulation was performed to achieve stereochemically correct models. For these local simulations an environment of 5 Å around the clash was selected (command 'same residue as within 5 of clash'). A preceding 2000-step minimization was performed to subsequently simulate a MD run for 2 ps.

IMD was also used for simulating the movement of *ES27^L* (see section 9.3.2.3) and H43-H44 (see section 10.4) of the large subunit.

¹⁵<http://www.ks.uiuc.edu/Research/namd/>

¹⁶<http://www.ks.uiuc.edu/Research/vmd/>

¹⁷<http://www.charmm.org>

¹⁸<http://ndbserver.rutgers.edu/services/download/>

¹⁹<http://www.ks.uiuc.edu/Research/vmd/imd/>

²⁰<http://www.ks.uiuc.edu/Research/vmd/plugins/autopsf/>

²¹<http://www.ks.uiuc.edu/Research/vmd/imd/>

Global MDFF simulation of the ribosome

An all-atom MDFF simulation of the 80S ribosome was performed including all 78 proteins, 1800 nts of the SSU rRNA, 3554 nts of the LSU rRNA, 75 nts of the tRNA and 21 nts of the mRNA. The first simulation incorporated the EM-map which was filtered between 5–6 Å. For the second MD simulation an EM-map filtered between 5 – 5.5 Å was used.

For the global MDFF simulation of the complete ribosome including the tRNA and the mRNA all separate pdb files had to be merged with the *'autopsf'*-command²² in VMD. The resulting *psf* and *pdb* had a special VMD format²³ because the complex had more than 100'000 atoms which would be invalid for the PDB format²⁴.

7.4. Visualization and structure analysis

The cryo-EM maps and models were visualized using Chimera²⁵ [153, 177], VMD²⁶ [172], Coot²⁷ [178] and PyMol²⁸. The rRNA secondary structure maps were derived from the Comparative RNA website (CRW)²⁹ [27]. The rRNA secondary structures of *S. cerevisiae*, *T. aestivum* and *O. sativa* were adjusted using Corel Draw X3³⁰. The pictures for the expansion segment inventory were edited with Adobe Photoshop³¹ and Adobe Illustrator³².

²²<http://www.ks.uiuc.edu/Research/vmd/plugins/autopsf/>

²³<http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node21.html>

²⁴<http://www.wwpdb.org/docs.html>

²⁵<http://www.cgl.ucsf.edu/chimera/>

²⁶<http://www.ks.uiuc.edu/Research/vmd/>

²⁷<http://www.ytbl.york.ac.uk/~emsley/coot/>

²⁸ <http://www.pymol.org/>

²⁹<http://www.rna.ccbb.utexas.edu/>

³⁰<http://www.corel.com>

³¹<http://www.adobe.com>

³²<http://www.adobe.com>

Part III.

Results

8. Workflow for modeling and fitting large RNA molecules

Modeling ribosomal RNA (rRNA) has a total input size of about 5500 nts (120 nts 5S rRNA, \sim 1800 nts 18S rRNA and \sim 3500 nts 28S rRNA). However, no protocol for large-scale modeling of RNA has been established, to date (see section 2.5). Therefore a new modeling workflow was developed to model the rRNA of *S. cerevisiae* and *T. aestivum* 80S ribosomes. The development was done in collaboration with Prof. Dr. Eric Westhof and Dr. Fabrice Jossinet (University of Straßburg, France). All the implementations for S2S and Assemble were done by Dr. Fabrice Jossinet.

This new approach combines novel modeling techniques with a novel flexible fitting into cryo-EM density maps. In general the modeling effort is splitted into homology modeling and *de novo* modeling depending on the presence of alignable structural templates. RNA parts that can be aligned to a structural template (Figure 8.0.1 top) are modeled by homology (Figure 8.0.1 bottom left; see section 8.1). Stretches of RNA that can not be aligned to a structural template are modeled *de novo* (Figure 8.0.1 bottom middle; see section 8.2). Finally, the constructed models are refined and fitted into the cryo-EM map (Figure 8.0.1 bottom right). So far each modeling step needs manual input.

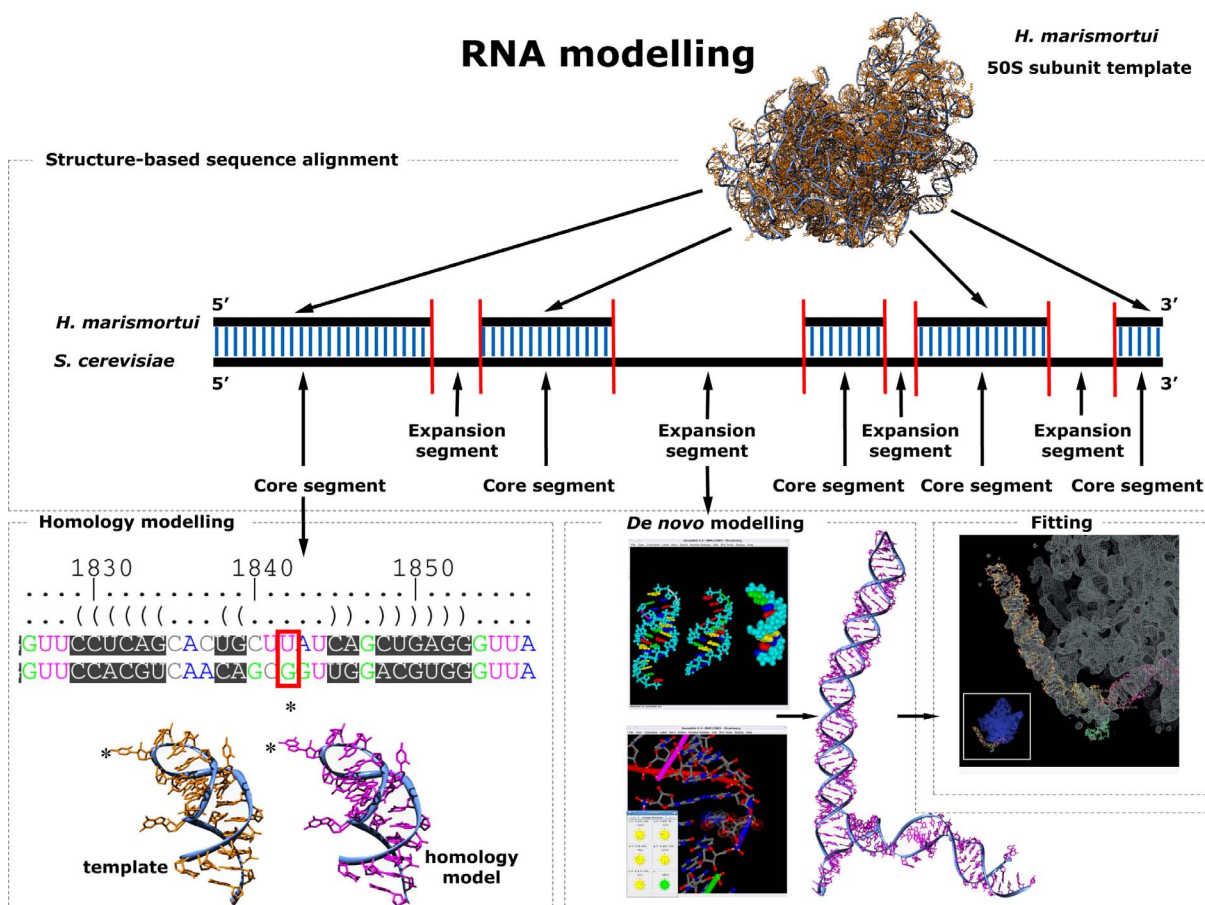


Figure 8.0.1.: **Scheme for modeling RNA in 3D using cryo-EM maps.** (**Top**) The RNA sequence is aligned to the structural template. (**Bottom left**). The RNA parts that can be aligned to a template are modeled by homology. (**Bottom middle**) The RNA parts that are not aligned are modeled *de novo*. (**Bottom right**) If an electron density is available the RNA will be modeled into it. A refinement and flexible fit is done using Molecular Dynamics Flexible Fitting (MDFF).

8.1. RNA homology modeling

The idea of RNA homology modeling is based on protein homology modeling. For proteins the essential criteria for a template-based modeling is the degree of *sequence similarity/sequence identity*, which is determined by the sequence alignment. For RNA the sequence similarity criteria is not applicable since co-varying base substitutions and isosteric base pair substitutions do not necessarily have sequence similarity. Moreover, similar structural motifs are not conserved in terms of sequence similarity and/or sequence length.

The first step in this approach is to perform a structure-based sequence alignment with S2S (see section 7.1.2; section 8.1.1). Based on this structure-based sequence alignment (Figure 8.1.1 A) the homology model for the aligned regions is deduced from the template (Figure 8.1.1 B; see section 8.1.2). The homology model is generated from the alignment automatically followed by

a subsequent refinement.

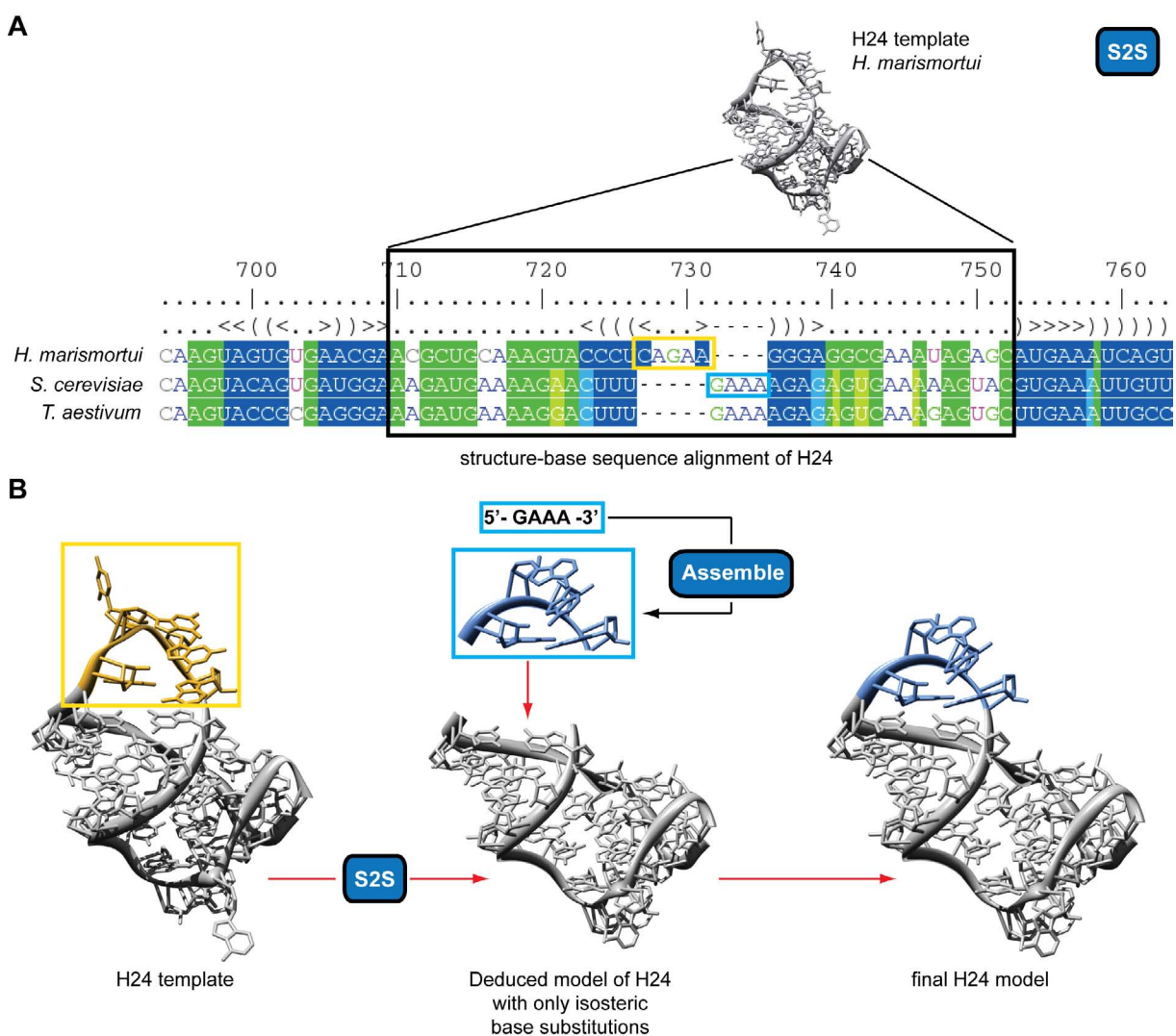


Figure 8.1.1.: **Example for the RNA homology modeling of H24.** (A) Structure-based sequence alignment between the template of *H. marismortui* and the sequences of *S. cerevisiae* and *T. aestivum*. The region for H24 is highlighted. Aligned residues of secondary structure keeping the same isostericity class like the template are shown in blue. Aligned residues of tertiary interactions keeping the same isostericity class like the template are shown in green. (B) Crystal structure of *H. marismortui* H24 (template) with the pentaloop highlighted in yellow (left). Homology model deduced from S2S of the aligned parts of H24 for *S. cerevisiae* (gray). The non-aligned GNRA-tetraloop (blue) is modeled separately in Assemble (middle). Refined model of H24 for *S. cerevisiae* (right).

8.1.1. Structure-based sequence alignment

The structure-based sequence alignment between the template and the sequence of interest is generated manually using S2S (see section 7.1.2). The template is loaded with its 3D coordinates and the secondary structure is annotated automatically by RNAview [171]. Now the secondary structure of the target is used to align the target sequence to the template. In the process a base pair is only aligned if the base pair substitution fulfills the isostericity criterion (see section 2.1; Figure 8.1.1 A; see section 2.1). The resulting structure-based sequence alignment is then used

to deduce the coordinates of the aligned residues (see section 8.1.2).

8.1.2. Deducing the homology model

Deducing a homology model from the template works in three steps:

1. **Backbone:** For each aligned residue the sugar-phosphate backbone coordinates will be copied from the template to the model.
2. **Base:** Upon an identical base the coordinates for the base itself is copied. Otherwise, the coordinates for the substituted base are generated automatically by keeping the same glycosidic bond orientation (Figure 8.1.1 B left panel). The generation of the coordinates is done by Assemble.
3. **Refinement:** After generating the coordinates the homology model is refined automatically in Assemble (see section 7.2.3). By changing the bases of an isosteric base pair with a subsequent automatic generation of coordinates, minor suboptimal distances between atoms (clashes) can occur. These errors are solved by the refinement.

8.1.3. Homology modeling of structural motifs

Non-aligned residues of loops or innerhelical motifs are folded separately and automatically based on known structural motifs (Figure 8.1.1 B middle panel; see section 6.5). Therefore, Assemble was adapted with a repository of structural motifs that can be extended by the users. The structural motif is then manually applied to the target sequence. Minor refinement is necessary if the target sequence differs from the template sequence.

8.2. RNA *de novo* modeling

RNA parts lacking any structural template had to be modeled *de novo*. This new *de novo* modeling approach can be seen as a 5-step procedure:

1. **Input:** The primary sequence of the target is used as an input for Assemble. The input format is the FASTA-format¹ (Figure 8.2.1 A).
2. **Secondary structure assignment:** Assemble was adapted to predict an RNA secondary structure by RNAfold [179] however, different algorithms or an existing secondary structure can be used as an input (Figure 8.2.1 B-D). In addition to the secondary structure, additional Watson-Crick base pairs and non-Watson-Crick base pairs can be assigned manually. All assigned base pairs will be taken into account within the refinement. The target secondary structure is organized in modules which are either RNA A-form helices or RNA single stranded region connecting the helices.
3. **Generation of A-form helices:** The construction of the 3D model starts with the automatic generation of the A-form helices (Figure 8.2.1 E, G). The helices are generated with the ideal torsion angles for A-form helices, however, all torsion angles can be adjusted manually.

¹<http://www.ncbi.nlm.nih.gov/BLAST/fasta.shtml>

8. Workflow for modeling and fitting large RNA molecules

- Folding of structural motifs:** Single stranded modules are preliminary generated as an A-form helix strand, but have to be adjusted manually (Figure 8.2.1 F, H). Upon the user's input the structural motifs are applied to the unfolded module. Subsequently, the RNA torsion angles can be adjusted arbitrarily.
- Incorporation of electron density maps:** An important new feature of this approach is the ability to incorporate electron density data in Assemble (Figure 8.2.1 J). This enables the model to be based on empirical data. The RNA modules can be fitted manually into the density and the strands can be connected.

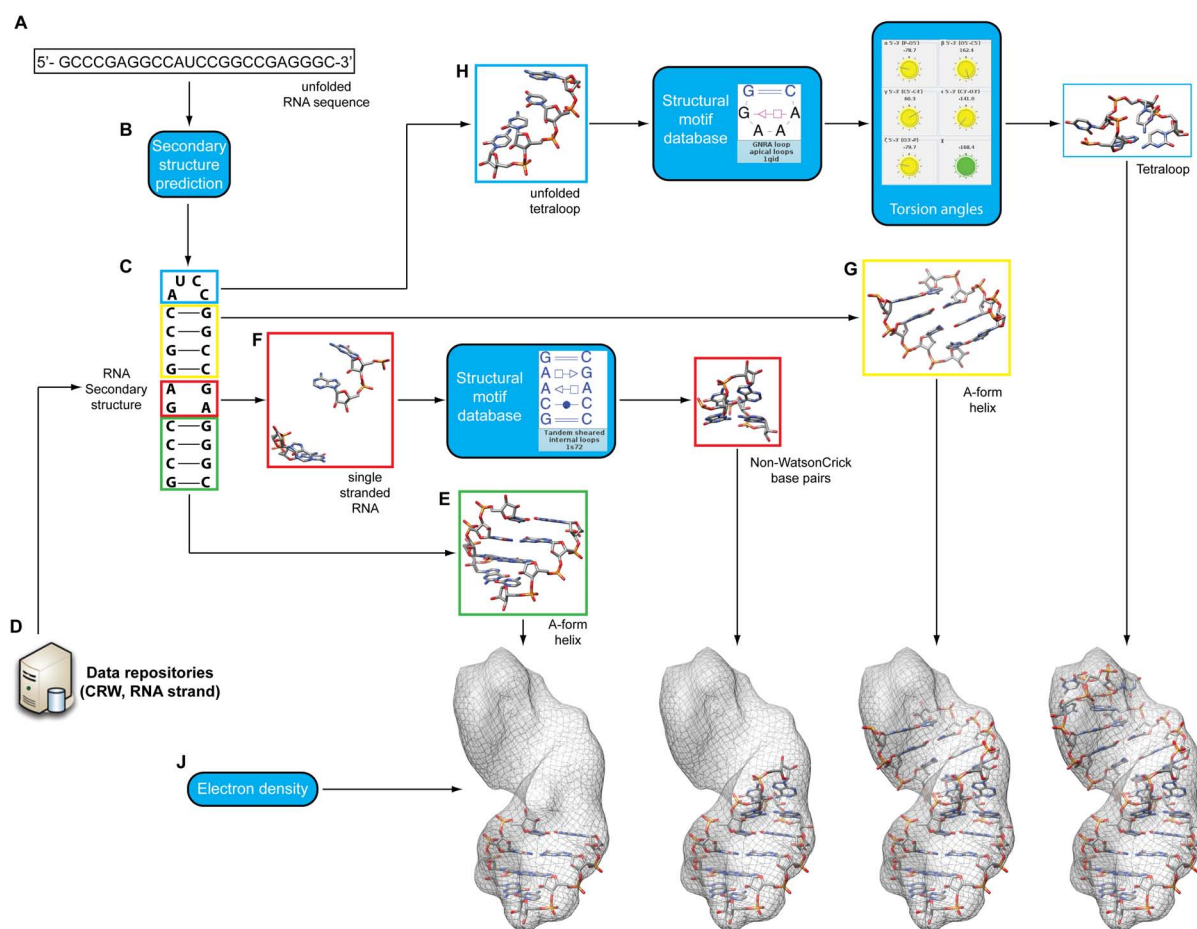


Figure 8.2.1.: **Example of RNA *de novo* modeling and fitting of expansion segment ES3^L.** (A) Primary sequence input for Assemble. (B) Secondary structure prediction by RNAfold or other algorithms. (C) Secondary structure prediction output. (D) External data that can be incorporated in the secondary structure. (E, G) Automatic generation of RNA A-form helices modules. (F) Generation of single stranded RNA followed by applying a structural motif. (H) Generation of an unfolded tetraloop followed by applying the GNRA-tetraloop motif and torsion angle adjustment. (J) Incorporation of electron density maps. The RNA modules can be modeled and fitted simultaneously.

8.3. Refinement and flexible fitting of RNA

Refining the molecular RNA models is a two-step procedure.

1. A geometric refinement within Assemble. Here the residues are optimized according to the ideal atom distances. Moreover non-Watson-Crick base pairs are refined to their ideal isostericity class (C1'-C1' distance and orientation of the glycosidic bond).
2. The second step will be applied if an electron density is present. This new approach combines an RNA refinement while flexibly fitting the molecular model into the electron density. For this approach the protocol of Molecular Dynamics Flexible Fitting (MDFF) [180, 168] was extended in collaboration with Dr. Elisabeth Villa. Therefore all 96 observed base pairs (according to reference [85, 173]) were annotated and each hydrogen bonding scheme included in the MDFF protocol.

The simulation is performed on all atoms including hydrogens. Water molecules coordinated by a base pair are not included. In all RNA simulations, neither solvation shells of RNA base pairs nor ions are included, resulting in simulations performed in vacuum. Therefore the simulations are restrained to maintain base pairing and secondary structure. In order to keep the planarity of bases and base pairs a script applies the ideal angles and distances to the base or base pair.

Another feature of this new approach is the ability to flexibly fit large complexes consisting of both RNA(s) and protein(s) simultaneously. This enables MDFF simulations of large complexes like ribosomes, spliceosomes or chromatin remodeling complexes.

Simulations of motions on molecular levels The RNA implementations for MDFF were applied to Interactive Molecular Dynamics (IMD)². A targeted flexible fitting of molecular models can be achieved by interactively applying forces to atoms and residues. In this way motions between different conformational states can be simulated.

This method can be also used for a novel user driven refinement and fitting. With this method the user can resolve interatomic clashes by interactively manipulating atoms or residues.

²<http://www.ks.uiuc.edu/Research/Categories/IMD/>

9. Model of *T. aestivum* and *S. cerevisiae* 80S ribosomes based on high-resolution cryo-EM structures

9.1. Cryo-EM reconstruction

The cryo-EM map at 6.1 Å resolution of the *S. cerevisiae* 80S RNC-Ssh1 complex [141] (Figure 9.1.1 A; EMDB ID 1669¹) was used for modeling. Further a cryo-EM reconstruction of an 80S yeast ribosome stalled by a synthetic stem loop mRNA in complex with the no-go decay factors Dom34 and Hbs1 (Figure 9.1.1 B) [181] enabled a more precise model of the 40S subunit head region, because the binding of these ligands cause a more stable 40S head conformation.

As described in section 5 the 2.108.230 of *T. aestivum* 80S ribosomes particles were sorted into un-programmed and programmed ribosomes. The subset of 1.362.920 programmed particles was used for the final CTF-corrected 3D-reconstruction and yielded a resolution of 5.5 Å, according to 0.5 FSC criterion (Figure 9.1.1 C).

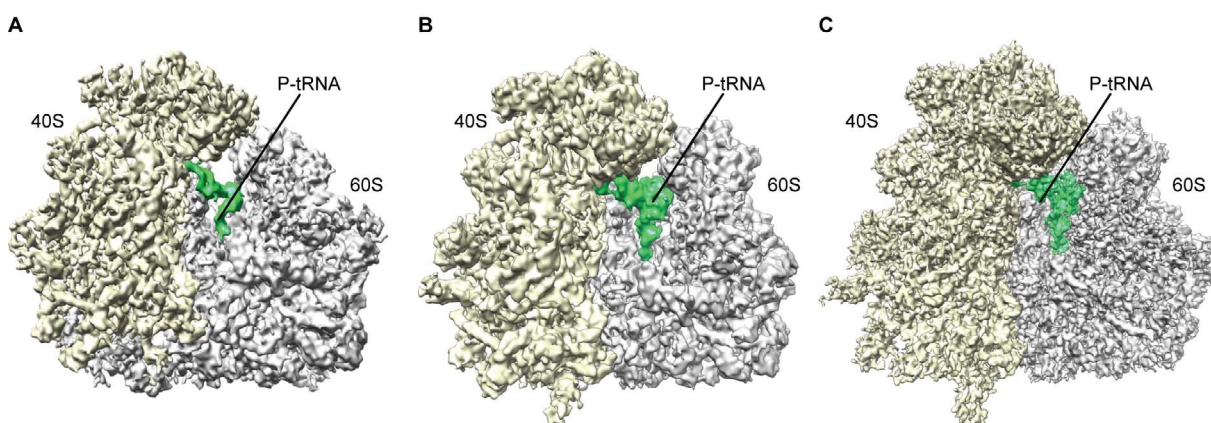


Figure 9.1.1.: **Cryo-EM reconstruction of eukaryotic ribosomes used for modeling.** (A and B) *S. cerevisiae* and (C) *T. aestivum* 80 ribosomes, with small (40S) and large (60S) subunits colored in yellow and gray, respectively and the P-tRNA in green. The figure was taken and adapted from [182].

¹http://www.ebi.ac.uk/pdbe-srv/emsearch/atlas/1669_summary.html

Characteristics of the *T. aestivum* map The final cryo-EM map of *T. aestivum* at 5.5 Å shows characteristics comparable to the early crystal structures of the 50S large subunit of *H. marismortui* [183] at 5 Å and the structures of the 30S small subunit and 70S ribosome of *T. thermophilus* [12, 184] at 5.5 Å. At this resolution, the double-stranded RNA can be observed as a smooth helical density, with a defined minor and major groove. The phosphate groups of the RNA backbone are represented as distinctive bumps in the density (Figure 9.1.2 A). In most cases, single stranded RNA can be traced (Figure 9.1.2 B) and bulged out nucleotides are resolved. Neither the different RNA bases and base modifications nor the type of base pair interactions are distinguishable, which is not expected at this resolution.

As protein secondary structure elements, α -helices are observed as rod-like densities (Figure 9.1.2 C) and β -sheets are represented by smooth and flat surfaces (Figure 9.1.2 D). The pitch of α -helices and the separation of β -sheets are not visible, which is also not expected at this resolution. The cryo EM density map of *T. aestivum* 80S-RNCs has been deposited in the 3D-Electron Microscopy Data Bank (EMDB²) with the EMDB ID: 1780³.

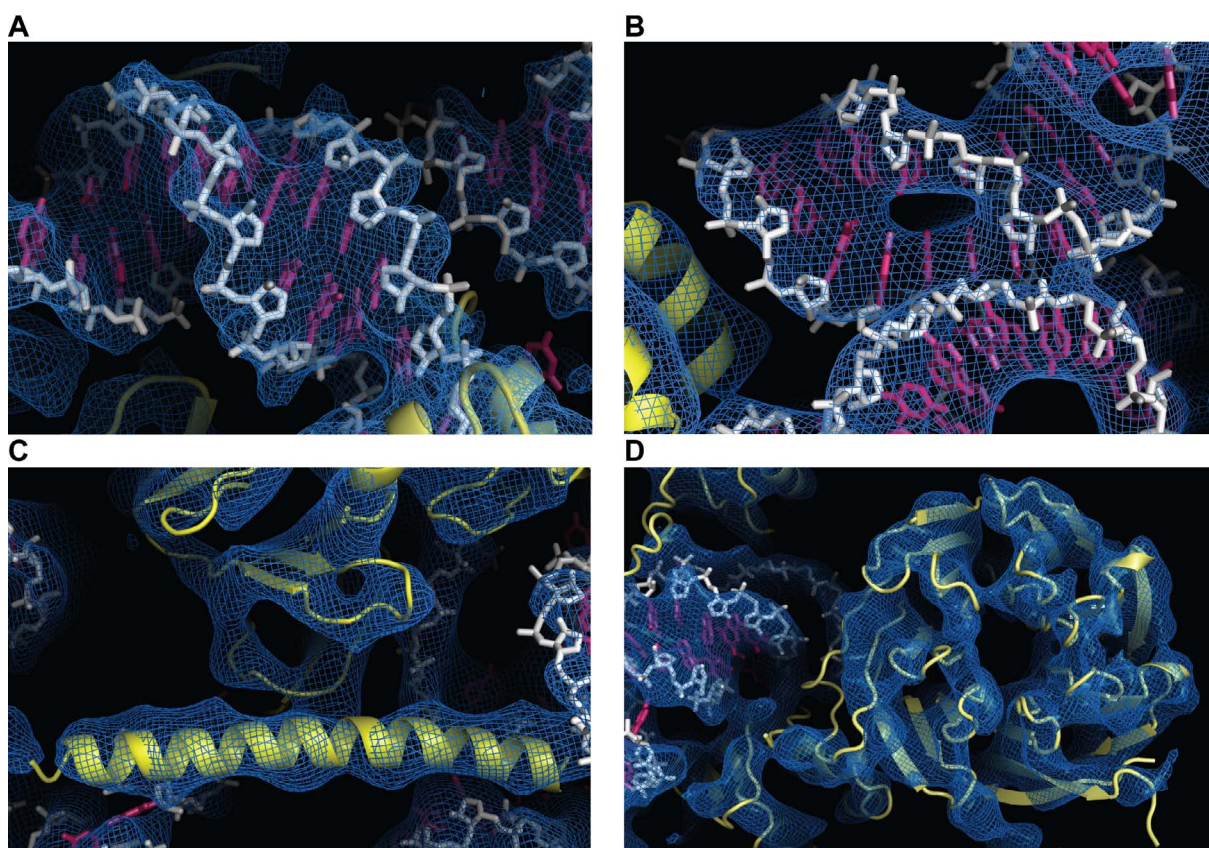


Figure 9.1.2.: **Features of the *T. aestivum* 80S cryo-EM map at 5.5 Å.** Selected views of the *T. aestivum* 80S density map (blue mesh) and the corresponding molecular model, with r-proteins in yellow and rRNA in white (backbone) and red (bases). The figure was taken from [182].

²<http://www.ebi.ac.uk/pdbe/emdb/>

³http://www.ebi.ac.uk/pdbe-srv/emsearch/atlas/1780_summary.html

9.2. Near complete model of the eukaryotic 80S ribosome

Eukaryotic 80S ribosomes are significantly larger than their bacterial counterparts, the *T. aestivum* ribosome contains 1.53 MDa (0.62 MDa/40S and 0.91 MDa/60S) of r-protein and 1.74 MDa (0.56 MDa/40S and 1.18 MDa/60S) of rRNA, thus totaling 3.27 MDa, whereas *E. coli* 70S ribosomes is in total 2.5 MDa (0.9 MDa/30S and 1.6 MDa/50S).

The molecular models are presented for translating *T. aestivum* (Figure 9.2.1 A) and *S. cerevisiae* (Figure 9.2.1 B) 80S ribosomes encompassing $\sim 98\%$ of the rRNA and 92.5% of the r-proteins.

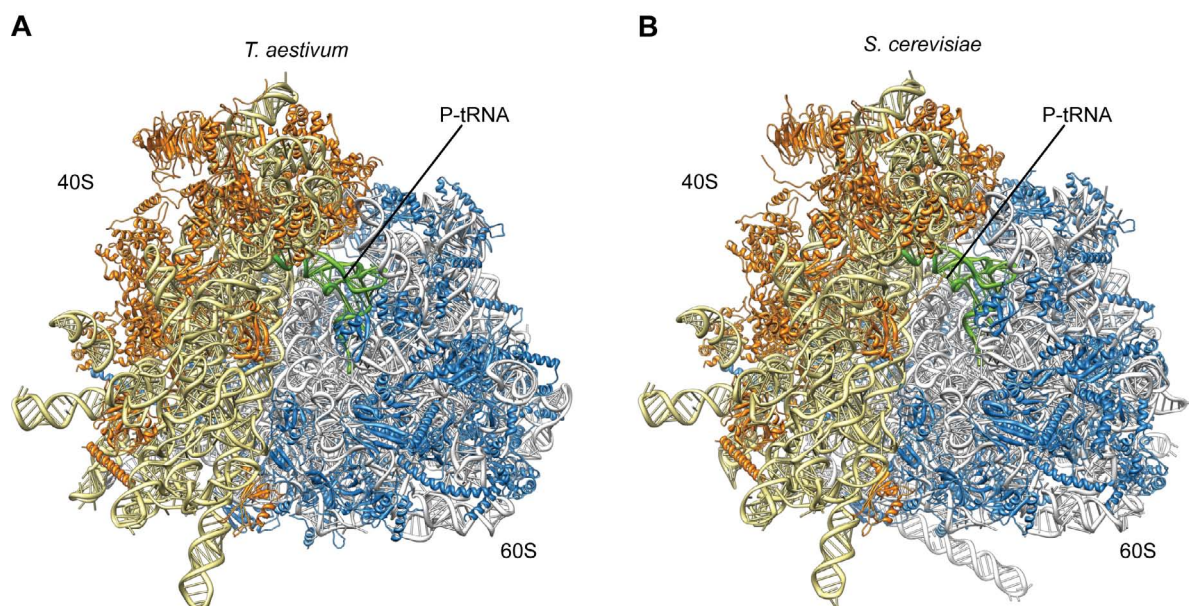


Figure 9.2.1.: **Structure of the *T. aestivum* and *S. cerevisiae* eukaryotic 80S ribosomes.** (A and B) Near-complete models for the (A) *T. aestivum* and (B) *S. cerevisiae* 80S ribosome, with rRNA and r-proteins shown in yellow and orange for the small subunit and gray and blue for the large subunit, respectively. The P-tRNA is shown in green. The figure was taken from [185].

The majority of the rRNA ($\sim 65\%$) model correspond to the conserved ribosomal core of *T. aestivum* and *S. cerevisiae* and was modeled based on homology of the eukaryotic rRNA with the available bacterial and *archaeal* ribosome structures (black in Figure 9.2.2 A and B, gray in Figure 9.2.2 C and D). The remaining parts ($\sim 35\%$) of the rRNA comprising structurally variable regions and rRNA expansion segments were modeled *de novo* (green in Figure 9.2.2 A-D). One-hundred sixteen (*T. aestivum*) nts and 122 (*S. cerevisiae*) nts, mostly single stranded linker regions, could not be modeled due to unreliable secondary structure predictions and/or ambiguity in the electron density (orange in Figures 9.2.3, 9.2.4, 9.2.5, 9.2.6, 9.2.7 and 9.2.8). Similarly, models for 44 of the 80 r-proteins of the *T. aestivum* 80S ribosome (gray in Figure 9.2.2 E and F) were built using the templates present in the bacterial and *archaeal* ribosome structures [12, 184], as well as 44 of 79 r-proteins of the yeast 80S ribosome [185]. Homologous proteins having an N-terminal or C-terminal extension were build *de novo* (green in Figure 9.2.2 E and F). A total of 27 r-proteins (12 on the small subunit, 17 on the large subunit; excluding P0, P1 and P2) that are not present in the crystal structures of the bacterial and *archaeal* ribosomes could be localized and modeled *de novo* (red in Figure 9.2.2 E, F).

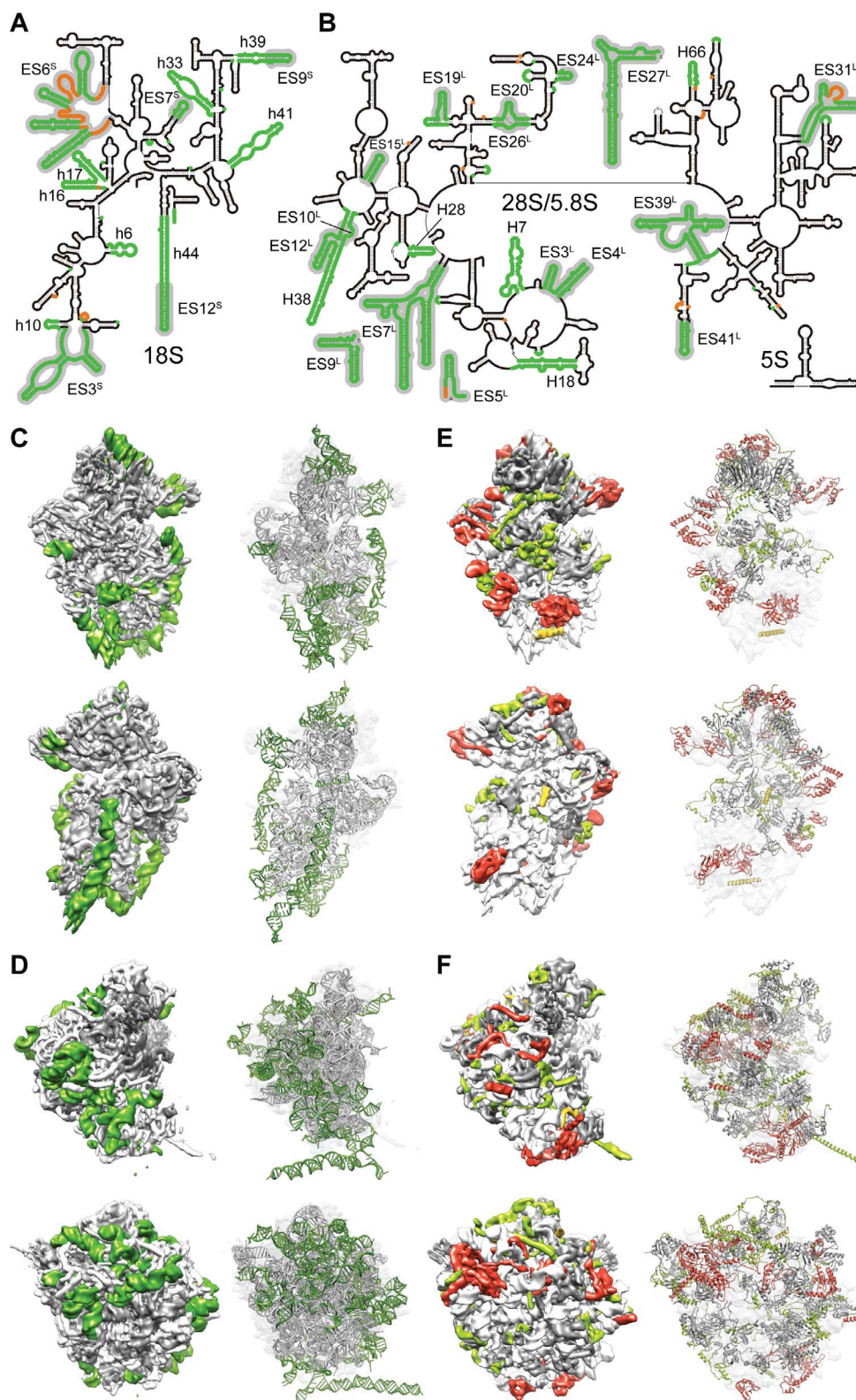


Figure 9.2.2.: **An atomic model for the *T. aestivum* 80S ribosome.**(A and B) Secondary structures for the (A) small (18S) and (B) large subunit (5S, 5.8S, and 28S) ribosomal RNAs, with the newly modeled regions colored in green. Expansion segments and variable regions are indicated in gray and un-modeled regions are orange. (C and D) Newly modeled regions of rRNA (green) are highlighted on the (C) small and (D) large subunit density map (Left) and as molecular models (Right). (E and F) Newly modeled proteins are highlighted on the (E) small and (D) large subunit density map (Left) and as molecular models (Right). Newly identified proteins are colored red, whereas *de novo* modeled extensions are colored light green, and modeled but unassigned proteins are yellow. The figure was taken from [182].

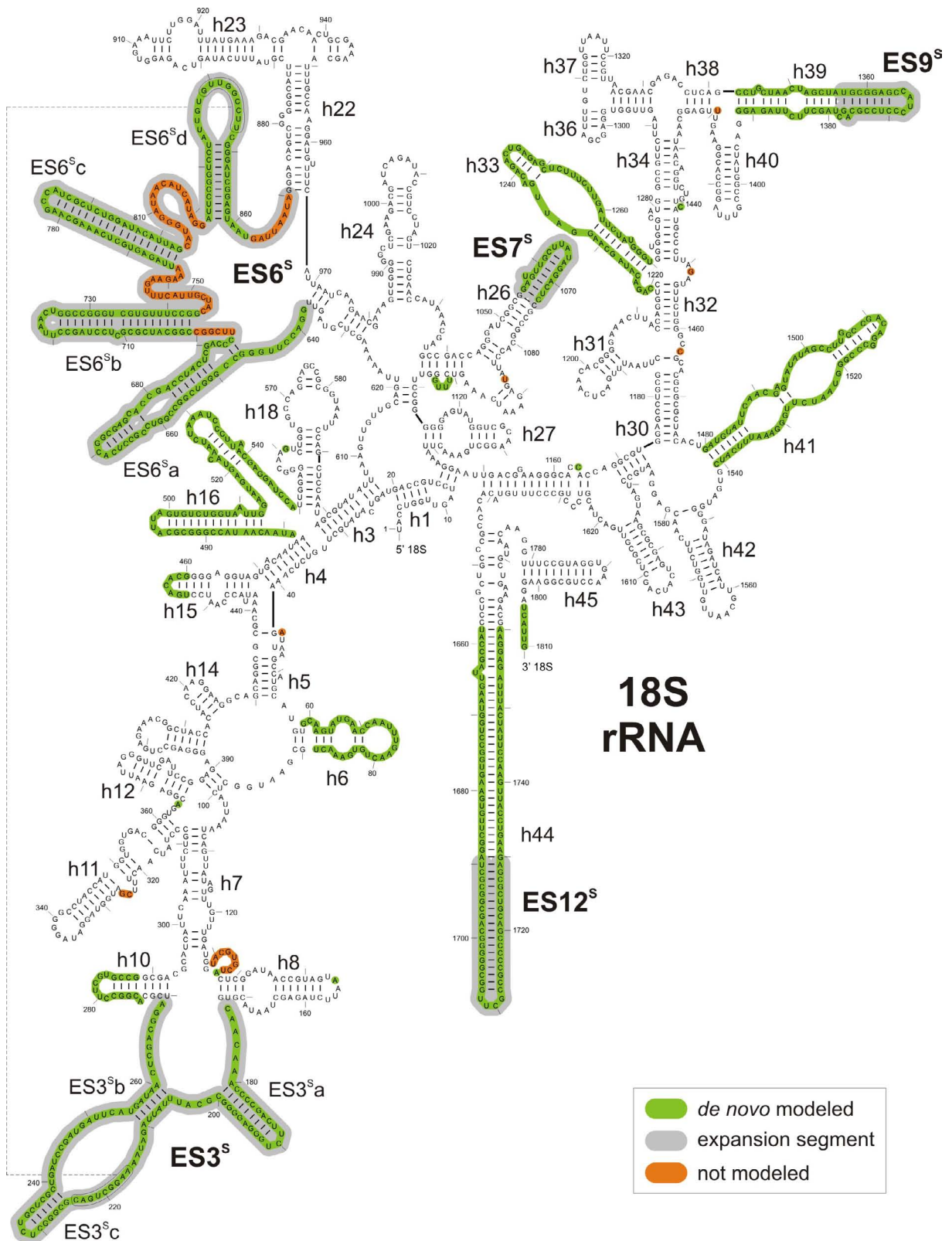


Figure 9.2.3.: **Secondary structure diagram for the small subunit (18S) rRNA of *T. aestivum*.** Green regions indicated *de novo* modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled. The diagram was modified from [27]. The figure was taken from [182].

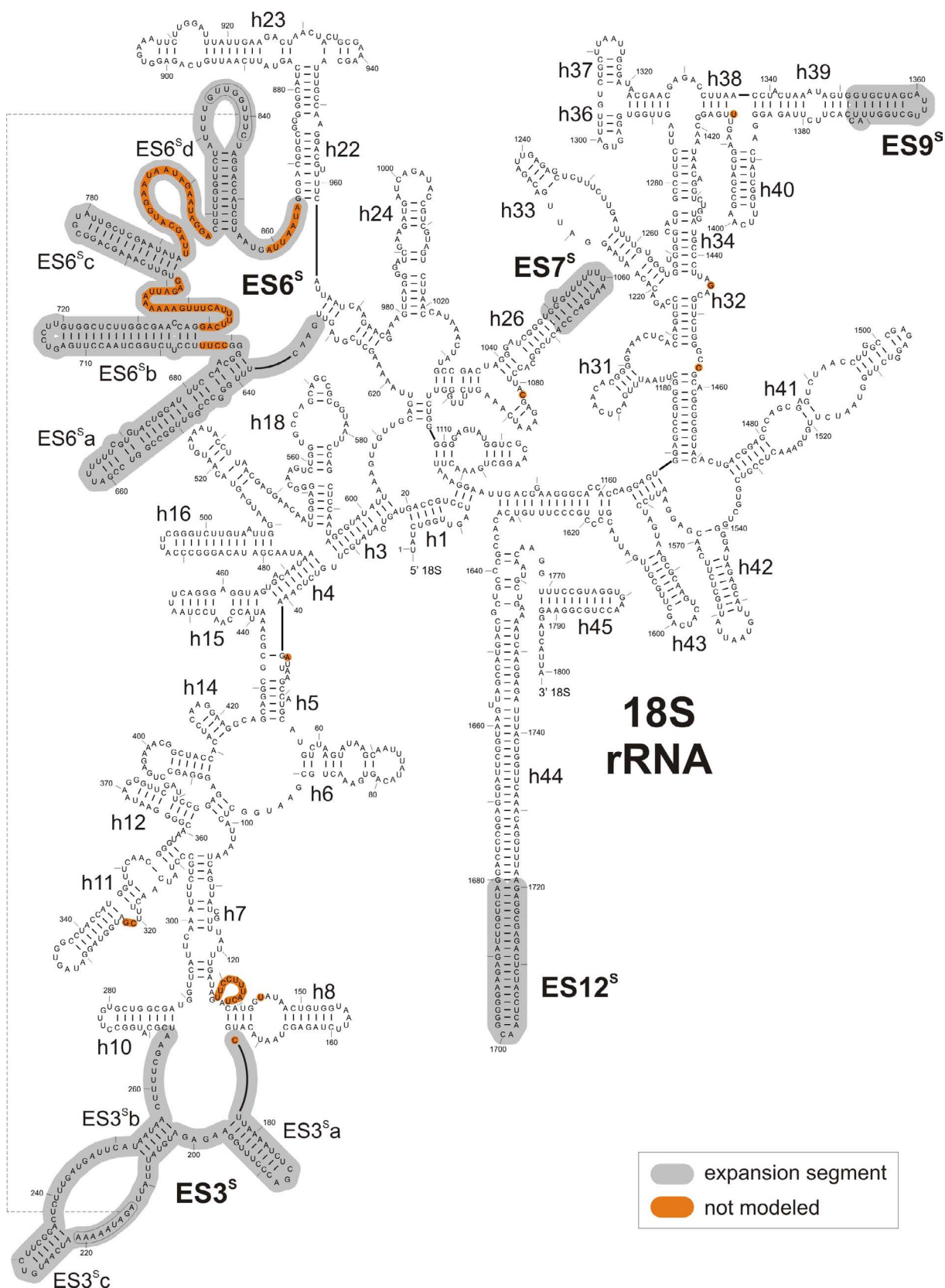


Figure 9.2.4.: **Secondary structure diagram for the small subunit (18S) rRNA of *S. cerevisiae*.** Gray regions indicate expansion segments that have been modeled *de novo*, whereas orange nucleotides were not modeled. The diagram was modified from [27]. The figure was taken from [182].

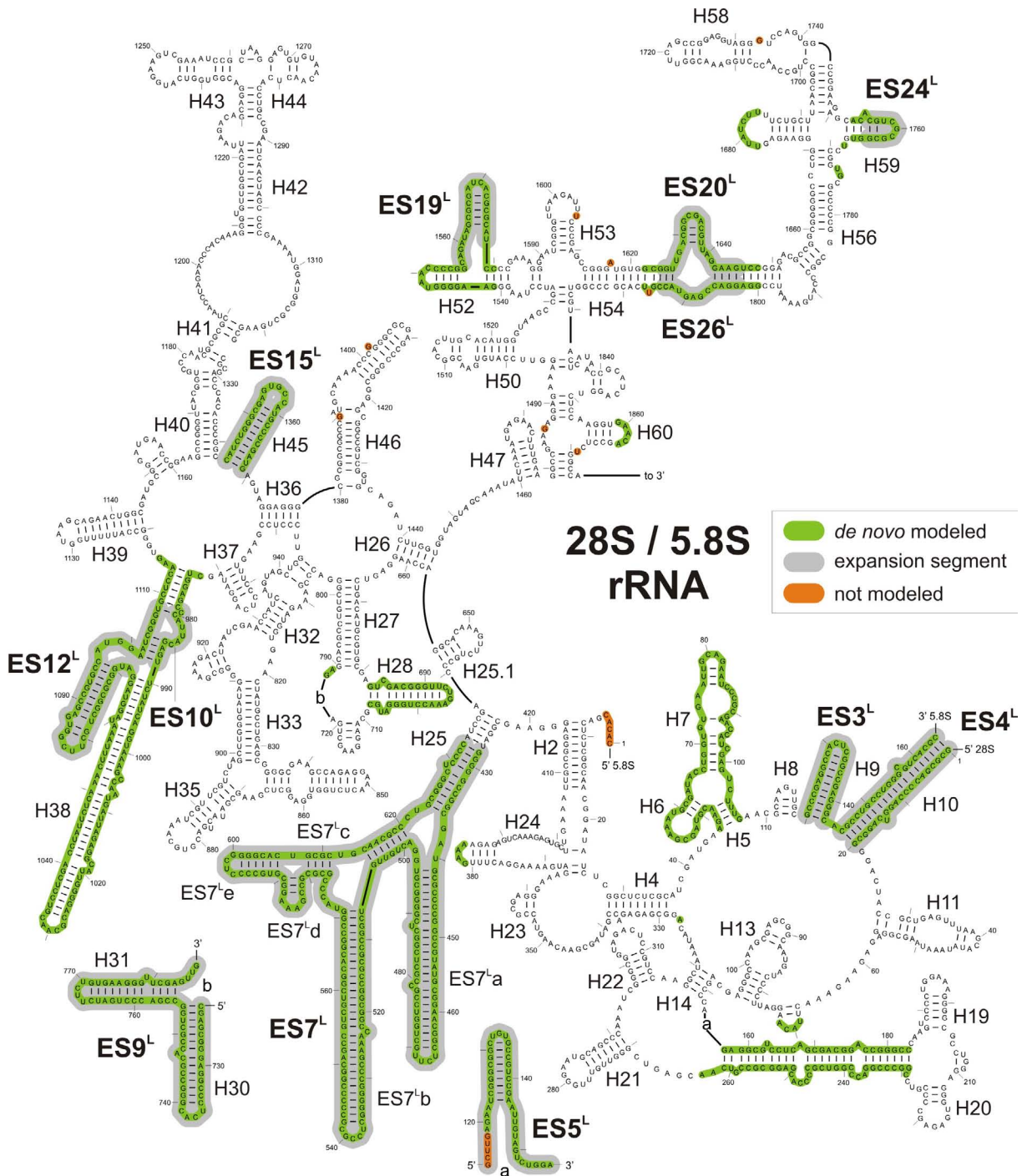


Figure 9.2.5.: Secondary structure diagram for the 5' region of the large subunit rRNAs (5.8S and 25S) of *T. aestivum*. Green regions indicated *de novo* modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled. The diagram was modified from [27]. The figure was taken from [182].

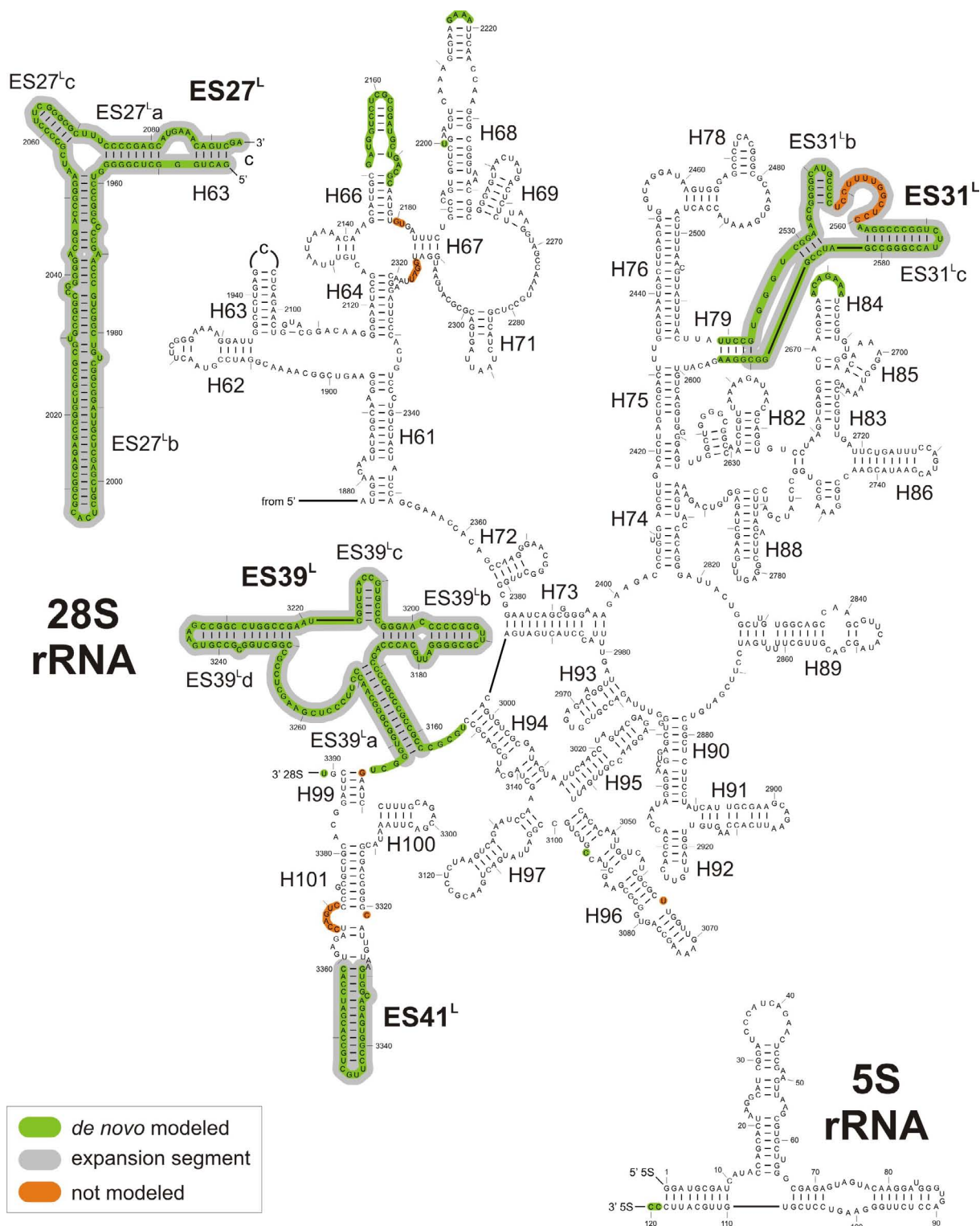


Figure 9.2.6.: Secondary structure diagram for the 3' region of the large subunit rRNAs (25S) of *T. aestivum*. Green regions indicated *de novo* modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled. The diagram was modified from [27]. The figure was taken from [182].

