Molecular Architecture of a 40S Ribosomal Subunit Maturation Intermediate

by

Cherisse Rae Loucks

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in The University of Michigan
2011

Doctoral Committee:

Assistant Professor Georgios Skiniotis, Chair
Professor Robert S. Fuller
Associate Professor John J.G. Tesmer
Associate Professor Raymond C. Trievel
Associate Professor Anne B. Vojtek

Dedication

I dedicate this thesis to Dr. Stephen Carper, may his positive attitude in any situation and kind words of wisdom live on in the hearts of those he has touched forever.

Acknowledgments

Being blessed with so many amazing and supportive people in my life makes it hard to know where I should begin, I guess where it all started… I want to thank my parents, Clyde and Arvette, for being my biggest fans, greatest supporters and best friends for the last nearly 27 years. Your constant, unconditional love for one another and myself fills me with joy everyday and reminds me nothing is impossible. I would apologize for all the shenanigans I've put you through, but I don't need to because you know they are part of my charm. My sister, Elli, we couldn't be more different but there are things about me that only you understand about me and I appreciate that immensely. Skylaar, Isaiah, Kiara and Gloria, you are all strong and beautiful and I would do anything for all of you, sorry your aunt has to live 4000 miles away. Mia, you are my sister and my soulmate, I am not exaggerating or being dramatic (imagine that) when I say, I would not be alive with you, I love you. Jaci, thank you for not allowing me to be a high school drop-out, I will always remember you bringing me my homework when I was too scared and anxious to leave my house, I am so proud to see you so happy and grounded, you deserve. Sheyna Lynn, your fearlessness and sense of adventure inspire me everyday, you always tell me how proud you are of me, I hope you

know I am so proud of you, blinding sequin shirt, tightest jeans in North America, huge coif of curls and all.

Alex, Leena and Callie, you guys were the people and escape I needed at the exact second I met all three of you, I will never, ever, ever forget you. Alex, you were my support group, my family, my best friend and shoulder to cry on throughout this stage of my life, I will never be able to repay you for that but hopefully I can be those things for you whenever you need them. Leena, I can't put into words how happy I am when I am around you, I feel like I have known you for my entire life and no matter how long we are apart, nothing ever changes between us. Kate for just getting it, even when I don't know what 'it' is, I don't need words when I am around you, you can sense my energy more than anyone I've ever known.

Lily and Matt, I was not kidding when I said you are my second family, I can't wait to tell your little rascal stories in 20 years about the time their mom wanted to kick everyone out of the bar for being too loud. Lily, continue to be strong there will be an end. Miss Heather, for her wholly Gemini existence, I love telling people that don't know you well about your bratty antics and watch them look completely shocked. Diane Mary and Katie Carr, for taking me in that first summer and your complete lack of judgment, I'll never forget telling you two the story of how Alex and I met. Ashley for never doubting anything I do and not killing Lily and I over the pizza boxes. Donald, for being a great roommate and taking care of me when

I really needed it, Anton is a lucky little guy for having you and Mary as parents. Aimee for just always being there and for your comparable love of Alkaline Trio. Celeste, you have taught me more than anyone the last two years and I will never leave you, just be a little further away. You are my little sister, my sidekick, my partner in crime, you have changed my entire outlook and perception of life, I love you, you are much wiser and grounded than you or anyone knows, but I do. Coco, you have 3 months to make your charm wear off on me, get to work! Jeff and Gerwin, you guys better take care of Lily while I'm gone. And Jeff, I'll find you a supermodel, jewish physicist who can cook amazingly and is as witty as you, mark my words. Finally, Yiorgo, as painful and frustrating our interaction was, I have to give credit for doing something correct, or I wouldn't be writing these acknowledgements today.

