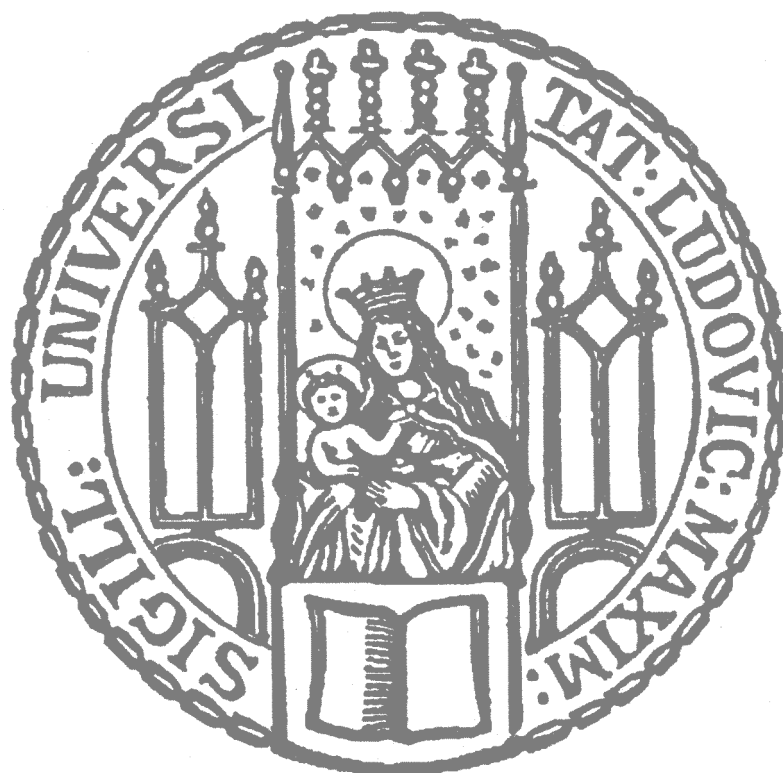


Dissertation zur Erlangung des Doktorgrades der Fakultät für Chemie und  
Pharmazie der Ludwig-Maximilians-Universität München

## **5.5 Å Structure of the Eukaryotic Ribosome**



Jean-Paul Armache

aus Łódź, Polen

2011

## **Erklärung**

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Roland Beckmann betreut.

## **Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfe erarbeitet.

München, am 25.02.2011

Jean-Paul Armache

Dissertation eingereicht am: 25.02.2011

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

Mündliche Prüfung am: 04.05.2011

# Table of Contents

Acknowledgements .....	6
Summary .....	8
1 Introduction .....	9
1.1 The Flow of Genetic Information .....	9
1.2 Ribosomes .....	10
1.3 Eukaryotic Ribosomes .....	11
1.4 Ribosome Biogenesis.....	11
1.5 Translation Mechanism .....	12
1.5 Peptide-bond Formation .....	14
1.6 Ribosomal proteins .....	15
1.7 Structures of Eukaryotic Ribosomes - Introduction.....	16
1.8 Aim (Outline of the Thesis).....	16
2 Materials and Methods .....	18
2.1 Sample Preparation.....	18
2.1.1 <i>Triticum aestivum</i> Sample Preparation (from Halic et al., 2004).....	18
2.1.2 <i>Saccharomyces cerevisiae</i> Knockout Strains Sample Preparation .....	19
2.1.3 Archaeal Ribosomes Sample Preparation (from Márquez et al., 2010).....	20
2.2 Cryo-Electron Microscopy .....	21
2.2.1 Sample Preparation .....	22
2.2.2 Image Generation.....	22
2.2.3 Image Processing.....	23
2.2.4 Resolution Determination.....	25
2.2.5 Map Deconvolution .....	25
2.3 Modeling.....	25

2.3.1 Protein Template Sequences .....	26
2.3.2 Comparative/Homology Modeling.....	26
2.3.3 Secondary Structure Prediction .....	27
2.3.4 Density-based Modeling with Secondary Structure Prediction.....	28
2.3.5 Rigid Body Fitting.....	29
2.3.6 Molecular Dynamics Flexible Fitting.....	29
2.3.6 Modeling Workflow Summary .....	29
2.4 Figures .....	31
3 Results.....	32
3.1 A Cryo-EM Map of a Eukaryotic Ribosome at 5.5 Å.....	32
3.1.1 Density .....	33
3.1.2 Feature resolution .....	33
3.2 Modeling into the Density.....	39
3.2.1 Conserved Core .....	39
3.2.2 Unlocalized Eukarya-Specific Proteins.....	42
3.2.3 Eukaryote-Specific Protein Extensions .....	47
3.3 Protein Localization.....	53
3.3.1 Small Subunit Proteins .....	54
3.3.1.3 Small Subunit Body.....	58
3.3.2 Large Subunit Proteins.....	58
3.3.3 The Remainder .....	68
4 Conclusions and Discussion .....	70
4.1 Cryo-Electron Microscopy Map of <i>Triticum aestivum</i> .....	70
4.1.1 Feature Visibility at Given Resolution.....	70
4.1.2 Improvements Over Existing Reconstructions of Eukaryotic Ribosomes.....	70
4.1.3 Cryo-EM and the future of intracellular investigation .....	71

4.2 Model .....	72
4.2.1 Localization of the Novel Proteins.....	72
4.2.2 Roles of Proteins.....	74
4.2.3 Interesting Folds and Interactions.....	83
4.2.4 Proteins on the ribosome .....	85
4.2.5 Reliability of the Model.....	88
4.2.6 Comparative Analysis with Crystal Structures .....	90
4.3 Ribosome in the Light of Evolution .....	94
5 Summary .....	98
6 Appendix .....	100
7 Curriculum vitae .....	127
8 References.....	129

## **Acknowledgements**

I would like to thank my supervisor, Prof. Roland Beckmann for allowing me to conduct research in his group in the Gene Center. I am grateful for his trust in a computer scientist working in Structural Biology, for the unique working conditions, atmosphere and his constant support.

Many thanks to Dr. Daniel Wilson for his patience, availability, openness to discussions and the immeasurable amount of time and work he invested in this project.

I wish to express a deep sense of gratitude to a number of people I had pleasure of working with. Dr. Thomas Becker, who taught me everything I know about cryo-electron microscopy and always had time to help me, no matter how busy he was; Dr. Shashi Bhushan, for the fruitful collaboration and the tremendous amount of high-quality data no one else could have acquired; Andreas Anger, Alexander Jarasch and Elizabeth Villa for all the time spent together, the time of crisis and the happy end; Dr. Viter Marquez and Sibylle Franckenberg for the very revealing archaeal ribosomes, which gave a lot of insight into protein distribution in Eukaryotes; Dr. Birgitta Beatrix, Dr. Eli van der Sluis and Dr. Soledad Funes for always being there with support and an answer to every challenging scientific question I could have; Dr. Otto Berninghausen and Charlotte Ungewickell for their constant work on improvement of the microscope; Andrea Gilmozzi, Joanna Musial, Heidi Sieber and Timo Weiler for all the hard work they do in the lab; Guelcin Dindar for the great ribosome purification she did for me; Ingegerd Walz who helped me when the I felt lost dealing with administrative issues.

I would like to thank Prof. Klaus Förstemann, Prof. Ulrich Hartl and Prof. Karl-Peter Hopfner for being in my thesis committee.

I am grateful to Dr. Daniel Wilson, Andreas Anger and Basma Abdel Motaal for critical comments on this thesis.

Many thanks to all the members of the Beckmann and Wilson lab, current and former, who made the time spent in the lab unique and fun.

My special appreciation goes to Agata, Anastasia, Andreas, Basma, Cilli, Daniel, Jens, Liza, Marko, Sofia and Thomas, who always had time to enjoy the life outside of science.

Najbardziej chciałbym podziękować mojej kochanej Rodzinie, która zawsze we wszystkim mnie wspiera i bez której niczego bym nie osiągnął: Annie, Akramowi i Karimowi Armache, a także Danieli i Wojciechowi Talar.

## Summary

Protein biosynthesis, the translation of the genetic code into polypeptides, occurs on ribonucleoprotein particles called ribosomes. Although X-ray structures of bacterial ribosomes are available, high-resolution structures of eukaryotic 80S ribosomes are lacking. Using cryo-electron microscopy and single-particle reconstruction we have determined the structure of a translating plant (*Triticum aestivum*) 80S ribosome at 5.5 Å resolution. This map, together with a 6.1 Å map of a *Saccharomyces cerevisiae* 80S ribosome, has enabled us to model ~98 % of the rRNA and localize 74/80 (92.5 %) of the ribosomal proteins, encompassing 11 archaeal/eukaryote-specific small subunit proteins as well as the complete complement of the ribosomal proteins of the eukaryotic large subunit. Near-complete atomic models of the 80S ribosome provide insights into the structure, function and evolution of the eukaryotic translational apparatus.



## 1 Introduction

### 1.1 The Flow of Genetic Information

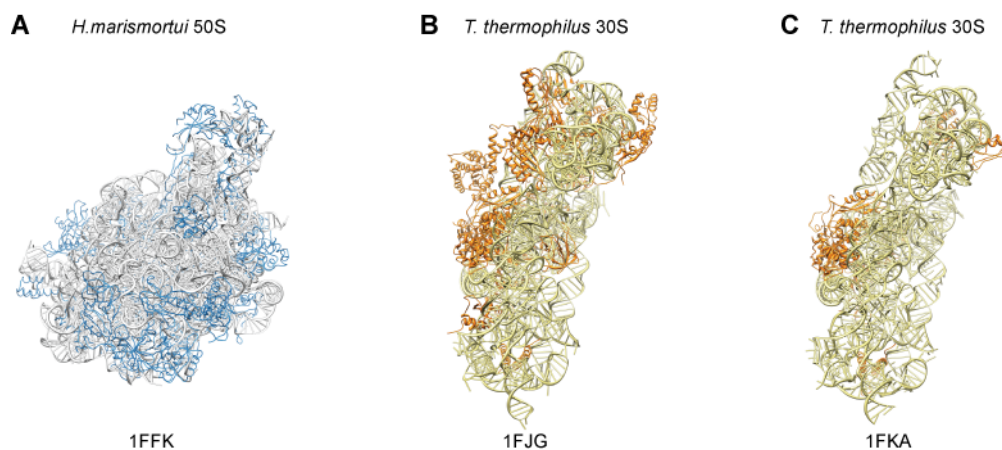
The main aim of all living organisms on Earth is to pass their genetic information to an offspring. Logically, this process cannot be random. Over the course of time, organisms organized themselves to store their basic data in genes – units of heredity information. Genes are bundled together on a larger storage space, a double helix constructed from deoxyribonucleic acid (DNA). This is a stable molecule that ensures safe storage of information in various conditions. Specialized machines, polymerases, are involved in making further copies of DNA in a process called replication. Some other polymerases are involved in assembling (transcribing) an information exchange ticket, a so-called messenger RNA, or mRNA. RNA, short for ribonucleic acid, is a molecule very similar to DNA, and has been speculated to have been the primary form of life, as we know it.

The mRNA is a transient structure on which one or more genes are written. It is then processed by another complex molecular machine, called the ribosome, which translates the message encoded in nucleotide-language to protein-language. This process is very important, since proteins constitute approximately half of the mass of a dry cell (Voet and Voet, 2004). They are therefore highly abundant and perform various tasks, both as free-proteins and as parts of larger assemblies. They take part in transcription, splicing, translation, regulation, transport and many other processes. They are the principle executors of biochemical activity.

These two steps – transcription of DNA into RNA, and then translation of RNA into proteins are essential and present in all living beings discovered to date. They differ in details, however, in essence, they are conserved, which infers a common origin. The most recent ancestor of all living organisms is called the Last Universal Ancestor (LUA) or Last Universal Common Ancestor (LUCA) (Glansdorff et al., 2008; Woese, 1998) and is estimated to have lived some 3.5 to 3.8 billion years ago. The fact that such an entity must have existed was first made clear in the 1960s when the genetic code was cracked and found to be universal. The initial split in lineages led to an enormous diversity, creating three separate domains of life, Bacteria, Archaea and Eukarya. However diverse, they still share enough common traits for the researchers to infer their functions by studying seemingly unrelated organisms.

## 1.2 Ribosomes

In all living cells the translation of mRNA into polypeptides occurs on ribosomes. The mRNA is organized in nucleotide triplets called codons. There are 64 of them, of which 61 encode the 20 standard amino acids, and three remaining ones constitute the stop codons. The decoding is conducted by pairing an anticodon of transfer RNA with a codon of mRNA. Ribosomes provide a platform upon which aminoacyl-tRNAs interact with the mRNA as well as position the aminoacylated-tRNAs for peptide-bond formation (Schmeing and Ramakrishnan, 2009). Ribosomes are composed of two subunits. The small subunit monitors the stereochemistry of the mRNA-tRNA codon-anticodon duplex to ensure fidelity of decoding and translation, while the large subunit contains the active site where peptide bond formation occurs. Both the small and large subunits are composed of RNA and protein: In eubacteria such as *Escherichia coli*, the small subunit contains one 16S ribosomal RNA (rRNA) and 21 ribosomal proteins (r proteins), while the large subunit contains 5S and 23S rRNAs and 33 r proteins. Generally, the large subunit is considered the core of the ancient ribosome, whereas the small subunit is regarded as a later addition.



**Figure 1.** First high-resolution X-ray crystal structures of the ribosomes. (A) Archaeal *Haloarcula marismortui* 50S, with RNA in white and proteins in blue; (B, C) Bacterial *Thermus thermophilus* 30S from groups of V. Ramakrishnan and A. Yonath, respectively, with RNA in yellow and proteins in orange.

In 2009, three scientists were awarded Nobel Prize in Chemistry for studies of the structure and function of the ribosome. Ada Yonath, Venkatraman Ramakrishnan and Thomas Steitz in the so-called “*annus mirabilis*” (Ramakrishnan and Moore, 2001), miraculous year 2000, reported high resolution structures of individual subunits of the prokaryotic ribosomes. The group of Thomas Steitz detailed the 50S structure from archaeal *H. marismortui* at 2.4 Å resolution (Ban et al., 2000; Nissen et al., 2000), while Ramakrishnan and Yonath reported

the 30S structure from bacterial *T. thermophilus* at 3.0 Å (Wimberly et al., 2000) and 3.3 Å (Schlunzen et al., 2000), respectively. Crystal structures of the complete bacterial 70S ribosome were initially reported at 5.5 Å (Yusupov et al., 2001), with an interpretation based on atomic models of the individual subunit structures (Ban et al., 2000; Wimberly et al., 2000), and are now available at atomic resolution (Selmer et al., 2006). These structures have provided unparalleled insight into the mechanism of different steps of translation (Schmeing and Ramakrishnan, 2009) as well as inhibition by antibiotics (Wilson, 2009).

### 1.3 Eukaryotic Ribosomes

Compared to the bacterial ribosome, the eukaryotic counterpart is more complicated, containing expansion segments (ES) and variable regions (VR) in the rRNA as well as many additional r proteins and r protein extensions. Plant and fungal 80S ribosomes contain ~5500 nucleotides (nts) of rRNA and 80 r proteins (see Table S 1 and Table S 2 for r protein nomenclature), whereas the bacterial 70S ribosomes comprise ~4500 nts and 54 r proteins. The additional elements present in eukaryotic ribosomes may reflect the increased complexity of translation regulation in eukaryotic cells, as evident for assembly, translation initiation, development as well as the phenomenon of localized translation (Freed et al., 2010; Richards et al., 2008; Sonenberg and Hinnebusch, 2009; Wang et al., 2010; Warner and McIntosh, 2009). Moreover, many of these Eukaryote-specific components have been associated with human disorders (Freed et al., 2010). Thus, insight into the structure and localization of these elements will be important to expand our understanding of eukaryotic translation regulation as well as diseases.

### 1.4 Ribosome Biogenesis

Genesis, from American Heritage Dictionary of the English Language, is “*the coming into being of something; the origin*”. Ribosome biogenesis is a procedure in which the pre-rRNA is processed, folded and assembled together with the ribosomal proteins. In Eukaryotes, this process is highly conserved and depends on a large number of small nucleolar RNAs and non-ribosomal factors, such as AAA-ATPases, ATP-dependent RNA helicases and kinases (Kressler et al., 2009). It occurs sequentially in the nucleolus, nucleoplasm and cytoplasm.

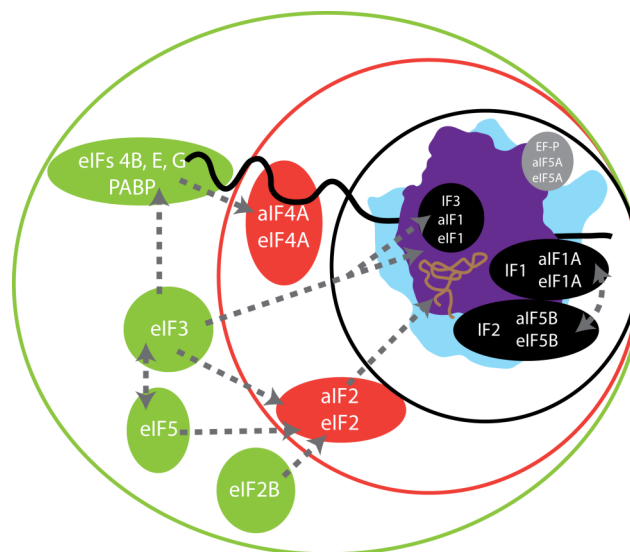
There are hundreds to thousands of genes that encode the pre-RNA. RNA polymerase I transcribes rDNA, which results in a 45S pre-rRNA (Lodish, 2008). Pre-ribonucleic particles

(pre-rRNPs) are formed through binding of proteins to the 45S pre-rRNA. A number of the proteins that form the pre-rRNPs remain associated with the mature ribosomes, however some are only involved in the assembly and restricted to the nucleolus.

45S pre-rRNA is subsequently processed in a series of cleavage and exonucleolytic steps into the mature 28S, 18S, and 5.8S rRNAs; in turn, RNA polymerase III transcribes 5S rRNA genes in nucleoplasm, from where, without further processing, 5S rRNA diffuses into the nucleolus (Lodish, 2008). There, the processed rRNA and ribosomal proteins get assembled into the 40S and 60S subunits, and consequently transported to the cytoplasm through the nuclear pore complexes (NPCs).

### 1.5 Translation Mechanism

There are four main phases through which protein synthesis proceeds: initiation, elongation, termination, and ribosome recycling. Of those four steps, only elongation involves similar factors and mechanisms in Prokaryotes and Eukaryotes, while mechanisms of initiation, termination, and ribosome recycling, as well as the factors involved, are different.



**Figure 2.** Conservation of a core set of translation-initiation factors through evolution in Bacteria, Archaea and Eukaryotes depicted based on their conservation through evolution. The three universally conserved initiation factors IF1/eIF1A, IF2/eIF5B, and IF3/eIF1 are in black and within the black circle; in gray, universally conserved factor EF-P/eIF5A. Existing only in Archaea/Eukarya, the DEAD-box RNA helicase eIF4A and the tRNA delivery factor eIF2 shown in the red circle. In green are the Eukarya-specific factors: eIF3, eIF4 family of factors and the proposed GAP (eIF5) and GEF (eIF2B) for eIF2. The arrows indicate protein-protein, protein-RNA (eIF2-tRNA), or factor-ribosome (eIF3-40S) interactions (from Hinnebusch et al., 2004).

The ribosome has three distinct tRNA binding sites: aminoacyl (A), peptidyl (P) and exit (E).

Initiation of translation is the process of assembly of elongation-competent ribosomes, in which the initiation codon and anticodon loop of initiator tRNA in the ribosomal P-site are base-paired. Even though initiation in Prokaryotes and Eukaryotes differs in mechanistic details, certain common patterns, like start site selection, can be found. As shown on Figure 2, there are three groups of initiation factors (Hinnebusch et al., 2004): (i) the ones present in all three domains; (ii) the ones present in Eukarya and in Archaea, but not in Bacteria; and (iii) specific only to Eukaryotes. The first group consists of three universally conserved initiation factors: IF1 (aIF1A, eIF1A), IF2 (aIF5B, eIF5B), and IF3 (aIF1, eIF1), which interact directly with the ribosome. The proposed grouping of IF3 and eIF1 is based on similar fold and function for the two factors (insuring the accurate Met-tRNA and start site selection in the ribosome P-site). Another factor is a universally conserved EF-P (aIF5A, eIF5A), whose function is still under investigation and might not necessarily participate in initiation. The second group of initiation factors consists of the tRNA delivery factor eIF2 and the DEAD-box RNA helicase eIF4A. The last group is present only in Eukarya: the eIF3 complex that facilitates both mRNA and tRNA binding to the 40S subunit; the eIF4 family of factors that function in mRNA binding; and the proposed GTPase Activity Proteins (GAP) (eIF5) and Guanine Exchange Factor (GEF) (eIF2B) for eIF2 (Hinnebusch et al., 2004).

The next step in the translation mechanism after initiation is called elongation. Messenger RNA is positioned for the next codon to be translated. The initiator tRNA is occupying the ribosomal P-site and the A-site is ready for accommodation of aminoacylated tRNA.

Elongation consists of three steps, namely: (i) decoding of an mRNA codon by the properly base-paired (cognate) aa-tRNA; (ii) formation of a peptide bond; (iii) translocation of the mRNA and tRNA - shifting mRNA by one codon and movement of peptidyl-tRNA from A-site to P-site. The decoding and translocation steps are catalyzed by elongation factors that are similar in prokaryotes and Eukaryotes, but can occur without them at a reduced rate (Cukras et al., 2003; Pestka, 1969; Rodnina and Wintermeyer, 2009).

When the ribosome reaches the end of the coding region and the mRNA termination codon is placed in the decoding site, elongation stops and the ribosome enters the third step in translation, called termination.

Termination differs between Eukaryotes and Prokaryotes. Eukaryotes have only two release factors, eRF1 and eRF3, while Bacteria contain three (Rodnina and Wintermeyer, 2009): RF1

and RF2, as well as RF3, which is not essential. eRF1 recognizes all three termination codons, whereas RF1 recognizes UAG/UAA and RF2 UAA/UGA. In Eukaryotes, eRF3 binds tightly to eRF1 and might enter the ribosome as a complex (Pisareva et al., 2007). Bacterial RF3 increases the rate of RF1/RF2 release from the ribosome, while eukaryotic eRF3 seems to ensure the efficient hydrolysis of peptidyl-tRNA by eRF1 (Rodnina and Wintermeyer, 2009).

The last step of translation is ribosome recycling, a different process in Bacteria and Eukaryotes. In the former, a factor called ribosome-recycling factor (RRF), coupled with EF-G, dissociates the ribosome into 50S and 30S-tRNA-mRNA. IF3 binding catalyses tRNA dissociation and destabilizes mRNA, which gets detached on its own spontaneously (Karimi et al., 1999; Peske et al., 2005). Eukaryotes lack RRF and require several initiation factors (Rodnina and Wintermeyer, 2009), like eIF3, which catalyzes the splitting of 80S ribosomes into 60S subunit and 40S-tRNA-mRNA. tRNA release from the complex is mediated by eIF1, while mRNA is dissociated by eIF3j (however, this might also require other factors).

Without changes in the structure of the ribosome between domains, emergence of novel factors and their binding sites would not be possible. Therefore, to understand the complete mechanism of eukaryotic translation, it is important to elucidate the atomic structure of the eukaryotic 80S ribosome.

### 1.5 Peptide-bond Formation

With respect to catalyzing peptide bond formation, the ribosome is a ribozyme (Cech, 2000), an enzyme deriving its power from ribonucleic acids, not proteins. The region in which this occurs is located on the large subunit and is called Peptidyl Transferase Center (PTC). Available crystal structures from Bacteria and Archaea revealed that it is constructed from RNA alone, with no proteins within a 18 Å radius that could participate in the catalysis (Nissen et al., 2000; Schmeing et al., 2005). As a catalyst, the ribosome increases the peptide-bond formation, achieving a  $2 \times 10^7$ -fold gain over an uncatalyzed reaction, entirely by lowering the entropy of activation (Sievers et al., 2004; Schmeing et al., 2002). It is however, still much less efficient than many protein enzymes, which can accelerate reactions up to  $10^{23}$ -fold.

Ribosomes ready for translation carry an initiator peptidyl tRNA in the P-site. New aminoacylated tRNA enters the T-site of the ribosome and pairs with mRNA. If an mRNA codon is properly base-paired with aa-tRNA anticodon, then a strong interaction occurs and aa-tRNA

moves in to the A-site. The A- and P-site 3'-terminal CCA residues are held in position through interaction with the large subunit rRNA. Peptide bond formation involves the nucleophilic attack of the alpha-amino group of the A-tRNA on the carbonyl-carbon of the P-tRNA. The outcome being that the peptide attached to the P-site tRNA and the amino acid in the A-tRNA form a peptide-bond, with the chain now attached to the A-site tRNA. The ester bond between the nascent chain and the P-site tRNA is now gone. After peptide-bond formation, the ribosome contains a peptidyl tRNA in the A-site and a deacylated tRNA in the P site. Following translocation, the P-site tRNA is relocated to the E-site and the A-site tRNA to the peptidyl P-site. Now the A-site is free to bind a new aa-tRNA at the A-site dependent on the codon of the mRNA displayed there.

### 1.6 Ribosomal Proteins

After the discovery of ribosomes and their function in protein synthesis, there was little surprise in the fact that they contained proteins. It was generally accepted that proteins are responsible for cellular enzymatic activity. If so, then it comes only natural to assume that it is up to proteins to catalyze peptide bond formation. However, the discovery of ribosomal RNA disturbed this easy assumption and gave birth to several theories (Moore and Steitz, 2002), such as: (i) rRNA is inactive and serves only as a scaffold; (ii) rRNA determines sequences of proteins to be constructed; (iii) rRNA itself is active and participates in protein synthesis. This last hypothesis was not easy to accept, since it challenged the belief that enzymes are only proteins. However, after the discovery of RNA that could catalyze chemical reactions (Cech et al., 1981; Guerrier-Takada et al., 1983), and with some other biochemical and genetic evidence (Gesteland and Atkins, 1993; Noller, 1991; Zimmermann and Dahlberg, 1996), this theory gained more ground.

At the beginning of the XXI century, a number of ribosomal structures appeared that left little to speculation. The ribosome emerged as a ribozyme (Cech, 2000), an enzyme built of ribonucleotides. It was proven that it is the ribosomal RNA that catalyzes the peptide-bond formation. So, why would the ribosome contain proteins?

In the 50S subunit of *Haloarcula marismortui*, proteins function primarily to stabilize inter-domain interactions that are necessary to maintain the subunit's structural integrity (Klein et al., 2004) or are instrumental in assembly (Rabl et al., 2011). It can be argued that this is a widespread notion in the ribosome across the three domains of life.

R proteins have been shown to carry different tasks: several of them possess distinct extra-ribosomal functions in apoptosis, DNA repair and transcription (Lindstrom, 2009); they also participate in ribosome biogenesis (Martin-Marcos et al., 2007; Moritz et al., 1990; Poll et al., 2009; Robledo et al., 2008), serve as signal transmitters (Gregory et al., 2009), factor binding partners (Andersen et al., 2006; Becker and al., 2011; Gao et al., 2009; Schmeing et al., 2009; Spahn et al., 2004b) and regulators (Sengupta et al., 2004). It is therefore very important to understand their structures, point out their interactions on the ribosome and the reasons why they were incorporated in the first place.

### 1.7 Structures of Eukaryotic Ribosomes - Introduction

Early models for Eukaryotic ribosomes were derived from electron micrographs of negative-stain or freeze-dried ribosomal particles and localization of r proteins was attempted using immuno-electron microscopy (EM) and cross-linking approaches (see for example Gross et al., 1983; Lutsch et al., 1990; Marion and Marion, 1987; Pisarev et al., 2008). The first cryo-EM reconstruction of a eukaryotic 80S ribosome was reported for wheat germ (*Triticum aestivum*) at 38 Å (Verschoor et al., 1996). Although reconstructions from a variety of different species with increasing resolution have been reported subsequently, few have attempted to present accompanying molecular models: Initial core models for the yeast 80S ribosome were built at 15 Å resolution (Spahn et al., 2001) by docking the rRNA structures of the bacterial small 30S subunit (Wimberly et al., 2000) and archaeal large 50S subunit (Ban et al., 2000), as well as docking of corresponding homology models of the r proteins. Recently, reconstructions at about 9 Å resolution of fungal and dog 80S ribosomes were used to extend the molecular models to include rRNA expansion segments (Chandramouli et al., 2008; Taylor et al., 2009).

### 1.8 Aim (Outline of the Thesis)

Despite a huge effort by the ribosomal community to obtain a complete structure of a eukaryotic ribosome, for years it was an insoluble task. Even though the crystals existed, the X-ray crystallographers were not able to resolve their phases to delineate the molecular structure. This is why it was reasonable to turn to cryo-EM. Only after successful publishing of the articles describing the first complete structure of the eukaryotic ribosome by the group of Prof. Roland Beckmann (Armache et al., 2010a, b), did the laboratory of Prof. Marat Yusupov succeed in publishing the results of a 4.15 Å *Saccharomyces cerevisiae* crystal



structure (Ben-Shem et al., 2010). Soon after, the group of Prof. Nenad Ban released a 3.9 Å *Tetrahymena thermophila* structure of a 40S-eIF1 complex (Rabl et al., 2011). They will be compared with our work in the *Conclusions and Discussion* section of this dissertation.

This dissertation is aimed at presenting the best-resolved cryo-EM structure of a translating ribosome to date, work done on localization of novel proteins and extension of known proteins. Later, possible functions and reasons for existence of the novel proteins and protein extensions that are specific to eukaryotic ribosomes are described.

## 2 Materials and Methods

### 2.1 Sample Preparation

A number of samples was prepared to conduct the research. In addition to *Triticum aestivum* sample preparation, knockout *Saccharomyces cerevisiae* strains and ribosomes from four different archaeal species were obtained. This was achieved through a combined effort of many people. *T. aestivum* samples were prepared by Dr. Shashi Bhushan and Dr. Soledad Funes, ribosomes from knockout *S. cerevisiae* strains were prepared by Gülcin Dindar under the supervision of Dr. Birgitta Beatrix and Andreas Anger; the archaeal cells were grown and supplied by Dr. Michael Thomm and ribosomes were purified in Dr. Daniel Wilson's laboratory by Dr. Viter Marquez.

The sample preparation will be described below.

It should be noted however, that this project in addition to obtaining *T. aestivum* cryo-EM reconstruction and building a molecular model, also involved creating a model of *S. cerevisiae* ribosome; the sample and a high resolution reconstruction of a *S. cerevisiae* ribosome were obtained by Dr. Thomas Becker and will not be described in the *Materials and Methods* section (for further information, see: Becker and al., 2011; Becker et al., 2009). In addition, *Pyrococcus furiosus* ribosomes were purified and reconstructed by Sibylle Franckenberg, and used in protein localization (Franckenberg, S., unpublished data).

#### 2.1.1 *Triticum aestivum* Sample Preparation (from Halic et al., 2004)

To generate purified RNCs, we used a homemade wheat germ in vitro translation system (based on Erickson and Blobel, 1983) programmed with truncated mRNA encoding the 120 N-terminal amino acids of DPAP-B from *Saccharomyces cerevisiae*. A DNA fragment with N-terminal histidine and HA tags were generated by polymerase chain reaction from yeast genomic DNA using forward (5' – taatcgactcactatagggaccaaaacaaaataaaacaaaaacacaat-gtctcatcatcatcatcattaccatagatgttccagattacgctgaaggtggcgaagaagaagttg – 3') and reverse (5' - ttgcagctcgtgatattgggatg – 3') primers. Uncapped mRNA was then synthesized from the PCR fragments using T7 RNA polymerase. For translation, six 200- $\mu$ l reactions were incubated for 30 min at 30°C and terminated with 2  $\mu$ l of 10 mg ml<sup>-1</sup> cycloheximide.

Reactions were spun through a high-salt sucrose cushion (50 mM Tris-HCl (pH 7.0), 500 mM potassium acetate, 25 mM magnesium acetate, 2 mM dithiothreitol (DTT), 1 M sucrose and

10 mg ml<sup>-1</sup> cycloheximide) at 355000 g for 45 min. Pellets were resuspended in ice-cold 250 buffer (50 mM Tris-HCl (pH 7.0), 250 mM potassium acetate, 25 mM magnesium acetate, 0.1% (w/v) Nikkol, 5 mM  $\beta$ -mercaptoethanol, 10  $\mu$ g ml<sup>-1</sup> cycloheximide and 250 mM sucrose) and transferred to 1.5 ml of Talon metal-affinity resin (Clontech). The resin was washed with 8 ml of 250 buffer, and 2 ml of 500 buffer (250 buffer containing 500 mM potassium acetate). RNCs were eluted with 100 mM imidazol (pH 7.1) in 250 buffer and spun through 500 ml of a high-salt sucrose cushion. The resulting pellet was slowly resuspended in G buffer (20 mM Tris-HCl (pH 7.0), 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 125 mM sucrose, 100  $\mu$ g ml<sup>-1</sup> cycloheximide, 0.05% (w/v) Nikkol and 0.03% (w/v) of an EDTA-free complete protease inhibitor pill (Roche) and 0.2 U  $\mu$ l<sup>-1</sup> RNasin (Ambion)), flash frozen and stored at 280 8C. From 1.2 ml of translation reaction, RNCs with an absorbance of 0.7 at 260 nm (0,15 pmol) were isolated (Beckmann et al., 2001).

### 2.1.2 *Saccharomyces cerevisiae* Knockout Strains Sample Preparation

For the purification of 80S ribosomes from r protein knockout strains, yeast cultures were grown to an OD<sub>600</sub> of 1.5-2.0 in YPD medium at 37°C. Cells were harvested by centrifugation for 10 min in a Sorvall SLC-6000 rotor at 4500 rpm, washed once with water followed by an additional wash step with 1% (w/v) KCl solution. Yeasts were resuspended in 100 mM Tris/HCl pH 8.0, 10 mM DTT and incubated at room temperature for 15 min. Cells were pelleted by centrifugation for 5 min at 5000 rpm in a Sorvall GS-3 rotor before resuspension in lysis buffer (20 mM Hepes/KOH pH 7.5, 100 mM KOAc, 7.5 mM Mg(OAc)<sub>2</sub>, 125 mM sucrose, 1 mM DTT, 0.5 mM PMSF supplemented with EDTA-free protease inhibitors (Roche)). Yeasts were lysed by using a Microfluidizer (Microfluidics M-110L Pneumatic) at 20,000 psi. The lysate was cleared by centrifugation for 15 min at 15500 rpm in a Sorvall SS-34 rotor, followed by a centrifugation at 100000 g at 4°C for 30 min in order to obtain the S100 fraction.

A crude ribosome fraction was prepared from S100 extracts by spinning through 1.5 M sucrose cushion in 20 mM Hepes/KOH pH 7.5, 500 mM KOAc, 7.5 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 0.5 mM PMSF for 45 min at 100000 rpm in a Beckman TLA-110 rotor. The ribosomal pellet was resuspended in buffer A (20 mM Hepes/KOH pH 7.5, 500 mM KOAc, 5 mM Mg(OAc)<sub>2</sub>, 125 mM sucrose, 1 mM DTT, 0.5 mM PMSF). Ribosomes were incubated on ice for 30 min under high salt / puromycin conditions (20 mM Hepes/KOH pH 7.5, 500 mM KOAc, 12.5 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 1 mM puromycin). Crude ribosomes were

subsequently loaded on 10% to 40% (w/v) linear sucrose gradients in 20 mM Hepes/KOH pH 7.5, 100 mM KOAc; 5 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 0.5 mM PMSF for further purification. The gradients were centrifuged for 4 h at 40000 rpm in a SW40 rotor (Beckman), followed by collection in 750 µl fractions while continuously recording of the absorption profile at 254 nm wavelength (A<sub>254</sub>). 80S ribosome fractions were pooled according to the A<sub>254</sub> profile and pelleted by centrifugation (100000 rpm, 45 min, 4°C, TLA110 rotor). The supernatant was quickly aspirated off and the ribosomal pellet re-suspended in 20 to 40 µl buffer B (20 mM Hepes/KOH pH 7.5, 100 mM sucrose, 50 mM KOAc; 7.5 mM Mg(OAc)<sub>2</sub>, 1 mM DTT, 0.5 mM PMSF). The ribosome concentration was determined photometrically by absorption at 260 nm (A<sub>260</sub>). Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

### 2.1.3 Archaeal Ribosomes Sample Preparation (from Márquez et al., 2010)

Cells of four different species of Archaea, representing two main phyla: Crenarchaeota (*Pyrobaculum aerophilum*, *Sulfolobus acidocaldarius* and *Staphylothermus marinus*) and Euryarchaeota (*Methanococcus igneus*) were obtained from the laboratory of Dr. Michael Thomm. The following procedure was then applied to obtain 50S (all Crenarchaeota species) and 70S (*M. igneus*) ribosomal particles by Dr. Viter Marquez.

Cell pellets were dissolved in Tico buffer (20 mM Hepes pH 7.5, 10 mM Mg(OAc)<sub>2</sub>, 30 mM NH<sub>4</sub>OAc, 4 mM β-Mercaptoethanol) at 4°C and subsequently disrupted by using a Microfluidizer (Microfluidics M-110L Pneumatic) at 18000 psi. The crude homogenate was centrifuged twice at 30000 g at 4°C for 30 min in order to obtain the S30 fraction. A crude ribosomal fraction was obtained by centrifugation at 100000g for 5 h at 4°C and dissolving the pellet in 1 volume of high salt wash (HSW) buffer (20 mM Hepes, 10 mM Mg(OAc)<sub>2</sub>, 500 mM NH<sub>4</sub>OAc, 4 mM β-Mercaptoethanol, pH 7.5). Large debris was removed by centrifuging the crude ribosomes for 5 min at 18000 g at 4°C. The clear supernatant was diluted 10-fold in HSW buffer and layered on top of 1.3 volumes of 25% (w/v) sucrose cushion prepared in HSW buffer and centrifuged at 100000 g for 7 h at 4°C. The pellet was resuspended in a minimal volume of Tico buffer and subsequently purified using a sucrose-density gradient centrifugation (10-40% sucrose in Tico buffer) at 46000 g for 17 h at 4°C. Fractions corresponding to the 50S (*P. aerophilum*, *S. acidocaldarius* and *S. marinus*) were pooled and pelleted at 140000 g for 12 h at 4°C (70S *M. igneus* at 140000 g for 2 h at 4°C) and resuspended in a minimal volume of Tico buffer.

## 2.2 Cryo-Electron Microscopy

Cryo-electron microscopy is a technique of studying a biological sample at low temperatures in vitreous ice on a transmission electron microscope. It allows observation of specimens that have not been fixed or stained, and presents an immediate advantage over X-ray crystallography in presenting a sample in its near-physiological state. The quality of the three-dimensional reconstructions has been steadily improving in the recent years. The usefulness of this technique has been proven repeatedly – it has been used for discovering the principles of binding of various ribosomal factors and ribosome-associated complexes (Andersen et al., 2006; Becker and al., 2011; Becker et al., 2009; Beckmann et al., 2001; Halic et al., 2004; Schuler et al., 2006; Sengupta et al., 2004); it was irreplaceable in studies on the nascent chain (Bhushan et al., 2010a; Bhushan et al., 2010b; Bhushan et al., 2010c; Seidelt et al., 2009) and the dynamic behaviour of the ribosome (ratchet) (Frank and Agrawal, 2001; Bhushan et al., 2010b); allowed for visualization of viral particles (Liu et al., 2010; Sachse et al., 2007) and, as one will learn from this dissertation, was the mean of obtaining the first complete model of a eukaryotic ribosome (Armache et al., 2010a, b).

The strength of this technique, apart of visualizing the particle in its native state, lies in the quantity of the sample. In comparison to other techniques, only tiny amounts of sample (picomol scale) are needed.

To obtain an image on a cryo-EM microscope, the sample has to be frozen on a holey carbon-coated grid. To ensure a thin vitreous ice layer, it has to be blotted before freezing. Then, it is flash plunged into liquid ethane to avoid formation of ice crystals. The result is a sample frozen in vitreous ice. The important assumption in Single Particle Reconstruction is that the particles in ice represent a homogenous population, which means that all the multiple copies preferably should be identical. However, there are computational techniques (*in silico* sorting/purification) that allow for separation of different states during 3D reconstruction.

Next, data is collected on a liquid nitrogen-cooled Transmission Electron Microscope (TEM) at low dose conditions. Depending on the acquisition, the data can be recorded either on a CCD or on a SO-163 Kodak film. If digitization is required, the film is scanned and transferred to the computer. Then, computational techniques are applied to derive the 3D orientation of the particles, which in turn leads to the 3D reconstruction of the sample. Iteratively, more precise maps are generated when the search range for the orientation of the particles is narrowed.

### 2.2.1 Sample Preparation

Samples were applied to carbon-coated holey grids as described (Wagenknecht et al., 1988).

All *T. aestivum* micrographs were recorded under low-dose conditions on a Tecnai F30 FEG (Field Emission Gun) electron microscope at 300 kV with low dose conditions, meaning an average of 20 e/Å<sup>2</sup> in a defocus range between 0.93 nm and 4.25 nm. The resulting images were recorded on a Kodak film and scanned on a Heidelberg drum scanner with a nominal pixel size of 1.2375 Å on the object scale.

Micrographs of the *S. cerevisiae* knockout samples, as well as all the archaeal samples were collected on a Tecnai G2 Spirit TEM at 120 kV at a nominal magnification of 90000 using an Eagle 4096 x 4096 pixel CCD camera (FEI), resulting in a pixel size of 3.31 Å pixel<sup>-1</sup>.

### 2.2.2 Image Generation

It can be assumed with a good approximation that the intensity observed in a cryo-EM image is a projection of the 3D Coulomb potential distribution corrupted by the wave aberrations of the objective lens. Image contrast is formed by interference of scattered and unscattered electron waves (Hanszen, 1971).

The Contrast Transfer Theorem states that the Fourier transform of the image is related to the Fourier transform of the object's Coulomb potential, multiplied by the electron microscope's Contrast Transfer Function (CTF) (Frank et al., 1995; Zhu et al., 1997), which in turn is dependent on the defocus settings. It becomes essential to compensate for the effect of the CTF, because biological samples in cryo-EM are weak phase objects with a poor phase contrast, resulting in a low overall contrast, contrary to amplitude objects in negative stain EM.

The position of the first zero of the CTF sets the limit on the resolution, which as mentioned, is determined by the defocus. Because of this, only by collecting data from a defocus range, substantial increase of the resolution can be achieved. If the defocus series are collected the way that the zeros of the corresponding CTFs do not overlap, then it is easy to compensate for the missing information and extend the resolution beyond the first zero of CTF (Zhu et al., 1997).

### 2.2.3 Image Processing

The micrographs were obtained on the microscope either on film or a CCD camera. After the digitization, CTFFIND (Mindell and Grigorieff, 2003) was used to determine the defocus values at which the data was collected. The resulting power spectra were visually inspected in WEB (Frank et al., 1996) and micrographs displaying high drift or astigmatism were not taken into further consideration. Micrographs selected for further processing were loaded into SIGNATURE (Chen and Grigorieff, 2007) and a semi-automated screening for particles was performed. The result of this action, particle coordinate files, were then loaded in SIGNATURE, evaluated and used in SPIDER (Frank, 1996), to perform boxing of the ribosomes. In order to have an initial set of correct particles, visual inspection was performed to eliminate false positives.

For the first alignment step, an existing map was selected to serve as a template. The choice was a map of *Triticum aestivum* solved to 7.4 Å (Halic et al., 2006) (EMD-1217). The particle images and the template were 3-fold decimated, to speed up the first steps of refinement. The particle images were then aligned to 2D projections of the 3D template volume, and assigned based on Cross Correlation Coefficient (CCC). After alignment, the initial back-projection was performed, a volume was created for each defocus range and CTF correction was applied to each of them. By merging all the sub-volumes, a complete ribosomal volume was created. From this step on, a result of each round served as a template volume for the next one. Iteratively, each round consisted of ever more restrictive alignment and a back-projection.

With the improving resolution, particle decimation was changed to two-fold and for the highest resolution, only undecimated images were used for the refinement.

The high-resolution map was obtained by combining several datasets of programmed (containing P-site tRNA) *Triticum aestivum* ribosomes, collected in 2008 and 2009. The reconstruction consisted of eight subsets:

- a. *T. aestivum* + P-tRNA + NC + SRP (Unpublished)
- b. *T. aestivum* + P-tRNA + NC + SRP + SR + Sec61 (Unpublished)
- c. *T. aestivum* + P-tRNA + AAP-NC + Sec61 (Bhushan et al., 2010c)
- d. *T. aestivum* + P-tRNA + CMV-NC + Sec61 (Bhushan et al., 2010c)
- e. *T. aestivum* + P-tRNA + Helix55-NC (In preparation)
- f. *T. aestivum* + P-tRNA + NC + 19Ala (In preparation)
- g. *T. aestivum* + P-tRNA + NC + 7L12Ala (In preparation)
- h. *T. aestivum* + P-tRNA + NC + Helix (Bhushan et al., 2010a)

They were joined together and aligned to the best map resulting from processing of (a). The reconstruction of (a) was started on April 2008, with the dataset consisting of 150.000 particles. The alignment and backprojection were performed very carefully to avoid overalignment and a characteristic ‘spikiness’ that results from that. This reconstruction reached 7.2 Å resolution (0.5 FSC cutoff), with an average S/N. However, it’s features pointed that the map would have a much better resolution had it contained more particles. For this purpose, it was decided that other datasets containing the same features would be added to the first reconstruction, which would serve as a template for alignment in projection matching.

The subsets constituting the 80S *T. aestivum* reconstruction were back-projected also separately. Their overall quality was better than if they had been refined separately, since in each round they were aligned to a map of an improved S/N ratio. It also introduced a direct basis for an independent map comparison, since there was only one template in projection matching.

While this technique seems straightforward, there is a problem that might be of dire consequences and affect the resolution. If the sample is not homogenous enough, one might encounter an obstacle in resolving the sample map to the desired resolution. This is why a so-called ‘sorting’ step is applied, which requires two templates for two parallel reconstructions. The particles are then assigned to be closer to one of the templates based on CCC. This technique was applied here, where the 1.362.920 out of a total of 2,108,230 particles were assigned to belong to one of the groups, the P-site programmed state.

The same scheme was applied to reconstructions of all available archaeal organisms and where necessary, sorting was used in addition. The initial and final number of particles and resolution of obtained reconstructions is summarized in Table 1.

**Table 1.** Summary of the reconstructions

Sample	Resolution, Å	Initial number of particles	Final number of particles
<i>T. aestivum</i>	5.5	2108230	1362920
<i>S. cerevisiae</i> ΔL29e	20.82	7272	7272
<i>S. cerevisiae</i> ΔL38e	21.04	10356	10356
<i>M. igneus</i>	20.7	9525	8932
<i>P. aerophilum</i>	30.1	9903	9183
<i>S. marinus</i>	24.16	11142	11142
<i>S. acidocaldarius</i>	26.06	10300	9301



### 2.2.4 Resolution Determination

In X-ray crystallography, the term ‘resolution’ refers to the highest resolvable peak in the diffraction pattern. The cryo-EM community developed an approach to calculate the resolution of the EM-density maps that would correlate with the X-ray definition.

The resolution in cryo-EM is measured by dividing the datasets into two random subsets, each reconstructed from half of the data and comparing them along shells in 3D Fourier space. The result is called Fourier Shell Correlation (FSC).

The community failed to establish one common way of resolution interpretation, however, with the most often used criterion being the 0.5 FSC cutoff value. This is to be interpreted as the resolution at which Signal to Noise value equals 1 (Frank, 2002) at CCC threshold value of 0.5.