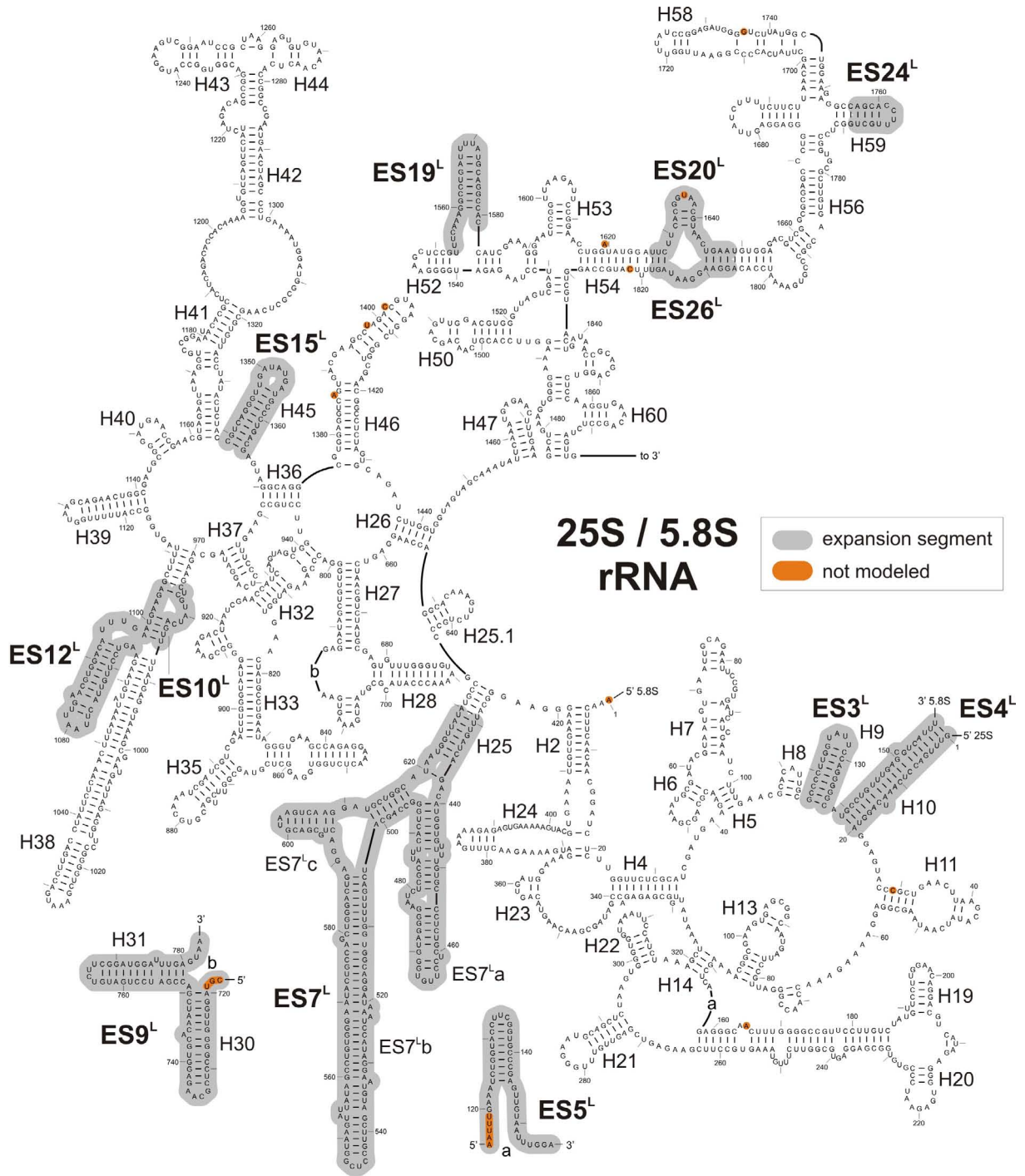


Figure 9.2.7.: **Secondary structure diagram for the 5' region of the large subunit rRNAs (5.8S and 25S) of *S. cerevisiae*.** Gray regions indicate expansion segments that have been modeled *de novo*, whereas orange nucleotides were not modeled. The diagram was modified from [27]. The figure was taken from [182].

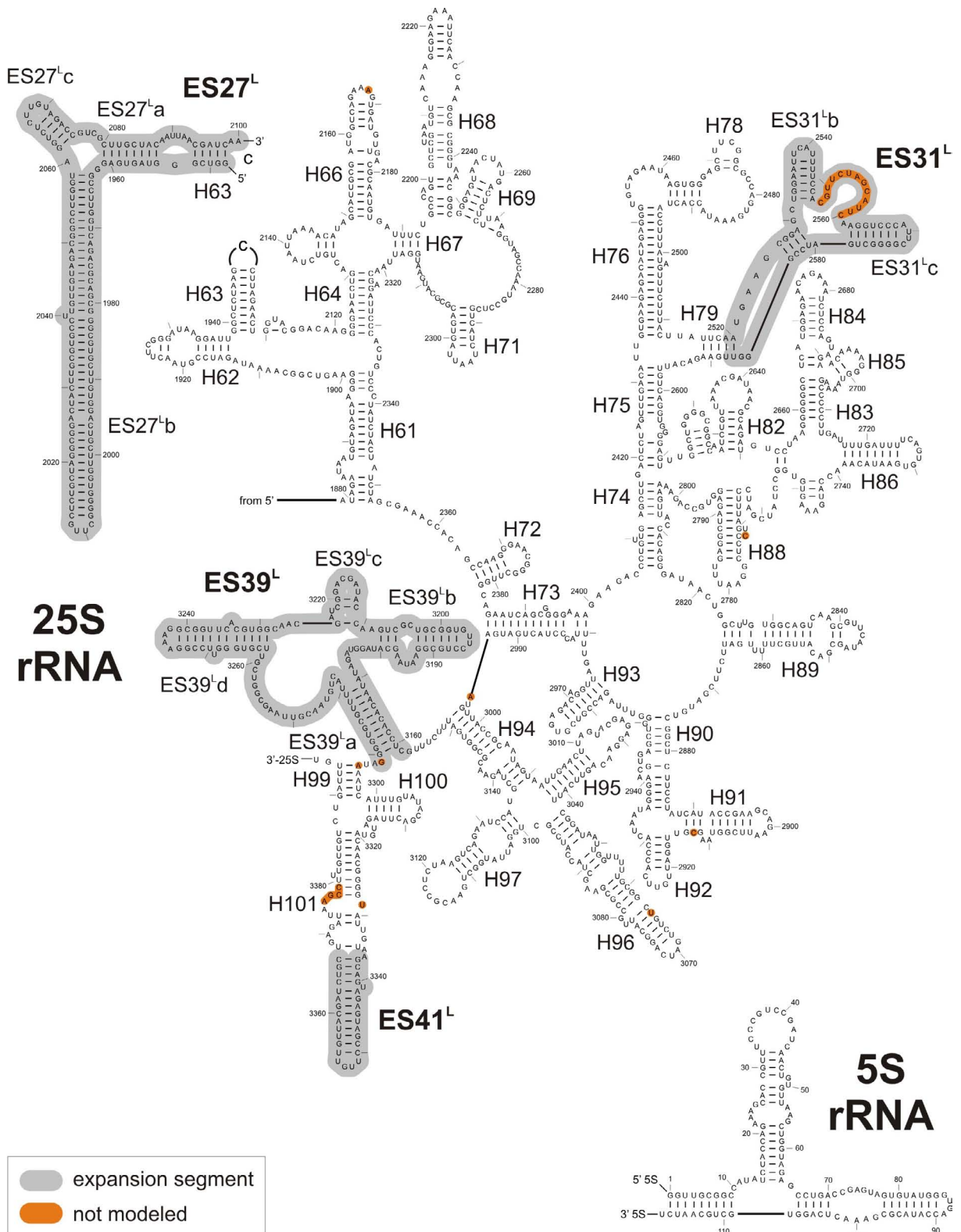


Figure 9.2.8.: **Secondary structure diagram for the 3' region of the large subunit rRNAs (25S) of *S. cerevisiae*.** Gray regions indicate expansion segments that have been modeled *de novo*, whereas orange nucleotides were not modeled. The diagram was modified from [27]. The figure was taken from [182].

9.2.1. Atomic model of the eukaryotic ribosomal RNA core structure

The structure of the ribosomal RNA core of prokaryotes and eukaryotes is very similar. In order to assess a rough structural similarity the rRNA (16S, 23S and 5S) of the crystal structures were rigidly fitted into the cryo-EM densities of *T. aestivum* and *S. cerevisiae* ('fit-in-map'-⁴command of Chimera [153]). Each RNA helix was inspected manually and secondary structure predictions of *T. aestivum*, *S. cerevisiae*, *T. thermophilus* and *H. marismortui* were compared. This revealed significant structural differences in helices h6, h9, h10, h16, h17, h21, h26, h33, h39, h41 and h44 on the small subunit as well as H7, H16, H17, H18, H28, H38, H45, H52, H59, H79 and H101 on the large subunit. Thus, these helices were not aligned and modeled by homology. A subsequent comparison between *T. aestivum* and *S. cerevisiae* revealed a higher similarity. On the small subunit *T. aestivum* and *S. cerevisiae* are only structurally different in the helices of *ES6^S*, h16 and minor differences in h39 and h41. On the large subunit significant structural differences occur at *ES3^L*, *ES7^L*, *ES27^L* and *ES39^L*.

As described in section 8.1.1, structure-based sequence alignments between sequences of *T. aestivum*, *S. cerevisiae* and the available *bacterial* and *archaeal* crystal structures were generated. Due to the better resolution of the *T. aestivum* cryo-EM map, the core model for *T. aestivum* was generated first. This core model was used as anchor point for the expansion segments and variable regions. The combined and flexible fitted RNA model of *T. aestivum* was then used as a structural template for modeling the RNA of *S. cerevisiae*.

Homology model of the *T. aestivum* ribosomal RNA core structure For *T. aestivum* 3584 nts (1051/40S, 2415/60S and 118/5S) of the 5484 nts were aligned, incorporating isosteric base substitutions [85, 86] (non-highlighted nucleotides in Figures 9.2.3, 9.2.5 and 9.2.6). One-hundred twenty nucleotides (2.2%) were not available in the sequences for *T. aestivum* and were therefore substituted with those from the closely related *Oryza sativa*. (see section 6.1).

Homology model of the *S. cerevisiae* RNA core structure The core model of *S. cerevisiae* was based on the RNA model of *T. aestivum*. This core model includes all RNA expansion segments except from *ES7^L*, *ES27^L* and *ES39^L*. For *S. cerevisiae* 4679 (1490/40S, 3071/60S and 118/5S) nts of 5472 nts were aligned, incorporating isosteric base substitutions [85, 86] (non-highlighted nucleotides in Figures 9.2.4, 9.2.7 and 9.2.8).

⁴<http://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/fitmaps/fitmaps.html>

9.2.2. Atomic models for eukaryote-specific ribosomal RNA expansion segments

Overview The *T. aestivum* and yeast 80S models contain all five expansion segments ($ES3^S$, $ES6^S$, $ES7^S$, $ES9^S$, and $ES12^S$ following the ES nomenclature of [23]) and five variable regions (h6, h16, h17, h33, and h41) of the small subunit, as well as the 16 expansion segments ($ES3^L$, $ES4^L$, $ES5^L$, $ES7^L$, $ES9^L$, $ES10^L$, $ES12^L$, $ES15^L$, $ES19^L$, $ES20^L$, $ES24^L$, $ES26^L$, $ES27^L$, $ES31^L$, $ES39^L$, and $ES41^L$) and two variable regions (H16–18 and H38) of the large subunit (Figure 9.2.9). On the small subunit, the majority of the additional rRNA is clustered at the spur or foot region, except for $ES9^S$ which is positioned at the head (Figure 9.2.9 C and E). On the large subunit, most ES are located on the back and sides of the particle, leaving the subunit interface and exit tunnel regions essentially unaffected (Figure 9.2.9 D and F).

For *T. aestivum* 1903 nts comprising structurally variable regions and expansion segments were modeled *de novo* (Figure 9.2.2 A-D and Figure 9.2.9 A-E) using Assemble [136]. The secondary structure was predicted using RNAfold [179] and RNASHapes [170]. The predictions were then compared to the predictions from the CRW database [27] and the corresponding EM density.

A secondary structure comparison between *S. cerevisiae* and *T. aestivum* revealed that in both species the expansion segments interrupt the universal core secondary structure at the same sites. Most of the expansion segments have the same or similar length and secondary structure prediction. Therefore, only 654 nts (171/40S, 483/60S) had to be modeled *de novo* in *S. cerevisiae*.

The next section will give a summary of all the expansion segments in *S. cerevisiae* and *T. aestivum*. The expansion segments $ES7^L$ and $ES27^L$ are presented separately for *T. aestivum* and *S. cerevisiae*, due to major differences in the two structures. The ribosomal proteins will be denoted with their family name. A table with the specific names for each organism can be found in tables 9.1 and 9.2.

9. Model of *T. aestivum* and *S. cerevisiae* 80S ribosomes based on high-resolution cryo-EM structures

	Family name	<i>S. cerevisiae</i> name	<i>O. sativa</i> name
1	S2p	rpS0	Sa
2	S3p	rpS3	S3
3	S3ae	rpS1	S3a
4	S4p	rpS9	S9
5	S4e	rpS4	S4
6	S5p	rpS2	S2
7	S6e	rpS6	S6
8	S7p	rpS5	S5
9	S7e	rpS7	S7
10	S8p	rpS22	S15a
11	S8e	rpS8	S8
12	S9p	rpS16	S16
13	S10p	rpS20	S20
14	S10e	rpS10	S10
15	S11p	rpS14	S14
16	S12p	rpS23	S23
17	S12e	rpS12	S12
18	S13p	rpS18	S18
19	S14p	rpS29	S29
20	S15p	rpS13	S13
21	S17p	rpS11	S11
22	S17e	rpS17	S17
23	S19p	rpS15	S15
24	S19e	rpS19	S19
25	S21e	rpS21	S21
26	S24e	rpS24	S24
27	S25e	rpS25	S25
28	S26e	rpS26	S26
29	S27e	rpS27	S27
30	S27ae	rpS31	S27a
31	S28e	rpS28	S28
32	S30e	rpS30	S30
33	RACK1	RACK1	RACK1

Table 9.1.: **Nomenclature for r-proteins of the small subunit of the *S. cerevisiae* and *T. aestivum*.**
The table was taken from [185].

9. Model of *T. aestivum* and *S. cerevisiae* 80S ribosomes based on high-resolution cryo-EM structures

	Family name	<i>S. cerevisiae</i> name	<i>O. sativa</i> name
1	L1p	rpL1	L1
2	L2p	rpL2	L2
3	L3p	rpL3	L3
4	L4e/L4p	rpL4	L4
5	L5p	rpL11	L11
6	L6p	rpL9	L9
7	L6e	rpL6	L6
8	L7ae	rpL8	L7a
9	L10p	rpP0	P0
10	L10e	rpL10	L10
11	L11p	rpL12	L12
12	L12p	rpP1/rpP2	P1/P2
13	L13p	rpL16	L13a
14	L13e	rpL13	L13
15	L14p	rpL23	L23
16	L14e	rpL14	L14
17	L15p	rpL28	L27a
18	L15e	rpL15	L15
19	L18p	rpL5	L5
20	L18e	rpL18	L18
21	L18ae	rpL20	L18a
22	L19e	rpL19	L19
23	L21e	rpL21	L21
24	L22p	rpL17	L17
25	L22e	rpL22	L22
26	L23p	rpL25	L23a
27	L24p	rpL26	L26
28	L24e	rpL24	L24
29	L27e	rpL27	L27
30	L28e	-	L28
31	L29p	rpL35	L35
32	L29e	rpL29	L29
33	L30p	rpL7	L7
34	L30e	rpL30	L30
35	L31e	rpL31	L31
36	L32e	rpL32	L32
37	L34e	rpL34	L34
38	L35ae	rpL33	L35a
39	L36e	rpL36	L36
40	L37e	rpL37	L37
41	L37ae	rpL43	L37a
42	L38e	rpL38	L38
43	L39e	rpL39	L39
44	L40e	rpL40	L40
45	L41e	rpL41	L41
46	L44e	rpL42	L44

Table 9.2.: Nomenclature for r-proteins of the large subunit of the *S. cerevisiae* and *T. aestivum*.
The table was taken from [185].

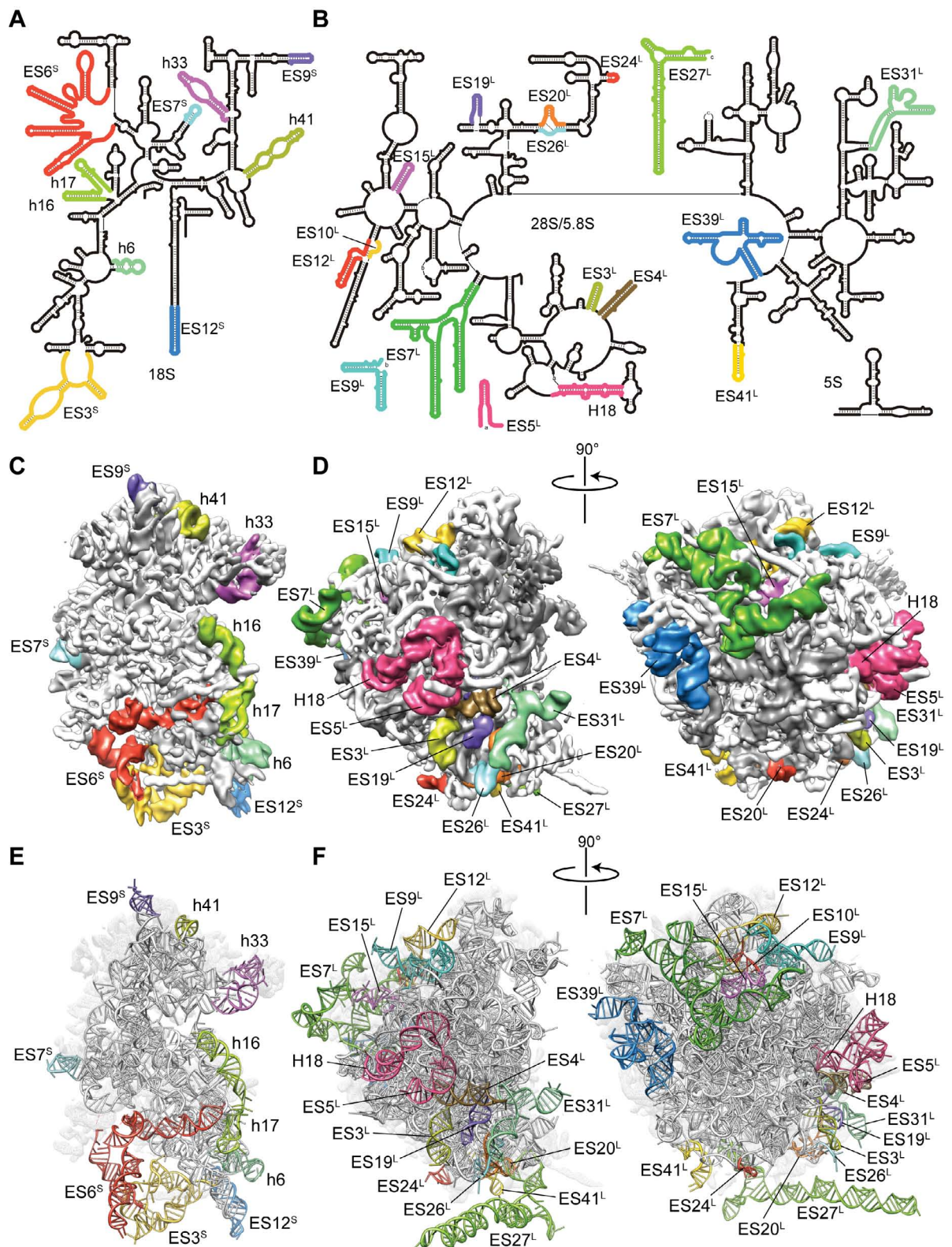


Figure 9.2.9.: **Ribosomal RNA expansion segments and variable regions.** (A and B) Secondary structures for the *T. aestivum* (A) small (18S) and (B) large subunit (5S, 5.8S, and 25S) ribosomal RNAs, with the expansion segments and variable regions colored distinctly. (C and D) Cryo-EM maps of the (C) small and (D) large subunits with assigned ES and VR colored as in A and B. (E and F) Molecular models of the ES and VR of rRNA colored as in C and D. The figure was taken from [182].

9.2.2.1. $ES1^S$ (h6)

Helix 6 (h6) denoted as $ES1^S$ in [23] is one of 5 variable regions for both *T. aestivum* and *S. cerevisiae*. It is located on the foot region of the small subunit and close to the inter subunit space (Figure 9.2.10 A). Helix 6 is part of the 5' domain of the 18S rRNA (Figure 9.2.10 B). In both species it is 33 nt long and shorter than in *T. thermophilus* (Figure 9.2.10 C-F). While the stem region seems to fold like the *bacterial* counterpart the fold of the loop region is ambiguous (Figure 9.2.10 E). The EM density in the loop region suggest the position for one of the non-identified novel proteins.

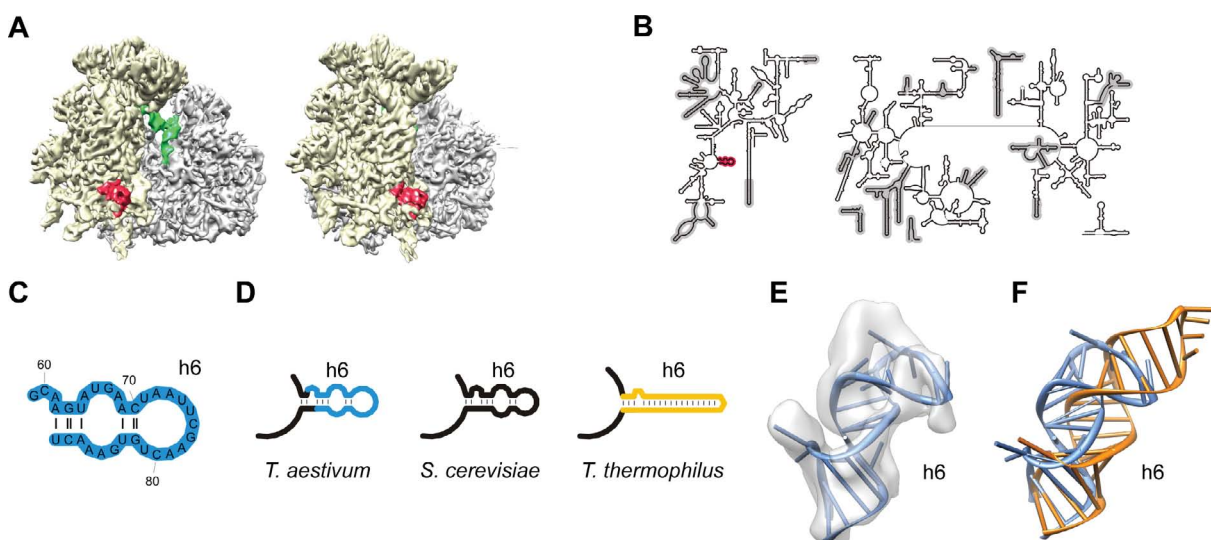


Figure 9.2.10.: **Expansion segment $ES1^S$.** (A) Location of $ES1^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES1^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). The table was taken from [185].

9.2.2.2. $ES3^S$ (h9, h10)

Expansion segment $ES3^S$ is located on the foot of the small subunit (Figure 9.2.11 A) close to expansion segment $ES6^S$. $ES3^S$ is part of the 5' domain of the 18S rRNA (Figure 9.2.11 B). In *T. aestivum* and *S. cerevisiae* helix $ES3^Sb$ substitutes h9 in *bacteria* (Figure 9.2.11 C-F) and keeps $ES3^Sa$ and $ES3^Sc$ in a close spatial arrangement. The 14-nt-long internal loop between helices $ES3^Sb$ and $ES3^Sc$ was modeled to interact with the loop of $ES6^Sd$ forming a 9-mer interhelical pseudoknot (Figure 9.2.11 C-F). The density for helix $ES3^Sc$ is weak suggesting a higher flexibility for the stem loop (Figure 9.2.11 E). $ES3^S$ interacts with the ribosomal protein S17p and the eukaryote-specific r-protein S4e. Helix 10 (h10) which is longer in eukaryotes and helix $ES3^Sa$ interact with an unidentified protein XS1 (Figure 9.2.11 G, H).

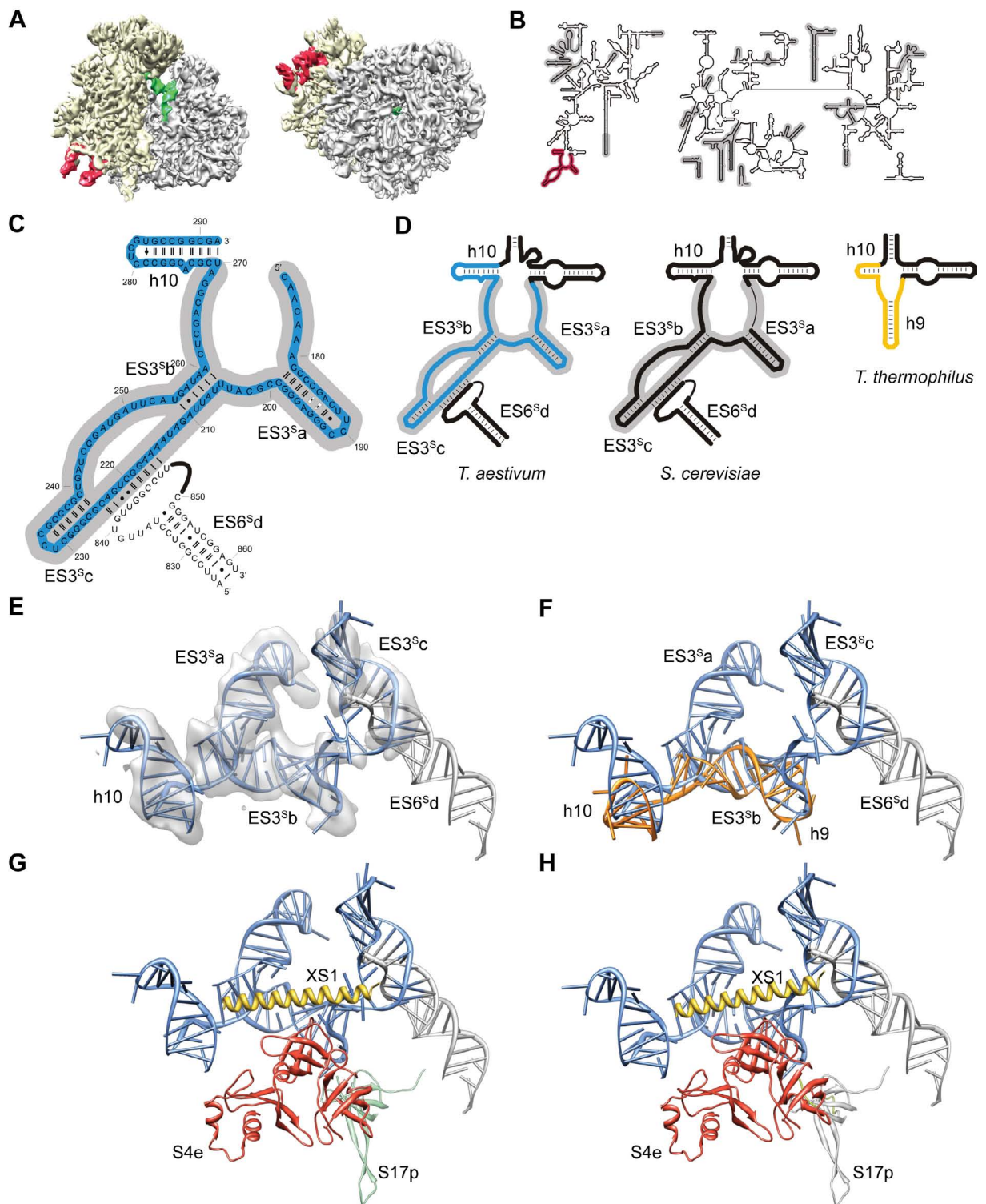


Figure 9.2.11.: **Expansion segment $ES3^S$.** (A) Location of $ES3^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES3^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES3^S$ (blue) with r-proteins S4e (red), S17p (green) and an unidentified protein XS1 (yellow). (H) Same as in (G) but eukaryote-specific proteins (red), protein extensions (light green) of bacterial homologous (gray) and unknown proteins (yellow).

9.2.2.3. $ES5^S$ (h16, h17)

The variable region helix 16 (h16) and helix 17 (h17) are denoted $ES5^S$ in [23]. They are located at the mRNA entry site on the small subunit. While h16 folds towards the head h17 folds towards the small subunit foot (Figure 9.2.12 A). Both helices are part of the 5' domain of the 18S rRNA (Figure 9.2.12 B). In *T. aestivum* and *S. cerevisiae* h16 is not longer compared to *T. thermophilus* (Figure 9.2.12 C-F).

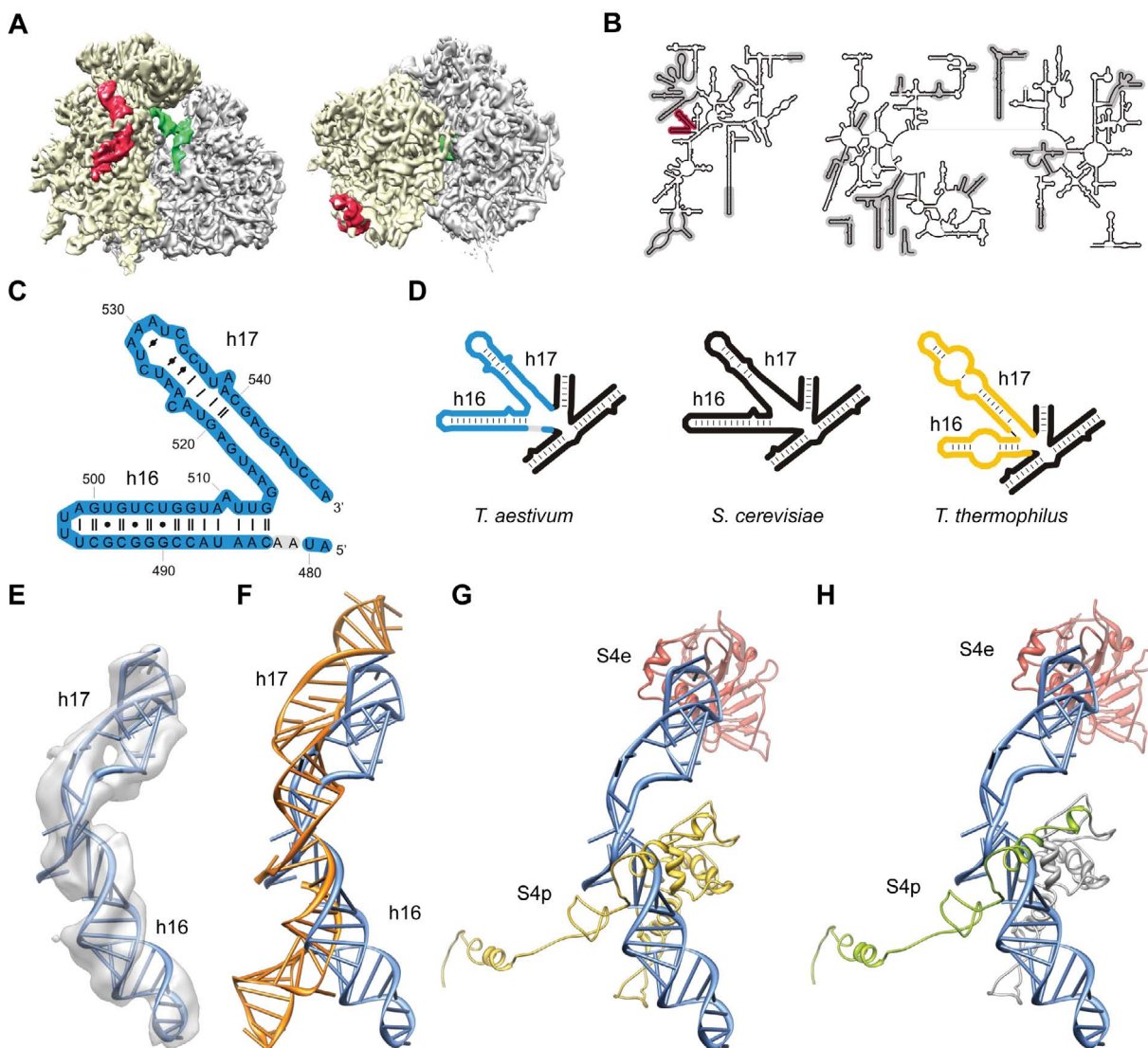


Figure 9.2.12.: **Expansion segment $ES5^S$.** (A) Location of $ES5^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES5^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES5^S$ (blue) with r-proteins S4e (red) and S4p (yellow). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

The innerhelical motif in *bacteria* (right panel in Figure 9.2.12 D) is not present in eukaryotes (Figure 9.2.12 D) and therefore h16 is rearranged (Figure 9.2.12 E, F). The rearrangement corresponds to the additional C-terminus of r-protein S4p which is also relocated in eukaryotes and stabilizes the single stranded linker region between h16 and h17 (Figure 9.2.12 G, H) . The secondary structure prediction of h17 is ambiguous. Like in the *bacterial* counterpart h17 seems to form several non-Watson-Crick base pairs. Helix 17 is shorter in *T. aestivum* and *S. cerevisiae* (Figure 9.2.12 E, F) and the fold different compared to those of *bacteria* (Figure 9.2.12 D-F). The eukaryote-specific r-protein S4e stabilizes the loop region of h17 (Figure 9.2.12 G, H).

9.2.2.4. $ES6^S$ (h21)

Expansion segment $ES6^S$ is the largest ES on the small subunit located on the platform (Figure 9.2.13 A). $ES6^S$ is part of the C-domain of the 18S rRNA (Figure 9.2.13 B). The secondary structure predictions for $ES6^S$ are ambiguous. In the top 10 of the minimum free energy predictions by RNAsHapes [170] structures with three up to five helices are predicted. In the EM densities for both *T. aestivum* and *S. cerevisiae* 4 helices are visible. The RNAsHapes predictions were then constraint to 'Shape type=1' and 'Match shape' was set to 4 helices. The tertiary interaction of $ES6^{Sd}$ with the single stranded region $ES3^S$ suggested in [28] constraint the secondary structure prediction and model of $ES6^{Sd}$ (Figure 9.2.13 C, D; Figure 9.2.14 E, F). Sequence alignments of the bacterial helix 21 with the sequences of $ES6^S$ in *T. aestivum* and *S. cerevisiae* suggest the first predicted helix $ES6^{Sa}$ to be eukaryotic counterpart (Figure 9.2.14 E, F).

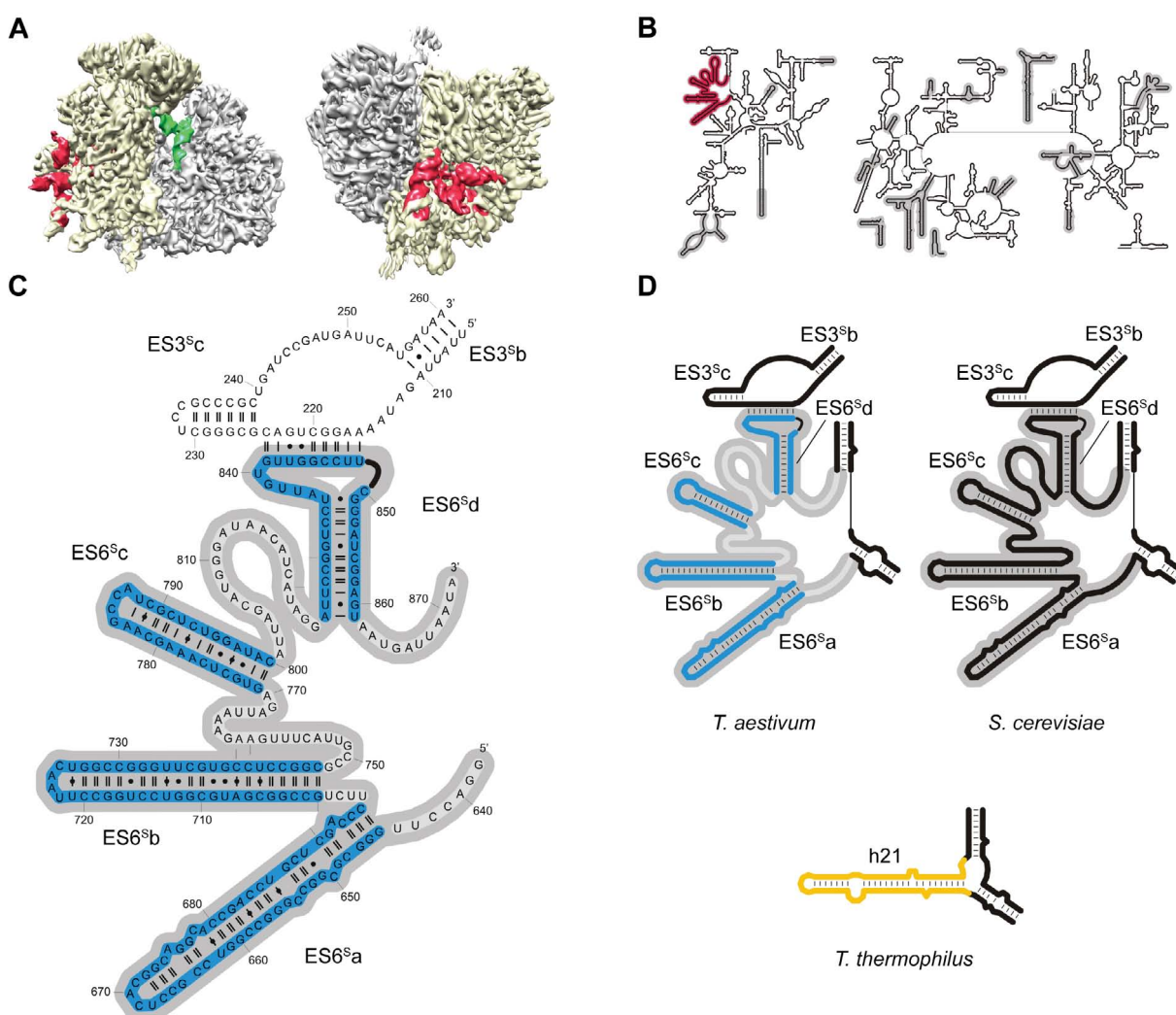


Figure 9.2.13.: **Expansion segment $ES6^S$.** (A) Location of $ES6^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES6^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively.

The density for helices $ES6^{Sb}$ and $ES6^{Sc}$ is observed as double-helical RNA. The densities are weak and not connected suggesting that the two helices are very flexible (Figure 9.2.14 E). Therefore $ES6^{Sb}$ and $ES6^{Sc}$ are modeled like regular A-form helices (Figure 9.2.14 E, F). The single stranded loop region of $ES6^{Sd}$ forms a pseudoknot with the single stranded region of $ES3^S$ (Figure 9.2.14 E-H). The single stranded parts between the four helices (gray in Figure 9.2.13 C) could not be modeled, because they are not resolved in the EM densities and secondary structure predictions are ambiguous.

The eukaryote-specific r-protein S4e stabilizes $ES6^{Sa}$. Helix $ES6^{Sb}$ interacts with S8p and the extension of S5p and helix $ES6^{Sc}$ interacts with r-protein S17p (Figure 9.2.14 G). The last α -helix of the C-terminal extension of L19e interacts with the $ES6^{Sd}$ (Figure 9.2.14 H).

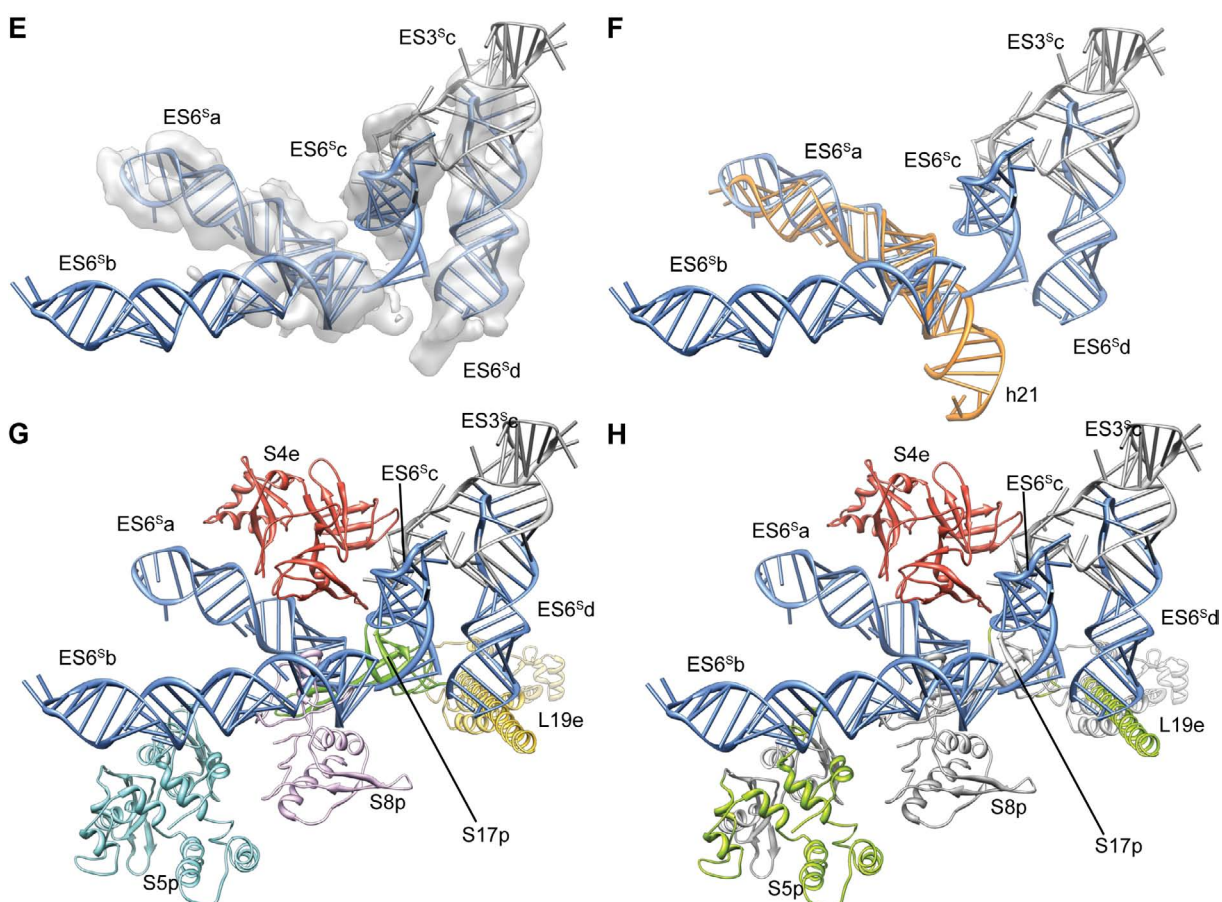


Figure 9.2.14.: **Expansion segment $ES6^S$.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES6^S$ (blue) with r-proteins S4e (red), L19e (yellow), S17p (green), S5p (blue) and S8p (pale purple). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

9.2.2.5. $ES7^S$ (h26)

The extension of helix 26, denoted as $ES7^S$, is located on the platform of the small subunit (Figure 9.2.15 A). $ES7^S$ is part of the central domain of the 18S rRNA (Figure 9.2.15 B). $ES7^S$ folds into a straight A-form helix (Figure 9.2.15 E), which interacts via the minor groove with the N-terminal extension of S2p where the bacterial h26 ends (Figure 9.2.15 F, H). Another two minor groove interactions are observed with the eukaryote-specific r-proteins S21e and S26e (Figure 9.2.15 G, H).

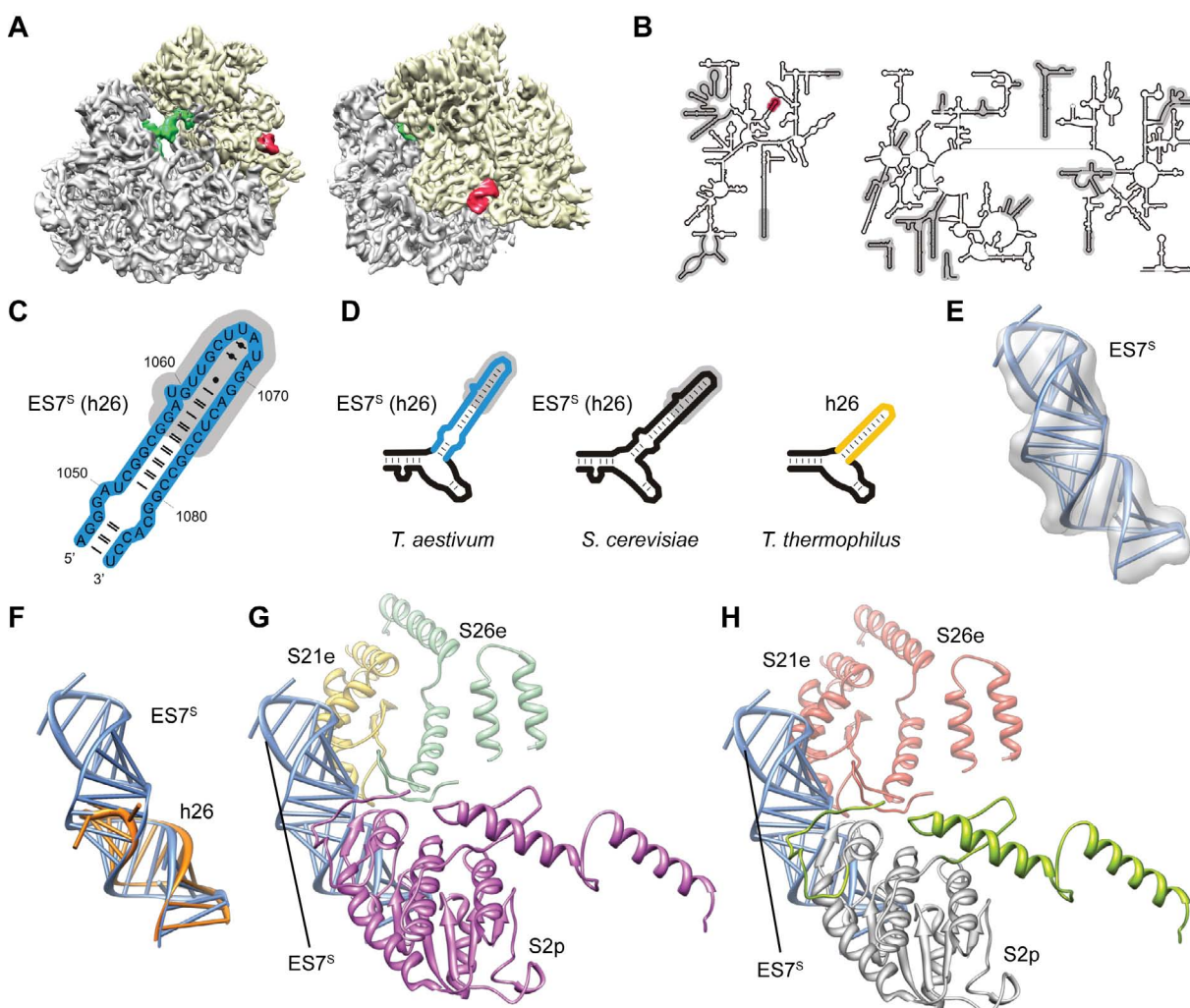


Figure 9.2.15.: **Expansion segment $ES7^S$.** (A) Location of $ES7^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES7^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES7^S$ (blue) with r-proteins S21e (yellow), S26e (blue) and S2p (purple). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

9.2.2.6. ES8^S (h33)

The variable region helix 33 (h33) is located on the beak of the small subunit (Figure 9.2.16 A). Helix 33 is part of the 3'M domain of the 18S rRNA (Figure 9.2.16 B). In *T. thermophilus* h33a is stacked on h33c while the loop of h33b interacts with the minor groove of h32 (Figure 9.2.16 F). In contrast to the bi-fork structure observed in *bacteria*, h33 in *T. aestivum* and *S. cerevisiae* consists of only two helices linked by single stranded RNA (Figure 9.2.16 C, D). The first stem mimics the bacterial fold of h33a while the loop region mimics the same minor groove interaction as h33b in *bacteria* (Figure 9.2.16 E). The single stranded parts kink the loop region $\sim 90^\circ$. Here the non-Watson-Crick interactions can not be explained due to ambiguous density. The N-terminal extension of S19p interacts with the innerhelical loop of h33. The eukaryote-specific r-proteins S17e and S30e replace the non-existing RNA helix h33c which only exists in *bacteria* and stabilizes the kink as well as the first stem (Figure 9.2.16 G, H).

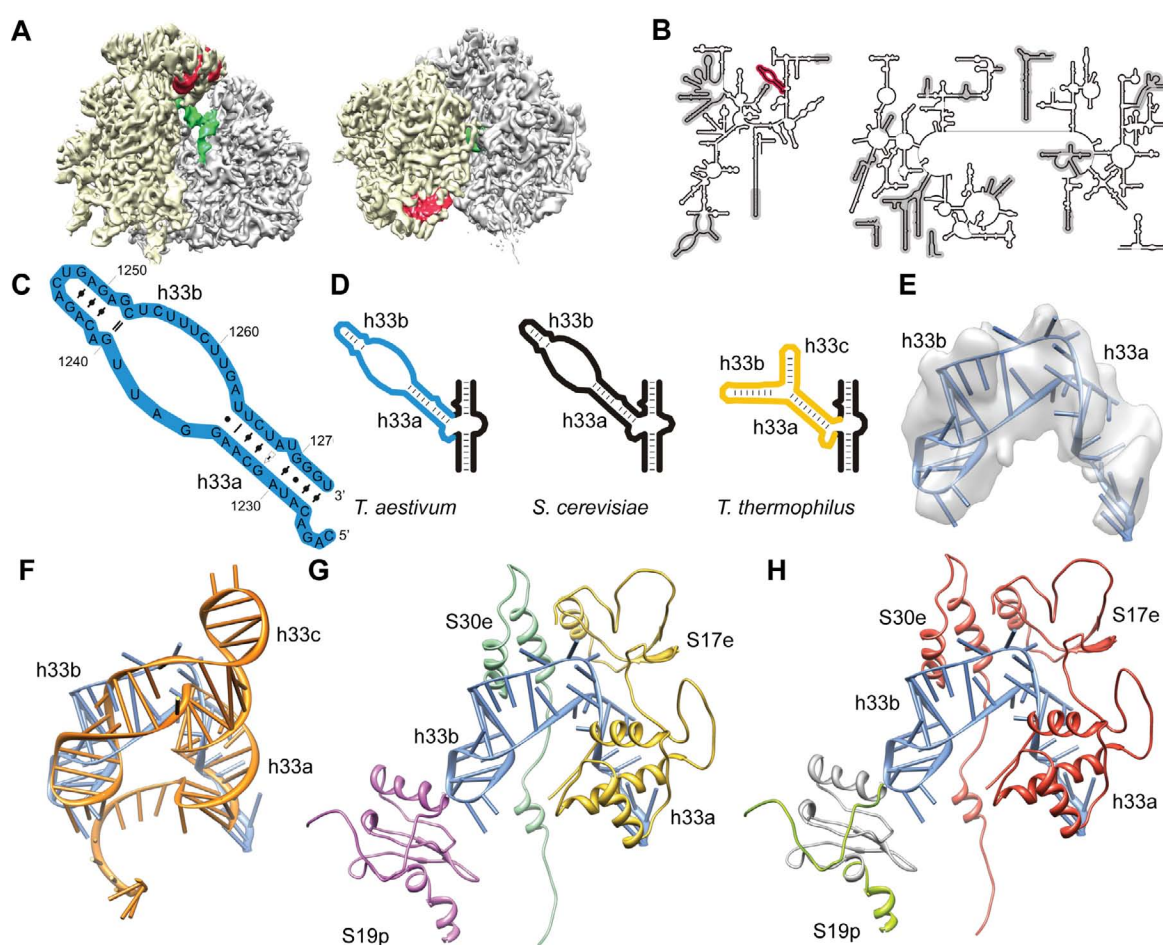


Figure 9.2.16.: **Expansion segment ES8^S**. (A) Location of ES8^S (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of ES8^S (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of ES8^S (blue) with r-proteins S17e (yellow), S30e (blue) and S19p (purple). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

9.2.2.7. $ES9^S$ (h39)

Expansion segment $ES9^S$ is the extension of helix 39 (h39) located next to h41 (see section 9.2.2.8) on the head of the small subunit (Figure 9.2.17 A). $ES9^S$ is part of the 3'M domain of the 18S rRNA (Figure 9.2.17 B). Although the secondary structure is predicted to be helical (Figure 9.2.17 C, D) h39 is bended like its *bacterial* counterpart (Figure 9.2.17 E, F). The loop of $ES9^S$ seems to be flexible, due to the weak density. Two eukaryote-specific r-proteins interact with $ES9^S$ namely RACK1 with the stem region and an α -helix of S19e with the minor groove of the $ES9^S$ loop region (Figure 9.2.17 G, H). The N-terminal extension of S10p (rpS20 in *S. cerevisiae*) contacts the backbone and minor groove of $ES9^S$ as well (Figure 9.2.17 G, H).