Table of Contents

Dedication ii

Acknowledgments iii

List of Figures xi

List of Tables xiv

List of Abbreviations xv

Chapter 1: Understanding Ribosomal Maturation with Electron Microscopy 1

1.1 Abstract *1*

1.2 Eukaryotic ribosome function *2*

1.3 High-resolution structure of the eukaryotic ribosome *2*

1.4 The structure of the 40S subunit *4*

1.5 Initiation of protein translation *5*

1.6 Ribosome biogenesis *8*

1.7 40S assembly *9*

1.8 40S assembly factors in yeast pre-40S *10*

Nob1 12

Rio2 13

Dim1 15

Pno1 16

Tsr1 18

Ltv1 18

Enp1 20

1.9 Electron Microscopy *21*

Basic principles 21

Classification by reference-free alignment 22

Contrast transfer function 22

Angular refinement using model-based projection matching 24

Multiple-reference angular refinement 25

Fitting of X-Ray crystal structures into EM densities 25

1.10 Discussion *26*

1.11 References *28*

Chapter 2: Localization of Assembly Factors on 40S Subunit Precursor **36**

2.1 Abstract *36*

2.2 Results *36*

Isolation of wild-type Rio2-TAP pre-40S 37

Negative stain EM of pre-40S complexes 39

Cryo-EM 3D structure of pre-40S complexes 40

Modeling of ribosomal proteins and RNA in pre-40S 42

Determination of densities unique to precursor 40S subunits 45

2.3 Experimental procedures *47*

Purification of pre-40S complexes 47

Mass spectrometry 49

Specimen preparation and EM imaging of pre-40S complexes 50

Cryo-EM sample preparation and imaging of pre-40S complexes 51

Single-particle reconstruction of pre-40S complexes 51

MDFF: Initial structure preparation 53

MDFF: Initial equilibration phase 57

MDFF simulation details 58

Difference mapping 62

2.4 Discussion *62*

2.5 Concluding remarks *64*

2.6 Acknowledgements *64*

2.7 References *66*

Chapter 3: Localization of Assembly Factors on 40S Subunit Precursor **69**

3.1 Abstract *69*

3.2 Results *70*

The 3D structure of recombinant Tsr1 70

Localization of Nob1 on pre-40S 73

Localization of Rio2 and Dim1 on pre-40S 74

Localization of Tsr1 and Pno1 77

Localization of Ltv1 and Enp1 82

3.3 Materials and methods *89*

Isolation of Nob1-depl, Rio2-depl, Tsr1-depl and ΔLtv1 89

Yeast strains 89

Mass spectrometry 90

Immuno-labeling of pre-40S particles 91

Purification of rTsr1 and MBP-rTsr1 91

Specimen preparation and negative stain EM imaging 92

3D reconstruction of rTsr1 93

Cryo-EM sample preparation and imaging of depletion/deletion pre-40S 94

Single-particle reconstruction of depletion/deletion pre-40S 95

Docking of assembly factors 97

3.4 Discussion *97*

Localization of AFs agrees with their interaction data 98

AFs located on the platform of pre-40S form a network of interactions 101

The presence of Tsr1 holds Helix 44 in its premature conformation 102

Assembly factors block binding sites of translation initiation factors 103

3.5 Concluding remarks *107*

3.6 Acknowledgements *108*

3.7 References *110*

Chapter 4: Tsr1 Prevents Precursor 40S Subunits from Binding 60S Subunits 113

4.1 Abstract *113*

4.2 Results *114*

4.3 Materials and Methods *119*

Tsr1-depl 80S 3D reconstruction 119

4.4 Discussion *120*

4.5 Acknowledgements *122*

4.6 References *123*

Chapter 5: Future Investigations into 40S Biogenesis **124**

5.1 Abstract *124*

5.2 Structural investigation of the Ltv1-Enp1-Rps3 subcomplex *125*

Rationale 125

Proposed investigation 126

5.4 Structural investigation into Tsr1 and helix 44 interactions *127*

Rationale 127

Proposed investigation 127

5.5 Investigating of signal transduction between Rio2 and Nob1 *128*

Rationale 128

Proposed investigations 128

5.3 Discussion *129*

5.4 Acknowledgements *130*

5.5 References *131*

List of Figures

Figure 1.1: High resolution structure of the eukaryotic ribosome. 4

Figure 1.2: 40S subunit structure. 5

Figure 1.3: Cooperative binding of eIF1/1A opens the 40S mRNA channel. 7

Figure 1.4: Binding of eIF1, tRNA and eIF3 on mature 40S. 7

Figure 1.5: 40S biogenesis occurs in multiple steps and subcellular locations. 9

Figure 1.6: PIN domain of Nob1 is predicted to form a homotetramer. 12

Figure 1.7: The crystal structure of Rio2. 14

Figure 1.8: Crystal structure of human Dim1. 15

Figure 1.9: The archaeal ortholog of Pno1 binds eIF2A and GAAUC. 17

Figure 2.1: SDS-PAGE analysis of AFs in pre-40S. 38

Figure 2.2: Negative stain EM analysis of pre-40S. - 40

Figure 2.3: Pre-40S cryo-EM 3D reconstruction. 41

Figure 2.4: Data and analysis of the pre-40S cryo-EM 3D reconstruction. 42

Figure 2.5: Fitting of rps and rRNA into our pre-40S 3D map. 44

Figure 2.6: Subtraction of the rRNA from the pre-40S density. 46

Figure 2.7: Molecular architecture of pre-40S. 47

Figure 2.8: Tandem affinity purification of pre-40S. 49

Figure 2.9: MDFF results for the fitting of 3O2Z into the pre-40S 3D map. 60

Figure 2.10: MDFF results for the fitting of 3JYV into the pre-40S 3D map. 61

Figure 3.1: Negative stain EM analysis of purified recombinant Tsr1. 71

Figure 3.2: An MBP-tag on rTsr1 allows for the localization of its N-terminus. 72

Figure 3.3: 3D reconstruction of Tsr1. 73

Figure 3.4: Localization of Nob1 on pre-40S. 74

Figure 3.5: Localization of Rio2 and Dim1 on pre-40S. 77

Figure 3.6: The conformation of helix 44 is altered in Tsr1-depl pre-40S. 79

Figure 3.7: Localization of Tsr1 on pre-40S. 80

Figure 3.8: Localization of Pno1 on pre-40S. 82

Figure 3.9: Localization of Ltv1 and Enp1 on pre-40S. 84

Figure 3.10: Multi-reference alignment of the ΔLtv1 projection dataset. 88

Figure 3.13: Structural model of a 40S subunit precursor. 98

Figure 3.14: Comparing AF locations with RNA interaction data. 100

Figure 3.15: Comparing AF locations with protein interaction data. 101

Figure 3.16: AFs obstruct binding sites for factors required for translation. 105

Figure 3.17: The binding of AFs inhibits the opening of the mRNA channel. 106

Figure 4.1: Northern blot analysis of Tsr1-depl pre-40S. 115

Figure 4.2: Raw cryo-EM micrograph of Tsr1-depl shows 80S-like particles. 116

Figure 4.3: Cryo-EM 3D map of 80S-like particles from pre-40S. 118

List of Tables

Table 2.1: Mass spectrometry analysis of rps in pre-40S. **39**

Table 2.2: Rps structures used in our pre-40S model. **70**

Table 2.3: Cross-correlation coefficient (CCC) values for MDFF runs. **62**

Table 3.1: Mass spectrometry analysis of rps in pre-40S particles. **90**

List of Abbreviations

C	Celsius
CBP	Calmodulin binding peptide
CCC	Cross-correlation coefficient
DTT	Dithiothreitol
EGTA	Ethylene glycol tetraacetic acid
EF-Tu	Elongation factor Tu (EF-Tu)
fs	Femtosecond
HA	Hemagluttin
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
kcal	Kilocalorie
KCl	Potassium chloride
kDa	Kilodaltons
L	Liter
LB	Luria Broth
LC	Liquid chromatography

ΔLtv1	Pre-40S complexes from yeast cells with Ltv1 genetically deleted
$MgCl_2$	Magnesium chloride
mM	Millimolar (mM)
mol	Moles
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
NaCl	Sodium chloride
Ni-NTA	Nickel nitrilotriacetic acid (Ni-NTA)
Nob1-depl	Pre-40S complexes from yeast cells with Nob1 depleted with a galactose-inducible promoter
NP-40	Nonyl phenoxypolyethoxylethanol 40
OD_{600}	Optical density at 600 nanometers
PMSF	Phenylmethanesulfonylfluoride (PMSF)
RMSD	Root mean squares deviation
Rio2-depl	Pre-40S complexes from yeast cells with Rio2 depleted with a galactose-inducible promoter
rTsr1	Recombinant *S. cerevisiae* Tsr1
T. lanuginosus	*Thermomyces lanuginosus*
TEV	Tobacco etch virus

TRIS-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
Tsr1-depl	Pre-40S complexes from yeast cells with Tsr1 depleted with a galactose-inducible promoter
Tsr1-depl 80S-like	Particles isolated from Tsr1-depl with similar appearance to WT 80S ribosomes
µL	Microliters
VMD	Visual Molecular Dynamics

Chapter 1
Understanding Ribosomal Maturation with Electron Microscopy

1.1 Abstract

The biogenesis of eukaryotic ribosomes is a highly regulated process and requires approximately 200 assembly factors (AFs). The process begins with transcription of ribosomal RNA (rRNA) in the nucleolus followed by maturation events in both the nucleolus and the cytoplasm. Although many assembly factors have been identified, detailed knowledge of their specific functions and binding locations on ribosomal precursor subunits is lacking. To understand the maturation of the 40S subunit (40S) and specific AF functions in more detail, structural characterization of a precursor 40S subunit is necessary. Cryo-EM is an invaluable tool for studying the 3D structures of these precursor complexes, as they are large, flexible and heterogeneous, characteristics making macromolecular crystallization difficult. Despite current limits on resolution levels attainable with cryo-EM, detailed structural models are possible by combining cryo-EM 3D maps with with X-ray crystallographic and molecular dynamics data.

1.2 Eukaryotic ribosome function

The ribosome is a large macromolecular complex comprised of a small (SSU) and large (LSU) subunit. The SSU and LSU are 40S and 60S in eukaryotic cells, and 30S and 50S in prokaryotic cells, respectively. The ribosome, 80S in eukaryotes and 70S in prokaryotes, catalyzes protein synthesis by translation of messenger RNA (mRNA) into polypeptides. Peptide bond formation occurs by messenger (mRNA) codon and amino-acylated transfer RNA (tRNA) anticodon recognition on the decoding site on 40S, comprised of rRNA nucleotides highly conserved in all living cells[1,2]. The 80S holoenzyme contains four rRNA transcripts and approximately 80 ribosomal proteins (rps). After transcription from a polycistronic gene, a 35S rRNA transcript undergoes multiple cleavage steps to generate all four rRNA transcripts found in the mature ribosome. Rps begin binding the rRNA in the nucleolus, facilitating rRNA folding events and stabilizing its structure[3]. Precursors to the mature 40S and 60S subunits are exported from the nucleolus and maturation is completed in the cytoplasm[4]. With the help of initiation factors, mature subunits join to form the 80S ribosome (the process of translation initation is reviewed by Kapp[5] and Sonnenburg[6]).