### 2.2.5 Map Deconvolution

Maps were sharpened using a non-negative deconvolution method (Hirsch, Schölkopf and Habeck, accepted) based on the multiplicative updates proposed in (Sha et al., 2007). As a blurring function, an isotropic Gaussian kernel (generated with the EMAN (Ludtke et al., 1999) software package command *pdb2mrc* for a Protein Data Bank file containing a single atom) was chosen. In addition, a non-negative background density was introduced to account for solvent contributions and other artifacts. The background was constrained to be uncorrelated with the deconvolved density map. Both the deconvolved map and the background density were then estimated simultaneously using interleaved multiplicative updates. The deconvolution algorithm was run for different kernel sizes and constraint strengths. The most informative density map was selected by visual inspection.

## 2.3 Modeling

In modeling of proteins into the maps, one starts with a search for an appropriate template. For ribosomal proteins the best-case scenario is to find an NMR or X-ray crystallography template of r protein, in the ribosome-bound or free-state. Comparative modeling is also applicable when one has a template of a more remote homolog of the particular protein, however the result might be less reliable. In addition to the template-based modeling, one can also recruit secondary-structure predictions, as well as electron density-based modeling. In

many cases during this project all of these approaches needed to be combined for modeling of a single protein.

### 2.3.1 Protein Template Sequences

*Saccharomyces cerevisiae* sequences were taken from the UniProt (<http://www.uniprot.org>) database of manually annotated and reviewed sequences. *Triticum aestivum* protein modeling was unfortunately slightly more complex, since only a number of sequences from this organism were available. Because of this, remaining sequences from closely-related organism, *Oryza sativa* were taken instead. To verify the viability of this approach, *T. aestivum* sequences were extracted from KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2006a; Kanehisa et al., 2006b) database and compared to their *O. sativa* counterparts. The results were clear: the proteins have very high similarity (>90% identity on average, see Table S 8). Therefore, at the given map resolution, this proves to be a valid approach.

### 2.3.2 Comparative/Homology Modeling

Polypeptides with similar sequences fold into similar structures. Protein amino acid sequence is less conserved than its tertiary structure. That means that in the course of evolution, proteins in different organism might have undergone dramatic sequence changes (yet still retaining a traceable similarity) but their overall 3D structures remains similar.

The idea behind homology modeling is that by having a solved NMR or X-ray crystallography structures of a protein (*template*), we could predict and build an atomic model of a related structure in a different organism (*target*). It takes time to obtain high-resolution structures using experimental methods, therefore it makes sense to obtain them using template-based modeling.

The quality of the homology model is dependent on the template structure and the quality of the sequence alignment. The quality of the model deteriorates with decreasing sequence identity. If two proteins share a reasonable sequence identity, the other determinant for the homology model quality is the template-target sequence alignment, which means that ‘the better the alignment, the better the model’. Of course, not only the protein sequence content varies between two homologous proteins, but also their length (see Table S 13 and

Table S 14). It is quite often that a gap is introduced; therefore a sequence alignment reveals which part of the target protein can be based on a template and which has to be modeled using other means.

Two main questions naturally emerge when dealing with homology modeling: (a) how to search for the appropriate templates and (b) how to ensure a correct target-template alignment.

Many tools have been developed to deal with the issue of template search. However the one used in this exercise, HHpred (Soding et al., 2005) proved to be the most useful for our purposes.

Based on the crystal structures of the archaeal 50S subunit (Ban et al., 2000) and the bacterial 70S ribosomal structures (Schuwirth et al., 2005; Selmer et al., 2006), it was possible to generate 45 *S. cerevisiae* and *T. aestivum* (or *O. sativa*) homology models (Tables S3–S6). In addition, there are also 14 structures of r proteins obtained from either X-ray or NMR structures in a non-ribosome-associated state (Table 4, Tables S3–S6). Screening of available structures and selecting the best ones was done the basis of both the sequence identity and their fit in the cryo-EM map. Sequence-to-structure matching was performed based on profile–profile alignments (Eswar et al., 2005; Marti-Renom et al., 2004). Alignments were acquired using a number of alignment servers, like ClustalW (Thompson et al., 1994), TCOffee (Notredame et al., 2000), MUSCLE (Edgar, 2004), and Mafft (Kato et al., 2002). Using Modeler (Sali and Blundell, 1993), numerous models were created and ranked based on the discrete optimized protein energy (DOPE) (Shen and Sali, 2006) score. From the top scoring models, two were chosen and rigidly fitted into the EM density using Chimera (Pettersen et al., 2004) and Coot (Emsley and Cowtan, 2004), and the best fit was taken for further refinement. Extended parts of the proteins that did not overlap with a template were truncated at this step. The rigidly fitted proteins were then assessed in their densities and adjusted manually.

### 2.3.3 Secondary Structure Prediction

Secondary structure prediction is a technique that, as the name indicates, predicts the secondary structure of proteins (or RNA) based on their primary structure (sequence). It predicts existence of alpha-helices, beta-strands, turns or disordered regions.

We used a program called Quick2D (Biegert et al., 2006) that accumulates predictions from servers like PSIPRED (Jones, 1999),(McGuffin et al., 2000), JNET (Cuff and Barton, 2000)

or DISOPRED (Ward et al., 2004) and displays them together for comparison. Each amino acid in the primary structure is assigned a confidence of how probable the secondary structure prediction is.

#### 2.3.4 Density-based Modeling with Secondary Structure Prediction

This relatively new approach in cryo-EM has been long used in the X-ray crystallography community for model building. Only in the past few years, cryo-EM has come to terms with a higher resolution structures in which secondary structure elements could be clearly distinguished, some of the viral reconstructions even surpassing the 4 Å resolution (Liu et al., 2010). With this 5.5 Å map, one should not be overconfident when it comes to modeling based only on density; the direction of the proteins and the side-chain location cannot be determined.

Whenever modeling of r protein extensions seemed possible on the basis of information in the cryo-EM map, secondary structure predictions were performed (Biegert et al., 2006), along with search for an appropriate template among existing structures using HHpred (Soding et al., 2005). This information, together with the density information in close proximity to the protein core was examined and, if possible, the extended part was modeled. In cases of ambiguous density, comparison with additional maps (*S. cerevisiae* (see Figure 10), deconvolved *T. aestivum* (Figure 9)) was used. Using this approach, a total of over 2000 amino acids were modeled de novo. The increasing number of modeled extensions allowed for iterative minimization of the amount of available density, thus providing constraints to find additional solutions to RNA modeling and protein localization.

Modeling of helices into rod-like density features and an attempt of beta-sheet modeling into flat non-RNA surfaces must be regarded as tentative. Ideal poly-alanine alpha helices were generated with *Build Structure* option in Chimera and subsequently fit into the density using the same program.

Beta-sheets were generated from existing structures, and mutated using Coot *Mutate residues* function to a poly-alanine sequence. Depending on the density representing the beta-sheet, an appropriate model was chosen and placed.

The fit was then evaluated in Coot and further connections were considered.

### 2.3.5 Rigid Body Fitting

Rigid body fitting is a method of finding the optimal orientation and position of a component in a density map by optimizing its position to enhance their CCC.

For this purpose, we used UCSF Chimera and its inbuilt method *Fit in Map*. It provides a fast local optimization of the fit. This has an advantage of speed over the programs using global optimization, however, it has a downside: the structure has to be initially fit manually or by other means into its approximate location to be further refined.

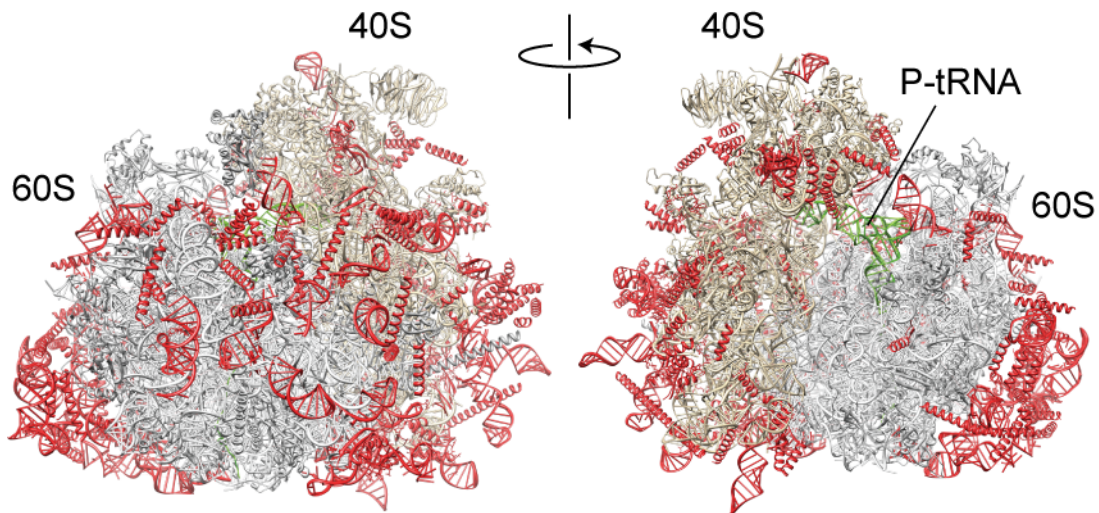
### 2.3.6 Molecular Dynamics Flexible Fitting

Because common methods for protein modeling are, to date, not capable of incorporating EM data or interaction with RNA directly in the modeling process, the proteins still needed to be flexibly fitted into the density and reconciled with RNA models. Thus, subsequent to the fitting and modeling of the rRNA, proteins were introduced in the model using Visual Molecular Dynamics (VMD) (Humphrey et al., 1996) software package. An interactive extended version of Molecular Dynamics Flexible Fitting (MDFF) was used to refine the proteins into the density using default parameters (Trabuco et al., 2008). In regions where the protein density was weak, the location of protein regions was determined by visual inspection, and harmonic constraints to the alpha carbons of those regions were imposed to preserve such location. This process resulted in a rearrangement of the proteins to fit the density, and to resolve protein–RNA and protein–protein clashes while preserving secondary structure. Further MDFF refinement was then applied to the entire 80S model.

### 2.3.6 Modeling Workflow Summary

At the start of the modeling, a considerable amount of time was spent on constructing rough, unreliable models for both proteins and RNA, which served as an aide in estimating the density limits for those entities. Initial homology models for proteins, without elongated eukarya-specific extensions, were rigidly placed in the density. A core RNA model was rigidly placed in the ribosome and A-form helices were put into the helical parts of the density. Clear RNA parts were then filled with RNA elements and adjusted by hand to better fit the density.

This step concluded with the removal of the modeled density from the pool of available one and examining the remains. Next, protein secondary structure elements were modeled into the remaining density. This step concluded with a distribution of alpha-helices, beta-sheets and unfolded loop regions in various parts of the density.

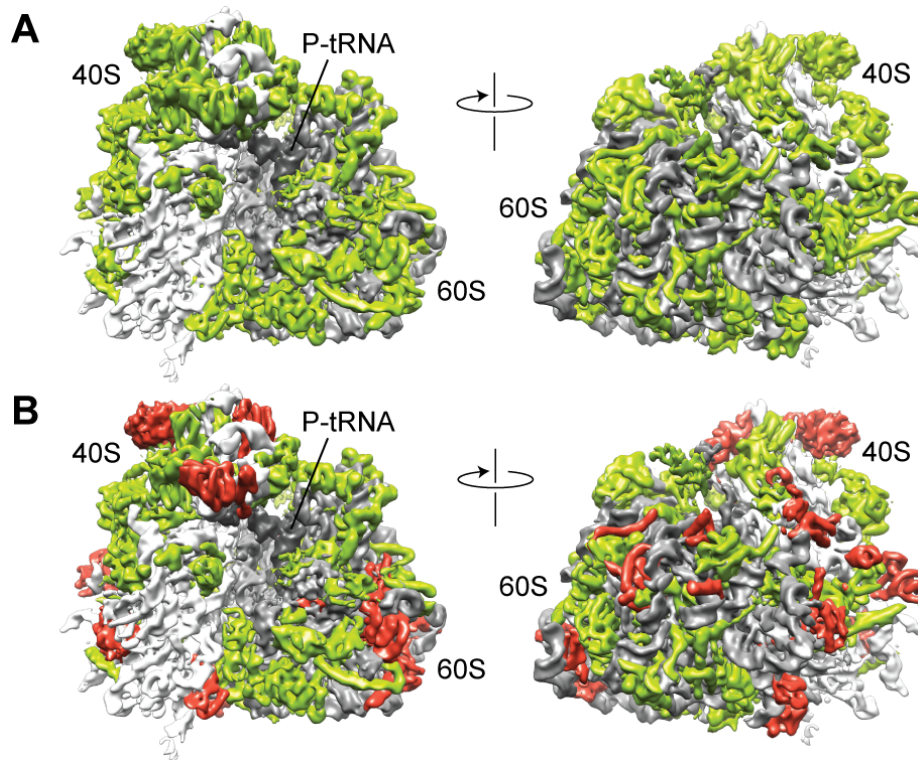


**Figure 3.** Initial model of a *T. aestivum* ribosome, created for the purpose of proper density separation. Core elements of 40S and 60S subunits are colored yellow and grey, respectively, with tRNA in green. Newly placed, Eukaryote-specific rRNA and r protein elements are colored in red.

Those two steps were repeated until little to no density was left. It resulted in a rough estimate of protein and RNA distribution, along with basic unconnected features like A-form helices, alpha-helices, loops and beta-sheets (Figure 3).

The next step involved building proper RNA and protein models. The core RNA model served as an anchor point to further expansion segments, while the initial protein homology models were adjusted into their respective densities. The space around the proteins was examined in search of possible extensions, and if found, they were left in the density.

At the same time, employing various methods mentioned in §3.3 *Protein Localization*, an effort was under way to localize archaeal and eukaryotic proteins that were not found in the previous studies (Ban et al., 2000; Chandramouli et al., 2008; Taylor et al., 2009). With every newfound protein, the search space was growing smaller, which in turn led to refining of protein extensions, RNA model and localization of further proteins (see Figure 4).



**Figure 4.** Cryo-EM map of the *T. aestivum* 80S ribosome, with rRNA colored gray and r protein colored green. (B) Same as A, but with localized r proteins colored red.

## 2.4 Figures

There were a number of tools used for image generation for this dissertation. The majority of the images were prepared using UCSF Chimera (Pettersen et al., 2004), with user-defined color schemes. This is a very fast Python-based visualization tool that has an advantage of user-friendliness and speed in displaying electron-density maps over most other available programs.

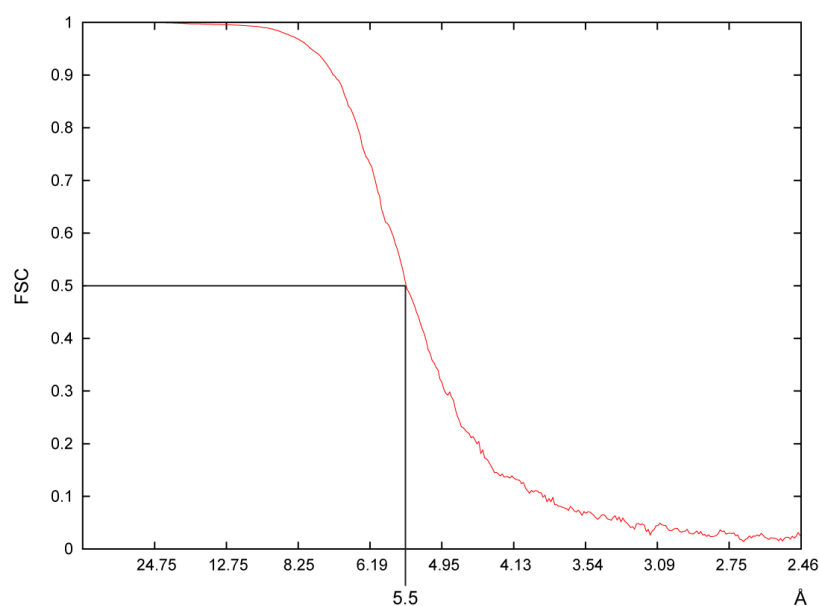
For Figure 18, PyMol (<http://www.pymol.org>), a very common visualization program, was employed. It has the largest number of display options and therefore is very often the preferred tool for image preparation for molecular models without density maps.

Secondary structure diagrams for rRNA (Figure S 1-6) were generated using a Gutell database (Cannone et al., 2002) as templates for *S. cerevisiae* and *O. sativa*. The initial pdf files were converted using a program *pdf2ps* and then imported to Corel Draw. The resulting files were then modified by hand to adjust the sequences and interactions.

### 3 Results

#### 3.1 A Cryo-EM Map of a Eukaryotic Ribosome at 5.5 Å

The final result was obtained by merging eight datasets of programmed (containing P-site tRNA) *Triticum aestivum* ribosomes (see §2.2.3 *Image Processing*). They were joined together, which resulted in a map with very improved Signal to Noise ratio, and a 0.5 FSC cutoff resolution of 5.5 Å (Figure 5). As of February 2011, this is an unprecedented resolution in single particle cryo-Electron Microscopy of asymmetrical molecules.

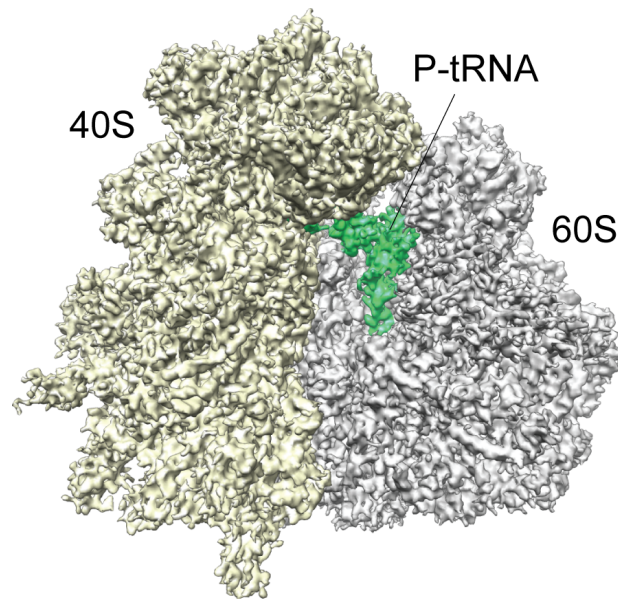


**Figure 5.** Fourier shell correlation (FSC) curve for the cryo-EM reconstruction of the *Triticum aestivum* 80S ribosome-nascent chain complex, with resolution of 5.5 Å according to a cut-off of the FSC at 0.5.

The resulting volume (Figure 6) is a mix of different datasets, which were backprojected all together, as well as separately. The high amount of particles stabilized the reconstruction, and improved the S/N ratio, thus allowing for a better high-frequency information alignment. This also led to the possibility of boosting higher frequencies during alignment, with less concern about aligning the noise.

Even though the stability of the reconstruction was ensured, there is one disadvantage of this approach. The resulting dataset consists of a mixed set of subsets, with different ligands and possible minuscule differences, which affects the homogeneity of the sample, probably limiting its resolution.





**Figure 6.** Cryo-EM reconstruction of eukaryotic 80S *T. aestivum* ribosome.

### 3.1.1 Density

The final resolution of the map reached 5.5 Å and was calculated using a ribosomal mask to eliminate non-ribosomal elements of density. The density represents an 80S ribosome with a tRNA in the P-site and a mix of nascent chains. Generally, the core of the ribosome is subject to the largest stability, with the surface areas, depending on the neighborhood, more or less flexible. If an area is stabilized by protein-protein, protein-RNA or RNA-RNA interactions, the surface areas are nicely solved, and the features are visualized in high detail. Parts like the PTC, ribosomal tunnel, ribosomal tunnel exit site, the large protuberance and many others have a high level of sophistication. However, not all parts of the map are equally well resolved. Flexible parts of the ribosome, like the L1 or P0-P1/P2 stalk, ES27<sup>L</sup> on the large subunit as well as the left and right foot, beak and mRNA exit on the small subunit are not as well resolved as those of the conserved core. They are more ambiguous and therefore more difficult to interpret.

### 3.1.2 Feature resolution

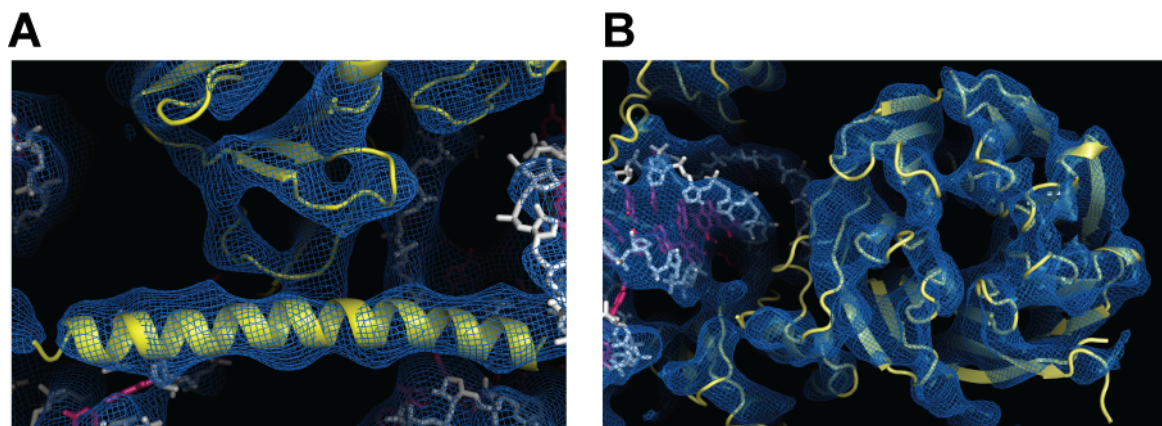
The density map presents a high degree of structural detail, both for proteins and RNA.

### 3.1.2.1 Proteins

For r proteins,  $\alpha$ -helices are observed as rod-like densities and  $\beta$ -sheets are represented by smooth surfaces (Figure 7). Most often, the  $\alpha$ -helix pitch is not visible. There are cases however, in which a ‘bump’ distribution on the density, which would correspond to the helix pitch, is recognizable.

The strands in  $\beta$ -sheets are generally indiscernible, usually represented as smooth surfaces. In rare situations when they are quite well resolved (which usually applies to the conserved core of the ribosome), they appear as surfaces with a characteristic distribution of holes.

Often boosting of the high frequencies of the map is required in order to ascertain the characteristic protein secondary structure elements.

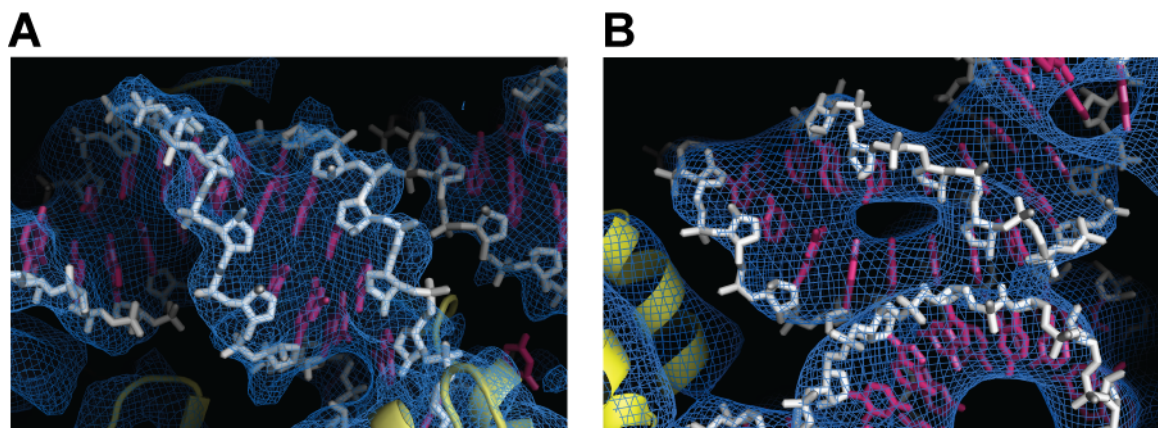


**Figure 7.** Selected views of the *T. aestivum* 80S density map (blue mesh) and corresponding molecular model, with r protein in yellow and rRNA in white (backbone).

### 3.1.2.2 RNA

In many regions, single-stranded rRNA sections are traceable and assignment of bulged nucleotides is possible, as reported previously for the 5.8 Å cryo-EM maps of TnaC-stalled (Seidelt et al., 2009) and 5.6 Å SecM-stalled (Bhushan et al., 2010b) bacterial 70S ribosomes.

RNA helices are clearly visible, containing heavy phosphorus, which strongly scatters electrons during data collection (Figure 8). Also, depending on the flexibility of the region, certain structure junctions can be resolved (see Figure 44).

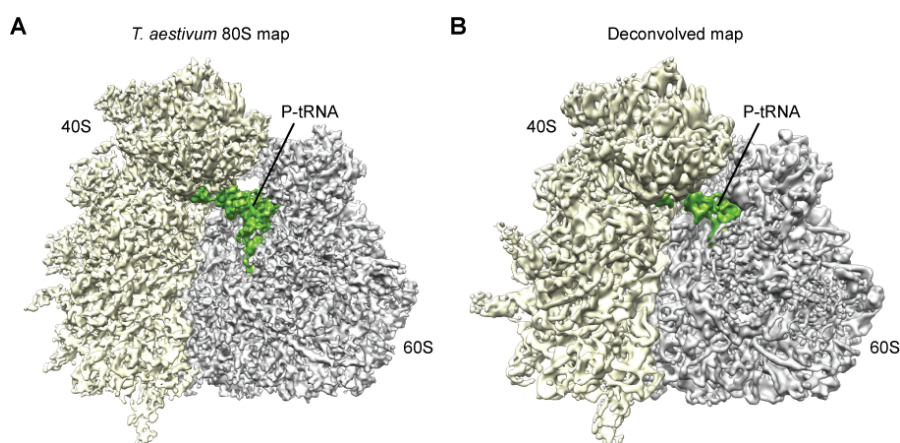


**Figure 8.** Selected views of the *T. aestivum* 80S density map (blue mesh) and corresponding molecular model, with r protein in yellow and rRNA in white (backbone).

### 3.1.2.3 Deconvolution of the Map

In proteins, many loops and extensions are discernible and modeling of connections is possible. The majority however, suffers from distortions and without a technique called deconvolution, would be impossible to even distinguish them from the noise.

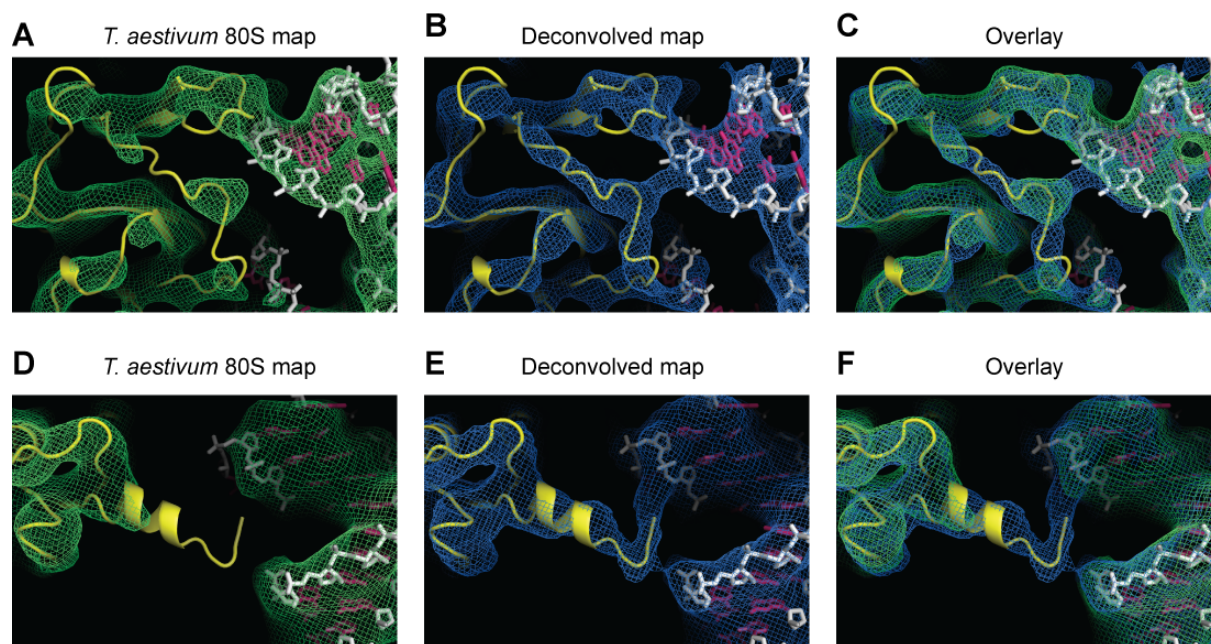
Their deconvoluted maps were generated from the best-resolved map of 80S *T. aestivum* (see Figure 9). The first striking feature is that the maps seem completely devoid of noise, independent of the levels.



**Figure 9.** Comparison of pre- and post-deconvolution sharpened maps. (A) pre- and (B) post-deconvolution cryo-EM maps of the *Triticum aestivum* 80S ribosome, with small and large subunits in yellow and gray, respectively, and P-tRNA colored green.

The connectivity of proteins is boosted, but it comes at a price. The quality of the map is reduced, and it becomes obvious that even though the connectivity is clearer, the secondary

structure protein elements, as well as rRNA, look as if they lost some high frequency information (see Figure 10).



**Figure 10.** Comparison of pre- and postdeconvolution sharpened maps. Examples of (A and D) pre- and (B and E) postdeconvolution maps (mesh), with overlays shown in C and F, respectively. R proteins are shown as yellow ribbons and rRNA nucleotides with white backbone and red bases.

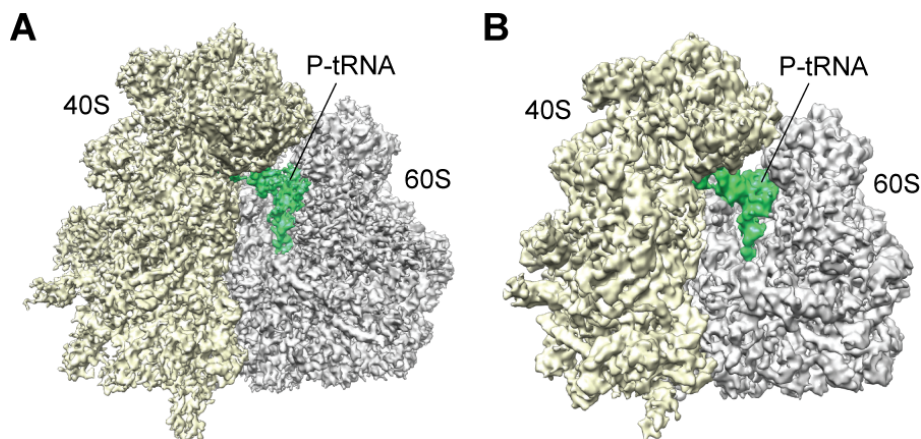
The deconvoluted map becomes useful during modeling when coupled with the non-deconvoluted maps. Alone, it is confusing, since all the possible paths of protein extension receive the same significance, resulting in potential connectivity artifacts. Therefore, they have to be used in conjunction with the unmodified maps.

#### 3.1.2.4 *Saccharomyces cerevisiae* Map Used in Modeling

For the modeling of the yeast ribosome, two cryo-electron microscopy maps of the *S. cerevisiae* programmed 80S ribosomes were used. They were obtained by Dr. Thomas Becker and described in two separate publications (Becker et al., 2009 and Becker and al., 2011). The first map, *S. cerevisiae* 80S-tRNA-NC-Ssh (Becker et al., 2009), later referred to as the Ssh map (EMD-1668), was solved to 6.1 Å, while the second, *S. cerevisiae* 80S-tRNA-NC-Dom34-Hbs1-Sec61 (Becker and al., 2011) was solved to 8.2 Å (Dom34 map). Counterintuitively, the quality of the Dom34 map was in majority of the cases superior to that of Ssh map, especially in the small subunit. The reason for such a discrepancy is that the

former map was reconstructed with a very small number of particles of extraordinary quality, resulting in high detail – but low Fourier Shell Correlation between the half volumes generated from an even smaller number of particles. Both maps were used for modeling of the large subunit proteins and RNA, while for the modeling of the small subunit only the Dom34 map was used. Also, when modeling more flexible parts of the proteins, the Dom34 map was noted for a higher stability and deemed the preferred one, also when compared in some parts to the *T. aestivum* 80S map. The best illustration for this was r protein S28e, localized at the mRNA exit site; its placement was performed based on the Dom34 map and transferred as such to the *T. aestivum* ribosome.

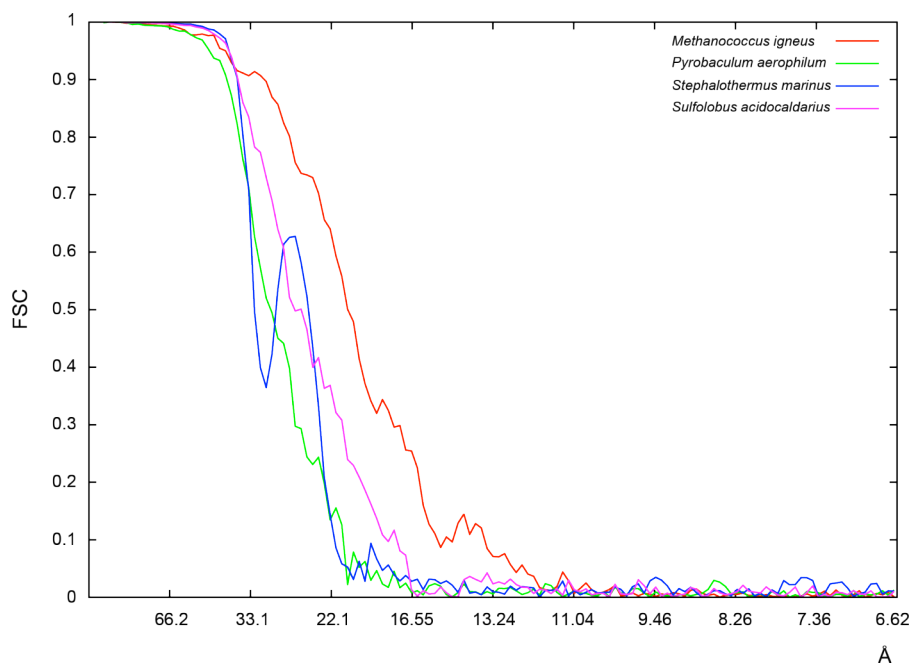
Figure 10 represents a side-by-side comparison of the *T. aestivum* and *S. cerevisiae* Dom34 map.



**Figure 10.** Cryo-EM reconstruction of eukaryotic 80S ribosomes. (A) *T. aestivum* and (B) *S. cerevisiae* 80S ribosomes, with small (40S) and large (60S) subunits colored yellow and gray, respectively and the P-tRNA, green.

### 3.1.2.5 Archaeal Ribosomes

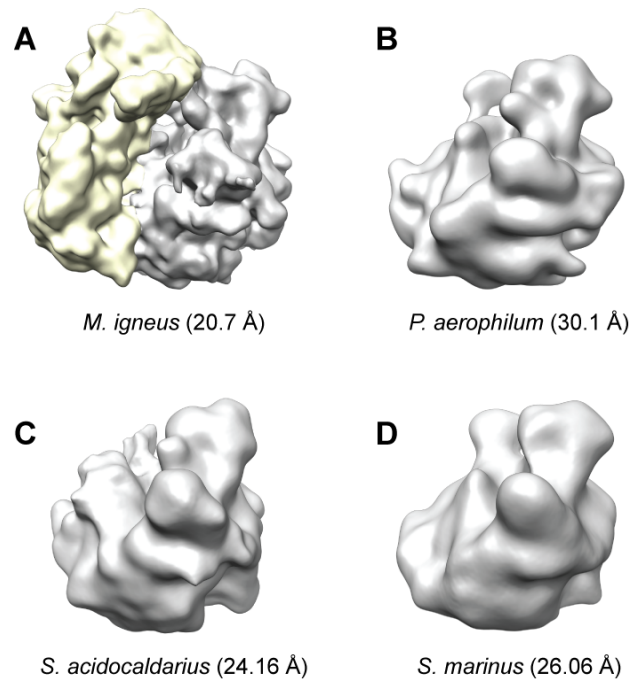
Four available archaeal species were reconstructed to a range of resolutions (see Table 1). Three of those species were available only as 50S subunits, with *Methanococcus igneus* reconstruction being the only one of a complete 70S ribosome.



**Figure 11.** Resolution curves of cryo-EM reconstructions of the archaeal species. The resolutions were based on the Fourier Shell Correlation with a cutoff value of 0.5. *M.igneus* (20.7 Å, red), *P. aerophilum* (30.1 Å, green) (Márquez et al., 2010), *S. marinus* (24.16 Å, blue), *S. acidocaldarius* (26.06 Å, pink) (Márquez et al., 2010).

When examining the resolution curves (Figure 11), two conclusions could be reached: (i) the 70S reconstruction is by far the best of them and (ii) *Staphylothermus marinus* curve has a missing gap. This cavity can be explained as an insufficiency of particles with information between 33 and 27 Å.

The results of these reconstructions are low-resolution density maps of archaeal ribosomes, which at later stages would be used for protein localization.



**Figure 12.** Cryo-EM reconstructions of a 70S ribosome from the Euryarchaeota *M. igneus* (A), and 50S subunits from three Crenarchaeota species: *P. aerophilum* (B) (Márquez et al., 2010), *S. acidocaldarius* (C) (Márquez et al., 2010) and *S. marinus* (D). The small and large subunits are shown in yellow and gray, respectively. The resolutions were calculated at 0.5 FSC cutoff.

### 3.2 Modeling into the Density

The protein modeling effort can be divided into three distinct groups:

- Modeling of conserved core
- Modeling of Eukaryote-specific protein extensions
- Modeling of unlocalized Eukaryote-specific proteins

A different approach was employed for each of those groups. Combined, this resulted in a total of over 5200 amino acids (see: Table S 11 and Table S 12) placed anew: either remodeled, modeled de novo or based on structures not reported in the published crystal structures of the bacterial/archaeal ribosomes (Ban et al., 2000; Schuwirth et al., 2005; Selmer et al., 2006; Yusupov et al., 2001).

#### 3.2.1 Conserved Core

Modeling of the conserved core was rather straightforward, since it employed homology modeling of proteins in a ribosomal environment. The core models were taken from crystal structures of the 50S ribosomal subunit of *Haloarcula marismortui* and 30S subunits of

*Thermus thermophilus* and *Escherichia coli*. The eukaryotic proteins enumerated in Table 2 and Table 3, could be based on the core proteins from those structures. This does not however reflect the fact that some of those proteins might undergo immense refolding (like protein S2p or S4p). This, most often, is caused by the changes in rRNA that are followed by accompanying changes in the protein number, length and structure.

**Table 2.** Small subunit proteins based on the core protein templates

Protein name	Protein family	Template	PDB id
Sa	S2p	<i>Thermus thermophilus</i>	2J00_B
S2	S5p	<i>Escherichia coli</i>	2QAL_E
S3	S3p	<i>Escherichia coli</i>	2QAL_C
S5	S7p	<i>Pyrococcus horikoshii</i>	1IQV_A
S9	S4p	<i>Thermus thermophilus</i>	2J00_D
S11	S17p	<i>Thermus thermophilus</i>	2J00_Q
S13	S15p	<i>Escherichia coli</i>	2QAL_O
S14	S11p	<i>Thermus thermophilus</i>	2J00_K
S15	S19p	<i>Escherichia coli</i>	2QAL_S
S16	S9p	<i>Thermus thermophilus</i>	2J00_I
S18	S13p	<i>Escherichia coli</i>	2QAL_M
S20	S10p	<i>Thermus thermophilus</i>	2J00_J
S15a	S8p	<i>Escherichia coli</i>	2QAL_H
S23	S12p	<i>Thermus thermophilus</i>	2J00_L
S29	S14p	<i>Thermus thermophilus</i>	2J00_N

Nearly all the core proteins in the small subunit are modeled based on either *E. coli* or *T. thermophilus*, with an exception of protein S7p (yeast rpS5), which was based on a non-ribosome-bound model (Hosaka et al., 2001). This template structure contained a domain that the bacterial one was missing. The rest of this template archaeal protein had identical features to the bacterial one. Thanks to this, one could model S7p including the missing domain.

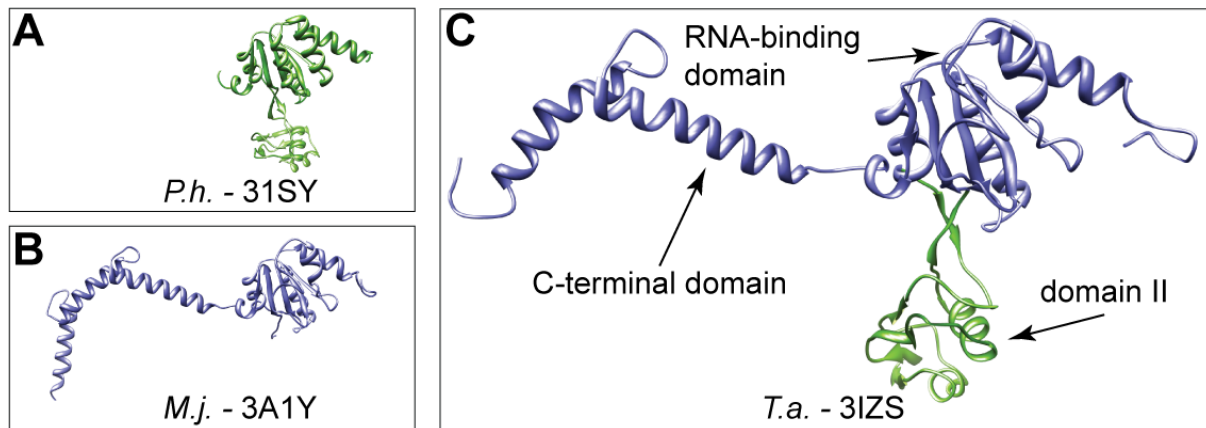
The large subunit was based primarily on the 50S subunit proteins of *H. marismortui*, with 3 exceptions, namely L1p (rpL1), L5p (rpL11) and L10p (rpP0). L1p is a protein that was not visualized in 50S subunit of *H. marismortui*, hence the *T. thermophilus* template was used and placed exactly like in the bacterial structure. L5p was based on two templates, since the archaeal template was incomplete, missing structural information on unresolved loops, which was filled in from the *T. thermophilus* one. L10p on the other hand, is a bit different: it was crystallized in the original structure of *H. marismortui* (Ban et al., 2000), however, it consisted only of two small N-terminal helices. Later, it was refined in (Diaconu et al., 2005)



and (Kavran and Steitz, 2007), but remained incomplete, lacking an important Eukarya-specific domain II (see §4.2.2.9. *P0-P1/P2 Stalk Proteins*) and an entire C-terminal part. Therefore, when two crystal structures in a non-ribosomal state appeared in 2010, they were overlain with the structures mentioned and used to complete the model (Figure 13).

**Table 3.** Large subunit proteins based on the core protein templates

Protein name	Protein family	Template	PDB id
L1	L1p	<i>Thermus thermophilus</i>	2HW8_A
L2	L2p	<i>Haloarcula marismortui</i>	1VQ8_A
L3	L3p	<i>Haloarcula marismortui</i>	1VQ8_B
L4	L4p/L4e	<i>Haloarcula marismortui</i>	1VQ8_C
L5	L18p	<i>Haloarcula marismortui</i>	1VQ8_N
L7	L30p	<i>Haloarcula marismortui</i>	1VQ8_W
L7a	L7ae	<i>Haloarcula marismortui</i>	1VQ8_F
L9	L6p	<i>Haloarcula marismortui</i>	1VQ8_E
L10	L10e	<i>Haloarcula marismortui</i>	3CC2_H
L11	L5p	<i>Haloarcula marismortui</i> & <i>Thermus thermophilus</i>	1VQ8_D & 2J01_G
L12	L11p	<i>Haloarcula marismortui</i>	2QA4_I
L15	L15e	<i>Haloarcula marismortui</i>	3CC2_M
L16	L13p	<i>Haloarcula marismortui</i>	1VQ8_J
L17	L22p	<i>Haloarcula marismortui</i>	1VQ8_R
L18	L18e	<i>Haloarcula marismortui</i>	1VQ8_O
L19	L19e	<i>Haloarcula marismortui</i>	1VQ8_P
L21	L21e	<i>Haloarcula marismortui</i>	1VQ8_Q
L23	L14p	<i>Haloarcula marismortui</i>	1VQ8_K
L24	L24e	<i>Haloarcula marismortui</i>	1VQ8_U
L25	L23p	<i>Haloarcula marismortui</i>	1VQ8_S
L26	L24p	<i>Haloarcula marismortui</i>	1VQ8_T
L27a	L15p	<i>Haloarcula marismortui</i>	1VQ8_L
L31	L31e	<i>Haloarcula marismortui</i>	1VQ8_X
L32	L32e	<i>Haloarcula marismortui</i>	1VQ8_Y
L35	L29p	<i>Haloarcula marismortui</i>	1VQ8_V
L37	L37e	<i>Haloarcula marismortui</i>	1VQ8_1
L39	L39e	<i>Haloarcula marismortui</i>	1VQ8_2
L42	L44e	<i>Haloarcula marismortui</i>	1VQ8_3
L43	L37ae	<i>Haloarcula marismortui</i>	3CC2_Z
P0	L10p	<i>Pyrococcus horikoshii</i> & <i>Methanocaldococcus janaschii</i>	3A1Y_G & 3JSY_A



**Figure 13.** Model of *Triticum aestivum* L10p (C), a composite of two structures: *Pyrococcus horikoshii* L10p containing a crystallized domain II (A) and *Methanocaldococcus janaschii* with the C-terminal domain (B).

### 3.2.2 Unlocalized Eukarya-Specific Proteins

To model the newly localized proteins, several approaches were used:

- Homology modeling based on structures of ribosomal proteins in an un-bound state
- Homology modeling based on non-ribosomal protein templates
- Modeling into the density without a template

Each of them will be further described in the following points.

#### 3.2.2.1 Homology Modeling Based on Existing Structures of Free Ribosomal Proteins

Prior X-ray crystallography and NMR efforts resulted in a number of free-state r protein structures. This makes it possible to use them as templates and, ultimately, fit them into density. Those proteins might however undergo structural rearrangements in order to be accommodated into the ribosome. That means that even though the location of the protein is known, as well as the fold, the exact way in which it interacts with the ribosome is not as straightforward as when modeling the elements of the conserved core.

As one can observe from Table 4, there are 14 such structures, seven in the small and seven in the large subunit, not including P1/P2 dimer. Of those, locations of three were known prior to this investigation: L30e (rpL30) (Halic et al., 2005), S19e (rpS19) (Taylor et al., 2009) and RACK1 (Sengupta et al., 2004).

**Table 4.** Summary for r proteins based on existing free ribosomal proteins

Protein name	Protein family	Template	PDB id
S4	S4e	<i>Thermoplasma acidophilum</i>	3KBG_A
S17	S17e	<i>Methanobacterium thermoautotrophicum</i>	1RQ6_A
S19	S19e	<i>Pyrococcus abyssi</i>	2V7F_A
S24	S24e	<i>Pyrococcus abyssi</i>	2V94_A
S27	S27e	<i>Archeoglobus fulgidus</i>	1QXF_A
S28	S28e	<i>Pyrococcus horikoshii</i>	1NY4_A
RACK1	RACK1	<i>Mus musculus</i>	2PBI_B
L6	L6e	<i>Sulfolobus solfataricus</i>	2JOY_A
L14	L14e	<i>Sulfolobus solfataricus</i>	2JOY_A
L18a	L18ae	<i>Methanobacterium thermoautotrophicum</i>	2JXT_A
L27	L27e	<i>Sulfolobus solfataricus</i>	2JOY_A
L30	L30e	<i>Saccharomyces cerevisiae</i>	1CN7_A
L35a	L35ae	<i>Pyrococcus furiosus</i>	1SQR_A
L40	L40e	<i>Sulfolobus solfataricus</i>	2AYJ_A

It should be noted here that the same template was used for 3 large subunit proteins, namely L6e (rpL6), L14e (rpL14) and L27e (rpL27). This would point to a gene duplication event, which resulted in copies following their own line of evolution. All of them share the same core, however their extensions that could not be based on any template had to be modeled *de novo* into the density. S4e, S19e, S24e, S27e, S28e, RACK1, L30e, L35a and L40e r proteins were modeled with little to no extensions; they were placed into their respective densities and mildly adjusted to reflect the fit. S17e, L18a and aforementioned proteins L6e, L14e and L27e were placed into the density, adjusted and, based on the availability of the options, further extended.