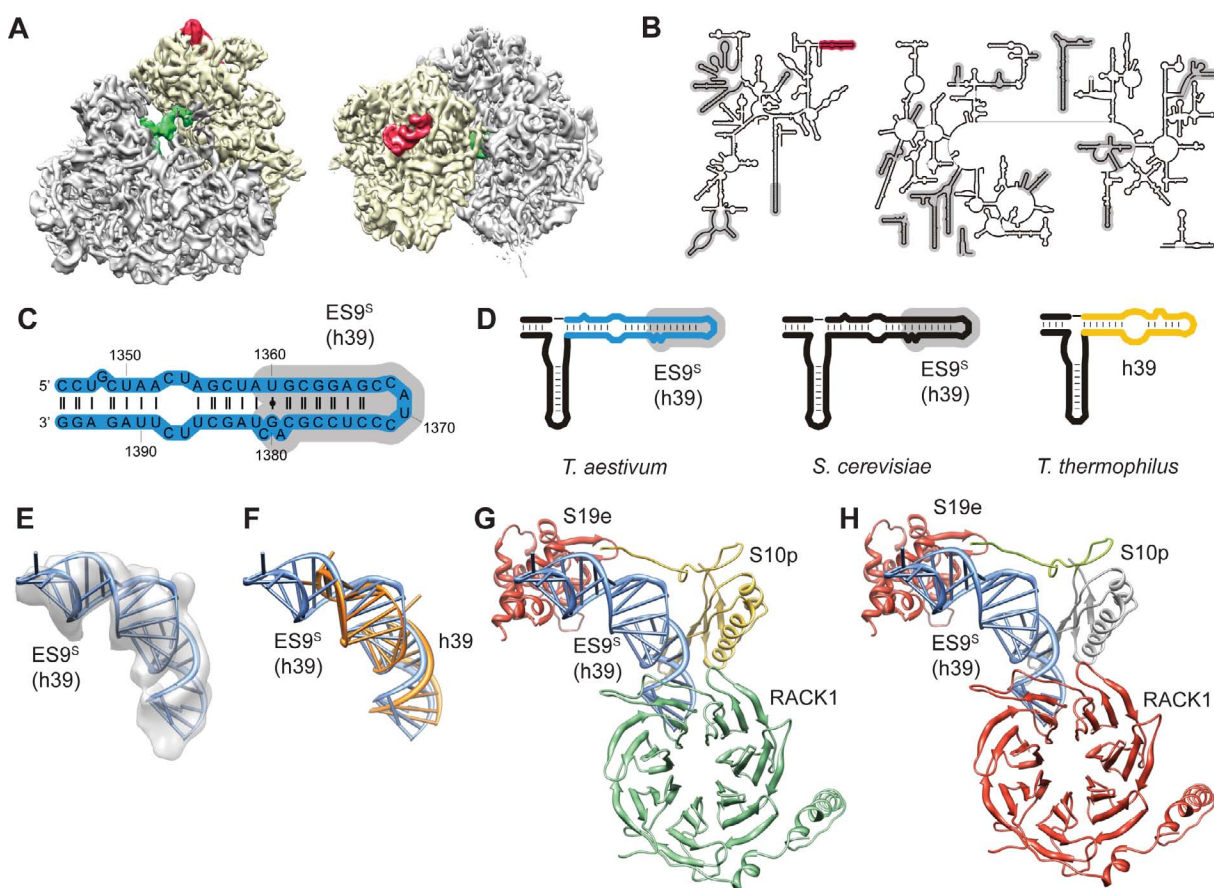


Figure 9.2.17.: **Expansion segment $ES9^S$.** (A) Location of $ES9^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES9^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES9^S$ (blue) with r-proteins S19e (red), S10p (yellow), S30e (blue) and RACK1 (green). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

9.2.2.8. $ES10^S$ (h41)

The variable region helix 41 (h41) is denoted as $ES10^S$ in [23]. It is located next to $ES9^S$ (see section 9.2.2.7) on the head of the small subunit (Figure 9.2.18 A). Helix 41 is part of the 3'M domain of the 18S rRNA (Figure 9.2.18 B). In both *T. aestivum* and *S. cerevisiae* a similar kink-turn-fold as in *T. thermophilus* is observed (Figure 9.2.18 E, F). The length difference between both eukaryotic species is only 2 nts located in the single stranded region (Figure 9.2.18 E). Helix 41 is closely intertwined by several proteins. The newly identified eukaryote-specific r-proteins S19e and S25e interact with the lower stem region of h41 via the backbone and minor groove (Figure 9.2.18 G, H). Further S7p, S9p, the N-terminal extension of S10p, S3p and S13p interacted closely with h41. This results in a tight packing and stabilization of the kink-turn structure (Figure 9.2.18 G, H).

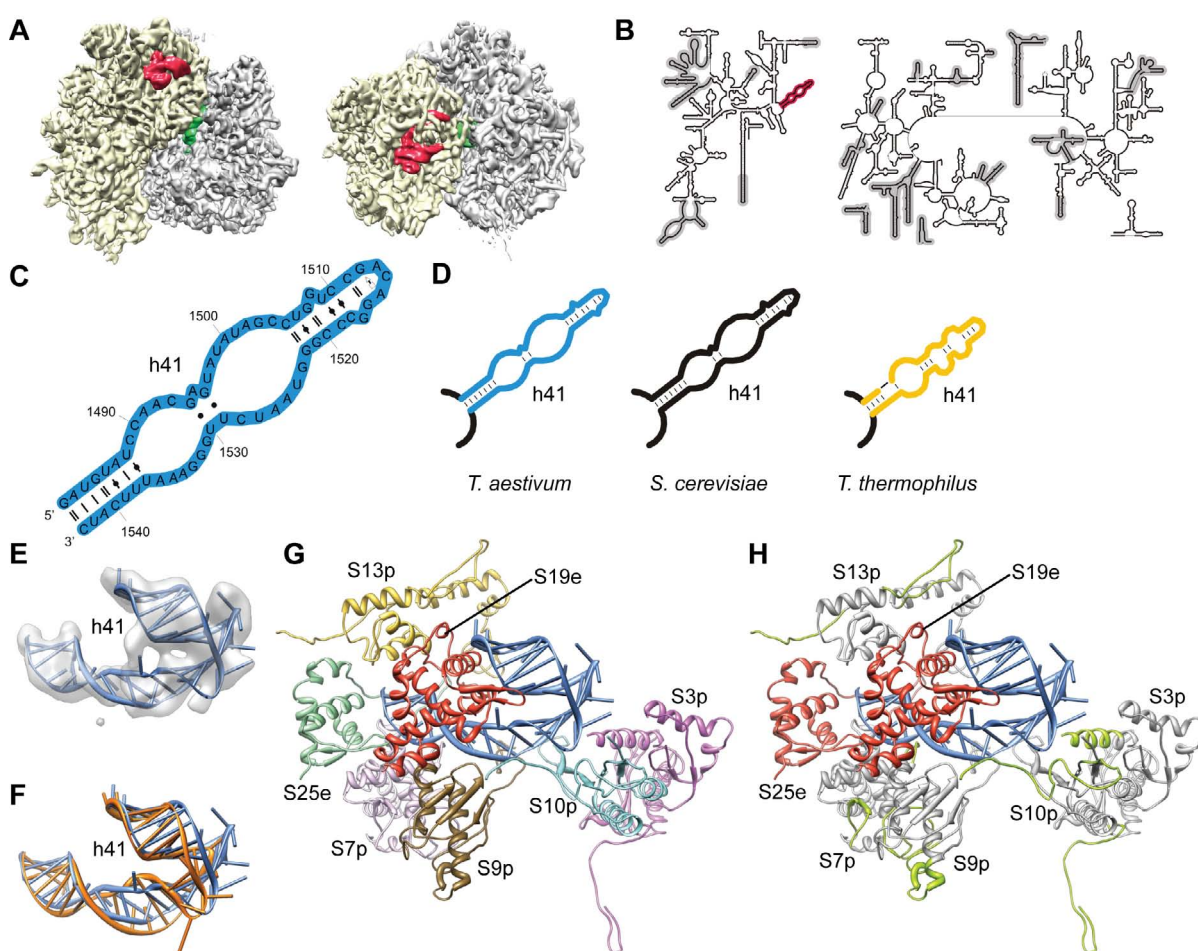


Figure 9.2.18.: **Expansion segment $ES10^S$.** (A) Location of $ES10^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES10^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES10^S$ (blue) with r-proteins S19e (red), S13p (yellow), S10p (blue), S9p (brown), S7p (pale purple), S25e (green) and S3p (purple). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

9.2.2.9. $ES12^S$ (h44)

Expansion segment $ES12^S$ is the long helical extension of helix 44 (h44) located on the inner site of the small subunit (9.2.19 A). $ES12^S$ /h44 is part of the 3'm domain of the 18S rRNA (Figure 9.2.19 B) and reaches from the decoding site out of the ribosome along the inter subunit space between the small and the large subunit. The secondary structure of $ES12^S$ is interrupted by several non-Watson-Crick base pairs (Figure 9.2.19 C, D) leading to a slightly bended helix (Figure 9.2.20 E, F). Moreover, these bases face their Watson-Crick edge towards helices of the large subunit (H71, H64, H62 and H27) enabling important contacts (inter subunit bridges). The density for the last seven base pairs and the loop region is weak suggesting that $ES12^S$ is very flexible at the lower region. $ES12^S$ /h44 interacts with the N-terminal extension of S12p via the major groove, the eukaryote-specific r-protein S24e and an unknown r-protein XS2. (Figure 9.2.20 G, H). Moreover the large subunit ribosomal proteins L14p and the eukaryote-specific r-protein L24e interact with the backbone of h44 (Figure 9.2.20 G, H).

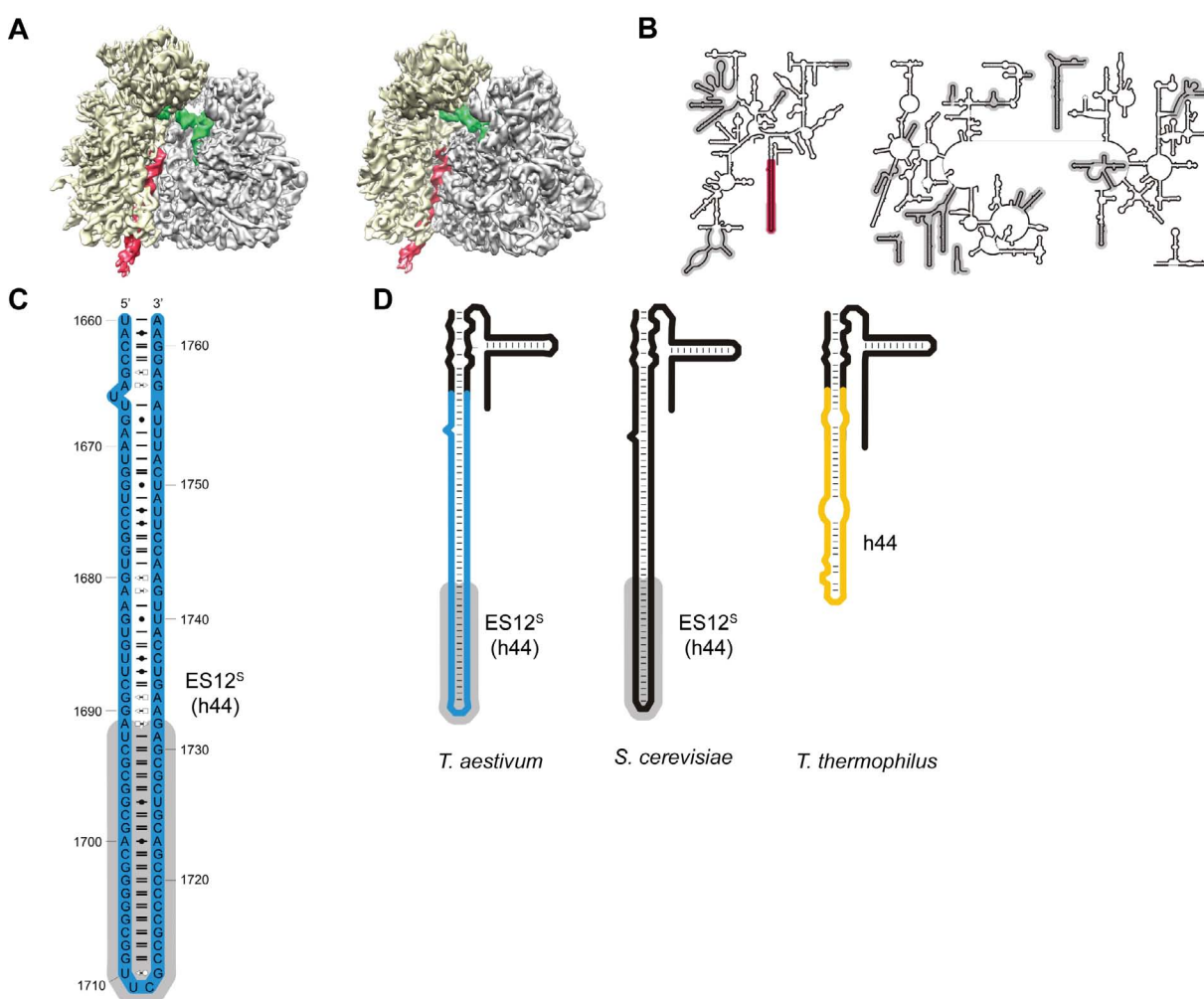


Figure 9.2.19.: **Expansion segment $ES12^S$.** (A) Location of $ES12^S$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES12^S$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *T. thermophilus* (yellow), respectively.

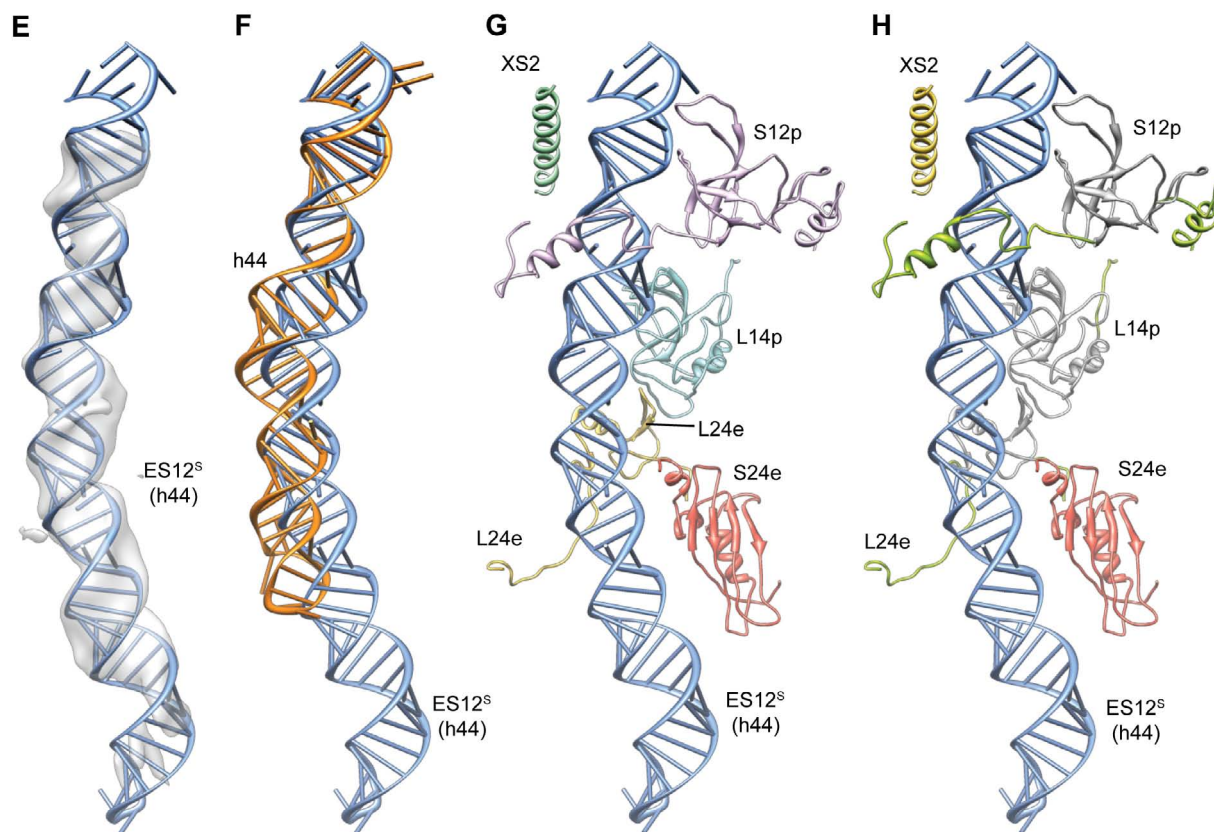


Figure 9.2.20.: **Expansion segment $ES12^S$.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its bacterial counterpart (orange). (G and H) Interaction of $ES12^S$ (blue) with r-proteins S24e (red), L24e (yellow), L14p (blue), XS2 (green) and S12p (pale purple). (H) Same as in (G) but eukaryote-specific proteins (red) and protein extensions (light green) of bacterial homologous (gray).

9.2.2.10. $ES3^L$ / $ES4^L$ (H9, H10)

The adjacent expansion segment $ES3^L$ and $ES4^L$ are located on the back of the large ribosomal subunit (Figure 9.2.21 A). $ES3^L/ES4^L$ are part of domain I of the 5.8S/25S rRNA (Figure 9.2.21 B, C). $ES3^L$ and the 5' strand of $ES4^L$ are part of the 5.8S rRNA while the 3' strand of $ES4^L$ is part of 25S rRNA (Figure 9.2.21 B-D). In both EM densities $ES3^L$ and $ES4^L$ are observed as straight helices for both species (Figure 9.2.21 E). $ES3^L$ is the extension of helix 9 (H9) which is present in *E. coli* but not in *H. marismortui*. In *S. cerevisiae* $ES3^L$ is formed by only Watson-Crick base pairs, while $ES3^L$ in *T. aestivum* has two innerhelical non-Watson-Crick A-G base pair (tHS) (Figure 9.2.21 C). $ES4^L$ is the extension of helix 10 (H10) folding perpendicular to $ES3^L$ (Figure 9.2.21 E, F). The N-terminal extension of L23p interacts with the backbone of $ES3^L$ and with the major and minor groove of $ES4^L$ (Figure 9.2.21G, H). The C-terminal extension of L29p interacts with the backbone of $ES4^L$. The ribosomal protein L15e stabilizes $ES4^L$ via backbone and minor groove interactions (Figure 9.2.21 G, H).

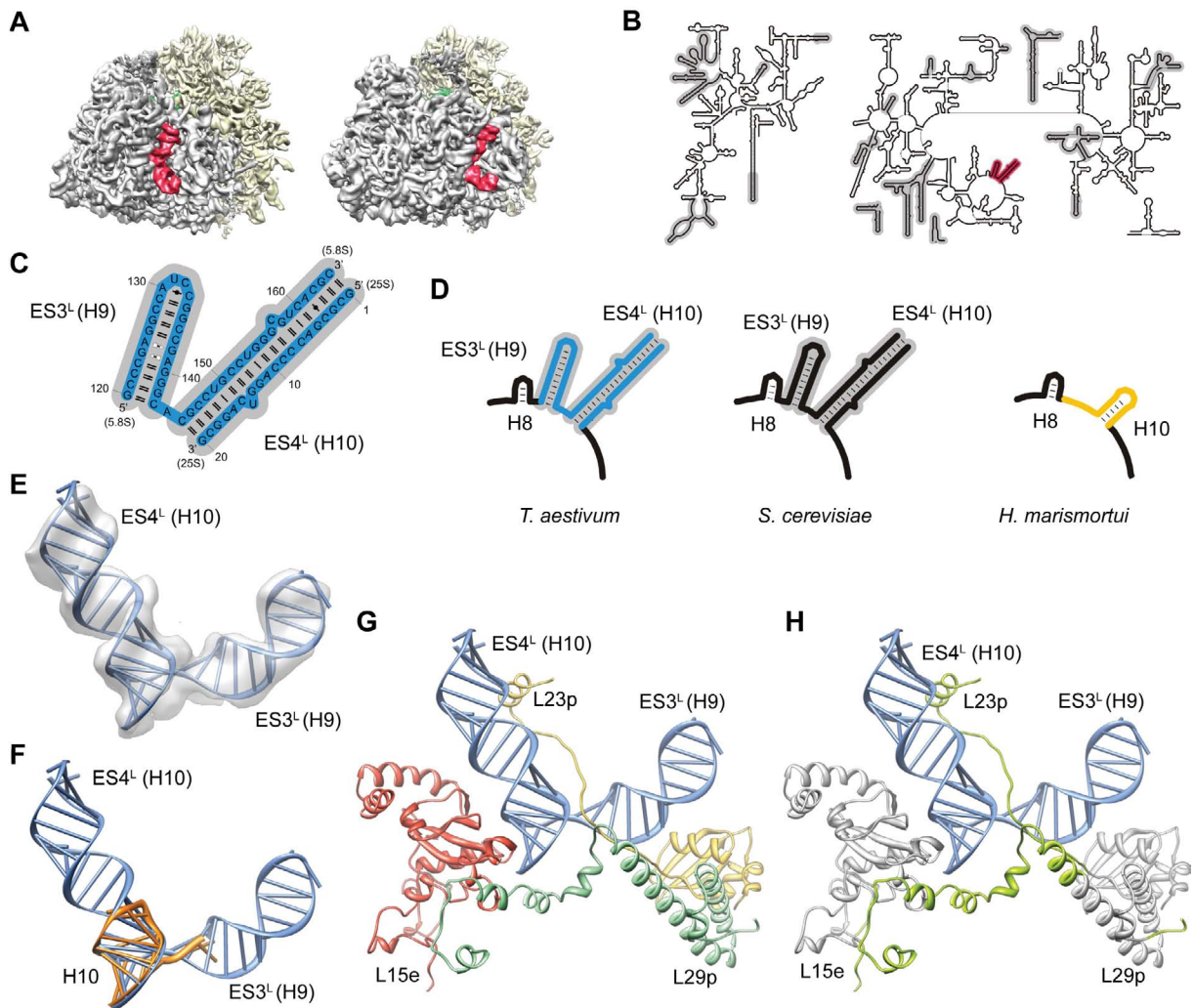


Figure 9.2.21.: **Expansion segment $ES3^L$ and $ES4^L$.** (A) Location of $ES3^L/ES4^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES3^L/ES4^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES3^L/ES4^L$ (blue) with r-proteins L15e (red), L23p (yellow) and L29p (green). (H) Same as in (G) but eukaryote-specific protein extensions (light green) of bacterial homologous (gray).

9.2.2.11. $ES5^L$ (H15, H16-H18)

Expansion segment $ES5^L$ and the adjacent variable region of H16-H18 are located on the back of the large subunit (Figure 9.2.22 A). $ES5^L$ is the extension of helix 15 (H15) and part of domain I of the 25S rRNA (Figure 9.2.22 B). The junction between H14 and H15 could not be modeled, due to ambiguous density. The EM densities for both species show a direct interaction between $ES5^L$ and the variable region of H16-H18 (Figure 9.2.22 E). This interaction was identified to be the loop of $ES5^L$ and the minor groove of H17 and explains the rearrangement of the variable region (H16-H18) in *S. cerevisiae* and *T. aestivum* (Figure 9.2.22 E, F). $ES5^L$ interacts via the minor groove with the adjacent minor groove of $ES4^L$. The junction between $ES5^L$ and H16 is stabilized by L7ae, L15e and the eukaryote-specific r-protein L36e (Figure 9.2.22 G, H). The α -helix of the novel r-protein L41e interacts with the major groove of $ES5^L$. The C-terminal extension of L29p stabilizes the interaction between the loop of $ES5^L$ and the rearranged H17. The altered conformation of H16-H18 is also caused by the presence of the newly identified r-protein L13e, which pushes H17 towards $ES5^L$ (Figure 9.2.22 G, H). The bended H18 is stabilized by the C-terminal extended helix of L24p.

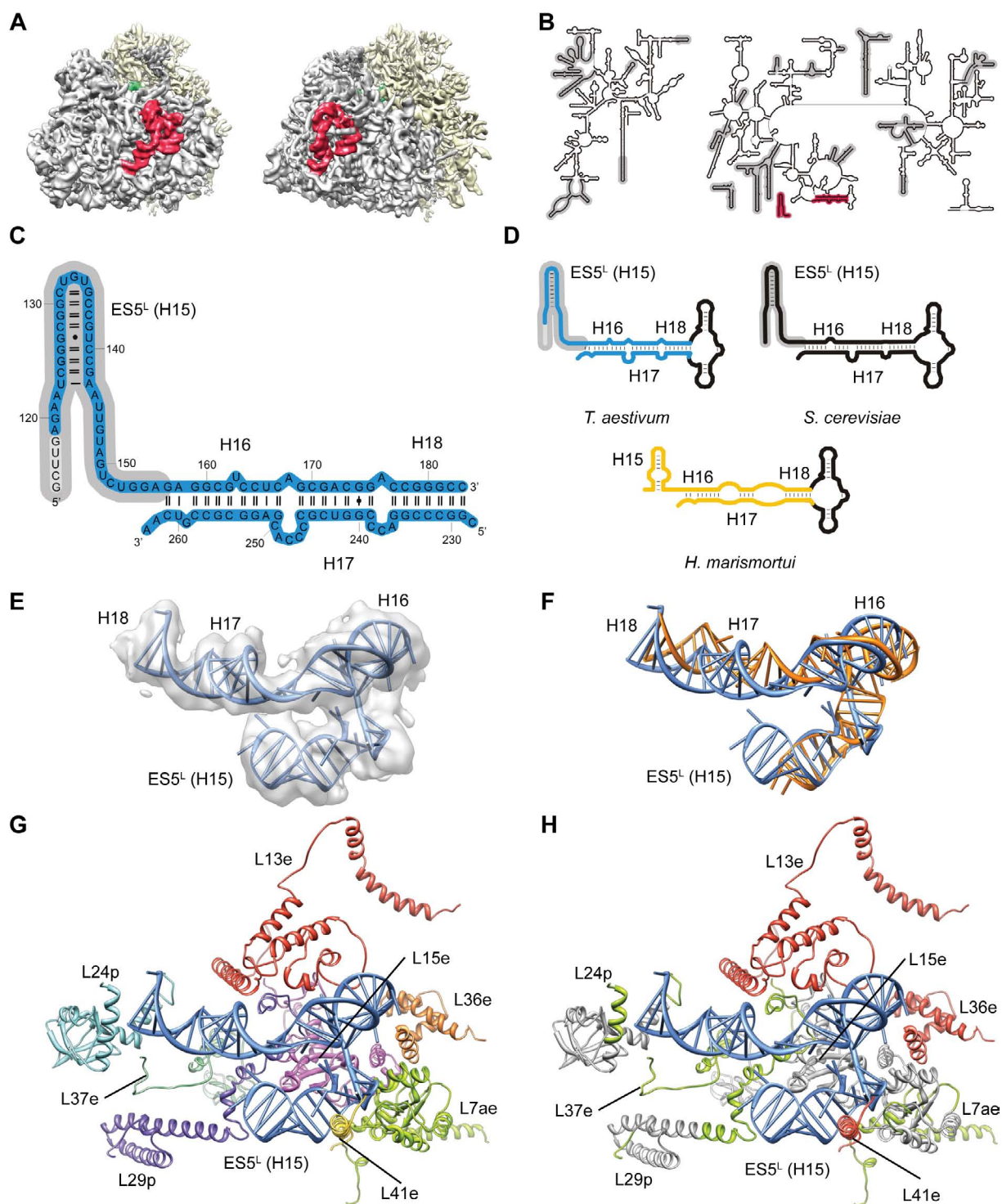


Figure 9.2.22.: **Expansion segment $ES5^L$ and variable region H16-H18.** (A) Location of $ES5^L$ /H16 – H18 (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES5^L$ /H16 – H18 (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES5^L$ /H16 – H18 (blue) with r-proteins L13e (red), L41e (yellow), L36e (orange), L15e (magenta), L29p (purple), L24p (blue) and L7ae (green). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

9.2.2.12. $ES7^L$ of *T. aestivum*

The largest expansion segment on the large subunit is $ES7^L$ located at the back of the 60S subunit (Figure 9.2.23 A). $ES7^L$ emerges from H25 and is part of domain I of the 25S rRNA (Figure 9.2.23 B). $ES7^L$ consists of 6 helices, $ES7^La - ES7^Le$, and three RNA three-way junctions (Figure 9.2.23 B-D). The first three-way junction spans between H25, $ES7^La$ and $ES7^Lb$; the second one between the first and the second part of $ES7^Lb$ and $ES7^Lc$ and the last three-way junction between $ES7^Lc$, $ES7^Ld$ and $ES7^Le$ (Figure 9.2.23 C).

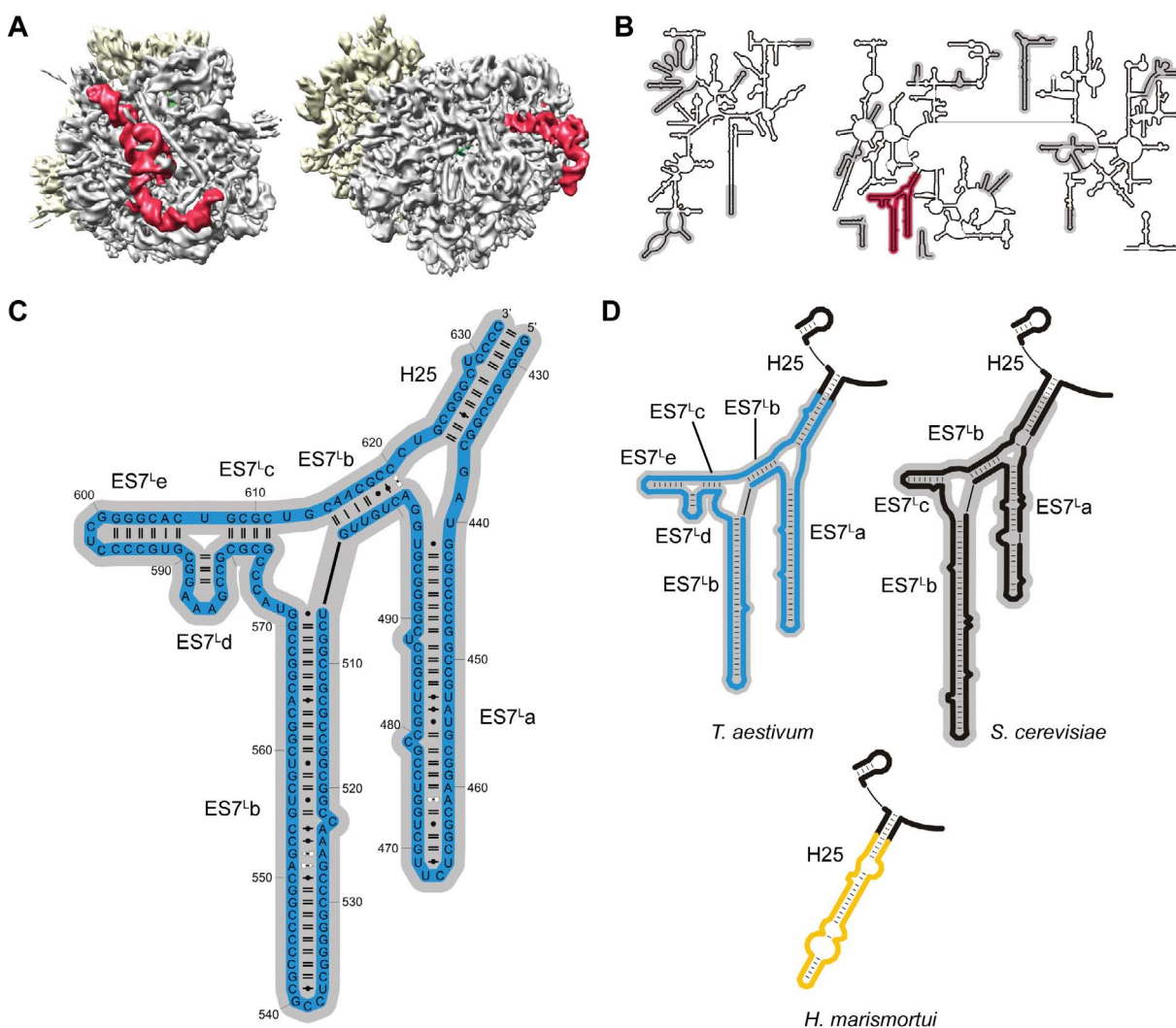


Figure 9.2.23.: **Expansion segment $ES7^L$ of *T. aestivum*.** (A) Location of $ES7^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES7^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively.

The first helix $ES7^L a$ is observed to be bended in the *T. aestivum* EM density (Figure 9.2.24 E). $ES7^L b$ is the eukaryotic counterpart of the bacterial H25. $ES7^L c$ folds perpendicular to $ES7^L b$ and positions the three-helix-bundle consisting of $ES7^L c$, $ES7^L d$ and $ES7^L e$ above $ES7^L b$ (Figure 9.2.24 E, F). All helices are clearly visible suggesting that $ES7^L$ forms a stable structure together with the intertwined r-proteins. The r-protein L13p contacts the backbone of H25. The eukaryote-specific r-protein L35a interacts with H25 and the backbone of $ES7^L b$. The first three-way junction and the lower region of $ES7^L a$ is stabilized by the eukaryote-specific r-protein L32e, while the upper region of $ES7^L a$ is stabilized by another eukaryote-specific r-protein, L28e, which is only present in *T. aestivum*. Helix $ES7^L b$ interacts with five r-proteins, namely the eukaryote-specific r-proteins L6e, L14e and L18a, as well as eukaryote-specific N-terminal extension of L30p and the first helix of the C-terminal extension of L4p (Figure 9.2.24 G, H). Further this C-terminal extension of L4p interacts with the loop of $ES7^L e$. The long N-terminal α -helix of L30p interacts with the minor groove of $ES7^L e$ and the loop of $ES7^L a$. The eukaryote-specific r-protein L6e plays a special role in stabilizing $ES7^L$. Its N-terminal α -helix inserts through the third three-way junction of $ES7^L$. The second α -helix interacts with the loop of $ES7^L d$ and the second three-way junction. Finally, the C-terminal domain of L6e interacts with the first three-way junction consisting of $ES7^L a$, $ES7^L b$ and H25 (Figure 9.2.24 G, H).

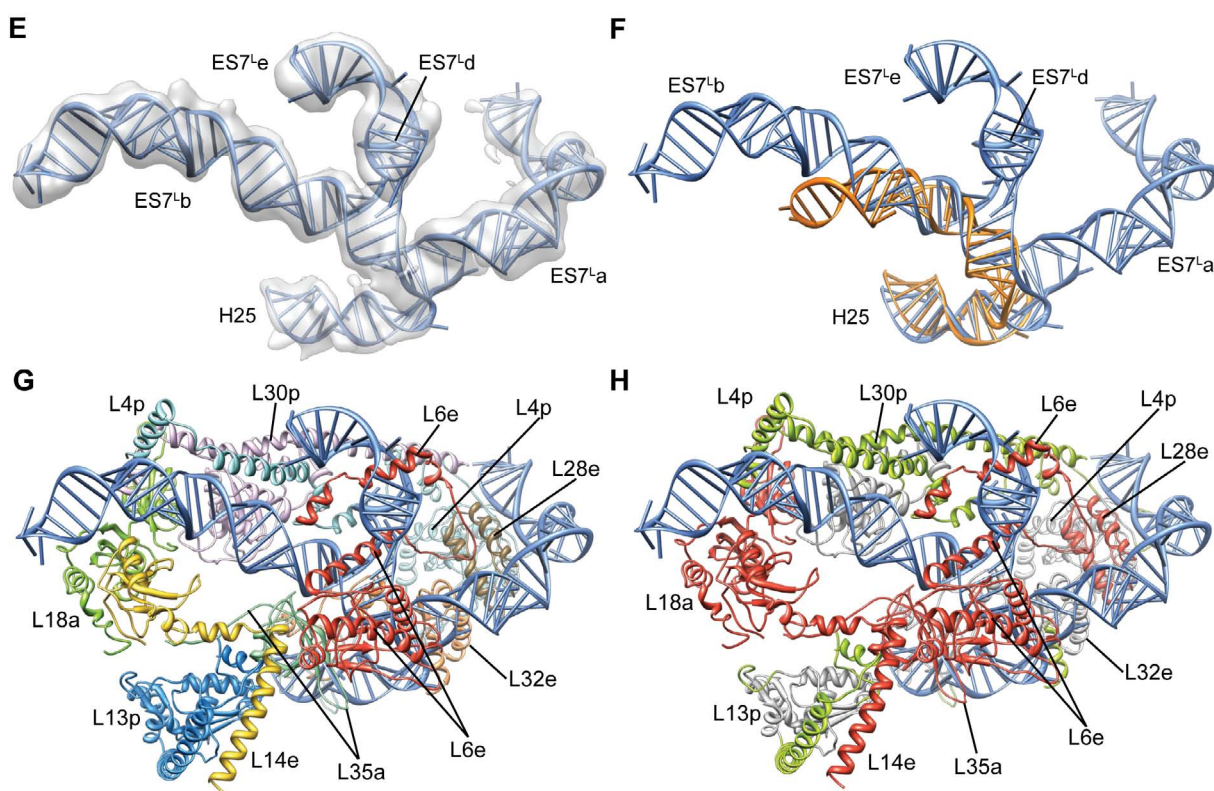


Figure 9.2.24.: **Expansion segment $ES7^L$ of *T. aestivum*.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES7^L$ (blue) with r-proteins L6e (red), L14e (yellow), L32e (orange), L4p (cyan), L30p (pale purple), L13p (blue), L35a (mint green), L28e (brown) and L18a (green). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

9.2.2.13. $ES7^L$ of *S. cerevisiae*

Expansion segment $ES7^L$ is the largest ES located on the back of the 60S subunit (Figure 9.2.25 A). $ES7^L$ emerges from H25 and is part of domain I of the 25S rRNA (Figure 9.2.25 B). $ES7^L$ consists of three main helices $ES7^L_a$, $ES7^L_b$ and $ES7^L_c$. There are two RNA three-way junctions. The first three-way junction is between helices H25, $ES7^L_a$ and $ES7^L_b$. The second spans between the first and second part of $ES7^L_b$ and $ES7^L_c$ (Figure 9.2.25 B-D).

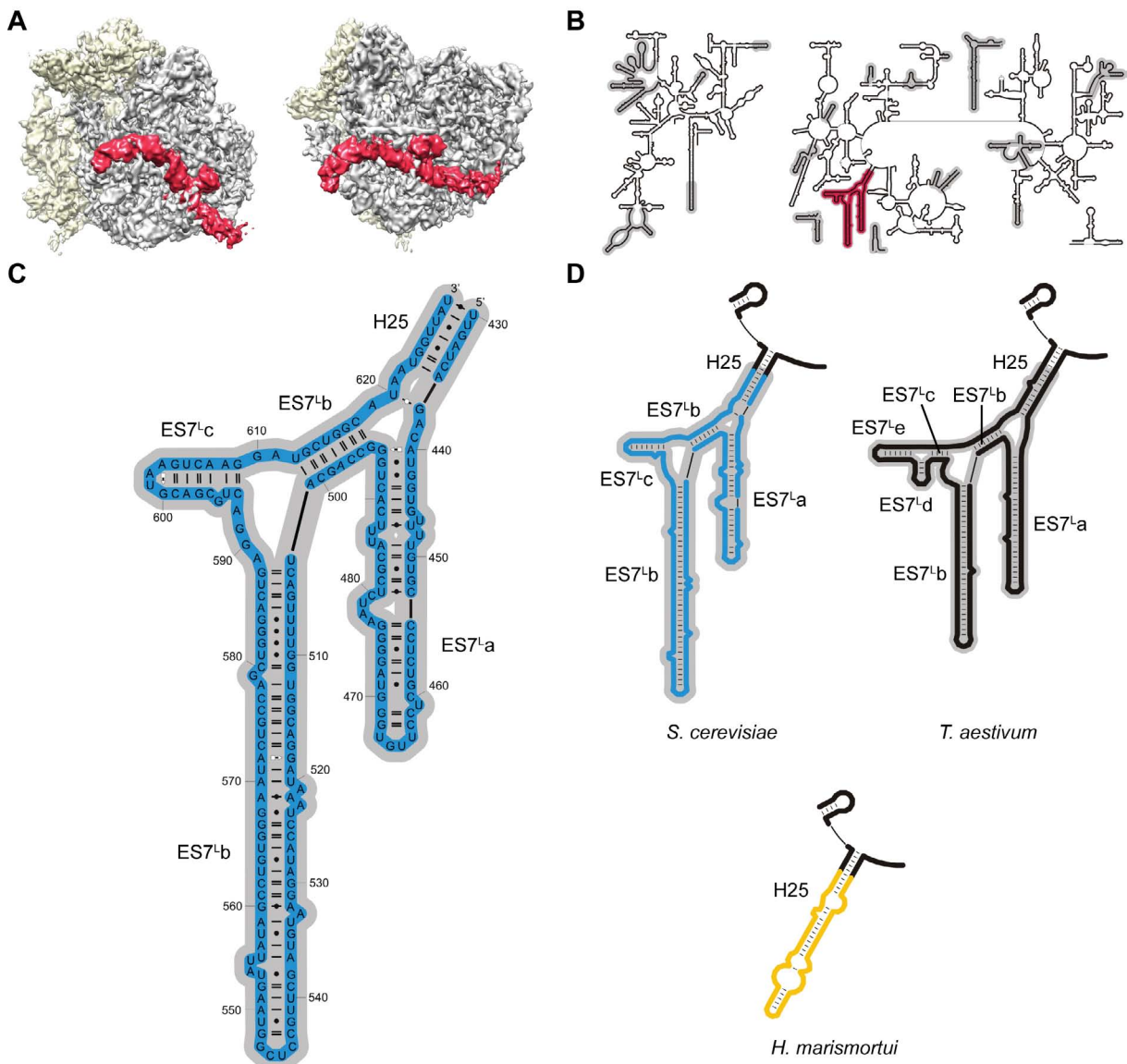


Figure 9.2.25.: **Expansion segment $ES7^L$ of *S. cerevisiae*.** (A) Location of $ES7^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES7^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *S. cerevisiae* model. (D) RNA secondary structure diagram. *S. cerevisiae* (blue), *T. aestivum* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively.

The first helix $ES7^L a$ is delocalized and not entirely visible in the *S. cerevisiae* cryo-EM map (Figure 9.2.26 E). After two helical turns the helix kinks about 90° which is only observed at lower thresholds in the density (Figure 9.2.27 A-C). Like in *T. aestivum*, $ES7^L b$ is the eukaryotic counterpart of the bacterial H25 and heavily bended. $ES7^L c$ folds perpendicular to $ES7^L b$ (Figure 9.2.26 E, F). $ES7^L$ is closely intertwined with eight r-proteins. The ribosomal protein L13p and the eukaryote-specific r-protein L35a interact with H25 and the backbone of $ES7^L b$. The first three-way junction and the lower region of $ES7^L a$ is stabilized by the eukaryote-specific r-protein L32e (Figure 9.2.26 G, H). Two other r-proteins specific for eukaryotes namely L14e and L18a interact with $ES7^L b$. The long N-terminal α -helix of L30p interacts with the backbone of $ES7^L b$ and the major groove of $ES7^L c$. The C-terminal extension of L4p contacts the backbone and major groove of $ES7^L b$ as well as the minor groove of $ES7^L c$ (Figure 9.2.26 G, H). The N-terminal part of the eukaryote-specific r-protein L6e interacts with $ES7^L c$ while the C-terminal domain stabilizes the first three-way junction (Figure 9.2.26 G, H).

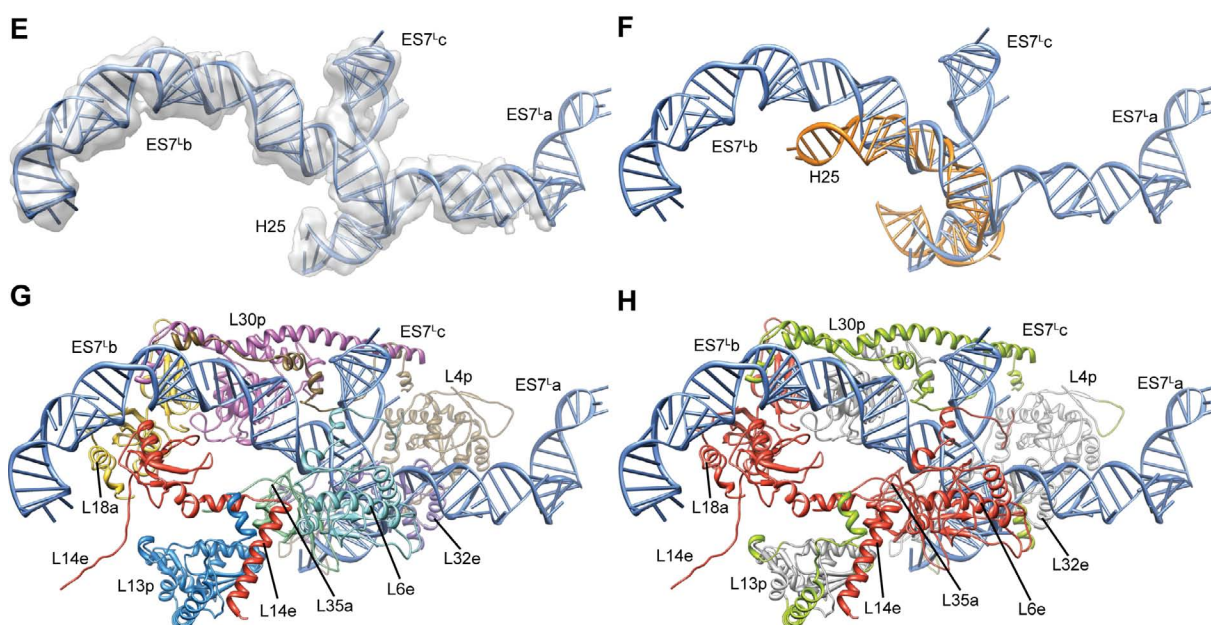


Figure 9.2.26.: **Expansion segment $ES7^L$ of *S. cerevisiae*.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES7^L$ (blue) with r-proteins L14e (red), L18a (yellow), L4p (brown), L30p (pink), L13p (blue) and L35a (mint green), L6e (cyan) and L32e (purple). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

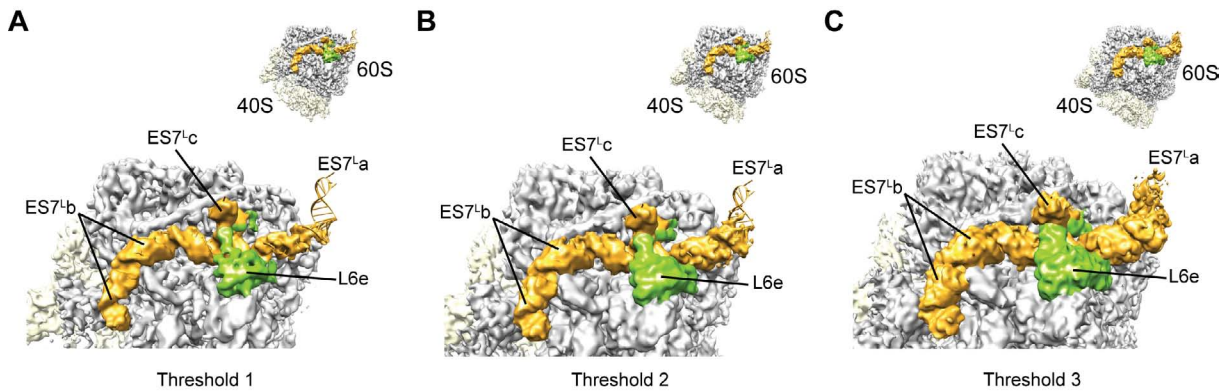


Figure 9.2.27.: **Flexibility of expansion segment $ES7^L$ of *S. cerevisiae*.** Visualization of density for (A-C) *S. cerevisiae* $ES7^L$ (yellow) and r-protein L6e (green) at different contour levels. The density for $ES7^La$ is observed at lower thresholds (C). R-protein L6e interacts with $ES7^Lb, c$. The figure was taken from [182].

9.2.2.14. $ES9^L$ (H31, H28-H30)

Expansion segment $ES9^L$ is located on the back of the large subunit between $ES7^L$ and below the central protuberance (Figure 9.2.28 A). $ES9^L$ is the extension of the bacterial H31 and part of domain II of the 25S rRNA (Figure 9.2.28 B). Compared to the bacterial template helices H28 and H30 have a slightly different secondary structure (Figure 9.2.28 B-D).

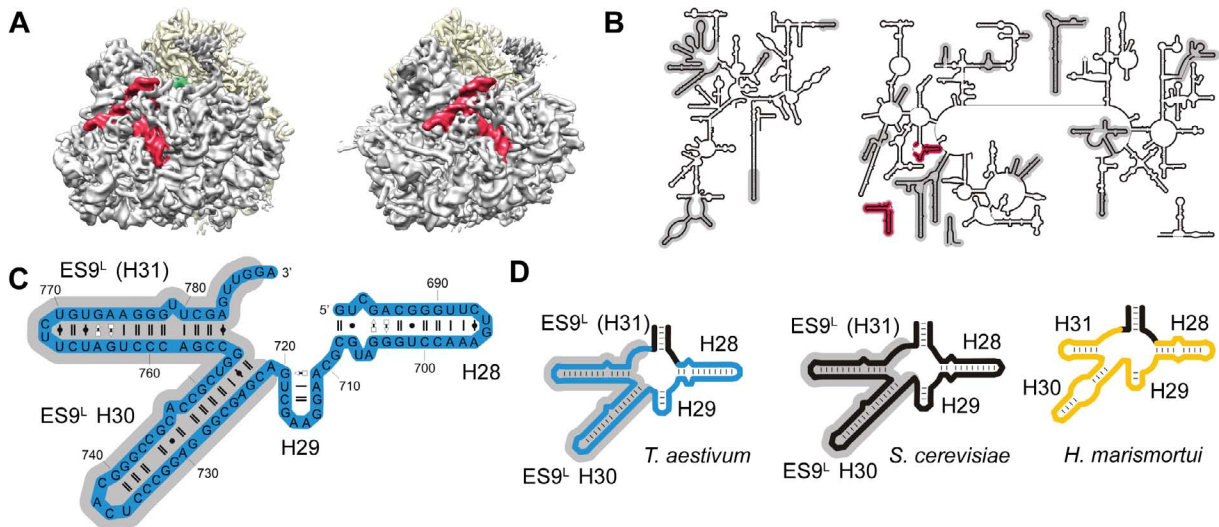


Figure 9.2.28.: **Expansion segment $ES9^L$.** (A) Location of $ES9^L/H28 - H30$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES9^L/H28 - H30$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively.

The loop of H30 is not resolved in the structural template (Figure 9.2.29 F). The model of H28 and H29 folds very similar to the bacterial counterpart (Figure 9.2.29 F). The loop region of H30 is resolved in the cryo-EM density (Figure 9.2.29 E) and the loop was modeled according to the structural templates (see section 6.5). $ES9^L$ coaxial stacks on H30 and is stabilized by the GNRA-tetraloop of H29 (Figure 9.2.29 E). H28 is wrapped by the N-terminal part of L4p, L15p and the N-terminal part of the eukaryote-specific L13e. Helix 29 is pushed towards $ES9^L$ by r-protein L15p (Figure 9.2.29 G, H). The C-terminal extensions of L18e and L21e stabilize the coaxial stacked H30 and H31 (Figure 9.2.29 G, H). The two C-terminal α -helices of L13e interact with the major groove and the loop $ES9^L$. An unknown r-protein named XL2 interacts with the lower region of $ES9^L$ (Figure 9.2.29 G, H).