1.3 High-resolution structures of the eukaryotic ribosome

Recently, several groups have calculated the structure of the eukaryotic ribosome using X-ray crystallography and electron cryomicroscopy (cryo-EM). Although

the crystal structure of the prokaryotic 70S ribosome had been solved to atomic resolution in 2000[7,8,910], knowledge of the eukaryotic ribosome structure had relied on 3D maps calculated from cryo-EM studies until late 2010. The first 80S ribosome cryo-EM 3D reconstructions with moderate resolutions (below 30 Angstroms (Å)) were determined using *Saccharomyces cerevisiae (S. cerevisiae)*[11] and *Tetrahymena thermophila* (thermophilic yeast)[12] ribosomes, which were solved to 17.5 Å and 8.9 Å, respectively. The resolution of these 3D reconstructions prevented localization and modeling of all ribosomal proteins and RNA. More detailed structural information was facilitated with a cryo-EM 3D map of a yeast 80S solved to 5.5 Å by the Beckmann group[13] (Protein Databank (PDB) identification codes (IDs): 3O2Z and 3O58). The crystal structure of the macromolecular machine from *S. cerevisiae* was solved in November 2010 to a resolution of 4.15 Angstroms (Å)[14]. Shortly after, the 40S subunit from *T. thermophila* with eukaryotic initiation factor (eIF) 1 bound (PDB ID: 2XZM) was solved to 3.9 Å[15]. These crystal structures allowed nearly all the rps to be identified and structurally characterized in the 80S ribosome.

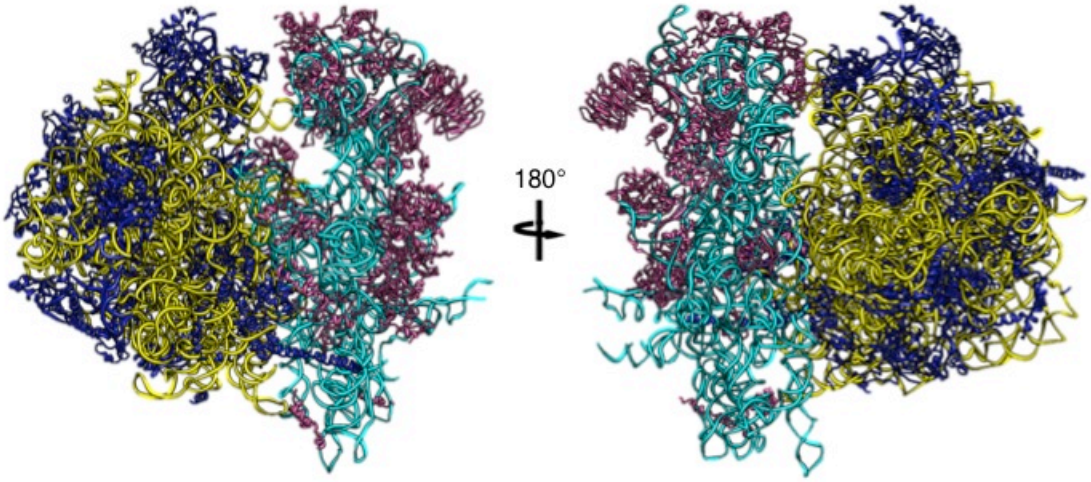


Figure 1.1: High resolution structure of the eukaryotic ribosome. Ribbon representation of the 4.15 Å *S. cerevisiae* 80S ribosome. Rps are shown in magenta and yellow, and rRNA is shown in cyan and navy blue, for 40S and 60S, respectively (PDB IDs: 3O2Z (40S) and 3O58 (60S)).

1.4 The structure of the 40S subunit

Investigation of the eukaryotic 40S ribosomal subunit (40S) shows a 'structural core' formed by the rRNA with the rps binding peripherally on the rRNA. Rps play roles in structural maintenance and rRNA processing of the small subunit during assembly[3]. The 40S subunit possesses four distinct structural regions: the head, the beak, the platform and the foot (**Figure 1.2**). The solvent exposed side of pre-40S is shown in the left panel of **Figure 1.2** and the 60S-binding side of the complex is shown in the right panel of **Figure 1.2**. The binding interface contains a large network of secondary structure elements responsible for the bridging of the SSU and LSU[16]. The mRNA channel extends perpendicular to the long-axis of the particle and is located on the platform, the area below the head on the binding interface. The binding sites for multiple eukaryotic initiation factors

(eIFs), as well as the tRNA amino-acyl, peptidyl and exit sites (A-site, P-site, E-site; respectively), are also on the platform[14,17,18,19,20,21]. The foot is made from the bottom of helix 44 (numbered helices refer to secondary structure elements of the 18S rRNA). Helix 44 contains the active site nucleotides of the decoding site[22,23].

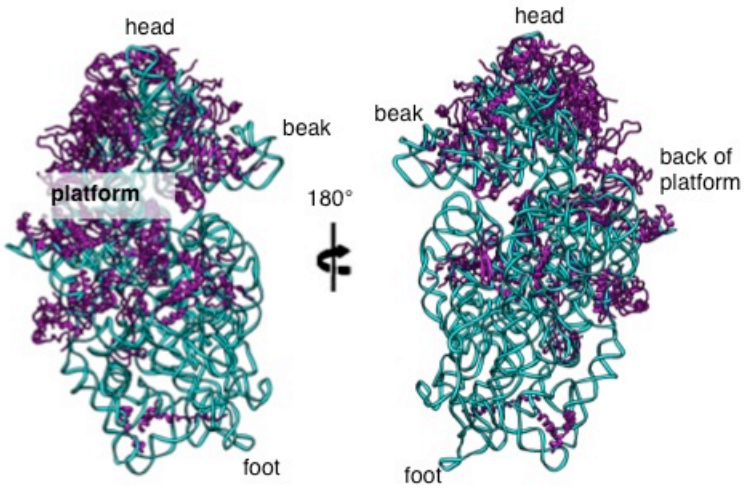


Figure 1.2: 40S subunit structure. Ribbon representation of the 40S subunit alone from the 4.15 Å *S. cerevisiae* 80S ribosome crystal structure (shown in **Figure 1.1**)[14]. Rps are magenta and rRNA shown in cyan. The four characteristic structural features of 40S (head, platform, beak and foot) are labeled. The view in the left panel shows the solvent-exposed side and the view in the right panel shows the subunit-binding interface.

1.5 Initiation of protein translation

Translation initiation is a multiple step process that requires the binding of many factors, including 12 identified eukaryotic initiation factors (eIFs) (reviewed by Kapp[5]). First, the ternary complex (TC) of GTP bound eIF2 and amino-acylated initiator tRNA ($tRNA_i$) is delivered to the P-site of the mature 40S subunit, making

pre-43S initiation complex (PIC)[24]. eIF1, eIF1A, eIF3 and the poly(A)-binding protein all assist in the binding of the 5'-end of the mRNA to the PIC. The small subunit scans the mRNA for the start codon followed by the eIF5-catalyzed conversion of guanine triphosphate (GTP) to guanine diphosphate (GDP) on eIF2, releasing eIF2-GDP[25,26]. At this stage, the tRNAi is bound to the mRNA start codon and, with the help of eIF5B, the 40S can join the 60S, forming the translation elongation competent 80S complex[25,27,28].

An understanding of the process of translation initiation was aided by cryo-EM and X-ray crystallographic structures of the 40S complex with various initiation factors bound[14,17,18,19,20,21]. A series of low-resolution maps show that the cooperative binding of eIF1 and eIF1A is required for the opening of the mRNA channel. In addition to the mRNA channel being opened, a connection is formed between the head and body on the solvent interface of 40S, stabilizing the open conformation of the channel[18] (**Figure 1.3A**). The binding of the factors and channel opening allows the complex to scan the mRNA[29-32] (**Figure 1.3A)**. Additional cryo-EM 3D maps show eIF1 binds 40S on the platform near the beak[15] (**Figure 1.4A),** and eIF3 binds on the back of the platform, opposite the beak[17] (**Figure 1.4B)** Despite substantial structural knowledge of the mature ribosome and 40S translation initiation intermediates, the assembly of the 40S subunit is not well understood.