### 3.2.2.2 Homology Modeling Based on Non-Ribosomal Protein Templates

This modeling involves using templates that are not ribosomal proteins, yet retain the structure and sufficient sequence relation to the target molecules. It was used in six cases, with one protein in the small and five in the large subunit (see Figure 14).

Protein S25e (rpS25), located on the head of the small subunit, was modeled based on a structure of a homolog of a transcriptional regulator protein from *Pyrococcus horikoshii* (Okada et al., 2006). The template-target structural alignment (Figure 14F) reveals no differences in the core domain, which consists of helix-turn-helix and KH-domains. The N-terminal end is flipped in respect to the template and harbors an elongated disordered part and

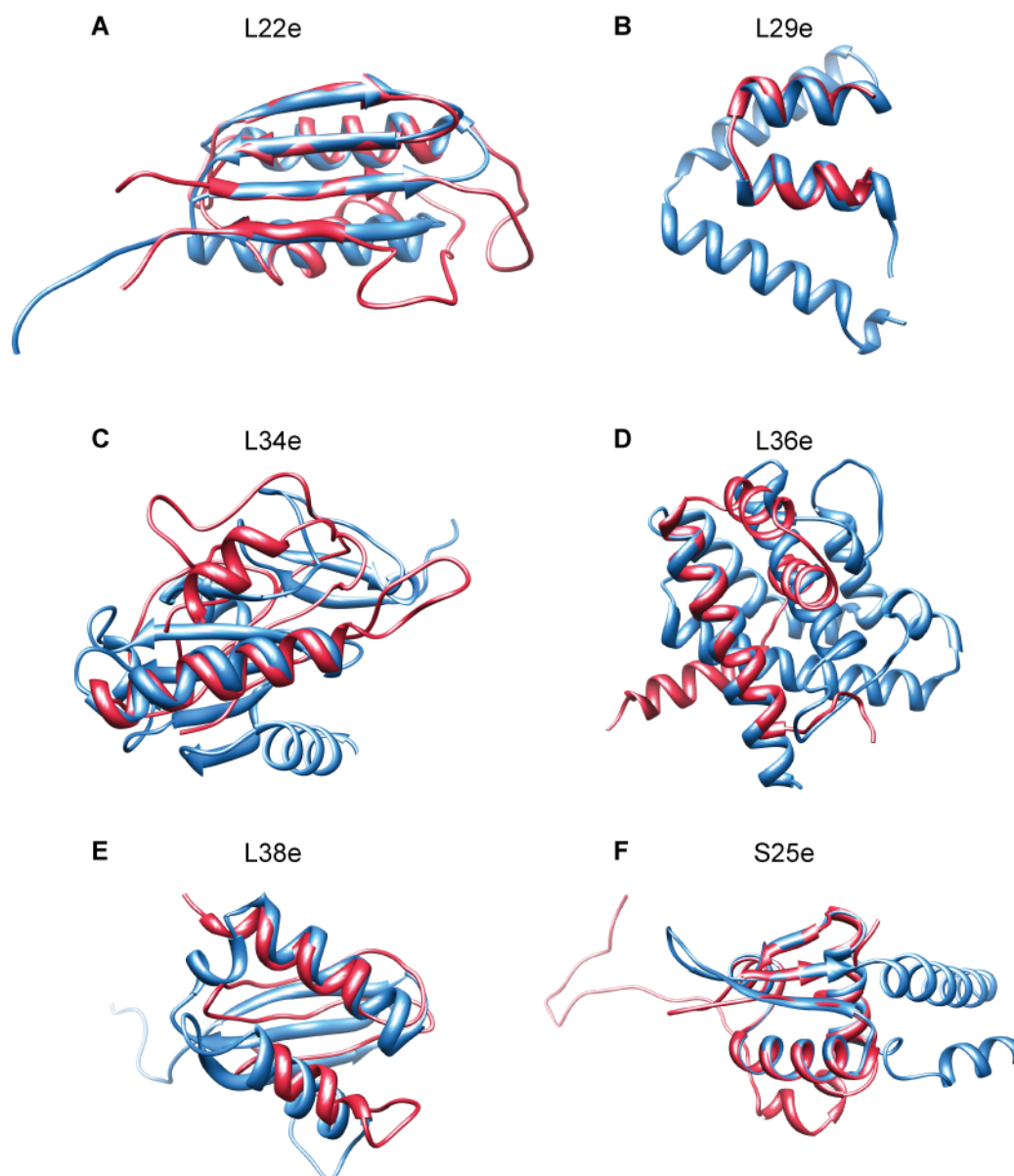
a helix. The disordered part can be only visualized in *T. aestivum* map, there is however, no trace of it in yeast. The carboxy-terminal end in turn is shorter and does not contain the helix present in the template.

**Table 5.** Summary of r proteins based on non-ribosomal templates

Protein name	Protein family	Template	PDB id
S25	S25e	<i>Pyrococcus horikoshii</i>	1UB9_A
L22	L22e	Artificial gene	2KL8_A
L29	L29e	<i>Oryctolagus cuniculus</i>	1UTG_A
L34	L34e	<i>Rhodobacter capsulatus</i>	2PPT_A
L36	L36e	<i>Archeoglobus fulgidus</i>	2OEB_A
L38	L38e	<i>Homo sapiens</i>	1WH9_A

At the bottom of the large subunit, protein L22e was located and modeled. Solution NMR structure of “*de novo* designed ferredoxin-like-fold” protein served as a template. The modeled protein harbors a large resemblance to its template, with a small number of modifications to better fit into the density. In comparison to the template (Figure 14A), the homology model contains four elongated loops, which were manually fitted to the density. The largest difference lies in the rotation of the second helix by approximately 30 degrees in order to properly accommodate it into the density. The N- and C-terminal parts of the protein are predicted to contain disordered regions, and were not revealed to be present during density examination. Those parts were not modeled.

Protein L29e (rpL29) consists of only 59 amino acids in yeast (60 in *T. aestivum*). It was located at the stalk base, in a small RNA pocket. The model was based on an N-terminal part of a crystal structure of uteroglobin (Morize et al., 1987) (see Figure 14B). The model represents the C-terminal part of L29e and closely resembles the template, consisting of a single helix-loop-helix motif. The first 37 amino acids were not modeled due to high disorder in the map in the P0-P1/P2 stalk base region.



**Figure 14.** Panel with protein templates (blue) and homology models based on them (red). The templates are enumerated in Table 5; For detailed description about the modeling, see paragraph §3.2.2.2.

Two models, L34e (rpL34) and L36e (rpL36) were modeled freely based on a template structure (Figure 14C-D). For both of them the homology modeling served as a seed for a following modeling and fitting effort. L34e was based on a crystal structure of thioredoxin-2 from *Rhodobacter capsulatus*, L36e on a crystal structure of gene product Af1862 from *Archaeoglobus fulgidus*. As can be seen on the panel (Figure 14C and D) the models do not really represent their templates, with the exception of certain key features, like a prominent helix in L34e, common N-terminal helix in L36e and it's general helicity. The models have to

be considered as less reliable than those of S25e, L22e or L29e, however more reliable than proteins modeled completely *de novo*.

The last protein modeled using a non-ribosomal template was L38e (rpL38). The template was a solution structure of the KH domain of human ribosomal protein S3p (Figure 14E). When viewed, the general features of the template are noticeable: two helices and a beta-sheet. However, when examining the density, changes had to be introduced, to reflect the position of the loops and the C-terminal helix. Summarizing, the model seems to be of a moderately high reliability, especially when comparing the structures of L38e in *S. cerevisiae* and *T. aestivum*. A loop region between residues 30-41 in yeast is elongated in respect to the same region in *T. aestivum*, and is reflected in the distribution of density in the available maps.

### 3.2.2.3 Modeling into the Density Without a Template

In paragraph §2.3.4 (*Density-based Modeling with Secondary Structure Prediction*), it was explained how modeling was performed when no information could be derived from previously solved structures. During the course of the exercise, densities assigned to the newly localized proteins were filled with the secondary structure elements they represented. If the secondary structure prediction matched the filled density, then the sequence was assigned to the protein, since they were in agreement. Otherwise, the protein was modeled as a poly-alanine chain.

Table 6 presents proteins modeled using this approach. Out of seven proteins, four were modeled as a poly-Alanine chain and the rest with their respective sequences

**Table 6.** Summary of proteins that are modeled *de novo*, without a template

Protein name	Protein family	Template
S7	S7e	Poly-alanine
S21	S21e	Poly-alanine
S26	S26e	Poly-alanine
S30	S30e	<i>de novo</i>
L13	L13e	Poly-alanine
L28	L28e	<i>de novo</i>
L41	L41e	<i>de novo</i>

Little could be done in terms of reliability to model protein S7e (rpS7). There was no available structure, low quality homology models did not fit the density, and in addition, secondary structure prediction-based modeling did not yield results. Therefore, the final model is a density-based approximation, which in no way can be treated as the ‘real’ structure of S7e, rather as a ‘space fill’. The same can be said about protein S21e (rpS21), which was based roughly on the density assigned to it and loosely follows secondary structure prediction. Protein S26e (rpS26) is modeled with slightly higher reliability, since it followed a secondary structure prediction more strictly, in addition fitted well into the density. As for the last small subunit protein modeled this way, protein S30e (rpS30), it fits the density and predictions reasonably well.

In the large subunit, there are three proteins that were modeled using this approach: L13e (rpL13), L28e (absent in *S. cerevisiae* genome (Lecompte et al., 2002), L28 in *T. aestivum*) and L41e (rpL41). Model for L13e fits the density well, which applies partially also to the secondary structure predictions. When both organisms are taken into consideration, one can see the sequence differences between them visualized on the maps. This might indicate that, at least in certain parts, the model could be representative. Protein L28e was modeled based primarily on the initial density assignment. Then, examination of the secondary structure prediction was performed to select for the part that would fit the map. As such, the quality of the model depends highly on the reliability of the localization of the protein. The least problematic of those proteins was protein L41e, predicted as a straight alpha-helix. The part of the map into which it was modeled was represented by a rod-like density, which is in agreement with the ss predictions.

One has to be aware that this technique is rather an approximation and does not yield very reliable models. Since no template for modeling is selected, the directionality of the protein (N- and C-terminus) might be inverted; the secondary structure prediction might be misleading; and the density for the modeling might contain false connectivity or be falsely assigned by the modeling person.

### **3.2.3 Eukaryote-Specific Protein Extensions**

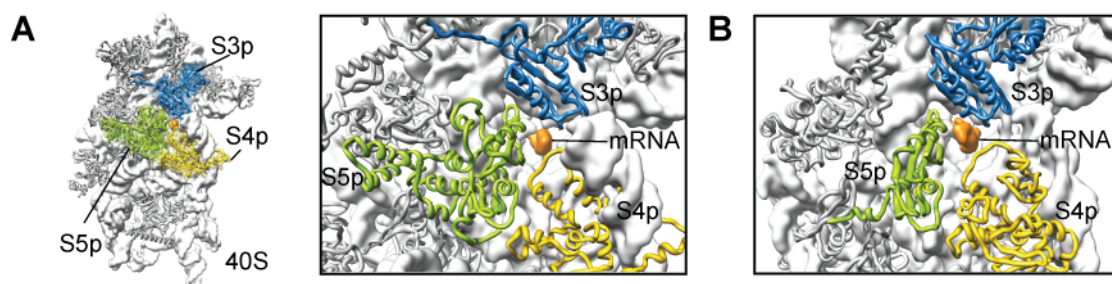
The same methods as described above were applied to modeling Eukaryote-specific protein extensions. These are the parts of eukaryotic proteins that extend beyond their bacterial or/and archaeal templates. Modeling of those extensions should be considered slightly more reliable, since the anchor points and the directionality are known.

Tables S 11-12 present protein extensions, which were modeled into the density.

### 3.2.3.1 Small Subunit

#### 3.2.3.1.1 mRNA Entry

In Prokaryotes, the mRNA entry site is formed by S3p (rpS3), S4p (rpS9) and S5p (rpS2) r proteins (see Figure 15). Small subunit rRNA h16 assumes a closed conformation, with its tip in the proximity of S3p (Ben-Shem et al., 2010). In Eukaryotes, this helix is bent and adopts an open state, leading it away from the subunit body (Spahn et al., 2001). In Bacteria, an S4p domain composed of two alpha helices covers a large part of h16, forming a strong interaction (Jenner et al., 2010; Selmer et al., 2006). This is not the case in Eukaryotes, where the S4p carboxy-terminal domain becomes relocated. Thus, the strong interaction of h16 with S4p in Bacteria is not present in Eukarya and h16 is more likely to rotate around its own base.

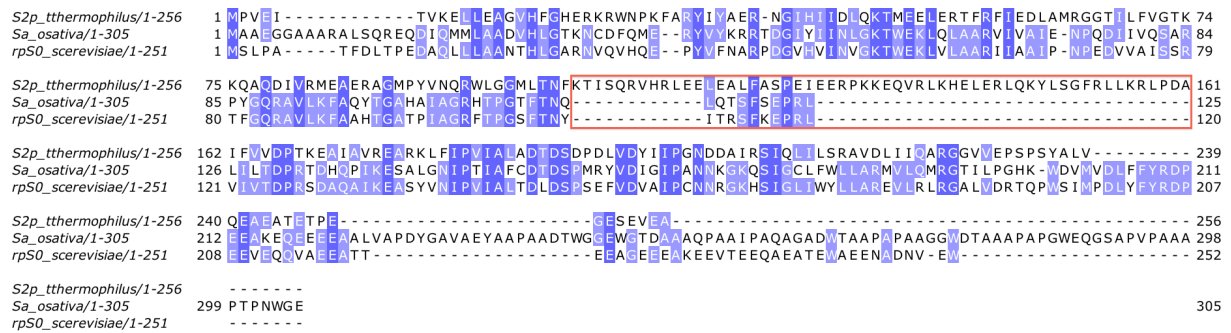


**Figure 15.** (A) Eukaryotic small subunit mRNA entry site overview (thumbnail, Left; zoom, Right). (B) Comparative view of the bacterial 30S subunit mRNA entry site. In A and B, mRNA (orange) is shown for reference.

#### 3.2.3.1.2 mRNA Exit

R protein S2p (rpS0) is an important part of the mRNA exit site. It is present in all three domains, but its structure is different in Bacteria and Archaea/Eukarya. A whole part of the protein, approximately between residues 106 and 161 of *Thermus thermophilus* (and by extension, in Bacteria) is missing in Eukarya (Figure 16). Instead, eukaryotic organisms harbor an extended carboxy-terminus in this ribosomal protein.





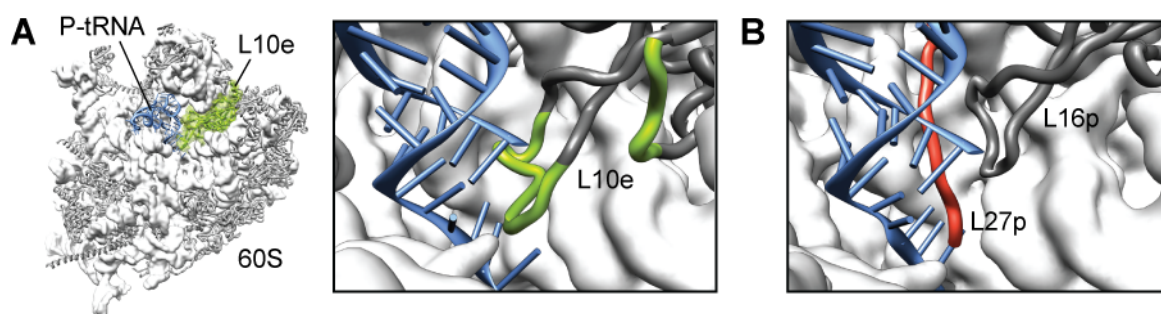
**Figure 16.** Sequence alignment of *T. thermophilus* (Bacteria), *O. sativa* and *S. cerevisiae* (both Eukarya). There is a missing part in the eukaryotic sequences, approximately at residues 106-161 of *T. thermophilus* (marked with red)

This extended part represents a feature that is specific to Eukaryotes in this region of the ribosome - the direct contact between mRNA entry and exit sites (Ben-Shem et al., 2010). This is achieved by a strong interaction between mRNA entry tunnel protein S5p (rpS2) and the aforementioned Eukarya-specific extension of the mRNA exit site S2p (rpS0).

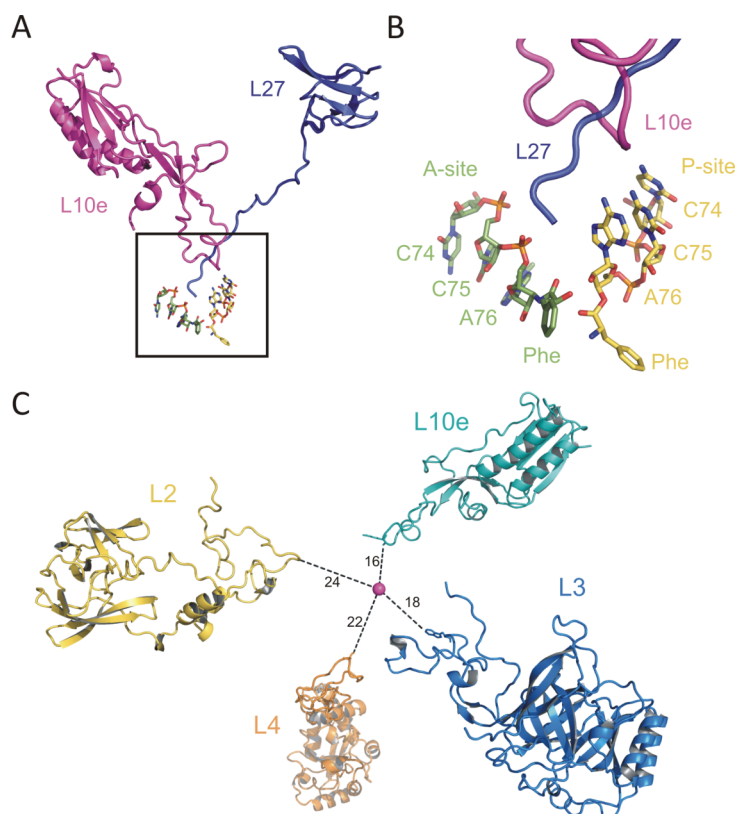
### 3.2.3.2 Large Subunit

#### 3.2.3.2.1 Peptidyl Transferase Center

At the peptidyl-transferase center on the large subunit, direct interaction is observed between the loop of r protein L10e and the CCA-end of a peptidyl-tRNA at the P-site (Figure 17). Based on this model, the loop of L10e is now the r protein region that comes closest ( $\sim 16$  Å) to the site of peptide-bond formation (Figure 18). This loop was disordered and not visualized in the crystal structure of the archaeal 50S subunit (Ban et al., 2000) and is absent in the bacterial homologue, L16p. In bacterial ribosomes, the amino-terminal extension of the r protein L27p occupies a similar but distinct position instead (Maguire et al., 2005; Voorhees et al., 2009) (Figure 17, Figure 18).



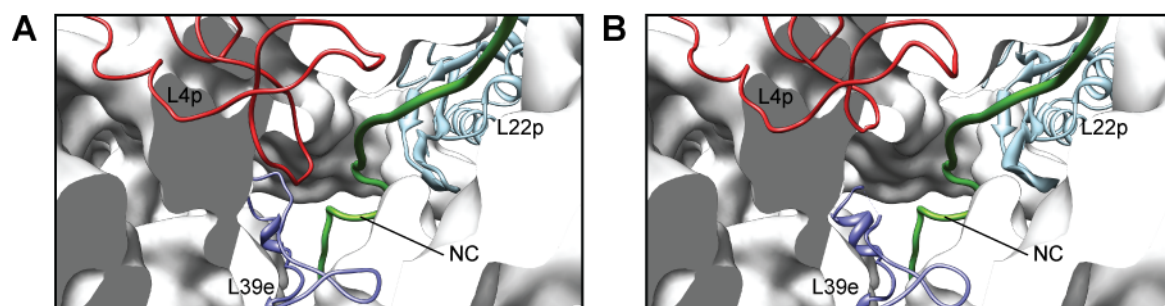
**Figure 17.** (A) Large 60S subunit with Eukaryote-specific extension of L10e (green) highlighted (thumbnail, Left; zoom, Right). (B) Comparative view of the bacterial 50S subunit with bacterial-specific L27p colored red (Selmer et al., 2006). In A and B, the acceptor-stem of the P-tRNA (blue) is shown for reference.



**Figure 18.** Ribosomal proteins that approach the peptidyl transferase center of the ribosome. (A and B) Comparison of the relative positions of the N terminus of bacterial r protein L27 (blue) (Voorhees et al., 2009) and eukaryotic L10e (magenta) with the CCA-ends of tRNA mimics in A- (green) and P site (yellow) (Hansen et al., 2002). (C) R proteins L2p (yellow), L3p (blue), L4p (orange), and L10e (aqua) come within approximately 24, 22, 18, and 16 Å of the site of peptide bond formation, based on ref. Nissen et al., 2000.

### 3.2.3.2.2 Ribosomal Tunnel

The ribosomal tunnel is a universal feature of the ribosome, constituted primarily by ribosomal RNA (Ban et al., 2000; Frank et al., 1995; Halic et al., 2006; Morgan et al., 2000), with extensions of ribosomal proteins L4p (rpL4) and L22p (rpL17) contributing to formation of the tunnel wall. Those extensions form a so-called ‘constriction’ where the tunnel narrows. Near the tunnel exit resides Archaea/Eukarya-specific protein L39e (rpL39), substituting two bacterial-specific protein entities: extended loop of L23p (rpL25) and a protein L34p. Proteins L4p, L22p and L39e were previously crystallized and visualized in Archaea (Ban et al., 2000). In addition, L4p in *T. aestivum* and *S. cerevisiae* contains certain Eukarya-specific features, which we were able to analyze and model. The tunnel part of this protein is elongated in respect to the archaeal and bacterial templates and is located in the loop region interacting with the descending nascent chain (Figure 19). The C-terminal elongated part is situated outside the tunnel and is most probably involved in scaffolding. Protein L22p has the same structure as its template, with a short carboxy-terminal elongation, while L39e has the same structure as in Archaea, except for the ‘tunnel’ part of the protein, which was previously not resolved ((Ban et al., 2000), see Figure 19).

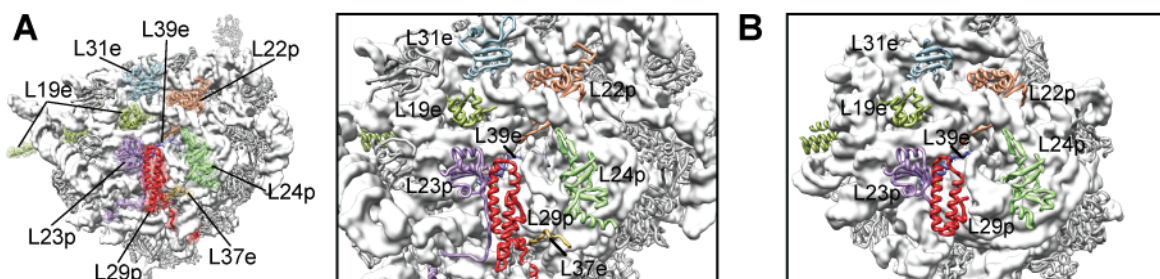


**Figure 19.** Comparison of three proteins in the tunnel in *T. aestivum* (A) and *H. marismortui* (B): L4p (red), L22p (cyan) and L39e (blue) with a nascent chain (NC) in green for overview.

### 3.2.3.2.3 Ribosomal Tunnel Exit Site

The tunnel exit site, with few exceptions, is generally a structurally conserved region of the ribosome. It consists both of proteins and RNA. On the RNA side, it is constructed of H7, H24, H50 and H59, while on the protein one from L19e, L22p, L23p, L24p, L29p and L31e (Nissen et al., 2000). While this also true in Eukarya, a number of extensions of those proteins were modeled (see Figure 20). A protein extension of L37e was found to emerge close to the tunnel exit, on the side of H7. The location of protein L39e should also classify it to be a part of the tunnel-tunnel exit. All proteins have extensions leading away from the exit site, leaving

this spot conserved from Archaea to Eukarya and accessible to complexes such as SRP or Sec61 translocon (Becker et al., 2009; Halic et al., 2006). Protein L19e passes through a large part of the 60S rRNA core and emerges close to the small subunit (to which the extension was assigned before (Chandramouli et al., 2008), forming a SSU-LSU bridge. The extension to r protein L23p was modeled to a limited extent, due to the predicted flexibility of the missing carboxy-terminal. The modeled extension interacts with ES4<sup>L</sup>, ES19<sup>L</sup> and a part of ES31<sup>L</sup>. L24p and L31e have very short extensions, therefore leaving them practically unchanged as compared to Archaea. The extension of protein L29p (about 60 amino acids) interacts with remodeled rRNA stretch, H15-H18, interestingly, a variable part of the core.



**Figure 20.** (A) Tunnel exit site of the large 60S subunit with Eukaryote-specific extension to the proteins highlighted (thumbnail, Left; zoom, Right). (B) Comparative view of the archaeal *H. marismortui* 50S subunit exit site (right) with eukaryotic *T. aestivum* (left)

### 3.2.3.2.4 ES7<sup>L</sup>-ES15<sup>L</sup>-ES39<sup>L</sup>

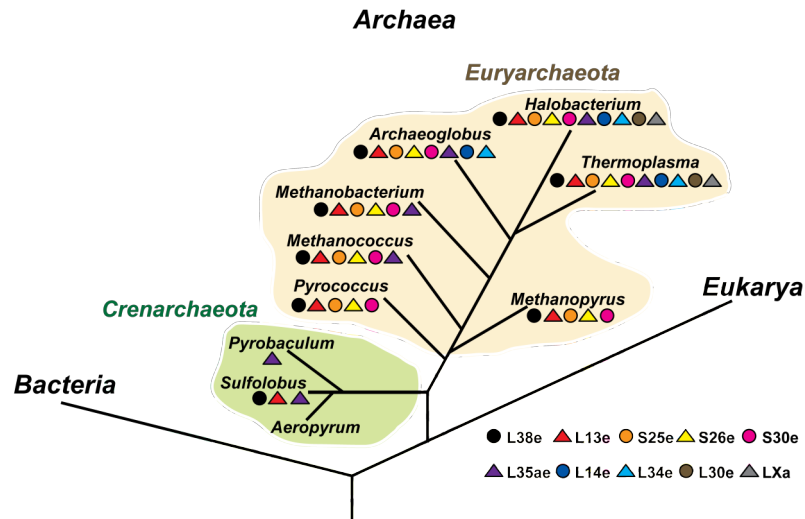
The region neighboring large rRNA expansions, ES7<sup>L</sup> and ES39<sup>L</sup>, and a smaller ES15<sup>L</sup>, is a place of dramatic gain in protein content, both as extended parts of previously known proteins, as well as a number of new ones (see §3.3.2.1 *ES7<sup>L</sup>-ES39<sup>L</sup> Region* and Figure 29). The extensions of four proteins – L4p, L13p, L18e and L30p are modeled completely anew, totaling in this region to almost 270 newly placed amino acids (L4p – 102, L13p – 53, L18e – 18 and L30p – 85). What is most astounding is the fact that those protein extensions are predominantly alpha-helical and in the direct vicinity of rRNA. Extensions of L4p and L30p interact with ES7<sup>L</sup> and ES15<sup>L</sup>, carboxy-terminal part of L13p interacts with a part of ES39<sup>L</sup> and the amino-terminal part of L18e interacts only with ES15<sup>L</sup>. Those parts were created iteratively by extending each of them while comparing the resulting model with a secondary structure prediction. The extension to protein L30p was placed as one of the first, since the prediction was clear – the N-terminal extended part was to be an alpha-helix. The same applies to L13p, whose C-terminal part is almost completely alpha-helical. Small extension of L18e was placed into a free density nearby, while extension to protein L4p was modeled,

starting from the C-terminus. The starting and ending anchor points of this extension were known, the question remained – whether there would be enough space to accommodate around a hundred amino acids. Consulting secondary structure predictions at every step, it was possible to model this part, without the fragment connecting the core of L4p to its extension.

### 3.3 Protein Localization

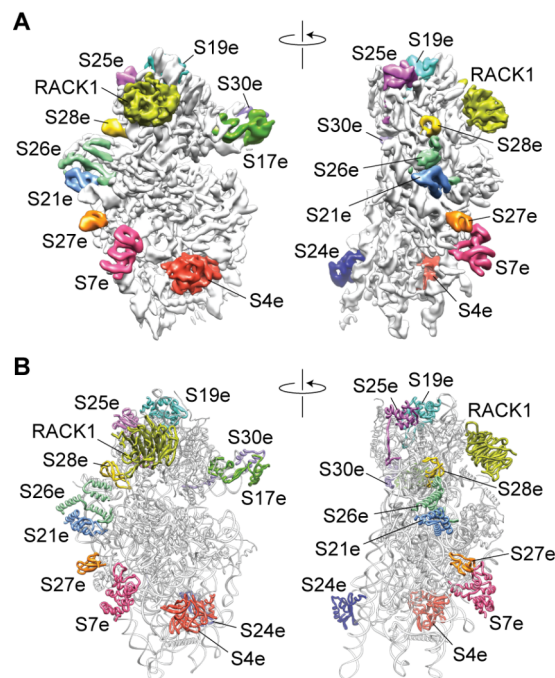
The main basis for the localization of r proteins in the cryo-EM reconstructions of the 80S ribosomes was the excellent agreement between the density features in the maps with distinctive protein-fold characteristics of the X-ray structures and homology models. Additional supporting information was utilized for the localization of r proteins, particularly those modeled *ab initio*.

There is a wealth of low resolution data available on the spatial arrangement of r proteins in eukaryotic ribosomes that derives from a variety of different approaches, such as order of assembly (Ferreira-Cerca et al., 2007), accessibility of particular r proteins to proteolysis, cross-linking (Bulygin et al., 2005; Gross et al., 1983; Nishiyama et al., 2007; Pisarev et al., 2008) (see Figure 23) and immuno-EM studies (Bommer et al., 1991; Marion and Marion, 1987) (see Figure 24, Figure 28). In conjunction with additional analysis, such as matching density features in the cryo-EM maps with (i) distinctive protein-fold characteristics, (ii) species-specific differences in length between r proteins of wheat germ, yeast, and archaeal ribosomes (Lecompte et al., 2002) (see Figure 21), (iii) secondary structure predictions, as well as (iv) knock-out mutants of particular r proteins (Table S 10), it was possible to localize a total of 27 r proteins (excluding P0, P1 and P2) that are not present in the crystal structures of bacterial or archaeal ribosomes. This encompasses twelve small subunit r proteins (Figure 22 - S4e, S7e, S17e, S19e, S21e, S24e, S25e, S26e, S27e, S28e, S30e and RACK1) and fifteen large subunit r proteins (Figure 27 - L6e, L13e, L14e, L18ae, L22e, L27e, L28e, L29e, L30e, L34e, L35ae, L36e, L38e, L40e and L41e).



**Figure 21.** Heterogeneous distribution of archaeal ribosomal proteins (Marquez, 2010, (Márquez et al., 2010) modified from Lecompte et al., 2002.

### 3.3.1 Small Subunit Proteins

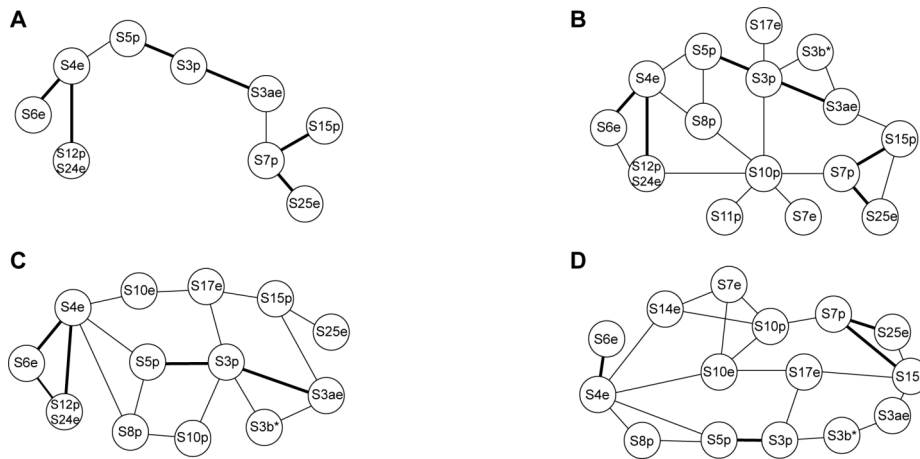


**Figure 22.** Localization of Eukaryote-specific r proteins. Cryo-EM maps of the *T. aestivum* 40S subunit (A) and (B) molecular models of r proteins, with newly identified r proteins colored distinctly.

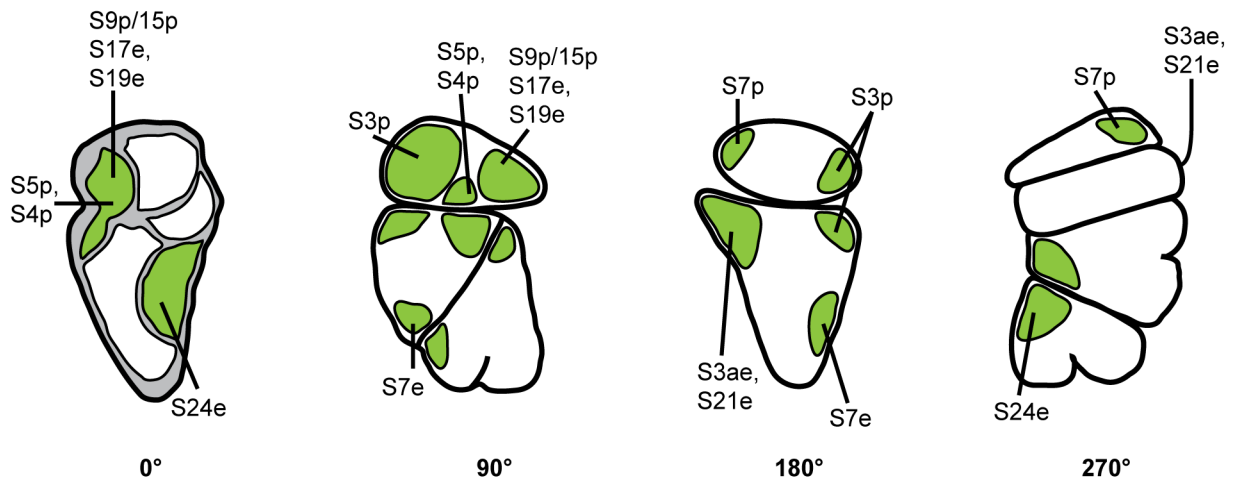
Bacterial ribosomal small subunit contains up to 23 proteins, of which 8 exist only in this domain. Eukaryotic one in turn comprises 33 proteins, of which 13 are Archaea/Eukarya-specific and 5 are found only in the latter. We managed to assign the location of twelve small

subunit proteins, of which 2 are Eukarya-only and the rest exist both in Eukaryotes and Archaea.

The assignment of location of S19e (rpS19) to the head of the 40S subunit was based on the order of assembly of precursor particles (Ferreira-Cerca et al., 2007). Additional data, like the structural information provided by X-ray and NMR, as well as immuno-EM and cross-links (Figure 23, Figure 24), contributed to localization of the other 11 small subunit proteins.



**Figure 23.** Model of the spatial arrangement of proteins within the small ribosomal subunit based on cross-linking and immuno-EM data (adapted from (Gross et al., 1983)). (A-D) represent different views of the SSU.



**Figure 24.** Four different views of the 40S subunit model indicating the location of ribosomal proteins, based on immuno-electron microscopy studies (adapted from Bommer et al., 1991).

### 3.3.1.1 Head of the Small Subunit

The position of RACK1 is in agreement with localization of this protein to the back of the head of the small subunit in fungi and mammalian ribosomes (Chandramouli et al., 2008; Sengupta et al., 2004), however we do not observe any evidence for asymmetry of the RACK1  $\beta$ -propellers (Figure 7B), as suggested previously for the canine 80S ribosome (Chandramouli et al., 2008). Moreover, we can assign the unidentified protein interaction partner of RACK1 as being the Eukaryote-specific C-terminal extension of r protein S2p.

The localization of S19e to the head of the 40S subunit is consistent with biochemical data of assembly precursor particles formed in vivo (Ferreira-Cerca et al., 2007). In addition, an X-ray structure of S19e from *P. abyssi* (Gregory et al., 2007) revealed a unique fold, which immediately pointed to the exact location of this protein.

The loops of S19e located between  $\alpha 1$  and  $\alpha 2$  as well as  $\alpha 4$  and  $\alpha 5$  are disordered in the crystal structure (Gregory et al., 2007), but become ordered upon ribosome binding where they interact with the variable region of helix41 (h41) of the 18S rRNA. Mutations in S19e found in Diamond-Blackfan anemia (DBA) patients are clustered around  $\alpha 3$  (Gregory et al., 2007), which is also seen to interact with h41 in the *T. aestivum* and *S. cerevisiae* 80S models. DBA is an inherited bone marrow failure syndrome that results from defects in ribosomal assembly (Freed et al., 2010).

#### 3.3.1.1.1 mRNA Entry

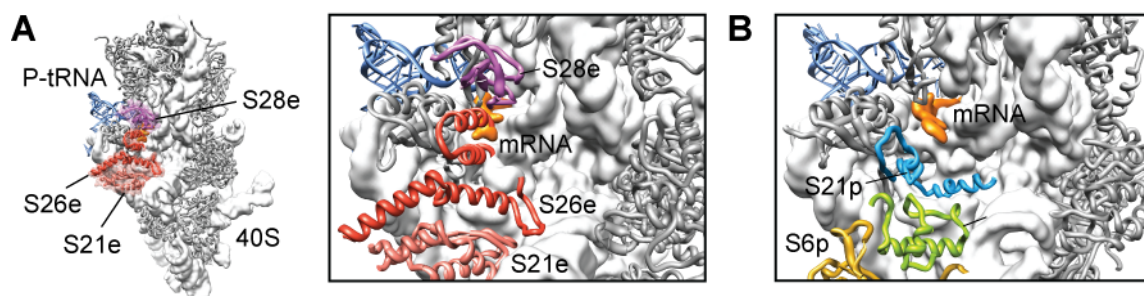
The interactions between the beak and h18 of the small subunit form the mRNA entry tunnel latch. The beak is described as a variable region (Armache et al., 2010a) and has a substantially different structure to that of the Prokaryotes. It is also a place where a new Eukaryote-specific protein, S17e (rpS17) was placed tentatively.

#### 3.3.1.1.2 mRNA Exit

Three Eukaryote-specific r proteins, S21e, S26e, and S28e, were identified at the mRNA exit site between the platform and head of 40S subunit (see Figure 25A). Both S26e and S28e have been cross-linked from positions (-6 and -7/-10, respectively) in the 5' untranslated region (UTR) of mRNA (Pisarev et al., 2008). The equivalent region of bacterial 30S subunits is occupied by bacterial-specific r proteins S6, S8, as well as S21 in *E. coli* (Schuwirth et al.,



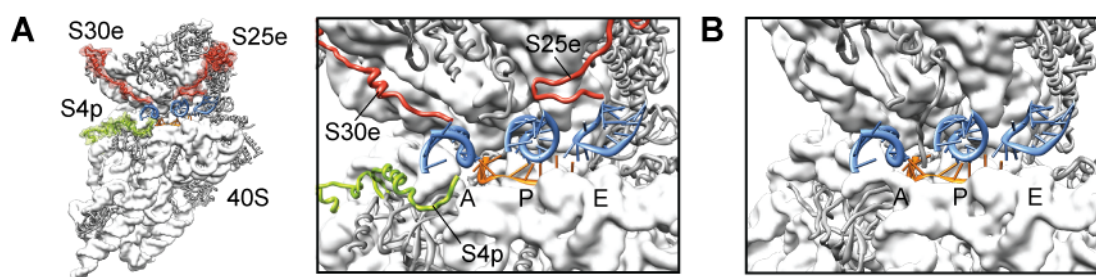
2005; Wimberly et al., 2000) (see Figure 25B). These differences may reflect the distinct elements found in the 5' UTRs of bacterial and eukaryotic mRNAs, as well as the divergence in the translation initiation phase (Sonenberg and Hinnebusch, 2009).



**Figure 25.** (A) Small 40S subunit with newly modeled r proteins S21e, S26e, and S28e colored distinctly (thumbnail, Left; zoom, Right). (B) Comparative view of the bacterial 30S subunit with bacterial-specific S18p shown in green (Selmer et al., 2006). In A and B, the P-tRNA (blue) and mRNA (orange) are shown for reference.

### 3.3.1.1.3 Decoding Center

Although the active sites of the ribosome - the decoding site on the small subunit and the site of peptide-bond formation on the large subunit - are composed largely of rRNA, they are not completely devoid of r proteins (see Figure 17 and Figure 26). Compared with bacterial 30S subunits, eukaryotic 40S subunits contain two additional r proteins, S25e and S30e, with extensions that reach into the decoding and tRNA binding sites (see Figure 26A and B). Consistent with this localization, S30e was cross-linked to the 4-thiouridine containing UGA stop codon of mRNA positioned at the A-site (Bulygin et al., 2005).



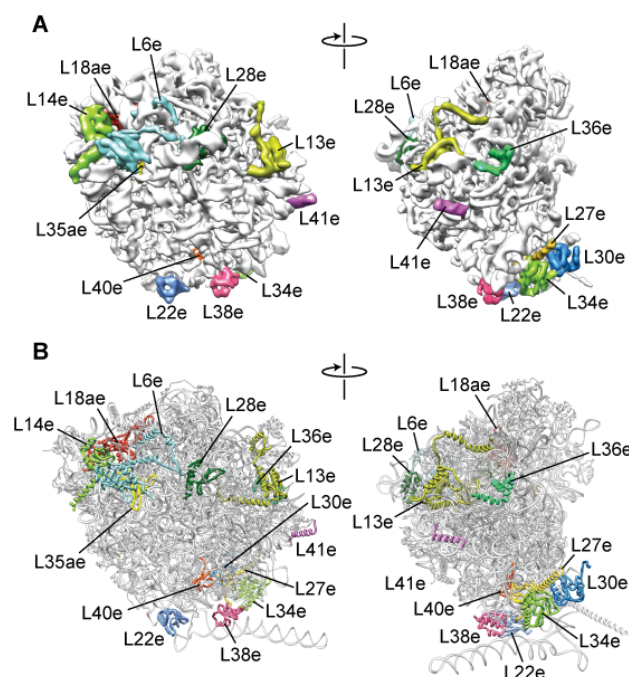
**Figure 26.** (A) Small 40S subunit with newly modeled r proteins S30e and S25e (red) and Eukaryote-specific extension of S4p (green) highlighted (thumbnail, Left; zoom, Right). (B) Comparative view of the bacterial 30S subunit decoding site (Jenner et al., 2010; Selmer et al., 2006). In A and B, the anticodon-stem-loops of A-, P- and E-tRNAs (blue) and mRNA (orange) are shown for reference.

### 3.3.1.3 Small Subunit Body

Four small subunit proteins were localized on the body of the 40S subunit, namely S4e, S7e, S24e and S27e. Information about their position was derived from the immuno-electron microscopy studies (for all of them, see Figure 24), as well as the structural data available for S4e, S24e (Choismel et al., 2008) and S27e (Herve du Penhoat et al., 2004). A general position of those proteins was established and the available structures were tentatively fit into the density. This resulted in points of references for localization of S7e into the only remaining density in the region pointed out by immuno-EM.

### 3.3.2 Large Subunit Proteins

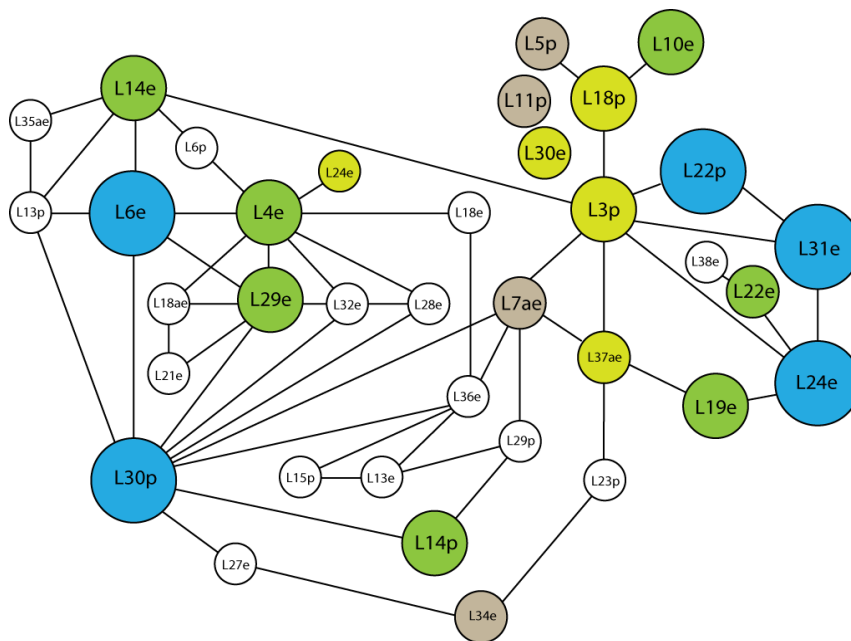
Bacterial ribosomal large subunit contains up to 34 proteins, of which 15 are specific to this domain. The eukaryotic one in turn comprises 47-51 proteins (number depending on the number of P1/P2 dimers), of which 20 are Archaea/Eukarya-specific and 7 are found only in the latter. We managed to assign the location of fifteen large subunit proteins, of which seven exist only in Eukarya and the rest can be found both in Eukaryotes as well as Archaea.



**Figure 27.** Localization of Eukaryote-specific r proteins. Cryo-EM maps of the *T. aestivum* 60S subunit (A) and (C) molecular models of r proteins, with newly identified r proteins colored distinctly.

All previously mentioned techniques were employed to localize proteins in the large ribosomal subunit. In addition, ribosomes were purified from *S. cerevisiae* knockout strains and revealed the location of L29e and L38e. Variation in distribution of gene encoding r protein L28e between *S. cerevisiae* and *T. aestivum* was exploited for localizing this protein. Reconstructions of ribosomes from four archaeal species were used to localize proteins L13e, L18ae, L34e, L35ae and L36e. All those methods, in conjunction with immuno-EM studies (Marion and Marion, 1987) were applied to localize all the large subunit proteins.

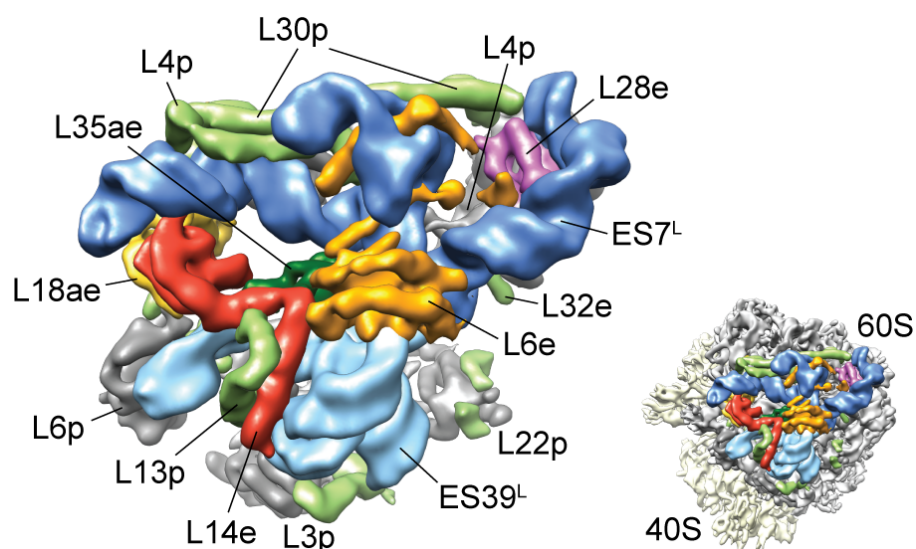
The summary for localization of the proteins can be found on Figure 28.