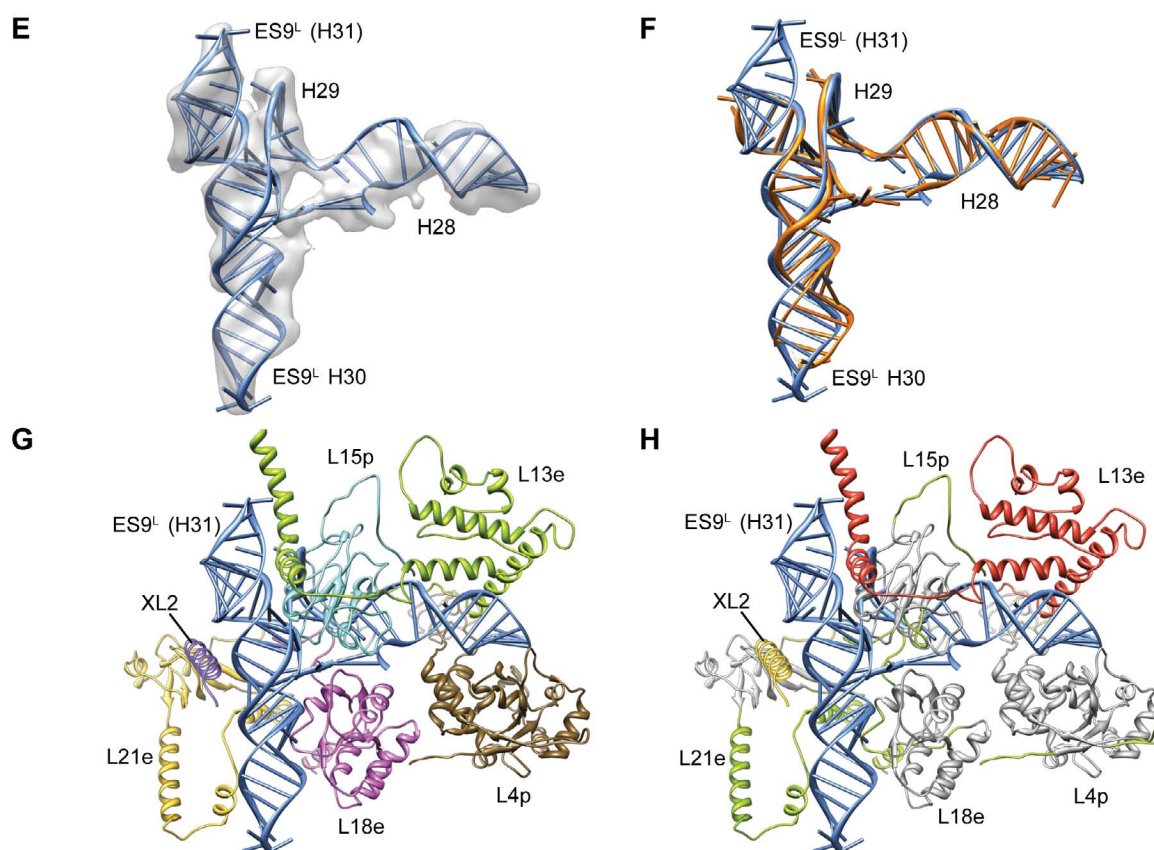


Figure 9.2.29.: **Expansion segment $ES9^L$.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES9^L/H28 - H30$ (blue) with r-proteins L13e (green), L21e (yellow), L18e (dark pink), L15p (cyan), L4p (brown) and XS2 (dark purple). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

9.2.2.15. $ES10^L$ / $ES12^L$ (H38)

Expansion segments $ES10^L$ and $ES12^L$ are located on the back of the large subunit above $ES7^L$ and below the central protuberance (Figure 9.2.30 A). $ES10^L/ES12^L/H38$ are part of domain II of the 25S rRNA (Figure 9.2.30 B). The A-site finger (H38) reaches over the inter subunit space to the small subunit.

In *T. aestivum* and *S. cerevisiae* $ES10^L$ has no secondary structure suggesting that $ES10^L$ is single stranded or is involved in only non-Watson-Crick interactions (Figure 9.2.30 B-D). $ES12^L$ is predicted as helical with a closing tetraloop (Figure 9.2.30 B-D).

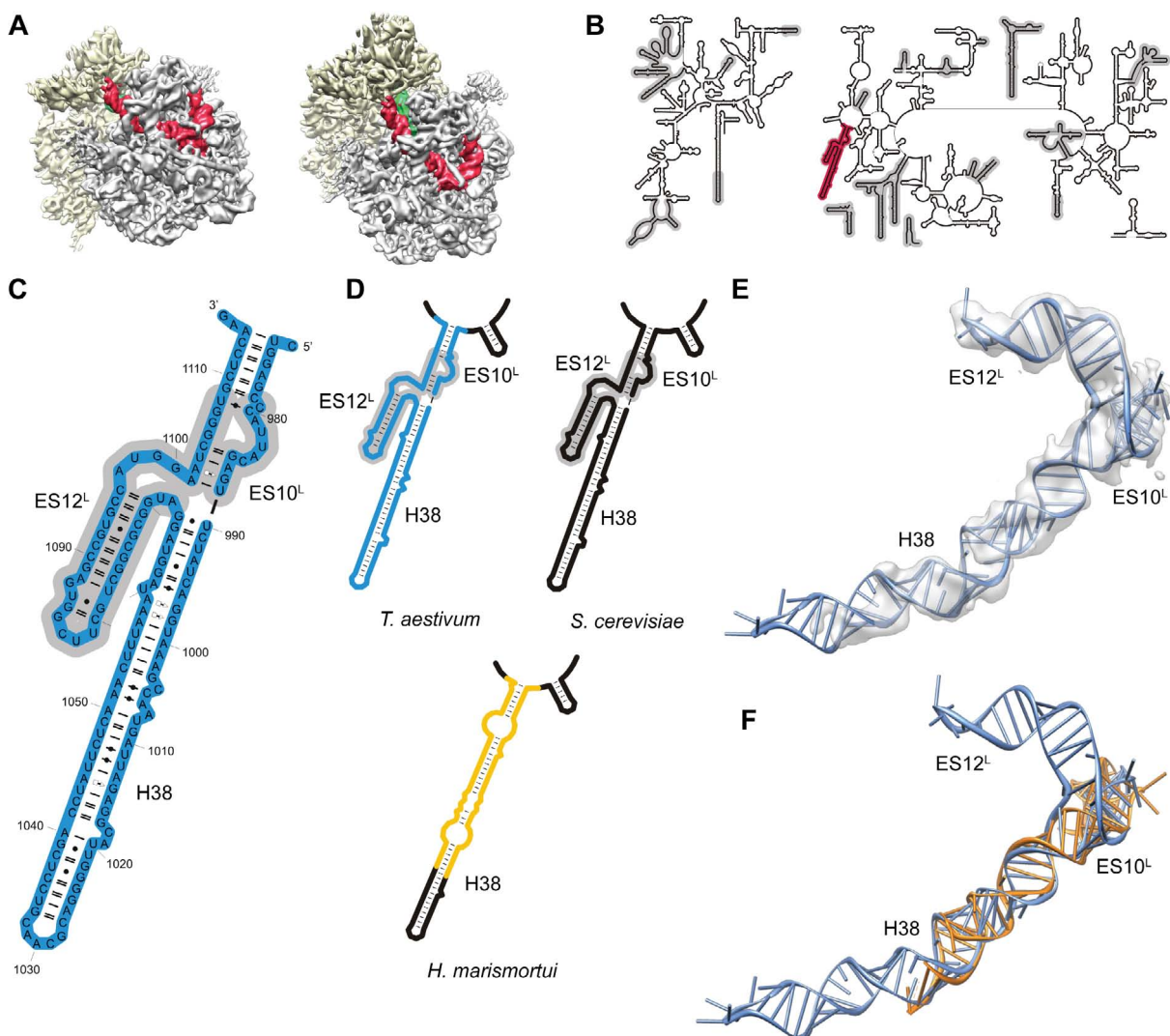


Figure 9.2.30.: **Expansion segment $ES10^L/ES12^L$ and H38.** (A) Location of $ES10^L/ES12^L$ and H38 (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES10^L/ES12^L$ and H38 (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange).

According to the secondary structure prediction $ES10^L$ was modeled as single stranded region (Figure 9.2.30 E). The loop of H38 is neither resolved in the crystal structure of *H. marismortui* (Figure 9.2.30 F, PDB 1FFK, PDB 1VQ8) [7] nor it is clearly resolved in the EM density for both *T. aestivum* and *S. cerevisiae*. The model is based on a EM density filtered between 8-9 Å. $ES12^L$ folds perpendicular to H38 (Figure 9.2.30 E, F). $ES10^L$ interacts with the N-terminal part of L18e. The C-terminal domain of L30p and the C-terminal extension of L21e wrap around the three-way junction between H38, $ES10^L$ and $ES12^L$ (Figure 9.2.31 G, H). H38 interacts with the N-terminal domain of L21e, the C-terminal extension of L10e and L18p. Further, the extensions of small subunit r-proteins S19p (C-terminal) and S13p (N-terminal) seem to interact with the loop region of H38 (Figure 9.2.31 G, H). $ES12^L$ is intertwined by L18p, the C-terminal extension of L18e, the C-terminal extension of L21e and an unknown r-protein XL2 (Figure 9.2.31 G, H).

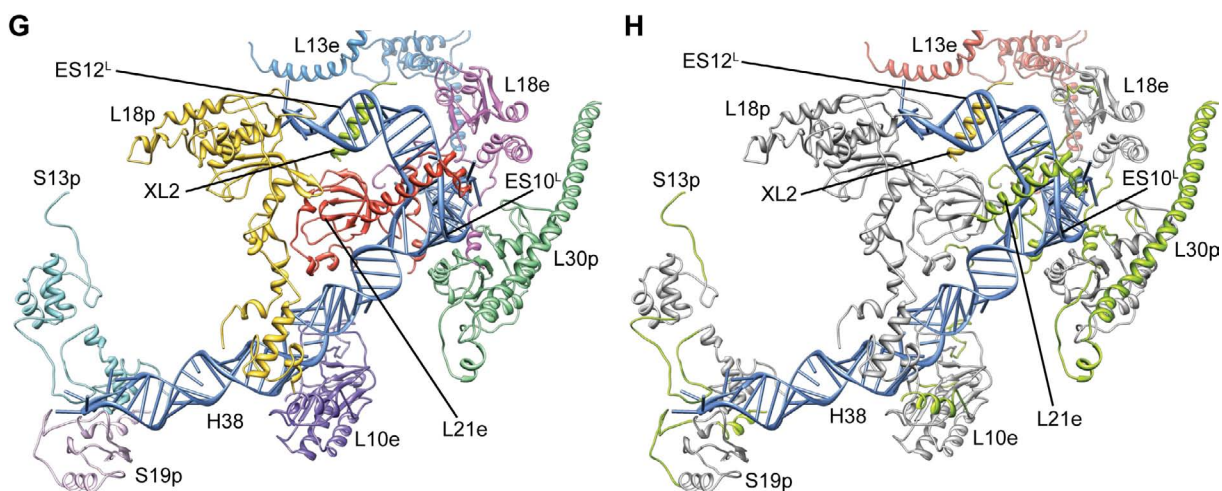


Figure 9.2.31.: **Expansion segment $ES10^L/ES12^L$ and H38.** (G and H) Interaction of $ES10^L/ES12^L$ and H38 (blue) with r-proteins L21e (red), L18p (yellow), L30p (mint green), L13e (blue), S13p (cyan), L10e (purple), S19p (pale purple), L18e (magenta) and XL2 (green). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

9.2.2.16. $ES15^L$ (H45)

Expansion segments $ES15^L$ is located on the back of the large subunit buried between $ES7^Lc$ – $ES7^Le$ and the ribosomal protein L28e (Figure 9.2.32 A). $ES15^L$ is part of domain II of the 25S rRNA (Figure 9.2.32 B). In *T. aestivum* and *S. cerevisiae* $ES15^L$ is a short extension of helix 45 (H45) folding as a normal A-form helix (Figure 9.2.32 B-F). The EM density for the loop region of $ES15^L$ is ambiguous (Figure 9.2.32 E). The stem part of $ES15^L/H45$ is stabilized by short α -helices of L4p (C-terminal extension) and L18e (N-terminal extension) which interact with the minor groove. The loop of $ES15^L$ interacts with the long N-terminal extended α -helix of L30p, the C-terminal extension of L4p, the N-terminal extension of L18e and the eukaryote-specific r-protein L28e (Figure 9.2.32 G, H).

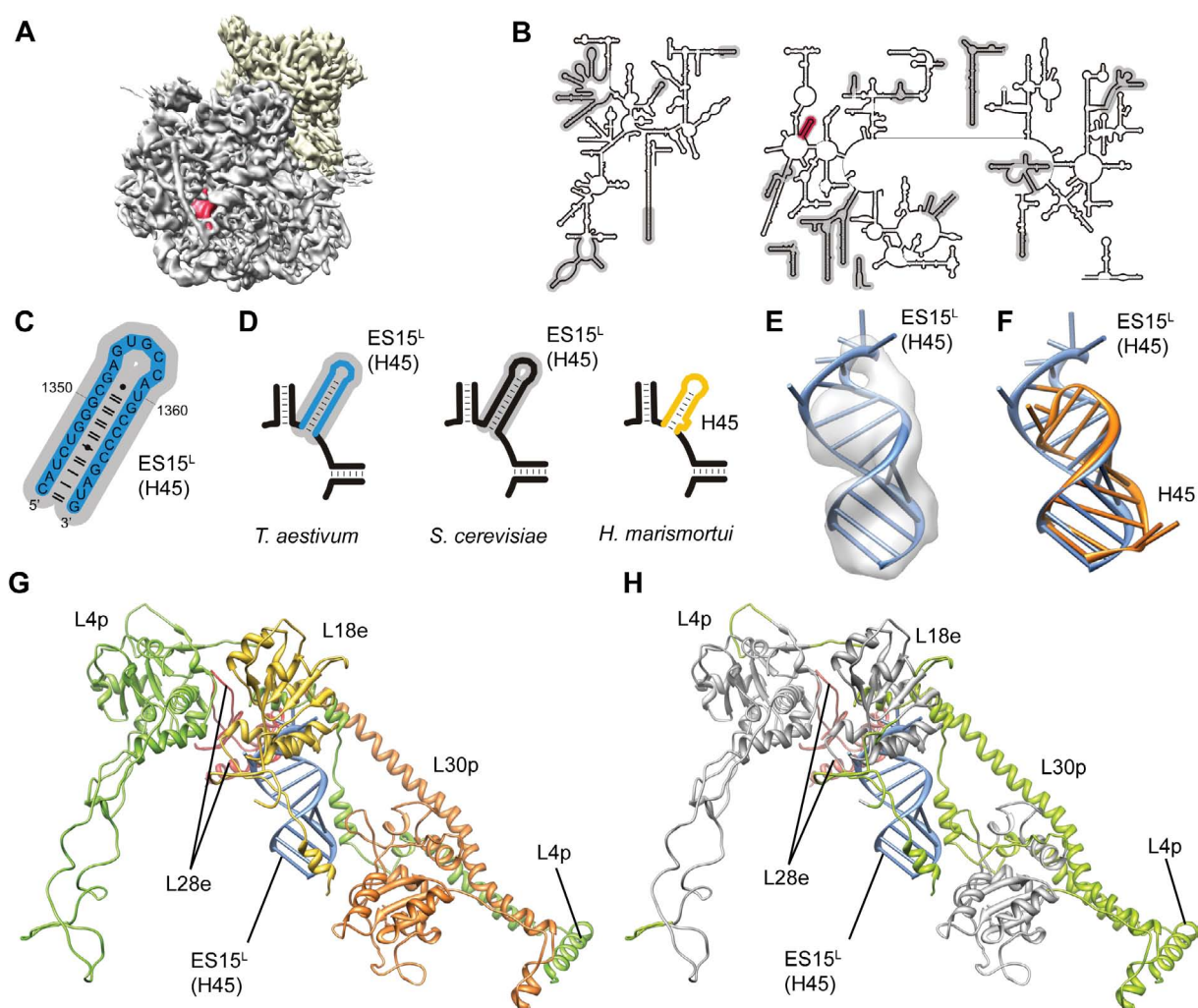


Figure 9.2.32.: **Expansion segment $ES15^L$.** (A) Location of $ES15^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES15^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES15^L$ (blue) with r-proteins L28e (red), L30p (orange), L18e (yellow) and L4p (green). (H) Same as in (G) but eukaryote-specific protein extensions (light green) of bacterial homologs (gray).

9.2.2.17. $ES19^L$ (H52)

Expansion segment $ES19^L$ is located at the bottom of the large subunit between $ES3^L$ and $ES31^L$ (Figure 9.2.33 A). In the RNA secondary structure, $ES19^L$ is part of domain III (Figure 9.2.33 B) and is predicted to be helical (Figure 9.2.33 C, D). $ES19^L$ and the two helical parts of helix 52 (H52) are arranged in an RNA three-way junction (H52a, H52b, $ES19^L$). The bacterial GNRA-tetraloop of H52 is substituted by another tetraloop in both species. $ES19^L$ folds inverse to H52 (Figure 9.2.33 E, F). In *S. cerevisiae*, $ES19^L$ is three base pairs longer and bended towards $ES3^L$ suggesting an RNA-RNA interaction between the two expansion segments. $ES19^L$ interacts with the C-terminal extension of L7ae and the N-terminal extension of L23p (Figure 9.2.33 G, H).

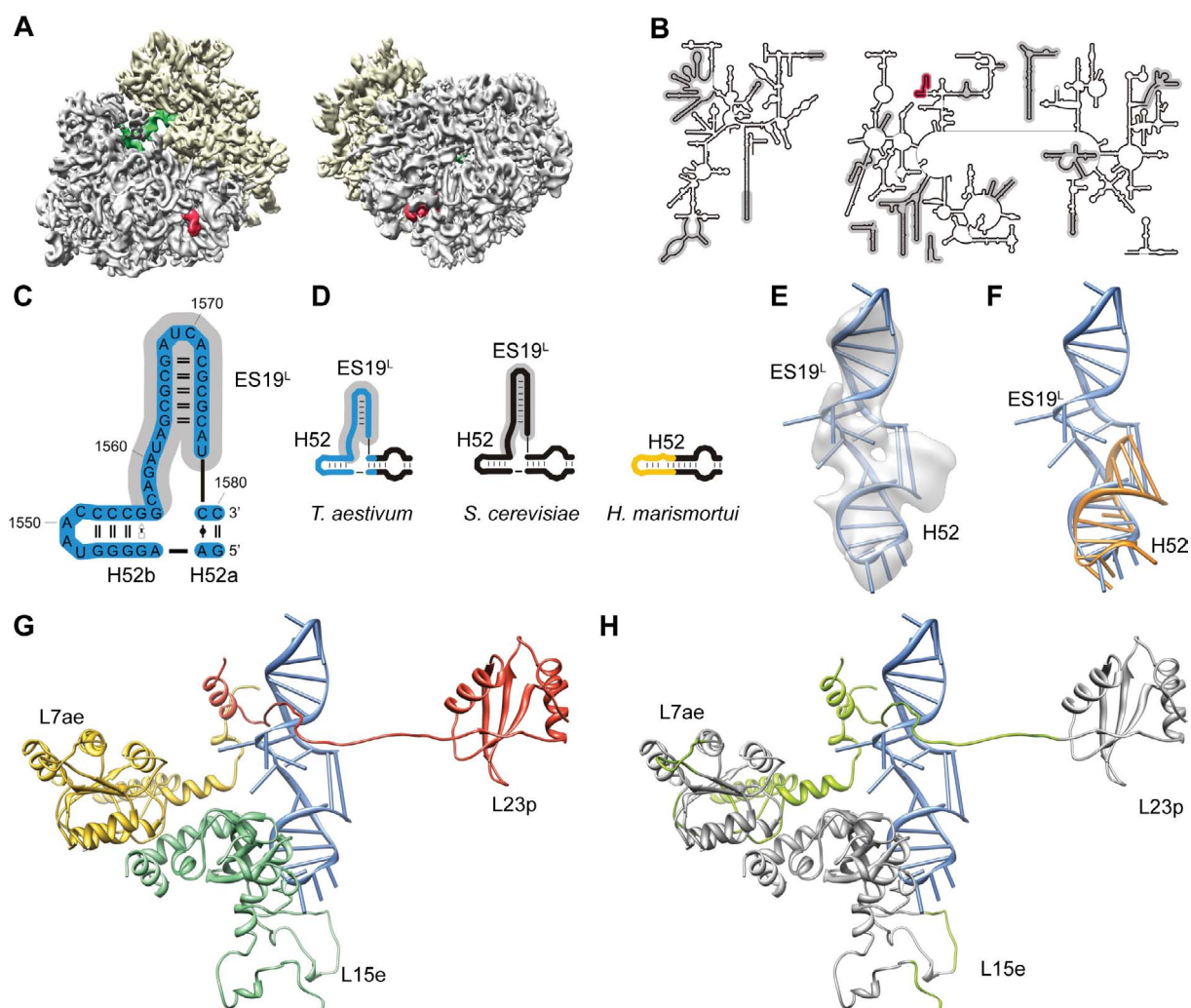


Figure 9.2.33.: **Expansion segment $ES19^L$.** (A) Location of $ES19^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES19^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES19^L$ (blue) with r-proteins L23p (red), L7ae (yellow) and L15e (mint green). (H) Same as in (G) but eukaryote-specific protein extensions (light green) of bacterial homologous (gray).

9.2.2.18. $ES20^L$ / $ES26^L$ (H54, H55)

In eukaryotes helices H54 and H55 are interrupted by two expansion segments namely $ES20^L$ and $ES26^L$. They are part of domain III and located at the bottom of the 60S subunit (Figure 9.2.34 A, B). $ES20^L$ is a small helix closed by a GNRA-tetraloop in *T. aestivum* and a GNRA-like pentaloop in *S. cerevisiae*, respectively (Figure 9.2.34 C-F). $ES26^L$ folds as single stranded part suggesting to be involved in tertiary non-Watson-Crick interactions (Figure 9.2.34 C-F). $ES20^L$ is wrapped by two eukaryote-specific r-proteins L27e and L34e which interact via their β -sheet structures (Figure 9.2.34 G, H).

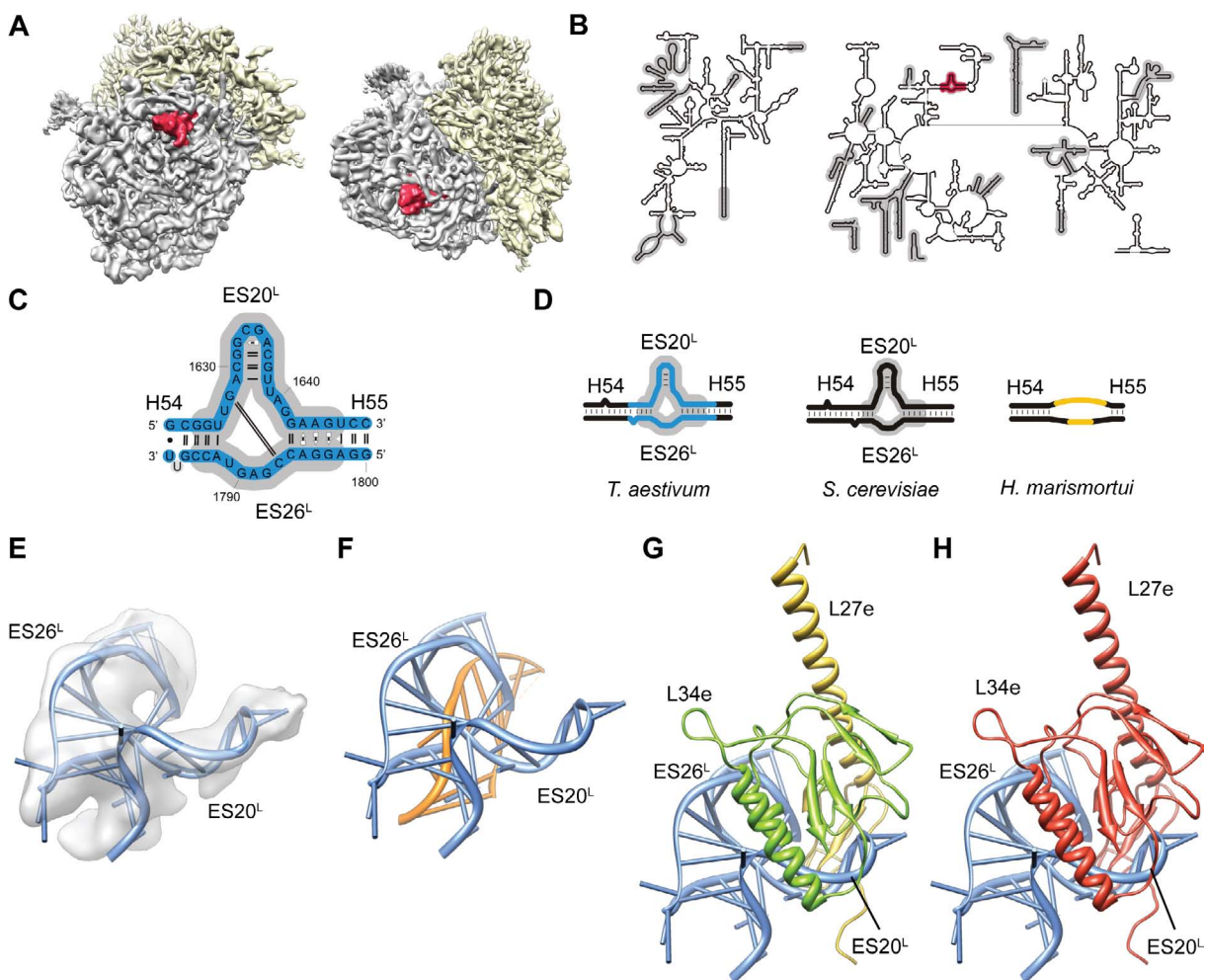


Figure 9.2.34.: **Expansion segment $ES20^L/ES26^L$.** (A) Location of $ES20^L/ES26^L$ (red) on the ribosome. The small and large subunit in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES20^L/ES26^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (blue) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES20^L/ES26^L$ (blue) with r-proteins L34e (green) and L27e (yellow). (H) Same as in (G) but eukaryote-specific r-proteins (red).

9.2.2.19. $ES24^L$ (H59)

Expansion segment $ES24^L$ is located near the tunnel exit site of the large subunit (Figure 9.2.35 A). $ES24^L$ is the extension of H59 and part of domain III (Figure 9.2.35 B). In *T. aestivum* $ES24^L$ extends H59 by three base pairs closed by a pentaloop (Figure 9.2.35 E, F). In *S. cerevisiae* the extension consists of 5 base pairs closing with a tetraloop. In both cryo-EM densities $ES24^L$ is clearly visible and interacts with L19e via the backbone and with the eukaryote-specific r-protein L22e via the minor groove, respectively (Figure 9.2.35 G, H).

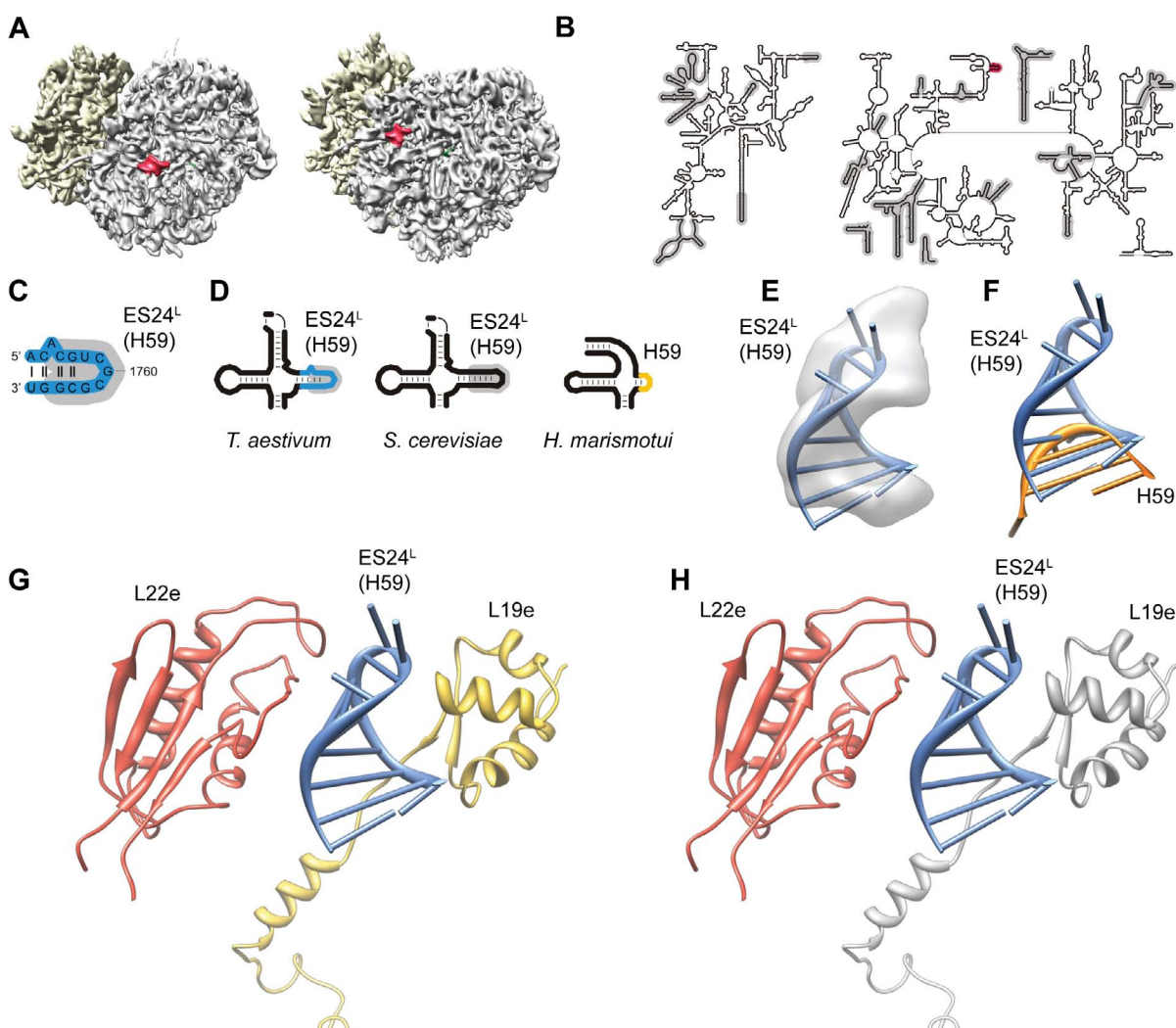


Figure 9.2.35.: **Expansion segment $ES24^L$.** (A) Location of $ES24^L$ (red) on the ribosome. The small and large subunit in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES24^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES24^L$ (blue) with r-proteins L22e (red) and L19e (yellow). (H) Same as in (G) but eukaryote-specific protein extensions (light green) of bacterial homologous (gray) and eukaryote-specific r-proteins (red).

9.2.2.20. $ES27^L$ of *T. aestivum*

Expansion segment $ES27^L$ is located on the opposite site of the central protuberance on the large subunit (Figure 9.2.36 A). $ES27^L$ emerges from H63 and is part of domain IV (Figure 9.2.36 B). In *T. aestivum* $ES27^L$ is 150 nts long and consists of three helices $ES27^L_a$, $ES27^L_b$ and $ES27^L_c$ arranged in a three-way junction (Figure 9.2.36 C, D). The cryo-EM density for H63 is strong while the density for $ES27^L$ is weak (Figure 9.2.36 E). Only at lower contour level the orientation of $ES27^L_b$ is traceable (Figure 9.2.38 A-C) suggesting a preferred conformation of $ES27^L_b$ above r-protein L38e.

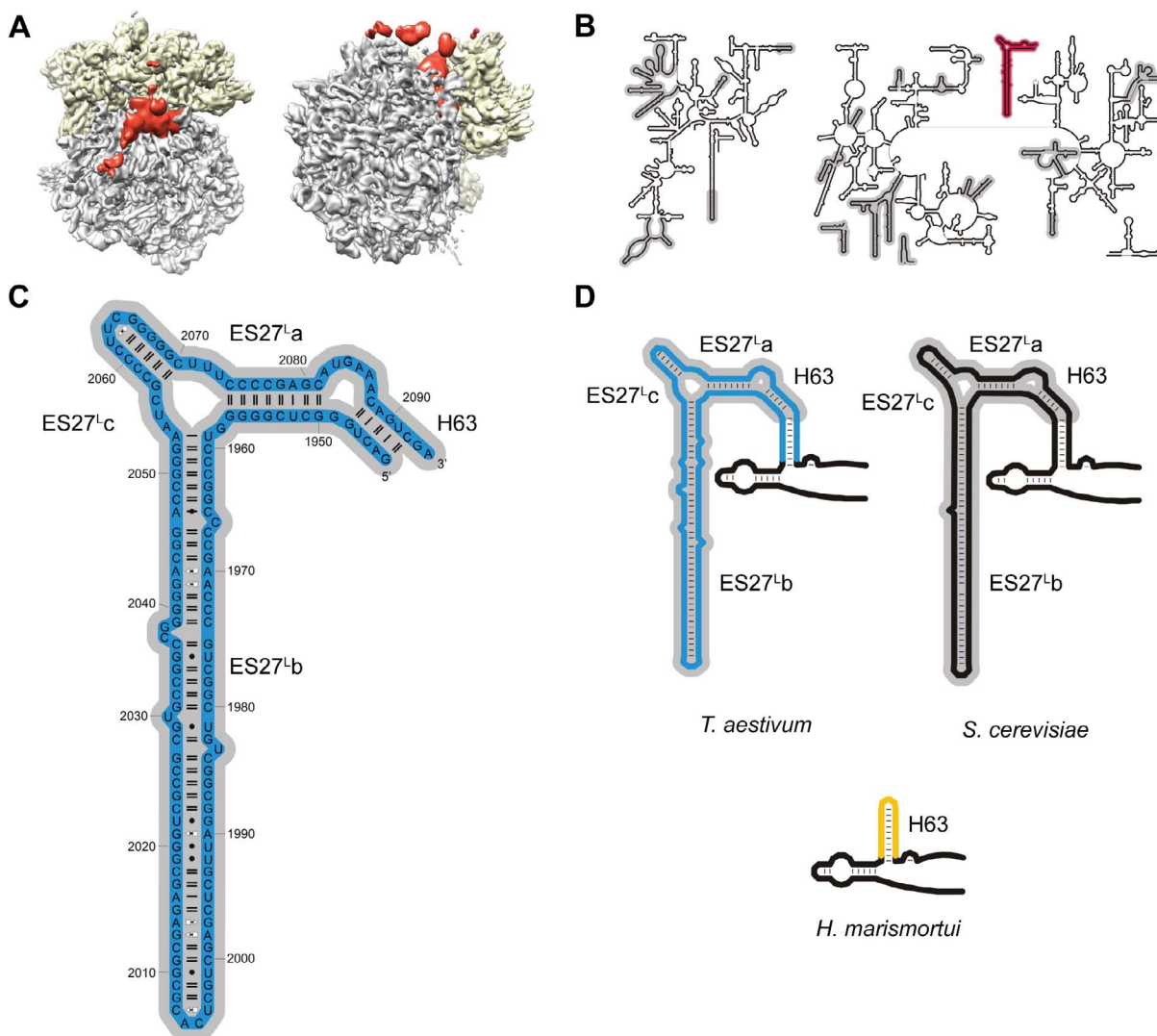


Figure 9.2.36.: **Expansion segment $ES27^L$ of *T. aestivum*.** (A) Location of $ES27^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES27^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively.

Helix 63 is stabilized by L19e (major groove interaction), S17p (backbone interaction) and the eukaryote-specific r-protein L22e interacting with the backbone and minor groove (Figure 9.2.37 G, H). The tight interactions prevent H63 to move and explain the stable conformation of H63 (Figure 9.2.37 E, F). *ES27^{Lb}* interacts with L38e and L34e (Figure 9.2.37 G, H and Figure 9.2.38 A-C), which are only present in eukaryotes.

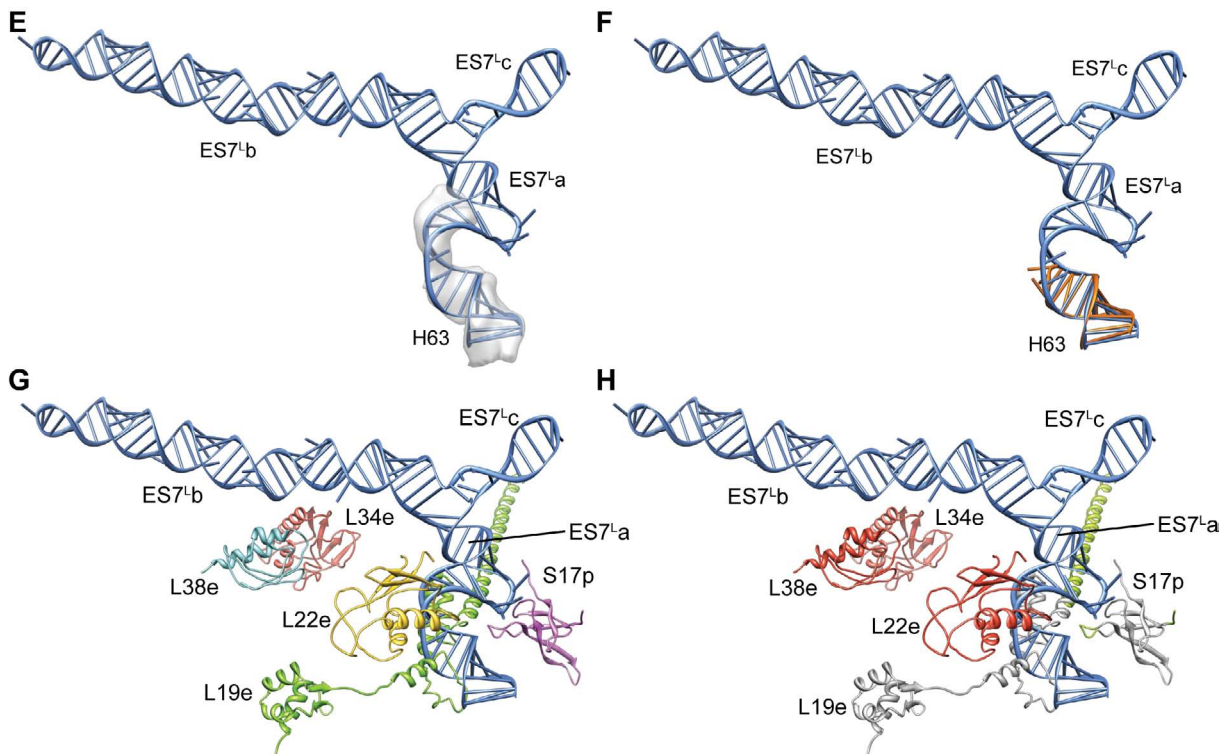


Figure 9.2.37.: **Expansion segment *ES27^L* of *T. aestivum*.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of *ES27^L* (blue) with r-proteins L34e (red), L22e (yellow), L38e (cyan), L19e (green) and S17p (purple). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

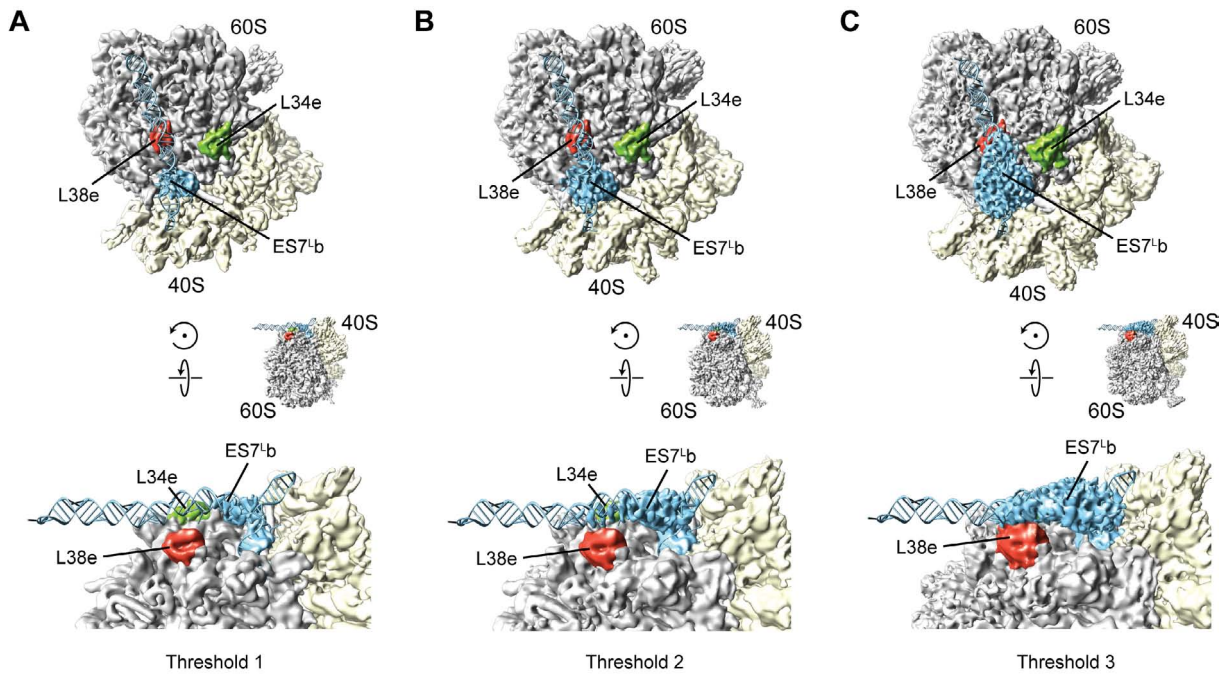


Figure 9.2.38.: **Flexibility of expansion segment $ES27^L$ of *T. aestivum*.** Visualization of density for (A-C) $ES27_{int}^L$ (blue) and r-proteins L38e (red) and L34e (green). In the *T. aestivum* 80S reconstruction, the density for $ES27^{Lb}$ is observed at lower contour levels. The interaction with r-protein L38e (red) with $ES27^{Lb}$ is evident (C). The figure was taken from [182].

9.2.2.21. $ES27^L$ of *S. cerevisiae*

In *S. cerevisiae* expansion segment $ES27^L$ is located on the opposite of the central protuberance on the 60S subunit and close to the 40S subunit (Figure 9.2.39 A). $ES27^L$ emerges from H63 and is part of domain IV (Figure 9.2.39 B). $ES27^L$ is 152 nts long and consists of three helices $ES27^L_a$, $ES27^L_b$ and $ES27^L_c$ arranged in a three-way junction (Figure 9.2.39 C, D). In *S. cerevisiae* $ES27^L$ is observed in two preferred orientations denoted as $ES27^L_{in}$ oriented towards the L1 stalk (Figure 9.2.41 A); and $ES27^L_{out}$ which is oriented towards the tunnel exit site (Figure 9.2.41 B).

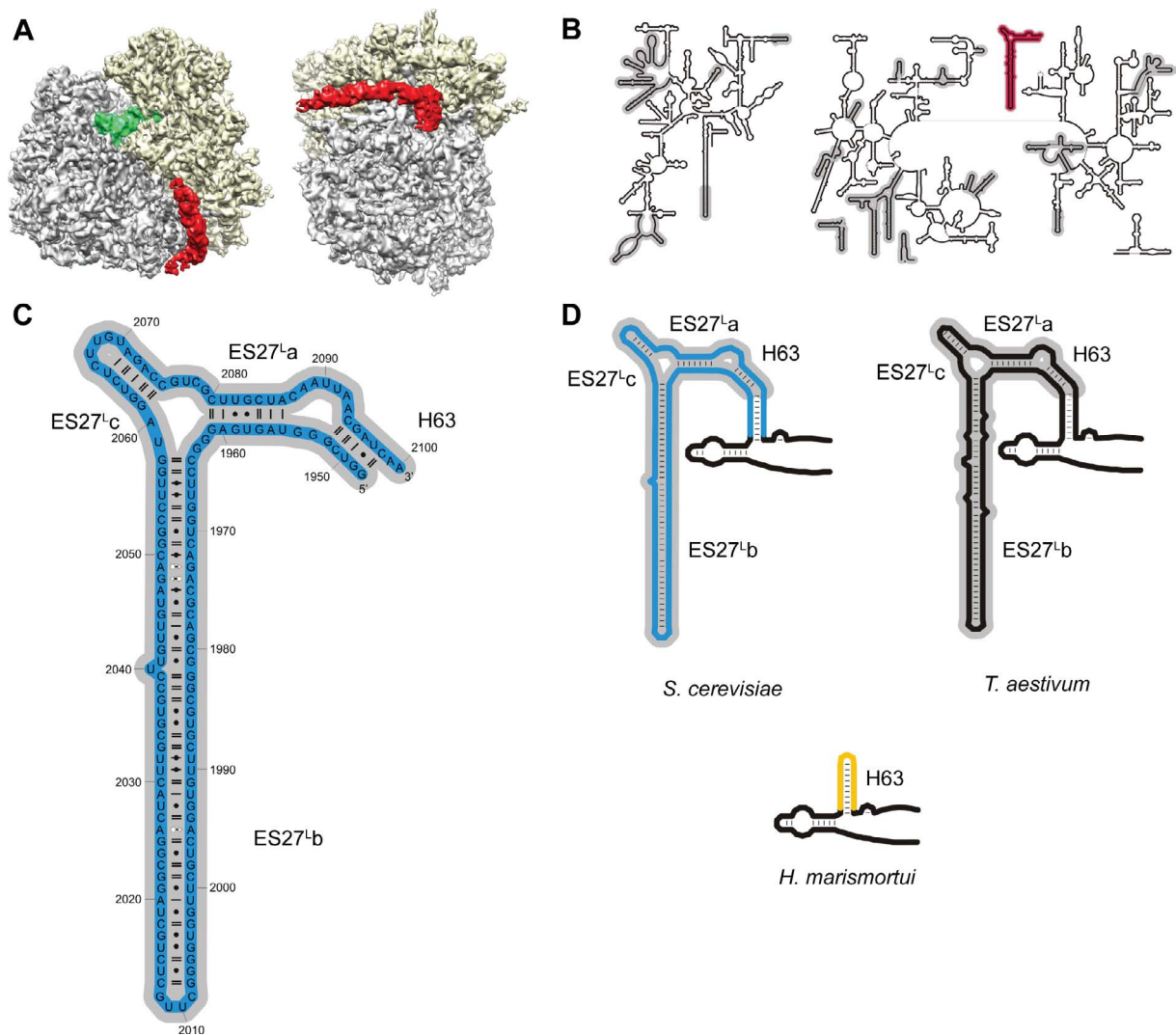


Figure 9.2.39.: **Expansion segment $ES27^L$ of *S. cerevisiae*.** (A) Location of $ES27^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES27^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *S. cerevisiae* model. (D) RNA secondary structure diagram. *S. cerevisiae* (blue), *T. aestivum* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively.

$ES27^L_a$ is connected to H63 via six single stranded nucleotides and stacks on H63. The bended helix $ES27^L_b$ folds perpendicular to $ES27^L_a$ and $ES27^L_c$, respectively. On the other hand $ES27^L_c$ folds perpendicular to $ES27^L_a$ and $ES27^L_b$, respectively (Figure 9.2.40 E, F). The three-way junction is of class A (according to [186]) and explains the the high flexibility of $ES27^L$.

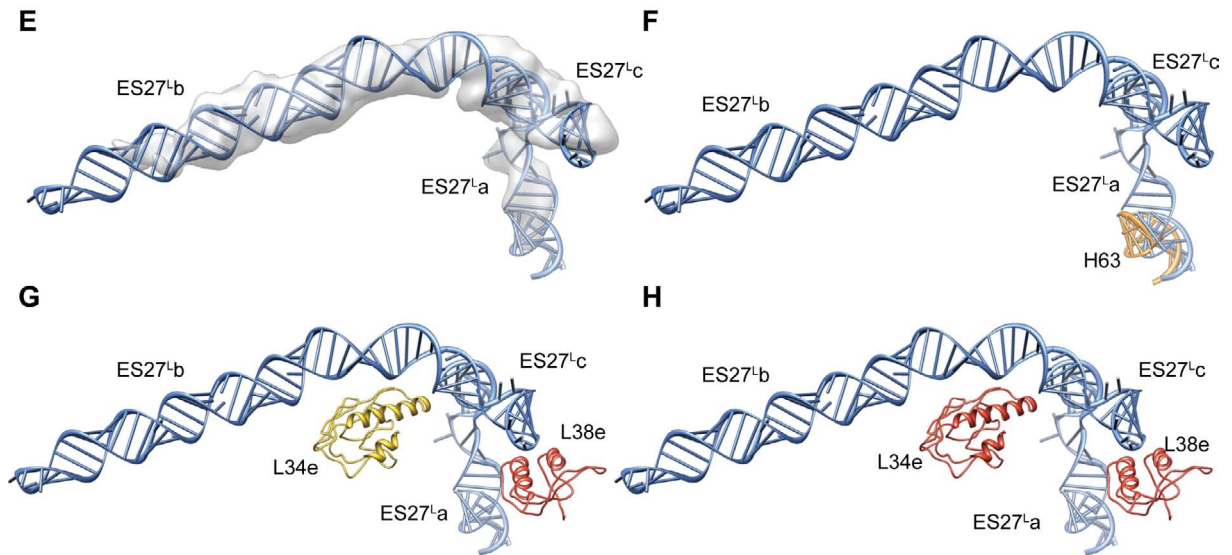


Figure 9.2.40.: **Expansion segment $ES27^L$ of *S. cerevisiae*.** (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES27^L$ (blue) with r-proteins L34e (red), L22e (yellow), L38e (cyan), L19e (green) and S17p (purple). (H) Same as in (G) eukaryote-specific proteins (red), but eukaryote-specific protein extensions (light green) of bacterial homologous (gray).

The conformation $ES27_{in}^L$ The long helix $ES27^Lb$ has an innerhelical part of non-Watson-Crick base pairs. In $ES27_{in}^L$ conformation the non-WC part of helix $ES27_{in}^Lb$ interacts via its backbone with the eukaryote-specific r-protein L34e (Figure 9.2.40 G, H). The short helix $ES27_{in}^Lc$ interacts with another r-protein specific to eukaryotes namely L38e (Figure 9.2.40 G, H).

The conformation $ES27_{out}^L$ In the $ES27_{out}^L$ conformation the backbone of helix $ES27_{out}^Lb$ interacts with the eukaryote-specific r-protein L38e. In this conformation helix $ES27_{out}^Lc$ is not stabilized by a r-protein.

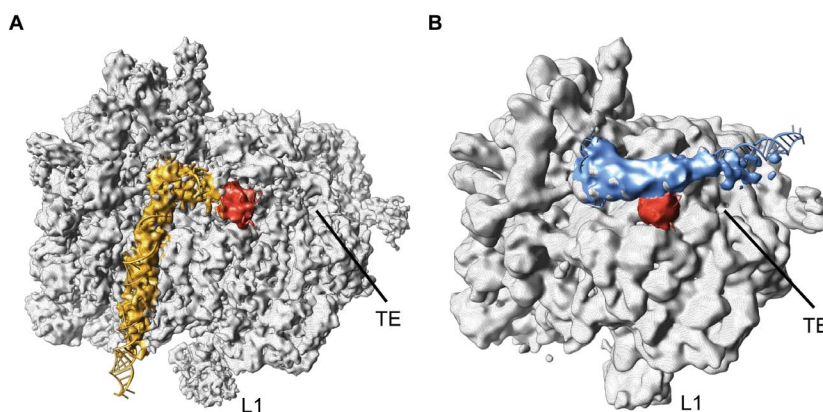


Figure 9.2.41.: **Expansion segment $ES27^L$ of *S. cerevisiae*.** (A) $ES27^L$ (gold) in the conformation oriented towards the L1 stalk denoted as $ES27_{in}^L$ and interacting with L38e (red). (B) $ES27^L$ (blue) in the conformation oriented towards the tunnel exit site of the ribosome denoted as $ES27_{out}^L$ and interacting with L38e (red). The figure was adapted from [182].

9.2.2.22. $ES31^L$ (H79)

Expansion segment $ES31^L$ is located on the side of the the 60S subunit and close to $ES4^L$, $ES19^L$, and $ES20^L$ (Figure 9.2.42 A). $ES31^L$ emerges from H79 and is part of domain V of the 25S rRNA (Figure 9.2.42 B-D). In both species *T. aestivum* and *S. cerevisiae* $ES31^L$ consists of three helices $ES31^La$, $ES31^Lb$ and $ES31^Lc$ forming an RNA three-way junction (Figure 9.2.42 C-F). The helices are clearly identified in the cryo-EM maps (Figure 9.2.42 E). The single stranded region between $ES31^Lb$ and $ES31^Lc$ (13 nt) could not be modeled, due to ambiguous density. The three-way junction is stabilized by the N-terminal extension of L7ae (Figure 9.2.42 G, H). $ES31^La$ interacts with L2p and the N-terminal extension of L23p. The eukaryote-specific r-protein L34e interacts with the minor groove and loop the of $ES31^Lc$ (Figure 9.2.42 G, H).

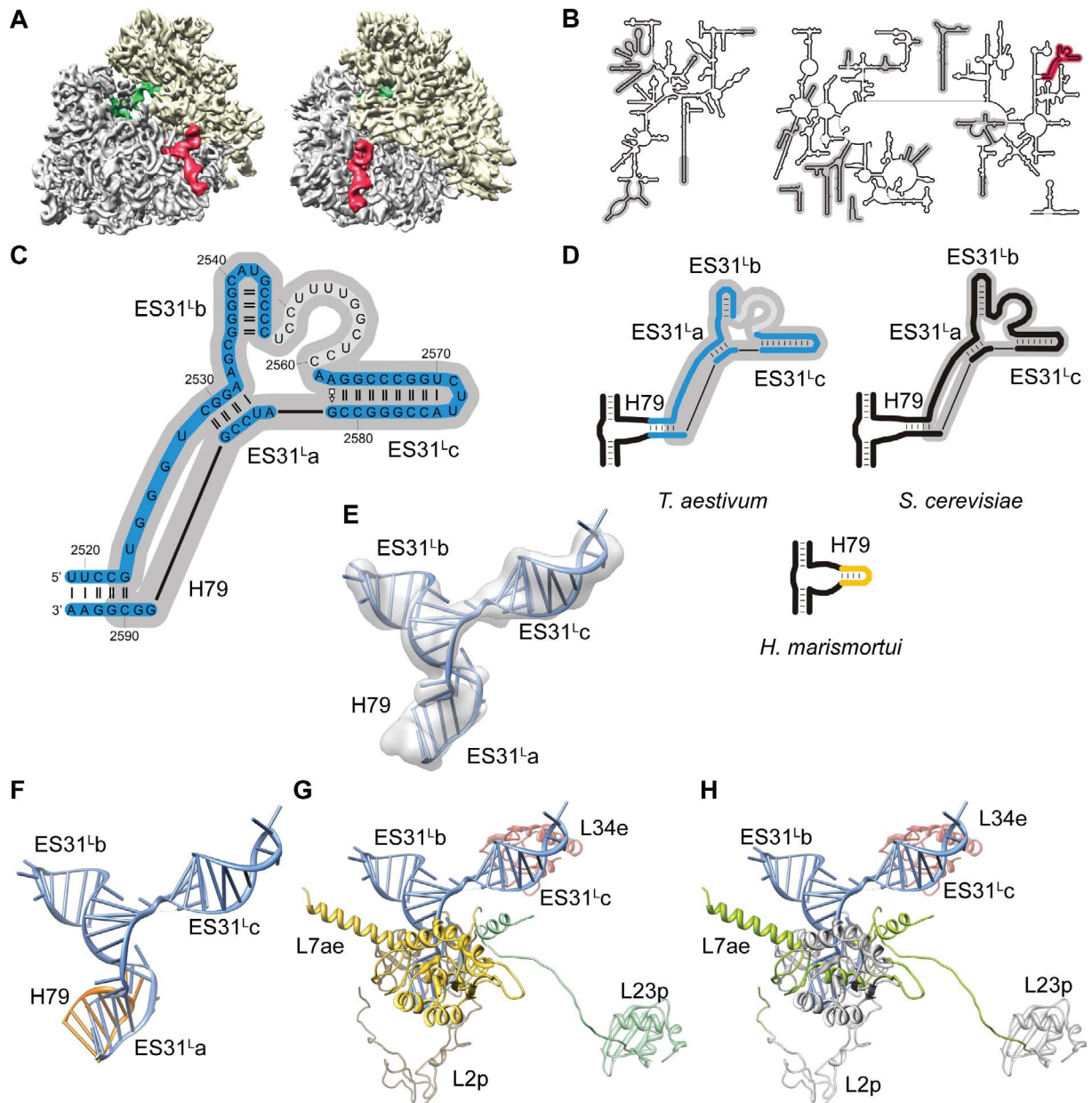


Figure 9.2.42.: **Expansion segment** $ES31^L$. Location of $ES31^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES31^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES31^L$ (blue) with r-proteins L34e (red), L7ae (yellow), L2p (pale purple) and L23p (mint green). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologs (gray).