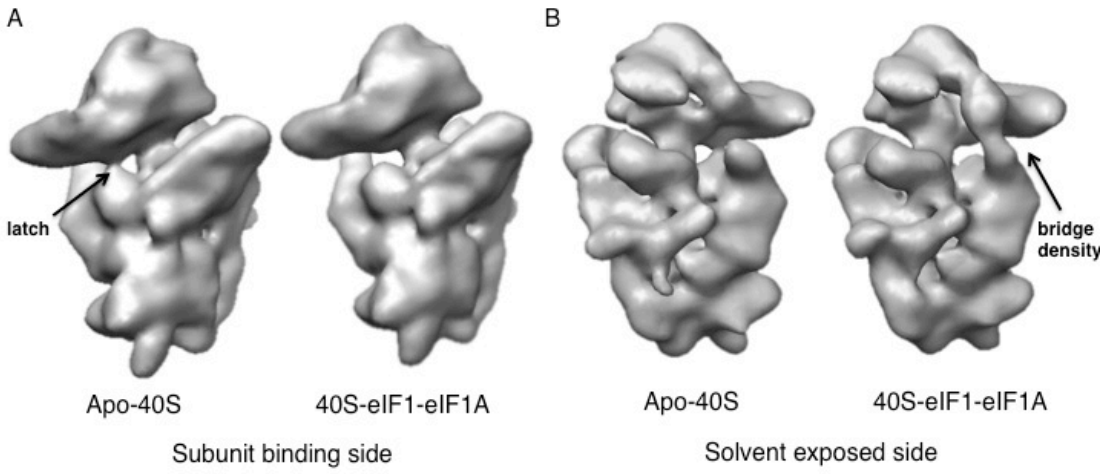


Figure 1.3: Cooperative binding of eIF1/1A opens the 40S mRNA channel. The subunit binding side of mature 40S contains a 'latch' keeping the mRNA channel closed (left panel, EMDB ID: 1346). Cooperative bind of eIF1 and eIF1A bind to 40S opens the mRNA channel by releasing the latch (right panel, EMDB ID: 1347). (B) The opening of the latch is stabilized by a new connection between the beak and platform on the solvent side (arrow in right panel) in 40S-eIF1-eIF1A[18].

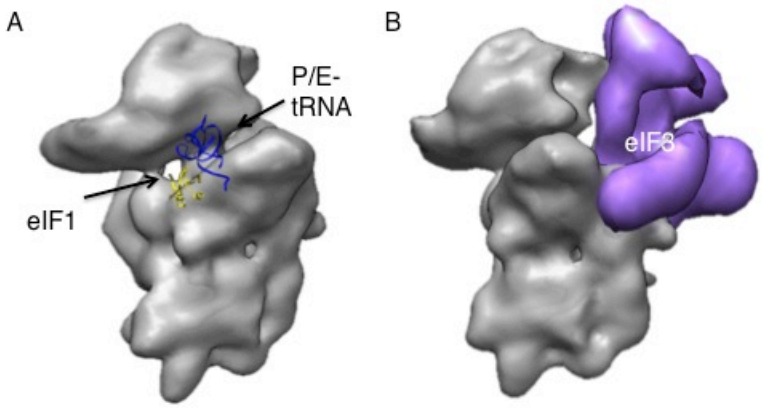


Figure 1.4: Binding of eIF1, tRNA and eIF3 on mature 40S. (A) Mature 40S (gray) with eIF1[15] (yellow) and P/E-tRNA[12] (royal blue) bound, both bind the 40S near the mRNA channel on the platform. (B) eIF3 (purple) binds at the back of the platform on 40S (gray)[17].

1.6 Ribosome biogenesis

Ribosome biogenesis is a highly regulated process, which begins with transcription of the 35S rRNA from a polycistronic gene in the nucleolus and continues through export of precursor subunits to the cytoplasm[4] (**Figure 1.5**). The 35S precursor rRNA contains 5' and 3' external transcribed spacers (ETS) and two internal transcribed spacers (ITS), ITS-1 and ITS-2. Multiple cleavage events yield the 18S rRNA found in the mature 40S subunit, and the 5S, 5.8S and 25S rRNAs associated with the mature 60S[33-35]. rRNA cleavage steps are facilitated by ribosomal proteins and trans-acting non-ribosomal assembly factors (AFs), which are also required for the proper formation of mature subunits[36,37]. Approximately 200 AFs have been identified in biogenesis of the subunits[38]. The AFs transiently associate with the maturation intermediates at various stages in the pathway, with binding observed in both the nucleolus and cytoplasm[39,40]. AFs are rarely found in polysomes[41], indicating their dissociation is required for the binding of eIFs, tRNAi and the mRNA to 40S, as well as joining of the small and large subunits. In addition to potentially serving as quality control measures by inhibiting the binding of translation initiation factors to immature subunits, AFs have a wide variety of functions in the maturation process, including RNA methylation and cleavage, protein phosphorylation, and precursor subunit nuclear export (reviewed by Strunk et al.[42]).