**Figure 28.** Model of the spatial arrangement of proteins within the 60S rat liver ribosomal subunit. Proteins are represented by circles, and the order of protein accessibility is: blue > green > yellow > tan, with white representing the most inaccessible proteins. Adapted from Marion and Marion, 1987.

### 3.3.2.1 ES7<sup>L</sup>-ES39<sup>L</sup> Region

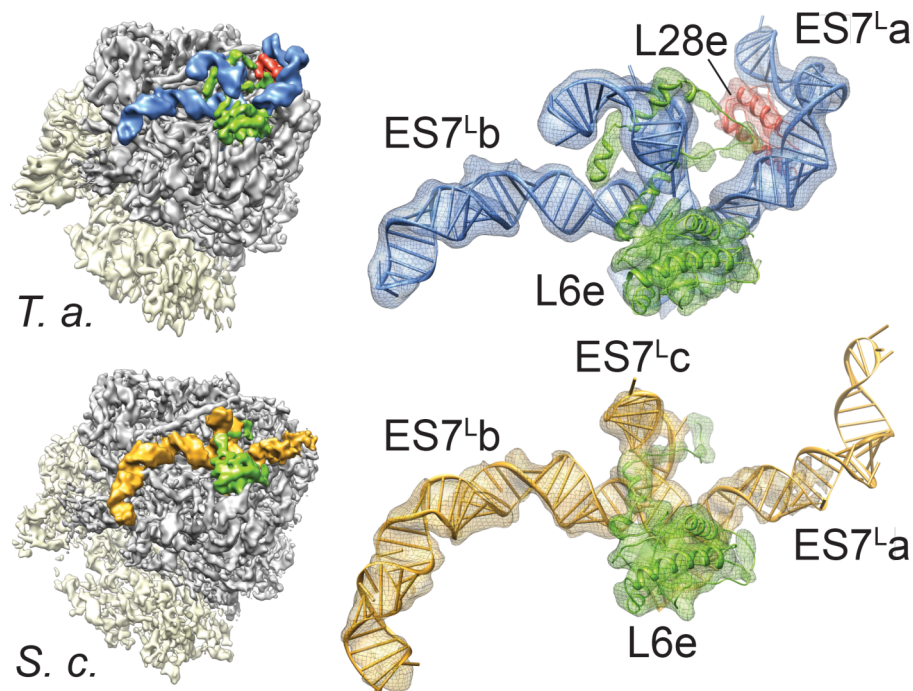
A large mass is found on the back of the 60S subunit comprising ES7<sup>L</sup>, ES39<sup>L</sup> and five eukaryotic r proteins L6e (rpL6), L14e (rpL14), L18ae (rpL20), L28e (L28 in *T. aestivum*) and L35ae (rpL33) (see Figure 29).



**Figure 29.** View of the intertwined region of ES7<sup>L</sup> (dark blue) and ES39<sup>L</sup> (light blue), with core r proteins (gray), Eukaryote-specific r protein extensions (pale green), and r proteins (L6e, orange; L14e, red; L18ae, yellow; L28e, pink; L35ae, green) highlighted. Inset shows relative position to 40S (yellow) and 60S subunits (gray).

L6e and L14e are the eukaryotic ribosomal proteins with possibly the largest size diversity between the species. L6e is 176 aa long in *S. cerevisiae*, 219 aa in *T. aestivum*, 243 aa in *Drosophila melanogaster* and 288 in *Homo sapiens*, while L14e - 134, 138, 166 and 215 aa long, respectively. Interestingly, L6e, L14e and L27e all adopt the same SH3-like barrel fold, possibly reflecting their origin due to gene duplication events (Edmondson et al., 2009).

Localization of protein L6e was based primarily on cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). Since L6e is Eukaryote-specific, initial candidate location was screened against reconstructions of archaeal ribosomes. Empty spaces in all available reconstructions of ribosomes of the archaeal species in this location confirmed the preliminary fit. The core of L6e was placed into the density and extensions were considered. The N-terminus of L6e was assigned based on differences between the densities and the sequences (43 aas) of *T. aestivum* and *S. cerevisiae*.

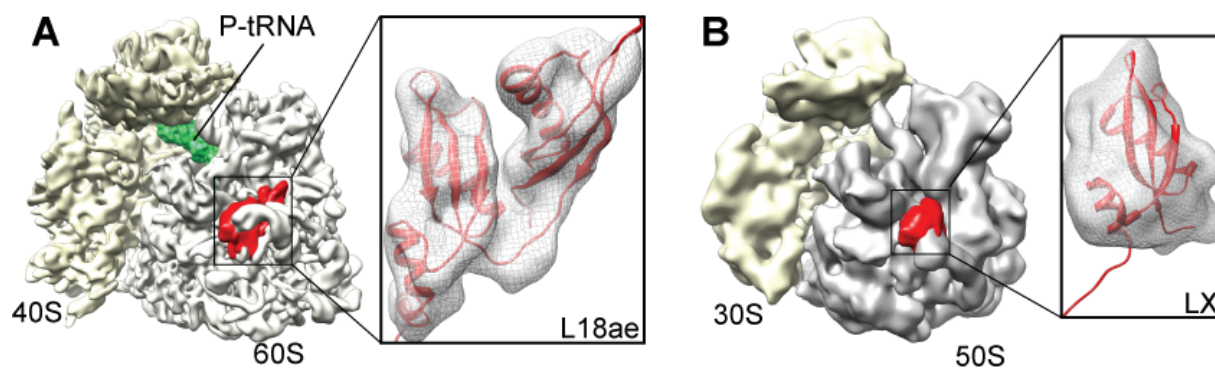


**Figure 30.** Isolated density for ES7<sup>L</sup> from *T. aestivum* (*T. a.*, blue) and *S. cerevisiae* (*S. c.*, gold) on the 80S ribosome (Left) and transparent with a molecular model (Center). Ribosomal proteins L28e (red) stabilizes ES7<sup>L</sup><sub>a</sub> in the *T. aestivum* 80S ribosome

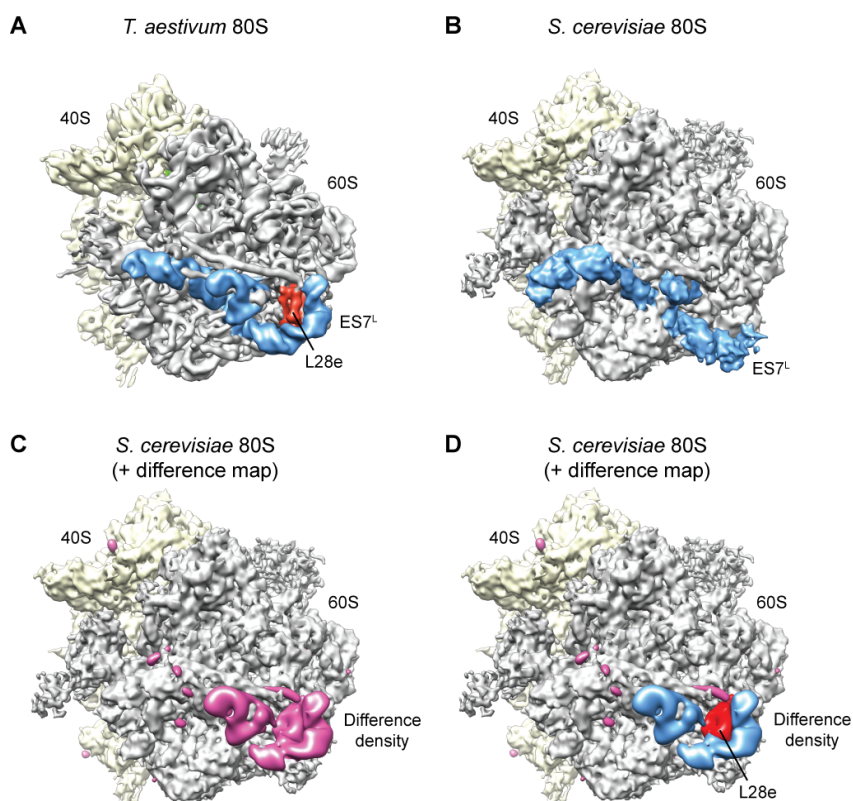
Once the location of protein L6e was established, similar approach was applied to finding L14e, which was predicted to reside in its vicinity (Marion and Marion, 1987) (see Figure 28). The protein model was fit and adjusted into the density. The alignment between *S. cerevisiae* and *T. aestivum* revealed that C-terminus of the latter is longer, while yeast harbors an extended N-terminus. A careful examination of the densities confirmed this, further supporting the localization.

L18ae has been described as a Eukaryote-only r protein. Fold search revealed it consists of two domains with a distinct LX motif. Protein LX is known to be an Archaea-specific ribosomal protein, with a heterogeneous distribution within this domain of life. L18ae therefore is probably a result of LX gene duplication.

This protein was localized by exploiting differences between the domains of life. In this part of the ribosome, the putative densities in Eukarya seem roughly twice as large as those in Archaea. This would correspond to protein L18ae consisting of two LX domains, with only one LX protein present in Archaea. When visualized in the archaeal reconstructions, protein L18ae reveals to have only one of its LX domains fitting, which supports the localization (Figure 31).



**Figure 31.** A side-by-side comparison of *T. aestivum* location of protein L18ae (A) and a respective putative position of protein LX in *M. igneus* (B).



**Figure 32.** Difference maps generated between *Triticum aestivum* and *Saccharomyces cerevisiae* identify position of r protein L28e. (A) *T. aestivum* 80S reconstruction with ES7<sup>L</sup> (blue) and L28e (red) positions highlighted. (B) *S. cerevisiae* 80S reconstruction with ES7<sup>L</sup> (blue) highlighted. (C) *S. cerevisiae* 80S reconstruction superimposed with the difference density (magenta) calculated between the (A) *T. aestivum* map and (B) the *S. cerevisiae* map. (D) Same as C but with the regions of the difference density corresponding to ES7<sup>L</sup> in *T. aestivum* colored blue, leaving a large region of extra density (red) that was assigned as r protein L28e.

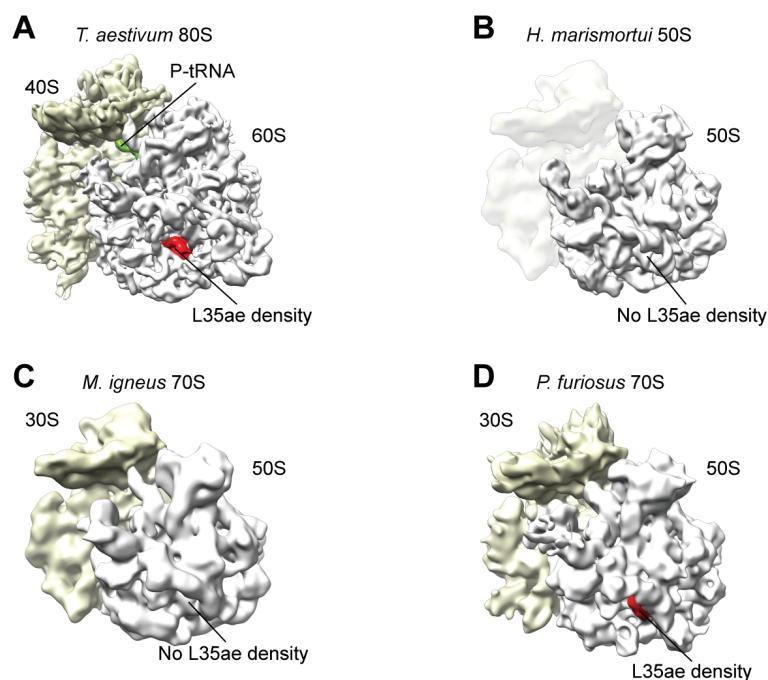
R protein L28e is not present in the *S. cerevisiae* genome, and therefore its localization was possible by generating difference maps between yeast and *T. aestivum* ribosomes (Figure 32).

The difference map clearly shows that the two organisms have a rather large disparity in this region. This is due to dissimilarity in rRNA ES7<sup>L</sup>, as well as in the protein content. This, coupled with the immuno-EM studies (Marion and Marion, 1987) pointed out the position of L28e.

L35ae is a ribosomal protein found both in Archaea and in Eukarya, with a heterogenous distribution within the former domain. As shown in (Figure 21, Table 7), it exists in *Pyrococcus* and *Aeropyrum* species, but is absent in *Methanococcus*, *Sulfolobus* as well as *Pyrobaculum* species. This allows for refining the initial protein localization based on cross-linking and accessibility to proteolysis (Marion and Marion, 1987) by comparing available archaeal density maps (see Figure 33).

	<i>Sulfolobus</i> species	<i>Pyrococcus</i> species	<i>Methanococcus</i> species	<i>Pyrobaculum</i> species	<i>Aeropyrum</i> species (incl. <i>Staphylothermus</i> <i>marinus</i> )	Length, aa
L13e						~90
L14e						~95
L30e						~110
L34e						~90
L35ae						~90
L38e						~70
LX						~80
S25e						~110
S26e						~100
S30e						~55

**Table 7.** Distribution of ribosomal proteins in archaeal species available for reconstruction.

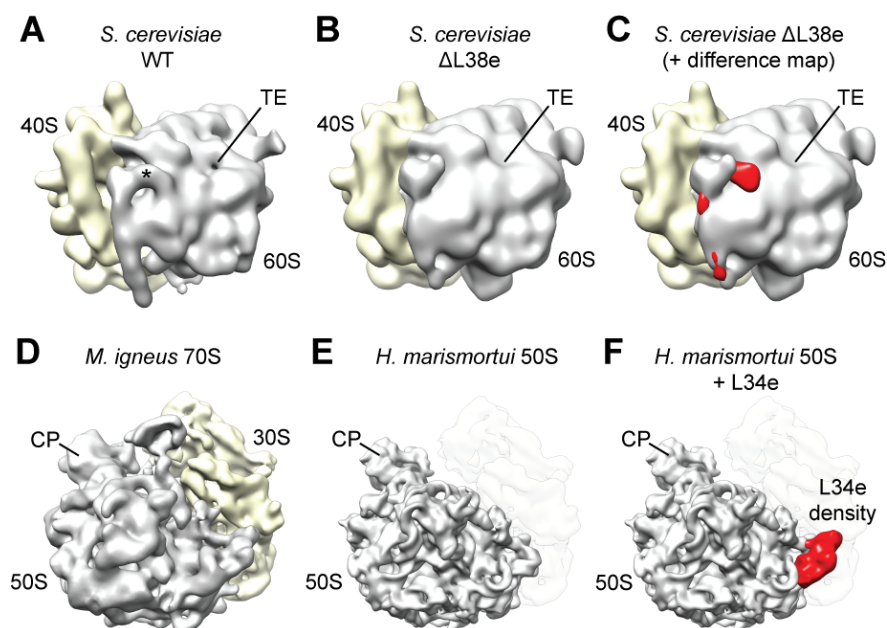


**Figure 33.** Identification of r protein L35ae based on heterogenous distribution in Archaea (see Table 7). *T. aestivum* 80S ribosome (A) compared to (B) X-ray structure of the 50S subunit from *H. marismortui* filtered to a similar resolution, (C) *M. igneus* 70S ribosome (both lacking protein L35ae) and to (D) *P. furiosus* (courtesy of Sibylle Franckenberg) with protein L35ae highlighted in red.

### 3.3.2.2 ES27<sup>L</sup> Interacting Partners

ES27<sup>L</sup> is unique for its highly dynamic behavior, being found in two distinct positions in yeast 80S ribosomes (Beckmann et al., 2001); one oriented towards the L1 stalk, termed ES27<sup>L</sup><sub>in</sub> and one away from the L1 stalk but towards the tunnel exit, termed ES27<sup>L</sup><sub>out</sub> (Figure 40). Modeling of both conformations reveals that interchange between the ES27<sup>L</sup><sub>in</sub> (gold) and ES27<sup>L</sup><sub>out</sub> (blue) positions involves a rotation of  $\sim 110^\circ$  of ES27<sup>L</sup><sub>a-c</sub> relative to H63 (Figure 40). Weak density for ES27<sup>L</sup><sub>in</sub> the reconstruction of the *T. aestivum* ribosome suggests that ES27<sup>L</sup> exhibits a continuum of different conformational states.





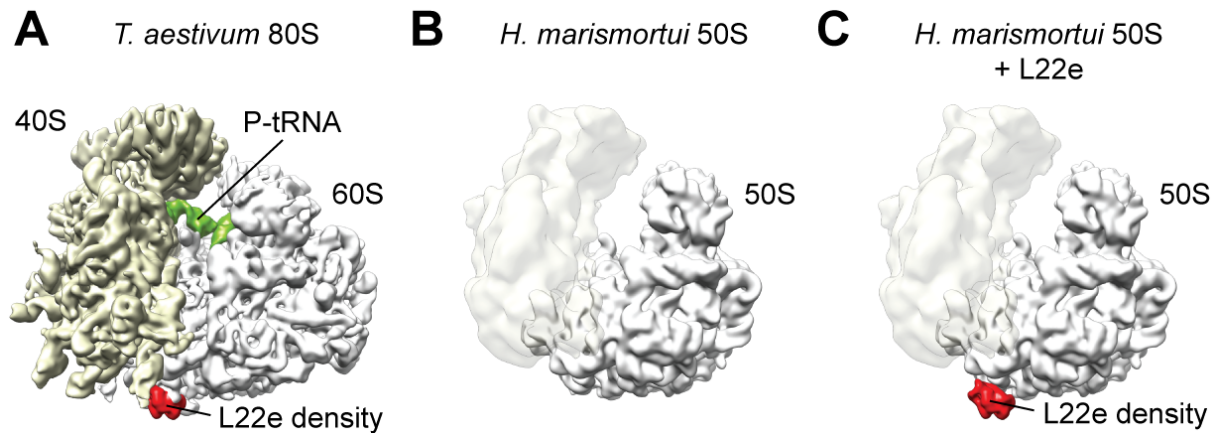
**Figure 34.** Identification of r proteins L38e and L34e. (A) *S. cerevisiae* WT 80S ribosome compared to (B) reconstruction of *S. cerevisiae* 80S ribosomes isolated from a strain lacking the gene for L38e. The asterisk indicates the position of additional density assigned to L38e, and the tunnel exit (TE) is shown for reference. (C) Difference density map calculated between A and B and shown superimposed on the map from B. Reconstruction of (D) *M. igneus* 70S ribosome, compared to (E) X-ray structure of the 50S subunit from *H. marismortui* filtered to a similar resolution. (F) Difference density map calculated between D and E and shown superimposed on the map from E identifying the location of r protein L34e (red).

Nevertheless, at low thresholds one preferential state is observed, intermediate in position ( $ES27^{L_{int}}$ ) to the yeast  $ES27^L$  in and  $ES27^{L_{out}}$  positions (Figure 40). All three positions appear to be stabilized through interaction with r proteins: The yeast  $ES27^{L_{out}}$  and the *T. aestivum*  $ES27^{L_{int}}$  conformations directly contact r protein L38e (Figure 40), which was assigned based on a reconstruction of yeast  $\Delta L38e$ -80S ribosome (Figure 34A-C). R protein L34e stabilizes the yeast  $ES27^{L_{in}}$  position and was assigned partly on the basis of its heterogeneous distribution in Archaea, being absent in *H. marismortui* 50S structure (Ban et al., 2000), but present in a reconstruction of the *Methanococcus igneus* and 70S ribosome (Figure 34D-F).

### 3.3.2.3 H57 Interactor

In Eukaryotes, the loop of rRNA H57 is conserved, unlike in Bacteria. This loop is an interaction partner for L22e, a protein localized primarily based on ref. Marion and Marion, 1987 (see Figure 28). The location of the protein was further confirmed via comparison of *S.*

*cerevisiae* and *T. aestivum* cryo-EM maps (which contain L22e) with the *H. marismortui* 50S X-ray structure and the *M. igneus* 70S cryo-EM structure (which lack L22e).



**Figure 35.** Identification of r protein L22e (A) *T. aestivum* 80S ribosome compared to (B) X-ray structure of the 50S subunit from *H. marismortui* filtered to a similar resolution. (C) Difference between A and B and shown superimposed on the map from B identifying the location of r protein L22e (red).

### 3.3.2.4 H58 Interactors

L27e is located below the L1 stalk on the opposite side of the ribosome from L6e and L14e, where it is sandwiched between H55 and H58. L27e and L34e overlap the position of H58 in the *E. coli* 70S ribosome, emphasizing the conformational rearrangements that relocate H58 in archaeal/eukaryotic compared to bacterial ribosomes. L27e was localized by (i) comparing reconstructions of archaeal ribosomes (L27e is absent) with their eukaryotic counterparts, where L27e is present; (ii) placing it in relation to other proteins (known from cross-linking studies (Marion and Marion, 1987), Figure 28). L34e in turn exists both in Eukarya and Archaea; it was assigned at the end, when most of the densities were already allocated to other entities. Protein L27e is a homolog of proteins L6e and L14e with which it shares a common core. It was built based on an L14e template, while L34e was modeled using a non-ribosomal template and heavily adjusted to the density.

### 3.3.2.5 L1 Stalk

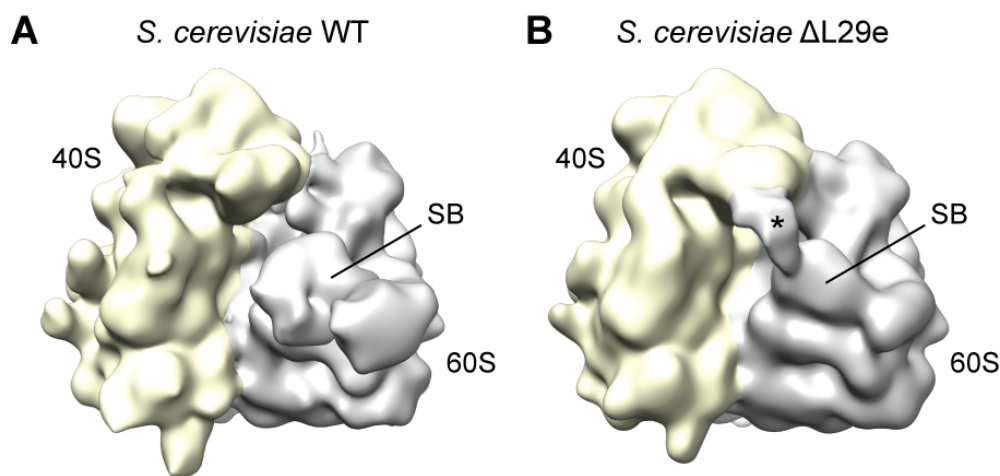
R protein L36e was located at the base of L1 stalk. It binds between H21 and H22, H68 and H76, close to a position that is occupied by protein L9 in Bacteria. L13e is a protein modeled close to H76, most probably interacting with the L1 stalk using its elongated C-terminal

extension (not modeled). The amino-terminus of L13e inserts a helix into the core (H12, H13, H28), most probably stabilizing it after a swivel of H15-H18 (as compared to Archaea).

Both proteins were localized based on cross-linking and access to proteolysis data (Marion and Marion, 1987). Archaeal reconstructions were used for assessment in both cases: L13e is heterogeneously distributed in Archaea (see Table 7); L27e is present in *S. cerevisiae* and *T. aestivum* cryo-EM map, but not in *H. marismortui* 50S X-ray structure and the *M. igneus* / *P. furiosus* 70S cryo-EM structure (it is Eukarya-specific).

### 3.3.2.5 P0-P1/P2 Stalk Base

The localization of L29e to this pocket was based partly on the observation that the stalk rearranges position to establish contact with the head of the 40S subunit in a reconstruction of the yeast  $\Delta$ L29e-80S (Figure 36), which has not been observed in any previous yeast 80S reconstructions. Moreover, the assigned position for L29e is in close proximity to L10e (L16p), which exhibits synthetic lethality with L29e in yeast (DeLabre et al., 2002).

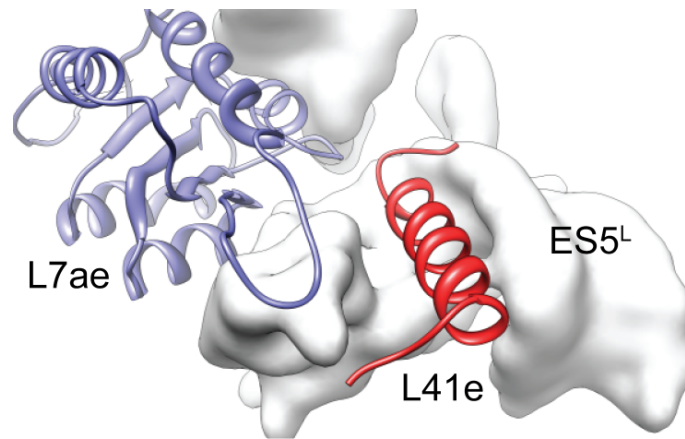


**Figure 36.** Localization of ribosomal protein L29e. Reconstruction of (A) *Saccharomyces cerevisiae* WT 80S ribosome, compared to (B) reconstruction of *S. cerevisiae* 80S ribosomes isolated from a strain lacking the gene for L29e. In B, the rearranged position of the stalk base (SB) on the large subunit (gray) leads to a contact between the stalk (\*) and the head of the small 40S subunit (yellow).

### 3.3.2.6 L40e and L41e

The localization of proteins L40e (rpL40) and L41e (rpL41) had the least supporting data. Protein L41e is a 25 amino acid-long protein that is conserved in Eukaryotes, with a secondary structure prediction of a straight helix. It was one of the last proteins to be

localized, due to its miniature size. During the course of examination of the leftover unassigned density, a rod-like density was found, which did not form any connectivity with other protein entities in the vicinity. Thus, L41e was modeled into this remaining density (Figure 37).



**Figure 37.** A model of protein L41e, the smallest Eukaryote-specific protein

The last protein to be found, L40e, had its structure solved using NMR (Wu et al., 2008a). It was assigned because of its location: an isolated density that is unlikely to be RNA or an r protein extension. It is inserted into the conserved rRNA core, in between H33, H53, H56 and H57.

### 3.3.3 The Remainder

Six small subunit r proteins (S3ae, S6e, S8e, S10e, S12e, and S27ae) could not be localized and were therefore not modeled. Their location can be roughly estimated to lie in the foot and the spur of the small subunit, since the leftover protein density is located primarily there. Except for protein S27ae (rpS31), all the proteins lack a reliable base for modeling and therefore without knowing their structure or the location, little can be done to predict that.

## 3.4 RNA

It is a must, when writing about the ribosome, to mention the ribosomal RNA, especially since there is a quite a difference in the rRNA structure between Prokaryotes and Eukaryotes. The majority of the rRNA modeling effort was done by Alexander Jarasch and Andreas Anger.

The majority of the conserved core of the *T. aestivum* and yeast 80S ribosomes was modeled based on homology of the eukaryotic rRNA with the available bacterial and archaeal ribosome structures using Assemble (Jossinet et al., 2010). On this basis, it was possible to generate a template-based model for the *T. aestivum* 80S with a total of 3,466 (1,051/40S and 2,415/60S) nts of the 5485 rRNA, incorporating isosteric base substitutions (Stombaugh et al., 2009). Nucleotides that were not available in the sequences for *T. aestivum* (120 nts, 2.2%) were substituted with those from the closely related *Oryza sativa*. One hundred sixteen (67 and 49) nts, mostly single stranded linker regions, could not be modeled, due to unreliable secondary structure predictions and/or ambiguity in the electron density. The remaining 1,903 nts comprising structurally variable regions and rRNA expansion segments were modeled de novo using Assemble (Jossinet et al., 2010), RNAfold (Hofacker, 2003), and RNASHapes (Steffen et al., 2006).

## 4 Conclusions and Discussion

### 4.1 Cryo-Electron Microscopy Map of *Triticum aestivum*

#### 4.1.1 Feature Visibility at Given Resolution

The resulting cryo-EM map displays characteristics similar to X-ray crystallographic maps of ribosomes at similar resolution, namely the *Haloarcula marismortui* 50S subunit at 5 Å (Ban et al., 1999), and the *Thermus thermophilus* 30S (Clemons et al., 1999) and 70S structures (Yusupov et al., 2001) at 5.5 Å.

As stated, in the crystal structures, “*at a resolution of 5.5 Å, the phosphate backbone of the ribosomal RNA is visible, as are the  $\alpha$ -helices of the ribosomal proteins, enabling double-helical regions of RNA to be identified*” (Clemons et al., 1999).

At this resolution, well-resolved density for double-helical RNA is observed; with defined minor and major grooves as well as distinctive bumps indicative of phosphate groups located along the backbone ridges (Figure 7 and Figure 8).

At close examination, some parts of the map are visibly better resolved, while others are very distorted. The former can be observed in parts of the ribosome where proteins and RNA form an entangled, stabilized region, like the tunnel exit-site or the ribosome active site; the latter applies especially to the flexible regions of the ribosome, like the L1 stalk or the P0/P1-P2 stalk. Therefore it has to be annotated that while the overall resolution of the ribosome is declared at 5.5 Å, some parts might be resolved to a slightly better or worse resolution, depending on the region's flexibility.

#### 4.1.2 Improvements Over Existing Reconstructions of Eukaryotic Ribosomes

Previous eukaryotic ribosome reconstructions were reported at a range of resolutions, over the years reaching sub-nanometer resolutions - up to 6.1 Å in 2009 (Becker et al., 2009). The first model of a eukaryotic ribosome was constructed in 2001 (Spahn et al., 2001), by placement of models based on prokaryotic elements into a 15 Å yeast density, without further modeling of extensions and expansion segments. A second one was constructed in 2004 (Spahn et al., 2004a) to the 11.7 Å map of an 80S-eEF2-sordarin complex from yeast, by docking protein and rRNA models into the density, not supplying details on novel proteins or expansion segments either.

Over the years, the quality of the maps was improving, resulting in a number of reconstructions with resolution below 10 Å (a selection is shown in Table 8). Only two more attempts were undertaken at constructing models into a cryo-EM density: in 2008 into 8.7 Å map of a dog ribosome (Chandramouli et al., 2008) and in 2009 to 8.9 Å map of a thermophilic fungi (Taylor et al., 2009). Despite other, better resolved maps, no further effort beside ours was undertaken to model the ribosome based on the cryo-EM densities.

**Table 8.** A selection of sub-nanometer cryo-EM maps reported in the years 2006-2010

Year	Kingdom	Organism	Resolution, Å	EMD ID	Author
2006	Fungi	<i>S. cerevisiae</i>	9.9	1233	Andersen et al., 2006
2006	Fungi	<i>S. cerevisiae</i>	7.3	1285	Schuler et al., 2006
2006	Plantae	<i>T. aestivum</i>	7.4	1217	Halic et al., 2006
2007	Fungi	<i>T. lanuginosus</i>	8.9	1345	Taylor et al., 2007
2008	Mammalia	<i>C. familiaris</i>	8.7	1480	Chandramouli et al., 2008
2009	Plantae	<i>T. aestivum</i>	6.5	1652, 1768	Becker et al., 2009, Bhushan et al., 2010c
2009	Fungi	<i>S. cerevisiae</i>	6.1	1667	Becker et al., 2009
2010	Plantae	<i>T. aestivum</i>	5.5	1780	Armache et al., 2010a, b

When compared to our 5.5 Å reconstruction, a few conclusions emerge. The large number of particles in our structure yielded an unparalleled clarity of the map, showing a much better assignment of densities, resulting in better protein/RNA differentiation. In most of the maps, alpha-helices are much smoother, without the characteristic ‘bumps’ (pitch of the helix), beta-sheets are also not really distinguishable. Loops and less ordered regions are either absent or, when resolved, highly ambiguous. In contrast to our map, in the majority of the reconstructions the typical pitch for phosphates, visible at 5.5 Å, is absent. Head swivel is also an issue in some of the reconstructions, restricting resolution of the small subunit, unlike in our map.

#### 4.1.3 Cryo-EM and the Future of Intracellular Investigation

For years, cryo-Electron Microscopy has been at the frontline of intracellular investigation of molecular machines. Ribosome, polymerase, spliceosome, various viruses, they have all been visualized using this technique, having a clear advantage of showing a state in progress. The structures of a Sec61 translocon or SRP bound to the translating ribosomes (Becker et al., 2009; Halic et al., 2004) were visualized only thanks to this approach.

As the technique is maturing, the resolution of the molecules is increasing, and in some time, cryo-EM might be a realistic alternative to X-ray crystallography in terms of resolution. Already, some publications point to the resolution border being passed. Ref. Harrison, 2010, discusses two articles reporting a structure of a human adenovirus, Liu et al., 2010, using cryo-EM and Reddy et al., 2010, using X-ray crystallography. Harrison states: “*The cryo-EM density map of Liu et al. appears to be substantially clearer and more interpretable than the x-ray density map of Reddy et al. Several other icosahedral virus structures have shown that cryo-EM single-particle analysis now rivals x-ray crystallography when applied to large, homogeneous, highly symmetric objects. It can be argued that only conquerable computational barriers now prevent extension to less symmetric structures, although rigidity and conformational homogeneity (qualities experimentally selected when growing crystals) will continue to be important*”.

Our 5.5 Å map should be therefore regarded as an intermediate step between low-resolution cryo-EM maps and high-resolution crystallography data. In the future, with the emergence of more sophisticated computational methods for ensuring homogeneity, superior microscope stability and better sample quality, it might result in asymmetrical structures reaching resolutions below 4 Å.

## 4.2 Model

### 4.2.1 Localization of the Novel Proteins

A total of 17 r proteins (see Tables S 1 and S 2 for r protein family nomenclature), 7 (S4e, S17e, S19e, S24e, S27e, S28e, and RACK1) from the 40S subunit, and 10 (L4e, L6e, L14e, L18ae, L27e, L30e, L35ae, P0, P1, and P2) from the 60S subunit were modeled using available X-ray and NMR structures of free r proteins (see Table 4). Homology models for six r proteins (S25e, L22e, L29e, L34e, L36e, and L38e) were built using HHpred (Soding et al., 2005) and Modeller (Sali and Blundell, 1993) on the basis of similarity with domains of proteins of known structure, for example, S25e and L38e were predicted to have helix-turn-helix and K-homology domains, both of which are known to interact with RNA. Seven r proteins (S7e, S21e, S26e, S30e, L13e, L28e, and L41e) were tentatively modeled *ab initio* on the basis of secondary structure predictions and density characteristics, and six small subunit r proteins (S3ae, S6e, S8e, S10e, S12e, and S27a) could not be localized and were therefore not modeled. This constitutes 74 from a total of 80 proteins (92.5 %) that were localized and



modeled, with 27 proteins that were not crystallized on the ribosome and 24 that were previously not reported.

#### 4.2.1.1 Reliability of the Localization

It is important to assess the reliability of the protein localization, based on the number of constraints mentioned in Table S 10.

At the given resolution, it should be noted that the degree of accuracy and reliability of the assignments varies for the different r proteins. The fold and location of ribosomal core proteins - the ones derived from available ribosomal crystal structures, cannot be disputed. Their location and fold has been repeatedly analyzed and confirmed. Proteins modeled on the basis of available unbound X-ray and NMR r protein structures will have a higher degree of accuracy than those generated using remote homology or *ab initio* modeling.

The newly localized proteins can be divided into three groups, based on the quality and quantity of available data and the author's subjective assessment: (i) highly reliable, there is no doubt about their location; (ii) reliable, assuming that certain conditions are fulfilled; and (iii) unreliable, but possible.

Nine novel proteins populate the first group: S4e, S19e, S25e, S28e, L14e, L18ae, L22e, L28e and L38e. The reason for the choice depends from case to case. There are structures of S4e, S19e, S28e and L14e in solution, and the proteins fit their respective densities. R proteins S25e, L18ae, L22e and L38e have reliable homology models, that support the fit, and, in case of protein L38e, a reconstruction of a knockout strain. Density for protein L28e is shown as the only one on the ribosome to harbor this particular entity, its model strictly based on the ss prediction. All the proteins also fit their respective location data.

The second group consists of ten proteins, three in the small subunit (S7e, S26e and S27e) and seven in the large (L6e, L13e, L27e, L29e, L34e, L35ae and L36e). Four members of this group can be described as being more reliable than others, namely L27e, L34e, L35ae and L36e, because their location is backed by the available archaeal reconstructions, furthermore strengthened by available solution structures of L27e and L35ae.

The last group consists of proteins that were positioned and modeled, but the basis for their location was vague and ambiguous, hence may or may not be true. Consequently, they should be treated as tentatively placed. This applies both to the small (S17e, S21e, S24e and S30e;

see §4.2.6.2 *The Structure of a 40S-eIF1 T. thermophila Complex at 3.9 Å*) and large subunit (L40e, L41e; see §3.3.2.6 *L40e and L41e*). Even though for proteins S17e, S24e and L40e structures were available, their localization was not as clear as in the second group, whereas no homologues were found for S21e and S30e.

#### 4.2.2 Roles of Proteins

There are 80 proteins in the eukaryotic ribosome, acquired on different stages of the evolution. They have a range of functions, such as assembly initiation, RNA processing, signal transmission, GTPase activation or scaffolding (see Table S 16 and Table S 17). A number of them will be described below in more detail.

##### 4.2.2.1 mRNA Entry

The mRNA entry is constructed by proteins S3p, S4p and S5p. S4p and S5p play an important role in maintaining translational accuracy (see Table S 16) (Alksne et al., 1993), while S3p is a protein through which stability of the ribosome is maintained. This cluster of proteins at the mRNA entry site might have also another implication in decoding. In the cell, all mRNA form some secondary structure (Doty et al., 1959). For the translation, a single stranded mRNA needs to enter the decoding site; thus, any secondary structure present in mRNA has to be melted for it to pass through the entry pore (a helicase activity, Takyar et al., 2005). In Eukaryotes, initiation starts with the binding of the “small subunit-initiation factors-GTP-initiator tRNA” complex to the 5' cap of mRNA and scanning until the first AUG is reached. Thus, the mRNA entry pore is also involved in the initiation.

Additionally, the C-terminus of r protein S4p is relocated in Eukaryotes, due to corresponding rearrangements in h16/17, and reaches from the globular domain on the solvent side right into the decoding site of the small subunit (Figure 26A). Thus, together with the extensions and loops of eukaryotic homologues to the bacterial S7p, S9p, S11p, S12p and S13p r proteins (Jenner et al., 2010; Selmer et al., 2006), at least seven different r proteins can interact and modulate the binding of tRNAs to the 40S subunit.

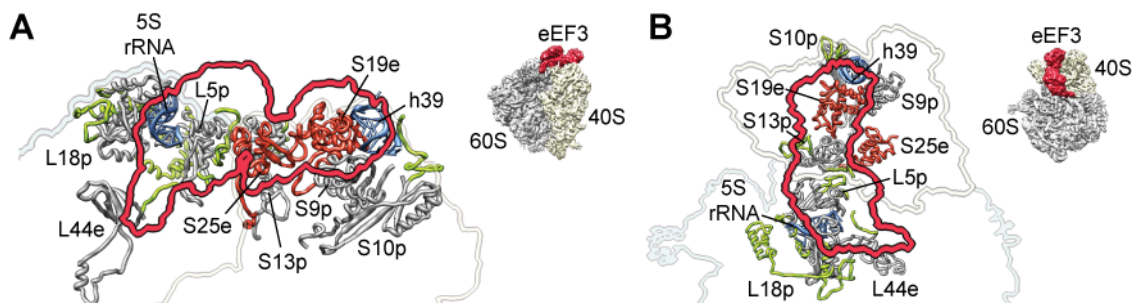
Eukarya-specific extensions to those proteins might point to the ways the ribosome adapted to the growing complexity of the translation initiation, inducing new regulatory functions.

#### 4.2.2.2 mRNA Exit

The structures of mRNA exit in Bacteria and Eukarya differ to a large extent. Since translation initiation in the latter does not involve Shine-Dalgarno sequence and bacterial-specific protein S1 (Kapp and Lorsch, 2004), the 3' end of their 18S rRNA does not contain a complementary anti-Shine-Dalgarno. The relative space in the 40S subunit is deprived of both of those entities and partially filled by r proteins S26e and S28e (Figure 25). This part of 18S rRNA in Eukarya is stabilized by interactions with S26e, as opposed to flexible 3' end of 16S in Bacteria (Rabl et al., 2011). Thus, the change in the protein composition in the mRNA entry may reflect differences in translation initiation in Eukaryotes.

#### 4.2.2.3 Head of the Small Subunit (eEF3 Binding Site)

We can now identify the eEF3 interaction partners in the yeast 80S, previously assigned as rpSX1 and rpSX2 (Andersen et al., 2006), as being r-protein S19e and S25e, respectively, both of which are located in the head of the 40S subunit (Figure 38). In addition, r-protein L44e as well as Eukaryote-specific extensions of r proteins L5p and L18p located within the central protuberance of the 60S subunit also comprise the eEF-3 binding site (Figure 38).



**Figure 38.** The binding site of eEF3 on the *S. cerevisiae* 80S ribosome, with (A) side and (B) top views (see insets) showing the binding site of eEF3 as a red outline and molecular models of ribosomal components that comprise the eEF3 binding site. Newly identified proteins are shown in red (S19e, S25e) and newly modeled r protein extensions in green, whereas core r proteins are colored gray. Modified from ref. Andersen et al., 2006.

The head of the small subunit harbors also a unique protein, for a long time not regarded as associated to the ribosome. Receptor for activated protein kinase C (RACK1) in the cytoplasm exists in two states: associated and dissociated (from the ribosome). When dissociated, it might prevent certain genes from being translated; associated, it might serve as a signaling scaffold between ribosome and Scp160 (Coyle et al., 2009), PKC $\beta$ , and others.

#### 4.2.2.4 Peptidyl Transferase Center

Protein L10e, known in Bacteria as L16, was also crystallized in *H. marismortui*, however without a very interesting loop region (*H.m.* 102-114). This loop was visualized in our reconstruction and further modeled, indicating that it acquires a definite position in the presence of tRNA, and in its absence is disordered. This loop in bacterial L16p is shorter and, as was pointed before, substituted by an amino-terminal extension of bacterial-specific L27p (Maguire et al., 2005; Voorhees et al., 2009).

In Eukarya, it is highly conserved and mutations or deletions in this loop are lethal (Hofer et al., 2007), suggesting that it may play an important role in tRNA positioning, as proposed for the N terminus of L27 (Maguire et al., 2005; Voorhees et al., 2009). It may be involved in proper positioning of the acceptor stem of A- and P-site tRNAs (Wilson and Nierhaus, 2005) and involved in GTPase regulation (Johnson, A., unpublished data). It is also the closest protein entity in Eukaryotes to approach the site of peptide bond formation (approximately 16 Å, see Figure 18).

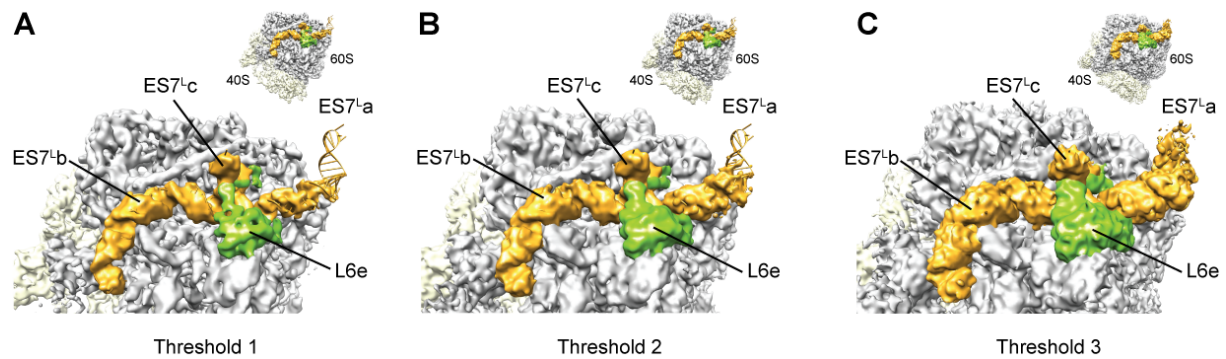
#### 4.2.2.5 ES7<sup>L</sup>-ES15<sup>L</sup>-ES39<sup>L</sup> Stabilization

The largest expansion segment in the *T. aestivum* and yeast ribosomes is ES7<sup>L</sup>, which is located at the back of the 60S subunit (see Figure 29).

Novel proteins in the ES7<sup>L</sup>-ES15<sup>L</sup>-ES39<sup>L</sup> region are all interacting with the expansion segments that are unique to Eukaryotes. A core domain of L6e is located exactly in between ES7<sup>L</sup> and ES39<sup>L</sup> and most probably serves as a stabilizing partner for those expansion segments. Protein L14e stabilizes ES7<sup>L</sup><sub>b</sub> and ES39<sup>L</sup><sub>b</sub>, inserting a carboxy-terminal alpha-helix in between ES39<sup>L</sup><sub>b</sub> and ES39<sup>L</sup><sub>d</sub>. Protein L18ae acts as a clamp for three rRNA entities, namely 5S, ES7<sup>L</sup><sub>b</sub> and helix 42. The placement of extensions of proteins L4p, L13p, L18e and L30p, along with the newly localized proteins (see §3.2.3.2.4 and §3.3.2.1) suggests a scaffolding function. When viewed with the *H. marismortui* rRNA, one can notice that extended parts of those proteins would be ‘drifting in space’, with no RNA support.

As shown, along with the expanding rRNA structure in this region, novel proteins appear and extensions to the existing proteins emerge to hold down and stabilize the growing rRNA entity.

#### 4.2.2.6 Stabilization of ES7<sup>L</sup><sub>a</sub> (L28e)



**Figure 39.** Density for ES7<sup>L</sup><sub>a</sub> in yeast is only seen at low thresholds, suggesting it to be more flexible than in *T. aestivum*, where it is stabilized through interaction with L28e.

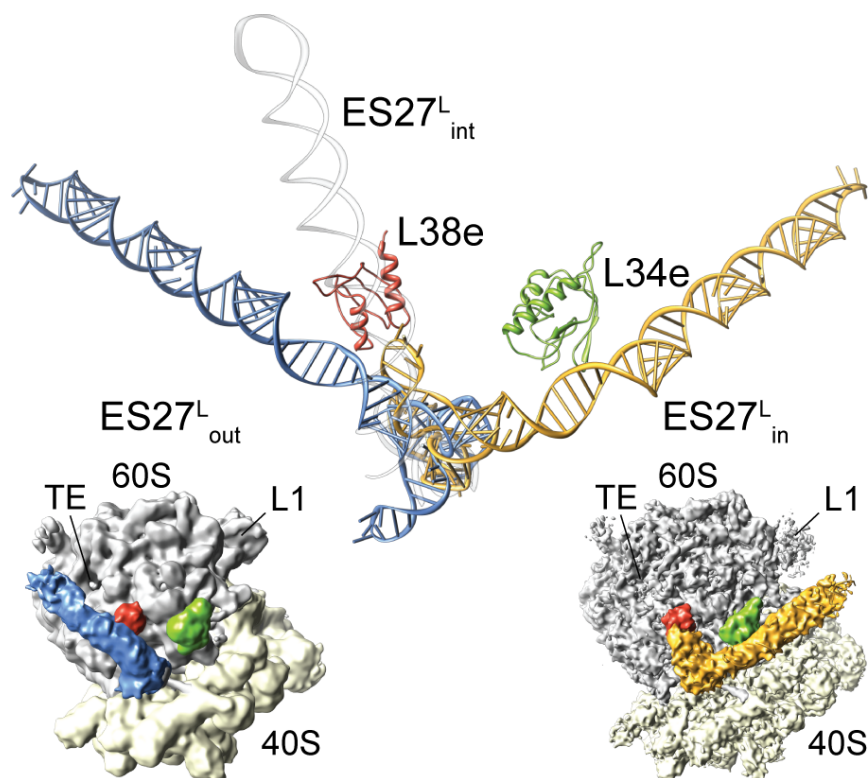
ES7<sup>L</sup> in *T. aestivum* and yeast share a number of common features, however, one large difference is evident: density for ES7<sup>L</sup><sub>a</sub> in yeast is only seen at low thresholds (Figure 39), suggesting it to be more flexible than in *T. aestivum* (see Figure 30). The reason for this appears to be that ES7<sup>L</sup><sub>a</sub> in *T. aestivum* is stabilized through an interaction with r protein L28e, which, as mentioned in (Lecompte et al., 2002), does not exist in yeast. This feature was also observed in *Thermomyces lanuginosus* (Taylor et al., 2009), therefore it is safe to assume that this protein serves as a scaffold for rRNA, gluing together ES15<sup>L</sup> and ES7<sup>L</sup><sub>a</sub>. It is not known however to play any other role in the ribosome.