9.2.2.23. ES39^L (H98)

Expansion segment *ES39^L* is located on the back of the large subunit (Figure 9.2.43 A). *ES39^L* emerges from H98 and is part of domain VI of the 25S RNA (Figure 9.2.43 B). *ES39^L* consists of four helices *ES39^La*, *ES39^Lb*, *ES39^Lc* and *ES39^Ld*. A single stranded region of 18 nt (*T. aestivum*) and 19 nt (*S. cerevisiae*), respectively, connect *ES39^Ld* and *ES39^La* (Figure 9.2.43 C, D). *ES39^La* is the counterpart of the bacterial H98 which is not present in *H. marismortui* (Figure 9.2.43 D, F). The innerhelical non-Watson-Crick base pairs and the single stranded part in *ES39^Lb* is mostly likely a kink-turn, but of unknown structure. *ES39^Ld* coaxial stacks on *ES39^Lc* (Figure 9.2.43 E, F). The single stranded region between *ES39^Ld* and *ES39^La* is modeled to be closely intertwined with r-proteins. Ribosomal proteins L22p and L3p interact with *ES39^La/H98*. The linker region between *ES39^La* and helix *ES39^Lb* are stabilized by L35ae. The C-terminal extension of L13p interacts with the innerhelical kink-turn-like fold of *ES39^Lb* (Figure 9.2.43 G, H). On the other side of this helix two eukaryote-specific r-proteins interact with the backbone and minor groove of *ES39^Lb*, namely L14e and L18a. The loop region of helix *ES39^Lb* is stabilized by L6p. The GNRA-tetraloop of *ES39^Lc* interacts with the α -helix of the eukaryote-specific r-protein L6e. The C-terminal helix of L3p interacts with the GNRA-tetraloop of *ES39^Ld* (Figure 9.2.43 G, H).

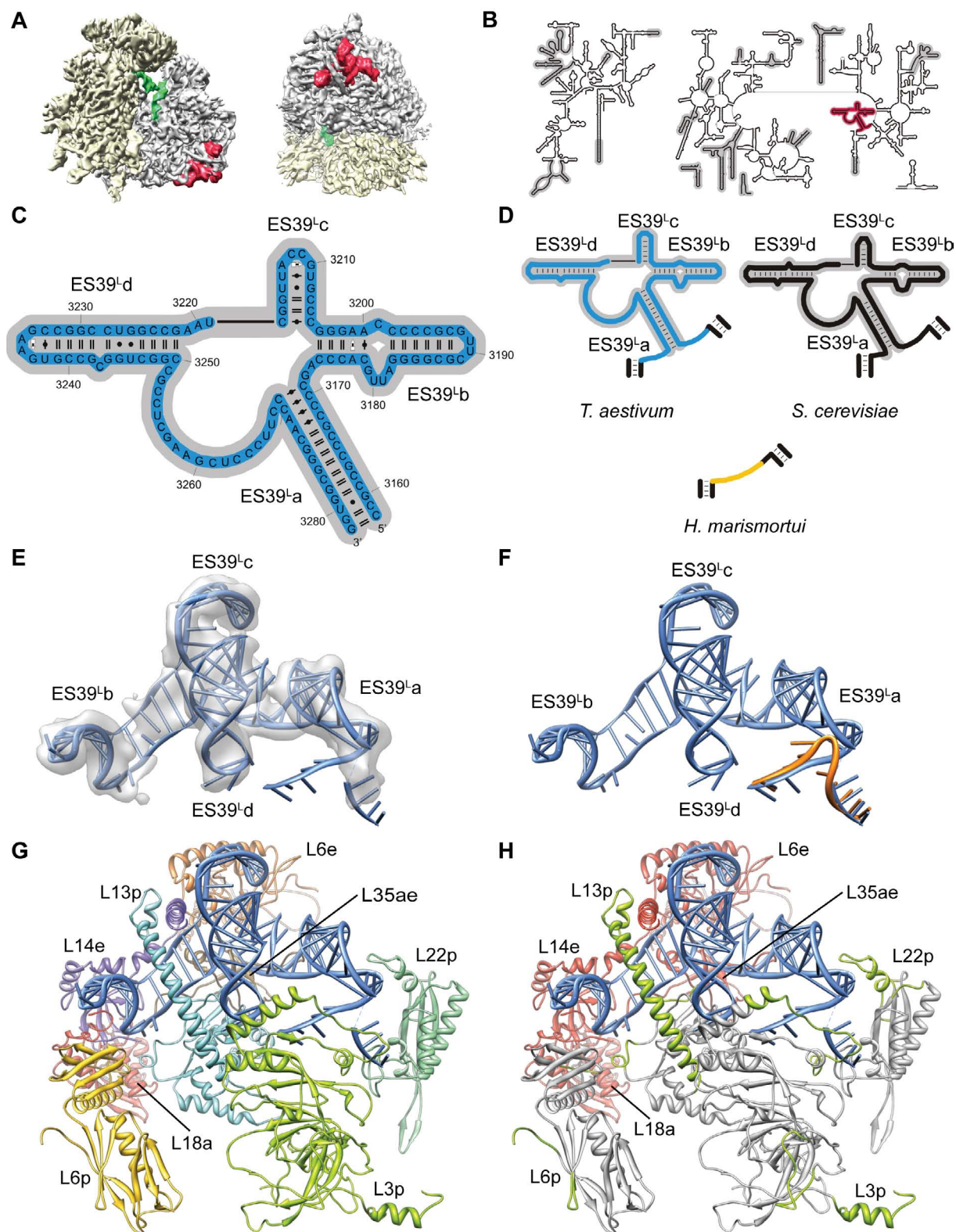


Figure 9.2.43.: **Expansion segment $ES39^L$.** Location of $ES39^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES39^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES39^L$ (blue) with r-proteins L18e (red), L6e (orange), L6p (yellow), L3p (green), L22p (mint green), L13p (cyan), L14e (purple) and L35e (brown). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

9.2.2.24. $ES41^L$ (H101)

Expansion segment $ES41^L$ is located on the side of the large subunit and close to the small subunit (Figure 9.2.44 A). $ES41^L$ is the helical extension of H101 and is part of domain VI of the 25S rRNA (Figure 9.2.44 B). $ES41^L$ adopts a helical fold and interacts via its innerhelical non-Watson-Crick base pairs with H63 (Figure 9.2.44 E, F). The C-terminal extension of L24e interacts with the minor groove and the backbone of $ES41^L$. The loop region of the eukaryote-specific r-protein L22e interacts with the minor groove of $ES41^L$ (Figure 9.2.44 G, H).

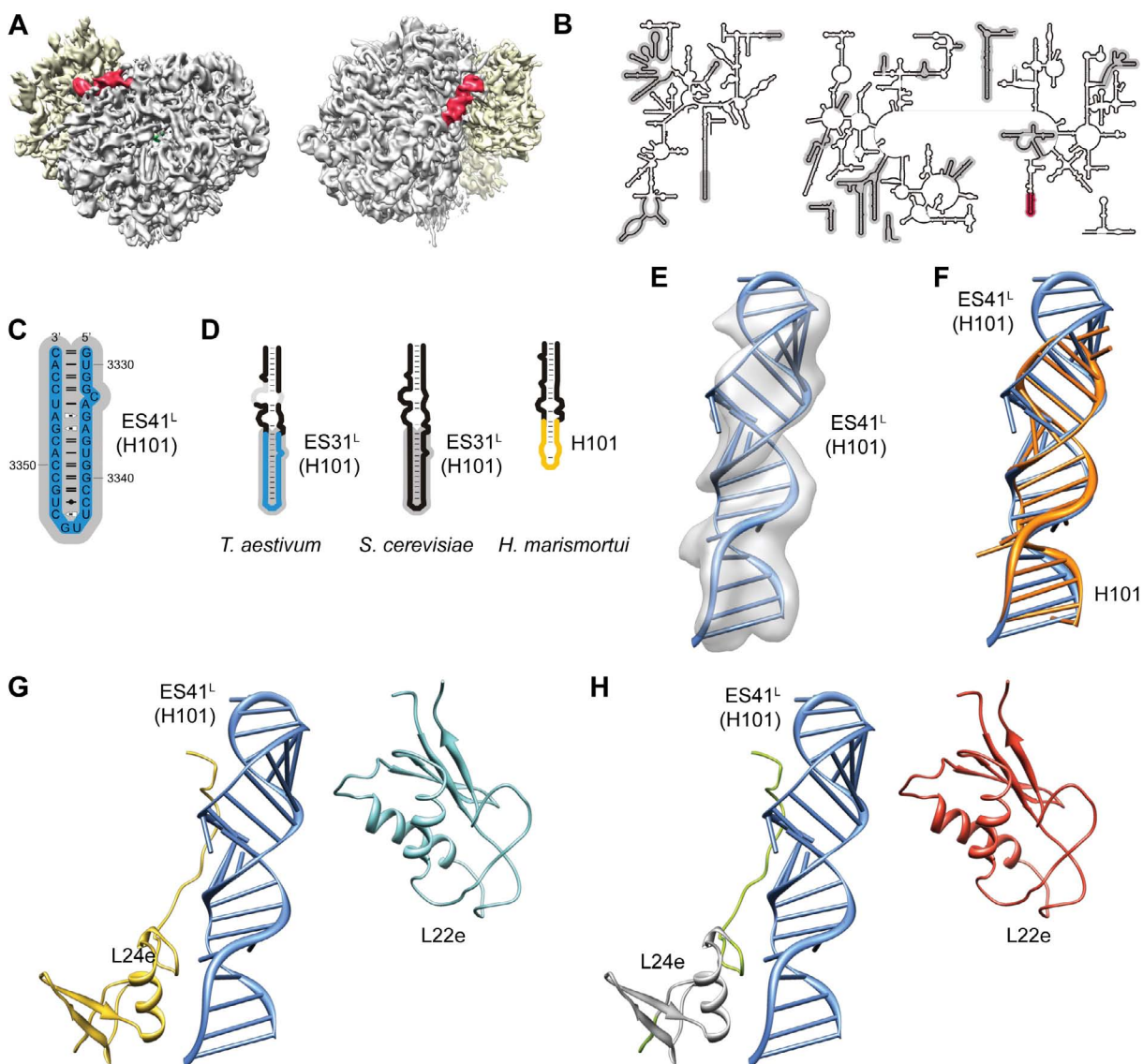


Figure 9.2.44.: **Expansion segment $ES41^L$.** Location of $ES41^L$ (red) on the ribosome. The small and large subunit are colored in yellow and gray, respectively. The P-tRNA is colored in green. (B) Location of $ES41^L$ (red) in the secondary structure diagram. (C) RNA secondary structure diagram of the *T. aestivum* model. (D) RNA secondary structure diagram. *T. aestivum* (blue), *S. cerevisiae* (black) and the corresponding RNA in *H. marismortui* (yellow), respectively. (E) RNA model (blue) in cryo-EM density. (F) RNA model (blue) on top of its archaeal counterpart (orange). (G and H) Interaction of $ES41^L$ (blue) with r-proteins L22e (cyan) and L24e (yellow). (H) Same as in (G) but eukaryote-specific r-proteins (red), eukaryote-specific r-protein extensions (light green) of bacterial homologous (gray).

9.3. Comparison between *S. cerevisiae* and *T. aestivum*

9.3.1. Comparison of the cryo-EM densities

The cryo-EM maps of *T. aestivum* at 5.5 Å and *S. cerevisiae* at 6.1 Å were used to build molecular models for a eukaryotic translating ribosome. The overall shape of the two cryo-EM densities is very similar. The densities corresponding to the 5S rRNA and 18S rRNA are almost identical in both species. The densities corresponding to the 28S rRNA show minor differences in *ES3^L*, H16-H18, *ES19^L* and *ES39^L* as well as major differences in the regions of *ES7^L* and *ES27^L*.

9.3.2. Comparison of expansion segments

The common structure of 4679 (1490/40S, 3071/60S and 118/5S) nts between *T. aestivum* and *S. cerevisiae* support a common evolutionary origin. Both species have the same amount of RNA expansion segments - five expansion segments (*ES3^S*, *ES6^S*, *ES7^S*, *ES9^S*, and *ES12^S* following the ES nomenclature of [23]) of the small subunit, as well as the 16 expansion segments (*ES3^L*, *ES4^L*, *ES5^L*, *ES7^L*, *ES9^L*, *ES10^L*, *ES12^L*, *ES15^L*, *ES19^L*, *ES20^L*, *ES24^L*, *ES26^L*, *ES27^L*, *ES31^L*, *ES39^L*, and *ES41^L*) (Figure 9.2.9). The rRNA of both species were compared on the level of secondary and tertiary structure. Interestingly, *T. aestivum* and *S. cerevisiae* share the same interaction of *ES3^S* and *ES6^S*. Significant differences between the corresponding expansion segments are only present in *ES7^L* and *ES27^L*.

9.3.2.1. Interaction of $ES3^S$ and $ES6^S$

The cryo-EM reconstructions for both *T. aestivum* and *S. cerevisiae* support a direct interaction of $ES3^S$ and $ES6^S$ (Figure 9.3.1 A). The interaction between these two expansion segments was previously predicted by Alkemar and Nygård [28]. A covariation analysis also suggests a direct interaction of $ES3^S$ and $ES6^S$, which is conserved in six different eukaryotic taxa including mammals [28].

In the 3-dimensional model the single stranded region between $ES3^{Sb}$ and $ES3^{Sc}$ interacts with the loop of $ES6^{Sd}$ (Figure 9.3.1 B; see 9.2.2.2 and 9.2.2.4). The helix is formed by 9 bp in both species. In *T. aestivum* consists of 9 Watson-Crick base pairs (Figure 9.3.1 C; Figure 9.2.3). In contrast the helix in *S. cerevisiae* consists of 5 Watson-Crick base pairs and 4 non-Watson-Crick base pairs (Figure 9.2.4). The helix is terminated by a non-Watson-Crick $A-G$ (tHS) base pair, which are often found at helix termini in ribosomal RNA [3, 7, 187].

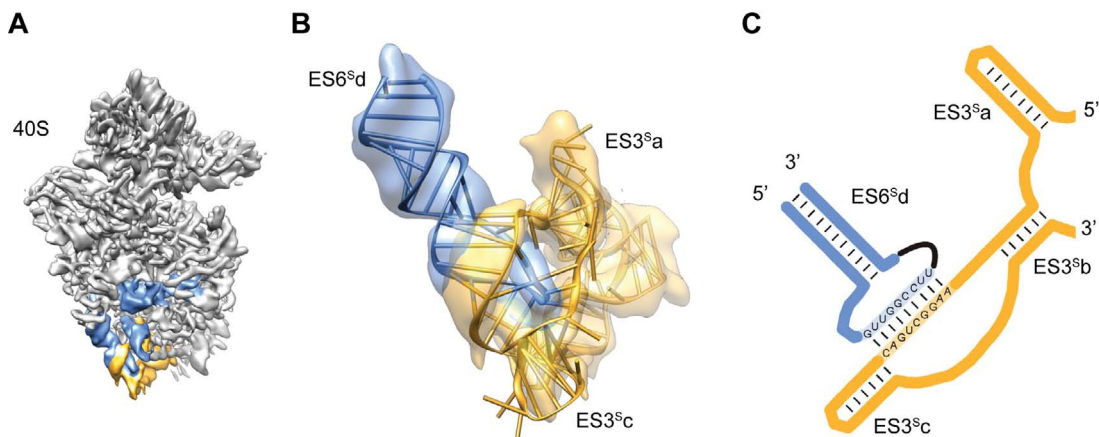


Figure 9.3.1.: **Interaction of $ES3^S$ and $ES6^S$.** (A) Cryo-EM density of *T. aestivum* 40S subunit with highlighted density for $ES3^S$ (yellow) and $ES6^S$ (blue). (B) Isolated density for $ES6^{Sd}$ (blue) and $ES3^{Sa}$ (yellow) transparent with molecular models. (C) Secondary structure prediction highlighting the interaction between the loop of $ES6^{Sd}$ (blue) and the bulge of $ES3^S$ (yellow). The figure was taken from [182].

9.3.2.2. Comparison between $ES7^L$ in *T. aestivum* and *S. cerevisiae*

The largest expansion segment on the large ribosomal subunit of *T. aestivum* and *S. cerevisiae* is $ES7^L$, which is located on the back of the 60S subunit (Figure 9.3.2 A and D). The overall shape of $ES7^L$ is similar between *T. aestivum* and *S. cerevisiae*, however, some clear differences are evident: In *S. cerevisiae* the density for $ES7^L_a$ is only visible at lower contour levels (Figure 9.2.27), suggesting that it is more flexible than in *T. aestivum* (Figure 9.3.2 B). The reason for this flexibility appears to be that $ES7^L_a$ in *T. aestivum* is stabilized by r-protein L28e (Figure 9.3.2 B), which is absent in the *S. cerevisiae* genome [37]. Further the bended fold of $ES7^L_a$ in *T. aestivum* adopts a kinked conformation by $\sim 85^\circ$ in *S. cerevisiae* (Figure 9.3.2 B, E). A second difference is the length of $ES7^L_b$. In *S. cerevisiae* $ES7^L_b$ is longer than in *T. aestivum* and fold around r-protein L18a. The third significant difference is the presence of an RNA three-way junction in *T. aestivum* formed by $ES7^L_{c-e}$, whereas this architecture is not present in *S. cerevisiae*, due to the absence of $ES7^L_{d,e}$ (Figure 9.3.2 and Figures 9.2.5, 9.2.7). The N-terminal extension of *T. aestivum* r-protein L6e, which is shorter in *S. cerevisiae*, appears to insert through the three-way junction formed by $ES7^L_{c-e}$ (Figure 9.3.2 C). The rod-like density for L6e suggests an α -helical fold of the N-terminus. Such kind of RNA-protein interaction has not been reported previously.

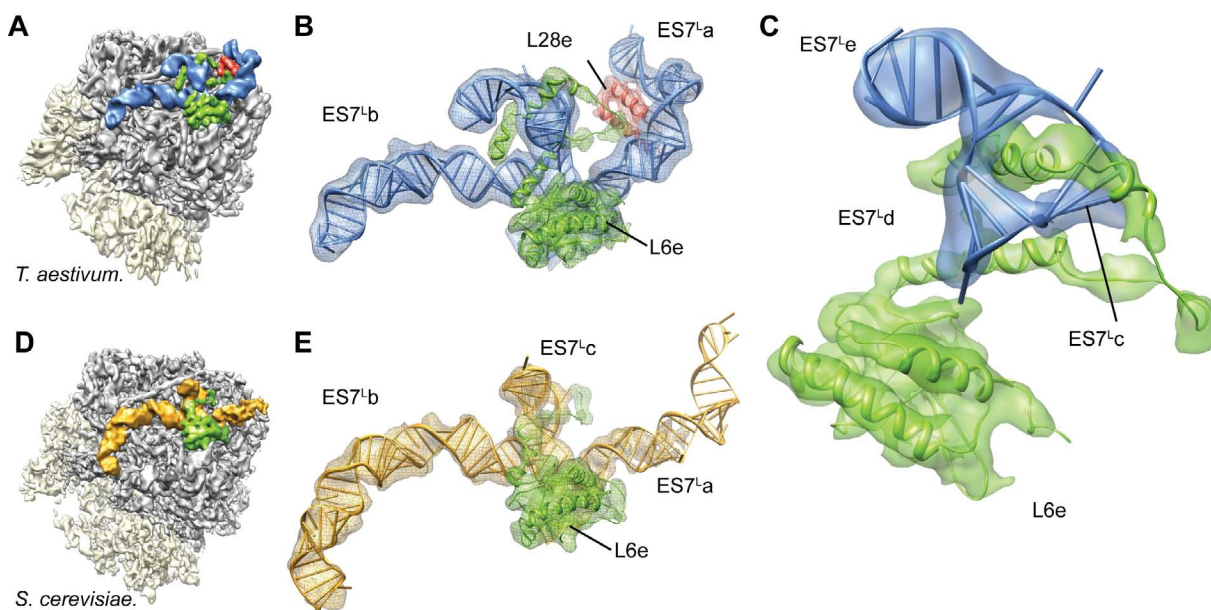


Figure 9.3.2.: **Molecular models for $ES7^L$ in *T. aestivum* and *S. cerevisiae*.** (A and D) Cryo-EM densities of 80S ribosomes with isolated densities for $ES7^L$ from *T. aestivum* (blue) and *S. cerevisiae* (gold). (B and C) Transparent densities with molecular models for $ES7^L$ from *T. aestivum* (blue) and *S. cerevisiae* (gold). Ribosomal-proteins L28e (red) stabilizes $ES7^L_a$ in *T. aestivum*, L6e (green). (E) Ribosomal protein L6e (green) appears to pass through the the RNA three-way junction formed by $ES7^L_{c-e}$. The figure was taken from [182].

9.3.2.3. Motion of $ES27^L$

$ES27^L$ is known to be highly dynamic. In *S. cerevisiae* $ES27^L$ is observed in two preferred orientations in the 80S ribosome [59] denoted as $ES27_{out}^L$ and $ES27_{in}^L$ (Figure 9.3.3 A; see also section 9.2.2.21). The $ES27_{in}^L$ conformation is better resolved than the $ES27_{out}^L$ conformation. In the *T. aestivum* $ES27^L$ is observed in an intermediate position denoted as $ES27_{int}^L$ (Figure 9.3.3 A; see also section 9.2.38).

Modeling the two conformation in *S. cerevisiae* reveals an interchange of $ES27^L$ between the $ES27_{out}^L$ (blue in Figure 9.3.3 A) and $ES27_{in}^L$ (gold in Figure 9.3.3 A) states, causing a $\sim 110^\circ$ rotation of $ES27^L a - c$ relative to H63 (Figure 9.3.3 B). The intermediate position $ES27_{int}^L$ observed in the *T. aestivum* EM density suggests a continuous motion of $ES27^L$ between the conformational states. The key players for the motion are the single stranded part at the 3'-end of $ES27^L$ connected to H63 and the three-way junction formed by $ES27^L a - c$ (Figure 9.3.3 B). Hereby, the three-way junction keeps the three helices together while the connecting single stranded part is flexible and rotates $ES27^L$ around the 5' connected single stranded part. Although $ES27^L$ is highly flexible the three preferential observed states ($ES27_{in}^L$, $ES27_{out}^L$ and $ES27_{int}^L$) are stabilized through the interaction with newly identified eukaryote-specific r-proteins L34e and L38e (described in section 9.2.37 and 9.2.40).

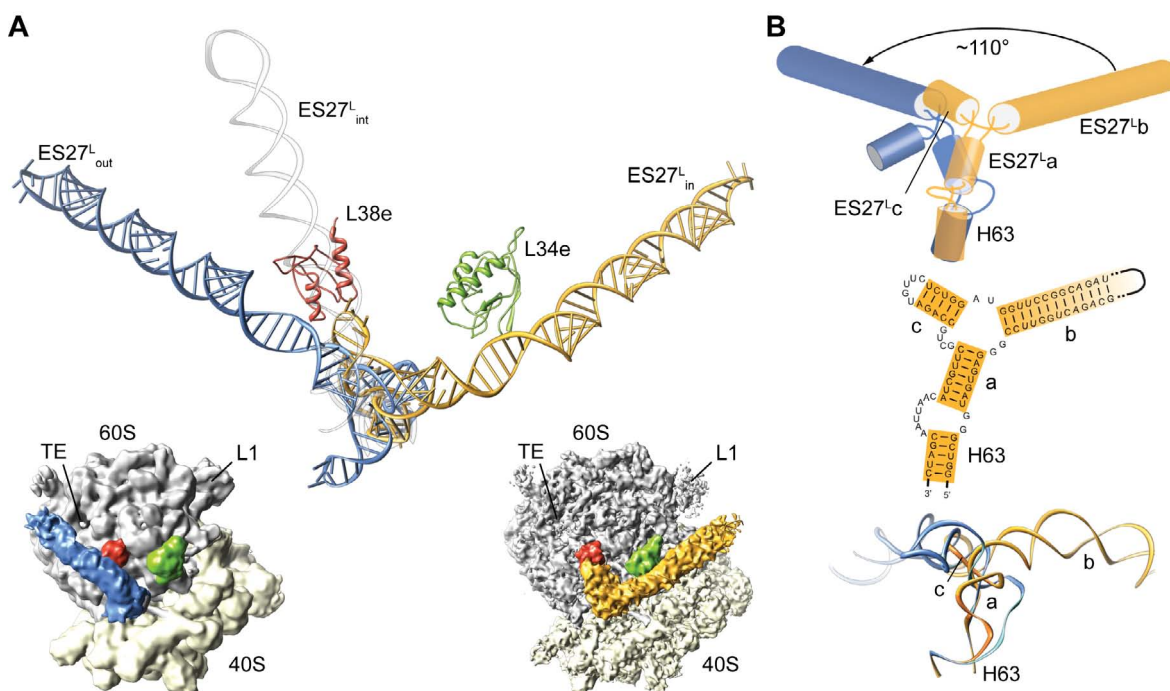


Figure 9.3.3.: **Motion of $ES27^L$.** (A) Molecular models for $ES27_{in}^L$ (gold), $ES27_{out}^L$ (blue) as observed in *S. cerevisiae* 80S ribosomes (Thumbnail insets) and the intermediate position $ES27_{int}^L$ (*T. aestivum* transparent gray) observed in *T. aestivum* 80S ribosomes. In *S. cerevisiae*, r-proteins L38e (red) and L34e (green) interact with the different conformational states. (B) Schematic (Top) and molecular model (Bottom) for $ES27^L$ indicating the interchange between $ES27_{in}^L$ (gold) and $ES27_{out}^L$ (blue) involves a rotation of $\sim 110^\circ$ rotation of $ES27^L a - c$ relative to H63. The RNA secondary structure for the junction between $ES27^L$ and H63 and the three-way junction of $ES27^L a - c$ in *S. cerevisiae* (Middle). The figure was taken from [182].

10. Molecular interpretation of functional sites of the eukaryotic 80S ribosome

The molecular models of the *T. aestivum* and *S. cerevisiae* 80S ribosomes were used to analyze the functional sites of the eukaryotic 80S ribosome.

10.1. The decoding site and the PTC

The active sites of the ribosome, namely the PTC on the large subunit and the decoding site on the small subunit, are largely composed of rRNA. However, they are not completely devoid of r-proteins. In principle, the eukaryotic rRNA is highly conserved, while some r-proteins are different.

10.1.1. The decoding site

The decoding site is highly conserved. Nucleotides G530, A1492 and A1493 are invariant and do not change between prokaryotes and eukaryotes. Compared to the bacterial 30S subunit, the eukaryotic 40S subunit contains two additional r-proteins, S25e and S30e (Figure 10.1.1 A, B). Both r-proteins contain extensions, which reach into the decoding and tRNA binding sites. In addition the C-terminus of r-protein S4p is relocated in eukaryotes, due to the corresponding rearrangement of h16/h17, and reaches from the globular domain on the solvent side right into the decoding site of the small subunit (Figure 10.1.1 A).

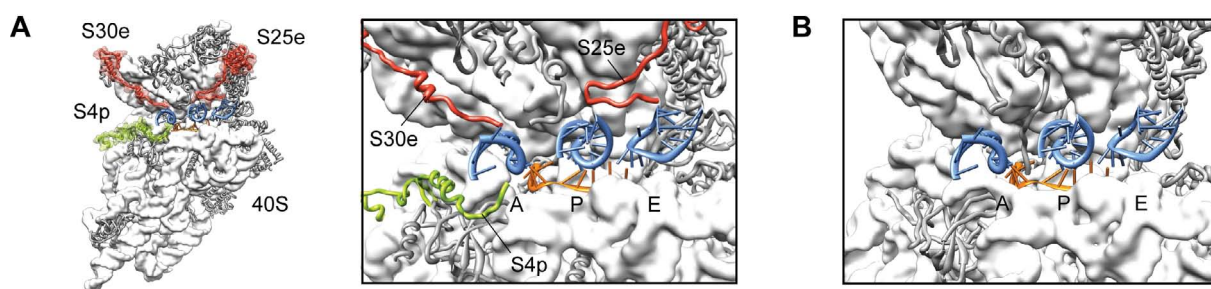


Figure 10.1.1.: **The decoding site in eukaryotes.** (A) The 40S subunit with newly modeled eukaryote-specific r-proteins S30e and S25e (red) and eukaryote-specific extension of S4p (green) highlighted in the thumbnail (left) and zoom (right). (B) Comparative view of the bacterial 30S subunit decoding site [14]. (A and B) The anticodon-stem loops of A-, P-, and E-site are colored in blue, the mRNA in orange. The rRNA is shown as surface representation and colored in gray. Bacterial r-proteins are represented in ribbons and colored in gray. The figure was taken from [185].

10.1.2. The PTC

The central loop of domain V in the 28S rRNA, which forms the PTC, is highly conserved. The length of all the single stranded RNA parts is identical. The A- and P-loop are also identical. There are only some isosteric base pair substitutions at residues that do not participate at peptide bond formation. The first substitution occurs in the second single stranded part of the central loop (5' direction). Compared to *H. marismortui* and *S. cerevisiae*, the *T. aestivum* 28S rRNA has a U at position 2821 instead of an adenine. At position 2824 (*T. aestivum*)/2822 (*S. cerevisiae*) both eukaryotes have a U instead an adenine in *bacteria*. In the third single stranded part of the central loop the two eukaryotes have two base substitutions: The bacterial adenine is substituted by a U at position 2868 (*T. aestivum*) and C2866 (*S. cerevisiae*), respectively. The C2534 (*H. marismortui*) is substituted by U2870 (*T. aestivum*) and U2868 (*S. cerevisiae*), respectively. The last two base substitutions occur in the fourth single stranded region of the central loop. Here G2642 (*H. marismortui*) is substituted by A2978 (*T. aestivum*) and U2976 (*S. cerevisiae*), respectively. Moreover, C2644 (*H. marismortui*) is substituted by U2980 (*T. aestivum*) and U2978 (*S. cerevisiae*), respectively. The involved RNA regions are part of the common core structure which was modeled by homology and based on the structure-based sequence alignment.

At the eukaryotic PTC, a direct interaction between the loop of r-protein L10e and the CCA-end of the P-site tRNA is observed (Figure 10.1.2 A). Based on the model, the loop of L10e is now the r-protein region that comes closest (~ 16 Å) to the site of peptide-bond formation. This loop is disordered and not modeled in the crystal structures of the *archaeal* 50S subunit [7] (Figure 10.1.2 B).

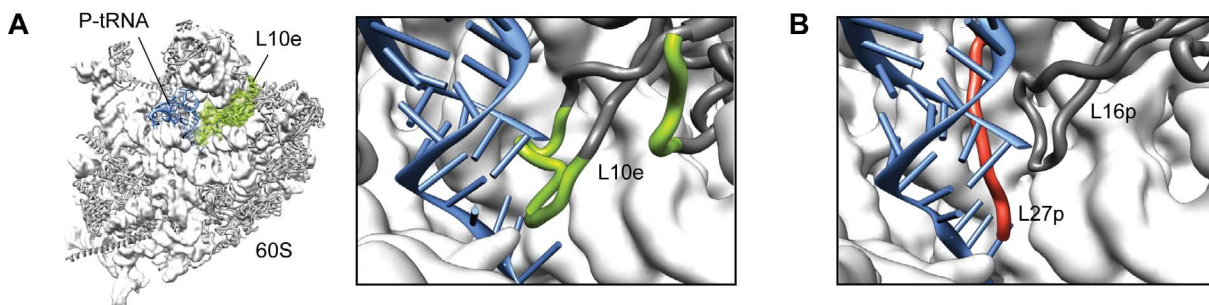


Figure 10.1.2.: **The eukaryotic Peptidyl-Transferase-Center.** (A) The 60S subunit with the eukaryote-specific extension of r-protein L10e (green) highlighted in the thumbnail (left) and zoom (right). (B) Comparative view of the *archaeal* 50S subunit with the bacterial-specific L27p colored in red and L16p in dark gray [14]. (A and B) The acceptor-stem of the P-tRNA is shown in blue. The rRNA is shown as surface representation and colored in gray. Bacterial r-proteins are represented in ribbons and colored in gray. The figure was taken from [185].

10.2. Model of the eukaryotic ribosomal tunnel

The molecular model of the *T. aestivum* 80S ribosome was used to analyze the architecture and dimensions of the eukaryotic ribosomal tunnel. In general, the composition and dimensions of the eukaryotic tunnel are very similar compared to those in *bacteria*. The eukaryotic ribosomal tunnel is $\sim 80 - 100$ Å long, about $10 - 20$ Å in diameter and stretches from the PTC to the tunnel exit site (TE). The narrowest part of the tunnel is the constriction, which is formed by r-proteins L4 (in *bacteria* L4) and L17 (in *bacteria* L22), approaching the tunnel from opposite sides.

Composition The tunnel is predominately composed of rRNA. The involved helices and loops are the same rRNA helices (H2, H4, H6, H7, H23, H24, H25.1, H26, H32, H33, H35, H35a, H39, H47, H50, H59, H59.1, H61, H64, H65, H72, H73, H74, H80, H89, H90, H92 and H93) and RNA loops as in *bacteria*. These helices and loops are structurally conserved except for H7 and H59. In eukaryotes H7 is a variable region (according to [23]) and folds in an overall similar kink-turn shape compared to its *bacterial* counterpart. Helix 59 is a eukaryotic expansion segment, termed *ES24^L* (see section 9.2.2.19). The tunnel surface is largely hydrophilic and solvent accessible. Although, the major part of the tunnel is composed of rRNA significant contributions are also made by the non-globular regions of r-proteins L4, L17 and the eukaryote-specific r-protein L39e.

10.2.1. Interaction of the ribosomal tunnel with nascent chains

The ribosomal tunnel was believed to be a passive conduit for all nascent chains [7]. However, it has been suggested that at least for some nascent chains (NC) the tunnel has a more active role and interacts with the amino acids of the nascent polypeptide. For example, the tunnel might play a role in folding of α -helical secondary structure (see section 1.2.3). In order to directly visualize an α -helix in the ribosomal tunnel, two cryo-EM reconstructions of ribosomes carrying a nascent chain with high α -helical propensity were analyzed [188].

Samples The two samples were based on *T. aestivum* 80S RNCs with a dipeptidylaminopeptidase B (DPAP-B) nascent chain. For the first construct, called the 80S-helix1 RNC construct, the NC was modified by substituting parts of the DPAP-B by the helix forming sequence positioned at residues 72 – 96. For the second construct, called 80S-helix2 RNC construct, the helix forming sequence was located at residues 83 – 108. Helix1 was expected to form in the lower region of the tunnel, near the exit, whereas helix2 would theoretically form in the constriction region. The samples were prepared by Dr. Shashi Bhushan, the protein models were prepared by J.-P. Armache and data processing was done by Marco Gartmann.

Nascent chain-ribosome interactions within the ribosomal tunnel The two cryo-EM reconstructions for both the 80S-helix1 construct and the 80S-helix2 construct, as well as the 80S-RNC bearing a DPAP-B NC, were analyzed at resolutions between 7.1-7.3 Å [188]. The most striking observation is an additional density within the ribosomal tunnel which is assigned to nascent polypeptide chain (Figure 10.2.2 A-F). The model of the *T. aestivum* 80S ribosome was then

used to dissect the molecular interactions between the nascent chain and the ribosomal components in both 80S-helix RNC reconstructions. The densities for P-tRNA and the associated nascent chains were isolated for the 80S-helix1 RNC, the 80S-helix2 RNC and the 80S-DPAP RNC (Figure 10.2.1 A-C). The molecular models for the P-tRNA and the NC were fitted into the isolated densities of 80S-helix1 RNC (Figure 10.2.1 D). The CCA-end of the tRNA was used to locate the C-terminus of the NC. For the 80S-helix1 RNC, the strong density for the N-terminal region of the NC is consistent with assumed α -helix conformation in this region and coincides perfectly with the expected location in the tunnel (Figure 10.2.1 D, E).

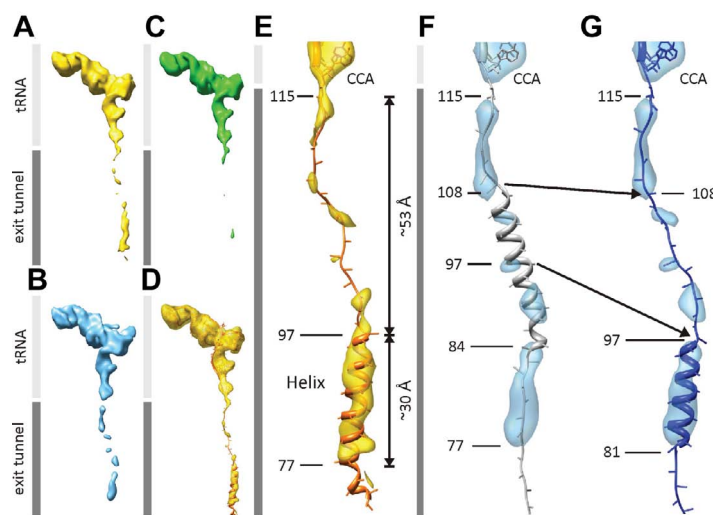


Figure 10.2.1.: **Comparison of the tRNA-nascent chains from the RNCs.** Isolated density for P-tRNA and NC from (A) 80S-helix1 RNC, (B) 80S-helix2 RNC, and (C) 80S-DPAP RNC. (D) Transparent density of (A) with molecular model for tRNA and NC. (E) Zoom of (D) with residue numbering, as well as distances between C α if residues 97-115 and 77-97. (F, G) Zoom of (B) with alternative models for the helix2 NC. Arrows indicate the corresponding region (residues 97-108) that is modeled as helical (F) or extended (G). The figure was taken from [188].

Different from an expected strong density in the middle of the tunnel, a strong density for the 80S-helix2 RNC is only observed in the lower region of the tunnel which is similar to the 80S-helix1 RNC (Figure 10.2.1 F). Therefore, a model of a NC with a distal portion of the remaining α -helical stretch (residues 81-97) was placed (Figure 10.2.1 G). For the 80S-helix1 RNC, three connections in the upper tunnel region are observed with strong density between the NC and the ribosomal tunnel wall (Figure 10.2.3 A). Adjacent to the PTC, the NC seems to contact A2062 (*E. coli* numbering throughout) of the 28S rRNA. Adenine A2062 was fitted into the density, suggesting a stabilization of the NC by the distinct conformation of A2062 (Figure 10.2.3 A, C). This contact is not observed in the 80S-helix2 RNC (Figure 10.2.3 E).

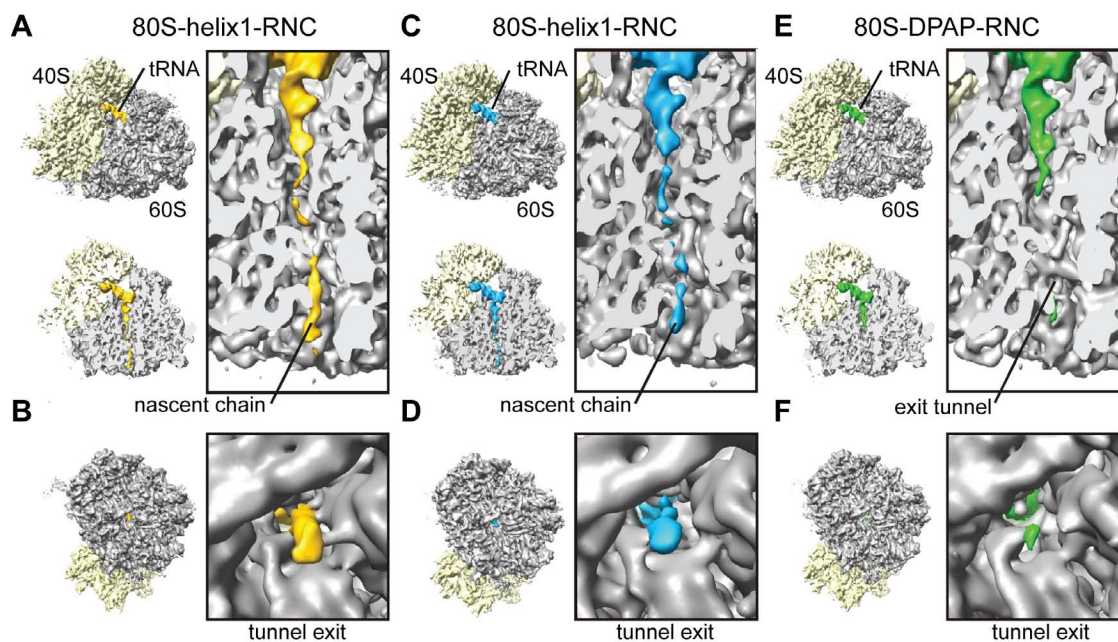


Figure 10.2.2.: **Cryo-EM reconstructions of RNCs.** (A) Cryo-EM reconstruction of the 80S-helix1-RNC. (B) Cryo-EM reconstruction of the 80S-helix2-RNC. (D) Cryo-EM reconstruction of the 80S-DPAP-RNC. For each reconstruction the thumbnail at the top left indicates the overall orientation of the ribosome. The small subunit is colored in yellow, the large subunit in gray and the P-site tRNAs and nascent chains in gold, blue and green, respectively. (A, C, E) View on the cross sections of the ribosomal tunnel. (B, D, F) View on the tunnel exit site. Right-Hand panels show a zoom of the respective cross sections and tunnel exit views. The figure was taken from [188].

Deeper in the tunnel, the NC contacts nucleotide A751, which is observed in both 80S-helix1 and 80S-helix2 density maps (Figure 10.2.3 A, E). The extension of r-protein L4 contact the 80S-helix1 RNC nascent chain at two positions, one near the constriction (Figure 10.2.3 A, C) and second close to the end of the α -helix (Figure 10.2.3 B, D). This first L4 contact at the constriction is significantly weaker in the 80S-helix2 RNC density (Figure 10.2.3 E). The second contact close to the α -helix is absent in the 80S-helix2 RNC (Figure 10.2.3 F). A negligible density between the two L4 contacts is observed in both nascent chains, indicating that they do not adopt a single distinct conformation in this region (Figure 10.2.3 A, E). An additional contact is observed between the NC and the tip of the β -hairpin of L17 (Figure 10.2.3 A, C, E).

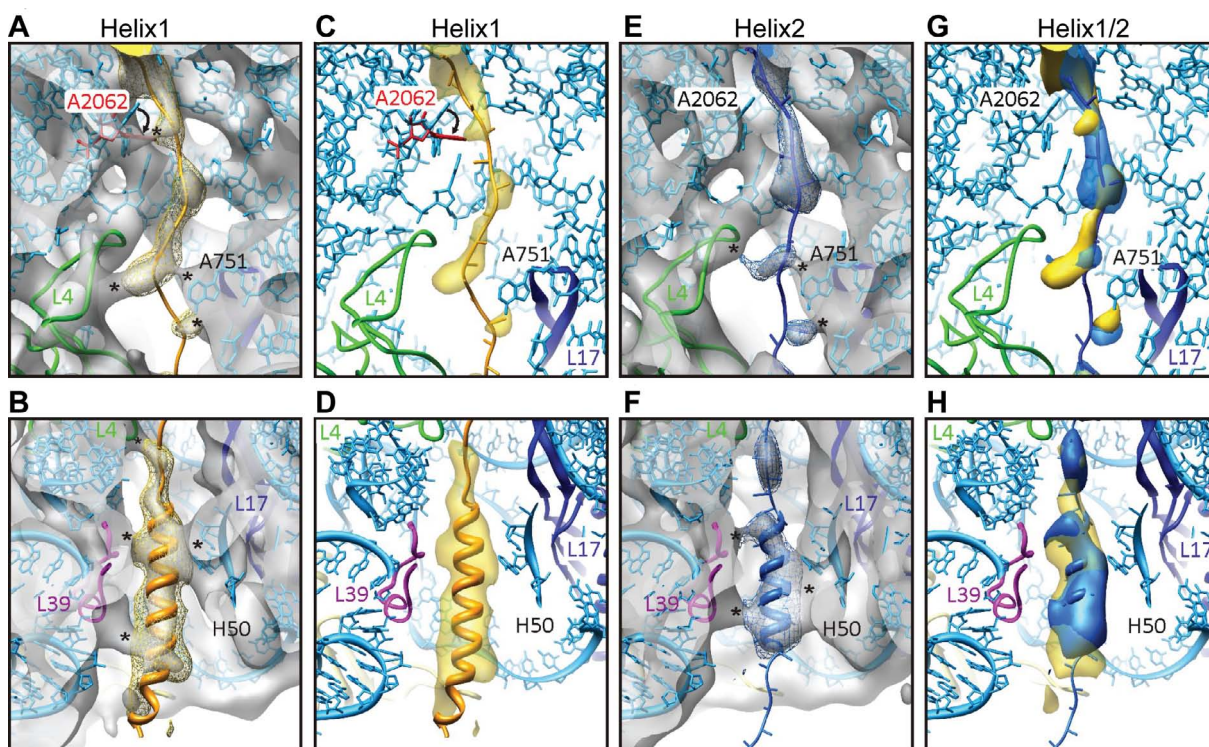


Figure 10.2.3.: **Interactions of the helix-nascent chain with the components of the ribosomal tunnel.**

Cross sections of the upper (top panels) and lower (bottom panels) part of the tunnel regions. The contacts between the NC and the tunnel are indicated with asterisks (*). The cryo-EM maps of 80S-helix1-RNC and 80S-helix2-RNC are shown in gray with isolated densities for the NCs (gold and blue, respectively). The nucleotides of the 28S rRNA are shown as blue sticks with ribbon backbone. The r-proteins L4 is colored in green, L17 in blue and L39 in magenta and shown as ribbons. In (A) and (C) the nucleotide A2062 (*E. coli* numbering) is shown in two conformations, taken from PDB 1VQN (red, [189]) and PDB 1S72 (blue, [7]). (A-D) 80S-helix1-RNC with nascent chain in gold. (E, F) 80S-helix2-RNC with nascent chain in blue. (G, H) Comparison of 80S-helix1-RNC and 80S-helix2-RNC. The figure was taken from [188].

α -helix formation in the ribosomal tunnel For the 80S-helix1 RNC density map, the strongest density is observed in the N-terminal region of the nascent chain near the tunnel exit (Figure 10.2.3 B, F). This coincides well with the helix-forming sequence for the 80S-helix1 RNC. Here, the density corresponding to the α -helix is sandwiched between the ribosomal protein L39e and the internal loop region of helix H50 of the 28S rRNA (Figure 10.2.3 B, D, F). Two strong contacts between L39e and the α -helix are observed, one close contact and another more distal to the N-terminus (Figure 10.2.3 B, D, F). Surprisingly, for the 80S-helix2 RNC no α -helix formation is observed in the constriction. Strong density is observed near the tunnel exit in the same position as helix1. Only the contacts to H50 appear to be different between the two structures (Figure 10.2.3 B, D, F, H). A detailed interaction table of the ribosomal tunnel and the nascent chain is summarized in table 10.1.

	Ribosomal contact	80S-helix1 RNC	80S-helix2 RNC
Upper tunnel	U2585	100	100
	G2505/U2506	99	99
	A2062	96	95/94
	A2058/9	(94)	94
Constriction	A751	(92)	92
	L4: 71/75	91	91
	L17: 136	(88)	88
	A1614	88	88
	L4: 88-89	85	86
Lower tunnel	L39e: 35	79/75	80
	H50: 1320-1321	77/73	77/75/70
	L39e: 23/26	68/64	71

Table 10.1.: Interactions between the ribosomal tunnel and the nascent chain.

10.3. The eukaryotic ribosomal tunnel exit site

The molecular models of the *S. cerevisiae* and *T. aestivum* 80S ribosomes were used to analyze the architecture of the eukaryotic ribosomal tunnel exit site. The tunnel exit site consists of rRNA helices H6, H7, H24, H47, H50, H53 and H59 and the r-proteins L4e (rpL4), L19e (rpL19), L22p (rpL17), L23p (rpL25), L24p (rpL26), L29p (rpL35), L31e (rpL31) and L39e (rpL39) (Figure 10.3.1 A, B). The overall architecture of the tunnel exit site is very similar compared to *bacteria*, however, significant differences in H7, H24 and H59 are present. In both, *S. cerevisiae* and *T. aestivum*, H7 adopts a newly and unidentified kink-turn fold comparable to *E. coli*. As in *bacteria* H7 makes a pseudoknot interaction with the adjacent H6. The interaction between the loops of H6 and H7 is stabilized by the eukaryote-specific r-protein L39e which replaces the loop of the bacterial counterpart L23. Another difference is observed for the H24 loop region. In *bacteria* the loop of H24 is a pentaloop. In contrast, the loop of H24 in fungal and plant ribosomes is a GNRA-tetraloop, which is clearly visible in the cryo-EM densities. Although the loop motif and structure of the loop changes, the stacking interaction with the loop of H47 is conserved. The expansion segment $ES24^L$ is the extension of H59. The loop of $ES24^L$ is a tetraloop of unknown structure in *S. cerevisiae* and a pentaloop in *T. aestivum*, respectively.

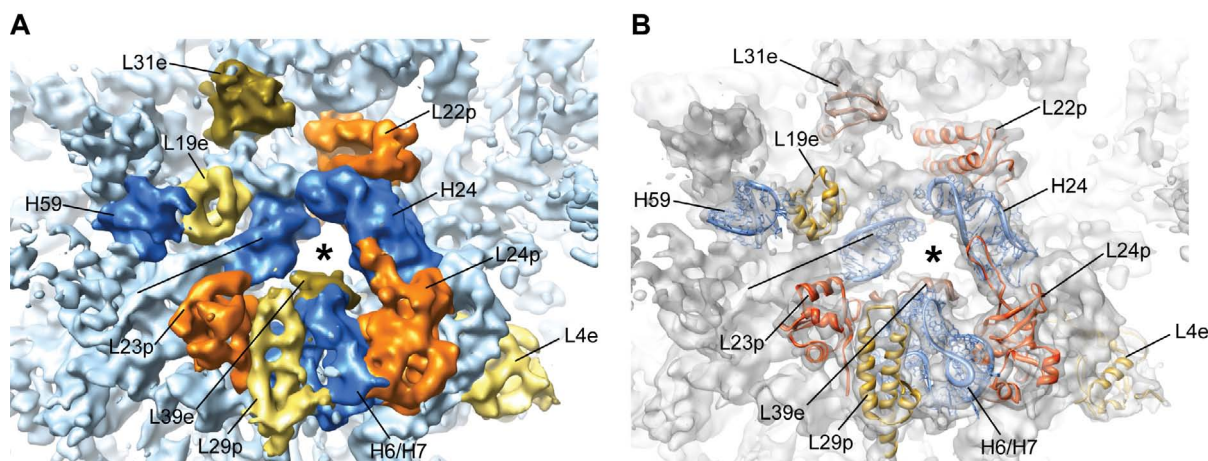


Figure 10.3.1.: **Visualization of the eukaryotic ribosomal exit tunnel.** (A) Bottom view of the *S. cerevisiae* cryo-EM map at 6.1 Å with highlighted rRNA (blue) and r-proteins (yellow, orange and brown, respectively). The asterisk indicates the tunnel exit. (B) Same as in (A), showing the molecular models for rRNA helices (blue) H6, H7, H24, H50 and H59 as well as r-proteins L4e, L29p and L19e (in yellow), L22p, L23p and L24p (in orange) and L31e and L39e (in brown). The r-proteins are named by their family name. The figures were taken from [141].

10.3.1. Interaction of the protein-conducting channel and the tunnel exit site

One of the main interaction partners with the ribosomal exit site is the the protein-conducting channel (PCC) studied in this work [141]. The PCC of the canonical secretory pathway is formed in all cells by the Sec61/SecY complex. The PCC enables post- and co-translational translocation of secretory proteins into the membrane of the endoplasmic reticulum in eukaryotes and plasma membrane of *bacteria* [190, 191].

On the basis of the molecular model it was possible to characterize the interactions between eukaryotic actively translating ribosomes and the PCC [141].

Sample 80S ribosomes of *T. aestivum* and *S. cerevisiae* were programmed with a truncated mRNA coding for the first 120 amino acids of DPAP-B (DP120) carrying a signal sequence for co-translational protein translocation. These RNCs were *in vitro* reconstituted with purified Ssh1p or mammalian Sec61 (mSec61) complexes embedded in a detergent micelle. Cryo-EM structures of the *T. aestivum* DP120-RNC-mSec61 complex and the *S. cerevisiae* DP120-RNC-Ssh1p complex were obtained at 6.5 Å and 6.1 Å, respectively. At this resolution, the extended NC could be visualized from the PTC to the tunnel exit site for the first time. Moreover, for the mammalian Sec61 complex most of the transmembrane helices could be resolved and assigned. On this basis, a single Sec61 heterotrimer surrounded by a mixed detergent/lipid micelle was identified. This could also be shown for the Ssh1p and allowed the interpretation of the ribosome-PCC interaction on a molecular level. The *S. cerevisiae* RNC-Ssh1 complex was prepared by Dr. Thomas Becker (RNCs) and Dr. Elisabeth Menden (Ssh1p) and the *T. aestivum* RNC-mSec61 complex was prepared by Dr. Shashi Bhushan (RNCs) and Dr. Soledad Funes (mSec61). The processing was done by Dr. Thomas Becker. The model of the PCC was constructed by Dr. Thomas Becker. The models for the r-proteins were constructed by J.-P. Armache.