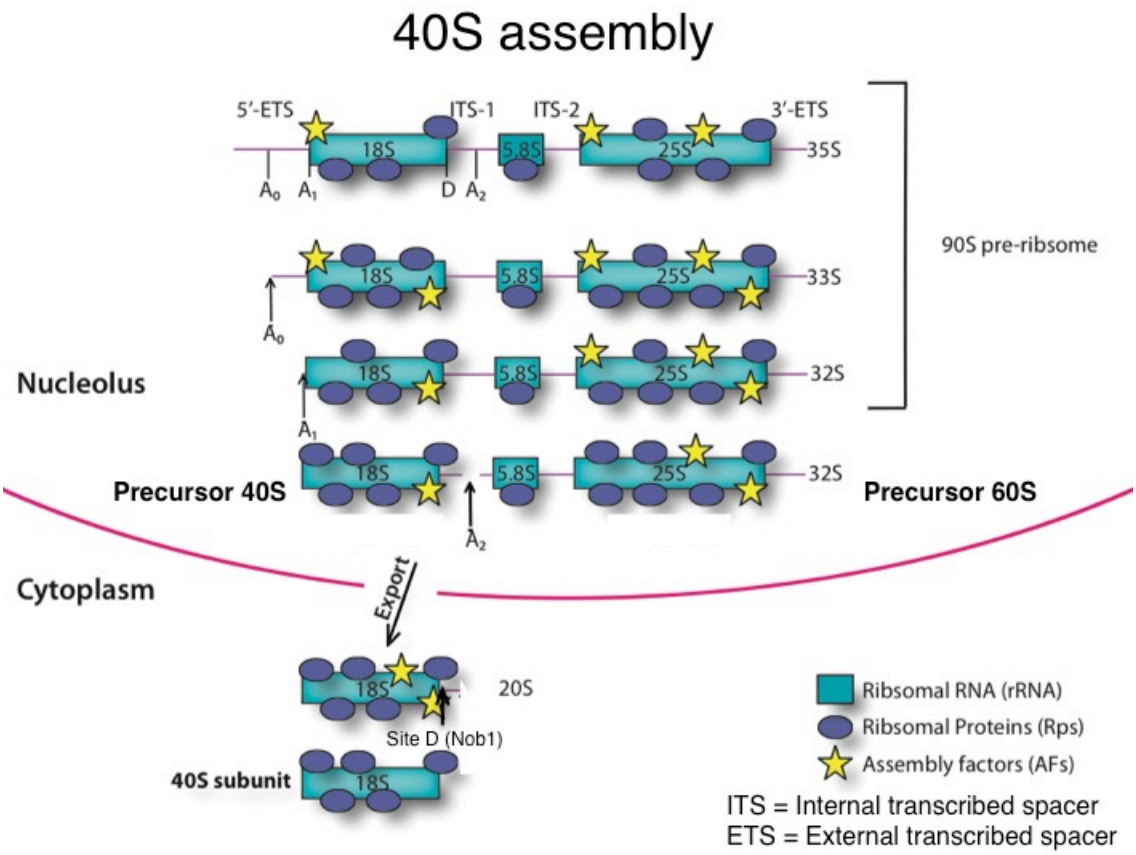


Figure 1.5: 40S biogenesis occurs in multiple steps and subcellular locations. Ribosome biogenesis begins in the nucleolus with the transcription of the 35S rRNA (green boxes with ITS and ETS shown as pink lines), most rps (represented by purple ovals) and some AFs (yellow stars) bind the rRNA in the nucleolus. After cleavage at sites A_0, A_1 and A_2 the premature 40S is exported to the cytoplasm containing the premature 20S rRNA precursor. Additional AFs, including Nob1 which cleaves the 20S rRNA at site D to yield the mature 18S rRNA, associate in the cytoplasm. AFs dissociate prior to incorporation of subunits into 80S complexes.

1.7 40S assembly

Cleavage of the 35S rRNA transcript at sites, A_0 and A_1 in ETS-1, and A_2 in ITS-1[43,44] yields the 20S rRNA precursor found in the pre-mature 40S. The 40S precursor intermediate containing the 20S rRNA is exported to the cytoplasm

with most of the mature rps, and the AFs that bind the complex in the nucleolus[41,45,46]. After export, additional AFs bind the premature 40S subunit (pre-40S) in the cytoplasm[41,44,47-49]. The final cleavage step to generate the mature 18S from the 20S rRNA occurs at site D, located at the 3' end of the 18S rRNA, by the endonuclease AF, Nob1[44,50,51]. Additional AFs are required for site D and earlier cleavage events, as shown by the accumulation of multiple precursor rRNAs in yeast cells with the factors depleted[41]. The functional roles in maturation and their binding sites, effectors and substrates of AFs on the premature subunits is poorly understood. Aberrations in the production of human ribosomes can lead to countless pathologies including Diamond-Blackfan anemia[4,52], myelodysplastic syndromes[53], dyskeratosis congenita, cartilage hair hypoplasia and Treacher Collins syndrome[54]. To better understand the process and regulation of ribosome biogenesis, a structural understanding of the 40S premature subunit intermediates is required. The localization of the AFs on 40S precursors is an important step in delineating their interactions with rRNA and other proteins, potentially narrowing down their functions, mechanisms, regulators and downstream targets.

1.8 40S assembly factors in yeast pre-40S

Protein families of AFs responsible for ribosome biogenesis include, but are not limited to, GTPases, ATPases, helicases, kinases, export adaptors and RNA-

modifying enzymes, such as methylases (reviewed by Strunk et al.[42]). Similar to their binding on 40S, dissociation of AFs occurs at different stages and subcellular locations[41], contributing to different populations of 40S maturation intermediates in cells. To isolate a late cytoplasmic premature 40S (pre-40S) from yeast cells, Schaefer and colleagues added a tandem affinity purification (TAP) tag to the C-terminal end of the kinase, Rio2, which preferentially associates with the cytoplasmic precursor 40S subunits over those in the nucleolus[41,46]. Seven AFs are associated with this particle: Rio2, a serine kinase with an unknown substrate; Ltv1, which is expected to be nuclear export factor; Tsr1 and Enp1, proteins with unknown functions; Nob1, the endonuclease responsible for cleavage of the 20S rRNA at site D; Dim1, an RNA methylase which modifies two adenosine residues near site D; and Pno1, a protein containing three K homology (KH) domains, that are implicated in nucleotide-binding. Northern blot analysis confirmed the isolated late-cytoplasmic complex contains the 20S precursor rRNA. Biochemical data has provided a rough estimate of the interactions of the AFs[2,51,55,56] on pre-40S, but determination of their specific binding sites will provide a better understanding of the overall assembly of 40S.

Nob1

Nob1 produces the mature 18S rRNA by endonucleolytic cleavage of the 20S rRNA at site D[44,50,51]. Nob1 binds pre-40S in the cytoplasm and contains a N-terminal PIN domain (amino acids 1-130), which is homologous to members of the RNAse H superfamily[41,57-59]. This domain is required for cleavage at site D[74]. The crystal structure of an archaeal PIN domain has been solved[58], and was crystallized as a tetramer[44]. The center of the ring formed by PIN domains is lined by four conserved metal-chelating acidic residues (D15/E43/D92/D110) required for site D cleavage, and it is hypothesized that the single-stranded ITS1 passes through the channel (**Figure 1.6**).

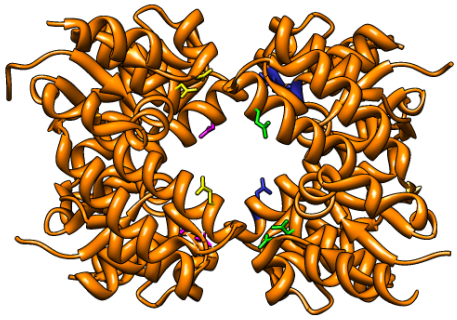


Figure 1.6: PIN domain of Nob1 is predicted to form a homotetramer. Crystal structure of homotetramer formed by an archaea PIN domain[57] (PDB: 1V8P). Active site residues are differentially colored and shown in as sticks.

On the pre-40S complex, RNA footprinting data shows Nob1 binds at cleavage site D and to nucleotides in ITS-1[51]. Both regions are located on the back of the platform in the mature 40S subunit. This location is near Rps5 and Rps14, ribosomal proteins that are required for site D cleavage[60] and interact with

Nob1[55]. However, cross-linking analysis indicates Nob1 also binds helix 40, located in the head domain, approximately 33 Å away from site D[56]. These conflicting localization data might indicate a shift in the positioning of Nob1 on pre-40S closer to site D immediately prior to cleavage. Furthermore, Nob1 can be found on pre-40S complexes containing the precursor 20S rRNA[41] indicating that an activation event or conformational change is required for cleavage. Further understanding of specific binding site of Nob1 on pre-40S and its distance from site D may elude the signals responsible for activation or movement of Nob1 on pre-40S.

Rio2

Rio2 is a serine kinase and a member of the Right Open (RiO) reading frame family of atypical protein kinases[61]. The crystal structure of archaeal Rio2 reveals a RiO kinase domain and an N-terminal domain with a winged-helix fold[62] (PDB ID: 1ZAO) (**Figure 1.7**). Although Rio2 possesses a canonical RiO kinase fold, phosphorylation targets have not been identified on pre-40S. In human cells, the kinase activity of Rio2 is required for the release of Pno1, Ltv1 and Nob1 from pre-40S particles[63], indicating Rio2 activity may signal for an event, such as a conformational change, that releases the AFs by altering the conformation of pre-40S at their binding sites. Rio2 is conserved in eukaryotic cells and its depletion causes the accumulation of the 20S rRNA in both yeast[64-66] and human[67] cells.

Rio2 interacts with Rps14 and Rps5[55], located on the platform and the head, respectively. Rio2 crosslinks to helix 31[56], located in the head near the mRNA channel, specifically at the P-site. Binding of Rio2 to this region would block the binding sites on mature 40S for several factors required for initiation of protein translation, including the mRNA, tRNA and eIF1. Determining the location of Rio2 on the pre-40S complex will confirm whether its association with the premature complex blocks binding sites for initiation factors. Its localization will also allow a more thorough understanding of the upstream signal(s) required for activation of the kinase and the downstream target(s) it modifies to regulate site D cleavage and recycle AFs on pre-40S.