#### 4.2.2.7 Coordination of ES27<sup>L</sup> (L38e & L34e)

Ribosomal proteins L34e and L38e were discovered at the bottom of the large ribosomal subunit, in the vicinity of ES27<sup>L</sup>. Their position and the three states of the rRNA expansion segment indicate that they might be involved in stabilization of ES27<sup>L</sup>. L38e was observed to be in the immediate neighborhood of ES27<sup>L</sup><sub>out</sub> and ES27<sup>L</sup><sub>int</sub>. The location of this r protein suggests that it supports the whole range of ES27<sup>L</sup> positions. In turn, protein L34e was observed in the direct vicinity of ES27<sup>L</sup><sub>in</sub>, but not in interaction with ES27<sup>L</sup><sub>int</sub>. However, its location might suggest that there are also some intermediate positions of this expansion segment to interact with L34e.

This would lead to a conclusion that proteins L38e and L34e play a structural role on the ribosome, with a possibility of their function in coordinating the positions of ES27<sup>L</sup>. In

*Tetrahymena*, deletion of ES27<sup>L</sup> is lethal (Sweeney et al., 1994), suggesting a functionally important role for this RNA insertion. Despite the high variability in length of ES27<sup>L</sup>, ranging from ~150 nucleotides in *T. aestivum* and yeast to ~700 nucleotides in mammals (Cannone et al., 2002), the ES27<sup>L</sup> deletion can be complemented with a corresponding ES27<sup>L</sup> from other species (Sweeney et al., 1994). ES27<sup>L</sup> has been suggested to play a role in coordinating the access of non-ribosomal factors, such as chaperones, modifying enzymes, signal recognition particle or translocon, to the emerging nascent polypeptide chain at the tunnel exit (Beckmann et al., 2001). It cannot be excluded that r protein L38e passes ES27<sup>L</sup> to L34e during the stage of binding of the translocon or SRP.



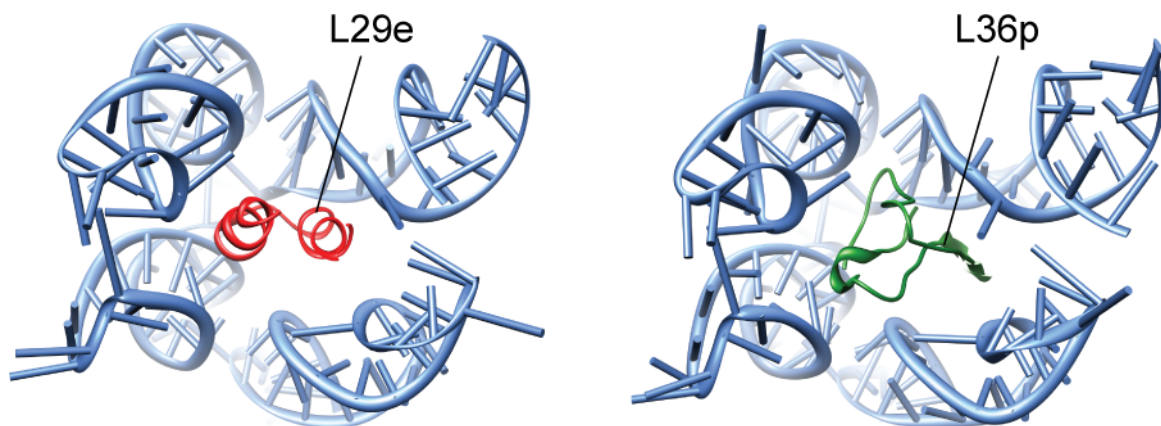
**Figure 40.** Molecular models for the ES27<sup>L</sup><sub>in</sub> (gold) and ES27<sup>L</sup><sub>out</sub> (blue) positions (Left), as observed in *S. cerevisiae* 80S ribosomes (*Thumbnail Insets*) (Beckmann et al., 2001) and an intermediate position (ES27<sup>L</sup><sub>int</sub>, gray) observed in the *T. aestivum* 80S ribosome. In yeast, r protein L34e (green) and L38e (red) interact with the ES27<sup>L</sup><sub>in</sub> and ES27<sup>L</sup><sub>out</sub> positions, respectively. The tunnel exit (TE) and L1 stalk (L1) are indicated for reference.

#### 4.2.2.8 P0-P1/P2 Stalk Base

The localization of protein L29e might be considered speculative, since the only indication of the possible location of this protein is an unusual behavior of the stalk in the yeast  $\Delta$ L29e-80S (see Figure 36). However, this is further backed by the fact that L10e (L16p), being in close

proximity of the assigned position of L29e, exhibits synthetic lethality with L29e in yeast (DeLabre et al., 2002).

L29e sits in a small RNA pocket at the stalk base, which is occupied by L36p in Bacteria (Figure 41). Surprisingly, the same pocket in Archaea is deprived of any proteins (Klein et al., 2004). Protein L29e, like its bacterial-specific counterpart, seems to be involved in rRNA tertiary structure stabilization. Removal of L36p affects a proper positioning of the rRNA regions that contact factors and thus might have an influence on their function, having an effect on the lowered efficiency (Maeder and Draper, 2005). Lack of L29e (as well as L36p) impairs proper protein assembly to the 60S at the interface with 40S, a defect that disrupts subunit joining and diminishes protein synthesis (DeLabre et al., 2002; Dresios et al., 2006; Valle et al., 2003).

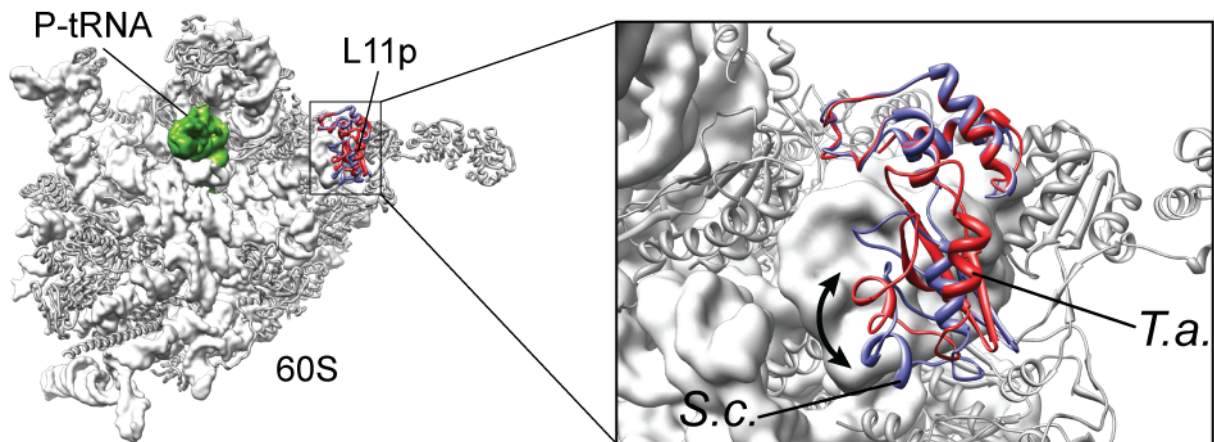


**Figure 41.** Comparison of relative positions of L29e (red) in yeast/*T. aestivum* 80S (Left) with L36p (green) in Bacteria (Selmer et al., 2006) (Right).

A second protein at the stalk base, L11p, is highly conserved and comprises two globular domains connected by a short linker. In our models, we noticed an amino-terminal domain movement between *T. aestivum* and yeast (see Figure 42), which is supported by the behavior of L11p described in ref. Kavran and Steitz, 2007. At the same time, the C terminal domain binds with a higher affinity to the rRNA (Gonzalo and Reboud, 2003) and does not undergo any visible structural changes.

In ref. Becker and al., 2011, L11p is shown to approach the carboxy-terminal part of Dom34, a homologue of eRF1. In Bacteria, r protein L11p, together with Sarcin-Ricin Loop (SRL) and L7/L12 stalk are responsible for activation of GTPase (Agrawal et al., 2001), with L11p rather stimulating the rate of protein synthesis (Uchiumi et al., 2002). Therefore L11p may play a

more specialized role in the factor binding cycle (Kavran and Steitz, 2007), despite being a non-vital protein (Dabbs, 1978), serving rather as a catalyst to a process that could otherwise be diffusion-controlled.



**Figure 42.** Movement of N-terminal part of protein L11p between *T. aestivum* (red) and *S. cerevisiae* (blue) models, as observed also in ref. Kavran and Steitz, 2007.

#### 4.2.2.9 P0-P1/P2 Stalk Proteins

The lateral stalk of the ribosome is a flexible part of the large subunit, constructed from different components, depending on the organism. In Bacteria, the stalk consists of a single L10 and two to three pairs of dimerizing L7/L12 proteins. In Archaea and Eukarya, it is built of a single P0, a homologue of L10 and two to three copies of the P1/P2 dimer (L12p homologues). Protein rpP0 (L10p) is connected to the H42 of large subunit RNA and to a protein L11p (rpL12) at the base of the stalk (Gonzalo and Reboud, 2003).

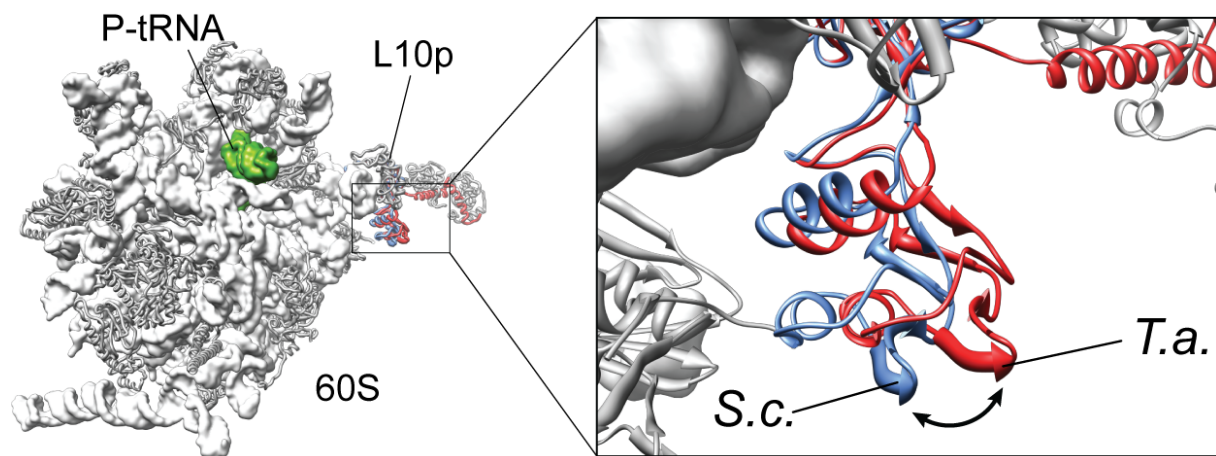
It was shown, that the overall model of this protein is a conjunction of several crystal structures. Initial placement of the domain I of the model from *P. horikoshii* L10p protein template (Naganuma et al.) was based on the position in *H. marismortui* (Kavran and Steitz, 2007). However, in that work the structure of domain II of archaeal L10p was not solved. It should be noted that this region is specific to Archaea and Eukarya only.

A structure of this domain was solved in the archaeon *M. jannaschii* (Kravchenko et al., 2010). Domain II was named ‘internal disorder region’ (Naganuma et al., 2010) and it was shown that a deletion of this region had an effect on the GTPase activity, eEF-2 binding and polyphenylalanine synthesis. From this it was concluded that this domain in Archaea and Eukaryotes might be an aid for eEF2-dependent GTPase turnover.



When the models of yeast and *T. aestivum* r protein P0 are overlaid, it can be noticed that their domain II was positioned differently (see Figure 43). In ref. Kravchenko et al., 2010, the authors suggested that exactly the same movement of this domain might occur. When placed in *T. thermophilus*, the N terminal domain of protein L11p overlaps with domain II of protein P0, and in the immediate vicinity of the domain V and domain G of EF-G, possibly in the position that could inhibit binding of this factor to the ribosome.

Therefore, it does not seem too far-fetched to agree with the statement about domain II that “this domain can be involved in discriminating between archaeal/eukaryotic and bacterial elongation factors” (Kravchenko et al., 2010).



**Figure 43.** A comparison of P0 (L10p) domain II locations between *T. aestivum* (red) and *S. cerevisiae* (blue)

From our structures, it is impossible to derive any functions of P1/P2 proteins, however, previous studies supply certain notion as to their significance. A number of publications indicate that the C-terminal domains of P1/P2 proteins might be interacting with elongation factors (Kopke et al., 1992; Kotliansky et al., 1978). In ref. Bargis-Surgey et al., 1999, a direct interaction in solution was shown between P proteins and eEF-2. In ref. Gonzalo and Reboud, 2003 authors suggested that P proteins might act as GTPase Activity Proteins for eEF-2. Probably a parallel structural evolution occurred between the P proteins and the elongation factors to improve translation efficiency (Gonzalo and Reboud, 2003). In Bacteria, protein L7/12 is probably the most important ribosomal element activating the GTPase of EF-G (Rodnina et al., 2000).

A very interesting fact is that in *S. cerevisiae* mutants lacking all P1 and P2 proteins compared to the wildtype translate selectively a different subset of mRNAs (Remacha et al., 1995; Tchorzewski, 2002).

#### 4.2.2.10 L1 Stalk

The L1 stalk is a flexible lateral protuberance consisting of a single protein, L1p (rpL1) attached to H76 of the 28S rRNA. It is involved in the release of a deacylated tRNA from the ribosomal E-site (Gonzalo and Reboud, 2003; Nikulin et al., 2003). As in all the reported maps, the stalk in the 5.5 Å map confirms the high flexibility of this region. Since the flexibility in cryo-EM maps reduces the resolution and inhibits interpretation, the protein was placed as in the template crystal structures, with disregard to the density distribution. This presents us with a model that does not reflect the true position of L1 stalk in the programmed eukaryotic ribosome, but gives an approximation of its location in one of the states.

#### 4.2.2.11 Examples of Ribosomal Proteins Involved in Diseases

Ribosomal proteins play important (ribosomal and extra-ribosomal) roles in diseases (Freed et al., 2010). We will focus on the former role in two examples – Diamond-Blackfan-Anemia and 5q- syndrome.

We have localized a number of proteins associated with Diamond-Blackfan-Anemia (DBA), a bone-marrow-failure syndrome, associated with mutations in a few ribosomal proteins (Gazda et al., 2008). In addition to a previously reported protein S19e (rpS19), we have localized the other major protein components associated with DBA, such as S17e (rpS17) at the beak of the 40S subunit, S24e (rpS24) at the 40S interface side bridging h8 and h44 and L35ae (rpS33), a 60S protein located between ES7<sup>L</sup> and ES39<sup>L</sup>. Also, ref. Gazda et al., 2008 shows that mutations in L18p (rpL5) and L5p (rpL11) are associated with cleft palate and abnormal thumbs in patients with DBA. Candidate mutations in proteins S7e (rpS7) on the platform at the base of ES6<sup>S</sup>, L36e (rpL36) at the base of the L1 stalk, S19p (rpS15) on the head of the 40S subunit and in r protein S27ae (rpS31 - which has not been localized) were revealed by large-scale screening of ribosomal protein genes in DBA populations (Gazda et al., 2008; Narla and Ebert, 2010).

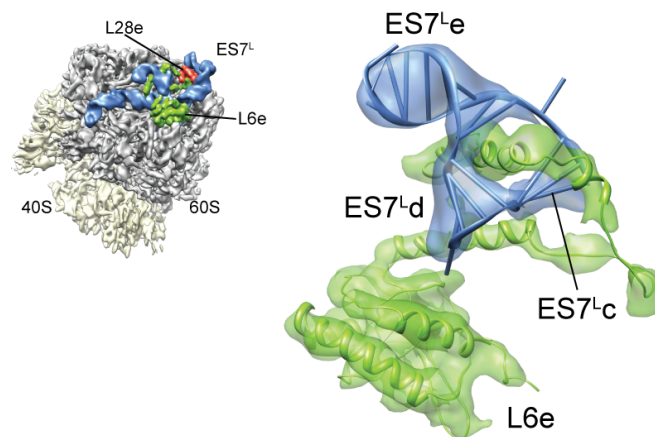
Another disease, the 5q- syndrome, results in refractory anemia and origins in deletion of the long arm of Chromosome 5 (Ebert et al., 2008; Narla and Ebert, 2010). One allele of r protein

S11p (rpS14) is deleted and patients with the 5q- syndrome suffer from haploinsufficient expression of this protein. The decrease in expression of S11p results in impaired erythropoiesis, as well as blocks processing of 18S rRNA and formation of the 40S subunit.

### 4.2.3 Interesting Folds and Interactions

#### 4.2.3.1 L6e Inserts a Helix Into a 3-way Junction of ES7<sup>L</sup>

Protein L6e, a probable gene duplication of protein L14e was found to form a novel protein-RNA interaction. In *T. aestivum*, ES7<sup>L</sup> contains a three-way junction formed by ES7<sup>L</sup><sub>c-e</sub>, through which L6e inserts its C-terminal tail. It was modeled as a helix, and was confirmed by secondary structure prediction and by the strong density (see Figure 44). However, this is so unusual, that it should to be treated with a certain dose of suspicion. To our knowledge, this is a unique, not seen before motif, which has to be further examined.



**Figure 44.** A carboxy-terminal extension of protein L6e in *T. aestivum* passing through three-way junction of ES7<sup>L</sup>

This motif has no counterpart in *S. cerevisiae*; in this organism ES7<sup>L</sup> is different and does not harbor a three-way junction due to the absence of ES7<sup>L</sup><sub>d,e</sub> (Figure 30; Figure S 3 and S 5); protein L6e is also different in this region, being over 40 amino acids shorter.

The motif of a protein inserting its extension through a three-way junction is a feature that is undeniable, independent of the quality of the protein assignment. We cannot neglect the fact that this occurs, whether this is a C-terminal part of L6e or a part of some other ribosomal protein.

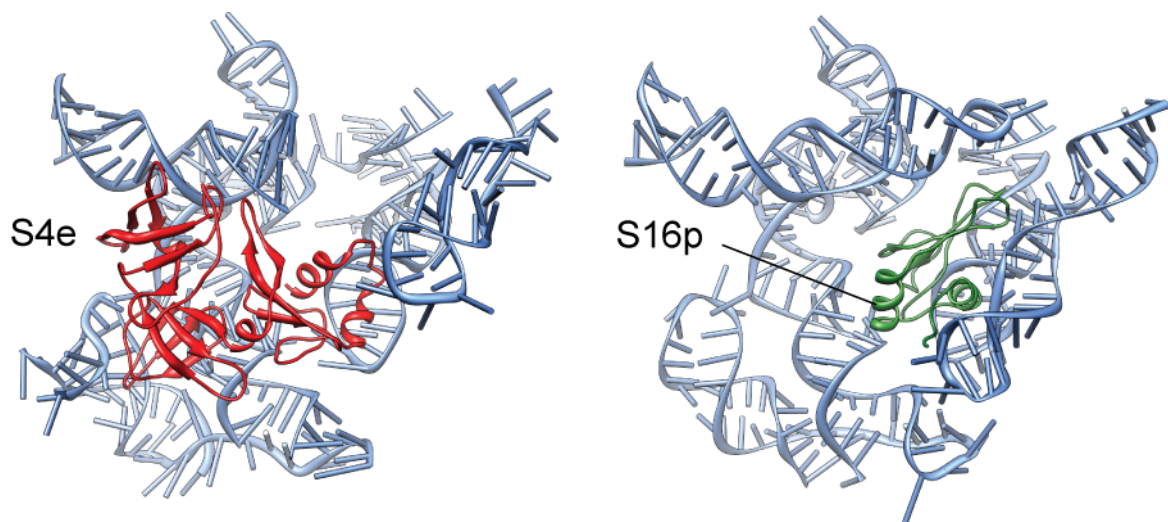
#### 4.2.3.2 L18ae Interaction with rRNA

R protein L18ae was shown to interact with three distinct parts of the large subunit rRNA: 5S, H42 and ES7<sup>L</sup><sub>b</sub>. Acting as a clamp between two parts of rRNA core and a Eukaryote-specific RNA extension, it must have been a rather late addition to the 80S ribosome. Interestingly, it was depicted as a protein that exists only in Eukaryotes, not in Bacteria or Archaea.

Computational methods reveal however that this might not be true. The protein comprises two Archaea-specific LX domains, which would signify that this is a product of gene duplication. What makes it more interesting is that, even though archaeal 70S ribosomes do not contain an extended ES7<sup>L</sup> in any form, this protein is still prevalent in all Archaea. Therefore the gene duplication event of LX could have occurred after the formation of ES7<sup>L</sup>, giving this extended rRNA mode stability (as observed on the example of r protein L28e).

#### 4.2.3.3 S4e Interaction With h16

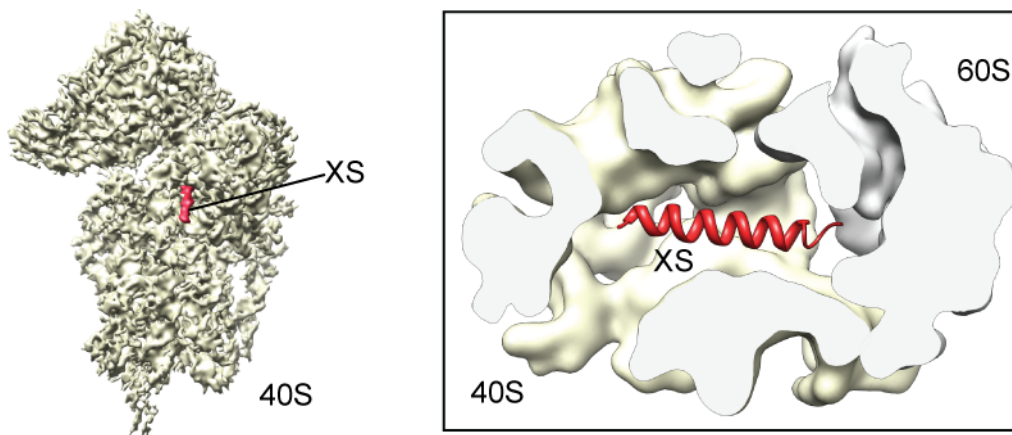
S4e is one among several ribosomal proteins that substitutes a bacterial-specific protein in Eukarya (see Table S 15, Figure 45). It is located at the back of the ribosomal small subunit and is attached to a large number of ribosomal rRNA: h7, h15, ES3S and ES6S, which might indicate its function as a structural scaffold to keep RNA integrity.



**Figure 45.** Comparison of relative positions of S4e (red) in yeast/*T. aestivum* 80S (Left) with S16p (green) in Bacteria (11) (Right).

#### 4.2.3.4 An Unknown Helical Entity Between 40S and 60S

One of the most interesting cases in the ribosome is a short helix located between 40S and 60S subunits, surrounded only by RNA. This helix, approximately 25 amino acids in length, was initially assigned to be L41e, based on its size. It was speculated that this particular protein might exist in two places on the ribosome, being a non-stoichiometric component. However, due to the fact that this helix was also found in a ribosome that should not contain protein L41e, this idea was reconsidered and rejected. Yet, in the recent 40S-eIF1 crystal structure (Rabl et al., 2011), no helix in this spot is reported, while in contrast, it is present in the crystal structure of an 80S (Ben-Shem et al., 2010), which would point to this: (i) being a 60S protein or (ii) being a 40S protein but needing an 80S to bind.



**Figure 46.** An unknown protein entity (XS, red) residing between 40S (yellow) and 60S (grey) subunits.

#### 4.2.4 Proteins on the ribosome

##### 4.2.4.1 Structural Differences and Similarities to Bacteria

Bacterial and eukaryotic ribosomes are in many ways both similar and different. The large (50S) subunit of bacterial ribosomes contains up to 34 proteins and two rRNA molecules: 23S and 5S, while eukaryotic one (60S) contains some 50 proteins and 3 rRNA molecules: 28S, 5.8S and 5S. The 28S and 5.8S rRNA are both related to the bacterial 23S rRNA. The 5.8S rRNA is similar in sequence to the 5' end of the 23S rRNA, hence its existence is most probably a result of an ancient mutation that split the ancestral gene in two. The small subunit in Bacteria is constructed of a 16S rRNA and up to 24 proteins, with its eukaryotic counterpart consisting of 18S rRNA and maximum of 33 proteins. Due to heterogeneous distribution of proteins S1p, S21p, S22p, S31e, L30p, L7ae and L25p (also *Thermus*-specific Thx (Choli et al., 1993)), the exact number of proteins might differ between bacterial species.

There are two major structural differences between bacterial and eukaryotic ribosomes: (i) presence of domain-specific proteins and the (ii) large number of additional rRNA extensions in Eukarya called expansion segments. Bacterial ribosomes contain up to 24 Bacteria-specific proteins that are absent from Eukarya. At the same time, there are 33 proteins that are only present in Archaea and Eukarya and 12 that are only found in Eukaryotes (Lecompte et al., 2002; Sengupta et al., 2004). As stated in ref. Klein et al., 2004: “*The non-homologous proteins that bind the same sites in both archaeal and eubacterial large subunits raise intriguing questions concerning their evolution. A likely and fascinating possibility is that they evolved convergently from different ancestors and now perform the similar function of stabilizing the same region of 23 S RNA*”. There are many proteins that can serve as an example for such a case (Table S15). One such example, protein L44e (rpL42) is located in the same region that in bacterial ribosomes is occupied by L33p, exhibiting a similar fold.

Ribosomal RNA in Eukarya is generally longer and clustered rather on the surface. There are certain places, however, like h33 in the small subunit (the ‘beak’), that exhibit variability in the structure between domains, even though they are not expansion segments. They are called variable regions and are present on both subunits of the ribosome. One such region, H58 might serve as an example: in Bacteria it exhibits a completely different conformation to that of Archaea and Eukarya (see §3.3.2.4 *H58 Interactors*).

#### 4.2.4.2 Structural Differences and Similarities to Archaea

Much more than Bacteria, archaeal and eukaryotic ribosomes share a large number of common features, like the protein composition, RNA structure or factors (Table S 9). This means that the archaeal ribosomes could be regarded as very primitive eukaryotic ones, lacking certain ribosomal proteins and having an RNA structure deprived of expansion segments. It is stated in ref. Lecompte et al., 2002, that “*with the exception of LX, all the r proteins found in Archaea are also found in Eukarya and the archaeal ribosome, which is close to the bacterial one in terms of size, appears to be a small-scale model of the eukaryotic ribosome in term of r-protein composition*”. Lecompte goes farther to show that the difference in the protein distribution between Archaea and Eukarya, as well as between different Archaeal species is the “*first tangible example of reductive evolution observed at a primary domain scale*” (see Table S13 and S14). However, this publication analyzes only 66 different species, comprising 45 Bacteria, 14 Archaea and 7 Eukaryotes. In ref. Márquez et al., 2010, P.

*aerophilum* and *S. acidocaldarius* are shown to possess additional ribosomal proteins, unique to those species and not found in Eukaryotes, which contradicts Lecompte et al., 2002.

The primary rRNA composition in Archaea and Eukarya is not very different, with the former comprising three RNA molecules and the latter four. The large subunit in Archaea consists of 23S and 5S molecules (with an exception of *Aeropyrum pernix* which contains 5.8S rRNA), while the eukaryotic one of 25S (or 28S, depending on the species), 5.8S and 5S RNA; the small subunits are built from 16S and 18S RNA, respectively. Eukaryotic ribosomes are generally larger, having additional set of proteins as well as rRNA expansion segments. This difference is however, not always obvious, ie. *Giardia muris*, an intestinal protozoan (Eukarya) parasite of rodents. A comparison of 16S and 23S rRNA from *G. muris* and *H. marismortui*, reveals that the sizes in the former are 1432 and 3071 nt against 1472 and 2923 nt in the latter. The differences are larger with the more evolved organisms – 16S in *H. sapiens* is 1871 nt, 28S is 5184.

When compared in detail, the archaeal ribosome is puzzling. The 16S rRNA is very similar to that of Bacteria, with one exception, namely, h33 forming the beak. This region is identical in Archaea and Eukarya, and at the same time completely different to that of Bacteria. Therefore it can be said that the small subunit in Archaea is a chimera between Eukarya and Bacteria.

#### 4.2.4.3 Structural Differences and Similarities Between Eukarya

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 1.6 MDa (0.53 MDa/40S and 1.07 MDa/60S) of rRNA, thus totaling 3.11 MDa (see Table S 7), whereas *E. coli* 70S ribosomes total to ~2.5 MDa (0.9 MDa/30S and 1.6 MDa/50S). Figure 48 shows that the ES and additional r proteins/protein extensions (green and gold, respectively) form an intricate layer of additional RNA-protein mass that locates predominantly to the solvent surface of the ribosome. The intertwined nature of the additional rRNA ES and r proteins supports the idea that they are coevolving together (Yokoyama and Suzuki, 2008), which is exemplified by the large mass found on the back of the 60S subunit comprising ES7<sup>L</sup>, ES39<sup>L</sup>, and five eukaryotic r proteins (L6e, L14e, L18ae, L28e, and L35ae – see Figure 29).

The models of *S. cerevisiae* and *T. aestivum* eukaryotic ribosomes are surprisingly similar, showing signs of divergence in only a few places. In all Eukaryotes, the number of proteins is

stable, *S. cerevisiae* being a notable exception because of lack of r protein L28e; also, some plants contain additional P1/P2 dimers. Human ribosomes have the same 80 r proteins that are found in *T. aestivum* ribosomes and, in terms of rRNA, differ significantly only in the length of four ES on the large subunit (ES7<sup>L</sup>, ES15<sup>L</sup>, ES27<sup>L</sup> and ES39<sup>L</sup>). These are longer in human (~850, ~180, ~700, ~220 nts) than in *T. aestivum*/yeast (~200, ~20, ~150, ~120 nts, respectively) and cryo-EM reconstructions of mammalian ribosomes show that the longer ES in mammalian ribosomes are generally highly mobile elements for which little to no density is visible (Figure 49). The differences in ribosomal rRNA sizes within this domain are vast, as shown on *G. muris* and *H. sapiens*. Table S 13 and Table S 14 report the sizes of ribosomal protein in *S. cerevisiae*, *T. aestivum*, *D. melanogaster* and *H. sapiens* and confirm that the majority of them remain stable within their protein family. Some of the proteins located in the variable regions might differ between organisms (the extreme case is L6e), and could be explained as a support for growing RNA expansions.

#### 4.2.5 Reliability of the Model

When trying to assess the reliability of the model, confidence can unfortunately not be expressed with a number. Crystallographers use the R-factor, a measure of how the experimental diffraction data agrees with the model. In cryo-EM, the assessment is more difficult, and should rather be based on a comparison with a high-resolution X-ray model. However, without those in hand, one can only estimate it according to the available data: the reliability of the map, secondary structure predictions, homology searches, experimental data and the literature.

##### 4.2.5.1 Density Map

Calculation of resolution of cryo-EM maps represents an estimation of the quality of the reconstruction. It is assumed that the better the resolution, the better the quality, therefore a map at 4 Å should be better than another at 6 Å. However, the resolution does not always reflect the level of detail, hence to assess the quality of the map one has to perform an analysis of known features. Our structure was compared to known rRNA and r protein structures (*H. marismortui*, *T. thermophilus*), and indeed, it contained features that are visible at a nominal resolution of 5.5 Å. Several parts of the map, namely the beak, mRNA exit, left and right foot of the small, and both lateral stalks of the large subunit were of a lesser quality, due to a larger degree of flexibility. The core parts of the map possibly exhibit a better detail level, because



proteins and rRNA are more strongly stabilized. To ensure the highest map homogeneity, sorting for the P-site tRNA was performed. However, there are certainly small differences in conformations that cannot be resolved at this resolution. In addition to that, the map is a sum of eight constructs, which were merged to form the 5.5 Å map., which also has an influence on the overall reconstruction.

#### 4.2.5.4 Dependence on Cross-linking and Immuno-EM studies

When assessing the quality of the protein localization and, by extension, of modeling of those proteins, cross-linking and immuno-EM data has to be evaluated. This will be only possible when a high-resolution crystal structure of a eukaryotic ribosome appears. However, from assessment of the crystal structure of 40S-eIF1 *T. thermophilus* 3.9 Å (Rabl et al., 2011) it seems that those studies are compatible.

#### 4.2.5.5 Dependence on Computational Methods

Search for proper templates for proteins and RNA is dependent on computational analysis. Secondary structure predictions are a standard now, constantly improving, with PSI-PRED (Jones, 1999) averaging at 75-80% accuracy. A search for homologues using HHPred, is also very good, yielding trustworthy results, presenting a list of possible relatives of the requested sequences. However, even with the best initial model, interactions with rRNA and other r proteins might enforce a change to structure. The same applies to rRNA, since the interactions with other rRNA and proteins might potentially result in refolding or a change in conformation. Therefore the full assessment of the quality of the computational methods should be examined with high-resolution X-ray structure of an 80S ribosome.

#### 4.2.5.1 Modeling and Metal Ions

Modeling of our structure, due to a limited resolution, did not include placement of metal ions. In nature, metal ions are important both for RNA as well as for protein folding. As stated in ref. Draper, 2004: “*RNA folding into stable tertiary structures is remarkably sensitive to the concentrations and types of cations present*”. It is known that “*without cations, the repulsive forces generated in the close-packed structure [of RNA] would overwhelm the energetically favorable interactions that dictate the proper three-dimensional structure*” (Gesteland et al., 2006). In proteins, metal-dependent folding requires a coordination of metal ions (one or

more) to direct the folding of the polypeptide into the completely functional conformation. Among of the most common examples are the folding motifs of the Zinc-finger domains (i.e. protein S27ae, L40e). To compensate for the lack of the ions, secondary-structure and backbone restrictions were imposed both on RNA and proteins in Molecular Dynamic Flexible Fitting, which resulted in a fairly equilibrated structure.

#### **4.2.5.3 Interpretation**

The interpretation of the cryo-EM density map involves a human factor, making it a subjective process. Thus, in case of uncertainty during the modeling, the same task could be performed differently from one person to another. There is unfortunately no way to address the human factor influence in terms of mathematical analysis.

#### **4.2.6 Comparative Analysis with Crystal Structures**

During the course of writing this dissertation, a first crystal structure of 80S yeast ribosome at 4.15 Å (Ben-Shem et al., 2010) was released. Also, a 3.9 Å crystal structure of 40S-eIF1 complex from *T. thermophila* was published (Rabl et al., 2011). These publications allowed for a direct comparison of the rising-tide of cryo-EM modeling and established technique of X-ray crystallography.

The results of this comparison are presented below.

##### **4.2.6.1 The Structure of an 80S *S. cerevisiae* Ribosome at 4.15 Å**

In 2010, a structure of the yeast 80S empty ribosome was released as the first crystal structure of the eukaryotic ribosome. The assignments of the proteins, as well as protein and RNA extensions and RNA allow a comparison between the structure discussed in this dissertation and the crystal structure mentioned.

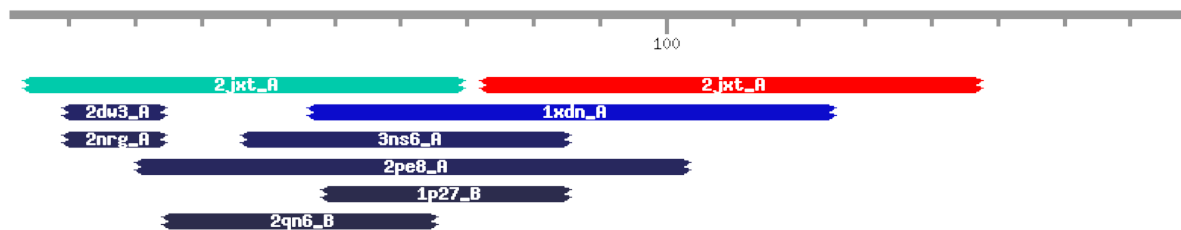
###### **4.2.6.1.1 Similarities and Differences**

The structure of yeast at 4.15 Å and our model at 5.5 Å share a surprising number of similarities, both in rRNA and in r protein content. The Yusupov group modeled all the proteins that have a prokaryotic template, in addition localizing and modeling several new

ones. When compared, the locations of protein S17e, S19e, S25e, S28e, L14e, L30e and RACK1 are the same, as is the majority of the protein extensions, like L10e, L13p or L30p. In both models, similar interactions are observed for the mRNA entry and exit, showing connectivity by a strong interaction between S2p and S5p.

The main difference that can be noticed is the positioning of protein L6e. This Eukaryote-specific protein was modeled in the position that in our model is partially occupied by protein L18ae. The reasons for our placement of protein L6e in a different location are shown in Table S 1 (cross-linking studies, differences between organisms). Localization of protein L18ae also had a strong backing, namely: search for a fold and comparisons between eukaryotic and archaeal reconstructions. Since protein L6e is a Eukarya-specific protein, in the archaeal ribosome the same position should not be occupied by any protein. This is due to the reason that except for protein LX, there are no other archaeal-specific proteins. We observe something different – there is a density present in a reconstruction of a 70S *M. igneus* (see Table 7, Figure 31). This means that in this position binds a protein that is present both in Eukarya and in Archaea (therefore, we would argue that that this cannot be protein L6e).

L18ae is not such a protein, however, HHPred reveals that it consists of two domains, each domain that has an average of 17% identity with protein LX, predicted with 99 and 60% probability respectively (see Figure 47). This suggests that protein L18ae in Eukaryotes emerged through a gene duplication event of protein LX. When LX is placed in the density in Archaea, the fold of the protein seems fitting (see Figure 31). Hence, we suggest that the assignment was not correct in the crystal structure.



**Figure 47.** HHPred prediction for the protein L18ae. The first hit shows that it consists of two domains, each domain being homologous to LX, first domain with 99% probability, the second with 60% (Biegert et al., 2006; Soding et al., 2005).

In addition, several protein extensions are placed slightly different, either due to different contacts in the crystal structure, mistakes in assignment on either side or the fact that disordered regions become visible in certain conditions, as in the case of protein L10e.

When comparing rRNA, one can see that in general, with the exception of h17, ES6<sup>S</sup> and ES39<sup>L</sup>, the structures are quite similar. In the region of ES39<sup>L</sup> and ES6<sup>S</sup>, we were not able to visualize single-stranded linkers, unlike in the X-ray structure. Because of this, helix assignment in ES6<sup>S</sup> in the cryo-EM density was more ambiguous and resulted in an inverted assignment of two helices (ES6<sup>S</sup><sub>a</sub> and ES6<sup>S</sup><sub>c</sub>). The rRNA core is identical in both models, with most of the expansion segments highly similar, especially ES4<sup>L</sup>, ES7<sup>L</sup>, ES9<sup>L</sup>, ES20<sup>L</sup>-ES26<sup>L</sup>, ES31<sup>L</sup> and ES41<sup>L</sup>. Minor register differences in the expansions result in differences in the loop regions, but in overall the structures exhibit a surprising likeness.

#### 4.2.6.1.2 Conclusions

Crystal structure at 4.15 Å is not considered to be the limit of the technique. At this resolution, there is little-to-no information about the position of the side chains. The map at this resolution is not significantly better than the one obtained by us, however, in most cases authors were more timid to localize and model proteins into the density.

The models for the ribosomal RNA as well as the proteins are surprisingly similar to results presented in this dissertation, despite a 1.3 Å difference in resolution. The major differences are observed mainly in ES39<sup>L</sup> and ES6<sup>S</sup>. In the crystal structure, an improvement in certainty of assignment of rRNA entities in these regions is seen due to the visualization of linker regions. All other rRNA regions are modeled reliably; ES7<sup>L</sup> in our structure is even more complete, due to the absence of the flexible arm of ES7<sup>L</sup><sub>a</sub> in the crystal structure, which we successfully modeled.

It seems that the authors of this structure focused predominantly on the rRNA and less on the proteins, most probably due to an upcoming 3 Å crystal structure (Ben-Shem, A., unpublished). However, based on the available models, it is safe to state that both structures are also quite similar when assigning the protein densities. The group of Prof. Yusupov modeled all the proteins that could be based on the prokaryotic templates, in addition localizing and confirming the location of some proteins that were not crystallized on the ribosome before. Except for the location of protein L6e, this agrees with our findings.

#### 4.2.6.2 The Structure of a 40S-eIF1 *T. thermophila* Complex at 3.9 Å

At the beginning of 2011, a structure of a 40S-eIF1 complex from *Thermomyces thermophila* at 3.9 Å was reported from the group of Prof. Nenad Ban. In this structure, the complete set of small subunit proteins, as well as the whole ribosomal SSU rRNA was modeled.

##### 4.2.6.2.1 Similarities and Differences

The ribosomal rRNA is extremely similar to that of the structure from the Yusupov group, therefore the same conclusions can be reached from the RNA point of view. An assignment of two helices of ES6<sup>S</sup> (ES6<sup>S</sup><sub>a</sub> and ES6<sup>S</sup><sub>c</sub>), the same in both crystal structures, in our model is switched, due to the absence of linker regions. The Ban group localized all the small subunit proteins; therefore a comparison of the assignments of those proteins is unavoidable. After careful examination, we confirmed our assignment of eight novel proteins: S4e, S7e, S19e, S25e, S26e, S27e, S28e and RACK1. For proteins S4e, S19e, S25e, S27e, S28e and RACK1, we were in possession of X-Ray or NMR structures (or very reliable homologues), thus there are no significant differences in their fold. That was not the case with S7e and S26e - even though their location prediction is correct, their fold is not, due to insufficient density connectivity and absence of reliable homologues. Proteins assigned differently were: S17e, S21e, S24e and S30e. Protein S17e was pointed by us, as well as the Yusupov group, to reside in the beak, based on the immuno-EM studies (Gross et al., 1983). There, a three-helix bundle density was found, which could accommodate the known structure (Wu et al., 2008b). However, a similar density was also found under RACK1 and was assumed to form an extension of protein S2p. In *T. thermophila*, the C-terminal part of S2p is much shorter than in *S. cerevisiae* and *T. aestivum*, but the density under RACK1 prevails. This means that this is not a location for S2p extensions, which is very likely disordered. In the 40S crystal structure, this exact density was used for S17e placement.

In place of protein S30e in our structure, a protein S27ae was modeled, with S12e in its direct vicinity. The beak in the cryo-EM density is a not a well-resolved region, which, in conjunction with the lack of sufficient structural data resulted in misassignment of this protein.

N-terminal part of protein S3ae was modeled in the 40S X-ray structure in place of protein S21e (but sharing a very similar fold to the one modeled *de novo* by our group). The location of N-terminal part of protein S6e modeled by the Ban group is where we placed protein S24e

(they have a similar fold). There are at least two reasons for the different assignment of the proteins. The resolution of the cryo-EM density map did not solve all the connectivity, even when using the deconvoluted map. The higher degree of flexibility of the small subunit makes it more difficult to model, for locations like the beak or the feet exhibit a more ambiguous density than that at the tunnel exit site or the large protuberance of the ribosomal large subunit. Another reason for the misassignment lies in a more general placement of the proteins in the resource publications like (Bommer et al., 1991; Gross et al., 1983; Lutsch et al., 1990). Hence, the amount of unassigned density, lack of more crystal structures to place and a certain ambiguity in some parts of the map lead to misassignment of a few small- and medium-sized proteins.

#### 4.2.6.2.2 Conclusions

The structure of the small subunit at 3.9 Å conveys more information than the one delivered from the group of Prof. Yusupov and the cryo-EM structure presented here. As shown in §4.2.1.1 *Reliability of the Localization*, the protein assignments that were depicted as incorrect, received low localization reliability, however, followed in general the immuno-EM and cross-linking studies done in the past. The 5.5 Å structure was of sufficient quality to localize eight novel small subunit proteins, model extensions that were confirmed by the Ban structure and to model the rRNA quite reliably. As for the localization of the proteins in the large subunit – there are many more studies on the fold, interaction and location of the proteins in the 60S. It is therefore very likely that the novel high-resolution structures will yield a confirmation of a much larger number of novel proteins in the ribosome.

### 4.3 Ribosome in the Light of Evolution

The ribosome is an enormous machine that has gone through many changes since the time of its conception. Not much is really known about its origins. The ribosome field has gone through many research studies to arrive at what is considered currently as the dominant theory: the ribosome originates in the so-called ‘RNA world’.

Before the catalytic RNA was discovered, a number of theories existed about the role of RNA in the ribosome. At the beginning it was clear that since known enzymes were proteins, the proteins are responsible for the peptide-bond-formation. It was puzzling therefore why the ribosome would need RNA. rRNA was considered either a scaffold or a determinant for the

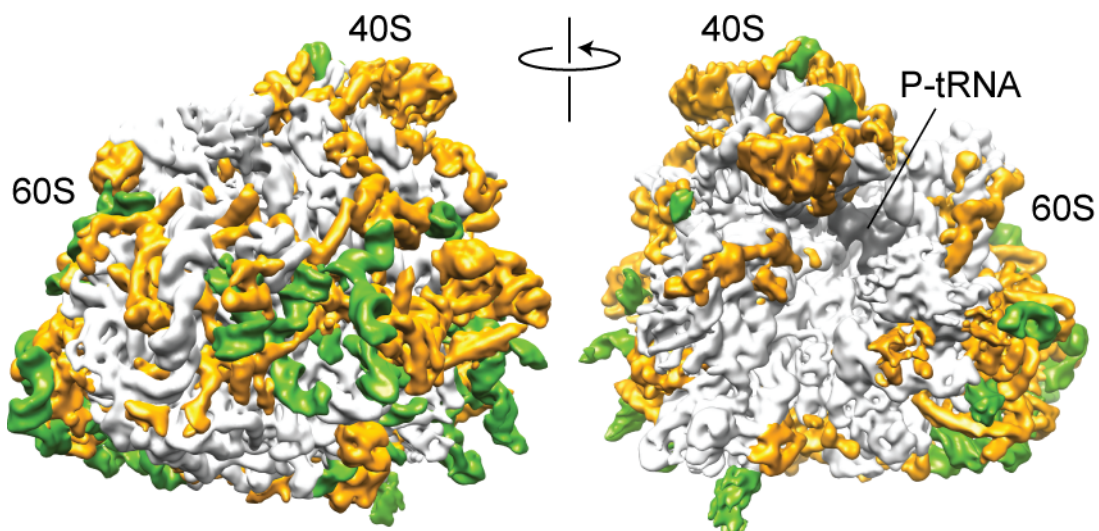
sequence of proteins the ribosome was producing. The first theory was considered to be faulty, since it made no evolutionary sense and when mRNA was discovered, the second theory was disproved as well. Therefore, the idea of RNA participating directly in the protein synthesis was steadily gaining ground until the discovery of RNA with a catalytic activity (Cech et al., 1981; Guerrier-Takada et al., 1983). The full certitude of the importance of RNA was reached with the emergence of high resolution structures pointing directly to rRNA as the active part in protein synthesis. The same structures pointed to the role of ribosomal proteins as largely structural (Moore and Steitz, 2002).

This makes us ask an unavoidable question: what is the start point of the ribosome, its ‘Big Bang’ and what is the future of the structure?