Interactions of the PCC with the ribosome Both the Ssh1p and mSec61 monomers use the universal ribosomal adapter site as a main contact. The main connections are established by the cytoplasmic loops L6 and L8 (Figure 10.3.2 A). In Ssh1p, loop L6 directly interacts with the loop of H7 (backbone). The loop L8 interacts with the loop of H6 (backbone) and H50 (minor groove and backbone). L8 is also stabilized by r-proteins L23p (rpL25) and L29p (rpL35). Two additional connections are established between i) the loop of the eukaryote-specific RNA expansion segment $ES24^L$ (H59), which interacts with the N-terminus of Ssh1p, and ii) H24 together with L24p (rpL26) stabilizing the C-terminus of Ssh1p (Figure 10.3.2 A, B).

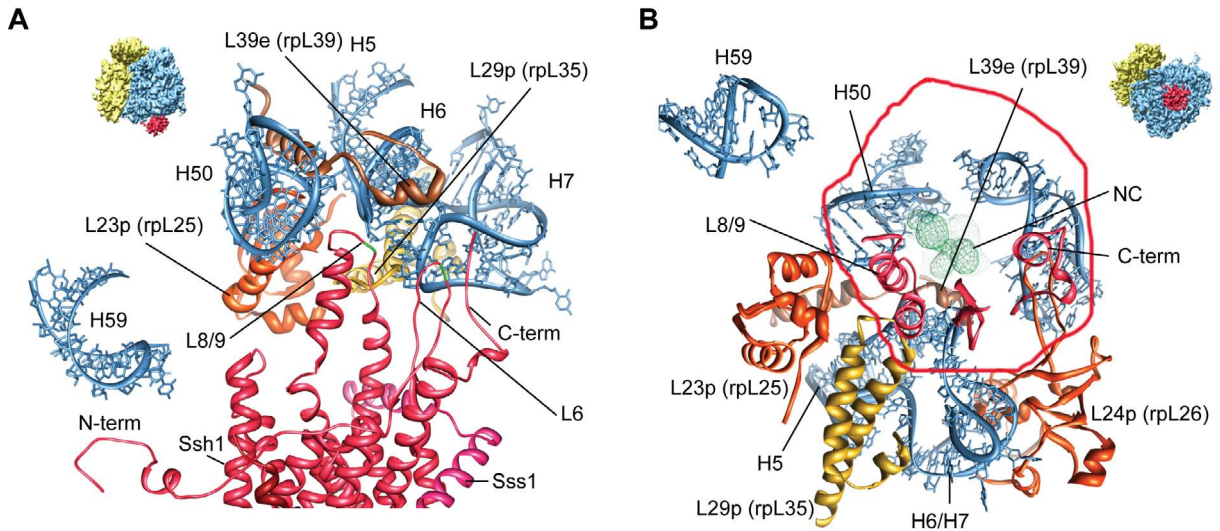


Figure 10.3.2.: **Interaction of Ssh1p with the ribosome.** (A and B) *S. cerevisiae* 80S ribosome (Thumb-nail insets). (A) Molecular models for rRNA (blue) and r-proteins L23p (orange), L29p (yellow) and L39e (brown). The cytosolic half of the Ssh1p and Sss1 model is shown in red and magenta, respectively. The positions of the conserved R278 and R411 are indicated (green). (B) Bottom view. The red line indicates the shape of the Ssh1p. The density for the nascent chain in the tunnel exit site is shown in green. The figures were taken from [141].

10. Molecular interpretation of functional sites of the eukaryotic 80S ribosome

The connections between the 80S ribosome and the mammalian Sec61 complex are similar to those of Ssh1p. Hereby, loop L6 interacts with the loop of H6, H50, L29p (rpL35) and L39e (rpL39). Loop L8 contacts L23p (rpL25), H50 and also H53. As in the Ssh1p homolog the N-terminus of mSec61 interacts with the loop of rRNA expansion segment $ES24^L$, while the C-terminus is stabilized by H24 together with L24p (rpL26) (Figure 10.3.3 A).

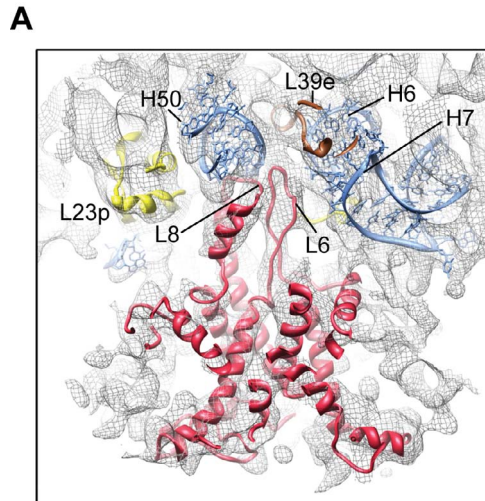


Figure 10.3.3.: **Interaction of mammalian Sec61 (mSec61) with the 80S ribosome.** (A) Side view of the *T. aestivum* 80S ribosome bound to the mSec61 complex. Molecular models for rRNA (blue) and r-proteins L23p (yellow) and L39e (brown). The mSec61 complex is shown in red. Loop L6 interacts with H6, H50, L29p and L39e. Loop L8 interacts with H50 and L23p. The figure was taken from [141].

10.4. Canonical Translation factor binding site

Translational stalling can result from mRNAs forming a stable stem loop or pseudoknot [192]. These stalled ribosomes are rescued by a process called No-Go mRNA decay (NGD), involving the factors Dom34 and Hbs1 [193]. Both factors are homologous to the eukaryotic release factors (Dom34 to eRF1; Hbs1 to eRF3) suggesting that these factors directly interact with the canonical translation factor binding site. Our model was used to analyze the canonical translation factor binding site in eukaryotes.

Sample NGD-intermediates were generated by using ribosomes, which are stalled by a stable stem-loop (SL) structure and subjected to cryo-EM and single particle analysis.

The model of the *S. cerevisiae* 80S ribosome was used for the molecular interpretation of a SL-RNC-Dom34-Hbs1 complex cryo-EM structure at 9.5 Å resolution (Figure 10.4.1 A) [181]. The homology models for Dom34 and Hbs1 were generated based on existing crystal structures of Dom34, Hbs1 and eRF3 [194, 195, 196, 197] by J.-P. Armache. (Figure 10.4.1 A). The samples were prepared by Dr. Thomas Becker and Heidemarie Sieber.

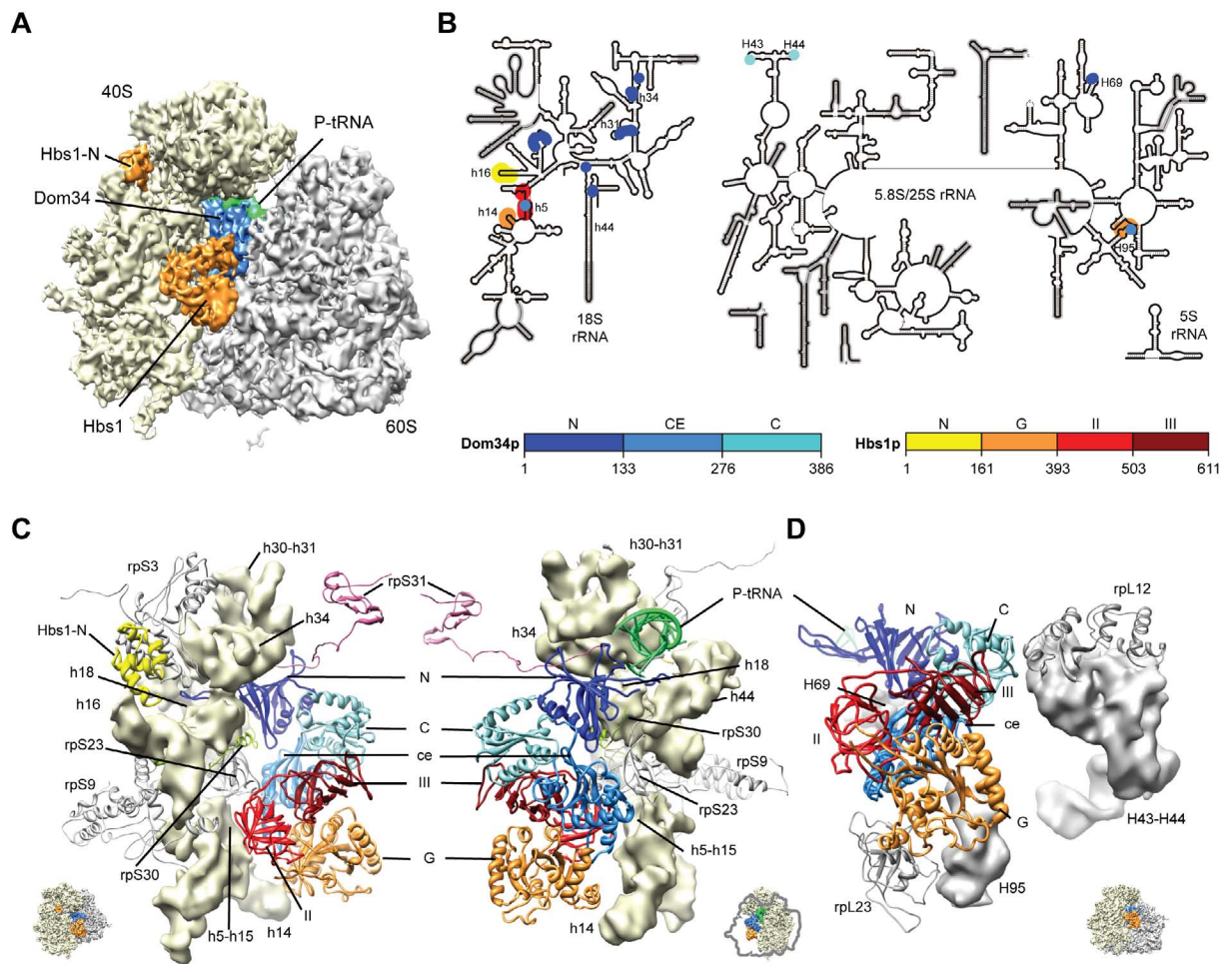


Figure 10.4.1.: **Interaction of Dom34/Hbs1 with the *S. cerevisiae* 80S ribosome.** (A) Cryo-EM reconstruction of stem-loop-RNC-Dom34/Hbs1 complex at 9.4 Å. The small subunit is colored in yellow, the large subunit in gray and the P-site tRNA in green. Additional densities in the canonical factor binding site and near the mRNA entry site were assigned to Dom34 (blue) and Hbs1 (orange). (B) Schematic secondary structure of the ribosomal 5S, 18S and 5.8S/25S rRNA. The protein-RNA interactions are highlighted in the colors for the individual domains of Dom34 and Hbs1, respectively. The color code is given by the schematic representation of the domain organization below. (C) Interactions of Dom34 (blue; color code as in B) and Hbs1 (orange; color code as in B) with the 40S subunit (yellow). Thumbnail of the ribosome (bottom). (D) Interactions of Dom34 (blue) and Hbs1 (orange) with the 60S subunit (gray). Thumbnail of the ribosome (bottom). The figures were adapted from [181].

Interactions of Dom34 with the 80S ribosome The N-terminal domain of Dom34 (Dom34-NTD) reaches deeply into the A-site decoding center of the 40S subunit and forms several contacts with the rRNA helix 18 (h18 or G530 loop), h30, h31, h34 and h44 as well as with r-proteins rpS23 (Figure 10.4.1 B, C). In addition, the NTD of Dom34 interacts with the newly identified eukaryote-specific r-protein rpS31 and the N-terminal extension of rpS30 [20]. In eukaryotes, the N-terminus of rpS30 is in close proximity to the functionally important rRNA nucleotides A1492/A1755 and A1493/A1756 (*E. coli* and *S. cerevisiae* numbering, respectively) of h44 (18S rRNA) and the tip of H69 of the 28S rRNA (Figure 10.4.1 B, C). The central domain of Dom34 forms only minor contacts with the ribosome. The C-terminal domain of Dom34 (Dom34-CTD)

10. Molecular interpretation of functional sites of the eukaryotic 80S ribosome

interacts with the stalk base (H43/H44 and rpL12) of the large subunit (Figure 10.4.1 B, D). The binding of the Dom34/Hbs1 complex causes an inward rotation by $\sim 20^\circ$ of the stalk base relative to H42 (Figure 10.4.2 A). This inward rotation can be observed in a difference map between the Dom34/Hbs1-SL-RNC and the SL-RNC without factors bound (Figure 10.4.3 A-C). A detailed interaction table of Dom34 and the ribosome is summarized in table 10.2.

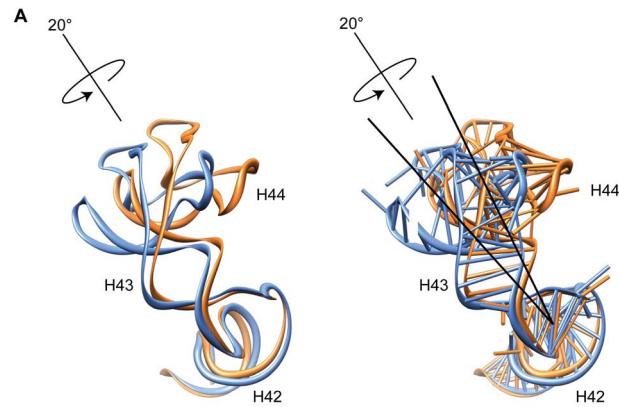


Figure 10.4.2.: **Movement of the stalk base upon binding of Dom34/Hbs1.** (A) Rotation of the stalk base (H43-H44) relative to H42. The stalk base conformation observed in *T. aestivum* 80S ribosome is colored in orange, the stalk base conformation observed in *S. cerevisiae* 80S ribosome bound to Dom34/Hbs1 is colored in blue.

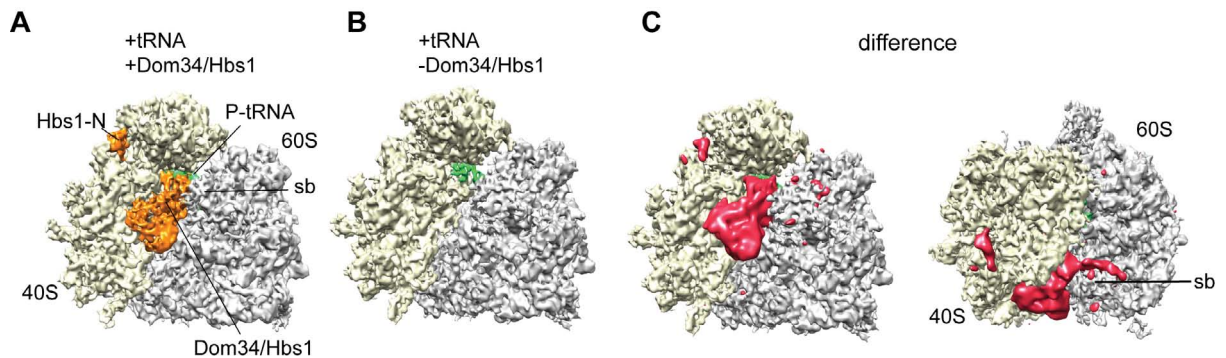


Figure 10.4.3.: **Movement of the stalk base.** Cryo-EM maps of the (A) SL-RNC-Dom34/Hbs1 complex, (B) the control SL-RNC, and (C) the difference map (side and top views) overlaid upon the control SL-RNC map. The stalk base (sb) moves inwards upon Dom34/Hbs1 binding as observed also for the eEF2 binding to 80S ribosomes [67]. The figure was taken from [181].

Interactions of Hbs1 with the 80S ribosome The C-terminal domain of Hbs1 (Hbs1-CTD) interacts with the ribosome in a similar fashion as EF-Tu in a ternary EF-Tu/tRNA/GTP complex. The N-terminal domain of Hbs1 (Hbs1-NTD) was assigned to the additional density bridging rRNA helix 16 (SSU) and rpS3 at the mRNA entry site (Figure 10.4.1 B, C). The Hbs1-NTD interacts with h16 and rpS3 (Figure 10.4.1 B, C). Helix 16 adapts a different conformation compared to h16 in *bacteria* (see also section 9.2.2.3) which provides space for Hbs1 to bind between h16 and rpS3 (Figure 10.4.1 B, C). The G-domain of Hbs1 contacts both the loop of h14 of the 40S subunit and the Sarcin-Ricin-Loop (SRL; H95) of the 60S subunit. The interaction between Hbs1 and the SRL is mediated by the highly conserved E-loop motif (or S-turn, Sarcin-Ricin-Loop) [198, 199, 200]. The G-domain forms multiple contacts with the central domain of Dom34 (Dom-ce). Domain II of Hbs1 is oriented towards the small subunit and interacts with the rRNA helices h5, h15, the r-protein rpS23 (S12p) and the central domain of Dom34. The Hbs1 domain III interacts with the central domain and the CTD of Dom34 and rpL12 at the stalk base of the 60S subunit (Figure 10.4.1 B, D). A detailed interaction table of Hbs1 and the ribosome is summarized in table 10.3.

10. Molecular interpretation of functional sites of the eukaryotic 80S ribosome

Ribosomal subunit	Ribosomal protein	Residue of ribosomal protein	Domain of Dom34	Residue of Dom34
SSU	rpS3 (S3p)	145-146	Dom34-N	53-54
	rpS23 (S12p)	45-46	Dom34-ce	201
		55-59	Dom34-N	46, 49, 58, 80
		85-87	Dom34-ce	167-171, 187, 190, 194
		134	Dom34-ce	11
		143-145	Dom34-ce	151, 169-174
	rpS31 (S27ae)	77-87	Dom34-ce	38-40, 68, 86-92, 120
rps30 (S30e)	1-13	Dom34-N	21, 44, 73-82, 108-114	
LSU	rpL12 (L11p)	24-26	Dom34-C	311-313, 374-379
	rpL23 (L14p)	72	Dom34-ce	232

Ribosomal subunit	Ribosomal RNA	<i>S. cerevisiae</i> rRNA numbering	<i>E. coli</i> rRNA numbering	Domain of Dom34	Residue of Dom34
SSU	h5	432	360	Dom34-ce	183
	h18	564-566	517-519	Dom34-N	47-50, 82, 103-106
		575-578	528-531		47-52, 58, 100, 103-104
	h30/h31	1179	954	Dom34-N	94
		1182-1183	957-958		10-12
		1187-1190	962-965		90-93
	h34	1271-1274	1051-1054	Dom34-N	50-51, 89, 100-102
		1427	1196		50-54, 57
	h28	1634-1635	1397-1398	Dom34-N	55-58
	h44	1756	1493	Dom34-N	46, 113
LSU	H43	1242	1067-1068	Dom34-ce	312, 345, 373-374
	H44	1270	1095-1096	Dom34-C	373-375
	H69	2256-2257	1913-1914	Dom34-N	62, 112
	H95	3029	2662	Dom34-ce	218

Table 10.2.: Interactions between r-proteins and rRNA and Dom34. The tables was taken from [181].

10. Molecular interpretation of functional sites of the eukaryotic 80S ribosome

Ribosomal subunit	Ribosomal protein	Residue of ribosomal protein	Domain of Hbs1	Residue of Hbs1
SSU	rpS3 (S3p)	52-56	Hbs1-N	28, 31, 37-40
		90-91		29-32, 35, 37
		94		29
		108		24
		117		20-21
		124-125		25
	rpS23 (S12p)	109	Hbs1-II	484
		130-145	Hbs1-II	408-409, 416-417, 439-442, 483-490
			Hbs1-III	531-532, 574-576
rpS30 (S30e)	54	Hbs1-N	22	
LSU	rpL23 (L14p)	127-131	Hbs1-G	205-211

Ribosomal subunit	Ribosomal RNA	<i>S. cerevisiae</i> rRNA numbering	<i>E. coli</i> rRNA numbering	Domain of Hbs1	Residue of Hbs1
SSU	h14	414-417	342-345	Hbs1-G	197-198, 201
	h5	430-432	358-360	Hbs1-II	412-413,
	h15	439-440	367-368	Hbs1-II	413, 460, 463
	h16	487-489	408-410	Hbs1-N	79-84
		492-495			75-81
LSU	H95	3022	2655	Hbs1-G	211, 222, 225, 354
		3026-3028	2659-2661		176-177, 281, 285

Table 10.3.: Interactions between r-proteins and rRNA and Hbs1. The tables was taken from [181].

Part IV.

Conclusion and discussion

11. Pipeline for modeling and fitting RNA

In section 8 a novel approach for modeling and fitting RNA was presented. This approach provides a new workflow starting from the primary sequence of an RNA molecule up to the refinement and flexible fitting of the 3D model into an electron density. The workflow can be applied to all kinds of structured RNA and is in general independent from the input size. This approach is built in a modular way, meaning that each step can be applied separately and independently from each other. Moreover, these modules can be extended or replaced by other more accurate modules or algorithms, which can be easily implemented within the current framework. In conclusion, the current workflow is more user-friendly, however, expert knowledge of RNA chemistry is still indispensable.

Manual alignment vs. automated alignment The workflow includes several manual steps and decisions. The manual construction of the structure-based sequence alignment with S2S ensures an accurate alignment which, in contrast to a simple sequence alignment tool like ClustalW2¹ [201], reflects also the structural constraints as well as non-Watson-Crick base pair interactions. An example for the requirement of a highly accurate alignment is rRNA helix 24 (H24) on the large subunit. The model was built based on two different multiple sequence alignments (MSA) between the template of *H. marismortui* and the target sequences of *D. melanogaster*, *S. cerevisiae* and *T. aestivum*. The first one is an alignment generated automatically by ClustalW2 [201] (Figure 11.0.1 A). The second alignment is constructed manually with S2S² [135] (Figure 11.0.1 B).

There are two significant differences in the alignments highlighted with red squares (Figure 11.0.1 A, B). ClustalW2 aligns the target sequences to the pentaloop (alignment position 727-731) present in the template structure. Thereby, the adenine residues get aligned to the last adenine residue in the template (alignment position 731). At the second position (alignment position 737) a gap is introduced in the target sequences while a guanosine, involved in a $U^{\circ}G$ base pair, is present in the template. Based on this alignment the generated model contains a pentaloop with the clearly visible bulged-out adenine; and a gap in a helical region of H24 (Figure 11.0.1 C).

In the manual alignment the tetraloop (target sequence) is not aligned to the pentaloop (template) since the GAAA sequence forms a GNRA-tetraloop, resulting in a different conformation and sequence length (Figure 11.0.1 B). The tetra-loop was folded separately since the structural motif changes (see also Figure 8.1.1 B). Moreover, the adenine being misaligned to the A731 by ClustalW2, is aligned to base pair with the U (alignment position 726) which leads to an isosteric base pair substitution from $U^{\circ}G$ to $U - A$. Consequently, the $U^{\circ}G$ (cWW) base pair

¹<http://www.ebi.ac.uk/Tools/msa/clustalw2/>

²<http://www.bioinformatics.org/S2S/>

11. Pipeline for modeling and fitting RNA

get aligned to an $A^{\circ}G$ (tHS) non-Watson-Crick base pair which is not isosteric, but does not disturb the 3-dimensional structure. Based on the manual alignment with S2S the generated model contains a GNRA-tetraloop and a non-Watson-Crick $A^{\circ}G$ (tHS) base pair (Figure 11.0.1 E). The differences between the models are evident and visible by superpositioning the two models (Figure 11.0.1 D).

Both models can be judged by the recently published low-resolution crystal structure of the *S. cerevisiae* 80S ribosome at 4.5 Å resolution [19]. The high structural similarity between the S2S-model and the X-ray structure become evident by superpositioning both of them (right panel of Figure 11.0.1 F) and support the manual alignment to be correct. On the other hand the superpositioning of the ClustalW2-model and the crystal structure reveals the structural errors of the model resulting from the automated alignment (left panel of Figure 11.0.1 F).

This comparison demonstrates that the automated alignment generation only partially works, and deducing models from these alignments leads only to rough approximations. As a consequence, the resulting model shows significant errors in the loop region and the helical region. In contrast, the structural-based sequence alignment with S2S, which allows a manual alignment and thus the incorporation of expert knowledge, leads to a correct model (middle panel of Figure 11.0.1 F). This is a tremendous advantage over simple sequence alignment tools, since they only use sequence information and consequently do not incorporate any knowledge of structural motifs.

11. Pipeline for modeling and fitting RNA

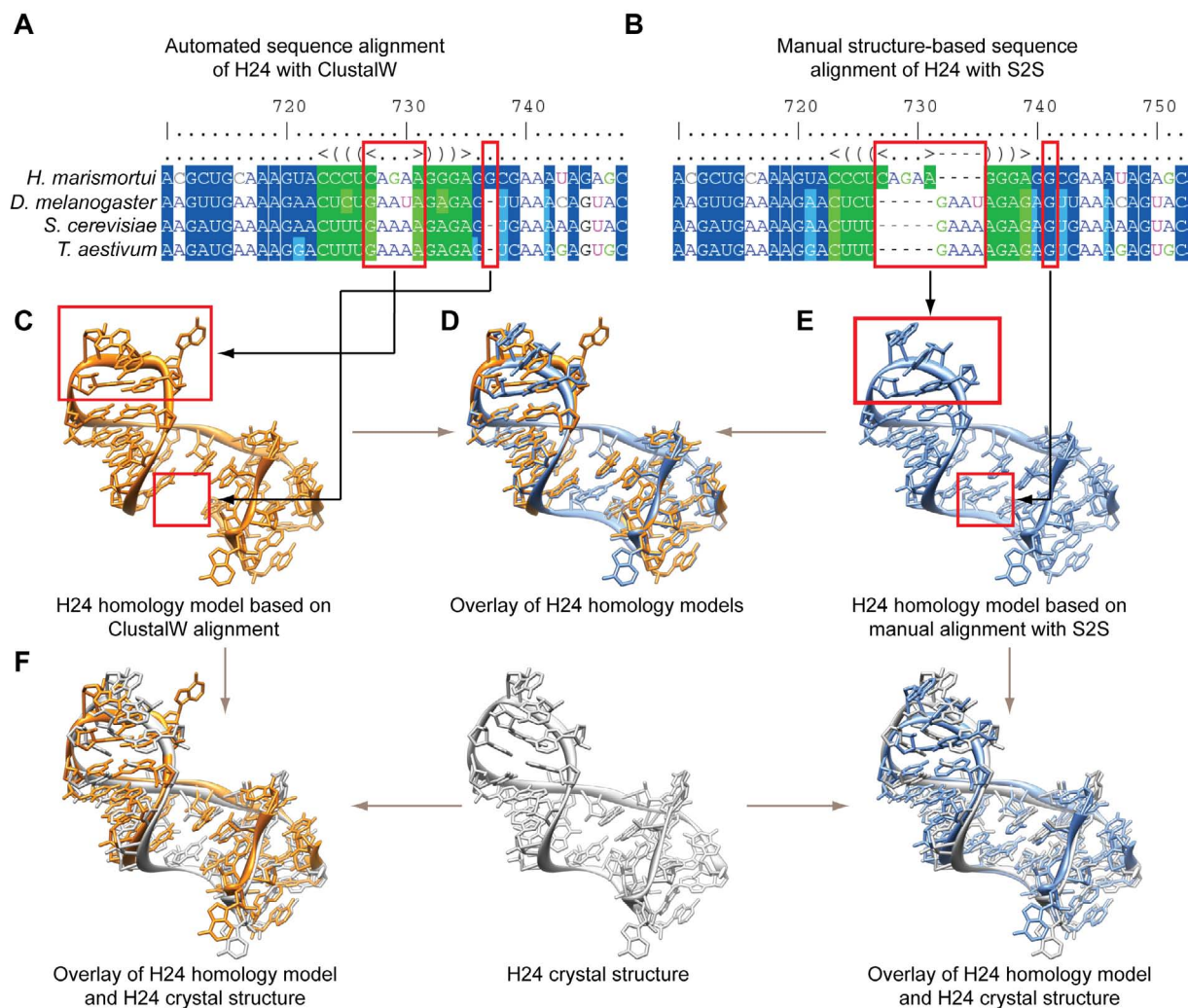


Figure 11.0.1.: **Comparison between manual and automated alignment.** (A) Automated sequence alignment of H24 with ClustalW2 and (B) manual structure-based sequence alignment of H24 with S2S of the target sequences from *D. melanogaster*, *S. cerevisiae* and *T. aestivum* are aligned to the template sequence from *H. marismortui*. The alignment differences are highlighted by a red square. (C) Homology model of H24 deduced from the template *H. marismortui* which results from the sequence alignment by ClustalW2. (D) Superpositioning of the models from (C) and (E). (E) Homology model of H24 deduced from the template *H. marismortui* which results from the structure-based sequence alignment by S2S. The loop was adapted separately using Assemble. (F) Superpositioning of the model from (C) and the crystal structure (left panel), the X-ray structure itself, and the superpositioning of the model from (E) and the crystal structure.

Homology modeling of RNA with S2S and Assemble Building RNA homology models based on structure-based sequence alignments was presented in section 8.1 and is performed semi automatically using S2S and Assemble. The resulting model not only includes accurate canonical Watson-Crick base pairs but also non-Watson-Crick interactions involved in structural motifs and tertiary interactions. The refinement includes all 96 observed base pair combinations (according to the geometric nomenclature [85, 173]) which ensures the correct hydrogen bonding between the bases.

De novo modeling of RNA Currently, algorithms predicting the entire 3-dimensional structure of an RNA molecule from its primary sequence are restricted by the input size and limits are already reached by molecules larger than 30 nts [125, 132]. RNA structure predictions with MC-fold and MC-Sym are able to fold human precursor microRNAs (50-70 nts), however, the lowest free-energy states determined by MC-fold often differ from the active or experimental states [126].

The *de novo* modeling approach presented in this work is able to build large RNA models (i.e. ~ 200 nts for *ES7^L* in *T. aestivum*; ~ 150 nts for *ES27^L* in *S. cerevisiae*). RNA models can be freely adjusted with respect to base pairing in the 2D map and the 3D model.

As prominent new features empirical data (i.e. cryo-EM density maps) and templates for structural RNA motifs (i.e. kink-turn motifs and loop motifs) can be incorporated while model building. However, manual construction of models requires time and expert knowledge but leads to more reliable RNA models based on empirical data.

Flexible fitting of RNA Because common methods for RNA modeling are, to date, not capable of incorporating electron density data we developed a new approach (section 8.3) that combines refinement and flexible fitting of RNA models. This approach incorporates the geometric optimization of distances between RNA residues according to the isostericity class, and provides a new flexible fitting method into EM densities. In addition to RNA refinement, the new approach is able to refine and fit RNA and proteins together, which leads to more accurate models as compared to rigid body fitting of RNA models used in previous studies [22, 21].

Because the MD simulations are performed without water and ions, the RNA secondary structure and base pairing have to be restrained manually.

MDFF simulations on the 80S ribosome and RNA restraints We performed an all-atom MDFF simulation for both the *T. aestivum* and *S. cerevisiae* 80S ribosome models into the cryo-EM density maps. Both all-atom simulations were performed with hydrogens leading to an input size of 349824 atoms for the *T. aestivum* 80S ribosome and 343881 atoms for the *S. cerevisiae* 80S ribosome, respectively. The calculation time for each ribosome model was ~ 166 hours and was performed on 8 CPU cores with 24GB of RAM summing up to ~ 1300 CPU-hours. For all calculations water molecules and ions were excluded, however, the calculation time is in principle feasible. Due to the exclusion of water and ions, we improved the RNA restraints used in the MDFF simulations by introducing all hydrogen bonding schemes of all Watson-Crick and non-Watson-Crick base pairs.

The two MDFF simulations on 80S ribosomes including hydrogens are proof that large-scale flexible fitting is feasible for huge RNA-protein complexes. Compared to the MDFF simulations of bacterial 70S ribosomes [168, 202], the simulations we presented, consist of an additional ~ 1000 nts and 20 proteins.

12. Molecular model of the 80S ribosome

Modeling effort for the ribosomal RNA Modeling the eukaryotic rRNA has a total input size of about 5500 nts (120 nts 5S rRNA, ~ 1800 nts 18S rRNA and ~ 3500 nts 28S rRNA). The majority of the rRNA (~ 65%) was modeled based on homology to the *bacterial* 16S rRNA and the *archaeal* 23S rRNA templates. The remaining parts (~ 35%) of the rRNA comprising structurally VR and ES were modeled *de novo*. Only ~ 2.2% of the rRNA were not modeled. Compared to other structured RNAs, i.e. the let-7 miRNA (33 nts), the Asp-tRNA (76 nts), the TPP riboswitch (80 nts) and to the mammalian SRP (108 nts) the input size is significantly larger (50-fold compared to mammalian SRP, 71-fold compared to Asp-tRNA/TPP-riboswitch, and 166-fold compared to the let-7 miRNA) for a eukaryotic ribosome (Figure 12.0.1).

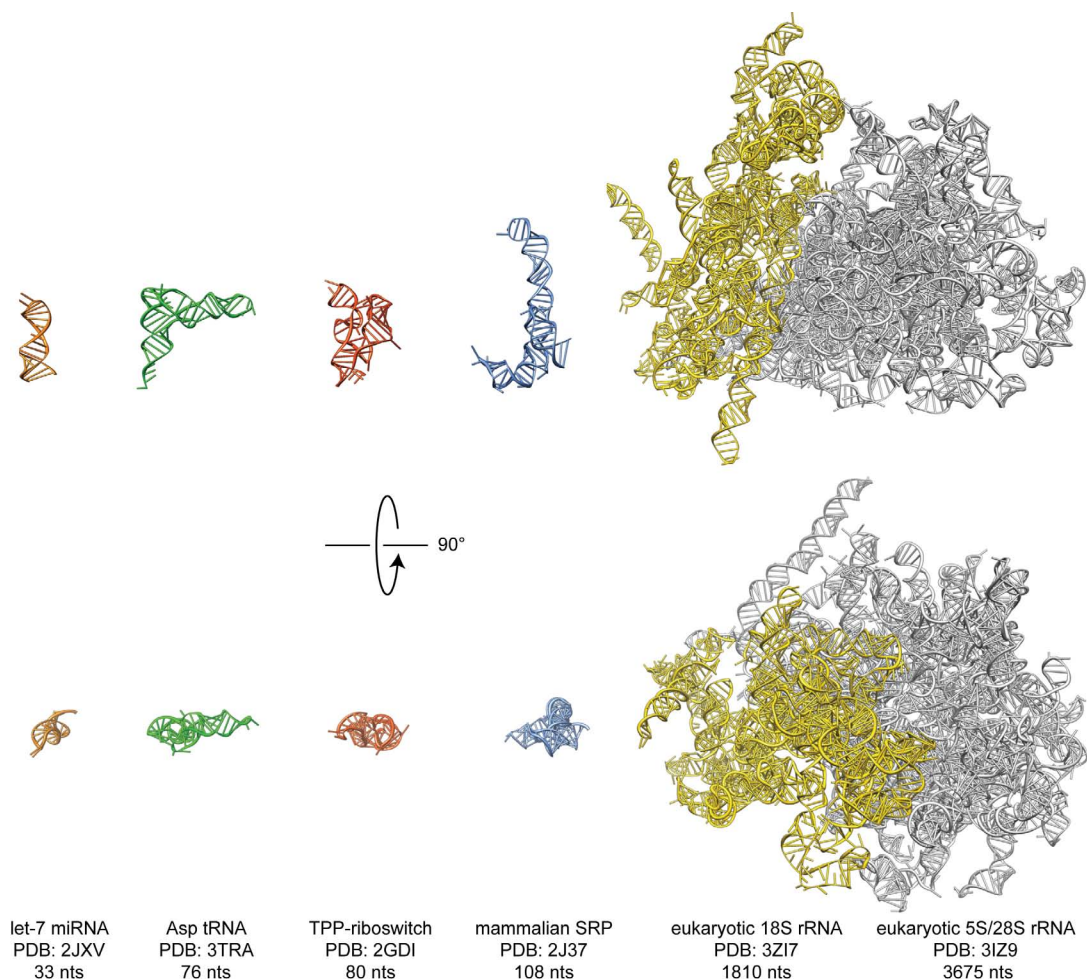


Figure 12.0.1.: **Size comparison of structured RNAs.** Size comparison (side and top views, rotated by 90°) between the let-7 miRNA (orange, 33nts), the Asp-tRNA (green, 76 nts), the TPP-riboswitch (red, 80 nts) and the mammalian SRP (blue, 108 nts) with our rRNA model consisting of the 18S rRNA (yellow, 1810 nts) and the 5S/28S rRNA (gray, 3675 nts).

12.1. Evaluation of RNA models for the eukaryotic ribosome

The molecular models for both the *T. aestivum* and *S. cerevisiae* 80S ribosomes are presented in this work. Based on the two cryo-EM reconstructions it was possible to model $\sim 98\%$ of the rRNA and $\sim 93\%$ of the ribosomal proteins.

To evaluate the quality of our RNA model, we compare it to previously published cryo-EM models [21, 22], and two very recent crystal structures [19, 20].

12.1.1. Comparison to the RNA model of the *mammalian* 80S ribosome

A cryo-EM reconstruction at 8.7 Å of a canine ribosome was used for building a model of the canine 80S ribosome [21]. In this study the 16S rRNA from the crystal structures of *T. thermophilus* [3] and the 5S/23S rRNA from the *H. marismortui* [7] were fitted into the density as rigid bodies. Some non-fitting parts were moved as rigid bodies relative to the rest. Based on tracing the electron density $\sim 50\%$ of the expansion segments were modeled by fitting of A-form helices into ES densities [21].

The models presented in our work are superior compared to the models of the canine 80S ribosome in terms of completeness, accuracy and reliability. Regarding the ribosomal core, the canine model does not reflect the sequence of the ribosomal core. The RNA expansion segments were modeled as A-form helices fitted roughly into the electron density. The ES are not connected to the core model which leads to an inaccurate placement of the RNA parts. Moreover, most single stranded parts of the ES are not modeled, which partly results from ambiguous or non-existing electron density and limited by the resolution (8.7 Å). Highly flexible ES like *ES27^L* could not be modeled due to non-existing density; and intermolecular contacts between *ES3^S* and *ES6^S*, as suggested by Alkemar & Nygård [28], were not modeled. In addition, at the given resolution identification of eukaryote-specific ribosomal proteins was not possible.

12.1.2. Comparison to the RNA model of the *Thermomyces lanuginosus* 80S ribosome

A more comprehensive model for *S. cerevisiae* was built and fitted into a 8.9 Å cryo-EM reconstruction of the thermophilic fungus *Thermomyces lanuginosus* [22]. This fungus shares an rRNA sequence identity of only $\sim 85\%$ with *S. cerevisiae*. The model of the thermophilic fungus included all ES and VR except for *ES10^L*, *ES27^L* and the tip of *ES15^L*. The modeling effort was more complete, however, several significant differences in the structure between the *S. cerevisiae* model presented by Taylor et al. [22] and the *S. cerevisiae* model presented in this work are evident. (Figure 12.1.1 A-E). For example, in our model the r-protein S7e was assigned to a density close to expansion segment *ES6^S* (Figure 12.1.1 A right panel), where Taylor et al. modeled a small helix of *ES6^S* in the same density (Figure 12.1.1 A left panel). Moreover, we assigned a rod-like density for r-protein L19e close *ES6^S* whereas another loop of *ES6^S* is assigned to this density in the other model.

In our model the variable region of H16-H18 and expansion segment *ES5^L* were entirely fitted and we identified the eukaryote-specific r-protein L13e (Figure 12.1.1 B right panel). In the model by Taylor et al. neither the model of *ES5^L* is fitted to the corresponding density, nor the

12. Molecular model of the 80S ribosome

r-protein L13e was identified. Further, the VR H18 was not fitted correctly. Expansion segment $ES12^L$ was modeled similar, however, the connecting single stranded region to H38 and $ES9^L$ were modeled differently. We modeled the single stranded region along the unoccupied density between $ES12^L$ and r-protein L21e, whereas the model by Taylor et al. is not fitted into any density (Figure 12.1.1 C).

In our model the GNRA-tetraloop of $ES20^L$ was modeled to interact with the minor groove of H58 and the eukaryote-specific r-protein L34e (Figure 12.1.1 D right panel). In contrast, the model by Taylor et al. $ES20^L$ was modeled to interact with a single stranded part of H58. Further, the five nucleotides corresponding to $ES26^L$ were modeled as helical parts and placed out of density (Figure 12.1.1 D left panel).

Another result of our model is the identification of the eukaryote-specific r-protein which was assigned in the vicinity of $ES39^L$. Helix $ES39^La$ was fitted into the corresponding density, whereas the model by Taylor et al. does not fit. Moreover, the single stranded region between $ES39^La$ and $ES39^Lb$ was not placed in any density (Figure 12.1.1 E left panel).

We further provided models for $ES10^L$, $ES15^L$ and $ES27^L$ which we modeled in their entirety. In addition, we modeled the interaction between $ES3^S$ and $ES6^S$ as suggested by Alkemar & Nygård [28].

12. Molecular model of the 80S ribosome

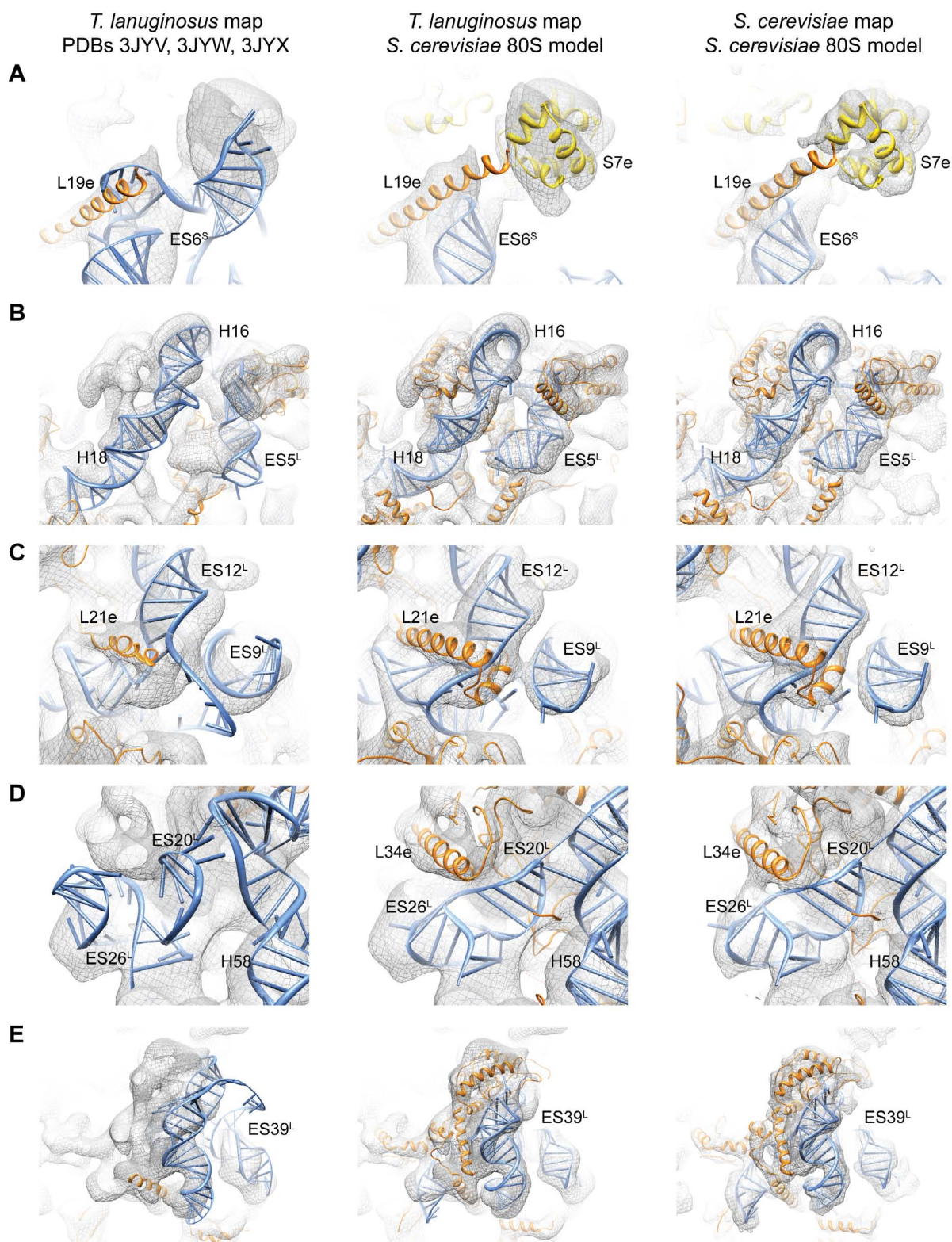


Figure 12.1.1.: **Comparison of *S. cerevisiae* model and fit in *T. lanuginosus* and *S. cerevisiae* 80S cryo-EM maps.** (Left column) PDB 3JYV, 3JYW and 3JYX in the associated cryo-EM map of *T. lanuginosus* 80S ribosome from [22]. (Middle Column) The fit of the *S. cerevisiae* model of this work into the cryo-EM map of *T. lanuginosus* 80S ribosome and into the *S. cerevisiae* 80S ribosome reconstruction (Right column). (A-E) Molecular models of rRNA (blue) and r-proteins (orange and yellow, respectively). The figure was taken from [182]

12.1.3. Comparison to the X-ray structure of the *S. cerevisiae* 80S ribosome at 4.5 Å

During the writing of this manuscript a crystal structure of the *S. cerevisiae* 80S ribosome at 4.5 Å [19] (PDB: 3O2Z¹, 3O30², 3O58³ and 3O5H⁴) and a crystal structure of the *Tetrahymena thermophila* 40S ribosomal subunit in complex with eukaryotic initiation factor 1 (eIF1) at a resolution of 3.9 Å (PDB: 2XZM⁵ and 2XZN⁶) [20] were published. Our models for the expansion segments and variable regions are mainly confirmed by these X-ray structures. For example our model for the *S. cerevisiae* *ES7^L* (Figure 12.1.2 A) is extremely similar compared to the X-ray structure at 4.5 Å (Figure 12.1.2 B, C). Beyond that, we provide a more complete model of *ES7^L* by modeling the highly flexible helix *ES7^La* which was not observed in the X-ray structure. Our model for *ES31^L* (Figure 12.1.2 D) is in good agreement with the X-ray structure (Figure 12.1.2 B, C), however we could not model the single stranded RNA region (13 nts) between the helices *ES31^Lb* and *ES31^Lc* due to unambiguous density. The adjacent expansion segments *ES3^L* and *ES4^L* (Figure 12.1.2 G) also show the high quality of our model. Compared to the X-ray structure our model has a small deviation at the end of *ES3^L* which can be explained by the flexibility of the helical end. Another small conformational deviation is observed at our model of *ES4^L* which adopts a slightly different conformation as compared to the X-ray structure. On the small subunit we observe differences in h17, which mainly consists of non-Watson-Crick interactions; the order of helices in *ES6^S* (see section 12.2.2); and in the single stranded parts of the VR h33. On the large subunit differences are observed at the loop of *ES15^L*, which is not resolved in our cryo-EM maps; and single stranded parts of *ES39^L*, which are also not resolved in our cryo-EM maps.

¹<http://www.pdb.org/pdb/explore/explore.do?structureId=302Z>

²<http://www.pdb.org/pdb/explore/explore.do?structureId=3030>

³<http://www.pdb.org/pdb/explore/explore.do?structureId=3058>

⁴<http://www.pdb.org/pdb/explore/explore.do?structureId=305H>

⁵<http://www.pdb.org/pdb/explore/explore.do?structureId=2XZM>

⁶<http://www.pdb.org/pdb/explore/explore.do?structureId=2XZN>

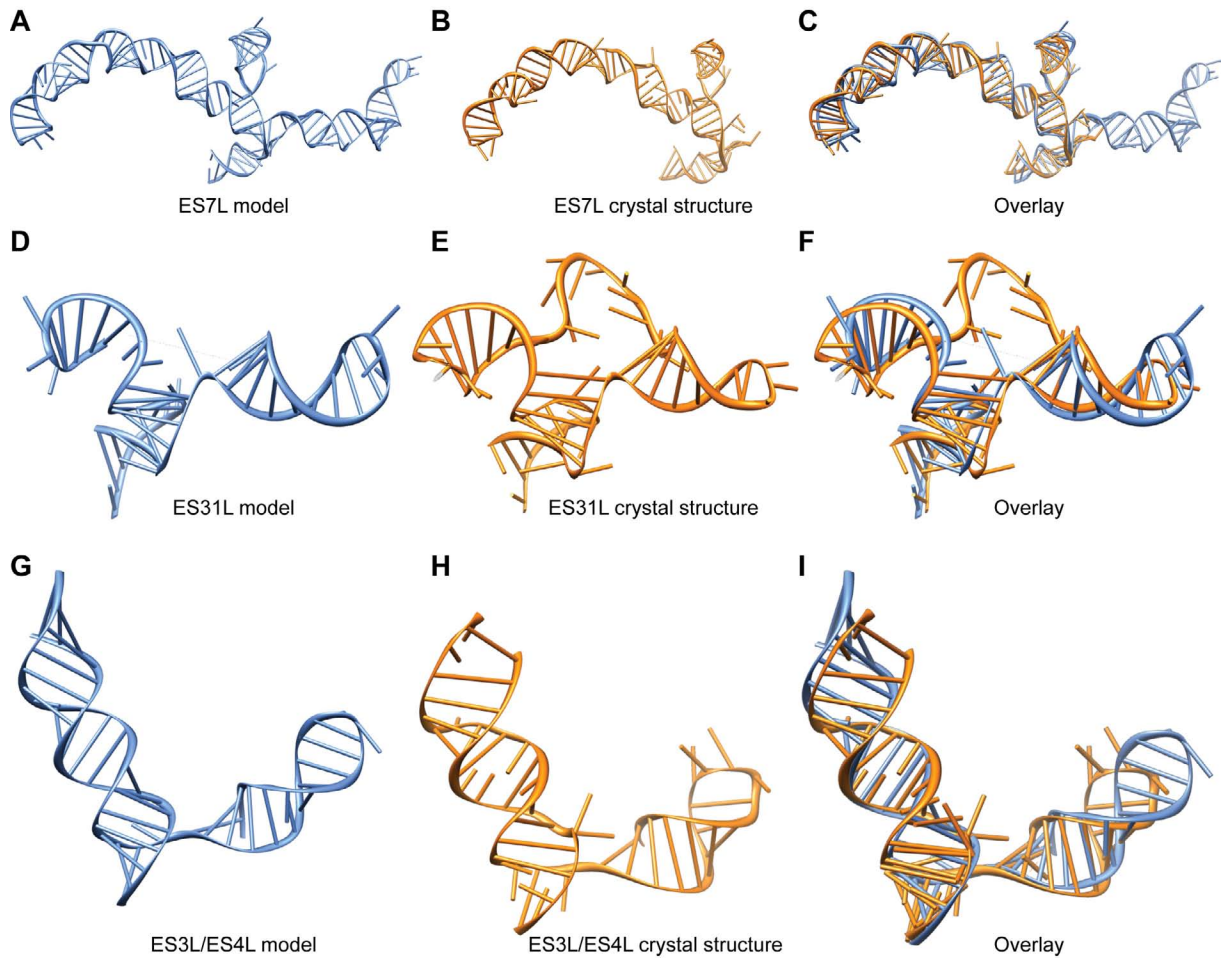


Figure 12.1.2.: **Comparison of *S. cerevisiae* RNA model and crystal structure.** Comparison of $ES7^L$ of (A) the model this work, (B) the X-ray structure of $ES7^L$ and (C) the overlay of both structures.

Comparison of $ES31^L$ of (D) the model of this work, (E) the X-ray structure of $ES31^L$ and (F) the overlay of both structures.

Comparison of $ES3^L/ES4^L$ of (G) the model of this work, (H) the X-ray structure and (I) the overlay of both structures. The coordinates from the crystal structure are taken from PDB: 3O5H [19].

In general, our model of the *S. cerevisiae* 80S ribosome and the crystal structure are extremely similar. Small structural deviations are evident, but expected at our resolution between 5.5 Å - 6.1 Å. For example, in some parts of the structure we observe a register shift by half of a nucleotide or one nucleotide, respectively, which was expected due to fact that we could not discriminate two adjacent bases in the cryo-EM density (Figure 9.1.2).

Nevertheless, for some parts of the RNA our model is better than the crystal structure. For example we provide molecular models for $ES27^L$ and $ES7^La$ which have functional implications (see section 1.1.5 and references [33, 34]) but are not resolved in the crystal structure.

Taken together, modeling of RNA based on high-resolution cryo-EM maps leads to highly accurate models which are comparable to crystal structures. Yet, only modeling of single stranded RNA regions and junctions between RNA helices are still limited.

12.2. The ribosomal expansion segments and variable regions

With this work we provide models for all expansion segments and variable regions of the fungal and plant ribosome which are based on a 6.1 Å and 5.5 Å cryo-EM map. For each expansion segment we provide the secondary structure, a molecular model and the interacting r-proteins (see section 9.2) which can be used as an inventory for studying the eukaryote-specific expansion segments.

By comparing each ES in both species we can conclude that all expansion segments, except for *ES7^L*, *ES27^L* and *ES39^L*, are very similar regarding to size and fold.

In addition to the ES, we also provide molecular models for the structurally variable regions (VR) in the ribosome. We also adapt the secondary structure maps according to our models of ES and VR (Figures 9.2.3, 9.2.4, 9.2.5, 9.2.6, 9.2.7 and 9.2.8).

In general, the biological function of all ES is unknown, however, there are functional implications for *ES7^L* and *ES27^L* [33, 34].