Figure 1.7: The crystal structure of Rio2. Crystal structure from yeast Rio2's archaeal ortholog shows an N-terminal winged helix fold domain and C-terminal RiO kinase-like fold[61] (PDB ID: 1ZAO).

Dim1

Dim1 is a S-adenosyl-L-methionine (SAM) transferase that is conserved from bacteria to humans[2,68]. Dim1 transfers methyl groups to two adjacent adenosine nucleotides in the loop between helices 44 and 45 near site D[69,70]. The modification occurs in the cytoplasm[39] and is required for cleavage by Nob1[51]. The crystal structure of human Dim1 has been solved[71] (PDB ID: 2ZQ9) (**Figure 1.8**) and contains two K homology (KH) domains. The KH domains are highly conserved and found in the crystal structure of the bacterial ortholog of Dim1, KsgA[2]. The larger N-terminal domain consists of mixed α-helical and β-sheet secondary structure elements, and the C-terminal domain contains only α-helical elements. A flexible linker connects the domains and the C-terminal domain sits on top of the N-terminal domain. The active site residues are positioned in the interface of the domains.

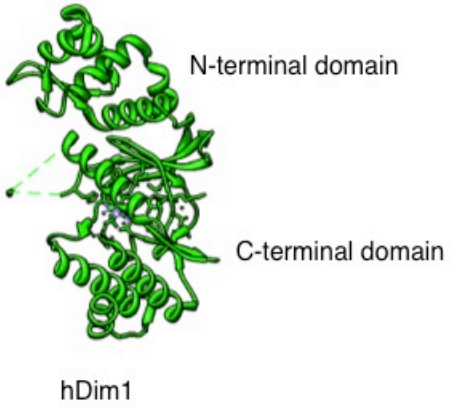


Figure 1.8: Crystal structure of human Dim1. The N-terminal KH domain sits on top of the C-terminal KH domain, the active site residues are located at the interface of the domains (PDB: 2ZQ9)[2].

Footprinting experiments show KsgA interacts with premature 30S complexes at the top of helix 44[2]. Dim1's adenosine substrates are in the loop between helices 44 and 45 located on the platform of pre-40S. Dim1 cross-links to helices 2 and 28, which make the central pseudo-knot of the small subunit[56], also located on the platform. Helix 28 is required for initiator methionine tRNA ($tRNA_{iMet}$) binding[72], indicating the binding site of Dim1 may overlap that of $tRNA_{iMet}$. Localization of Dim1 on pre-40S will indicate if association of the methylase prevents binding of factors required for translation initiation. Localization of Dim1 will also provide insight into the signaling pathway activating the methylase and the downstream targets of this modification.

Pno1

The specific function of Pno1 in 40S maturation is unknown. It has been shown that Pno1 associates with both pre-90S and pre-40S complexes[73] and is required for rRNA cleavage at sites A_1, A_2[74] and site D[73,75,76]. Pno1 contains 3 KH domains, which include the highly conserved nine residue amino acid sequence VIGxxGxxI (V=valine, I=isoleucine, G=glycine, x=any amino acid) implicated in RNA-binding[73]. The central KH domain binds to the PIN domain of Nob1[74] and is required for Nob1 cleavage. It is also required for rRNA methylation by Dim1[73].

Pno1's archaeal ortholog, Dim2p, forms a trimeric complex with the universally conserved GGAUC sequence at the 3' end of the 16S rRNA and eIF2α. This complex has been crystallized, revealing that Dim2p interacts directly with both molecules[77] (**Figure 1.9**)(PDB ID: 2E3U). eIF2 (comprised of α, β, and γ subunits) facilitates binding of initiator $tRNA_i$ to the small subunit. $tRNA_i$ binds the 40S in the mRNA channel region, between the platform and the head. Localization of Pno1 on the back of the platform places it near the binding site of eIF2 and the 3' end of the 18S rRNA. Determining the precise location of Pno1 on pre-40S may reveal it blocks the binding of $tRNA_i$ and/or eIF2. It may also allow an understanding of its mechanism of regulating the binding and/or activity of Dim1 and Nob1.

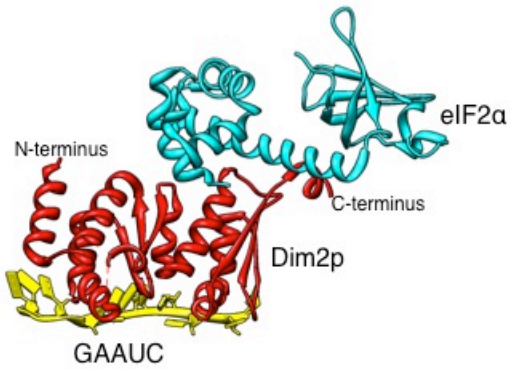


Figure 1.9: The archaeal ortholog of Pno1 binds eIF2A and GAAUC. The Pno1 archaeal ortholog, Dim2p (red), was crystallized in complex with eIF2A (cyan) and the universally conserved sequence, GAAUC (yellow) found near the 3' of the 18S rRNA. The crystal structure contains two of the three KH domains present in Pno1, it is lacking the C-terminal KH domain.

Tsr1

Similar to other 40S assembly factors, Tsr1 is required for generation of the 18S rRNA from its 20S precursor[48]. Tsr1 is found in both the nucleolus and cytoplasm indicating it binds pre-40S in the nucleolus[41]. Tsr1 contains sequence homology to GTPases but its activity has not been shown. The structure of Tsr1 is also unknown. Tsr1 is one of the last AFs to be dissociated from the pre-40S but is not found polysomes[41], in contrast to Nob1 and Pno1[41]. Protein pull-down experiments show that Tsr1 interacts with Rio2[55], which crosslinking data has localized near the platform. RNA-protein cross-linking data reveals Tsr1 interacts with helix 44[56], also on the subunit binding side and near the platform of pre-40S. By determining the specific location of Tsr1 we can determine whether it prevents binding of initiation factors with other AFs predicted to be found on the platform. Its localization may also allow the determination of its function and role in regulating cleavage by Nob1.

Ltv1

Ltv1 was initially predicted to be a nuclear export adaptor for pre-40S, due to its interaction with Crm1, a factor necessary for the export of both the pre-40S[45,76] and premature 60S subunits[78] In yeast cells, the genetic deletion of Ltv1 (*Δltv1*) resulted in the accumulation of Rps3 in the nucleus[45] which was used as a marker for the localization of 40S. Additional studies attempted to confirm this role of Ltv1 by mutating or deleting a leucine-rich region of Ltv1, expected to

function as its nuclear export sequence (NES). The NES-mutant Ltv1 proteins were incorporated into pre-40S complexes that were still capable of export to the cytoplasm but site D cleavage was blocked[79]. Interestingly, this phenotype is not observed in *Δltv1* yeast cells[45]. Fassio and colleagues[79] proposed two possible structural explanations for this observation. First, the NES-mutant Ltv1 is preventing the release of one or multiple AFs from pre-40S, blocking the exposure of site D and sterically inhibiting cleavage. A potential candidate is Tsr1 because it remains in the cytoplasm when the Ltv1 NES-mutant is over-expressed in yeast cells[79]. The second hypothesis is that the Ltv1 NES-mutant prevents an overall pre-40S conformational change, preventing exposure of site D or the binding of a protein required for cleavage.