As stated in ref. Tamura and Alexander, 2004: “[...] *emergence from the RNA world into the protein world could have been mediated by catalytic RNAs, first in a nontemplated fashion, then according to a developing genetic code*”. In ref. Bokov and Steinberg, 2009, it was shown that the primary domain for the ribosome activity was a part of the domain V, in which the peptide bond formation occurs. The ribosome possibly was a simple catalytic RNA, (ie. hammerhead RNA (Pley et al., 1994)), which through sequential additions of layers of proteins and RNA was gaining more stability, functions and precision. It was maybe also better in dealing with undesired molecules – an added layer of proteins or RNA that are not active during the translation might supply the binding surface for antibiotics, without influencing the synthesis.

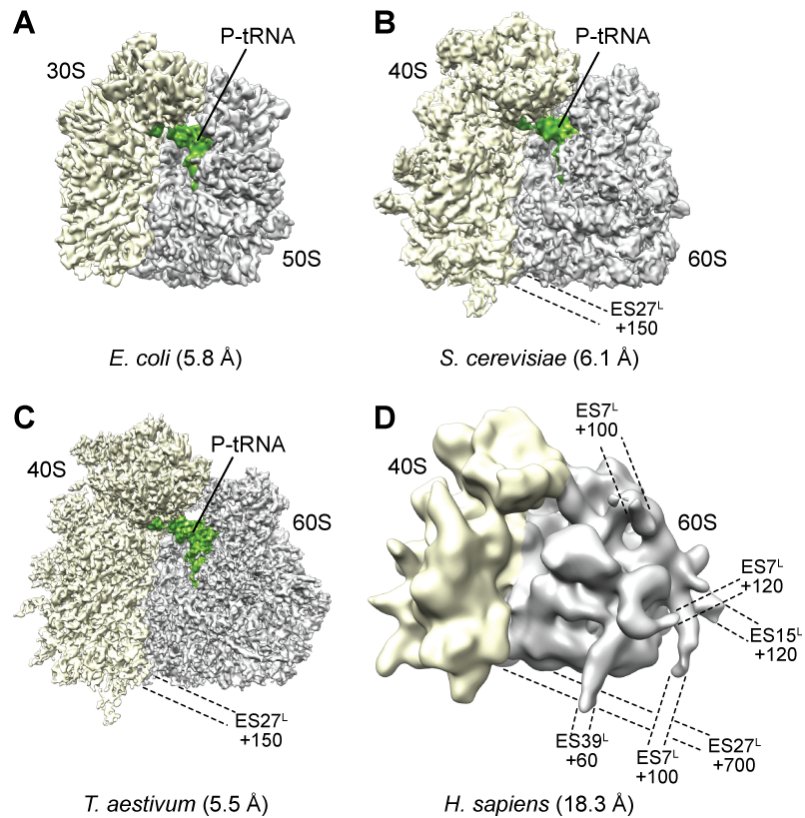
A previous comparison of archaeal and bacterial large subunits illustrated examples of potential convergent evolution, where evolutionarily unrelated r proteins have evolved to stabilize the same region of 23S rRNA (Klein et al., 2004). Many such examples are also found by comparing the models of the yeast and *T. aestivum* 80S ribosome with the archaeal and bacterial crystal structures: The N-terminal domain of S4e overlaps the binding position of S16p (Figure 45), and the extended N-terminus of L32e overlaps regions of bacterial-specific r proteins L20p and L21p. Likewise, L18ae has two ubiquitin-like  $\alpha/\beta$  roll domains (ULDs), with the N-terminal ULD overlapping bacterial L25p, and like L25p also interacting with the 5S rRNA, whereas alpha-helix 1 of the C-terminal ULD inserts in the minor groove of H41. Furthermore, L29e sits in an RNA pocket at the P0-P1/P2 stalk base, which in Bacteria is occupied by L36p (Figure 41) and devoid of proteins in Archaea (Klein et al., 2004). A comparison of genomic sequences from diverse organisms, ranging from Bacteria to mammals, indicates additional mass with increasing organism complexity (Figure 49).

However, the composition of mammalian ribosomes, e.g. from human, is surprisingly similar to those of other Eukaryotes, such as yeast and plants described here. Evolution has, thus, favored the development of two apparently distinct layers of mass gain for the ribosome: A first layer of tightly intertwined additional proteins and rRNA expansions rigidly positioned on the subunit surfaces (with the only exception of the mobile ES27<sup>L</sup>), that was followed by a second layer comprising a few drastically extended highly mobile rRNA elements with hitherto unknown function (Figure 48). The information gained from the *T. aestivum* and yeast 80S models (Figure 50) should, therefore, not only provide a resource for researchers working with these model organisms, but may also provide useful information when studying mammalian systems. However, it will not shed light on whether there is a size limit to the ribosome. Will it continue to grow forever, continuously gaining size and functions for the extensions, acquiring new proteins and more factors needed for the control over synthesis, or has it reached its limit?



**Figure 48.** Cryo-EM map of the *T. aestivum* 80S ribosome, with rRNA ES and variable regions colored green and Eukaryote-specific r proteins and extensions colored orange





**Figure 49.** Cryo-EM reconstructions of ribosomes from (A) the eubacterium *Escherichia coli* (Seidelt et al., 2009), (B) the yeast *S. cerevisiae* (Becker et al., 2009), (C) *T. aestivum* this work), and (D) *Homo sapiens* (Spahn et al., 2004b). The small and large subunits are shown in yellow and gray, respectively and the P-tRNA (green) is indicated for reference. The dashed lines and numbers indicate the number of nucleotides of the rRNA expansion segments that are not visualized.

## 5 Summary

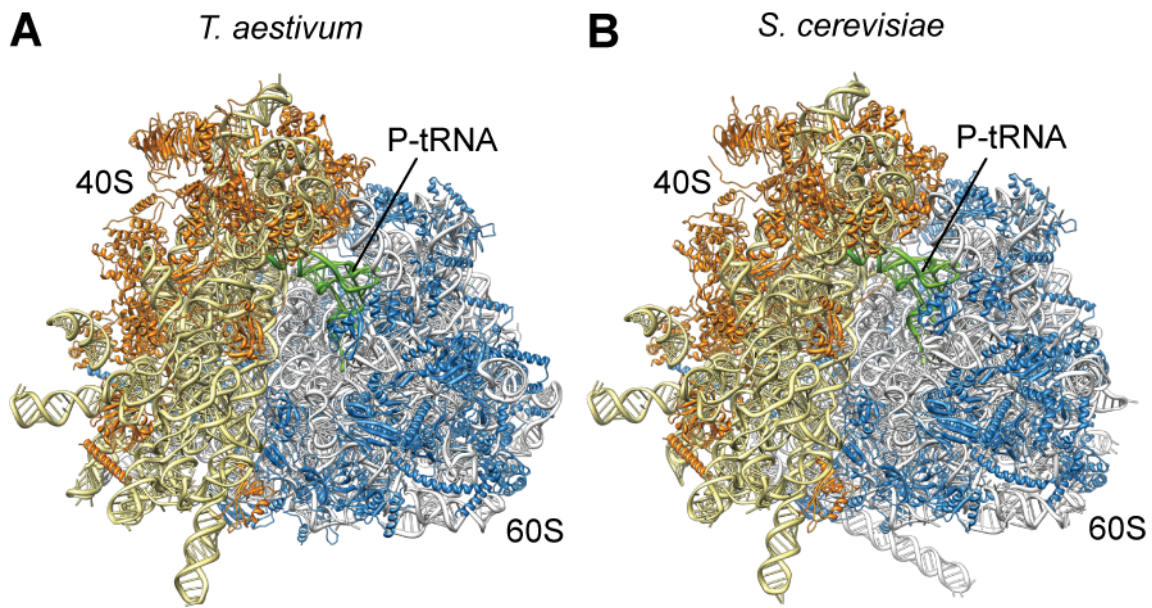
Macromolecular particles called ribosomes exist in all domains of life, performing one common task – protein synthesis from an mRNA template, called translation.

All ribosomes are composed of RNA and proteins. Their final architecture became clear in the recent years. 70S bacterial ribosomes and a 50S archaeal ribosomal subunit were crystallized in the last decade, however an accurate atomic structure of a eukaryotic 80S ribosome has been unavailable.

In this dissertation, the most complete structure of a eukaryotic 80S ribosome to date was presented (Figure 50), refined to 5.5 Å resolution using cryo-electron microscopy and single particle reconstruction. Coupled with RNA and protein homology modeling, molecular models of *S. cerevisiae* and *T. aestivum* ribosomes were obtained. All Eukaryote-specific RNA expansion segments were modeled, constituting approximately 98% of the RNA mass. We have localized and modeled 27 novel proteins (15 in the 60S subunit and 12 in the 40S subunit, 74 out of 80 r proteins) and modeled de novo over 2000 amino acids of extensions of known proteins.

The approach that led to obtaining the highest-resolution cryo-EM single particle reconstruction of a Eukaryotic ribosome is described. Density maps of ribosomes from yeast knockout strains and archaeal organisms, genomic differences between *T. aestivum* and *S. cerevisiae*, as well as literature resources were presented as means for protein localization and elongation.

With the structure of eukaryotic ribosomes, a comparison with archaeal and bacterial ribosomes was performed, pointing the most interesting regions and supplying first explanations for the encountered differences. Furthermore, a tentative analysis and summary of the functions of all the modeled ribosomal entities was undertaken.



**Figure 50.** Structures of wheat germ and yeast eukaryotic 80S ribosomes. (A and B) Near-complete molecular models for the (A) *T. aestivum* and (B) *S. cerevisiae* 80S ribosome, with rRNA and protein shown in yellow and orange for the small subunit and gray and blue for the large subunit, respectively.

Taken together, this analysis provides a near-complete structural inventory of the 80S eukaryotic translational apparatus. Studies of higher resolution are required in order to confirm assignments, identify errors and fill-in the remaining gaps. Nevertheless, the present work will set the stage for a molecular understanding of the eukaryotic ribosome's mode of interaction with its diverse ligands and its dynamic behavior. As stated by Venkatraman Ramakrishnan (Ramakrishnan, 2011): "*This structure [of 40S-eIF1 complex from *T. thermophila*], together with a crystal structure of the eukaryotic ribosome from the yeast *Saccharomyces cerevisiae* at a resolution of 4.15 Å (Ben-Shem et al., 2010) and a cryo-electron microscopy (cryo-EM) structure of a translating plant ribosome from *Triticum aestivum* at 5.5 Å (Armache et al., 2010a, b), represents a breakthrough in studying translation in eukaryotes.*"

## 6 Appendix

---

AA	<u>A</u> mino <u>A</u> cid
AAP	<u>A</u> rginine <u>A</u> ttenuator <u>P</u> eptide
A-tRNA	<u>A</u> -site tRNA
Å	<u>Å</u> ngstrom
CMV	<u>C</u> ytomegalovirus
CTF	Contrast Transfer Function
Da	<u>D</u> alton
DNA	<u>D</u> eoxyribonucleic <u>A</u> cid
EM	<u>E</u> lectron <u>M</u> icroscopy
ES	<u>E</u> xpansion <u>S</u> egment
E-tRNA	<u>E</u> -site tRNA
mRNA	<u>M</u> essenger <u>R</u> NA
NC	<u>N</u> ascent <u>C</u> hain
NT	<u>N</u> ucleotide
PTC	Peptidyl Transferase Center
P-tRNA	<u>P</u> -site tRNA
RNA	<u>R</u> ibonucleic <u>A</u> cid
rp(s)/r protein(s)	<u>R</u> ibosomal <u>P</u> rotein(s)
rRNA	<u>R</u> ibosomal <u>R</u> NA
S/N	<u>S</u> ignal to <u>N</u> oise
SRP	<u>S</u> ignal <u>R</u> ecognition <u>P</u> article
SS	<u>S</u> econdary <u>S</u> tructure
tRNA	<u>T</u> ransfer <u>R</u> NA
WT	<u>W</u> ild <u>T</u> ype

---

**Table S 1.** Nomenclature for small subunit r proteins of the *Saccharomyces cerevisiae* and *Oryza sativa/Triticum aestivum*

	Family name	<i>S. cerevisiae</i> name	<i>Oryza sativa</i> name
	Family name	<i>S. cerevisiae</i> name	<i>Triticum aestivum</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 S 2.** Nomenclature for large subunit r proteins of the *Saccharomyces cerevisiae* and *Oryza sativa/Triticum aestivum*.

	Family name	<i>S. cerevisiae</i> name	<i>Oryza sativa</i> name
	Family name	<i>S. cerevisiae</i> name	<i>Triticum aestivum</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 S 3.** Summary of modeled *T. aestivum* small subunit r proteins. In bold – newly localized proteins.

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length, aa	Modeled range, aa	Percent modeled, %	Template	PDB ID
Sa	S2p	<i>Oryza sativa</i>	Q8H3I3	305	260	1-260	85%	<i>Thermus thermophilus</i>	2J00_B
S2	S5p	<i>Oryza sativa</i>	Q84M35	274	263	1-263	96%	<i>Escherichia coli</i>	2QAL_E
S3	S3p	<i>Triticum aestivum</i>	Q8L804	227	208	12-219	92%	<i>Escherichia coli</i>	2QAL_C
<b>S4</b>	<b>S4e</b>	<b><i>Oryza sativa</i></b>	<b>P49398</b>	<b>265</b>	<b>200</b>	<b>43-242</b>	<b>75%</b>	<b><i>Thermoplasma acidophilum</i></b>	<b>3KBG_A</b>
S5	S7p	<i>Oryza sativa</i>	Q93VC6	200	191	10-200	96%	<i>Pyrococcus horikoshii</i>	1IQV_A
<b>S7</b>	<b>S7e</b>	<b><i>Triticum aestivum</i></b>	<b>Q517K2</b>	<b>192</b>	<b>143</b>	<b>1-143</b>	<b>74%</b>	<b>Poly-alanine</b>	-
S9	S4p	<i>Oryza sativa</i>	Q2R1J8	195	195	1-195	100%	<i>Thermus thermophilus</i>	2J00_D
S11	S17p	<i>Oryza sativa</i>	Q7XIK5	161	85	40-124	53%	<i>Thermus thermophilus</i>	2J00_Q
S13	S15p	<i>Oryza sativa</i>	Q69UI2	151	121	31-151	80%	<i>Escherichia coli</i>	2QAL_O
S14	S11p	<i>Oryza sativa</i>	Q6H7T1	150	119	32-150	79%	<i>Thermus thermophilus</i>	2J00_K
S15	S19p	<i>Oryza sativa</i>	P31674	154	91	58-148	59%	<i>Escherichia coli</i>	2QAL_S
S16	S9p	<i>Oryza sativa</i>	Q0IQF7	149	126	24-149	85%	<i>Thermus thermophilus</i>	2J00_I
<b>S17</b>	<b>S17e</b>	<b><i>Oryza sativa</i></b>	<b>Q7XEQ3</b>	<b>141</b>	<b>141</b>	<b>1-141</b>	<b>100%</b>	<b><i>Methanobacterium thermoautotrophicum</i></b>	<b>1RQ6_A</b>
S18	S13p	<i>Triticum aestivum</i>	Q8L806	152	152	1-152	100%	<i>Escherichia coli</i>	2QAL_M
<b>S19</b>	<b>S19e</b>	<b><i>Oryza sativa</i></b>	<b>P40978</b>	<b>146</b>	<b>146</b>	<b>1-146</b>	<b>100%</b>	<b><i>Pyrococcus abyssi</i></b>	<b>2V7F_A</b>
S20	S10p	<i>Oryza sativa</i>	P35686	128	128	1-128	100%	<i>Thermus thermophilus</i>	2J00_J
<b>S21</b>	<b>S21e</b>	<b><i>Oryza sativa</i></b>	<b>P35687</b>	<b>82</b>	<b>82</b>	<b>1-82</b>	<b>100%</b>	<b>Poly-alanine</b>	-
S15a	S8p	<i>Oryza sativa</i>	Q84AP1	130	130	1-130	100%	<i>Escherichia coli</i>	2QAL_H
S23	S12p	<i>Oryza sativa</i>	Q8L4F2	142	142	1-142	100%	<i>Thermus thermophilus</i>	2J00_L
S24	S24e	<i>Oryza sativa</i>	Q6H541	138	98	5-102	71%	<i>Pyrococcus abyssi</i>	2V94_A
<b>S25</b>	<b>S25e</b>	<b><i>Oryza sativa</i></b>	<b>Q53QG2</b>	<b>108</b>	<b>100</b>	<b>9-108</b>	<b>93%</b>	<b><i>Pyrococcus horikoshii</i></b>	<b>1UB9_A</b>
<b>S26</b>	<b>S26e</b>	<b><i>Oryza sativa</i></b>	<b>P49216</b>	<b>133</b>	<b>92</b>	<b>1-31 ; 73-133</b>	<b>69%</b>	<b>Poly-alanine</b>	-
<b>S27</b>	<b>S27e</b>	<b><i>Oryza sativa</i></b>	<b>Q6K5R5</b>	<b>86</b>	<b>50</b>	<b>33-82</b>	<b>58%</b>	<b><i>Archeoglobus fulgidus</i></b>	<b>1QXF_A</b>
<b>S28</b>	<b>S28e</b>	<b><i>Triticum aestivum</i></b>	<b>Q7X9K4</b>	<b>65</b>	<b>58</b>	<b>1-58</b>	<b>89%</b>	<b><i>Pyrococcus horikoshii</i></b>	<b>1NY4_A</b>
S29	S14p	<i>Triticum aestivum</i>	Q517K3	56	48	9-56	86%	<i>Thermus thermophilus</i>	2J00_N
<b>S30</b>	<b>S30e</b>	<b><i>Oryza sativa</i></b>	<b>Q6K853</b>	<b>62</b>	<b>62</b>	<b>1-62</b>	<b>100%</b>	<b>de novo</b>	-
RACK1	RACK1	<i>Triticum aestivum</i>	Q8LNY6	380	380	1-380	100%	<i>Mus musculus</i>	2PBI_B

**Table S 4.** Summary of modeled *T. aestivum* large subunit r proteins. In bold – newly localized proteins.

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length, aa	Modeled range, aa	Percent modeled, %	Template	PDB ID
L1	L1p	<i>Triticum aestivum</i>	Q517L3	216	216	1-216	100%	<i>Thermus thermophilus</i>	2HW8_A
L2	L2p	<i>Oryza sativa</i>	Q2QNF3	261	255	1-255	98%	<i>Haloarcula marismortui</i>	1VQ8_A
L3	L3p	<i>Triticum aestivum</i>	Q7X744	389	389	1-389	100%	<i>Haloarcula marismortui</i>	1VQ8_B
L4	L4p/L4e	<i>Oryza sativa</i>	Q6ZLB8	405	372	1-269 ; 303-405	92%	<i>Haloarcula marismortui</i>	1VQ8_C
L5	L18p	<i>Oryza sativa</i>	Q8L4L4	304	304	1-304	100%	<i>Haloarcula marismortui</i>	1VQ8_N
<b>L6</b>	<b>L6e</b>	<i>Triticum aestivum</i>	<b>Q517L4</b>	<b>219</b>	<b>219</b>	<b>1-219</b>	<b>100%</b>	<i>Sulfolobus solfataricus</i>	<b>2JOY_A</b>
L7	L30p	<i>Triticum aestivum</i>	Q517K6	244	244	1-244	100%	<i>Haloarcula marismortui</i>	1VQ8_W
L7a	L7ae	<i>Oryza sativa</i>	P35685	258	201	58-258	78%	<i>Haloarcula marismortui</i>	1VQ8_F
L9	L6p	<i>Oryza sativa</i>	P49210	190	190	1-190	100%	<i>Haloarcula marismortui</i>	1VQ8_E
L10	L10e	<i>Oryza sativa</i>	Q0ITS8	224	184	1-184	82%	<i>Haloarcula marismortui</i>	3CC2_H 1VQ8_D
L11	L5p	<i>Triticum aestivum</i>	Q517L2	180	170	1-170	94%	<i>Haloarcula marismortui</i> and <i>Thermus thermophilus</i>	2J01_G
L12	L11p	<i>Oryza sativa</i>	Q0JA12	166	128	12-139	77%	<i>Haloarcula marismortui</i>	2QA4_I
<b>L13</b>	<b>L13e</b>	<i>Oryza sativa</i>	<b>Q7XJB4</b>	<b>208</b>	<b>182</b>	<b>13-194</b>	<b>88%</b>	Poly-alanine	-
<b>L14</b>	<b>L14e</b>	<i>Oryza sativa</i>	<b>Q7X752</b>	<b>134</b>	<b>134</b>	<b>1-134</b>	<b>100%</b>	<i>Sulfolobus solfataricus</i>	<b>2JOY_A</b>
L15	L15e	<i>Oryza sativa</i>	Q8H8S1	204	194	1-194	95%	<i>Haloarcula marismortui</i>	3CC2_M
L13a	L13p	<i>Triticum aestivum</i>	Q517L1	206	206	1-206	100%	<i>Haloarcula marismortui</i>	1VQ8_J
L17	L22p	<i>Oryza sativa</i>	Q6ZLA1	171	171	1-171	100%	<i>Haloarcula marismortui</i>	1VQ8_R
L18	L18e	<i>Triticum aestivum</i>	Q517L0	188	163	1-163	87%	<i>Haloarcula marismortui</i>	1VQ8_O
<b>L18a</b>	<b>L18ae</b>	<i>Oryza sativa</i>	<b>Q7XY20</b>	<b>178</b>	<b>167</b>	<b>1-167</b>	<b>94%</b>	<i>Methanobacterium thermoautotrophicum</i>	<b>2JXT_A</b>
L19	L19e	<i>Triticum aestivum</i>	Q943F3	209	189	1-189	90%	<i>Haloarcula marismortui</i>	1VQ8_P
L21	L21e	<i>Triticum aestivum</i>	Q7XYC9	164	164	1-164	100%	<i>Haloarcula marismortui</i>	1VQ8_Q
<b>L22</b>	<b>L22e</b>	<i>Oryza sativa</i>	<b>Q6VSN0</b>	<b>130</b>	<b>108</b>	<b>14-121</b>	<b>83%</b>	Artificial gene	<b>2KL8_A</b>
L23	L14p	<i>Triticum aestivum</i>	Q517K4	140	140	1-140	100%	<i>Haloarcula marismortui</i>	1VQ8_K
L24	L24e	<i>Oryza sativa</i>	Q5N754	162	75	1-75	46%	<i>Haloarcula marismortui</i>	1VQ8_U
L23a	L23p	<i>Oryza sativa</i>	Q0JBZ7	152	122	31-152	80%	<i>Haloarcula marismortui</i>	1VQ8_S
L26	L24p	<i>Oryza sativa</i>	Q2QXN5	150	130	1-130	87%	<i>Haloarcula marismortui</i>	1VQ8_T
L27a	L15p	<i>Oryza sativa</i>	Q6EUQ7	144	144	1-144	100%	<i>Haloarcula marismortui</i>	1VQ8_L
<b>L27</b>	<b>L27e</b>	<i>Oryza sativa</i>	<b>Q7XC31</b>	<b>136</b>	<b>99</b>	<b>1-99</b>	<b>73%</b>	<i>Sulfolobus solfataricus</i>	<b>2JOY_A</b>
<b>L28</b>	<b>L28e</b>	<i>Oryza sativa</i>	<b>Q5TKP3</b>	<b>147</b>	<b>73</b>	<b>58-130</b>	<b>50%</b>	de novo	-
<b>L29</b>	<b>L29e</b>	<i>Oryza sativa</i>	<b>Q9FP55</b>	<b>60</b>	<b>23</b>	<b>38-60</b>	<b>38%</b>	<i>Oryctolagus cuniculus</i>	<b>1UTG_A</b>
L30	L30e	<i>Triticum aestivum</i>	Q517K9	112	112	1-112	100%	<i>Saccharomyces cerevisiae</i>	1CN7_A
L31	L31e	<i>Triticum aestivum</i>	Q6ZGV5	123	120	1-120	98%	<i>Haloarcula marismortui</i>	1VQ8_X
L32	L32e	<i>Oryza sativa</i>	Q3MST7	133	133	1-133	100%	<i>Haloarcula marismortui</i>	1VQ8_Y
<b>L34</b>	<b>L34e</b>	<i>Triticum aestivum</i>	<b>Q517K8</b>	<b>119</b>	<b>119</b>	<b>1-119</b>	<b>100%</b>	<i>Rhodobacter capsulatus</i>	<b>2PPT_A</b>
<b>L35a</b>	<b>L35ae</b>	<i>Oryza sativa</i>	<b>Q61608</b>	<b>111</b>	<b>104</b>	<b>1-104</b>	<b>94%</b>	<i>Pyrococcus furiosus</i>	<b>1SQR_A</b>
L35	L29p	<i>Triticum aestivum</i>	Q8L805	124	124	1-124	100%	<i>Haloarcula marismortui</i>	1VQ8_V
<b>L36</b>	<b>L36e</b>	<i>Triticum aestivum</i>	<b>Q517L5</b>	<b>112</b>	<b>77</b>	<b>27-103</b>	<b>69%</b>	<i>Archaeoglobus fulgidus</i>	<b>2OEB_A</b>
L37	L37e	<i>Oryza sativa</i>	Q6Z8Y5	94	94	1-94	100%	<i>Haloarcula marismortui</i>	1VQ8_1
L38	L38e	<i>Oryza sativa</i>	Q8GVY2	69	69	1-69	100%	<i>Homo sapiens</i>	1WH9_A
L39	L39e	<i>Triticum aestivum</i>	Q517K7	51	51	1-51	100%	<i>Haloarcula marismortui</i>	1VQ8_2
<b>L40</b>	<b>L40e</b>	<i>Oryza sativa</i>	<b>P35296</b>	<b>53</b>	<b>41</b>	<b>13-53</b>	<b>77%</b>	<i>Sulfolobus solfataricus</i>	<b>2AYJ_A</b>
<b>L41</b>	<b>L41e</b>	<i>Oryza sativa</i>	<b>P62125</b>	<b>25</b>	<b>25</b>	<b>1-25</b>	<b>100%</b>	de novo	-
L42	L44e	<i>Oryza sativa</i>	Q8H5N0	105	105	1-105	100%	<i>Haloarcula marismortui</i>	1VQ8_3
L43	L37ae	<i>Oryza sativa</i>	Q5QM99	92	92	1-92	100%	<i>Haloarcula marismortui</i>	3CC2_Z
P0	L10p	<i>Oryza sativa</i>	P41095	319	262	1-262	82%	<i>Pyrococcus horikoshii</i> and <i>Methanocaldococcus janaschii</i>	3A1Y_G and 3JSY_A
P1	L12p	<i>Triticum aestivum</i>	Q517K5	110	58	6-63	53%	<i>Pyrococcus horikoshii</i>	3A1Y_E
P2	L12p	<i>Triticum aestivum</i>	Q7X729	112	59	1-59	53%	<i>Pyrococcus horikoshii</i>	3A1Y_F



**Table S 5.** Summary of modeled yeast small subunit r proteins. In bold – newly localized proteins.

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length, aa	Modeled range, aa	Percent modeled, %	Template	PDB ID
rpS0	S2p	<i>Saccharomyces cerevisiae</i>	P32905	252	252	1-252	100%	<i>Thermus thermophilus</i>	2J00_B
rpS2	S5p	<i>Saccharomyces cerevisiae</i>	P25443	254	254	1-254	100%	<i>Escherichia coli</i>	2QAL_E
rpS3	S3p	<i>Saccharomyces cerevisiae</i>	P05750	240	204	12-215	85%	<i>Escherichia coli</i>	2QAL_C
<b>rpS4</b>	<b>S4e</b>	<i>Saccharomyces cerevisiae</i>	<b>P05753</b>	<b>261</b>	<b>200</b>	<b>43-242</b>	<b>77%</b>	<i>Thermoplasma acidophilum</i>	<b>3KBG_A</b>
rpS5	S7p	<i>Saccharomyces cerevisiae</i>	P26783	225	199	27-225	88%	<i>Pyrococcus horikoshii</i>	1IQV_A
<b>rpS7</b>	<b>S7e</b>	<i>Saccharomyces cerevisiae</i>	<b>P26786</b>	<b>190</b>	<b>143</b>	<b>1-143</b>	<b>75%</b>	Poly-alanine	-
rpS9	S4p	<i>Saccharomyces cerevisiae</i>	O13516	197	197	1-197	100%	<i>Thermus thermophilus</i>	2J00_D
rpS11	S17p	<i>Saccharomyces cerevisiae</i>	P26781	156	85	39-123	54%	<i>Thermus thermophilus</i>	2J00_Q
rpS13	S15p	<i>Saccharomyces cerevisiae</i>	P05756	151	121	31-151	80%	<i>Escherichia coli</i>	2QAL_O
rpS14	S11p	<i>Saccharomyces cerevisiae</i>	P06367	137	119	19-137	87%	<i>Thermus thermophilus</i>	2J00_K
rpS15	S19p	<i>Saccharomyces cerevisiae</i>	Q01855	142	88	49-136	62%	<i>Escherichia coli</i>	2QAL_S
rpS16	S9p	<i>Saccharomyces cerevisiae</i>	P40213	143	126	18-143	88%	<i>Thermus thermophilus</i>	2J00_I
<b>rpS17</b>	<b>S17e</b>	<i>Saccharomyces cerevisiae</i>	<b>P02407</b>	<b>136</b>	<b>136</b>	<b>1-136</b>	<b>100%</b>	<i>Methanobacterium thermoautotrophicum</i>	<b>1RQ6_A</b>
rpS18	S13p	<i>Saccharomyces cerevisiae</i>	P35271	146	140	7-146	96%	<i>Escherichia coli</i>	2QAL_M
<b>rpS19</b>	<b>S19e</b>	<i>Saccharomyces cerevisiae</i>	<b>P07280</b>	<b>144</b>	<b>144</b>	<b>1-144</b>	<b>100%</b>	<i>Pyrococcus abyssi</i>	<b>2V7F_A</b>
rpS20	S10p	<i>Saccharomyces cerevisiae</i>	P38701	121	113	9-121	93%	<i>Thermus thermophilus</i>	2J00_J
<b>rpS21</b>	<b>S21e</b>	<i>Saccharomyces cerevisiae</i>	<b>P0C0V8</b>	<b>87</b>	<b>87</b>	<b>1-87</b>	<b>100%</b>	Poly-alanine	-
rpS22	S8p	<i>Saccharomyces cerevisiae</i>	P0C0W1	130	130	1-130	100%	<i>Escherichia coli</i>	2QAL_H
rpS23	S12p	<i>Saccharomyces cerevisiae</i>	P32827	145	145	1-145	100%	<i>Thermus thermophilus</i>	2J00_L
<b>rpS24</b>	<b>S24e</b>	<i>Saccharomyces cerevisiae</i>	<b>P26782</b>	<b>135</b>	<b>96</b>	<b>1-96</b>	<b>71%</b>	<i>Pyrococcus abyssi</i>	<b>2V94_A</b>
<b>rpS25</b>	<b>S25e</b>	<i>Saccharomyces cerevisiae</i>	<b>Q3E792</b>	<b>108</b>	<b>85</b>	<b>24-108</b>	<b>79%</b>	<i>Pyrococcus horikoshii</i>	<b>1UB9_A</b>
<b>rpS26</b>	<b>S26e</b>	<i>Saccharomyces cerevisiae</i>	<b>P39938</b>	<b>119</b>	<b>92</b>	<b>1-31 ; 59-119</b>	<b>77%</b>	Poly-alanine	-
<b>rpS27</b>	<b>S27e</b>	<i>Saccharomyces cerevisiae</i>	<b>P35997</b>	<b>82</b>	<b>50</b>	<b>31-80</b>	<b>61%</b>	<i>Archeoglobus fulgidus</i>	<b>1QXF_A</b>
<b>rpS28</b>	<b>S28e</b>	<i>Saccharomyces cerevisiae</i>	<b>Q3E7X9</b>	<b>67</b>	<b>60</b>	<b>1-60</b>	<b>90%</b>	<i>Pyrococcus horikoshii</i>	<b>1NY4_A</b>
rpS29	S14p	<i>Saccharomyces cerevisiae</i>	P41057	56	48	9-56	86%	<i>Thermus thermophilus</i>	2J00_N
<b>rpS30</b>	<b>S30e</b>	<i>Saccharomyces cerevisiae</i>	<b>Q12087</b>	<b>63</b>	<b>63</b>	<b>1-63</b>	<b>100%</b>	de novo	-
RACK1	RACK1	<i>Saccharomyces cerevisiae</i>	P38011	319	319	1-319	100%	<i>Mus musculus</i>	2PBI_B

**Table S 6.** Summary of modeled yeast large subunit r proteins. In bold – newly localized proteins.

Protein name	Protein family	Organism	Acc. no.	Size, aa	Modeled length, aa	Modeled range, aa	Percent modeled, %	Template	PDB ID
rpL1	L1p	<i>Saccharomyces cerevisiae</i>	P53030	217	217	1-217	100%	<i>Thermus thermophilus</i>	2HW8_A
rpL2	L2p	<i>Saccharomyces cerevisiae</i>	P05736	254	254	1-254	100%	<i>Haloarcula marismortui</i>	1VQ8_A
rpL3	L3p	<i>Saccharomyces cerevisiae</i>	P14126	387	387	1-387	100%	<i>Haloarcula marismortui</i>	1VQ8_B
rpL4	L4p/L4e	<i>Saccharomyces cerevisiae</i>	P10664	362	329	1-261 ; 295-362	91%	<i>Haloarcula marismortui</i>	1VQ8_C
rpL5	L18p	<i>Saccharomyces cerevisiae</i>	P26321	297	297	1-297	100%	<i>Haloarcula marismortui</i>	1VQ8_N
<b>rpL6</b>	<b>L6e</b>	<i>Saccharomyces cerevisiae</i>	<b>Q02326</b>	<b>176</b>	<b>176</b>	<b>1-176</b>	<b>100%</b>	<i>Sulfolobus solfataricus</i>	<b>2JOY_A</b>
rpL7	L30p	<i>Saccharomyces cerevisiae</i>	P05737	244	239	6-244	98%	<i>Haloarcula marismortui</i>	1VQ8_W
rpL8	L7ae	<i>Saccharomyces cerevisiae</i>	P17076	256	197	60-256	77%	<i>Haloarcula marismortui</i>	1VQ8_F
rpL9	L6p	<i>Saccharomyces cerevisiae</i>	P05738	191	191	1-191	100%	<i>Haloarcula marismortui</i>	1VQ8_E
rpL10	L10e	<i>Saccharomyces cerevisiae</i>	P41805	221	181	1-181	82%	<i>Haloarcula marismortui</i>	3CC2_H
rpL11	L5p	<i>Saccharomyces cerevisiae</i>	P0C0W9	174	168	1-168	97%	<i>Haloarcula marismortui</i> and <i>Thermus thermophilus</i>	1VQ8_D
rpL12	L11p	<i>Saccharomyces cerevisiae</i>	P17079	165	127	12-138	77%	<i>Haloarcula marismortui</i>	2QA4_I
<b>rpL13</b>	<b>L13e</b>	<i>Saccharomyces cerevisiae</i>	<b>Q12690</b>	<b>199</b>	<b>169</b>	<b>14-182</b>	<b>85%</b>	Poly-alanine	-
<b>rpL14</b>	<b>L14e</b>	<i>Saccharomyces cerevisiae</i>	<b>P36105</b>	<b>138</b>	<b>138</b>	<b>1-138</b>	<b>100%</b>	<i>Sulfolobus solfataricus</i>	<b>2JOY_A</b>
rpL15	L15e	<i>Saccharomyces cerevisiae</i>	P05748	204	193	1-193	95%	<i>Haloarcula marismortui</i>	3CC2_M
rpL16	L13p	<i>Saccharomyces cerevisiae</i>	P26784	199	199	1-199	100%	<i>Haloarcula marismortui</i>	1VQ8_J
rpL17	L22p	<i>Saccharomyces cerevisiae</i>	P05740	184	170	1-170	92%	<i>Haloarcula marismortui</i>	1VQ8_R
rpL18	L18e	<i>Saccharomyces cerevisiae</i>	P07279	186	161	1-161	87%	<i>Haloarcula marismortui</i>	1VQ8_O
rpL19	L19e	<i>Saccharomyces cerevisiae</i>	P05735	189	189	1-189	100%	<i>Haloarcula marismortui</i>	1VQ8_P
<b>rpL20</b>	<b>L18ae</b>	<i>Saccharomyces cerevisiae</i>	<b>P0C210</b>	<b>172</b>	<b>167</b>	<b>1-167</b>	<b>97%</b>	<i>Methanobacterium thermoautotrophicum</i>	<b>2JXT_A</b>
rpL21	L21e	<i>Saccharomyces cerevisiae</i>	Q02753	160	160	1-160	100%	<i>Haloarcula marismortui</i>	1VQ8_Q
<b>rpL22</b>	<b>L22e</b>	<i>Saccharomyces cerevisiae</i>	<b>P05749</b>	<b>121</b>	<b>105</b>	<b>6-110</b>	<b>87%</b>	Artificial gene	<b>2KL8_A</b>
rpL23	L14p	<i>Saccharomyces cerevisiae</i>	P04451	137	131	7-137	96%	<i>Haloarcula marismortui</i>	1VQ8_K
rpL24	L24e	<i>Saccharomyces cerevisiae</i>	P04449	155	73	1-73	47%	<i>Haloarcula marismortui</i>	1VQ8_U
rpL25	L23p	<i>Saccharomyces cerevisiae</i>	P04456	142	122	21-142	86%	<i>Haloarcula marismortui</i>	1VQ8_S
rpL26	L24p	<i>Saccharomyces cerevisiae</i>	P05743	127	123	1-123	97%	<i>Haloarcula marismortui</i>	1VQ8_T
<b>rpL27</b>	<b>L27e</b>	<i>Saccharomyces cerevisiae</i>	<b>P0C2H6</b>	<b>136</b>	<b>95</b>	<b>5-99</b>	<b>70%</b>	<i>Sulfolobus solfataricus</i>	<b>2JOY_A</b>
rpL28	L15p	<i>Saccharomyces cerevisiae</i>	P02406	149	149	1-149	100%	<i>Haloarcula marismortui</i>	1VQ8_L
<b>rpL29</b>	<b>L29e</b>	<i>Saccharomyces cerevisiae</i>	<b>P05747</b>	<b>59</b>	<b>22</b>	<b>38-59</b>	<b>37%</b>	<i>Oryctolagus cuniculus</i>	<b>1UTG_A</b>
rpL30	L30e	<i>Saccharomyces cerevisiae</i>	P14120	105	105	1-105	100%	<i>Saccharomyces cerevisiae</i>	1CN7_A
rpL31	L31e	<i>Saccharomyces cerevisiae</i>	P0C2H8	113	110	1-110	97%	<i>Haloarcula marismortui</i>	1VQ8_X
rpL32	L32e	<i>Saccharomyces cerevisiae</i>	P38061	130	130	1-130	100%	<i>Haloarcula marismortui</i>	1VQ8_Y
<b>rpL33</b>	<b>L35ae</b>	<i>Saccharomyces cerevisiae</i>	<b>P05744</b>	<b>107</b>	<b>100</b>	<b>1-100</b>	<b>93%</b>	<i>Pyrococcus furiosus</i>	<b>1SQR_A</b>
<b>rpL34</b>	<b>L34e</b>	<i>Saccharomyces cerevisiae</i>	<b>P87262</b>	<b>121</b>	<b>118</b>	<b>1-118</b>	<b>98%</b>	<i>Rhodobacter capsulatus</i>	<b>2PPT_A</b>
rpL35	L29p	<i>Saccharomyces cerevisiae</i>	P39741	120	118	3-120	98%	<i>Haloarcula marismortui</i>	1VQ8_V
<b>rpL36</b>	<b>L36e</b>	<i>Saccharomyces cerevisiae</i>	<b>P05745</b>	<b>100</b>	<b>77</b>	<b>24-100</b>	<b>77%</b>	<i>Archaeoglobus fulgidus</i>	<b>2OEB_A</b>
rpL37	L37e	<i>Saccharomyces cerevisiae</i>	P49166	88	88	1-88	100%	<i>Haloarcula marismortui</i>	1VQ8_1
<b>rpL38</b>	<b>L38e</b>	<i>Saccharomyces cerevisiae</i>	<b>P49167</b>	<b>78</b>	<b>78</b>	<b>1-78</b>	<b>100%</b>	<i>Homo sapiens</i>	<b>1WH9_A</b>
rpL39	L39e	<i>Saccharomyces cerevisiae</i>	P04650	51	51	1-51	100%	<i>Haloarcula marismortui</i>	1VQ8_2
<b>rpL40</b>	<b>L40e</b>	<i>Saccharomyces cerevisiae</i>	<b>P14796</b>	<b>52</b>	<b>40</b>	<b>13-52</b>	<b>77%</b>	<i>Sulfolobus solfataricus</i>	<b>2AYJ_A</b>
<b>rpL41</b>	<b>L41e</b>	<i>Saccharomyces cerevisiae</i>	<b>P05746</b>	<b>25</b>	<b>25</b>	<b>1-25</b>	<b>100%</b>	de novo	-
rpL42	L44e	<i>Saccharomyces cerevisiae</i>	P02405	106	106	1-106	100%	<i>Haloarcula marismortui</i>	1VQ8_3
rpL43	L37ae	<i>Saccharomyces cerevisiae</i>	P49631	92	92	1-92	100%	<i>Haloarcula marismortui</i>	3CC2_Z
rpP0	L10p	<i>Saccharomyces cerevisiae</i>	P05317	312	257	1-257	82%	<i>Pyrococcus horikoshii</i> and <i>Methanocaldococcus janaschii</i>	3A1Y_G
rpP1	L12p	<i>Saccharomyces cerevisiae</i>	P05318	106	58	5-62	55%	<i>Pyrococcus horikoshii</i>	3A1Y_E
rpP2	L12p	<i>Saccharomyces cerevisiae</i>	P05319	106	58	1-58	55%	<i>Pyrococcus horikoshii</i>	3A1Y_F

**Table S 7.** Calculated mass for entities in the *T. aestivum* ribosome.

Subunit	Entity	Organism	Length	Mol. weight (kDa)
60S	28S rRNA	<i>T. aestivum</i> + gaps filled from <i>O. sativa</i>	3391 nt	988.81
60S	5.8S rRNA	<i>T. aestivum</i>	164 nt	48.04
60S	5S rRNA	<i>T. aestivum</i>	119 nt	34.94
40S	18S rRNA	<i>T. aestivum</i>	1810 nt	525.75
80S	80S rRNA	<i>T. aestivum/O. sativa</i>	5648 nt	1596.79
60S	60S proteins	<i>T. aestivum/O. sativa</i>	8025 aa	905.93
40S	40S proteins	<i>T. aestivum/O. sativa</i>	5504 aa	619.33
80S	80S proteins	<i>T. aestivum/O. sativa</i>	12397 aa	1525.26
60S	60S subunit	<i>T. aestivum/O. sativa</i>	3674 nt & 8025 aa	1977.72
40S	40S subunit	<i>T. aestivum/O. sativa</i>	1810 nt & 4372 aa	1145.08
80S	80S ribosome	<i>T. aestivum/O. sativa</i>	5648 nt & 12397 aa	3122.05

**Table S 8.** Table showing % identity between used *Oryza sativa* protein sequences and *Triticum aestivum* sequences of similar length from KEGG database.

Protein	% identity	Protein	% identity
S3p	96%	L1p	90%
S4p	95%	L3p	95%
S4e	83%	L5p	91%
S5p	92%	L6e	88%
S7p	89%	L10e	91%
S7e	92%	L13p	95%
S10p	91%	L14p	99%
S11p	81%	L14e	94%
S12p	100%	L15e	94%
S13p	92%	L18e	90%
S14p	91%	L18ae	93%
S17e	85%	L19e	88%
S19p	90%	L21e	90%
S21e	85%	L22e	89%
S27e	98%	L27e	88%
S28e	100%	L30p	87%
S30e	96%	L30e	89%
RACK1	93%	L31e	89%
		L32e	96%
		L34e	94%
		L35ae	81%
		L36e	91%
		L37e	94%
		L37ae	100%
		L38e	94%
		L39e	96%
		L40e	98%

**Table S 9.** Orthologous or functionally homologous factors are aligned. Adapted from Rodnina and Wintermeyer, 2009

Translation step	Bacteria	Archaea	Eukarya
Initiation	IF1	aIF1A	eIF1A
	IF2	aIF5B	eIF5B
	IF3	aIF1	eIF1
		aIF2 $\alpha$	eIF2 $\alpha$
		aIF2 $\beta$	eIF2 $\beta$
	aIF2 $\gamma$	eIF2 $\gamma$	
	aIF2B $\alpha$	eIF2B $\alpha$	
		eIF2B $\beta$	
		eIF2B $\gamma$	
		eIF2B $\delta$	
		eIF2B $\epsilon$	
		aIF2B $\delta$	eIF3 (13 subunits)
		aIF4A	eIF4A
			eIF4B
		eIF4E	
		eIF4G	
		eIF4H	
	aIF5	eIF5	
	aIF6	eIF6	
		PABP	
Elongation	EF-TU	eEF1 $\alpha$	eEF1A
	EF-Ts	aEF1B	eEF1B (2 or 3 subunits)
	SelB	SelB	eEFSec
			SBP2
	EF-G	aEF2	eEF2
		eEF3 (fungi)	
Termination	RF1	aRF1	eRF1
	RF2		
	RF3		eRF3
Recycling	RRF		eIF3
	EFG		eIF3j
			eIF1A
			eIF1

**Table S 10.** Localization of unassigned eukaryotic 80S r proteins

Protein family	<i>T. aestivum</i> name	<i>S. cerevisiae</i> name	Proof
L6e	L6	rpL6	(i) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). (ii) Comparison of <i>S. cerevisiae</i> and <i>T. aestivum</i> cryo-EM maps (which contain L6e) with the <i>Haloarcula marismortui</i> 50S X-ray structure and the <i>Pyrococcus furiosus</i> 70S cryo-EM structure (which lack L6e). (iii) The N terminus of L6e was assigned based on differences between the density and the sequence of <i>T. aestivum</i> and <i>S. cerevisiae</i> .
L13e	L13	rpL13	(i) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). (ii) Heterogeneous distribution in Archaea (Table 7).
L14e	L14	rpL14	(i) Fold search and secondary structure prediction. (ii) Length differences between <i>S. cerevisiae</i> and <i>T. aestivum</i> L14e sequences, i.e., C terminus is longer in <i>T. aestivum</i> L14e and N terminus is longer in <i>S. cerevisiae</i> .
L18ae	L18a	rpL20	(i) The difference between Archaea and Eukaryotes, namely, that majority of this density existed only on eukaryotic ribosomes (Figure 31). (ii) Fold search revealed that the protein consists of two domains with a distinct LX motif (Figure 47).
L22e	L22	rpL22	(i) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). (ii) Comparison of <i>S. cerevisiae</i> and <i>T. aestivum</i> cryo-EM maps (which contain L22e) with the <i>H. marismortui</i> 50S X-ray structure and the <i>P. furiosus</i> 70S cryo-EM structure (which lack L22e) (Figure 35, Table 7).
L27e	L27	rpL27	(i) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). (ii) Comparison of <i>S. cerevisiae</i> and <i>T. aestivum</i> cryo-EM maps (which contain L27e) with the <i>H. marismortui</i> 50S X-ray structure and the <i>P. furiosus</i> 70S cryo-EM structure (which lack L27e) (Table 7).
L28e	L28	-	(i) Difference map between <i>S. cerevisiae</i> 80S cryo-EM map with <i>T. aestivum</i> 80S cryo-EM maps, because L28e does not exist in <i>S. cerevisiae</i> , but is present in <i>T. aestivum</i> (Lecompte et al., 2002) (see Figure 32). (ii) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28).
L29e	L29	rpL29	(i) Cryo-EM reconstruction of $\Delta$ L29e-80S ribosome at 20.5 Å and comparison with yeast 80S ribosome from wild-type strain (Figure 36).
L34e	L34	rpL34	(i) Based on the fact that it exists in Eukarya and Archaea, on the fold of the model, and the fact that this was the only major density left unassigned.
L35ae	L35a	rpL33	(i) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). (ii) Heterogeneous distribution in Archaea.
L36e	L36	rpL36	(i) Cross-linking and accessibility to proteolysis (Marion and Marion, 1987) (see Figure 28). (ii) Comparison of <i>S. cerevisiae</i> and <i>T. aestivum</i> cryo-EM maps (which contain L36e) with the <i>H. marismortui</i> 50S X-ray structure and the <i>P. furiosus</i> 70S cryo-EM structure (which lack L36e).
L38e	L38	rpL38	(i) Cryo-EM reconstruction of $\Delta$ L38e-80S ribosome at 21 Å and comparison with yeast 80S ribosome from wildtype strain (Figure 34A-C).
L40e	L40	rpL40	(i) Fold and the size of the protein.
L41e	L41	rpL41	(i) Size and density features: L41e is only 25 amino acids. (ii) Location: isolated density that is unlikely to be an RNA or an r protein extension
S4e	S4	rpS4	(i) Cross-linking (Gross et al., 1983) (see Figure 23). (ii) Structural information from <i>Thermoplasma acidophilum</i> PDB 3KBG.
S7e	S7	rpS7	(i) Immuno-EM (Bommer et al., 1991) (see Figure 24).
S17e	S17	rpS17	(i) Cross-linking (Gross et al., 1983). (ii) Localization: All remaining density on the head of the small subunit was already assigned. (iii) Structural information (Wu et al., 2008b).
S19e	S19	rpS19	(i) Location: Assembly precursors indicate S19e to be associated with the head of the small subunit (Ferreira-Cerca et al., 2007). (ii) Structural information (Gregory et al., 2007). (iii) Subsequently localized in the fungi 80S ribosome (Taylor et al., 2009).
S21e	S21	rpS21	(i) Immuno-EM (Bommer et al., 1991) (see Figure 24).
S24e	S24	rpS24	(i) Immuno-EM (Bommer et al., 1991) (see Figure 24).
S25e	S25	rpS25	(i) Cross-linking to IRES elements (Nishiyama et al., 2007).
S26e	S26	rpS26	(i) Cross-linking to mRNA (Pisarev et al., 2008).
S27e	S27	rpS27	(i) Structural information (Herve du Penhoat et al., 2004).
S28e	S28	rpS28	(i) Cross-linking to mRNA (Pisarev et al., 2008). (ii) Structural information (Aramini et al., 2003).
S30e	S30	rpS30	(i) Cross-linking to mRNA (Bulygin et al., 2005; Takahashi et al., 2002).