12.2.1. Interaction between *ES3^S* and *ES6^S*

Originating from cryo-EM studies in 2001, expansion segments *ES3^S* and *ES6^S* on the 40S subunit are suggested to interact with each other (see 1.1.4 C) [16]. Moreover, in 2003 sequence studies of *ES3^S* and *ES6^S* from more than 2900 discrete eukaryotic species were analyzed for a possible sequence complementarity between the two ES, suggesting that *ES3^S* and *ES6^S* could interact by forming a helix consisting of seven to nine base pairs [28]. A particular region between *ES3^{Sb}* and *ES3^{Sc}* which has been suggested to be involved in forming the hybrid helix with *ES6^S*, was falsely predicted to base pair with another region of *ES3^S* by the CRW database⁷ [27]. Although the *ES3^S/ES6^S* interaction seems to be conserved, no such interaction between the two ES has been modeled in in the fungal or canine 80S ribosomes [21, 22].

In this work we present the first molecular model of the interaction between the *ES3^S* and *ES6^S* which is observed in both cryo-EM density maps of the *T. aestivum* and the *S. cerevisiae* 80S reconstructions [182] (Figure 12.2.1 A). Our model presents a new pseudoknot which is formed by the internal loop between *ES3^{Sb}* and *ES3^{Sc}* and the loop of *ES6^{Sd}*, as predicted before [28]. This model and the type of interaction is largely confirmed by the recently published crystal structure of the 40S subunit at a resolution of 3.9 Å [20] (Figure 12.2.1 B, C).

⁷<http://www.rna.cccb.utexas.edu/>

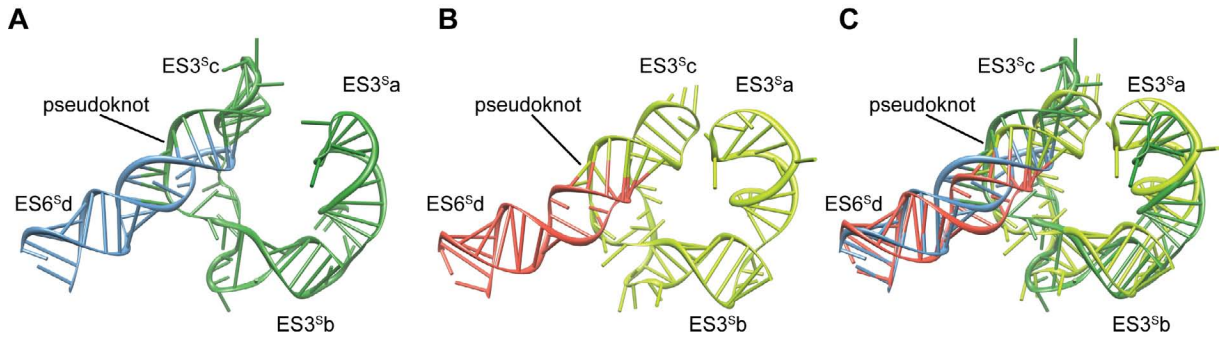


Figure 12.2.1.: **Comparison of interaction between ES3S and ES6S.** (A) Cryo-EM model of the interaction between $ES3^S$ and $ES6^{Sd}$. (B) Crystal structure of the interaction between $ES3^S$ and $ES6^{Sd}$ in *S. cerevisiae* (PDB: 2XZM) [20]. (C) Overlay of cryo-EM model and the X-ray structure.

12.2.2. Expansion segment $ES6^S$

Unlike most of the expansion segments, $ES6^S$ is present in all eukaryotic 16S-like rRNAs, suggesting that it may have a functional role [203]. $ES6^S$ has an average length of ~ 250 nts, but can contain up to 800 nts in protists and insects [204, 205, 206]. $ES6^S$ can be divided in two halves, being different in sequence variability characteristics. Secondary structure predictions from the CRW database⁸, which are based on covariation analysis, predict two helices for the 5'-half [27]. The 3'-half of $ES6^S$ is relatively invariable and makes secondary structure predictions based on covariation challenging. Several secondary structure predictions were suggested with 5 – 8 helices [207, 208, 209, 210, 211]. In 2001, $ES6^S$ was observed to emerge at the solvent side of the platform and appearing to branch into two irregular helices (see 1.1.4 C) [16].

In our *T. aestivum* and the *S. cerevisiae* 80S cryo-EM reconstructions we observe densities accounting for only four helices. Therefore we built a 3D model, which was based on secondary structure predictions with RNashapes that were constrained to contain only four helices. This model was fit into the density in a way that the pseudoknot with $ES3^S$ could be established (Figure 12.2.2 A). The placement of remaining three helices, however was difficult, since in the cryo-EM densities no connections between the four helices were resolved, which made it impossible to assign the order of the helices. Moreover, we were not able to identify the helix which corresponds to the h21 counterpart in *bacteria*. We therefore assigned the helix $ES6^{Sa}$ at the 5'-end to be the h21 counterpart, and $ES6^{Sb}$ and $ES6^{Sc}$ were placed arbitrarily into the EM map.

Compared to the X-ray structure of the 40S subunit at 3.9 Å [20] we correctly assigned the helices $ES6^{Sb}$ and $ES6^{Sd}$ at the 3'-end of the expansion segment, however, helices $ES6^{Sa}$ and $ES6^{Sc}$ were switched due to the absence of density in the helix connecting regions (Figure 12.2.2 B, C). It turned out that $ES6^{Sc}$ is the eukaryotic counterpart of h21 but contains an additional small helix located perpendicular to the h21 analogous part of $ES6^S$ resulting in a new three-way junction (indicated by an asterisk in Figure 12.2.2 B, C). This three-way junction has not been predicted by any secondary structure prediction program and thus could not be modeled.

⁸<http://www.rna.cccb.utexas.edu/>

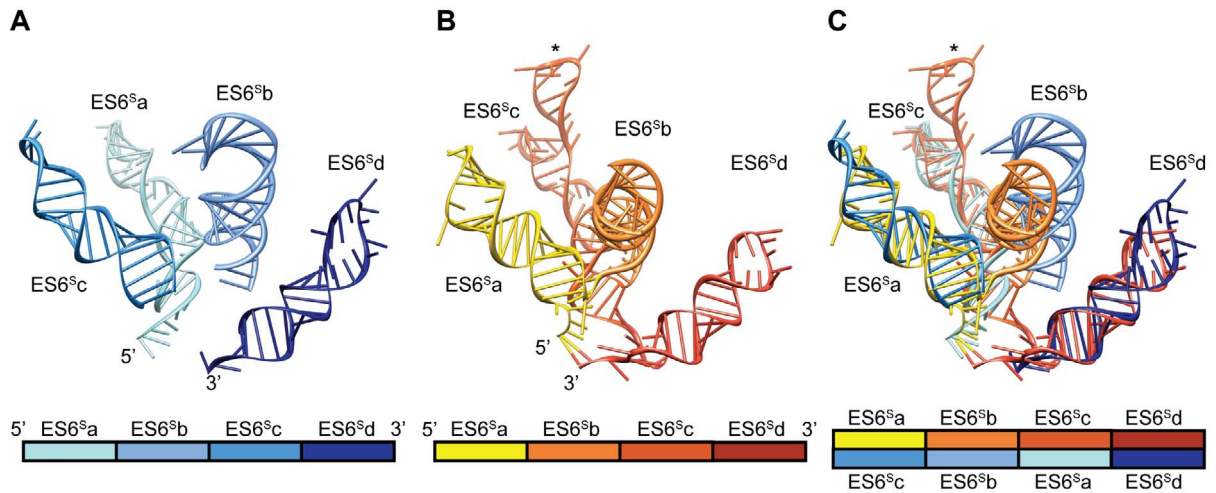


Figure 12.2.2.: **Order of the helices in ES6S.**(A) Order of the $ES6^{Sd}$ helices in the cryo-EM model. (B) Order of the $ES6^{Sd}$ helices in the X-ray structure in *S. cerevisiae* (PDB: 2XZM) [20]. (C) Overlay of $ES6^{Sd}$ cryo-EM model and the $ES6^{Sd}$ X-ray structure.

12.2.3. Expansion segment $ES7^L$

In sections 9.2.2.12 and 9.2.2.13 we present two models for expansion segment $ES7^L$ (for *T. aestivum* and *S. cerevisiae*) and made a comparison (see section 9.3.2.2). In both models we observe a new kink-turn motif in $ES7^L$, which is stabilized in case of *T. aestivum* by L28e. However, it is not possible to relate this kink-turn motif to a known one, since our resolution does not allow base pair assignments.

In a comparison between the two species we identified the presence of a 3-helix bundle comprised of $ES7^{Lc} - e$ arranged in a three-way junction in *T. aestivum*. By tracing the density for the N-terminus of r-protein L6e we suggest that the N-terminal helix inserts through the three-way junction of $ES7^L$ (Figure 9.3.2 C). To our knowledge, this kind of RNA-protein interaction has not been reported previously.

12.2.4. Expansion segment $ES27^L$

In our work we present the first molecular model for the highly flexible expansion segment $ES27^L$. The models are presented in three distinct conformations, $ES27_{in}^L$, $ES27_{out}^L$ and $ES27_{int}^L$ (see section 9.2.2.20, 9.2.2.21). We suggest a continuous motion of $ES27^L$ relative to H63 between the two positions $ES27_{in}^L$ and $ES27_{out}^L$ (Figure 9.3.3A, B). This motion involves a $\sim 110^\circ$ rotation and is mediated by the single stranded region located at the 3'-end of $ES27^L$ as well as the RNA three-way junction.

The model of $ES27^L$ in the $ES27_{out}^L$ conformation also confirms that the position of $ES27^L$ is close to the tunnel exit. Although it has been suggested that $ES27^L$ can directly interact with the nascent chain [17], we do not observe a direct connection. However, our model is in good agreement with this suggestion since the loop of $ES27_{out}^L b$ reaches to the tunnel exit (Figure 12.2.3 A).

Previously, the $ES27_{in}^L$ conformation (L1 position) was only observed in reconstructions of ribosome-channel complexes [17]. With our models of $ES27^L$ in the $ES27_{out}^L$ position (towards

the tunnel exit) and the model of the Sec61/Ssh1 translocon we can indeed observe an interference between helix $ES27^{Lb}$ and the PCC and the membrane (Figure 12.2.3 B). Consequently, we can also conclude that $ES27^L$ does not adopt the $ES27_{out}^L$ position while being bound to the PCC, but rather adopts the $ES27_{in}^L$ position as suggested by Beckmann et al. [17].

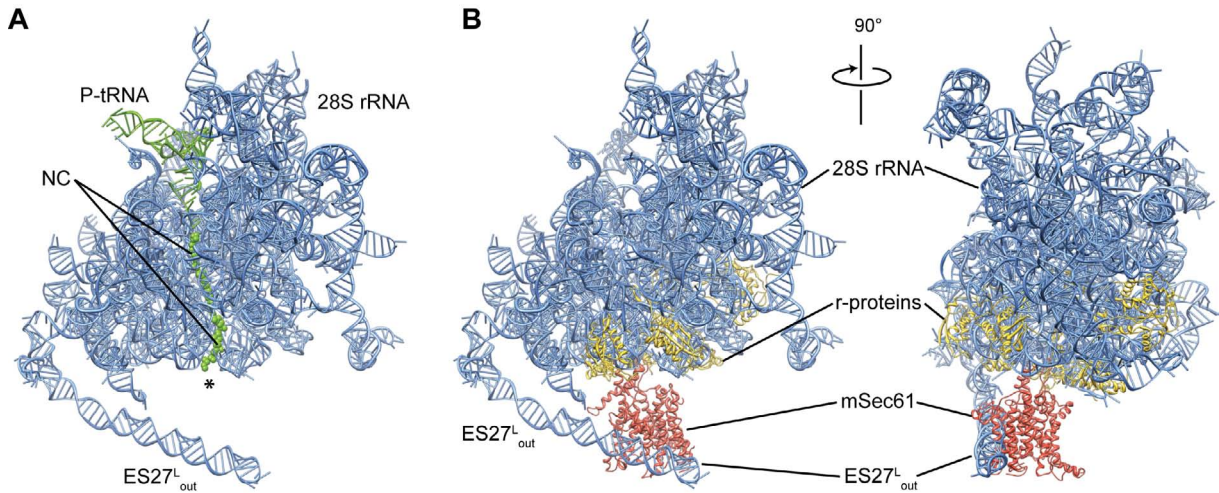


Figure 12.2.3.: **Model of $ES27^L$ in the tunnel exit conformation.** (A) Cross section of the *S. cerevisiae* 50S subunit with the 28S rRNA in blue, the P-tRNA in green and the nascent chain as surface representation in green. Expansion segment $ES27^L$ is shown in the tunnel exit position ($ES27_{out}^L$) and reaches until the tunnel exit site. The asterisk indicates the tunnel exit site. (B) View of the *S. cerevisiae* 50S subunit in overlay with the position of the PCC. The 28S rRNA is colored in blue, the r-proteins of the tunnel exit site in gold and the mSec61 in red. Expansion segment $ES27^L$ is shown in the tunnel exit position ($ES27_{out}^L$) and interferes with the PCC.

It has been also suggested that $ES27^L$ may play a role in coordinating access of non-ribosomal factors, such as chaperones, modifying enzymes or the Signal Recognition Particle (SRP) to the exit site [17], however, with our model we can not deduce any functional role of $ES27^L$. Since $ES27^L$ is essential and the deletion or substitution of $ES27^L$ in the *T. thermophila* 26S rRNA with an unrelated sequence of similar length is lethal [33], a functional role of $ES27^L$ in important cellular processes seems very likely.

12.3. Functional sites on the 80S ribosome

12.3.1. The decoding site and the PTC

Although the rRNA of the PTC and the decoding site are highly conserved, the r-proteins are different.

The decoding site Compared to bacteria, eukaryotes contain two additional r-proteins, S25e and S30e, at the decoding site of the 40S subunit (see section 10.1.1 and Figure 10.1.1 A). The localization of S30e is consistent with crosslinking experiments of S30e to the 4-thiouridine containing UGA stop codon of the mRNA positioned at the A-site of the human 80S ribosome

[212]. The extension of S4p is relocated in eukaryotes. Thus, together with the extensions and loops of the eukaryotic homologous to the bacterial S7, S9, S11, S12, and S13 r-proteins, at least seven different r-proteins can interact and modulate the binding of tRNAs to the 40S subunit [14, 213] (Figure 10.1.1 A).

The PTC The loop of L10e has been observed to interact with the CCA-end of the P-site tRNA (see section 10.1.2 and Figure 10.1.2 A). This loop is disordered and not modeled in the X-ray structure of the *archaeal* 50S subunit [7] and is absent in the bacterial homolog L16. Instead, the N-terminal extension of the r-protein L27 occupies a similar but not distinct position in bacterial ribosomes [214, 215]. The loop of L10e is highly conserved and mutations or deletions in this loop are lethal [216], suggesting that it may play an important role in tRNA positioning, as proposed for the N-terminus of L27 [214, 215].

12.3.2. The eukaryotic ribosomal tunnel

We present here the first molecular model of the eukaryotic ribosomal tunnel and the tunnel exit site. The given resolution of between 5.5 Å and 7.1 Å allowed us to build this model, which enabled us to distinguish between density corresponding to the ribosomal tunnel and the density corresponding to the nascent polypeptide chain. Indeed, we observe that the architecture and molecular environment of the tunnel is highly conserved. Interestingly, we could also visualize almost the entire nascent chain for the yeast DP120 RNC. On basis of this observation, we set out to investigate α -helix formation of the NC, which has been previously suggested by FRET studies, indicating a compaction of a transmembrane sequence as it travels through the tunnel [52]. However, we choose a very hydrophilic sequence with high helical propensity and placed it at different positions at the tunnel. We show that this α -helix can form in the lower region, but not in the central tunnel region (constriction) (Figure 12.3.1 A). This is consistent with biochemical studies that identified specific regions, called folding zones, within the tunnel that promote compaction [54]. The fact that α -helix formation within the tunnel is indeed possible in distinct regions raises the question of what functional significance this observation has. First, one can speculate if the promotion of α -helix formation in the tunnel has an impact on protein folding whereas β -sheet formation is not possible. This would imply that α -helix formation would occur first wherever possible and so result in a hierarchy of secondary structure formation. By assuming this scenario, the complexity of the theoretical conformational space that must have been sampled before in order to adopt the correct conformation, would be significantly reduced. Indeed, the tertiary structure formation, i.e. α - and β -hairpins, has already been observed to occur near the tunnel exit, where the tunnel widens substantially to form a vestibule (Figure 12.3.1 A) [43, 217]. Second, α -helix formation in the tunnel has an influence for proteins containing a membrane insertion domain, and indeed, compaction of transmembrane domains in the ribosomal tunnel has been reported [52, 55]. Co-translational targeting by SRP may be promoted, as (i) the presence of a signal-anchor sequence within the tunnel promotes binding of SRP to the ribosome [218] and (ii) α -helicity of the signal sequence is important for its recognition by SRP (Figure 12.3.1 B) [219]. A compaction of the signal-anchor sequence has been observed by cryo-EM to bind in the vestibule at the end of the ribosomal tunnel on *E. coli* ribosomes [58].

Recently we have shown that secondary structure formation in the tunnel may play a role in regulation of eukaryotic gene expression. Based on our tunnel model we provided the structural basis for studying direct interactions of eukaryotic leader peptides with the ribosomal tunnel and thereby causing translational arrest, i.e. the arginine attenuator peptide (AAP), which stalls in the presence of arginine; or the human cytomegalovirus (hCMV) gp48 upstream open reading frame 2 (uORF2) [49]. Also here, regions of compaction within the stalling sequence and direct interactions between the NC and components of the tunnel (i.e. r-protein L17) could be identified, leading to a rearrangement of the PTC which results in stalling of translation [49]. Taken together, we show that the tunnel provides more than only a passive conduit for nascent chains to pass through the ribosome. Our studies directly show that nascent chains can directly interact with components of the tunnel, presumably via side chain interactions with tunnel components. We showed that A2062 (*E. coli* numbering) interacts with the nascent chain in the upper part of the tunnel. Nucleotide A2062 is known to be flexible, adopting a position flat to the tunnel wall, as in the *H. marismortui* apo-50S X-ray structure [7], or protruding into the tunnel lumen when P-site tRNA mimics are bound [189]. Deeper in the tunnel we observed an interaction between the NC and adenine A751. Insertions at A751 have been reported in the literature to eliminate tryptophan induction associated with TnaC-mediated translational stalling [220]. Additional contacts between the NC and L17 have been observed at the constriction. In *E. coli*, mutations within this region of L22 (L17 is the homolog of L22) can relieve TnaC- and SecM stalling [47, 220]. Moreover, compacted transmembrane containing nascent chain photo-cross-links to L17 [52]. The contacts observed in the two 80S-helix RNCs are non stalling sequences but are nevertheless similar in location to those predicted for some of the known stalling leader peptides. Thus, this region of the tunnel seems to represent a functional hot spot for tunnel-nascent chain interaction. A very important region is the narrow constriction formed by the proteins L4 and L17, which has been previously shown to play an important role for SecM stalling [56, 221]. In the lower region we identified the conserved H50 as a general interaction site for all nascent chains. Chemical properties of nascent chains, like regions for α -helix formation or specific stalling sequences could be probed by the tunnel wall. Therefore, the tunnel most likely plays an active role in protein folding and gene regulation, compared to the previously suggested passive role of the tunnel, which has been described as “ an inert Teflon wall” [8].

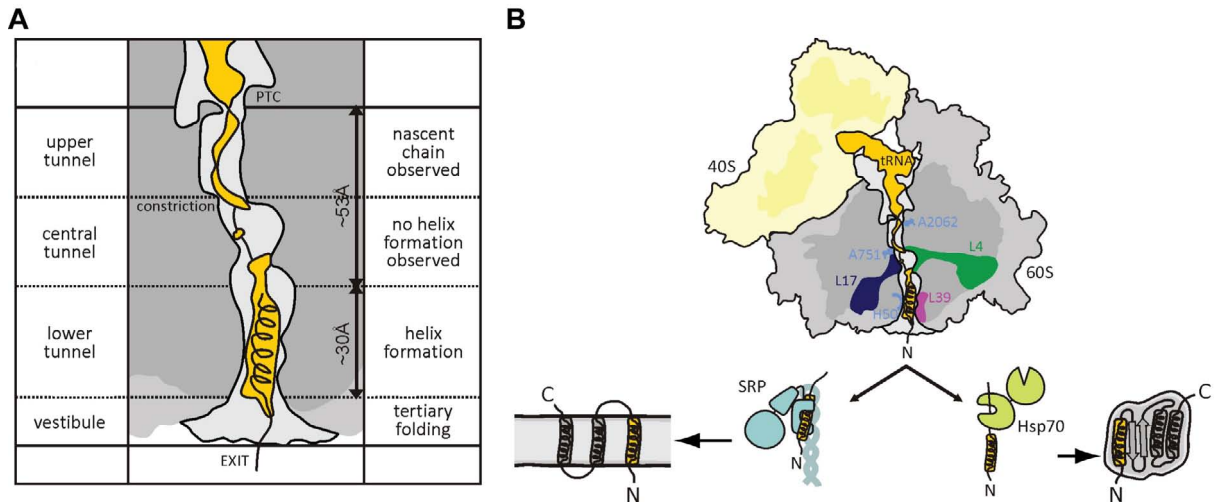


Figure 12.3.1.: **Implications for helix formation within the ribosomal tunnel.** (A) Schematic representation of the cross section of the 80S-helix1 RNC showing the regions where α -helix formation and tertiary structure formation is observed. (B) Schematic representation of the cross section of the 80S-helix1 RNC. The contact sites between the tunnel components and the NC (gold) are indicated: 28S rRNA nucleotides A2062, A751 (blue), r-proteins L4 (green), L17 (dark blue) and L39 (magenta). The helix forms interactions with H50 (blue) and L39. Helix formation may assist membrane insertion, as shown for the SRP-dependent pathway, and/or may promote correct or more efficient folding of cytoplasmic proteins with/without the aid of chaperones like the heat shock protein 70 (HSP70). The figure was adapted from [188].

12.3.3. The eukaryotic tunnel exit site

The tunnel exit site (TE) is the site, where the nascent polypeptide chain emerges from the ribosome and thus has a particular functional relevance. Here, events like co-translational protein folding or protein sorting and translocation take place. The eukaryotic TE is mostly conserved, however, shows some important differences to prokaryotes, especially for the rRNA (H7, H24 and H59). In this work we provide the first molecular model of the eukaryotic tunnel exit site. In eukaryotes H7 interacts with H6 via a new kink-turn motif and is stabilized by the eukaryote-specific r-protein L39e. We identify H7 as being part of the main interaction site of the PCC. In contrast to the bacterial pentaloop of H24, the fungal and plant H24 have a GNRA-tetraloop. Based on molecular dynamics simulations the loop of H24 has been suggested to act like gate, in which the last residue of the H24-pentaloop (A497) is supposed to block the tunnel exit in prokaryotes [222]. In our two models we only observed the GNRA-tetraloop, which is known to fold into a compact and stable structure. We therefore do not observe the adenine A497 acting like a gate, because it is absent in the fungal and plant ribosome (see Figure 11.0.1; Figure 8.1.1). Further, the last adenine residue of the GNRA-tetraloop is not bulged out, but interacts with the guanosine via a tHS interaction. Moreover, the last adenine is stabilized by stacking interactions with surrounding residues. Therefore, the last adenine residue is also not able to act like a gate. In eukaryotes, H59 is an expansion segment, called *ES24^L*. In low-resolution cryo-EM maps of 80S-Sec61-complexes, H59 has been suggested to interact with the PCC [59, 223, 224]. High-resolution cryo-EM structures, however, show that this interaction is established between H59 and the micelle around the PCC, but not the PCC itself [141]. Also in the bacterial system, the

H59-lipid interaction is observed in a cryo-EM reconstruction of the SecYEG-complex embedded in a native membrane environment (nanodisc) bound to a translating 70S ribosome [225]. Here, H59 is supposed to attract lipids and thus establishes another binding and interaction site between the ribosome and a PCC embedded in a nanodisc.

In this work we used the model of the tunnel exit for a detailed description of the PCC-ribosome interaction. Both, mSec61 and Ssh1 interact with the first universal adapter (L23/L29) site involving a newly identified interaction with H7 and L39e. At subnanometer resolution we could identify most of the mSec61 transmembrane α -helices and unambiguously determine the oligomeric state of the PCC as a monomer [141]. This monomer is surrounded by a mixed detergent lipid micelle which makes the complex to appear larger at lower resolution. This might explain the misassignment of the PCC-oligomeric state in previous publications [224, 226].

12.3.4. Canonical translation factor binding site

The canonical translation factor binding site is an important interaction platform for initiation, elongation and termination factors. This site is mainly conserved between eukaryotes and prokaryotes, however detailed interactions between translation factors and the ribosome are mainly described for the prokaryotic system (see section 1.2.4). We determined the structure of the ribosome bound Dom34/Hbs1 complex, which is a complex homologous to the eukaryotic release factors eRF1/eRF3 [181]. As expected for a translational GTPase, Hbs1 shows the same pattern of interactions with the ribosome as other GTPases, involving the E-loop motif of the SRL and the stalk base. This is similar to eEF2 [67] and EF-G [66] indicating a conserved mode of binding of translational GTPases in prokaryotes and eukaryotes. As for eEF2, we also observe an inward movement of the stalk base with Dom34/Hbs1. We modeled the stalk base in both conformations; one observed in the translating *T. aestivum* 80S ribosome reconstruction, and the other conformation observed with an 20° inward rotation relative to H42 in the *S. cerevisiae* 80S ribosome in complex with the Dom34/Hbs1 complex (see section 10.4, and Figure 10.4.2).

12.4. Co-evolution

The eukaryotic 80S ribosome is significantly larger than its bacterial counterpart. The additional mass corresponds to the RNA expansion segments, the N- and C-terminal extensions of homologous r-proteins and eukaryote-specific r-proteins, which form an intricate layer of additional RNA-protein mass. This additional layer is predominantly located on the solvent surface of the ribosome (Figure 12.4.1 A; green and gold parts). The eukaryote-specific RNA expansion segments can be seen as elements that have been gradually added to the evolving ribosome structure (see section 1.3.1 and [81]). The additional ES are intertwined with eukaryote-specific r-proteins and protein extensions and support the idea that they are co-evolving together [227]. For example the large additional mass found on the back of the 60S subunit comprises of *ES7^L*, *ES39^L* and five eukaryote-specific r-proteins L6e, L14e, L18ae, L28e and L35ae (Figure 12.4.1 B). L27e is sandwiched between H55, which is the adjacent helix of *ES20^L/ES26^L*, and H58 which adopts a slightly different conformation in eukaryotes. L27e and L34e overlap with the position of H58 in the *E. coli* ribosome and emphasizes the rearrangement of H58 in eukaryotes.

Interestingly, the loop of H57 which interacts with the eukaryote-specific r-protein L22e is longer and conserved in eukaryotes, but not in *bacteria*, which lack this r-protein.

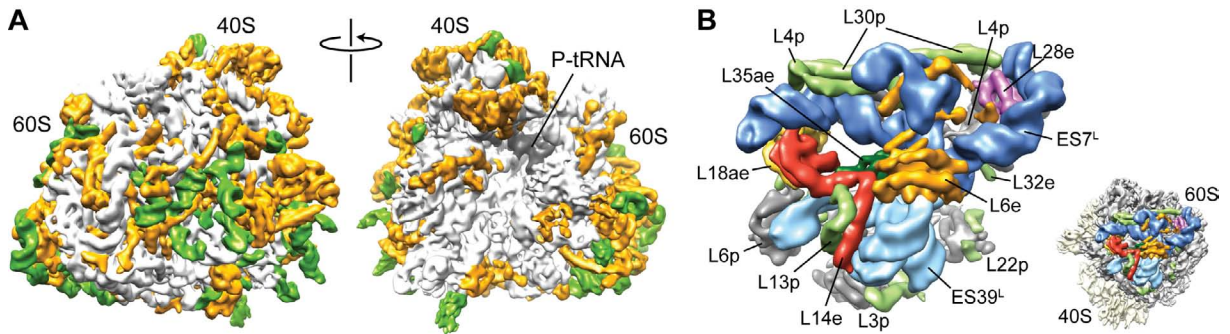


Figure 12.4.1.: **Co-evolution of rRNA expansion segments and r-proteins in the 80S ribosome.** (A) Cryo-EM map of the *T. aestivum* 80S ribosome, with rRNA ES and VR colored green and eukaryote-specific r-proteins and protein extensions colored gold. (B) View of the highly intertwined region of $ES7^L$ (dark blue) and $ES39^L$ (light blue), with core r-proteins (gray), eukaryote-specific r-proteins (L6e, orange; L14e, red; L18ae, yellow; L28e, magenta; L35ae, dark green) and r-protein extensions (pale green) highlighted. The figure was taken from [185].

12.5. Evolution of ribosomes

Comparing the genomic sequences from diverse organisms, ranging from *bacteria* to *mammals*, indicates that the additional mass increases with the complexity of the organism (Figure 12.5.1 A-D). However, the composition of mammalian ribosomes, such as from human, is surprisingly similar to those of other eukaryotes, such as yeast and plants described in this work. Human 80S ribosomes share the same 80 r-proteins that are found in *T. aestivum* ribosomes. Regarding the ribosomal RNA, only four ES differ significantly in length, namely $ES7^L$, $ES15^L$, $ES27^L$ and $ES39^L$. These ES are longer in *human* (~ 850 nts, ~ 180 nts, ~ 700 nts and ~ 220 nts) than in *T. aestivum* (~ 200 nts, ~ 20 nts, ~ 150 nts and ~ 120 nts). Cryo-EM reconstructions of mammalian ribosomes ([15, 21, 228, 229]) show that the longer ES are generally more flexible and therefore little to no density is visible (Figure 12.5.1 D). Thus, evolution has favored the development of two apparently distinct layers of mass gain for the ribosome: the first layer of rigidly positioned ES which tightly intertwined with r-proteins on the ribosomal surface and the additional second layer comprising of a few drastically extended and highly flexible expansion segments with hitherto unknown function [182, 185]. One can only speculate about a possible function of the ES. For example, the event of initiation is much more complex in eukaryotes involving more than a dozen initiation factors, which have to be coordinated and orchestrated during the assembly of the pre-initiation complex. It appears likely that also ES play a role here, but so far no experiments have been done to examine this. Another possible role might be that ES provide an interaction platform for regulatory factors associated with the ribosome. In order to find out more about the functional and possible interaction partner, ES could be expressed and immobilized in order to perform pullouts with whole cell extracts.

12. Molecular model of the 80S ribosome

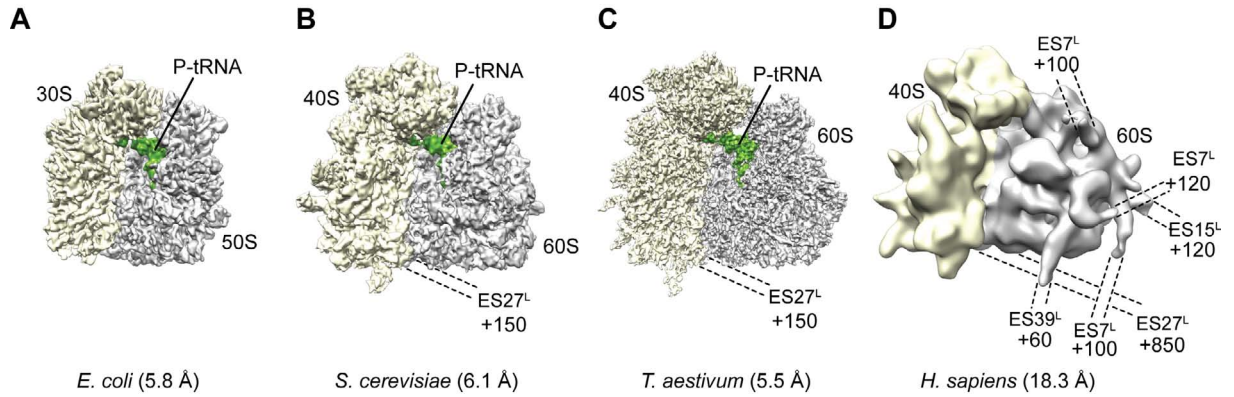


Figure 12.5.1.: **Cryo-EM reconstructions of ribosomes** of (A) the eubacterium *E. coli* [230] (B) the fungus *S. cerevisiae* [141], (C) the wheat germ *T. aestivum* (this work), and (D) *Homo sapiens* [229]. The figure was taken from [182].

This second layer of drastically extended ES is in good agreement with the hypothesis by Bokov & Steinberg [81]. These ES emerge from only the last three layers that have been added to the 23S rRNA (see also Figure 1.3.1 A, B).

13. Summary

In this dissertation we present a new approach for modeling and fitting large structured RNAs into cryo-EM density maps. Our homology modeling approach led to highly accurate models of the common ribosomal core structure. Combined with cryo-EM and single particle analysis we were able to build *de novo* molecular models for the eukaryotic RNA expansion segments. The models are very similar compared to the recently published crystal structure at 4.5 Å resolution (see section 12.1.3, see Figure 12.1.2), however, structural differences are evident (see section 12.2.2), which were expected at our resolution of 5.5 Å. Some of our models, like the highly flexible *ES7^L* and *ES27^L*, are even more complete than the recent crystal structure, but need to be further validated by future high-resolution crystal structures or cryo-EM maps. Taken together, this approach of *de novo* and homology RNA modeling combined with cryo-EM data is extremely valid for further modeling, for example the few drastically extended and highly flexible expansion segments in *Drosophila melanogaster* or *Homo sapiens*, which certainly will not be crystallized in the near future. Moreover, the modeling approach can be applied to other large RNA complexes.

Based on our cryo-EM reconstruction at 5.5 Å and 6.1 Å resolution, respectively, we provide models for *T. aestivum* and *S. cerevisiae* 80S ribosomes. Based on our RNA homology model of the common ribosomal core we could assign the structural similarities between *bacteria* and eukaryotes. In addition we modeled all the eukaryote-specific RNA expansion segments and variable regions leading to a ~98% completeness of the rRNA. Together with the 27 eukaryote-specific r-proteins that we localized, we provide the most complete model of a eukaryotic 80S ribosome which has been published to date (see Figure 9.2.1). This work provides a structural inventory for all eukaryote-specific RNA expansion segments and variable regions. The analysis of the expansion segments and variable regions in *T. aestivum* and *S. cerevisiae* revealed a very high similarity between both species, except for *ES7^L*, *ES27^L* and *ES39^L*.

The model was used for the molecular interpretation of the functionally important regions of the eukaryotic 80S ribosome and translation factors (see sections 10.2, 10.3 and 10.4). Moreover, the models were used to interpret the movement of the highly flexible expansion segments.

Bibliography

- [1] T. M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation., *Nature* 461 (7268) (2009) 1234–1242. doi:10.1038/nature08403.
URL <http://dx.doi.org/10.1038/nature08403>
- [2] P. B. Moore, T. A. Steitz, The structural basis of large ribosomal subunit function., *Annu Rev Biochem* 72 (2003) 813–850. doi:10.1146/annurev.biochem.72.110601.135450.
URL <http://dx.doi.org/10.1146/annurev.biochem.72.110601.135450>
- [3] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vornrhein, T. Hartsch, V. Ramakrishnan, Structure of the 30S ribosomal subunit., *Nature* 407 (6802) (2000) 327–339. doi:10.1038/35030006.
URL <http://dx.doi.org/10.1038/35030006>
- [4] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath, Structure of functionally activated small ribosomal subunit at 3.3 Å resolution., *Cell* 102 (5) (2000) 615–623.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11007480>
- [5] D. E. Brodersen, W. M. Clemons, A. P. Carter, B. T. Wimberly, V. Ramakrishnan, Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: Structure of the proteins and their interactions with 16S RNA., *J Mol Biol* 316 (3) (2002) 725–768. doi:10.1006/jmbi.2001.5359.
URL <http://dx.doi.org/10.1006/jmbi.2001.5359>
- [6] J. M. Ogle, A. P. Carter, V. Ramakrishnan, Insights into the decoding mechanism from recent ribosome structures., *Trends in biochemical sciences* 28 (5) (2003) 259–266.
URL <http://view.ncbi.nlm.nih.gov/pubmed/12765838>
- [7] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* 289 (5481) (2000) 905–920. doi:10.1126/science.289.5481.905.
URL <http://dx.doi.org/10.1126/science.289.5481.905>
- [8] P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis, *Science* 289 (5481) (2000) 920–930. doi:10.1126/science.289.5481.920.
URL <http://dx.doi.org/10.1126/science.289.5481.920>

Bibliography

- [9] D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz, The kink-turn: a new RNA secondary structure motif., *EMBO J* 20 (15) (2001) 4214–4221.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11483524>
- [10] P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, RNA tertiary interactions in the large ribosomal subunit: the A-minor motif., *Proc Natl Acad Sci U S A* 98 (9) (2001) 4899–4903. doi:10.1073/pnas.081082398.
URL <http://dx.doi.org/10.1073/pnas.081082398>
- [11] D. J. Klein, P. B. Moore, T. A. Steitz, The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit., *J Mol Biol* 340 (1) (2004) 141–177. doi:10.1016/j.jmb.2004.03.076.
URL <http://dx.doi.org/10.1016/j.jmb.2004.03.076>
- [12] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. D. Cate, H. F. Noller, Crystal structure of the ribosome at 5.5 Å resolution, *Science* 292 (5518) (2001) 883–896. doi:10.1126/science.1060089.
URL <http://dx.doi.org/10.1126/science.1060089>
- [13] B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, J. H. Cate, Structures of the bacterial ribosome at 3.5 Å resolution., *Science (New York, N.Y.)* 310 (5749) (2005) 827–834. doi:10.1126/science.1117230.
URL <http://dx.doi.org/10.1126/science.1117230>
- [14] M. Selmer, C. M. Dunham, F. V. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, V. Ramakrishnan, Structure of the 70S ribosome complexed with mRNA and tRNA., *Science (New York, N.Y.)* 313 (5795) (2006) 1935–1942. doi:10.1126/science.1131127.
URL <http://dx.doi.org/10.1126/science.1131127>
- [15] P. Dube, M. Wieske, H. Stark, M. Schatz, J. Stahl, F. Zemlin, G. Lutsch, M. van Heel, The 80S rat liver ribosome at 25 Å resolution by electron cryomicroscopy and angular reconstitution., *Structure* 6 (3) (1998) 389–399.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9551559>
- [16] C. M. T. Spahn, R. Beckmann, N. Eswar, P. A. Penczek, A. Sali, G. Blobel, J. Frank, Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions, *Cell* 107 (3) (2001) 373–386. doi:10.1016/S0092-8674(01)00539-6.
URL [http://dx.doi.org/10.1016/S0092-8674\(01\)00539-6](http://dx.doi.org/10.1016/S0092-8674(01)00539-6)
- [17] R. Beckmann, C. M. Spahn, N. Eswar, J. Helmers, P. A. Penczek, A. Sali, J. Frank, G. Blobel, Architecture of the protein-conducting channel associated with the translating 80S ribosome, *Cell* 107 (3) (2001) 361–372. doi:10.1016/S0092-8674(01)00541-4.
URL [http://dx.doi.org/10.1016/S0092-8674\(01\)00541-4](http://dx.doi.org/10.1016/S0092-8674(01)00541-4)
- [18] J. Nilsson, J. Sengupta, R. Gursky, P. Nissen, J. Frank, Comparison of fungal 80S ribosomes by cryo-EM reveals diversity in structure and conformation of rRNA expansion segments.,

Bibliography

- J Mol Biol 369 (2) (2007) 429–438. doi:10.1016/j.jmb.2007.03.035.
URL <http://dx.doi.org/10.1016/j.jmb.2007.03.035>
- [19] A. Ben-Shem, L. Jenner, G. Yusupova, M. Yusupov, Crystal structure of the eukaryotic ribosome., Science (New York, N.Y.) 330 (6008) (2010) 1203–1209. doi:10.1126/science.1194294.
URL <http://dx.doi.org/10.1126/science.1194294>
- [20] J. Rabl, M. Leibundgut, S. F. Ataide, A. Haag, N. Ban, Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1., Science (New York, N.Y.) 331 (6018) (2011) 730–736. doi:10.1126/science.1198308.
URL <http://dx.doi.org/10.1126/science.1198308>
- [21] P. Chandramouli, M. Topf, J.-F. Ménétrete, N. Eswar, J. J. Cannone, R. R. Gutell, A. Sali, C. W. Akey, Structure of the mammalian 80S ribosome at 8.7 Å resolution, Structure 16 (4) (2008) 535–548. doi:10.1016/j.str.2008.01.007.
URL <http://dx.doi.org/10.1016/j.str.2008.01.007>
- [22] D. J. Taylor, B. Devkota, A. D. Huang, M. Topf, E. Narayanan, A. Sali, S. C. Harvey, J. Frank, Comprehensive Molecular Structure of the Eukaryotic Ribosome, Structure 17 (12) (2009) 1591–1604.
URL [http://www.cell.com/structure/abstract/S0969-2126\(09\)00416-X](http://www.cell.com/structure/abstract/S0969-2126(09)00416-X)
- [23] S. A. Gerbi, Expansion segments: Regions of variable size that interrupt the universal core secondary structure of ribosomal RNA , CRC Press, 1996, Ch. 4, pp. 71–87.
- [24] C. G. Clark, B. W. Tague, V. C. Ware, S. A. Gerbi, *Xenopus laevis* 28S ribosomal RNA : A secondary structure model and its evolutionary and functional implications., Nucleic acids research 12 (15) (1984) 6197–6220.
URL <http://view.ncbi.nlm.nih.gov/pubmed/6147812>
- [25] J. M. Hancock, G. A. Dover, Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs., Mol Biol Evol 5 (4) (1988) 377–391.
URL <http://view.ncbi.nlm.nih.gov/pubmed/3405077>
- [26] I. L. Gonzalez, C. Chambers, J. L. Gorski, D. Stambolian, R. D. Schmickel, J. E. Sylvester, Sequence and structure correlation of human ribosomal transcribed spacers., Journal of molecular biology 212 (1) (1990) 27–35. doi:10.1016/0022-2836(90)90302-3.
URL [http://dx.doi.org/10.1016/0022-2836\(90\)90302-3](http://dx.doi.org/10.1016/0022-2836(90)90302-3)
- [27] J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D’Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Müller, N. Pande, Z. Shang, N. Yu, R. R. Gutell, The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs., BMC bioinformatics 3 (1).
URL <http://view.ncbi.nlm.nih.gov/pubmed/11869452>
- [28] G. Alkemar, O. D. D. Nygård, A possible tertiary rRNA interaction between expansion segments ES3 and ES6 in eukaryotic 40S ribosomal subunits, RNA 9 (1) (2003) 20–24.

Bibliography

- doi:10.1261/rna.2108203.
URL <http://dx.doi.org/10.1261/rna.2108203>
- [29] S. L. Larsson, O. Nygård, Proposed secondary structure of eukaryote specific expansion segment 15 in 28S rRNA from mice, rats, and rabbits., *Biochemistry*. 40 (10) (2001) 3222–3231.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11258939>
- [30] O. Nygård, G. Alkemar, S. L. Larsson, Analysis of the secondary structure of expansion segment 39 in ribosomes from fungi, plants and mammals, *Journal of Molecular Biology* 357 (3) (2006) 904–916. doi:10.1016/j.jmb.2006.01.043.
URL <http://dx.doi.org/10.1016/j.jmb.2006.01.043>
- [31] C. G. Clark, On the evolution of ribosomal RNA ., *Journal of molecular evolution* 25 (4) (1987) 343–350.
URL <http://view.ncbi.nlm.nih.gov/pubmed/3118048>
- [32] R. Sweeney, M. C. Yao, Identifying functional regions of rRNA by insertion mutagenesis and complete gene replacement in *Tetrahymena thermophila*., *The EMBO journal* 8 (3) (1989) 933–938.
URL <http://view.ncbi.nlm.nih.gov/pubmed/2542027>
- [33] R. Sweeney, L. Chen, M. C. Yao, An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence., *Molecular and cellular biology* 14 (6) (1994) 4203–4215.
URL <http://view.ncbi.nlm.nih.gov/pubmed/8196658>
- [34] R. E. Jeeninga, Y. Van Delft, M. de Graaff-Vincent, A. Dirks-Mulder, J. Venema, H. A. Raué, Variable regions V13 and V3 of *Saccharomyces cerevisiae* contain structural features essential for normal biogenesis and stability of 5.8S and 25S rRNA., *RNA (New York, N.Y.)* 3 (5) (1997) 476–488.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9149229>
- [35] H. F. Noller, RNA structure: Reading the Ribosome, *Science* 309 (5740) (2005) 1508–1514. doi:10.1126/science.1111771.
URL <http://dx.doi.org/10.1126/science.1111771>
- [36] E. R. Dabbs, Selection for *Escherichia coli* mutants with proteins missing from the ribosome., *Journal of bacteriology* 140 (2) (1979) 734–737.
URL <http://view.ncbi.nlm.nih.gov/pubmed/387748>
- [37] O. Lecompte, R. Ripp, J.-C. C. Thierry, D. Moras, O. Poch, Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale., *Nucleic acids research* 30 (24) (2002) 5382–5390.
URL <http://nar.oxfordjournals.org/cgi/content/abstract/30/24/5382>
- [38] J. M. Ogle, D. E. Brodersen, W. M. Clemons, M. J. Tarry, A. P. Carter, V. Ramakrishnan, Recognition of cognate transfer RNA by the 30S ribosomal subunit., *Science (New York,*

Bibliography

- N.Y.) 292 (5518) (2001) 897–902. doi:10.1126/science.1060612.
URL <http://dx.doi.org/10.1126/science.1060612>
- [39] J. M. Ogle, F. V. Murphy, M. J. Tarry, V. Ramakrishnan, Selection of tRNA by the ribosome requires a transition from an open to a closed form., *Cell* 111 (5) (2002) 721–732.
URL <http://view.ncbi.nlm.nih.gov/pubmed/12464183>
- [40] T. M. Schmeing, R. M. Voorhees, A. C. Kelley, Y.-G. Gao, F. V. Murphy, J. R. Weir, V. Ramakrishnan, The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA., *Science (New York, N.Y.)* 326 (5953) (2009) 688–694. doi:10.1126/science.1179700.
URL <http://dx.doi.org/10.1126/science.1179700>
- [41] T. A. Steitz, P. B. Moore, RNA, the first macromolecular catalyst: The ribosome is a ribozyme, *Trends in Biochemical Sciences* 28 (8) (2003) 411–418. doi:10.1016/S0968-0004(03)00169-5.
URL [http://dx.doi.org/10.1016/S0968-0004\(03\)00169-5](http://dx.doi.org/10.1016/S0968-0004(03)00169-5)
- [42] W. D. Picking, W. L. Picking, O. W. Odom, B. Hardesty, Fluorescence characterization of the environment encountered by nascent polyalanine and polyserine as they exit *Escherichia coli* ribosomes during translation., *Biochemistry* 31 (8) (1992) 2368–2375.
URL <http://view.ncbi.nlm.nih.gov/pubmed/1540593>
- [43] J. Lu, C. Deutsch, Electrostatics in the ribosomal tunnel modulate chain elongation rates, *Journal of Molecular Biology* 384 (1) (2008) 73–86. doi:10.1016/j.jmb.2008.08.089.
URL <http://dx.doi.org/10.1016/j.jmb.2008.08.089>
- [44] J. Lu, W. R. Kobertz, C. Deutsch, Mapping the electrostatic potential within the ribosomal exit tunnel., *Journal of molecular biology* 371 (5) (2007) 1378–1391. doi:10.1016/j.jmb.2007.06.038.
URL <http://dx.doi.org/10.1016/j.jmb.2007.06.038>
- [45] T. Tenson, M. Ehrenberg, Regulatory nascent peptides in the ribosomal tunnel., *Cell* 108 (5) (2002) 591–594.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11893330>
- [46] F. Gong, C. Yanofsky, Instruction of translating ribosome by nascent peptide., *Science (New York, N.Y.)* 297 (5588) (2002) 1864–1867. doi:10.1126/science.1073997.
URL <http://dx.doi.org/10.1126/science.1073997>
- [47] H. Nakatogawa, K. Ito, The ribosomal exit tunnel functions as a discriminating gate., *Cell* 108 (5) (2002) 629–636.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11893334>
- [48] N. Vazquez-Laslop, C. Thum, A. S. Mankin, Molecular mechanism of drug-dependent ribosome stalling., *Molecular cell* 30 (2) (2008) 190–202. doi:10.1016/j.molcel.2008.02.026.
URL <http://dx.doi.org/10.1016/j.molcel.2008.02.026>

Bibliography

- [49] S. Bhushan, H. Meyer, A. L. Starosta, T. Becker, T. Mielke, O. Berninghausen, M. Sattler, D. N. Wilson, R. Beckmann, Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide., *Molecular cell* 40 (1) (2010) 138–146. doi:10.1016/j.molcel.2010.09.009.
URL <http://dx.doi.org/10.1016/j.molcel.2010.09.009>
- [50] N. R. Voss, M. Gerstein, T. A. Steitz, P. B. Moore, The geometry of the ribosomal polypeptide exit tunnel., *Journal of molecular biology* 360 (4) (2006) 893–906. doi:10.1016/j.jmb.2006.05.023.
URL <http://dx.doi.org/10.1016/j.jmb.2006.05.023>
- [51] B. Hardesty, G. Kramer, Folding of a nascent peptide on the ribosome., *Progress in nucleic acid research and molecular biology* 66 (2001) 41–66.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11051761>
- [52] C. A. Woolhead, P. J. McCormick, A. E. Johnson, Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins., *Cell* 116 (5) (2004) 725–736.
URL <http://view.ncbi.nlm.nih.gov/pubmed/15006354>
- [53] A. Kosolapov, L. Tu, J. Wang, C. Deutsch, Structure acquisition of the T1 domain of Kv1.3 during biogenesis., *Neuron* 44 (2) (2004) 295–307. doi:10.1016/j.neuron.2004.09.011.
URL <http://dx.doi.org/10.1016/j.neuron.2004.09.011>
- [54] J. Lu, C. Deutsch, Folding zones inside the ribosomal exit tunnel, *Nature Structural & Molecular Biology* 12 (12) (2005) 1123–1129. doi:10.1038/nsmb1021.
URL <http://dx.doi.org/10.1038/nsmb1021>
- [55] J. Lu, C. Deutsch, Secondary structure formation of a transmembrane segment in Kv channels., *Biochemistry* 44 (23) (2005) 8230–8243. doi:10.1021/bi050372q.
URL <http://dx.doi.org/10.1021/bi050372q>
- [56] C. A. Woolhead, A. E. Johnson, H. D. Bernstein, Translation arrest requires two-way communication between a nascent polypeptide and the ribosome., *Mol Cell* 22 (5) (2006) 587–598. doi:10.1016/j.molcel.2006.05.021.
URL <http://dx.doi.org/10.1016/j.molcel.2006.05.021>
- [57] K. Wild, M. Halic, I. Sinning, R. Beckmann, SRP meets the ribosome, *Nature Structural & Molecular Biology* 11 (11) (2004) 1049+. doi:10.1038/nsmb853.
URL <http://dx.doi.org/10.1038/nsmb853>
- [58] M. Halic, M. Blau, T. Becker, T. Mielke, I. Sinning, R. Beckmann, Following the signal sequence from ribosomal tunnel exit to signal recognition particle, *nature*.
- [59] R. Beckmann, C. M. Spahn, N. Eswar, J. Helmers, P. A. Penczek, A. Sali, J. Frank, G. Blobel, Architecture of the protein-conducting channel associated with the translating 80S ribosome., *Cell* 107 (3) (2001) 361–372.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11701126>

Bibliography

- [60] J.-F. Menetret, J. Schaletzky, Clemons, A. R. Osborne, S. S. Skanland, C. Denison, S. P. Gygi, D. S. Kirkpatrick, E. Park, S. J. Ludtke, T. A. Rapoport, C. W. Akey, Ribosome binding of a single copy of the secY complex: Implications for protein translocation, *Molecular Cell* 28 (6) (2007) 1083–1092. doi:10.1016/j.molcel.2007.10.034.
URL <http://dx.doi.org/10.1016/j.molcel.2007.10.034>
- [61] R. Kohler, D. Boehringer, B. Greber, R. Bingel-Erlenmeyer, I. Collinson, C. Schaffitzel, N. Ban, YidC and Oxa1 form dimeric insertion pores on the translating ribosome., *Molecular cell* 34 (3) (2009) 344–353. doi:10.1016/j.molcel.2009.04.019.
URL <http://dx.doi.org/10.1016/j.molcel.2009.04.019>
- [62] H. Patzelt, S. Rüdiger, D. Brehmer, G. Kramer, S. Vorderwülbecke, E. Schaffitzel, A. Waitz, T. Hesterkamp, L. Dong, J. Schneider-Mergener, B. Bukau, E. Deuring, Binding specificity of *Escherichia coli* trigger factor., *Proceedings of the National Academy of Sciences of the United States of America* 98 (25) (2001) 14244–14249. doi:10.1073/pnas.261432298.
URL <http://dx.doi.org/10.1073/pnas.261432298>
- [63] K. Peisker, D. Braun, T. Wölflé, J. Hentschel, U. Fünfschilling, G. Fischer, A. Sickmann, S. Rospert, Ribosome-associated complex binds to ribosomes in close proximity of Rpl31 at the exit of the polypeptide tunnel in yeast., *Molecular biology of the cell* 19 (12) (2008) 5279–5288. doi:10.1091/mbc.E08-06-0661.
URL <http://dx.doi.org/10.1091/mbc.E08-06-0661>
- [64] M. Pech, T. Spreter, R. Beckmann, B. Beatrix, Dual binding mode of the nascent polypeptide-associated complex reveals a novel universal adapter site on the ribosome., *The Journal of biological chemistry* 285 (25) (2010) 19679–19687. doi:10.1074/jbc.M109.092536.
URL <http://dx.doi.org/10.1074/jbc.M109.092536>
- [65] M. Halic, M. Gartmann, O. Schlenker, T. Mielke, M. R. Pool, I. Sinning, R. Beckmann, Signal recognition particle receptor exposes the ribosomal translocon binding site, *Science* 312 (5774) (2006) 745–747. doi:10.1126/science.1124864.
URL <http://dx.doi.org/10.1126/science.1124864>
- [66] Y.-G. Gao, M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley, V. Ramakrishnan, The structure of the ribosome with Elongation Factor G trapped in the posttranslocational state, *Science* 326 (5953) (2009) 694–699. doi:10.1126/science.1179709.
URL <http://dx.doi.org/10.1126/science.1179709>
- [67] C. M. Spahn, M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jørgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta, J. Frank, Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation., *The EMBO journal* 23 (5) (2004) 1008–1019. doi:10.1038/sj.emboj.7600102.
URL <http://dx.doi.org/10.1038/sj.emboj.7600102>

Bibliography

- [68] U. Rawat, H. Gao, A. Zavialov, R. Gursky, M. Ehrenberg, J. Frank, Interactions of the release factor RF1 with the ribosome as revealed by cryo-EM., *Journal of molecular biology* 357 (4) (2006) 1144–1153. doi:10.1016/j.jmb.2006.01.038.
URL <http://dx.doi.org/10.1016/j.jmb.2006.01.038>
- [69] M. Laurberg, H. Asahara, A. Korostelev, J. Zhu, S. Trakhanov, H. F. Noller, Structural basis for translation termination on the 70S ribosome, *Nature* 454 (7206) (2008) 852–857. doi:10.1038/nature07115.
URL <http://dx.doi.org/10.1038/nature07115>
- [70] U. B. S. Rawat, A. V. Zavialov, J. Sengupta, M. Valle, R. A. Grassucci, J. Linde, B. Vestergaard, M. Ehrenberg, J. Frank, A cryo-electron microscopic study of ribosome-bound termination factor RF2, *Nature* 421 (6918) (2003) 87–90. doi:10.1038/nature01224.
URL <http://dx.doi.org/10.1038/nature01224>
- [71] B. P. Klaholz, T. Pape, A. V. Zavialov, A. G. Myasnikov, E. V. Orlova, B. Vestergaard, M. Ehrenberg, M. van Heel, Structure of the *Escherichia coli* ribosomal termination complex with release factor 2., *Nature* 421 (6918) (2003) 90–94. doi:10.1038/nature01225.
URL <http://dx.doi.org/10.1038/nature01225>
- [72] A. Korostelev, H. Asahara, L. Lancaster, M. Laurberg, A. Hirschi, J. Zhu, S. Trakhanov, W. G. Scott, H. F. Noller, Crystal structure of a translation termination complex formed with release factor RF2., *Proceedings of the National Academy of Sciences of the United States of America* 105 (50) (2008) 19684–19689. doi:10.1073/pnas.0810953105.
URL <http://dx.doi.org/10.1073/pnas.0810953105>
- [73] A. Weixlbaumer, H. Jin, C. Neubauer, R. M. Voorhees, S. Petry, A. C. Kelley, V. Ramakrishnan, Insights into translational termination from the structure of RF2 bound to the ribosome., *Science (New York, N.Y.)* 322 (5903) (2008) 953–956. doi:10.1126/science.1164840.
URL <http://dx.doi.org/10.1126/science.1164840>
- [74] H. Gao, Z. Zhou, U. Rawat, C. Huang, L. Bouakaz, C. Wang, Z. Cheng, Y. Liu, A. Zavialov, R. Gursky, S. Sanyal, M. Ehrenberg, J. Frank, H. Song, RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors., *Cell* 129 (5) (2007) 929–941. doi:10.1016/j.cell.2007.03.050.
URL <http://dx.doi.org/10.1016/j.cell.2007.03.050>
- [75] A. Weixlbaumer, S. Petry, C. M. Dunham, M. Selmer, A. C. Kelley, V. Ramakrishnan, Crystal structure of the ribosome recycling factor bound to the ribosome, *Nature Structural & Molecular Biology* 14 (8) (2007) 733–737. doi:10.1038/nsmb1282.
URL <http://dx.doi.org/10.1038/nsmb1282>
- [76] R. K. Agrawal, M. R. Sharma, M. C. Kiel, G. Hirokawa, T. M. Booth, C. M. Spahn, R. A. Grassucci, A. Kaji, J. Frank, Visualization of ribosome-recycling factor on the *Escherichia*

Bibliography

- coli* 70S ribosome: Functional implications, Proceedings of the National Academy of Sciences 101 (24) (2004) 8900–8905. doi:10.1073/pnas.0401904101.
URL <http://dx.doi.org/10.1073/pnas.0401904101>
- [77] D. N. Wilson, F. Schluenzen, J. M. Harms, T. Yoshida, T. Ohkubo, R. Albrecht, J. Buerger, Y. Kobayashi, P. Fucini, X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit, The EMBO Journal aop (current). doi:10.1038/sj.emboj.7600525.
URL <http://dx.doi.org/10.1038/sj.emboj.7600525>
- [78] W. Gilbert, Origin of life: The RNA world, Nature 319 (6055) (1986) 618. doi:10.1038/319618a0.
URL <http://dx.doi.org/10.1038/319618a0>
- [79] H. F. Noller, The driving force for molecular evolution of translation., RNA (New York, N.Y.) 10 (12) (2004) 1833–1837. doi:10.1261/rna.7142404.
URL <http://dx.doi.org/10.1261/rna.7142404>
- [80] J. Doudna, Structure and function of the eukaryotic ribosome: The next frontier, Cell 109 (2) (2002) 153–156. doi:10.1016/S0092-8674(02)00725-0.
URL [http://dx.doi.org/10.1016/S0092-8674\(02\)00725-0](http://dx.doi.org/10.1016/S0092-8674(02)00725-0)
- [81] K. Bokov, S. V. Steinberg, A hierarchical model for evolution of 23S ribosomal RNA, Nature 457 (7232) (2009) 977–980. doi:10.1038/nature07749.
URL <http://dx.doi.org/10.1038/nature07749>
- [82] I. Agmon, A. Bashan, R. Zarivach, A. Yonath, Symmetry at the active site of the ribosome: structural and functional implications., Biological chemistry 386 (9) (2005) 833–844. doi:10.1515/BC.2005.098.
URL <http://dx.doi.org/10.1515/BC.2005.098>
- [83] B. Zhang, T. R. Cech, Peptide bond formation by in vitro selected ribozymes., Nature 390 (6655) (1997) 96–100. doi:10.1038/36375.
URL <http://dx.doi.org/10.1038/36375>
- [84] E. Westhof, V. Fritsch, RNA folding: beyond Watson-Crick pairs., Structure 8 (3).
URL <http://view.ncbi.nlm.nih.gov/pubmed/10745012>
- [85] N. B. Leontis, J. Stombaugh, E. Westhof, The non-Watson-Crick base pairs and their associated isostericity matrices., Nucleic Acids Res 30 (16) (2002) 3497–3531.
URL <http://view.ncbi.nlm.nih.gov/pubmed/12177293>
- [86] Stombaugh, Jesse, Zirbel, Craig L., Westhof, Eric, Leontis, Neocles B., Frequency and isostericity of RNA base pairs., Nucleic acids research 37 (7) (2009) 2294–2312. doi:10.1093/nar/gkp011.
URL <http://dx.doi.org/10.1093/nar/gkp011>
- [87] W. Saenger, Principles of Nucleic acid structure, Springer-Verlag, 1983.