Biochemical data exists for numerous interactions of Ltv1 on pre-40S. Ltv1 cross-links to nucleotides on helix 41[56] located at the head. Ltv1 interacts directly with Enp1[55] and Ltv1 forms a trimeric complex with Enp1 and Rps3 that is released when pre-40S is exposed to high salt concentrations[46]. In mature 40S, Rps3 is located in the head of pre-40S, near the density connecting the beak to the platform of the complex. Helix 41 is also located at the bridge and Rps15 is located in the head near the beak. All data are consistent with placement of Ltv1 in the region connecting the bridge and head of pre-40S. Determining the

precise location of Ltv1 will provide clues on how this AF causes a structural change or release of AFs to allow cleavage of site D.

Enp1

The assembly factor Enp1 is involved throughout the ribosomal maturation process, as indicated by its association with both the nucleolar premature 90S complex[41] and the late cytoplasmic pre-40S[36]. Depletion of Enp1 causes accumulation of 35S rRNA and Northern blot analysis confirmed it is essential for early rRNA processing at sites A_0, A_1 and A_2[80]. Enp1 interacts with the small nucleolar RNAs (snoRNAs) U3 and U14, also implicated in pre-rRNA processing[81,82]. Although Enp1 has roles in multiple steps of 40S biogenesis and possesses homologues in all eukaryotes, its function on pre-40S is not well understood. Enp1's role in the maintaining the structural integrity of the premature intermediates has been hypothesized due to its presence throughout assembly and its involvement in the formation of the beak structure[41]. In addition to forming a trimeric complex with Ltv1 and Rps3[46], Enp1 also cross-links to helix 33 and helix 34[56], which are found in the beak and bridge, respectively. Placement of Enp1 in the region near the bridge connecting the beak to the body agrees with each of these interactions. Structural characterization of Enp1 on pre-40S may provide additional evidence for its role in 40S assembly, potentially in formation of the beak structure.

1.9 Electron Microscopy

Basic principles

In transmission electron microscopy (TEM) the contrast of the image is primarily generated from the interference of scattered and unscattered electron waves[68]. Scattering is caused by deflection of electrons after interaction with the specimen. In biological studies, the specimen is the biological sample itself or a biological sample embedded in a negative stain solution containing heavy metal salts. TEM images are recorded on film or a charged-couple device (CCD) camera. Negative stain EM requires samples to be embedded in heavy metal salts on a carbon support, and as a result, the sample is collapsed at different degrees and frequently assumes a preferred orientation on the grid surface. In contrast, samples prepared by cryo-EM, are suspended in vitreous ice[83] allowing structural determination of complexes in their fully hydrated, near native states. Single-particle analysis EM is based on averaging multiple projections of particles in the same orientation to increase the signal-to-noise ratio of the resulting 2D averages or 3D reconstructions. Cryo-EM 3D reconstructions are calculated by refining hundreds to hundreds of thousands of particle projections against an initial model created using the single-line approach or random conical tilt method with projection-based angular refinement matching[84]. The advantages of using cryo-EM for structural biology investigations include the ability to study large, protein assemblies without the need for crystallization and observing these

complexes at a near native, hydrated state. Although the application of cryo-EM typically results in low to intermediate resolution structures (10-30 Å), its combination with high-resolution data from X-ray crystallography allows modeling of individual components into the 3D maps of large macromolecular assemblies, providing greater structural detail of the complexes.

Classification by reference-free alignment

Classification of 2D EM projections allows the separation of multiple views or classes of a particle dataset. The benefit of classification by reference-free alignment is the elimination of model bias in the resulting classes. In this process, all particles in a dataset are compared to randomly chosen particles from the dataset. These particle projections are then used as models and remaining particles are grouped with the model in which they are most similar, determined by cross-correlation values. Particle images within each group are averaged to calculate new references and the process is continued iteratively. The resulting classes show distinct views or subpopulations of the particle dataset. 2D averaging after classification results in increased signal-to-noise ratio, allowing for clearer visualization of the sample's representative features.

Contrast transfer function

The contrast transfer theory explains that the Fourier transform of an object's

image is related to the Fourier transform of an object's Coulombic potential. To correct for this the Fourier transform of the image must be multiplied by the contrast transfer function of the electron microscope used for imaging (reviewed by Wade[85]). The CTF modulates the amplitude and phases in the back focal plane of the microscope's objective lens. This modulation results in an artifact in the images that is dependent on the microscope's objective lens spherical aberration coefficient, its voltage, the defocus values used and the spatial frequency. CTF is a sinusoidal oscillating function that becomes attenuated with increasing spatial frequency. A CTF curve is plot against the resolution in inverse angstroms. The regions where the CTF passes through zero, no transmittance or contrast is transferred. In theory this first zero pass is the limit of the resolution of an EM 3D reconstruction. This limit can be avoided because the defocus value affects the CTF and images can be obtained at various defocus values. Acquiring micrographs for a given sample at a range of defocus values, shifts the resolution range where information is lost in the dataset[86]. Another approach to regaining information lost at the zero values is by inverting the negative values of the micrograph's CTF to obtain constant positive phase contrast. The simplest form of computational CTF correction, 'phase flipping', is performed by determining the values of the parameters contributing to the shape of the CTF for each micrograph[87]. Using these parameter values, the negative regions of the curve can be flipped to the positive phase. This correction can be

carried out automatically in single-particle processing software such as EMAN[88].

Angular refinement using model-based projection matching

This technique involves generating 2D projections of an initial model ('reprojections') that evenly cover the angular space[87,89]. The particle projections in the experimental dataset are compared with each reprojection, and are matched to the reprojection corresponding to the highest cross-correlation (CC) value. The particles matching a given reprojection are added to a specific class, and this matching correlates with the shift and rotation angle of the projection. Using these parameters, a new reconstruction is calculated and used as an initial model in subsequent iterations. Iterations are refined until the Fourier shell correlation (FSC) curve converges. The FSC measures the normalized cross-correlation values between corresponding densities in two 3D volumes in Fourier space and at each spatial frequency. Although preferred orientation bias of particles is decreased in cryo-EM, relative to negative stain EM, it is still not fully eliminated. 2D averages from particle images with preferred orientations will contain a relatively higher number of particle projections, increasing the signal-to-noise ratio for that class and potentially increasing the resolution in regions corresponding to preferred orientations. To determine the overall resolution of a final 3D map, the projection dataset is split randomly into two subpopulations and two 3D reconstructions are calculated. The value at which the FSC value for the

two reconstructions is equal to 0.5 (FSC=0.5)[90] is typically the value that is cited as the reconstruction's resolution in the literature.

Multiple-reference angular refinement

Multiple-reference angular refinement involves the quantitative separation of datasets that are heterogeneous in conformation and/or composition followed by angular refinement of the resulting subsets[91,92]. In this process, two or more initial models with expected similarities to the subsets are filtered to similar resolutions to prevent a bias toward the higher resolution model. Particle projections are aligned with all the angular reprojections of each initial model and assigned to a single model's class according to highest cross-correlation value. The first iteration is usually sufficient to separate the particles into the correct subpopulations. Subsequently, the subpopulations can be iteratively refined using a single model to generate final 3D maps. Differences in the 3D maps from each subset can be compared to determine the location of missing components or different conformations in the complex.

Fitting of X-Ray crystal structures into EM densities

With the majority of cryo-EM 3D maps at low to intermediate resolution, docking of crystal structures of individual components into the volume of macromolecule's 3D map allows a more detailed structural characterization of the complex. Rigid

body docking of a crystal structure to generate a reasonable model can be used when the component is similar structure and conformation in both the X-ray and EM experiments. If these differences are large, a reliable model may require optimization of the structure(s) docked into the cryo-EM density. Solvent explicit molecular dynamics flexible fitting (MDFF) can be used to improve the fit of initial model (for example the rigid-body docked crystal structure) by applying a force field to the model's atoms with proportional external forces corresponding to the map's density gradient (its Coloumbic potential). This procedure has been used to generate multiple models of different states of the eukaryotic ribosome from intermediate resolution cryo-EM 3D maps[12,13,93].