**Table S 11.** The summary for the number of amino acids modeled in each protein and overall in the *T. aestivum* ribosomal large subunit. The newly localized proteins (in bold), including S19e and RACK1, are calculated as a whole entity being modeled anew. In the core proteins, both the remodeled and newly modeled regions are counted.

Protein family	Protein name	Size, aa	Modeled, aa	Extended/Modified region	Modeled anew, aa
S0	S2p	305	260	1-19; 202-260	78
S2	S5p	274	263	1-118	118
S3	S3p	227	208	190-219	30
<b>S4</b>	<b>S4e</b>	<b>265</b>	<b>200</b>	<b>43-242</b>	<b>200</b>
S5	S7p	200	191	10-13; 37-62; 79-81; 122-133	55
<b>S7</b>	<b>S7e</b>	<b>192</b>	<b>143</b>	<b>1-143</b>	<b>143</b>
S9	S4p	195		1-18; 125-195	89
S11	S17p	161	85	40-42; 53-59; 120-124	15
S13	S15p	151	121	31-89; 135-138	63
S14	S11p	150	119	32-35; 68-70; 141-150	17
S15	S19p	154	91	58-74; 124-128; 138-148	43
S16	S9p	149	126	71-80	10
<b>S17</b>	<b>S17e</b>	<b>141</b>	<b>141</b>	<b>1-141</b>	<b>141</b>
S18	S13p	152	152	1-12; 73-97	37
<b>S19</b>	<b>S19e</b>	<b>146</b>	<b>146</b>	<b>1-146</b>	<b>146</b>
S20	S10p	127	128	1-29; 78-80	32
<b>S21</b>	<b>S21e</b>	<b>82</b>	<b>82</b>	<b>1-82</b>	<b>82</b>
S22	S8p	130	130	-	0
S23	S12p	142	142	1-34; 91-97; 127-142	57
<b>S24</b>	<b>S24e</b>	<b>138</b>	<b>98</b>	<b>5-102</b>	<b>98</b>
<b>S25</b>	<b>S25e</b>	<b>108</b>	<b>100</b>	<b>9-108</b>	<b>100</b>
<b>S26</b>	<b>S26e</b>	<b>133</b>	<b>92</b>	<b>1-31; 73-133</b>	<b>92</b>
<b>S27</b>	<b>S27e</b>	<b>86</b>	<b>50</b>	<b>33-82</b>	<b>50</b>
<b>S28</b>	<b>S28e</b>	<b>65</b>	<b>58</b>	<b>1-58</b>	<b>58</b>
S29	S14p	56	48	9-14	6
<b>S30</b>	<b>S30e</b>	<b>62</b>	<b>62</b>	<b>1-62</b>	<b>62</b>
<b>RACK1</b>	<b>RACK1</b>	<b>380</b>	<b>380</b>	<b>40-42; 53-59; 120-124</b>	<b>380</b>
					2202

**Table S 12.** The summary for the number of amino acids modeled in each protein and overall in the *T. aestivum* ribosomal large subunit. The newly localized proteins (in bold), including L30e, are calculated as a whole entity being modeled anew. This also applies to P proteins. In the core proteins, both the remodeled and newly modeled regions are counted.

Protein family	Protein name	Size, aa	Modeled, aa	Extended/Modified region	Modeled anew, aa
L1	L1p	216	216	-	0
L2	L2p	261	255	65-72; 246-255	18
L3	L3p	389	389	1-4; 118-145; 292-308; 367-399	82
L4	L4p/L4e	405	269	1-7; 16-23; 84-92; 303-405	127
L5	L18p	304	304	1-22; 113-142; 182-203; 233-304	146
<b>L6</b>	<b>L6e</b>	<b>219</b>	<b>219</b>	<b>1-219</b>	<b>219</b>
L7	L30p	244	244	1-85; 158-161; 172-182; 193-198	106
L8	L7ae	258	201	58-92; 115-125; 219-258	86
L9	L6p	190	190	1-5; 137-141; 185-190	16
L10	L10e	224	182	1-3; 102-115; 169-184	33
L11	L5p	180	170	1-8; 27-32; 90-94; 107-126	39
L12	L11p	166	128	-	0
<b>L13</b>	<b>L13e</b>	<b>208</b>	<b>182</b>	<b>13-194</b>	<b>182</b>
<b>L14</b>	<b>L14e</b>	<b>134</b>	<b>134</b>	<b>1-134</b>	<b>134</b>
L15	L15e	204	194	75-79; 191-194	9
L16	L13p	206	206	1-11; 153-206	65
L17	L22p	171	171	1-4; 74-76; 153-171	26
L18	L18e	188	163	1-19; 71-76; 91-93; 140-163	52
L19	L19e	209	189	144-189	46
<b>L20</b>	<b>L18ae</b>	<b>178</b>	<b>167</b>	<b>1-167</b>	<b>167</b>
L21	L21e	164	164	1-6; 79-84; 100-164	77
<b>L22</b>	<b>L22e</b>	<b>130</b>	<b>108</b>	<b>14-121</b>	<b>108</b>
L23	L14p	140	140	1-9	9
L24	L24e	162	75	1-6; 52-75	30
L25	L23p	152	122	31-69; 149-152	43
L26	L24p	150	130	1-9; 120-130	20
<b>L27</b>	<b>L27e</b>	<b>136</b>	<b>99</b>	<b>1-99</b>	<b>99</b>
L28	L15p	144	144	65-70; 127-144	24
<b>L28</b>	<b>L28e</b>	<b>147</b>	<b>73</b>	<b>58-130</b>	<b>73</b>
<b>L29</b>	<b>L29e</b>	<b>60</b>	<b>23</b>	<b>36-60</b>	<b>23</b>
<b>L30</b>	<b>L30e</b>	<b>112</b>	<b>112</b>	<b>1-112</b>	<b>112</b>
L31	L31e	123	120	1-17; 98-120	40
L32	L32e	133	133	82-89; 121-133	21
<b>L33</b>	<b>L35ae</b>	<b>111</b>	<b>104</b>	<b>1-104</b>	<b>104</b>
<b>L34</b>	<b>L34e</b>	<b>119</b>	<b>119</b>	<b>1-119</b>	<b>119</b>
L35	L29p	124	122	3-5; 38-44; 66-124	69
<b>L36</b>	<b>L36e</b>	<b>112</b>	<b>77</b>	<b>36-112</b>	<b>77</b>
L37	L37e	94	94	50-94	45
<b>L38</b>	<b>L38e</b>	<b>69</b>	<b>69</b>	<b>1-69</b>	<b>69</b>
L39	L39e	51	51	1-5; 32-37	11
<b>L40</b>	<b>L40e</b>	<b>53</b>	<b>41</b>	<b>13-53</b>	<b>41</b>
<b>L41</b>	<b>L41e</b>	<b>25</b>	<b>25</b>	<b>1-25</b>	<b>25</b>
L42	L44e	105	105	46-63; 96-105	28
L43	L37ae	92	92	1-9; 82-92	20
P0	L10p	319	262	1-262	262
P1	L12p	110	58	1-58	58
P2	L12p	112	59	1-59	59
					3336



**Table S 13.** Comparison of sizes of large subunit ribosomal proteins between different domains and organisms in those domains. (\*) points to proteins that are exceptionally large and do not follow the statistical size distribution within the protein family

Protein family	<i>T. thermophilus</i>	<i>H. marismortui</i>	<i>S. cerevisiae</i>	<i>T. aestivum</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>
L1p	229	212	217	216	217	217
L2p	276	240	254	261	256	257
L3p	206	338	387	389	416	403
L4p/L4e	210	246	362	405	401	427
L18p	112	187	297	304	299	297
L6e	-	-	176	219	243*	288*
L30p	60	154	244	244	252	248
L7ae	-	120	256	258	271	266
L6p	180	178	191	190	190	192
L10e	141 (L16p)	177	221	224	218	214
L5p	182	177	174	180	184	178
L11p	147	162	165	166	165	165
L13e	-	-	199	208	218	211
L14e	-	-	138	134	166	215*
L15e	-	194	204	204	204	204
L13p	140	145	199	206	205	203
L22p	113	155	184	171	186	184
L18e	-	116	186	188	188	188
L18ae	-	-	189	178	177	176
L19e	-	149	172	209	203	196
L21e	-	96	160	164	159	160
L22e	-	-	121	130	299*	128
L14p	122	132	137	140	140	140
L24e	-	67	155	162	155	157
L23p	96	85	142	152	277*	156
L24p	110	120	127	150	149	145
L15p	150	165	136	144	149	148
L27e	-	-	149	136	135	136
L28e	-	-	-	147	144	137
L29e	-	-	59	60	76	159
L30e	-	-	105	112	111	115
L31e	-	92	113	123	124	125
L32e	-	241	130	133	134	135
L34e	-	-	107	119	162*	117
L35ae	-	-	121	111	157*	110
L29p	72	71	120	124	123	123
L36e	-	-	100	112	115	105
L37e	-	57	88	94	93	97
L38e	-	-	78	69	70	70
L39e	-	50	51	51	51	51
L40e	-	-	52	53	52	52
L41e	-	-	25	25	25	25
L44e	-	92	106	105	104	106
L37ae	-	92	92	92	92	92
L10p	173	348	312	319	317	317
L12p	125	115	106	110	112	114
L12p	125	115	106	112	113	115

**Table S 14.** Comparison of sizes of small subunit ribosomal proteins between different domains and organisms in those domains. (\*) points to proteins that are exceptionally large and do not follow the average size limits within the protein family

Protein family	<i>T. thermophilus</i>	<i>H. marismortui</i>	<i>S. cerevisiae</i>	<i>T. aestivum/O. sativa</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>
S2p	256	267	252	305	313	295
S3ae	-	213	255	262	267	293*
S5p	162	212	254	274	246	243
S3p	239	304*	240	227	268	264
S4e	-	234	261	265	261	263
S7p	156	206	225	200	228	204
S6e	-	129	236	250	248	249
S7e	-	-	190	192	194	194
S8e	-	123	200	220	208	208
S4p	209	171	197	195	195	194
S10e	-	-	105	183	163	165
S17p	105	112	156	161	155	158
S12e	-	-	143	138	139	132
S15p	89	156	151	151	151	151
S11p	129	129	137	150	151	151
S19p	93	140	142	154	148	145
S9p	128	132	143	149	148	146
S17e	-	63	136	141	131	135
S13p	126	177	146	152	152	152
S19e	-	153	144	146	156	145
S10p	105	103	121	128	120	119
S21e	-	-	87	82	83	83
S8p	138	130	130	130	130	130
S12p	135	142	145	142	143	143
S24e	-	102	135	138	131	130
S25e	-	-	108	108	117	125
S26e	-	-	119	133	114	115
S27e	-	57	82	86	84	84
S28e	-	75	67	65	65	69
S14p	61	61	56	56	56	56
S30e	-	-	63	62	68	59
S27ae	-	44	76	79	80	80
RACK1	-	-	319	380*	318	317

**Table S 15.** A comparison of Archaea/Eukarya-specific substituting Bacteria-specific proteins on the ribosome

Protein family	Organism (Bacteria)	Residue range	Approximate residues (Bacteria)	<i>T. aestivum</i> (Eukarya)	Approximate residues ( <i>T. aestivum</i> )
L9p	<i>Escherichia coli</i>	1-149	5-9; 14-17; 25-37; 47-52	L7ae	86-91; 163-168; 222-228
L9p	<i>Escherichia coli</i>	1-149	38-40	L15e	17-19
L9p	<i>Escherichia coli</i>	1-149	44-46	L36e	99-101
L17p	<i>Thermus thermophilus</i>	2-118	7-59; 79-100; 110-118	L31e	16-85; 92-109
L17p	<i>Escherichia coli</i>	1-120	7-57; 86-99; 112-120	L31e	16-85; 92-109
L19p	<i>Thermus thermophilus</i>	1-138	1-14; 51-59	L3p	363-370; 306-322
L19p	<i>Escherichia coli</i>	1-114	1-11; 50-56	L3p	363-370; 306-322
L19p	<i>Thermus thermophilus</i>	1-137	48-104; 118-123	L24e	3-62
L19p	<i>Escherichia coli</i>	1-114	45-101	L24e	3-62
L20p	<i>Thermus thermophilus</i>	2-118	30-48	L32e	19-33; 47-52
L20p	<i>Escherichia coli</i>	1-117	29-47	L32e	19-33; 47-52
L20p	<i>Thermus thermophilus</i>	2-118	52-69; 72-75; 93-111	L35ae	9-11; 15-16; 35-41; 52-60; 63-72; 77-86; 100-104
L20p	<i>Escherichia coli</i>	1-117	51-68; 72-74; 92-110	L35ae	9-11; 15-16; 35-41; 52-60; 63-72; 77-86; 100-104
L21p	<i>Thermus thermophilus</i>	1-101	63-92	L32e	1-17; 36-55
L21p	<i>Escherichia coli</i>	1-103	65-94	L32e	1-17; 36-55
L25p	<i>Thermus thermophilus</i>	3-178	100-101; 116-120; 134-136	L10e	38-43; 179-181
L25p	<i>Thermus thermophilus</i>	3-178	13-17	L18ae	46-51
L25p	<i>Escherichia coli</i>	1-94	11-17	L18ae	45-49; 52-53
L25p	<i>Thermus thermophilus</i>	3-178	124-126; 162-163	L18p	291-293; 300-301
L27p	<i>Thermus thermophilus</i>	2-85	3-7	L10e	107-112
L27p	<i>Thermus thermophilus</i>	2-85	16-85	L21e	5-9; 31-76; 88-101
L27p	<i>Escherichia coli</i>	6-84	13-84	L21e	5-9; 31-76; 88-101
L28p	<i>Thermus thermophilus</i>	3-96	51-57; 75-76	L7ae	74-81; 153-156
L28p	<i>Thermus thermophilus</i>	3-96	3-15; 40-51; 58-74; 91-96	L15e	16-71; 89-95; 121-131
L28p	<i>Escherichia coli</i>	1-77	1-14; 25-77	L15e	16-71; 89-95; 121-131
L28p	<i>Thermus thermophilus</i>	3-96	18-37	L44e	31-52
L28p	<i>Escherichia coli</i>	1-77	16-23	L44e	46-52
L32p	<i>Thermus thermophilus</i>	2-60	15-27	L22p	55-65
L32p	<i>Escherichia coli</i>	1-56	19-30	L22p	55-65
L32p	<i>Thermus thermophilus</i>	2-60	52-57	L31e	1-3; 86-91
L32p	<i>Escherichia coli</i>	1-56	51-56	L31e	1-3; 86-91
L33p	<i>Thermus thermophilus</i>	5-54	5-54	L44e	1-29; 66-84; 88-99
L33p	<i>Escherichia coli</i>	3-52	3-52	L44e	1-29; 66-84; 88-99
L34p	<i>Thermus thermophilus</i>	1-49	1-49	L37e	3-60
L34p	<i>Escherichia coli</i>	1-46	1-46	L37e	3-60
L34p	<i>Thermus thermophilus</i>	1-49	47-48	L39e	9-10
L35p	<i>Thermus thermophilus</i>	2-65	8-13; 52-60	L15p	22-29
L35p	<i>Thermus thermophilus</i>	2-65	19-21; 48-50	L18e	157-163
L35p	<i>Thermus thermophilus</i>	2-65	33-36	L44e	83-86
L36p	<i>Thermus thermophilus</i>	1-37	1-37	L29e	38-60

L36p	<i>Escherichia coli</i>	1-38	1-38	L29e	38-60
S6p	<i>Thermus thermophilus</i>	1-101	46-58	S21e	5-11; 14-15; 40-41
S6p	<i>Escherichia coli</i>	1-100	45-56	S21e	5-11; 14-15; 40-41
S16p	<i>Thermus thermophilus</i>	1-84	49-84	S4e	43-112
S16p	<i>Escherichia coli</i>	1-82	49-82	S4e	43-112
S18p	<i>Thermus thermophilus</i>	19-88	36-82	S21e	15-63
S18p	<i>Escherichia coli</i>	19-73	24-73	S21e	15-63
S18p	<i>Thermus thermophilus</i>	19-88	85-88	S26e	101-104
S21p	<i>Escherichia coli</i>	3-53	27-35	S11p	135-148
S21p	<i>Escherichia coli</i>	3-53	3-22; 11-16	S26e	17-24; 94-113

**Table S 16.** Summary of r protein functions in the small subunit

Protein name	Protein family	Functions
Sa	S2p	(i) Forms a part of mRNA exit; (ii) The Rps0 proteins of <i>Saccharomyces cerevisiae</i> are components of the 40S ribosomal subunit required for maturation of the 3' end of 18S rRNA (Tabb-Massey et al., 2003); (iii) interaction of rpS0 with the TIF32 subunit of the translation initiation factor eIF3 may mediate eIF3 recruitment to mature 40S subunits (Valasek et al., 2003)
S2	S5p	(i) Forms mRNA entry site; (ii) Plays an important role in translational accuracy (Alksne et al., 1993)
S3	S3p	(i) Forms mRNA entry site; (ii) By maintaining stability of 40S ribosomal proteins such as rpS3, Hsp90 regulates the function of ribosomes (Kim et al., 2006)
S4	S4e	(i) Probably serving scaffolding role for h7, h15, ES3 <sup>s</sup> and ES6 <sup>s</sup> (i) In Bacteria, initiates assembly of the 16S rRNA (Held et al., 1974)
S5	S7p	(ii) SSU head component rpS5 and platform components as rpS14 are crucial constituents of a highly defined spatial arrangement in the head-platform interface of nascent SSUs, which is required for efficient processing of the SSU rRNA 3' end (Neueder et al., 2010)
S7	S7e	(i) In HeLa cells, the depletion of RPS7 and RPS28 leads to the accumulation of the early 30S rRNA precursor; (ii) In yeast, RPS7 and RPS28 affect the cytoplasmic maturation of the 18S rRNA and terminal assembly of the 40S subunit (Robledo et al., 2008).
S9	S4p	(i) Forms a part of the factor-binding site; (ii) In Bacteria, together with S7, it initiates the assembly of the 30S subunit and is absolutely essential to the assembly process (Held et al., 1974); (iii) In Bacteria, autogeneously regulates the expression of other ribosomal proteins by binding to polycistronic mRNA (Davies et al., 1998; Nomura et al., 1980); (iv) Plays an important role in translational accuracy (Alksne et al., 1993)
S11	S17p	(i) In Bacteria, one of the primary rRNA binding proteins, it binds specifically to the 5'-end of 16S ribosomal RNA
S13	S15p	(i) In Bacteria, binding of S15 to the central domain results in a conformational change in the RNA to form the platform of the 30S subunit (Jagannathan and Culver, 2003)
S14	S11p	(i) SSU head component rpS5 and platform components as rpS14 are crucial constituents of a highly defined spatial arrangement in the head-platform interface of nascent SSUs, which is required for efficient processing of the SSU rRNA 3' end (Neueder et al., 2010)
S15	S19p	(i) In humans RPS15 is essential for the final cytoplasmic maturation and assembly of the 40S subunit (Robledo et al., 2008) (ii) In yeast, depletion of RPS15 leads to the accumulation of the 20S in the nucleus, suggesting that RPS15 in yeast is essential for export from the nucleus (Leger-Silvestre et al., 2004)
S16	S9p	(i) RPS16 depletion impairs the processing of the 5' ETS, leading to the accumulation of an early precursor and decrease of the 20S (yeast)
S17	S17e	(i) Mutation in the gene is known to cause Diamond-Blackfan anemia by assembly disorders (Cmejla et al., 2007)
S18	S13p	(i) Located at the top of the head of the 30S subunit, contacts several helices of the 16S rRNA. (ii) In the 70S ribosome it contacts the 23S rRNA (bridge B1a) and protein L5 of the 50S subunit (bridge B1b), connecting the 2 subunits; these bridges are implicated in subunit movement
S19	S19e	(i) RPS19 depletion leads to the accumulation of the 20S/21S rRNA species, which is not processed further. (Robledo et al., 2008). (ii) Mutation in the gene is known to cause Diamond-Blackfan anemia by assembly disorders (Campagnoli et al., 2008)
S20	S10p	(i) Involved in binding of tRNA to the ribosomes
S21	S21e	(i) Involved in the maturation of the 3' end of 18S rRNA (Tabb-Massey et al., 2003)
S15a	S8p	(i) In Bacteria, S8p binding is not absolutely required for assembly of the platform, it appears to affect significantly the 16S rRNA environment of S15p by influencing central domain organization (Jagannathan and Culver, 2003)
S23	S12p	(i) Plays an important role in translational accuracy (Alksne et al., 1993)
S24	S24e	(i) RPS24 depletion impairs the processing of the 5' ETS, leading to the accumulation of an early precursor and decrease of the 20S (Robledo et al., 2008); (ii) Mutation in the gene is known to cause Diamond-Blackfan anemia by assembly disorders (Choesmel et al., 2008)
S25	S25e	(i) Depletion of RPS25 allows the assembly of the 40S subunit but impairs its association with the 60S subunit to form the mature 80S ribosome or that the 80S ribosome is unstable and is degraded (Robledo et al., 2008); (ii) It interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site (Nishiyama et al., 2007)
S26	S26e	(i) Forms a part of mRNA exit (Pisarev et al., 2008)
S27	S27e	(i) Ribosomal protein S27 affects ribosome assembly, as deletion of either one of the two genes encoding this essential protein leads to defects in pre-rRNA processing (Baudin-Baillieu et al., 1997; Dresios et al., 2006)

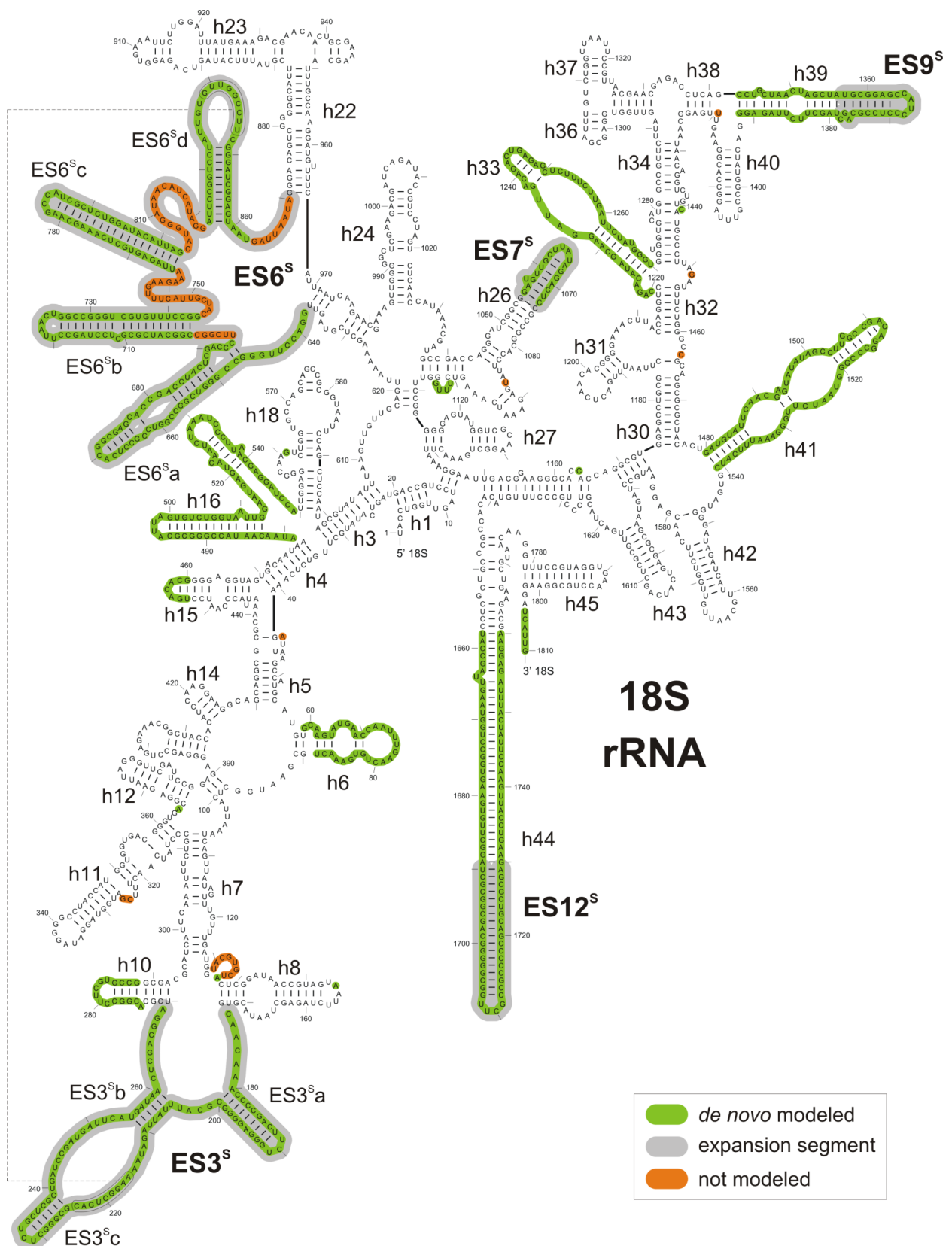
		(i) Forms a part of mRNA exit (Pisarev et al., 2008);
S28	S28e	(ii) In HeLa cells, the depletion of RPS7 and RPS28 leads to the accumulation of the early 30S rRNA precursor;
		(iii) In yeast, RPS7 and RPS28 affect the cytoplasmic maturation of the 18S rRNA and terminal assembly of the 40S subunit (Robledo et al., 2008).
S29	S14p	(i) In Bacteria: Binds 16S rRNA, required for the assembly of 30S particles
S30	S30e	(i) Cross-links to mRNA (Bulygin et al., 2005; Takahashi et al., 2002)
RACK1	RACK1	(i) Signaling scaffold, possible regulatory link between translation and signaling (Rabl et al., 2011)

**Table S 17.** Summary of r protein functions in the large subunit

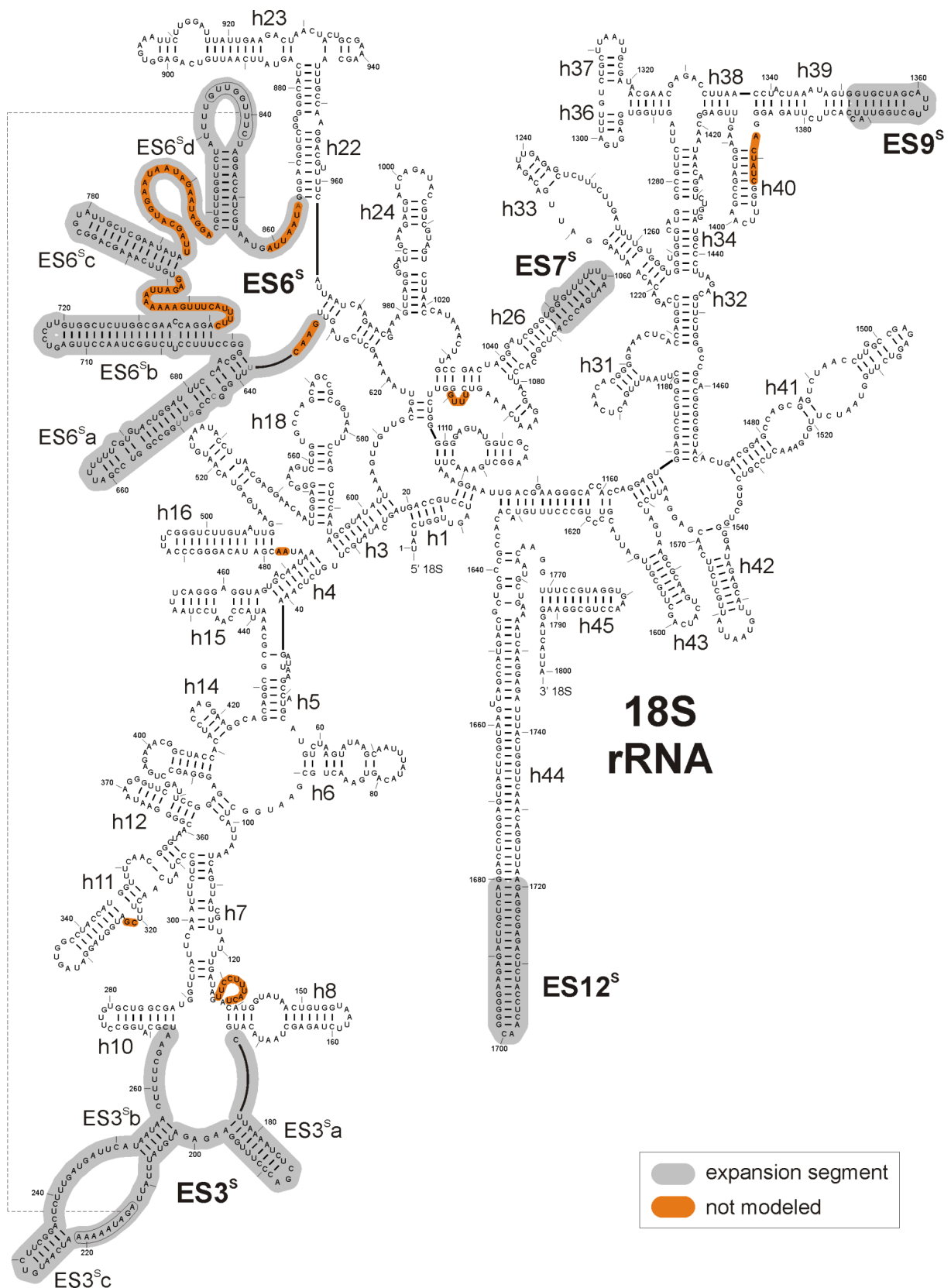
Protein name	Protein family	Functions
L1	L1p	(i) Probably involved in the removal of deacylated tRNA from the E site (Wilson and Nierhaus, 2005)
L2	L2p	(i) L2p contacts almost every domain of the large subunit rRNA; (ii) Involved in the association of the ribosomal subunits (Meskauskas et al., 2008), tRNA binding to A and P sites and peptidyl transfer (Diedrich et al., 2000)
L3	L3p	(i) Constitutes a part of peptidyl transferase center; (ii) Central extension of L3p may function as an allosteric switch in coordinating binding of the elongation factors (Meskauskas and Dinman, 2007)
L4	L4p	(i) The extension of the ribosomal proteins L4p contributes to formation of the tunnel wall and forms a so-called 'constriction' where the tunnel narrows (Ban et al., 2000; Wilson and Nierhaus, 2005) (i) Shown to interact with eEF-3 in yeast (Andersen et al., 2006);
L5	L18p	(ii) Mediates the principal interactions 5S RNA makes with the rest of the subunit (Klein et al., 2004); (iii) Helps to anchor the peptidyl-tRNA to the P-site of the ribosome (Meskauskas and Dinman, 2001)
L6	L6e	(i) Might stabilize the tertiary interaction between ES7 <sup>L</sup> and ES39 <sup>L</sup>
L7	L30p	(i) Mediates the principal interactions 5S RNA makes with the rest of the subunit (Klein et al., 2004)
L7a	L7ae	(i) Multifunctional RNA-binding - L7Ae has a role in RNPs acting in tRNA processing (RNase P), RNA modification (H/ACA, C/D snoRNPs), and translation (ribosomes) (Cho et al., 2010)
L9	L6p	(i) Forms a part of the factor-binding site at the edge of the intersubunit cleft of the ribosome (Brodersen and Nissen, 2005)
L10	L10e	(i) May be involved in correct positioning of the acceptor stem of A- and P-site tRNAs as well as RRF on the ribosome (Wilson and Nierhaus, 2005); (ii) Mediates the principal interactions 5 S RNA makes with the rest of the subunit (Klein et al., 2004); (iii) Important player in the joining of 60S to 40S subunits (iv) Interacts with the nuclear export factor Nmd3p and is necessary for the nuclear export 60S subunits (Karl et al., 1999; Dresios et al., 2006)
L11	L5p	(i) Shown to interact with eEF3 in yeast (Andersen et al., 2006); (ii) mediates the principal interactions 5 S RNA makes with the rest of the subunit (Klein et al., 2004)
L12	L11p	(i) In Bacteria, mutations in L11 or lack of the complete protein confer resistance against thiostrepton, an antibiotic that blocks the ribosomal transition from the pre- to post-translocational state and vice versa (Wilson and Nierhaus, 2005); (ii) Senses the presence of a deacylated tRNA in the A site (Wilson and Nierhaus, 2005); (iii) Mutations or the absence of the protein can cause a relaxed phenotype (relC) resulting from loss of stringent control (Wilson and Nierhaus, 2005); (iv) Forms a part of the factor-binding site at the edge of the intersubunit cleft of the ribosome (Brodersen and Nissen, 2005)
L13	L13e	(i) Extraribosomal: r protein L13 mRNA is up-regulated in response to DNA damage in hamster cells (Kobayashi et al., 2006); (ii) Extraribosomal: L13 expression is up-regulated in human gastrointestinal cancers (Kobayashi et al., 2006)
L14	L14e	(i) Might stabilize the tertiary interaction between ES7L and ES39L
L15	L15e	(i) Might stabilize the tertiary interaction between 28S and 5.8S
L13a	L13p	(i) Release of L13a from the 60S ribosomal subunit acts as a mechanism of transcript-specific translational control (Mazumder et al., 2003); (ii) L13a blocks 48S assembly (Kapasi et al., 2007)
L17	L22p	(i) In Bacteria, deletion of three amino acids in L22 confers erythromycin resistance without interfering with the binding of the drug (Wilson and Nierhaus, 2005); (ii) May interact with specific nascent chains to regulate translation; (iii) The extensions of the ribosomal protein L22p contributes to formation of the tunnel wall and forms a so-called 'constriction' where the tunnel narrows (Ban et al., 2000; Wilson and Nierhaus, 2005)
L18	L18e	(i) Stabilizes the tertiary rRNA structure within the 23S/28S rRNA domain (domain II) to which it binds
L18a	L18ae	(i) Might stabilize the tertiary interaction between 5S and 28S
L19	L19e	(i) Shown to be in the direct vicinity of Sec61 translocon (Becker et al., 2009) (ii) Forms a bridge with the small subunit
L21	L21e	(i) Mediates the principal interactions 5S RNA makes with the rest of the subunit (Klein et al., 2004)
L22	L22e	(i) Extraribosomal: EBER 1, a small noncoding viral RNA abundantly expressed in all cells transformed by Epstein-Barr virus (EBV), has been shown to associate with the human ribosomal protein L22 (Fok et al., 2006)
L23	L14p	(i) Forms a part of the factor-binding site at the edge of the intersubunit cleft of the ribosome (Brodersen and Nissen, 2005)
L24	L24e	(i) Deletion of L24e causes 60S subunits to bind inefficiently to the translation initiation machinery, causing accumulation of 43S translation pre-initiation complexes and a significant decrease in the amount of 80S ribosomes (Baronas-Lowell and Warner, 1990; Dresios et al., 2006); (ii) Forms a bridge with small subunit
L23a	L23p	(i) Shown to be a component of the chaperone trigger factor binding site on the ribosome (Wilson and Nierhaus, 2005); (ii) Shown to interact with SRP (Halic et al., 2006)

L26	L24p	(i) Assembly initiator protein; (ii) Present at the tunnel exit site and has been shown to interact with the translocon (Becker et al., 2009)
L27a	L15p	(i) Important in the early steps of LSU assembly
L27	L27e	(i) Might be involved in stabilization of H58
L28	L28e	(i) Involved in stabilization of RNA tertiary structure of ES7L a
L29	L29e	(i) Involved in stabilization of RNA tertiary structure; (ii) Absence of L29e impairs proper assembly of proteins onto the 60S subunit at the interface where the 60S subunit joins the 40S subunit, a defect that impedes subunit joining and reduces protein synthesis (DeLabre et al., 2002; Dresios et al., 2006)
L30	L30e	(i) L30 binds to eukaryotic selenocysteine insertion sequence in the 3' untranslated region of eukaryotic selenoprotein mRNA and stimulates UGA recoding activity in cells (Chavatte et al., 2005); (ii) Yeast ribosomal protein L30 binds to the transcript of its own gene to inhibit splicing to mature mRNA and to reduce translation (Dresios et al., 2006; Vilardell et al., 2000)
L31	L31e	(i) Shown to interact with SRP (Halic et al., 2006)
L32	L32e	(i) Involved in processing of pre-rRNA in the nucleolus, regulates splicing of the RPL32 transcript in the nucleus, and the translation of the spliced RPL32 mRNA in the cytoplasm (Vilardell and Warner, 1997)
L34	L34e	(i) Might be involved in ES27 <sup>L</sup> coordination and H58 stabilization
L35a	L35ae	(i) Might stabilize the tertiary interaction between ES7L and ES39L
L35	L29p	(i) Constitutes part of the binding site for the signal recognition particle and translocon (Becker et al., 2009; Halic et al., 2006)
L36	L36e	(i) Might be a protein stabilizing or coordinating the L1 stalk - maybe functionally similar to bacterial-specific L9
L37	L37e	(i) Zinc-finger
L38	L38e	(i) Might be involved in ES27 <sup>L</sup> coordination
L39	L39e	(i) Participates in the assembly of the 60S (Sachs and Davis, 1990) (ii) Single mutation in the gene for L39 (spb-2) suppresses the inhibition of translation initiation (Sachs and Davis, 1990); (iii) The lack of L39 leads to a fourfold increase in the error frequency (Dresios et al., 2000; Dresios et al., 2006); (iv) Absence L39 increases the rate of peptide bond formation (Dresios et al., 2001)
L40	L40e	(i) Might stabilize the tertiary interactions in 28S
L41	L41e	(i) May prevent spontaneous translocation; (ii) Absence of L41e lowers peptidyl-transferase activity 3-fold and increases resistance to cycloheximide (Dresios et al., 2003); (iii) Deletion of extra ribosomal protein L41 causes slight hyperaccuracy (Dresios et al., 2003); (iv) Deletion of both L41 genes results in increased paromomycin resistance (Dresios et al., 2006)
L42	L44e	(i) Interacts with the E-site tRNA (Deamer and Szostak, 2010)
L43	L37ae	(i) Zinc finger (Ban et al., 2000)
P0	L10p	(i) Forms a part of the factor-binding site at the edge of the intersubunit cleft of the ribosome (Brodersen and Nissen, 2005)
P1/P2	L12p	(i) Forms a part of the factor-binding site at the edge of the intersubunit cleft of the ribosome (Brodersen and Nissen, 2005); (ii) Involved in elongation-factor binding and GTPase activation

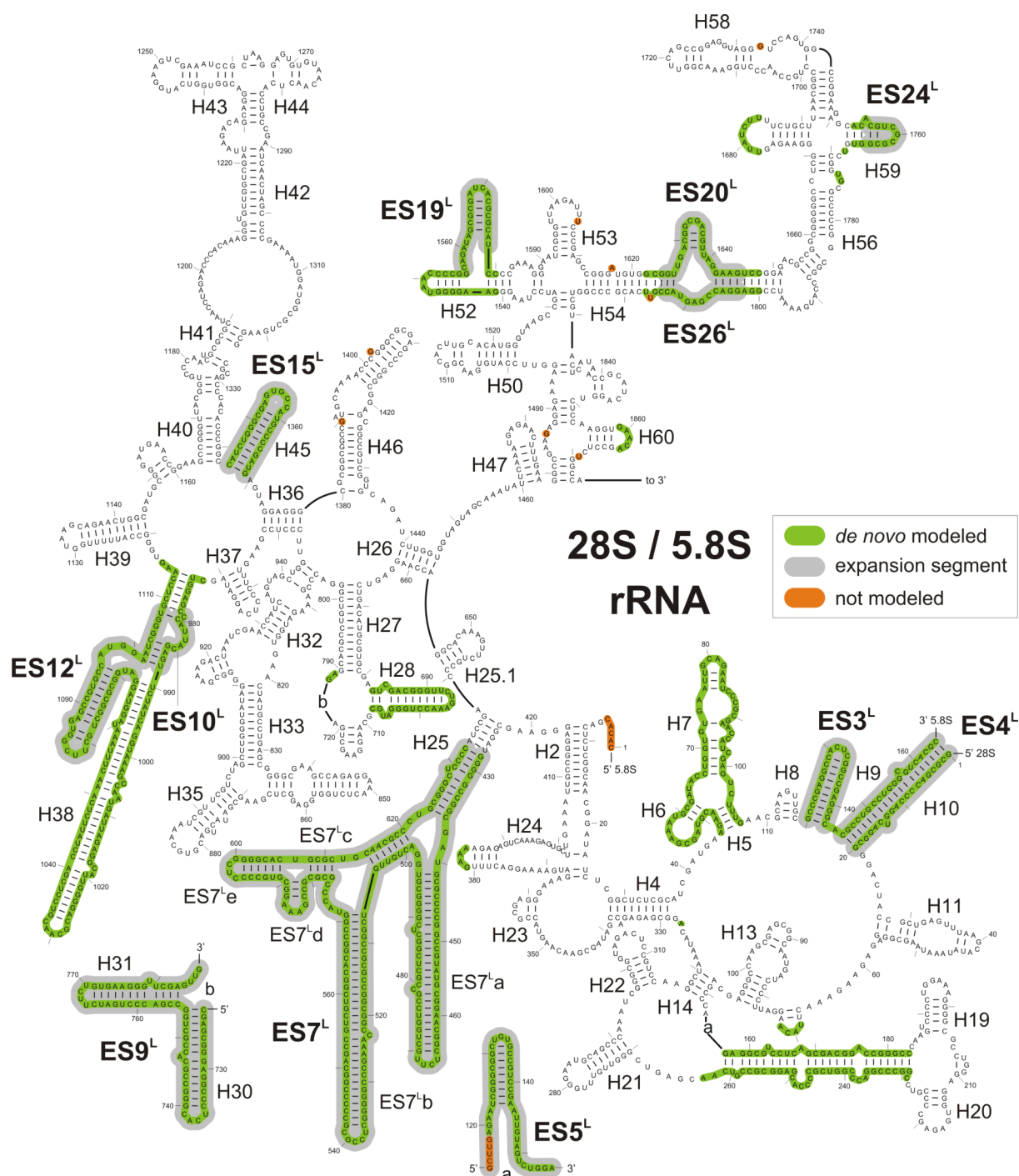




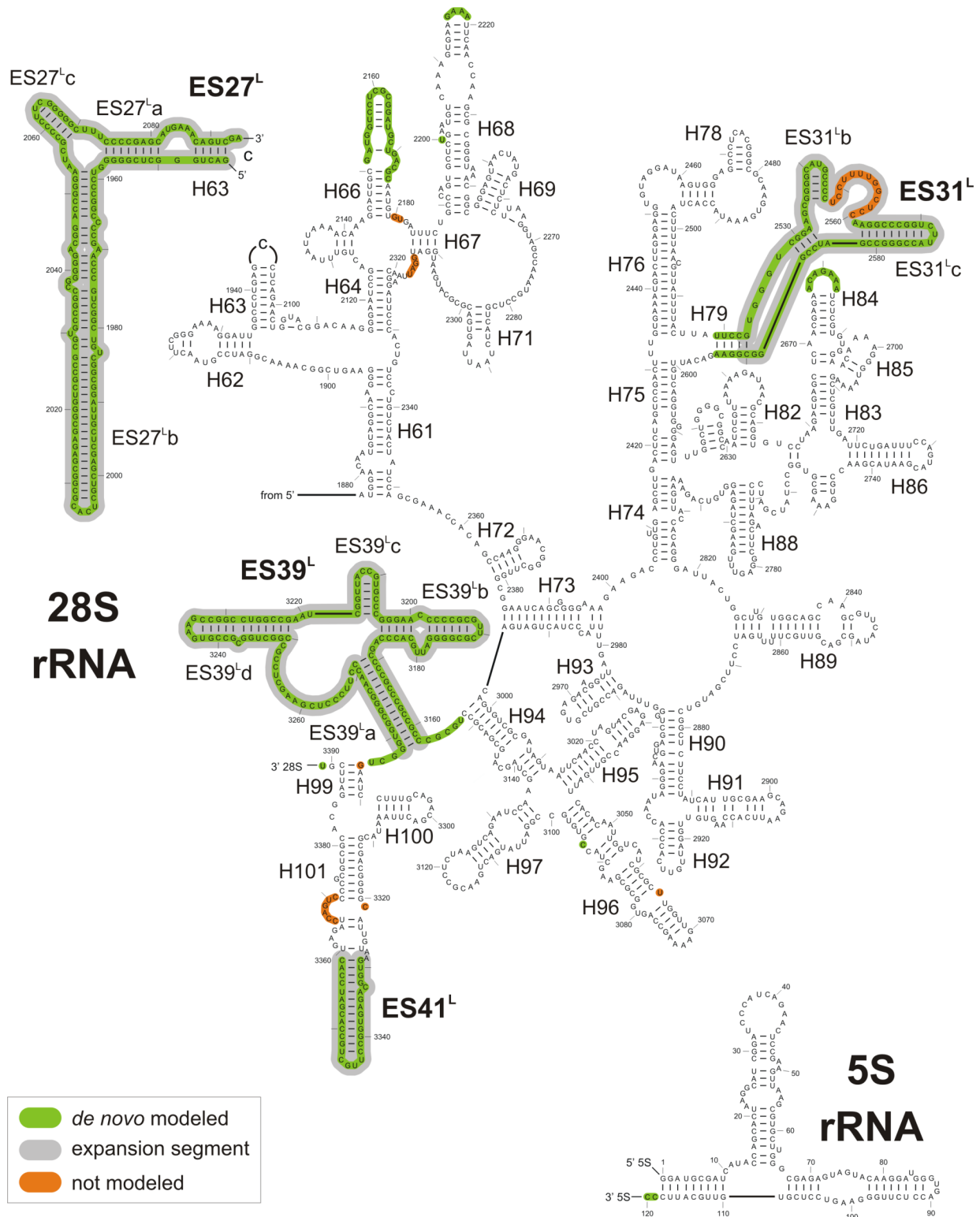
**Figure S 1.** Secondary structure diagram for the small subunit (18S) rRNA of *T. aestivum*, modified from Cannone et al., 2002. Green regions indicated *de novo* modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled



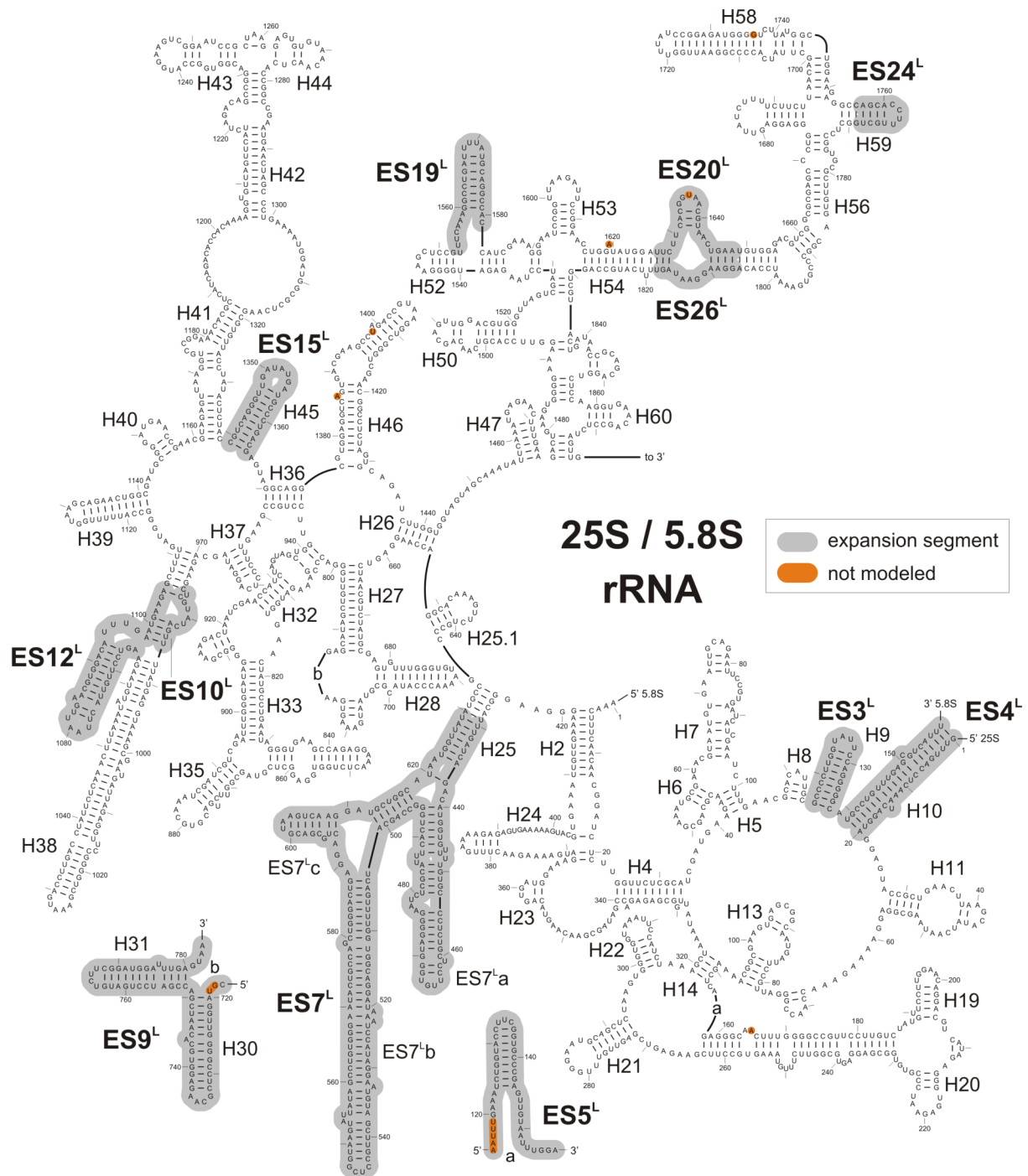
**Figure S 2.** Secondary structure diagram for the small subunit (18S) rRNA of *S. cerevisiae* modified from Cannone et al., 2002. Gray regions indicate expansion segments, whereas orange nucleotides were not modeled.