Bibliography

- [88] Y. Tanaka, S. Fujii, H. Hiroaki, T. Sakata, T. Tanaka, S. Uesugi, K.-i. Tomita, Y. Kyogoku, A'-form RNA double helix in the single crystal structure of r(UGAGCUUCGGCUC), *Nucleic Acids Research* 27 (4) (1999) 949–955. doi:10.1093/nar/27.4.949.
URL <http://dx.doi.org/10.1093/nar/27.4.949>
- [89] M. Popena, J. Milecki, R. W. Adamiak, High salt solution structure of a left-handed RNA double helix., *Nucleic acids research* 32 (13) (2004) 4044–4054. doi:10.1093/nar/gkh736.
URL <http://dx.doi.org/10.1093/nar/gkh736>
- [90] M. Vallazza, M. Perbandt, S. Klussmann, W. Rypniewski, H. M. Einspahr, V. A. Erdmann, Ch, First look at RNA in L-configuration., *Acta crystallographica. Section D, Biological crystallography* 60 (Pt 1) (2004) 1–7.
URL <http://view.ncbi.nlm.nih.gov/pubmed/14684885>
- [91] P. S. Klosterman, M. Tamura, S. R. Holbrook, S. E. Brenner, SCOR: a Structural Classification of RNA database., *Nucleic Acids Res* 30 (1) (2002) 392–394. doi:10.1093/nar/30.1.392.
URL <http://dx.doi.org/10.1093/nar/30.1.392>
- [92] D. K. Hendrix, S. E. Brenner, S. R. Holbrook, RNA structural motifs: building blocks of a modular biomolecule., *Q Rev Biophys* 38 (3) (2005) 221–243. doi:10.1017/S0033583506004215.
URL <http://dx.doi.org/10.1017/S0033583506004215>
- [93] D. E. Draper, A guide to ions and RNA structure, *RNA* 10 (3) (2004) 335–343. doi:10.1261/rna.5205404.
URL <http://dx.doi.org/10.1261/rna.5205404>
- [94] S. R. Holbrook, Structural principles from large RNAs., *Annual review of biophysics* 37 (1) (2008) 445–464. doi:10.1146/annurev.biophys.36.040306.132755.
URL <http://dx.doi.org/10.1146/annurev.biophys.36.040306.132755>
- [95] D. J. Klein, P. B. Moore, T. A. Steitz, The contribution of metal ions to the structural stability of the large ribosomal subunit., *RNA* 10 (9) (2004) 1366–1379. doi:10.1261/rna.7390804.
URL <http://dx.doi.org/10.1261/rna.7390804>
- [96] T. R. Cech, B. L. Bass, Biological catalysis by RNA ., *Annual review of biochemistry* 55 (1) (1986) 599–629. doi:10.1146/annurev.bi.55.070186.003123.
URL <http://dx.doi.org/10.1146/annurev.bi.55.070186.003123>
- [97] S. H. Kim, J. L. Sussman, F. L. Suddath, G. J. Quigley, A. McPherson, A. H. Wang, N. C. Seeman, A. RICH, The general structure of transfer RNA molecules., *Proceedings of the National Academy of Sciences of the United States of America* 71 (12) (1974) 4970–4974.
URL <http://view.ncbi.nlm.nih.gov/pubmed/4612535>
- [98] S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. Wang, N. C. Seeman, A. Rich, Three-dimensional tertiary structure of yeast phenylalanine transfer RNA

Bibliography

- ., *Science* (New York, N.Y.) 185 (149) (1974) 435–440.
URL <http://view.ncbi.nlm.nih.gov/pubmed/4601792>
- [99] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, J. A. Doudna, Crystal structure of a group I ribozyme domain: principles of RNA packing., *Science*. 273 (5282) (1996) 1678–1685. doi:10.1126/science.273.5282.1678.
URL <http://dx.doi.org/10.1126/science.273.5282.1678>
- [100] R. A. Friedman, B. Honig, A free energy analysis of nucleic acid base stacking in aqueous solution., *Biophysical journal* 69 (4) (1995) 1528–1535. doi:10.1016/S0006-3495(95)80023-8.
URL [http://dx.doi.org/10.1016/S0006-3495\(95\)80023-8](http://dx.doi.org/10.1016/S0006-3495(95)80023-8)
- [101] N. B. Leontis, E. Westhof, Analysis of RNA motifs, *Current Opinion in Structural Biology* 13 (3) (2003) 300–308. doi:10.1016/S0959-440X(03)00076-9.
URL [http://dx.doi.org/10.1016/S0959-440X\(03\)00076-9](http://dx.doi.org/10.1016/S0959-440X(03)00076-9)
- [102] Y. Xin, C. Laing, N. B. B. Leontis, T. Schlick, Annotation of tertiary interactions in RNA structures reveals variations and correlations., *RNA* (New York, N.Y.)doi:10.1261/rna.1249208.
URL <http://dx.doi.org/10.1261/rna.1249208>
- [103] A. Lescoute, N. B. Leontis, C. Massire, E. Westhof, Recurrent structural RNA motifs, isostericity matrices and sequence alignments., *Nucleic Acids Res* 33 (8) (2005) 2395–2409.
URL <http://view.ncbi.nlm.nih.gov/pubmed/15860776>
- [104] M. Tamura, D. K. Hendrix, P. S. Klosterman, N. R. Schimmelman, S. E. Brenner, S. R. Holbrook, SCOR: Structural Classification of RNA, version 2.0., *Nucleic acids research* 32 (Database issue). doi:10.1093/nar/gkh080.
URL <http://dx.doi.org/10.1093/nar/gkh080>
- [105] P. S. Klosterman, D. K. Hendrix, M. Tamura, S. R. Holbrook, S. E. Brenner, Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns., *Nucleic Acids Res* 32 (8) (2004) 2342–2352.
URL <http://view.ncbi.nlm.nih.gov/pubmed/15121895>
- [106] R. R. Gutell, B. Weiser, C. R. Woese, H. F. Noller, Comparative anatomy of 16-S-like ribosomal RNA ., *Progress in nucleic acid research and molecular biology* 32 (1985) 155–216.
URL <http://view.ncbi.nlm.nih.gov/pubmed/3911275>
- [107] R. R. Gutell, M. W. Gray, M. N. Schnare, A compilation of large subunit (23S and 23S-like) ribosomal RNA structures., *Nucleic acids research* 21 (13) (1993) 3055–3074.
URL <http://view.ncbi.nlm.nih.gov/pubmed/8332527>
- [108] M. Tamura, S. R. Holbrook, Sequence and structural conservation in RNA ribose zippers, *Journal of Molecular Biology* 320 (3) (2002) 455–474.

Bibliography

- URL <http://www.sciencedirect.com/science/article/B6WK7-46571GY-4/2/3efd7e8960b84e4d3eb17a3f858f993b>
- [109] S. A. Strobel, P. L. Adams, M. R. Stahley, J. Wang, RNA kink turns to the left and to the right., *RNA (New York, N.Y.)* 10 (12) (2004) 1852–1854. doi:10.1261/rna.7141504.
URL <http://dx.doi.org/10.1261/rna.7141504>
- [110] C. R. Woese, S. Winker, R. R. Gutell, Architecture of ribosomal RNA : Constraints on the sequence of "tetra-loops"., *Proc Natl Acad Sci U S A* 87 (21) (1990) 8467–8471.
URL <http://view.ncbi.nlm.nih.gov/pubmed/2236056>
- [111] O. C. Uhlenbeck, Tetraloops and RNA folding., *Nature* 346 (6285) (1990) 613–614. doi:10.1038/346613a0.
URL <http://dx.doi.org/10.1038/346613a0>
- [112] L. Jaeger, F. Michel, E. Westhof, Involvement of a GNRA tetraloop in long-range RNA tertiary interactions., *Journal of molecular biology* 236 (5) (1994) 1271–1276.
URL <http://view.ncbi.nlm.nih.gov/pubmed/7510342>
- [113] M. Costa, F. Michel, Frequent use of the same tertiary motif by self-folding RNA s., *The EMBO journal* 14 (6) (1995) 1276–1285.
URL <http://view.ncbi.nlm.nih.gov/pubmed/7720718>
- [114] M. A. Tanner, T. R. Cech, An important RNA tertiary interaction of group I and group II introns is implicated in gram-positive RNase P RNAs., *RNA (New York, N.Y.)* 1 (4) (1995) 349–350.
URL <http://view.ncbi.nlm.nih.gov/pubmed/7493313>
- [115] E. Lafuente, R. Ramos, E. Martínez-Salas, Long-range RNA -RNA interactions between distant regions of the hepatitis C virus internal ribosome entry site element., *The Journal of general virology* 83 (Pt 5) (2002) 1113–1121.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11961266>
- [116] A. Munishkin, I. G. Wool, The ribosome-in-pieces: binding of elongation factor EF-G to oligoribonucleotides that mimic the sarcin/ricin and thiostrepton domains of 23S ribosomal RNA ., *Proceedings of the National Academy of Sciences of the United States of America* 94 (23) (1997) 12280–12284.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9356440>
- [117] C. Zwieb, Conformity of RNA s that interact with tetranucleotide loop binding proteins., *Nucleic acids research* 20 (17) (1992) 4397–4400.
URL <http://view.ncbi.nlm.nih.gov/pubmed/1329024>
- [118] J.-L. L. Chen, C. W. Greider, Functional analysis of the pseudoknot structure in human telomerase RNA ., *Proceedings of the National Academy of Sciences of the United States of America* 102 (23) (2005) 8080–8085. doi:10.1073/pnas.0502259102.
URL <http://dx.doi.org/10.1073/pnas.0502259102>

Bibliography

- [119] D. W. Staple, S. E. Butcher, Pseudoknots: RNA structures with diverse functions., *PLoS biology* 3 (6). doi:10.1371/journal.pbio.0030213.
URL <http://dx.doi.org/10.1371/journal.pbio.0030213>
- [120] C. W. Pleij, K. Rietveld, L. Bosch, A new principle of RNA folding based on pseudoknotting., *Nucleic acids research* 13 (5) (1985) 1717–1731.
URL <http://view.ncbi.nlm.nih.gov/pubmed/4000943>
- [121] K. E. Berry, S. Waghray, J. A. Doudna, The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit., *RNA (New York, N.Y.)* 16 (8) (2010) 1559–1569. doi:10.1261/rna.2197210.
URL <http://dx.doi.org/10.1261/rna.2197210>
- [122] J. J. Ellis, M. Broom, S. Jones, Protein-RNA interactions: structural analysis and functional classes., *Proteins* 66 (4) (2007) 903–911. doi:10.1002/prot.21211.
URL <http://dx.doi.org/10.1002/prot.21211>
- [123] S. Woodson, RNA folding and ribosome assembly, *Current Opinion in Chemical Biology* 12 (6) (2008) 667–673. doi:10.1016/j.cbpa.2008.09.024.
URL <http://dx.doi.org/10.1016/j.cbpa.2008.09.024>
- [124] B. Shapiro, Y. Yingling, W. Kasprzak, E. Bindewald, Bridging the gap in RNA structure prediction, *Current Opinion in Structural Biology* 17 (2) (2007) 157–165. doi:10.1016/j.sbi.2007.03.001.
URL <http://dx.doi.org/10.1016/j.sbi.2007.03.001>
- [125] R. Das, D. Baker, Automated *de novo* prediction of native-like RNA tertiary structures, *Proceedings of the National Academy of Sciences* 104 (37) (2007) 14664–14669. doi:10.1073/pnas.0703836104.
URL <http://dx.doi.org/10.1073/pnas.0703836104>
- [126] M. Parisien, F. Major, The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, *Nature* 452 (7183) (2008) 51–55. doi:10.1038/nature06684.
URL <http://dx.doi.org/10.1038/nature06684>
- [127] C. Massire, E. Westhof, MANIP: an interactive tool for modelling RNA., *Journal of molecular graphics & modelling* 16 (4-6).
URL <http://view.ncbi.nlm.nih.gov/pubmed/10522239>
- [128] A. Malhotra, R. K. Tan, S. C. Harvey, Prediction of the three-dimensional structure of *Escherichia coli* 30S ribosomal subunit: a molecular mechanics approach., *Proceedings of the National Academy of Sciences of the United States of America* 87 (5) (1990) 1950–1954.
URL <http://view.ncbi.nlm.nih.gov/pubmed/2408047>
- [129] A. Malhotra, S. C. Harvey, A quantitative model of the *Escherichia coli* 16S RNA in the 30S ribosomal subunit., *Journal of molecular biology* 240 (4) (1994) 308–340. doi:10.1006/jmbi.1994.1448.
URL <http://dx.doi.org/10.1006/jmbi.1994.1448>

Bibliography

- [130] A. Malhotra, R. K. Tan, S. C. Harvey, Modeling large RNAs and ribonucleoprotein particles using molecular mechanics techniques., *Biophysical journal* 66 (6) (1994) 1777–1795.
URL <http://view.ncbi.nlm.nih.gov/pubmed/7521223>
- [131] J. Burks, C. Zwieb, F. Müller, I. Wower, J. Wower, Comparative 3-D modeling of tmRNA., *BMC molecular biology* 6 (1). doi:10.1186/1471-2199-6-14.
URL <http://dx.doi.org/10.1186/1471-2199-6-14>
- [132] H. M. Martinez, J. V. Maizel, B. A. Shapiro, RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA., *Journal of biomolecular structure & dynamics* 25 (6) (2008) 669–683.
URL <http://view.ncbi.nlm.nih.gov/pubmed/18399701>
- [133] Y. G. Yingling, B. A. Shapiro, The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation, *Journal of Molecular Graphics and Modelling* 25 (2) (2006) 261–274. doi:10.1016/j.jmgm.2006.01.003.
URL <http://dx.doi.org/10.1016/j.jmgm.2006.01.003>
- [134] D. Lambert, J. E. Heckman, J. M. Burke, Cation-specific structural accommodation within a catalytic RNA ., *Biochemistry* 45 (3) (2006) 829–838. doi:10.1021/bi0513709.
URL <http://dx.doi.org/10.1021/bi0513709>
- [135] F. Jossinet, E. Westhof, Sequence to structure (S2S): display, manipulate and interconnect RNA data from sequence to structure, *Bioinformatics* 21 (15) (2005) 3320–3321. doi:10.1093/bioinformatics/bti504.
URL <http://view.ncbi.nlm.nih.gov/pubmed/15905274>
- [136] F. Jossinet, T. E. Ludwig, E. Westhof, Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels, *Bioinformatics* 26 (16) (2010) 2057–2059. doi:10.1093/bioinformatics/btq321.
URL <http://dx.doi.org/10.1093/bioinformatics/btq321>
- [137] O. Llorca, Introduction to 3D reconstruction of macromolecules using single particle electron microscopy., *Acta Pharmacol Sin* 26 (10) (2005) 1153–1164.
URL <http://view.ncbi.nlm.nih.gov/pubmed/16174429>
- [138] J. Frank, Single-particle imaging of macromolecules by cryo-electron microscopy, *Annual Review of Biophysics and Biomolecular Structure* 31 (1) (2002) 303–319. doi:10.1146/annurev.biophys.31.082901.134202.
URL <http://dx.doi.org/10.1146/annurev.biophys.31.082901.134202>
- [139] S. J. Ludtke, M. L. Baker, D.-H. Chen, J.-L. Song, D. T. Chuang, W. Chiu, *De Novo* backbone trace of GroEL from single particle electron cryomicroscopy, *Structure* 16 (3) (2008) 441–448. doi:10.1016/j.str.2008.02.007.
URL <http://dx.doi.org/10.1016/j.str.2008.02.007>
- [140] X. Zhang, E. Settembre, C. Xu, P. R. Dormitzer, R. Bellamy, S. C. Harrison, N. Grigorieff, Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction,

Bibliography

- Proceedings of the National Academy of Sciences 105 (6) (2008) 1867–1872. doi:10.1073/pnas.0711623105.
URL <http://dx.doi.org/10.1073/pnas.0711623105>
- [141] T. Becker, S. Bhushan, A. Jarasch, J.-P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E. C. Mandon, R. Beckmann, Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome, *Science (New York, N.Y.)* 326 (5958) (2009) 1369–1373. doi:10.1126/science.1178535.
URL <http://dx.doi.org/10.1126/science.1178535>
- [142] R. Henderson, Realizing the potential of electron cryo-microscopy., *Quarterly reviews of biophysics* 37 (1) (2004) 3–13.
URL <http://view.ncbi.nlm.nih.gov/pubmed/17390603>
- [143] P. A. Penczek, C. Yang, J. Frank, C. M. T. Spahn, Estimation of variance in single-particle reconstruction using the bootstrap technique, *Journal of Structural Biology* 154 (2) (2006) 168–183. doi:10.1016/j.jsb.2006.01.003.
URL <http://dx.doi.org/10.1016/j.jsb.2006.01.003>
- [144] P. A. Penczek, J. Frank, C. M. Spahn, A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation., *Journal of structural biology* 154 (2) (2006) 184–194. doi:10.1016/j.jsb.2005.12.013.
URL <http://dx.doi.org/10.1016/j.jsb.2005.12.013>
- [145] J. Fu, H. Gao, J. Frank, Unsupervised classification of single particles by cluster tracking in multi-dimensional space., *Journal of structural biology* 157 (1) (2007) 226–239. doi:10.1016/j.jsb.2006.06.012.
URL <http://dx.doi.org/10.1016/j.jsb.2006.06.012>
- [146] A. H. Erickson, G. Blobel, Cell-free translation of messenger RNA in a wheat germ system., *Methods in enzymology* 96 (1983) 38–50.
URL <http://view.ncbi.nlm.nih.gov/pubmed/6656637>
- [147] M. Halic, T. Becker, M. R. Pool, C. M. Spahn, R. A. Grassucci, J. Frank, R. Beckmann, Structure of the signal recognition particle interacting with the elongation-arrested ribosome., *Nature* 427 (6977) (2004) 808–814. doi:10.1038/nature02342.
URL <http://dx.doi.org/10.1038/nature02342>
- [148] J. A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in electron microscopy., *Journal of structural biology* 142 (3) (2003) 334–347.
URL <http://view.ncbi.nlm.nih.gov/pubmed/12781660>
- [149] J. Z. Chen, N. Grigorieff, Signature: a single-particle selection system for molecular electron microscopy., *Journal of structural biology* 157 (1) (2007) 168–173. doi:10.1016/j.jsb.2006.06.001.
URL <http://dx.doi.org/10.1016/j.jsb.2006.06.001>

Bibliography

- [150] J. Frank, SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields, *Journal of Structural Biology* 116 (1) (1996) 190–199. doi:10.1006/jsbi.1996.0030.
URL <http://dx.doi.org/10.1006/jsbi.1996.0030>
- [151] T. R. Shaikh, H. Gao, W. T. Baxter, F. J. Asturias, N. Boisset, A. Leith, J. Frank, SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs, *Nature Protocols* 3 (12) (2008) 1941–1974. doi:10.1038/nprot.2008.156.
URL <http://dx.doi.org/10.1038/nprot.2008.156>
- [152] G. J. Kleywegt, T. A. Jones, Programs for Reformatting, Analysis and Manipulation of Biomacromolecular Electron-Density Maps and Reflection Data Sets, *Acta Crystallographica Section D* 52 (4) (1996) 826–828. doi:10.1107/S09074444995014983.
URL <http://dx.doi.org/10.1107/S09074444995014983>
- [153] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF chimera - A visualization system for exploratory research and analysis., *Journal of computational chemistry* 25 (13) (2004) 1605–1612. doi:10.1002/jcc.20084.
URL <http://dx.doi.org/10.1002/jcc.20084>
- [154] T. M. Schmeing, K. S. Huang, D. E. Kitchen, S. A. Strobel, T. A. Steitz, Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction., *Molecular cell* 20 (3) (2005) 437–448. doi:10.1016/j.molcel.2005.09.006.
URL <http://dx.doi.org/10.1016/j.molcel.2005.09.006>
- [155] M. A. Borovinskaya, R. D. Pai, W. Zhang, B. S. Schuwirth, J. M. Holton, G. Hirokawa, H. Kaji, A. Kaji, Jamie, Structural basis for aminoglycoside inhibition of bacterial ribosome recycling, *Nature Structural & Molecular Biology* 14 (8) (2007) 727–732. doi:10.1038/nsmb1271.
URL <http://dx.doi.org/10.1038/nsmb1271>
- [156] E. Westhof, P. Dumas, D. Moras, Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals., *Acta crystallographica. Section A, Foundations of crystallography.* 44 (Pt 2) (1988) 112–123.
URL <http://view.ncbi.nlm.nih.gov/pubmed/3272146>
- [157] C. H. Kim, C. C. Kao, I. Tinoco, RNA motifs that determine specificity between a viral replicase and its promoter., *Nature structural biology* 7 (5) (2000) 415–423. doi:10.1038/75202.
URL <http://dx.doi.org/10.1038/75202>
- [158] H. Wu, P. K. Yang, S. E. Butcher, S. Kang, G. Chanfreau, J. Feigon, A novel family of RNA tetraloop structure forms the recognition site for *Saccharomyces cerevisiae* RNase III., *EMBO J* 20 (24) (2001) 7240–7249.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11743000>

Bibliography

- [159] X. Ye, R. A. Kumar, D. J. Patel, Molecular recognition in the bovine immunodeficiency virus tat peptide-tar RNA complex., *Chemistry & biology* 2 (12) (1995) 827–840.
URL <http://view.ncbi.nlm.nih.gov/pubmed/8807816>
- [160] A. Bashan, I. Agmon, R. Zarivach, F. Schluenzen, J. Harms, R. Berisio, H. Bartels, F. Franceschi, T. Auerbach, H. A. Hansen, E. Kossoy, M. Kessler, A. Yonath, Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression., *Molecular cell* 11 (1) (2003) 91–102.
URL <http://view.ncbi.nlm.nih.gov/pubmed/12535524>
- [161] B. Hoffmann, G. T. Mitchell, P. Gendron, F. Major, A. A. Andersen, R. A. Collins, P. Legault, NMR structure of the active conformation of the *Varkud satellite* ribozyme cleavage site., *Proceedings of the National Academy of Sciences of the United States of America* 100 (12) (2003) 7003–7008. doi:10.1073/pnas.0832440100.
URL <http://dx.doi.org/10.1073/pnas.0832440100>
- [162] A. Nikulin, A. Serganov, E. Ennifar, S. Tishchenko, N. Nevskaya, W. Shepard, C. Portier, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, P. Dumas, Crystal structure of the S15-rRNA complex., *Nature structural biology* 7 (4) (2000) 273–277. doi:10.1038/74028.
URL <http://dx.doi.org/10.1038/74028>
- [163] H. Shi, P. B. Moore, The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited., *RNA* 6 (8) (2000) 1091–1105.
URL <http://view.ncbi.nlm.nih.gov/pubmed/10943889>
- [164] N. Amrani, S. Ghosh, D. A. Mangus, A. Jacobson, Translation factors promote the formation of two states of the closed-loop mRNP, *Nature* doi:10.1038/nature06974.
URL <http://dx.doi.org/10.1038/nature06974>
- [165] A. J. Lee, D. M. Crothers, The solution structure of an RNA loop-loop complex: the cole1 inverted loop sequence., *Structure (London, England : 1993)* 6 (8) (1998) 993–1005.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9739090>
- [166] F. Schluenzen, J. M. Harms, F. Franceschi, H. A. Hansen, H. Bartels, R. Zarivach, A. Yonath, Structural basis for the antibiotic activity of ketolides and azalides, *Structure* 11 (3) (2003) 329–338. doi:10.1016/S0969-2126(03)00022-4.
URL [http://dx.doi.org/10.1016/S0969-2126\(03\)00022-4](http://dx.doi.org/10.1016/S0969-2126(03)00022-4)
- [167] G. Blaha, R. E. Stanley, T. A. Steitz, Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome., *Science (New York, N.Y.)* 325 (5943) (2009) 966–970. doi:10.1126/science.1175800.
URL <http://dx.doi.org/10.1126/science.1175800>
- [168] L. G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics., *Structure (London, England : 1993)* 16 (5) (2008) 673–683. doi:10.1016/j.str.2008.03.005.
URL <http://dx.doi.org/10.1016/j.str.2008.03.005>

Bibliography

- [169] I. L. Hofacker, Vienna RNA secondary structure server, *Nucleic Acids Research* 31 (13) (2003) 3429–3431. doi:10.1093/nar/gkg599.
URL <http://dx.doi.org/10.1093/nar/gkg599>
- [170] P. Steffen, B. Voss, M. Rehmsmeier, J. Reeder, R. Giegerich, RNASHapes: an integrated RNA analysis package based on abstract shapes, *Bioinformatics* 22 (4) (2006) 500–503. doi:10.1093/bioinformatics/btk010.
URL <http://dx.doi.org/10.1093/bioinformatics/btk010>
- [171] H. Yang, F. Jossinet, N. Leontis, L. Chen, J. Westbrook, H. Berman, E. Westhof, Tools for the automatic identification and classification of RNA base pairs, *Nucl. Acids Res.* 31 (13) (2003) 3450–3460.
URL <http://nar.oxfordjournals.org/cgi/content/abstract/31/13/3450>
- [172] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics., *Journal of molecular graphics* 14 (1).
URL <http://view.ncbi.nlm.nih.gov/pubmed/8744570>
- [173] N. B. Leontis, E. Westhof, Geometric nomenclature and classification of RNA base pairs., *RNA* 7 (4) (2001) 499–512.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11345429>
- [174] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (16) (2005) 1781–1802. doi:10.1002/jcc.20289.
URL <http://dx.doi.org/10.1002/jcc.20289>
- [175] C. Grauffel, R. H. Stote, A. Dejaegere, Force field parameters for the simulation of modified histone tails, *Journal of Computational Chemistry* 31 (13) (2010) 2434–2451. doi:10.1002/jcc.21536.
URL <http://dx.doi.org/10.1002/jcc.21536>
- [176] P. Grayson, E. Tajkhorshid, K. Schulten, Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics., *Biophysical journal* 85 (1) (2003) 36–48. doi:10.1016/S0006-3495(03)74452-X.
URL [http://dx.doi.org/10.1016/S0006-3495\(03\)74452-X](http://dx.doi.org/10.1016/S0006-3495(03)74452-X)
- [177] E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, T. E. Ferrin, Tools for integrated sequence-structure analysis with UCSF chimera., *BMC bioinformatics* 7 (2006) 339+. doi:10.1186/1471-2105-7-339.
URL <http://dx.doi.org/10.1186/1471-2105-7-339>
- [178] Emsley, P., Cowtan, K., Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D: Biological Crystallography* 60 (1) (2004) 2126–2132. doi:10.1107/S0907444904019158.
URL <http://dx.doi.org/10.1107/S0907444904019158>

Bibliography

- [179] E. Bindewald, B. A. Shapiro, RNA secondary structure prediction from sequence alignments using a network of k-nearest neighbor classifiers, *RNA* 12 (3) (2006) 342–352. doi:10.1261/rna.2164906.
URL <http://dx.doi.org/10.1261/rna.2164906>
- [180] L. G. Trabuco, E. Villa, E. Schreiner, C. B. Harrison, K. Schulten, Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography, *Methods* 49 (2) (2009) 174–180. doi:10.1016/j.ymeth.2009.04.005.
URL <http://dx.doi.org/10.1016/j.ymeth.2009.04.005>
- [181] T. Becker, J. P. Armache, A. Jarasch, A. M. Anger, E. Villa, H. Sieber, B. Abdel Motal, T. Mielke, O. Berninghausen, R. Beckmann, Structure of the No-Go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome.
- [182] J.-P. Armache, A. Jarasch, A. M. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. N. Wilson, R. Beckmann, Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution., *Proceedings of the National Academy of Sciences of the United States of America* 107 (46) (2010) 19748–19753. doi:10.1073/pnas.1009999107.
URL <http://dx.doi.org/10.1073/pnas.1009999107>
- [183] N. Ban, P. Nissen, J. Hansen, M. Capel, P. B. Moore, T. A. Steitz, Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit., *Nature* 400 (6747) (1999) 841–847. doi:10.1038/23641.
URL <http://dx.doi.org/10.1038/23641>
- [184] H. F. Noller, M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, L. Lancaster, A. Dallas, K. Fredrick, T. N. Earnest, J. H. Cate, Structure of the ribosome at 5.5 Å resolution and its interactions with functional ligands., *Cold Spring Harbor symposia on quantitative biology* 66 (2001) 57–66.
URL <http://view.ncbi.nlm.nih.gov/pubmed/12762008>
- [185] J.-P. Armache, A. Jarasch, A. M. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. N. Wilson, R. Beckmann, Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome., *Proceedings of the National Academy of Sciences of the United States of America* 107 (46) (2010) 19754–19759. doi:10.1073/pnas.1010005107.
URL <http://dx.doi.org/10.1073/pnas.1010005107>
- [186] A. Lescoute, E. Westhof, Topology of three-way junctions in folded RNAs., *RNA* 12 (1) (2006) 83–93. doi:10.1261/rna.2208106.
URL <http://dx.doi.org/10.1261/rna.2208106>
- [187] T. Elgavish, J. J. Cannone, J. C. Lee, S. C. Harvey, R. R. Gutell, AA:AG@Helix.Ends: A:A and A:G base-pairs at the ends of 16S and 23S rRNA helices: A:A and A:G base-pairs

Bibliography

- at the ends of 16S and 23S rRNA helices, *Journal of Molecular Biology* 310 (4) (2001) 735–753.
URL <http://www.sciencedirect.com/science/article/B6WK7-457CYC2-DN/2/d4da26f149532c5746965d832a006843>
- [188] S. Bhushan, M. Gartmann, M. Halic, J.-P. P. Armache, A. Jarasch, T. Mielke, O. Berninghausen, D. N. Wilson, R. Beckmann, alpha-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel., *Nature structural & molecular biology* 17 (3) (2010) 313–317. doi:10.1038/nsmb.1756.
URL <http://dx.doi.org/10.1038/nsmb.1756>
- [189] T. M. Schmeing, K. S. Huang, S. A. Strobel, T. A. Steitz, An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA., *Nature* 438 (7067) (2005) 520–524. doi:10.1038/nature04152.
URL <http://dx.doi.org/10.1038/nature04152>
- [190] E. C. Mandon, S. F. Trueman, R. Gilmore, Translocation of proteins through the Sec61 and SecYEG channels., *Current opinion in cell biology* 21 (4) (2009) 501–507. doi:10.1016/j.ceb.2009.04.010.
URL <http://dx.doi.org/10.1016/j.ceb.2009.04.010>
- [191] T. A. Rapoport, Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes, *Nature* 450 (7170) (2007) 663–669. doi:10.1038/nature06384.
URL <http://dx.doi.org/10.1038/nature06384>
- [192] N. Hosoda, T. Kobayashi, N. Uchida, Y. Funakoshi, Y. Kikuchi, S. Hoshino, T. Katada, Translation termination factor eRF3 mediates mRNA decay through the regulation of deadenylation., *J Biol Chem* 278 (40) (2003) 38287–38291. doi:10.1074/jbc.C300300200.
URL <http://dx.doi.org/10.1074/jbc.C300300200>
- [193] M. K. Doma, R. Parker, Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation, *Nature* 440 (7083) 561–564. doi:10.1038/nature04530.
URL <http://dx.doi.org/10.1038/nature04530>
- [194] H. H. Lee, Y. S. Kim, K. H. Kim, I. Heo, S. K. Kim, O. Kim, H. K. Kim, J. Y. Yoon, H. S. Kim, d. o. . J. Kim, S. J. Lee, H. J. Yoon, S. J. Kim, B. G. Lee, H. K. Song, V. N. Kim, C. M. Park, S. W. Suh, Structural and functional insights into Dom34, a key component of No-Go mRNA decay., *Mol Cell* 27 (6) (2007) 938–950. doi:10.1016/j.molcel.2007.07.019.
URL <http://dx.doi.org/10.1016/j.molcel.2007.07.019>
- [195] M. Graille, M. Chaillet, H. van Tilbeurgh, Structure of yeast Dom34: a protein related to translation termination factor eRF1 and involved in No-Go decay., *J Biol Chem*doi: 10.1074/jbc.M708224200.
URL <http://dx.doi.org/10.1074/jbc.M708224200>

Bibliography

- [196] L. Chen, D. Muhrad, V. Hauryliuk, Z. Cheng, M. K. K. Lim, V. Shyp, R. Parker, H. Song, Structure of the Dom34-Hbs1 complex and implications for No-Go decay., *Nature structural & molecular biology* 17 (10) (2010) 1233–1240. doi:10.1038/nsmb.1922.
URL <http://dx.doi.org/10.1038/nsmb.1922>
- [197] K. Kobayashi, I. Kikuno, K. Kuroha, K. Saito, K. Ito, R. Ishitani, T. Inada, O. Nureki, Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1alpha complex., *Proceedings of the National Academy of Sciences of the United States of America* doi:10.1073/pnas.1009598107.
URL <http://dx.doi.org/10.1073/pnas.1009598107>
- [198] A. A. Szewczak, P. B. Moore, The Sarcin/Ricin Loop, a modular RNA., *Journal of Molecular Biology* 247 (1) (1995) 81–98. doi:10.1006/jmbi.1994.0124.
URL <http://dx.doi.org/10.1006/jmbi.1994.0124>
- [199] B. Wimberly, G. Varani, I. Tinoco, The conformation of loop e of eukaryotic 5S ribosomal RNA., *Biochemistry* 32 (4) (1993) 1078–1087.
URL <http://view.ncbi.nlm.nih.gov/pubmed/8424938>
- [200] C. C. Correll, I. G. Wool, A. Munishkin, The two faces of the *Escherichia coli* 23S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution., *Journal of molecular biology* 292 (2) (1999) 275–287. doi:10.1006/jmbi.1999.3072.
URL <http://dx.doi.org/10.1006/jmbi.1999.3072>
- [201] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson, Multiple sequence alignment with the Clustal series of programs., *Nucleic acids research* 31 (13) (2003) 3497–3500. doi:10.1093/nar/gkg500.
URL <http://dx.doi.org/10.1093/nar/gkg500>
- [202] E. Villa, J. Sengupta, L. G. Trabuco, J. LeBarron, W. T. Baxter, T. R. Shaikh, R. A. Grassucci, P. Nissen, M. Ehrenberg, K. Schulten, J. Frank, Ribosome-induced changes in Elongation factor Tu conformation control GTP hydrolysis., *Proceedings of the National Academy of Sciences of the United States of America* 106 (4) (2009) 1063–1068. doi:10.1073/pnas.0811370106.
URL <http://dx.doi.org/10.1073/pnas.0811370106>
- [203] G. Alkemar, O. Nygård, Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA, *Biochemistry* 45 (26) (2006) 8067–8078. doi:10.1021/bi052149z.
URL <http://dx.doi.org/10.1021/bi052149z>
- [204] J. M. Hancock, D. Tautz, G. A. Dover, Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*., *Mol Biol Evol* 5 (4) (1988) 393–414.
URL <http://view.ncbi.nlm.nih.gov/pubmed/3136295>
- [205] T. J. Crease, J. K. Colbourne, The unusually long small-subunit ribosomal RNA of the crustacean, *daphnia pulex*: sequence and predicted secondary structure., *Journal of molec-*

Bibliography

- ular evolution 46 (3) (1998) 307–313.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9493355>
- [206] C. P. Choe, J. M. Hancock, U. W. Hwang, W. Kim, Analysis of the primary sequence and secondary structure of the unusually long SSU rRNA of the soil bug, *Armadillidium vulgare.*, Journal of molecular evolution 49 (6) (1999) 798–805.
URL <http://view.ncbi.nlm.nih.gov/pubmed/10594181>
- [207] J. M. Neefs, R. De Wachter, A proposal for the secondary structure of a variable area of eukaryotic small ribosomal subunit RNA involving the existence of a pseudoknot., Nucleic acids research 18 (19) (1990) 5695–5704.
URL <http://view.ncbi.nlm.nih.gov/pubmed/2170942>
- [208] R. E. Ellis, J. E. Sulston, A. R. Coulson, The rDNA of *C. elegans*: sequence and structure., Nucleic acids research 14 (5) (1986) 2345–2364.
URL <http://view.ncbi.nlm.nih.gov/pubmed/3960722>
- [209] I. L. Gonzalez, R. D. Schmickel, The human 18S ribosomal RNA gene: evolution and stability., American journal of human genetics 38 (4) (1986) 419–427.
URL <http://view.ncbi.nlm.nih.gov/pubmed/2422931>
- [210] D. L. Nickrent, M. L. Sargent, An overview of the secondary structure of the V4 region of eukaryotic small-subunit ribosomal RNA., Nucleic acids research 19 (2) (1991) 227–235.
URL <http://view.ncbi.nlm.nih.gov/pubmed/2014163>
- [211] J. M. Hancock, A. P. Vogler, Modelling the secondary structures of slippage-prone hypervariable RNA regions: the example of the tiger beetle 18S rRNA variable region V4., Nucleic acids research 26 (7) (1998) 1689–1699.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9512540>
- [212] K. Bulygin, L. Chavatte, L. Frolova, G. Karpova, A. Favre, The first position of a codon placed in the a site of the human 80S ribosome contacts nucleotide C1696 of the 18S rRNA as well as proteins S2, S3, S3a, S30, and S15., Biochemistry 44 (6) (2005) 2153–2162. doi:10.1021/bi0487802.
URL <http://dx.doi.org/10.1021/bi0487802>
- [213] L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, Structural aspects of messenger RNA reading frame maintenance by the ribosome, Nature Structural & Molecular Biology 17 (5) (2010) 555–560. doi:10.1038/nsmb.1790.
URL <http://dx.doi.org/10.1038/nsmb.1790>
- [214] R. M. M. Voorhees, A. Weixlbaumer, D. Loakes, A. C. C. Kelley, V. Ramakrishnan, Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome., Nature structural & molecular biology doi:10.1038/nsmb.1577.
URL <http://dx.doi.org/10.1038/nsmb.1577>
- [215] B. A. Maguire, A. D. Beniaminov, H. Ramu, A. S. Mankin, R. A. Zimmermann, A protein component at the heart of an RNA machine: the importance of protein L27 for the function

Bibliography

- of the bacterial ribosome., *Molecular cell* 20 (3) (2005) 427–435. doi:10.1016/j.molcel.2005.09.009.
URL <http://dx.doi.org/10.1016/j.molcel.2005.09.009>
- [216] A. Hofer, C. Bussiere, A. W. Johnson, Mutational analysis of the ribosomal protein RpL10 from yeast, *Journal of Biological Chemistry* 282 (45) (2007) 32630–32639. doi:10.1074/jbc.M705057200.
URL <http://dx.doi.org/10.1074/jbc.M705057200>
- [217] A. Kosolapov, C. Deutsch, Tertiary interactions within the ribosomal exit tunnel., *Nature structural & molecular biology* 16 (4) (2009) 405–411. doi:10.1038/nsmb.1571.
URL <http://dx.doi.org/10.1038/nsmb.1571>
- [218] U. Berndt, S. Oellerer, Y. Zhang, A. E. Johnson, S. Rospert, A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel, *Proceedings of the National Academy of Sciences* 106 (5) (2009) 1398–1403. doi:10.1073/pnas.0808584106.
URL <http://dx.doi.org/10.1073/pnas.0808584106>
- [219] I. Mingarro, I. Nilsson, P. Whitley, G. von Heijne, Different conformations of nascent polypeptides during translocation across the ER membrane., *BMC cell biology* 1.
URL <http://view.ncbi.nlm.nih.gov/pubmed/11178101>
- [220] L. R. Cruz-Vera, S. Rajagopal, C. Squires, C. Yanofsky, Features of ribosome-peptidyl-tRNA interactions essential for tryptophan induction of tna operon expression., *Molecular cell* 19 (3) (2005) 333–343. doi:10.1016/j.molcel.2005.06.013.
URL <http://dx.doi.org/10.1016/j.molcel.2005.06.013>
- [221] A. Murakami, H. Nakatogawa, K. Ito, Translation arrest of SecM is essential for the basal and regulated expression of SecA, *PNAS* 101 (33) (2004) 12330–12335. doi:10.1073/pnas.0404907101.
URL <http://dx.doi.org/10.1073/pnas.0404907101>
- [222] P. M. Petrone, C. D. Snow, D. Lucent, V. S. Pande, Side-chain recognition and gating in the ribosome exit tunnel, *Proceedings of the National Academy of Sciences* 105 (43) (2008) 16549–16554. doi:10.1073/pnas.0801795105.
URL <http://dx.doi.org/10.1073/pnas.0801795105>
- [223] R. Beckmann, D. Bubeck, R. Grassucci, P. Penczek, A. Verschoor, G. Blobel, J. Frank, Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex., *Science (New York, N.Y.)* 278 (5346) (1997) 2123–2126.
URL <http://view.ncbi.nlm.nih.gov/pubmed/9405348>
- [224] J.-F. F. Ménétret, R. S. Hegde, S. U. Heinrich, P. Chandramouli, S. J. Ludtke, T. A. Rapoport, C. W. Akey, Architecture of the ribosome-channel complex derived from native membranes., *Journal of molecular biology* 348 (2) (2005) 445–457. doi:10.1016/j.jmb.

Bibliography

2005.02.053.

URL <http://dx.doi.org/10.1016/j.jmb.2005.02.053>

- [225] J. Frauenfeld, J. Gumbart, E. van der Sluis, S. Funes, M. Gartmann, B. Beatrix, O. Berninghausen, T. Becker, K. Schulten, R. Beckmann, Cryo-EM structure of the ribosome-SecYE complex in the membrane environment.
- [226] K. Mitra, C. Schaffitzel, T. Shaikh, F. Tama, S. Jenni, C. L. Brooks, N. Ban, J. Frank, Structure of the *E. coli* protein-conducting channel bound to a translating ribosome, *Nature* 438 (7066) (2005) 318–324. doi:10.1038/nature04133.
URL <http://dx.doi.org/10.1038/nature04133>
- [227] T. Yokoyama, T. Suzuki, Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments., *Nucleic acids research* 36 (11) (2008) 3539–3551. doi:10.1093/nar/gkn224.
URL <http://dx.doi.org/10.1093/nar/gkn224>
- [228] D. G. Morgan, J. F. Ménétret, M. Radermacher, A. Neuhof, I. V. Akey, T. A. Rapoport, C. W. Akey, A comparison of the yeast and rabbit 80S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments., *Journal of molecular biology* 301 (2) (2000) 301–321. doi:10.1006/jmbi.2000.3947.
URL <http://dx.doi.org/10.1006/jmbi.2000.3947>
- [229] C. M. Spahn, E. Jan, A. Mulder, R. A. Grassucci, P. Sarnow, J. Frank, Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor., *Cell* 118 (4) (2004) 465–475. doi:10.1016/j.cell.2004.08.001.
URL <http://dx.doi.org/10.1016/j.cell.2004.08.001>
- [230] B. Seidelt, C. A. Innis, D. N. Wilson, M. Gartmann, J.-P. P. Armache, E. Villa, L. G. Trabuco, T. Becker, T. Mielke, K. Schulten, T. A. Steitz, R. Beckmann, Structural insight into nascent polypeptide chain-mediated translational stalling., *Science (New York, N.Y.)* 326 (5958) (2009) 1412–1415. doi:10.1126/science.1177662.
URL <http://dx.doi.org/10.1126/science.1177662>

Acknowledgement

I would like to express my sincere gratitude to my Ph.D. supervisor Prof. Dr. Roland Beckmann for giving me the opportunity to work on this challenging and highly competitive project. I am grateful for his trust in me to independently work as a bioinformatician in his group, for providing an excellent working environment and money to travel to collaborators and conferences.

I am grateful for having been a member of this laboratory and funded by the cluster of excellence *CIPS^M* (Center for integrated protein science Munich).

Special thank go to Dr. Daniel Wilson and Dr. Thomas Becker for the vast amount of time and patience they spent on all the projects, productive discussions and their critical view of the RNA models.

I am deeply grateful to Prof. Dr. Eric Westhof for teaching me about RNA folding. I am really thankful for the discussions during all my visits in Straßburg and the e-mail correspondence during our collaboration. A special appreciation is dedicated to Dr. Fabrice Jossinet for our fruitful cooperation, the effort he made for constantly improving S2S and Assemble, for supporting my ideas and implementing most of them, for being a great host at all of my visits and being available countless nights for discussions and improvements as well as being critical concerning all talks, posters and figures.

Many thanks to Jean-Paul Armache who shared this challenging project with me, calculated the electron densities and modeled the r-proteins. I want to say thank you to Andreas Anger for suggesting ideas for some expansion segments as well as helping with secondary structure drawings; to Dr. Elisabeth Villa for introducing and scripting RNA restraints for MDFF; to Dr. Shashi Bhushan for the collaboration and tons of data; as well as to Thorsten Mielke and Jörg Bürger for microscope support and patience when my sample exploded. I am also grateful to Andreas Hauser for endless support of the MDFF simulations on the cluster.

I thank Prof. Dr. Patrick Cramer for recommending me to Roland Beckmann. I would further like to thank Prof. Dr. Karl-Peter Hopfner, Prof. Dr. Karl-Klaus Conzelmann and Prof. Dr. Klaus Förstemann for being in my thesis committee.

Furthermore, I deeply appreciate the critical reading of this thesis by Dr. Daniel Wilson, Dr. Thomas Becker and Daniela Dirndorfer.

Bibliography

I thank the whole AG Beckmann, AG Wilson and AG Wendler for the great atmosphere in the lab, which made some frustrations more bearable. I deeply appreciate the time with Dr. Jens Frauenfeld, Dr. Marco Gartmann, Dr. Birgit Seidelt, Christoph Leidig, Caroline Haas and Stephan Wickles, having tons of coffees, cakes, discussions, fun and especially sharing social activities. I would like to thank Christoph Leidig for outdoor activities, traveling and sharing such a nice time in the northern countries. I don't want to miss the billions of coffee breaks with Theresa Niederberger and Dr. Achim Tresch, which resulted in fruitful discussions and new ideas. I really appreciated the time in- and outside the lab with Cilli Sedlaczek and Dr. Soledad Funes.

Last but not least, I am deeply grateful to Daniela Dirndorfer, my sisters Nadine and Katja with Daniel and Roy, my brother Mario with Olga and my friends Franziska Misselwitz, Julia Wurzbacher, Anne Will, Jan Hoffmann, Viktor Rzesanke, Mario Jauch and Werner Will for all the support, patience, trust and encouragement.

Curriculum Vitae

Personal details

Name:	Alexander Jarasch
Date of birth:	1981/06/06
Place of birth:	Leipzig, Germany
Nationality:	German
Title:	Dipl.Bioinf. (equ. to M. Sc. in bioinformatics)

Education

2011/03 - 2007/05	Ph.D. student at Gene Center (LMU) <i>3D modeling of ribosomal RNA using cryo-electron microscopy density maps</i> (Adviser: Prof. Dr. Roland Beckmann)
2007/03	Diploma in bioinformatics
2007/03 - 2006/05	Research assistant and diploma thesis at the Max-Planck-Institute for Biochemistry and the Ludwig-Maximilians-Universität (LMU) <i>Similarity-based analysis of medium- and high-throughput screening data</i> (Dr. Thomas Mayer - Max Planck Institute) (Dr. Joannis Apostolakis- LMU) (Adviser: Prof. Dr. Ralf Zimmer - LMU)
2001/10	Studies of bioinformatics at both the Technical University of Munich (TUM) & Ludwig-Maximilians-Universität of Munich (LMU)
2001/04 - 2000/07	Military service
2000/06	Abitur at the Christian-Gottlieb-Reichard Gymnasium, Lobenstein