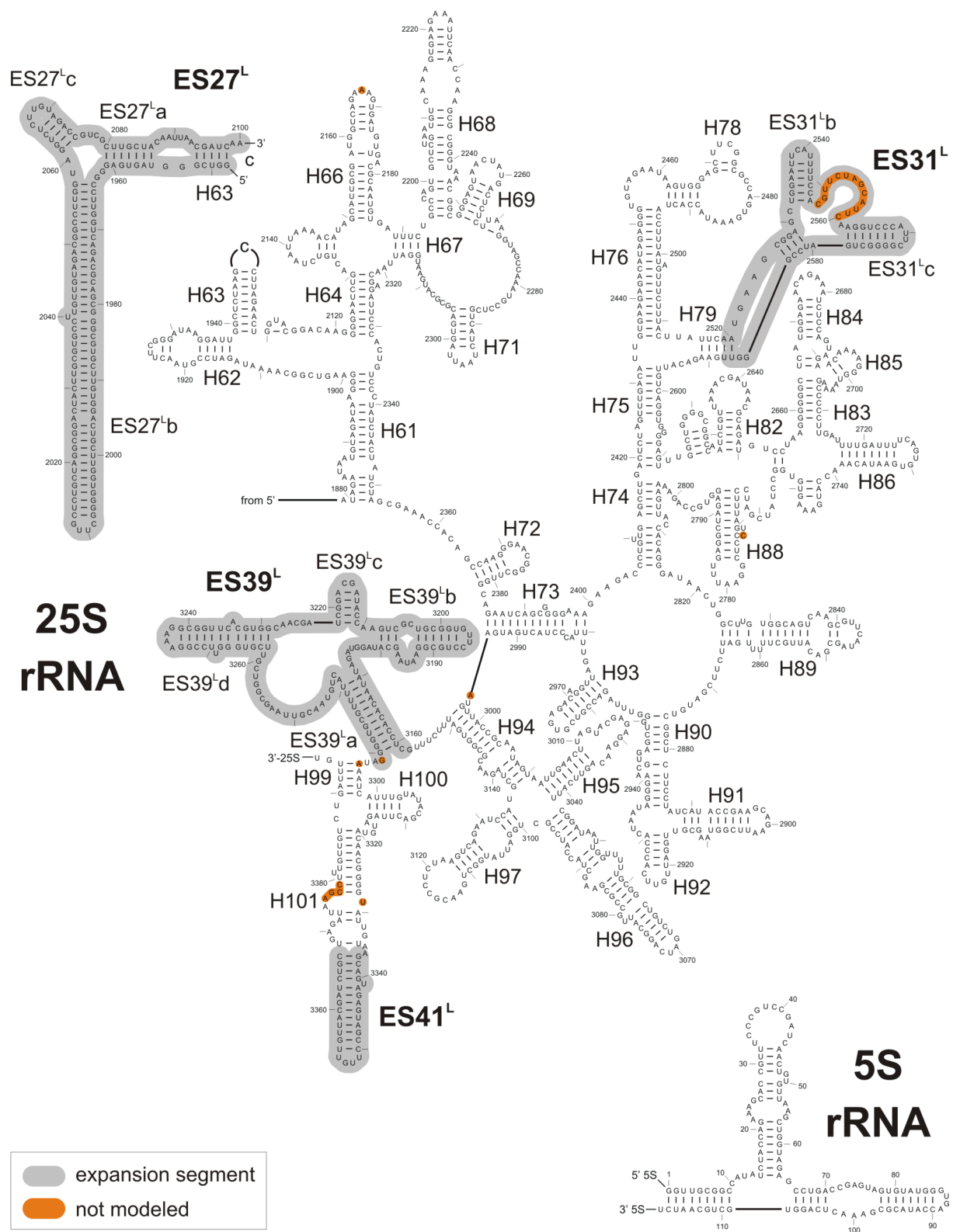
**Figure S 3.** Secondary structure diagram for the 5' region of the large subunit rRNAs (5.8S and 28S) of *T. aestivum* modified from Cannone et al., 2002. Green regions indicated *de novo* modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled.



**Figure S 4.** Secondary structure diagram for the 3' region of the large subunit rRNAs (5S and 28S) of *T. aestivum* modified from Cannone et al., 2002. Green regions indicated *de novo* modeled regions, gray regions are expansion segments, whereas orange nucleotides were not modeled.



**Figure S 5.** Secondary structure diagram for the 5' region of the large subunit rRNAs (5.8S and 25S) of *S. cerevisiae* modified from Cannone et al., 2002. Gray regions are expansion segments, whereas orange nucleotides were not modeled.



**Figure S 6.** Secondary structure diagram for the 3' region of the large subunit rRNAs (5S and 25S) of *S. cerevisiae* modified from Cannone et al., 2002. Gray regions are expansion segments, whereas orange nucleotides were not modeled.

## 7 Curriculum vitae

### Personal Details

---

Name	Jean-Paul Armache
Date of birth	December 2, 1982
Place of birth	Łódź, Poland
Nationality	Polish

### Education

---

2007 – 2011	Graduate studies in Biochemistry in the group of Prof. Roland Beckmann. Gene Center, Ludwig-Maximilians-Universität Munich, Germany Department of Biochemistry
2004 – 2005	Socrates/Erasmus exchange student in Institut National des Sciences Appliquées, Toulouse, France Department of Computer Science and Electrical Engineering Speciality: Computer Science.
2001 – 2006	Technical University of Łódź, Poland Department of Technical Physics, Computer Science and Applied Mathematics Speciality: Computer Science - Artificial Intelligence and Programming
1997 – 2001	XIII Liceum Ogólnokształcące im. Marii Piotrowiczowej, Łódź, Poland High school, Science baccalaureate, speciality in physics

### Conferences and Workshops

---

2010	Ribosomes 2010, Orvieto, Italy. Poster Presentation
2008	Computational Biology Conference 2008, Basel, Switzerland. Participant
2007	EMBO Course on Image Processing for Cryo EM, London, UK. Participant
2007	Structural Biology of Disease Mechanisms, Murnau, Germany. Participant

### Publications

---

2011	Becker, T., <u>Armache, J.-P.</u> , Jarasch, A., Anger, A.M., Villa, E., Sieber, H., Abdel Motaal, B., Mielke, T., Berninghausen, O., and Beckmann, R. (2011). <b><i>Structure of the No-Go mRNA decay complex Dom34/Hbs1 bound to a stalled 80S ribosome.</i></b> Nat Struct Mol Biol Accepted.
2010	Marquez, V., Frohlich, T., <u>Armache, J.-P.</u> , Sohmen, D., Donhofer, A., Mikolajka, A., Berninghausen, O., Thomm, M., Beckmann, R., Arnold, G.J., and Wilson, D.N. (2010). <b><i>Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins.</i></b> J. Mol. Biol. 405, 1215-1232.

Armache, J.-P.\*, Jarasch, A.\*, Anger, A.M.\*, Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., *et al.* (2010).

***Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution.***

Proc Natl Acad Sci U S A 107, 19748-19753.

\*(These authors contributed equally to this work)

Armache, J.-P.\*, Jarasch, A.\*, Anger, A.M.\*, Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., *et al.* (2010).

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

Proc Natl Acad Sci U S A 107, 19754-19759.

\*(These authors contributed equally to this work)

Gartmann, M., Blau, M., Armache, J.-P., Mielke, T., Topf, M., and Beckmann, R. (2010).

***Mechanism of eIF6-mediated inhibition of ribosomal subunit joining.***

J Biol Chem 285, 14848-14851.

Bhushan, S., Gartmann, M., Halic, M., Armache, J.-P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D.N., and Beckmann, R. (2010).

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

Nat Struct Mol Biol 17, 313-317.

2009

Becker, T., Bhushan, S., Jarasch, A., Armache, J.-P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., *et al.* (2009).

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

Science 326, 1369-1373.

Seidelt, B.\*, Innis, C.A.\*, Wilson, D.N., Gartmann, M., Armache, J.-P., Villa, E., Trabuco, L.G., Becker, T., Mielke, T., Schulten, K., *et al.* (2009).

***Structural insight into nascent polypeptide chain-mediated translational stalling.***

Science 326, 1412-1415.

\*(These authors contributed equally to this work)



## 8 References

- Agrawal, R.K., Linde, J., Sengupta, J., Nierhaus, K.H., and Frank, J. (2001). Localization of L11 protein on the ribosome and elucidation of its involvement in EF-G-dependent translocation. *Journal of Molecular Biology* *311*, 777-787.
- Alksne, L.E., Anthony, R.A., Liebman, S.W., and Warner, J.R. (1993). An accuracy center in the ribosome conserved over 2 billion years. *Proc Natl Acad Sci U S A* *90*, 9538-9541.
- Andersen, B.F., Becker, T., Blau, M., Anand, M., Halic, M., Balar, B., Mielke, T., Boesen, T., Pedersen, J.S., Spahn, C.M.T., *et al.* (2006). Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature* *433*, 663-668.
- Aramini, J.M., Huang, Y.J., Cort, J.R., Goldsmith-Fischman, S., Xiao, R., Shih, L.Y., Ho, C.K., Liu, J., Rost, B., Honig, B., *et al.* (2003). Solution NMR structure of the 30S ribosomal protein S28E from *Pyrococcus horikoshii*. *Protein Sci* *12*, 2823-2830.
- Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., *et al.* (2010a). Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc Natl Acad Sci U S A* *107*, 19748-19753.
- Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., *et al.* (2010b). Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *Proc Natl Acad Sci U S A* *107*, 19754-19759.
- Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B., and Steitz, T.A. (1999). Placement of protein and RNA structures into a 5 Å -resolution map of the 50S ribosomal subunit. *Nature* *400*, 841-847.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* *289*, 905-920.
- Bargis-Surgey, P., Lavergne, J.P., Gonzalo, P., Vard, C., Filhol-Cochet, O., and Reboud, J.P. (1999). Interaction of elongation factor eEF-2 with ribosomal P proteins. *Eur J Biochem* *262*, 606-611.
- Baronas-Lowell, D.M., and Warner, J.R. (1990). Ribosomal protein L30 is dispensable in the yeast *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* *10*, 5235-5243.
- Baudin-Baillieu, A., Tollervey, D., Cullin, C., and Lacroute, F. (1997). Functional analysis of Rrp7p, an essential yeast protein involved in pre-rRNA processing and ribosome assembly. *Molecular and Cellular Biology* *17*, 5023-5032.
- Becker, T., Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Sieber, H., Abdel Motaal, B., Mielke, T., Berninghausen, O., and Beckmann, R. (2011). Structure of the No-Go mRNA decay complex Dom34/Hbs1 bound to a stalled 80S ribosome. *Nat Struct Mol Biol* *Accepted*.
- Becker, T., Bhushan, S., Jarasch, A., Armache, J.-P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., *et al.* (2009). Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* *326*, 1369-1373.
- Beckmann, R., Spahn, C.M., Eswar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J., and Blobel, G. (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* *107*, 361-372.

- Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Crystal structure of the eukaryotic ribosome. *Science* *330*, 1203-1209.
- Bhushan, S., Gartmann, M., Halic, M., Armache, J.-P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D.N., and Beckmann, R. (2010a). alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* *17*, 313-317.
- Bhushan, S., Hoffman, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D.N., and Beckmann, R. (2010b). SecM-stalled ribosomes adopt an altered geometry at the peptidyltransferase center. *PLoS Biol*, in press.
- Bhushan, S., Meyer, H., Starosta, A., Becker, T., Mielke, T., Berninghausen, O., Sattler, M., Wilson, D., and Beckmann, R. (2010c). Structural basis for translational stalling by human cytomegalovirus (hCMV) and fungal arginine attenuator peptide (AAP). *Mol Cell* *40*, 138-146.
- Biegert, A., Mayer, C., Remmert, M., Soding, J., and Lupas, A.N. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* *34*, W335-339.
- Bokov, K., and Steinberg, S.V. (2009). A hierarchical model for evolution of 23S ribosomal RNA. *Nature* *457*, 977-980.
- Bommer, U.A., Lutsch, G., Stahl, J., and Bielka, H. (1991). Eukaryotic initiation factors eIF-2 and eIF-3: interactions, structure and localization in ribosomal initiation complexes. *Biochimie* *73*, 1007-1019.
- Brodersen, D., and Nissen, P. (2005). The social life of ribosomal proteins. *FEBS J* *272*, 2098-2108.
- Bulygin, K., Chavatte, L., Frolova, L., Karpova, G., and Favre, A. (2005). 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*, 2153-2162.
- Campagnoli, M.F., Ramenghi, U., Armiraglio, M., Quarello, P., Garelli, E., Carando, A., Avondo, F., Pavesi, E., Fribourg, S., Gleizes, P.E., *et al.* (2008). RPS19 mutations in patients with Diamond-Blackfan anemia. *Hum Mutat* *29*, 911-920.
- Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., *et al.* (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics* *3*, 2.
- Cech, T. (2000). Structural biology. The ribosome is a ribozyme. *Science* *289*, 878-879.
- Cech, T.R., Zaugg, A.J., and Grabowski, P.J. (1981). In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* *27*, 487-496.
- Chandramouli, P., Topf, M., Menetret, J.F., Eswar, N., Cannone, J.J., Gutell, R.R., Sali, A., and Akey, C.W. (2008). Structure of the mammalian 80S ribosome at 8.7 Å resolution. *Structure* *16*, 535-548.
- Chavatte, L., Brown, B.A., and Driscoll, D.M. (2005). Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nat Struct Mol Biol* *12*, 408-416.
- Chen, J.Z., and Grigorieff, N. (2007). SIGNATURE: a single-particle selection system for molecular electron microscopy. *J Struct Biol* *157*, 168-173.

- Cho, I.M., Lai, L.B., Susanti, D., Mukhopadhyay, B., and Gopalan, V. (2010). Ribosomal protein L7Ae is a subunit of archaeal RNase P. *Proc Natl Acad Sci U S A* *107*, 14573-14578.
- Choesmel, V., Fribourg, S., Aguisa-Toure, A.H., Pinaud, N., Legrand, P., Gazda, H.T., and Gleizes, P.E. (2008). Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. *Hum Mol Genet* *17*, 1253-1263.
- Choli, T., Franceschi, F., Yonath, A., and Wittmann-Liebold, B. (1993). Isolation and characterization of a new ribosomal protein from the thermophilic eubacteria, *Thermus thermophilus*, *T. aquaticus* and *T. flavus*. *Biol Chem Hoppe Seyler* *374*, 377-383.
- Clemons, W.M.J., May, J.L., Wimberly, B.T., McCutcheon, J.P., Capel, M.S., and Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* *400*, 833-840.
- Cmejla, R., Cmejlova, J., Handrkova, H., Petrak, J., and Pospisilova, D. (2007). Ribosomal protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. *Hum Mutat* *28*, 1178-1182.
- Coyle, S.M., Gilbert, W.V., and Doudna, J.A. (2009). Direct link between RACK1 function and localization at the ribosome in vivo. *Molecular and Cellular Biology* *29*, 1626-1634.
- Cuff, J.A., and Barton, G.J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* *40*, 502-511.
- Cukras, A.R., Southworth, D.R., Brunelle, J.L., Culver, G.M., and Green, R. (2003). Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA:tRNA complex. *Mol Cell* *12*, 321-328.
- Dabbs, E.R. (1978). Mutational Alterations in 50 Proteins of the *Escherichia coli* Ribosomes. *Mol Gen Genet* *165*, 73-78.
- Davies, C., Gerstner, R.B., Draper, D.E., Ramakrishnan, V., and White, S.W. (1998). The crystal structure of ribosomal protein S4 reveals a two-domain molecule with an extensive RNA-binding surface: one domain shows structural homology to the ETS DNA-binding motif. *EMBO J* *17*, 4545-4558.
- Deamer, D.W., and Szostak, J.W. (2010). *The origins of life : a subject collection from Cold Spring Harbor perspectives in biology* (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).
- DeLabre, M.L., Kessl, J., Karamanou, S., and Trumpower, B.L. (2002). RPL29 codes for a non-essential protein of the 60S ribosomal subunit in *Saccharomyces cerevisiae* and exhibits synthetic lethality with mutations in genes for proteins required for subunit coupling. *Biochim Biophys Acta* *1574*, 255-261.
- Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., Stark, H., Rodnina, M.V., and Wahl, M.C. (2005). Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* *121*, 991-1004.
- Diedrich, G., Spahn, C.M.T., Stelzl, U., Schäfer, M.A., Wooten, T., Bochariov, D.E., Cooperman, B.S., Traut, R.R., and Nierhaus, K.H. (2000). Ribosomal protein L2 is involved in the association of the ribosomal subunits, tRNA binding to A and P sites and peptidyl transfer. *EMBO J* *19*, 5241-5250.
- Doty, P., Boedtker, H., Fresco, J.R., Haselkorn, R., and Litt, M. (1959). Secondary Structure in Ribonucleic Acids. *Proc Natl Acad Sci U S A* *45*, 482-499.
- Draper, D.E. (2004). A guide to ions and RNA structure. *RNA* *10*, 335-343.

- Dresios, J., Derkatch, I.L., Liebman, S.W., and Synetos, D. (2000). Yeast ribosomal protein L24 affects the kinetics of protein synthesis and ribosomal protein L39 improves translational accuracy, while mutants lacking both remain viable. *Biochemistry* *39*, 7236-7244.
- Dresios, J., Panopoulos, P., Frantziou, C.P., and Synetos, D. (2001). Yeast ribosomal protein deletion mutants possess altered peptidyltransferase activity and different sensitivity to cycloheximide. *Biochemistry* *40*, 8101-8108.
- Dresios, J., Panopoulos, P., Suzuki, K., and Synetos, D. (2003). A dispensable yeast ribosomal protein optimizes peptidyltransferase activity and affects translocation. *J Biol Chem* *278*, 3314-3322.
- Dresios, J., Panopoulos, P., and Synetos, D. (2006). Eukaryotic ribosomal proteins lacking a eubacterial counterpart: important players in ribosomal function. *Mol Microbiol* *59*, 1651-1663.
- Ebert, B.L., Pretz, J., Bosco, J., Chang, C.Y., Tamayo, P., Galili, N., Raza, A., Root, D.E., Attar, E., Ellis, S.R., *et al.* (2008). Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* *451*, 335-339.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* *32*, 1792-1797.
- Edmondson, S.P., Turri, J., Smith, K., Clark, A., and Shriver, J.W. (2009). Structure, stability, and flexibility of ribosomal protein L14e from *Sulfolobus solfataricus*. *Biochemistry* *48*, 5553-5562.
- Emsley, P., and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D - Biological Crystallography* *60*, 2126-2132.
- Erickson, A.H., and Blobel, G. (1983). Cell-free translation of messenger RNA in a wheat germ system. *Methods Enzymol* *96*, 38-50.
- Eswar, N., Madhusudhan, M.S., Marti-Renom, M.A., and Sali, A. (2005). PROFILE\_SCAN: A module for fold assignment using profile-profile scanning in MODELLER. (<http://www.salilab.org/modeller>).
- Ferreira-Cerca, S., Poll, G., Kuhn, H., Neueder, A., Jakob, S., Tschochner, H., and Milkereit, P. (2007). Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol Cell* *28*, 446-457.
- Fok, V., Mitton-Fry, R.M., Grech, A., and Steitz, J.A. (2006). Multiple domains of EBER 1, an Epstein-Barr virus noncoding RNA, recruit human ribosomal protein L22. *RNA* *12*, 872-882.
- Frank, J. (1996). Three-dimensional electron microscopy of macromolecular assemblies. In *Three-dimensional electron microscopy of macromolecular assemblies* (New York).
- Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu Rev Biophys Biomol Struct* *31*, 303-319.
- Frank, J., and Agrawal, R. (2001). Ratchet-like movements between the two ribosomal subunits: their implications in elongation factor recognition and tRNA translocation. *Cold Spring Harb Symp Quant Biol* *66*, 67-75.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *Journal of Structural Biology* *116*, 190-199.

- Frank, J., Zhu, J., Penczek, P., Li, Y.H., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R.K., and Agrawal, R.K. (1995). A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* 376, 441-444.
- Freed, E.F., Bleichert, F., Dutca, L.M., and Baserga, S.J. (2010). When ribosomes go bad: diseases of ribosome biogenesis. *Mol Biosyst* 6, 481-493.
- Gao, Y.G., Selmer, M., Dunham, C.M., Weixlbaumer, A., Kelley, A.C., and Ramakrishnan, V. (2009). The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* 326, 694-699.
- Gazda, H.T., Sheen, M.R., Vlachos, A., Choismel, V., O'Donohue, M.F., Schneider, H., Darras, N., Hasman, C., Sieff, C.A., Newburger, P.E., *et al.* (2008). Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. *Am J Hum Genet* 83, 769-780.
- Gesteland, R.F., and Atkins, J.F. (1993). *The RNA world : the nature of modern RNA suggests a prebiotic RNA world* (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).
- Gesteland, R.F., Cech, T., and Atkins, J.F. (2006). *The RNA world : the nature of modern RNA suggests a prebiotic RNA world*, 3rd edn (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).
- Glansdorff, N., Xu, Y., and Labedan, B. (2008). The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct* 3, 29.
- Gonzalo, P., and Reboud, J. (2003). The puzzling lateral flexible stalk of the ribosome. *Biol Cell* 95, 179-193.
- Gregory, L.A., Aguisa-Toure, A.H., Pinaud, N., Legrand, P., Gleizes, P.E., and Fribourg, S. (2007). Molecular basis of Diamond-Blackfan anemia: structure and function analysis of RPS19. *Nucleic Acids Res* 35, 5913-5921.
- Gregory, S.T., Carr, J.F., and Dahlberg, A.E. (2009). A signal relay between ribosomal protein S12 and elongation factor EF-Tu during decoding of mRNA. *RNA* 15, 208-214.
- Gross, B., Westermann, P., and Bielka, H. (1983). Spatial arrangement of proteins within the small subunit of rat liver ribosomes studied by cross-linking. *EMBO J* 2, 255-260.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849-857.
- Halic, M., Becker, T., Frank, J., Spahn, C.M., and Beckmann, R. (2005). Localization and dynamic behavior of ribosomal protein L30e. *Nat Struct Mol Biol* 12, 467-468.
- Halic, M., Becker, T., Pool, M., Spahn, C., Grassucci, R., Frank, J., and Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature* 427, 808-814.
- Halic, M., Gartmann, M., Schlenker, O., Mielke, T., Pool, M.R., Sinning, I., and Beckmann, R. (2006). Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science* 312, 745-747.
- Hansen, J.L., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2002). Structural insights into peptide bond formation. *Proc Natl Acad Sci USA* 99, 11670-11675.
- Hanszen, K.J. (1971). The optical transfer theory of the electron microscope: fundamental principles and applications. *Advances in Optical and Electron Microscopy* 4, 1-84.
- Harrison, S.C. (2010). Virology. Looking inside adenovirus. *Science* 329, 1026-1027.

- Held, W.A., Ballou, B., Mizushima, S., and Nomura, M. (1974). Assembly mapping of 30 S ribosomal proteins from *Escherichia coli*. Further studies. *J Biol Chem* *249*, 3103-3111.
- Herve du Penhoat, C., Atreya, H.S., Shen, Y., Liu, G., Acton, T.B., Xiao, R., Li, Z., Murray, D., Montelione, G.T., and Szyperski, T. (2004). The NMR solution structure of the 30S ribosomal protein S27e encoded in gene RS27\_ARCFU of *Archaeoglobus fulgidis* reveals a novel protein fold. *Protein Sci* *13*, 1407-1416.
- Hinnebusch, A.G., Dever, T.E., and Sonenberg, N. (2004). Mechanism and Regulation of Protein Synthesis Initiation in Eukaryotes. In *Protein synthesis and ribosome structure: translating the genome*; , K.H. Nierhaus, and D.N. Wilson, eds. (Wennheim, Wiley-VCH).
- Hofacker, I.L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res* *31*, 3429-3431.
- Hofer, A., Bussiere, C., and Johnson, A.W. (2007). Mutational analysis of the ribosomal protein Rpl10 from yeast. *J Biol Chem* *282*, 32630-32639.
- Hosaka, H., Yao, M., Kimura, M., and Tanaka, I. (2001). The structure of the archaeobacterial ribosomal protein S7 and its possible interaction with 16S rRNA. *J Biochem* *130*, 695-701.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD - Visual Molecular Dynamics. *J Molec Graphics* *14*, 33-38.
- Jagannathan, I., and Culver, G.M. (2003). Assembly of the central domain of the 30S ribosomal subunit: roles for the primary binding ribosomal proteins S15 and S8. *J Mol Biol* *330*, 373-383.
- Jenner, L.B., Demeshkina, N., Yusupova, G., and Yusupov, M. (2010). Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol* *17*, 555-560.
- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* *292*, 195-202.
- Jossinet, F., Ludwig, T.E., and Westhof, E. (2010). Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*.
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* *28*, 27-30.
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2006a). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research* *34*, D355-360.
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006b). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* *34*, D354-357.
- Kapasi, P., Chaudhuri, S., Vyas, K., Baus, D., Komar, A.A., Fox, P.L., Merrick, W.C., and Mazumder, B. (2007). L13a blocks 48S assembly: role of a general initiation factor in mRNA-specific translational control. *Mol Cell* *25*, 113-126.
- Kapp, L.D., and Lorsch, J.R. (2004). The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* *73*, 657-704.
- Karimi, R., Pavlov, M., Buckingham, R., and Ehrenberg, M. (1999). Novel roles for classical factors at the interface between translation termination and initiation. *Mol Cell* *3*, 601-609.

- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* *30*, 3059-3066.
- Kavran, J.M., and Steitz, T.A. (2007). Structure of the base of the L7/L12 stalk of the *Haloarcula marismortui* large ribosomal subunit: analysis of L11 movements. *J Mol Biol* *371*, 1047-1059.
- Kim, T.S., Jang, C.Y., Kim, H.D., Lee, J.Y., Ahn, B.Y., and Kim, J. (2006). Interaction of Hsp90 with ribosomal proteins protects from ubiquitination and proteasome-dependent degradation. *Mol Biol Cell* *17*, 824-833.
- Klein, D., Moore, P., and Steitz, T. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *Journal of Molecular Biology* *340*, 141-177.
- Kobayashi, T., Sasaki, Y., Oshima, Y., Yamamoto, H., Mita, H., Suzuki, H., Toyota, M., Tokino, T., Itoh, F., Imai, K., *et al.* (2006). Activation of the ribosomal protein L13 gene in human gastrointestinal cancer. *Int J Mol Med* *18*, 161-170.
- Kopke, A.K.E., Leggatt, P.A., and Matheson, A.T. (1992). Structure Function Relationships in the Ribosomal Stalk Proteins of Archaeobacteria. *J Biol Chem* *267*, 1382-1390.
- Kotliansky, V.E., Domogatsky, S.P., and Gudkov, A.T. (1978). Dimer state of protein L7/L12 and EF-G-dependent reactions on ribosomes. *Eur J Biochem* *90*, 319-323.
- Kravchenko, O., Mitroshin, I., Nikonov, S., Piendl, W., and Garber, M. (2010). Structure of a two-domain N-terminal fragment of ribosomal protein L10 from *Methanococcus jannaschii* reveals a specific piece of the archaeal ribosomal stalk. *J Mol Biol* *399*, 214-220.
- Kressler, D., Hurt, E., and Bassler, J. (2009). Driving ribosome assembly. *Biochim Biophys Acta* *1803*, 673-683.
- Lecompte, O., Ripp, R., Thierry, J.C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* *30*, 5382-5390.
- Leger-Silvestre, I., Milkereit, P., Ferreira-Cerca, S., Saveanu, C., Rousselle, J.C., Choismel, V., Guinefoleau, C., Gas, N., and Gleizes, P.E. (2004). The ribosomal protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *EMBO J* *23*, 2336-2347.
- Lindstrom, M.S. (2009). Emerging functions of ribosomal proteins in gene-specific transcription and translation. *Biochem Biophys Res Commun* *379*, 167-170.
- Liu, H., Jin, L., Koh, S.B., Atanasov, I., Schein, S., Wu, L., and Zhou, Z.H. (2010). Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* *329*, 1038-1043.
- Lodish, H.F. (2008). *Molecular cell biology*, 6th edn (New York, W.H. Freeman).
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology* *128*, 82-97.
- Lutsch, G., Stahl, J., Kargel, H.J., Noll, F., and Bielka, H. (1990). Immunoelectron microscopic studies on the location of ribosomal proteins on the surface of the 40S ribosomal subunit from rat liver. *Eur J Cell Biol* *51*, 140-150.
- Maeder, C., and Draper, D.E. (2005). A small protein unique to bacteria organizes rRNA tertiary structure over an extensive region of the 50 S ribosomal subunit. *J Mol Biol* *354*, 436-446.

- Maguire, B.A., Beniaminov, A.D., Ramu, H., Mankin, A.S., and Zimmermann, R.A. (2005). A protein component at the heart of an RNA machine: the importance of protein l27 for the function of the bacterial ribosome. *Mol Cell* 20, 427-435.
- Marion, M.J., and Marion, C. (1987). Localization of ribosomal proteins on the surface of mammalian 60S ribosomal subunits by means of immobilized enzymes. Correlation with chemical cross-linking data. *Biochem Biophys Res Commun* 149, 1077-1083.
- Márquez, V., Frohlich, T., Armache, J.-P., Sohmen, D., Donhofer, A., Mikolajka, A., Berninghausen, O., Thomm, M., Beckmann, R., Arnold, G.J., *et al.* (2010). Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J Mol Biol* 405, 1215-1232.
- Marti-Renom, M.A., Madhusudhan, M.S., and Sali, A. (2004). Alignment of protein sequences by their profiles. *Protein Sci* 13, 1071-1087.
- Martin-Marcos, P., Hinnebusch, A.G., and Tamame, M. (2007). Ribosomal protein L33 is required for ribosome biogenesis, subunit joining, and repression of GCN4 translation. *Molecular and Cellular Biology* 27, 5968-5985.
- Mazumder, B., Sampath, P., Seshadri, V., Maitra, R.K., DiCorleto, P.E., and Fox, P.L. (2003). Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control. *Cell* 115, 187-198.
- McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404-405.
- Meskauskas, A., and Dinman, J.D. (2001). Ribosomal protein L5 helps anchor peptidyl-tRNA to the P-site in *Saccharomyces cerevisiae*. *Rna* 7, 1084-1096.
- Meskauskas, A., and Dinman, J.D. (2007). Ribosomal protein L3: gatekeeper to the A site. *Mol Cell* 25, 877-888.
- Meskauskas, A., Russ, J.R., and Dinman, J.D. (2008). Structure/function analysis of yeast ribosomal protein L2. *Nucleic Acids Res* 36, 1826-1835.
- Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology* 142, 334-347.
- Moore, P.B., and Steitz, T.A. (2002). The involvement of RNA in ribosome function. *Nature* 418, 229-235.
- Morgan, D.G., Menetret, J.F., Radermacher, M., Neuhof, A., Akey, I.V., Rapoport, T.A., and Akey, C.W. (2000). A comparison of the yeast and rabbit 80 S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments. *J Mol Biol* 301, 301-321.
- Moritz, M., Paulovich, A.G., Tsay, Y.F., and Woolford, J.L., Jr. (1990). Depletion of yeast ribosomal proteins L16 or rp59 disrupts ribosome assembly. *J Cell Biol* 111, 2261-2274.
- Morize, I., Surcouf, E., Vaney, M.C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E., and Mornon, J.P. (1987). Refinement of the C222(1) crystal form of oxidized uteroglobin at 1.34 Å resolution. *J Mol Biol* 194, 725-739.
- Naganuma, T., Nomura, N., Yao, M., Mochizuki, M., Uchiumi, T., and Tanaka, I. (2010). Structural basis for translation factor recruitment to the eukaryotic/archaeal ribosomes. *J Biol Chem* 285, 4747-4756.
- Narla, A., and Ebert, B.L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. *Blood*.



- Neueder, A., Jakob, S., Poll, G., Linnemann, J., Deutzmann, R., Tschochner, H., and Milkereit, P. (2010). A local role for the small ribosomal subunit primary binder rpS5 in final 18S rRNA processing in yeast. *PLoS One* 5, e10194.
- Nikulin, A., Eliseikina, I., Tishchenko, S., Nevskaya, N., Davydova, N., Platonova, O., Piendl, W., Selmer, M., Liljas, A., Drygin, D., *et al.* (2003). Structure of the L1 protuberance in the ribosome. *Nat Struct Biol* 10, 104-108.
- Nishiyama, T., Yamamoto, H., Uchiumi, T., and Nakashima, N. (2007). Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res* 35, 1514-1521.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920-930.
- Noller, H.F. (1991). Ribosomal RNA and Translation. *Annu Rev Biochem* 60, 191-227.
- Nomura, M., Yates, J.L., Dean, D., and Post, L.E. (1980). Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc Natl Acad Sci U S A* 77, 7084-7088.
- Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302, 205-217.
- Okada, U., Sakai, N., Yao, M., Watanabe, N., and Tanaka, I. (2006). Structural analysis of the transcriptional regulator homolog protein from *Pyrococcus horikoshii* OT3. *Proteins* 63, 1084-1086.
- Peske, F., Rodnina, M., and Wintermeyer, W. (2005). Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol Cell* 18, 403-412.
- Pestka, S. (1969). Studies on the formation of transfer ribonucleic acid-ribosome complexes. VI. Oligopeptide synthesis and translocation on ribosomes in the presence and absence of soluble transfer factors. *J Biol Chem* 244, 1533-1539.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J Comput Chem* 25, 1605-1612.
- Pisarev, A.V., Kolupaeva, V.G., Yusupov, M.M., Hellen, C.U., and Pestova, T.V. (2008). Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27, 1609-1621.
- Pisareva, V.P., Hellen, C.U., and Pestova, T.V. (2007). Kinetic Analysis of the Interaction of Guanine Nucleotides with Eukaryotic Translation Initiation Factor eIF5B. *Biochemistry* 46, 2622-2629.
- Pley, H.W., Flaherty, K.M., and McKay, D.B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* 372, 68-74.
- Poll, G., Braun, T., Jakovljevic, J., Neueder, A., Jakob, S., Woolford, J.L., Jr., Tschochner, H., and Milkereit, P. (2009). rRNA maturation in yeast cells depleted of large ribosomal subunit proteins. *PLoS One* 4, e8249.
- Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. *Science*.
- Ramakrishnan, V. (2011). Molecular biology. The eukaryotic ribosome. *Science* 331, 681-682.

- Ramakrishnan, V., and Moore, P.B. (2001). Atomic structures at last: the ribosome in 2000. *Curr Opin Struct Biol* *11*, 144-154.
- Reddy, V.S., Natchiar, S.K., Stewart, P.L., and Nemerow, G.R. (2010). Crystal structure of human adenovirus at 3.5 Å resolution. *Science* *329*, 1071-1075.
- Remacha, M., Jimenez-Diaz, A., Bermejo, B., Rodriguez-Gabriel, M.A., Guarinos, E., and Ballesta, J.P. (1995). Ribosomal acidic phosphoproteins P1 and P2 are not required for cell viability but regulate the pattern of protein expression in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* *15*, 4754-4762.
- Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C.J., *et al.* (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* *452*, 949-955.
- Robledo, S., Idol, R.A., Crimmins, D.L., Ladenson, J.H., Mason, P.J., and Bessler, M. (2008). The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA* *14*, 1918-1929.
- Rodnina, M.V., Stark, H., Savelsbergh, A., Wieden, H.J., Mohr, D., Matassova, N.B., Peske, F., Daviter, T., Gualerzi, C.O., and Wintermeyer, W. (2000). GTPases mechanisms and functions of translation factors on the ribosome. *Biol Chem* *381*, 377-387.
- Rodnina, M.V., and Wintermeyer, W. (2009). Recent mechanistic insights into eukaryotic ribosomes. *Curr Opin Cell Biol* *21*, 435-443.
- Sachs, A., and Davis, R. (1990). The poly(A)-binding protein is required for translation initiation and poly(A) tail shortening. *Mol Biol Rep* *14*, 73.
- Sachse, C., Chen, J.Z., Coureux, P.D., Stroupe, M.E., Fandrich, M., and Grigorieff, N. (2007). High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J Mol Biol* *371*, 812-835.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* *234*, 779-815.
- Schlunzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., *et al.* (2000). Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* *102*, 615-623.
- Schmeing, T.M., Huang, K.S., Strobel, S.A., and Steitz, T.A. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* *438*, 520-524.
- Schmeing, T.M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature* *461*, 1234-1242.
- Schmeing, T.M., Seila, A.C., Hansen, J.L., Freeborn, B., Soukup, J.K., Scaringe, S.A., Strobel, S.A., Moore, P.B., and Steitz, T.A. (2002). A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat Struct Biol* *9*, 225-230.
- Schmeing, T.M., Voorhees, R.M., Kelley, A.C., Gao, Y.G., Murphy, F.V.t., Weir, J.R., and Ramakrishnan, V. (2009). The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* *326*, 688-694.
- Schuler, M., Connell, S.R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroer, B., Mielke, T., Penczek, P.A., Westhof, E., and Spahn, C.M. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol* *13*, 1092-1096.

- Schuwirth, B., Borovinskaya, M., Hau, C., Zhang, W., Vila-Sanjurjo, A., Holton, J., and Cate, J. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827-834.
- Seidelt, B., Innis, C.A., Wilson, D.N., Gartmann, M., Armache, J.-P., Villa, E., Trabuco, L.G., Becker, T., Mielke, T., Schulten, K., *et al.* (2009). Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326, 1412-1415.
- Selmer, M., Dunham, C., Murphy, F.t., Weixlbaumer, A., Petry, S., Kelley, A., Weir, J., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935-1942.
- Sengupta, J., Nilsson, J., Gursky, R., Spahn, C., Nissen, P., and Frank, J. (2004). Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. *Nat Struct Mol Biol* 11, 957-962.
- Sha, F., Lin, Y., Saul, L.K., and Lee, D.D. (2007). Multiplicative updates for nonnegative quadratic programming. *Neural Comput* 19, 2004-2031.
- Shen, M.Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15, 2507-2524.
- Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R. (2004). The ribosome as an entropy trap. *Proc Natl Acad Sci U S A* 101, 7897-7901.
- Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33, W244-248.
- Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731-745.
- Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G., and Frank, J. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. *Cell* 107, 373-386.
- Spahn, C.M., Gomez-Lorenzo, M.G., Grassucci, R.A., Jorgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P., and Frank, J. (2004a). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J* 23, 1008-1019.
- Spahn, C.M., Jan, E., Mulder, A., Grassucci, R.A., Sarnow, P., and Frank, J. (2004b). 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, 465-475.
- Steffen, P., Voss, B., Rehmsmeier, M., Reeder, J., and Giegerich, R. (2006). RNASHAPES: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* 22, 500-503.
- Stombaugh, J., Zirbel, C.L., Westhof, E., and Leontis, N.B. (2009). Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* 37, 2294-2312.
- Sweeney, R., Chen, L.H., and Yao, M.C. (1994). An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. *Molecular and Cellular Biology* 14, 4203-4215.
- Tabb-Massey, A., Caffrey, J.M., Logsden, P., Taylor, S., Trent, J.O., and Ellis, S.R. (2003). Ribosomal proteins Rps0 and Rps21 of *Saccharomyces cerevisiae* have overlapping functions in the maturation of the 3' end of 18S rRNA. *Nucleic Acids Res* 31, 6798-6805.
- Takahashi, Y., Mitsuma, T., Hirayama, S., and Odani, S. (2002). Identification of the ribosomal proteins present in the vicinity of globin mRNA in the 40S initiation complex. *J Biochem* 132, 705-711.

- Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. *Cell* *120*, 49-58.
- Tamura, K., and Alexander, R.W. (2004). Peptide synthesis through evolution. *Cell Mol Life Sci* *61*, 1317-1330.
- Taylor, D.J., Devkota, B., Huang, A.D., Topf, M., Narayanan, E., Sali, A., Harvey, S.C., and Frank, J. (2009). Comprehensive molecular structure of the eukaryotic ribosome. *Structure* *17*, 1591-1604.
- Taylor, D.J., Nilsson, J., Merrill, A.R., Andersen, G.R., Nissen, P., and Frank, J. (2007). Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J* *26*, 2421-2431.
- Tchorzewski, M. (2002). The acidic ribosomal P proteins. *Int J Biochem Cell Biol* *34*, 911-915.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* *22*, 4673-4680.
- Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* *16*, 673-683.
- Uchiumi, T., Honma, S., Nomura, T., Dabbs, E.R., and Hachimori, A. (2002). Translation elongation by a hybrid ribosome in which proteins at the GTPase center of the Escherichia coli ribosome are replaced with rat counterparts. *J Biol Chem* *277*, 3857-3862.
- Valasek, L., Mathew, A.A., Shin, B.S., Nielsen, K.H., Szamecz, B., and Hinnebusch, A.G. (2003). The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. *Genes Dev* *17*, 786-799.
- Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003). Locking and unlocking of ribosomal motions. *Cell* *114*, 123-134.
- Verschoor, A., Srivastava, S., Grassucci, R., and Frank, J. (1996). Native 3D structure of eukaryotic 80S ribosome: Morphological homology with the *E. coli* 70S ribosome. *Journal of Cell Biology* *133*, 495-505.
- Villardell, J., and Warner, J.R. (1997). Ribosomal protein L32 of *Saccharomyces cerevisiae* influences both the splicing of its own transcript and the processing of rRNA. *Molecular and Cellular Biology* *17*, 1959-1965.
- Villardell, J., Yu, S.J., and Warner, J.R. (2000). Multiple functions of an evolutionarily conserved RNA binding domain. *Mol Cell* *5*, 761-766.
- Voet, D., and Voet, J.G. (2004). *Biochemistry*, 3rd edn (New York, J. Wiley & Sons).
- Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C., and Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* *16*, 528-533.
- Wagenknecht, T., Grassucci, R., and Frank, J. (1988). Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from *Escherichia coli*. *J Mol Biol* *199*, 137-147.
- Wang, D.O., Martin, K.C., and Zukin, R.S. (2010). Spatially restricting gene expression by local translation at synapses. *Trends in Neurosciences* *33*, 173-182.

- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* *337*, 635-645.
- Warner, J.R., and McIntosh, K.B. (2009). How common are extraribosomal functions of ribosomal proteins? *Mol Cell* *34*, 3-11.
- Wilson, D.N. (2009). The A-Z of bacterial translation inhibitors. *Crit Rev Biochem Mol Biol* *44*, 393-433.
- Wilson, D.N., and Nierhaus, K.H. (2005). Ribosomal Proteins in the Spotlight. *Crit Rev Biochem Mol Biol* *40*, 243-267.
- Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* *407*, 327-339.
- Woese, C. (1998). The universal ancestor. *Proc Natl Acad Sci U S A* *95*, 6854-6859.
- Wu, B., Lukin, J., Yee, A., Lemak, A., Semesi, A., Ramelot, T.A., Kennedy, M.A., and Arrowsmith, C.H. (2008a). Solution structure of ribosomal protein L40E, a unique C4 zinc finger protein encoded by archaeon *Sulfolobus solfataricus*. *Protein Sci* *17*, 589-596.
- Wu, B., Yee, A., Huang, Y.J., Ramelot, T.A., Cort, J.R., Semesi, A., Jung, J.W., Lee, W., Montelione, G.T., Kennedy, M.A., *et al.* (2008b). The solution structure of ribosomal protein S17E from *Methanobacterium thermoautotrophicum*: a structural homolog of the FF domain. *Protein Sci* *17*, 583-588.
- Yokoyama, T., and Suzuki, T. (2008). Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments. *Nucleic Acids Res* *36*, 3539-3551.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* *292*, 883-896.
- Yusupova, G.Z., Yusupov, M.M., Cate, J.H., and Noller, H.F. (2001). The path of messenger RNA through the ribosome. *Cell* *106*, 233-241.
- Zhu, J., Penczek, P.A., Schroder, R., and Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S *Escherichia coli* ribosome. *Journal of Structural Biology* *118*, 197-219.
- Zimmermann, R.A., and Dahlberg, A.E. (1996). *Ribosomal RNA : structure, evolution, processing, and function in protein biosynthesis* (Boca Raton, CRC Press).