1.10 Discussion

To understand the process of assembling the 40S ribosomal subunit, necessitates a detailed structural characterization of maturation intermediates, including the late-cytoplasmic precursor that will be described in this thesis. Similarly, localization of the specific binding sites of AFs on the complex will allow understanding of their upstream/downstream signals, substrates and functions. Cryo-EM investigation of the precursor yeast 40S subunit structures is invaluable understanding the structural biology of these large, heterogeneous and flexible complexes that are difficult or impossible to study by X-ray crystallography techniques. Docking of available atomic resolution crystal structures of the

complexes individual components (rRNA, rps and AFs) allows more detailed information about their interactions and conformations on premature 40S subunits. The fit of the model can be increased with MDFF to fit the crystal structures into the 3D map, which can lead to an increase in the level of understanding of the pre-40S 3D structure. The pre-40S model will also reveal differences in the premature and mature 40S, providing clues to how precursor 40S are prevented from incorporation into 80S complexes. In this body of work I will show the cryo-EM 3D structure of a late-cytoplasmic pre-40S complex, as well as the structure of four pre-40S complexes lacking one or more AFs. These data, combined with immuno-labeling experiments and published biochemical and interaction data, allowed the localization all seven AFs associated at this stage. To generate a model of the overall pre-40S complex, the crystal structure of the mature 40S subunit was docked into our pre-40S map density and the fit was optimized with MDFF. AFs with available full or partial crystal structures were docked into the 3D map as rigid bodies according to localization results. The findings highlight structural differences between the premature and mature 40S, reveal the specific interactions of AFs with 40S components, and ultimately shed more light to the function AFs during pre-40S maturation. This data presented in this thesis provides an overall model of a late-stage cytoplasmic 40S precursor, as well as reveals a novel function for assembly factors in the process of 40S maturation.

1.11 References

1 Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. Comparative anatomy of 16-S-like ribosomal RNA. *Progress in nucleic acid research and molecular biology* **32**, 155-216 (1985).

2 Xu, Z., O'Farrell, H. C., Rife, J. P. & Culver, G. M. A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nat Struct Mol Biol* **15**, 534-536 (2008).

3 Wool, I. G. Extraribosomal functions of ribosomal proteins. *Trends in biochemical sciences* **21**, 164-165 (1996).

4 Leger-Silvestre, I. *et al.* Specific Role for Yeast Homologs of the Diamond Blackfan Anemia-associated Rps19 Protein in Ribosome Synthesis. *J Biol Chem* **280**, 38177-38185 (2005).

5 Kapp, L. D. & Lorsch, J. R. The molecular mechanics of eukaryotic translation. *Annual review of biochemistry* **73**, 657-704, doi:10.1146/annurev.biochem.73.030403.080419 (2004).

6 Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745, doi:10.1016/j.cell.2009.01.042 (2009).

7 Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-920 (2000).

8 Schluenzen, F. *et al.* Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* **102**, 615-623 (2000).

9 Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327-339, doi:10.1038/35030006 (2000).

10 Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233-241 (2001).

11 Gomez-Lorenzo, M. G. *et al.* Three-dimensional cryo-electron microscopy localization of EF2 in the Saccharomyces cerevisiae 80S ribosome at 17.5 A resolution. *The EMBO journal* **19**, 2710-2718, doi:10.1093/emboj/19.11.2710 (2000).

12 Taylor, D. J. *et al.* Comprehensive molecular structure of the eukaryotic ribosome. *Structure* **17**, 1591-1604 (2009).

13 Armache, J. P. *et al.* Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. *Proc Natl Acad Sci U S A* **107**, 19754-19759, doi:1010005107,10.1073/pnas.1010005107 (2010).

14 Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203-1209, doi:330/6008/1203, 10.1126/science.1194294 (2010).

15 Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730-736, doi:science.1198308, 10.1126/science.1198308 (2010).
16 Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 A resolution. *Science* **292**, 883-896 (2001).
17 Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A. & Nogales, E. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310**, 1513-1515, doi:10.1126/science.1118977 (2005).
18 Passmore, L. A. *et al.* The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Molecular cell* **26**, 41-50, doi:10.1016/j.molcel.2007.03.018 (2007).
19 Mitra, K. *et al.* Structure of the E. coli protein-conducting channel bound to a translating ribosome. *Nature* **438**, 318-324, doi:10.1038/nature04133 (2005).
20 Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-1942, doi:10.1126/science.1131127 (2006).
21 Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**, 1065-1077, doi:10.1016/j.cell.2006.08.032 (2006).
22 Moazed, D. & Noller, H. F. Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16 S rRNA. *Journal of molecular biology* **211**, 135-145, doi:10.1016/0022-2836(90)90016-F (1990).
23 Dontsova, O. *et al.* Three widely separated positions in the 16S RNA lie in or close to the ribosomal decoding region; a site-directed cross-linking study with mRNA analogues. *The EMBO journal* **11**, 3105-3116 (1992).
24 Algire, M. A. *et al.* Development and characterization of a reconstituted yeast translation initiation system. *RNA* **8**, 382-397 (2002).
25 Algire, M. A., Maag, D. & Lorsch, J. R. Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Molecular cell* **20**, 251-262, doi:10.1016/j.molcel.2005.09.008 (2005).
26 Maag, D., Algire, M. A. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *Journal of molecular biology* **356**, 724-737, doi:10.1016/j.jmb.2005.11.083 (2006).
27 Das, S., Ghosh, R. & Maitra, U. Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *The Journal of biological chemistry* **276**, 6720-6726, doi:10.1074/jbc.M008863200 (2001).

28 Paulin, F. E., Campbell, L. E., O'Brien, K., Loughlin, J. & Proud, C. G. Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Current biology : CB* **11**, 55-59 (2001).

29 Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes & development* **16**, 2906-2922, doi:10.1101/gad.1020902 (2002).

30 Pestova, T. V., Borukhov, S. I. & Hellen, C. U. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**, 854-859, doi:10.1038/29703 (1998).

31 Pestova, T. V. *et al.* Molecular mechanisms of translation initiation in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7029-7036, doi:10.1073/pnas.111145798 (2001).

32 Maag, D., Fekete, C. A., Gryczynski, Z. & Lorsch, J. R. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Molecular cell* **17**, 265-275, doi:10.1016/j.molcel.2004.11.051 (2005).

33 Udem, S. A. & Warner, J. R. Ribosomal RNA synthesis in Saccharomyces cerevisiae. *Journal of molecular biology* **65**, 227-242 (1972).

34 Trapman, J. & Planta, R. J. Detailed analysis of the ribosomal RNA synthesis in yeast. *Biochimica et biophysica acta* **414**, 115-125 (1975).

35 Trapman, J., De Jonge, P. & Planta, R. J. On the biosynthesis of 5.8 S ribosomal RNA in yeast. *FEBS letters* **57**, 26-30 (1975).

36 Grandi, P. *et al.* 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Molecular cell* **10**, 105-115 (2002).

37 Ferreira-Cerca, S., Poll, G., Gleizes, P. E., Tschochner, H. & Milkereit, P. Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Molecular cell* **20**, 263-275, doi:10.1016/j.molcel.2005.09.005 (2005).

38 Fromont-Racine, M., Senger, B., Saveanu, C. & Fasiolo, F. Ribosome assembly in eukaryotes. *Gene* **313**, 17-42 (2003).

39 Stevens, A., Hsu, C. L., Isham, K. R. & Larimer, F. W. Fragments of the internal transcribed spacer 1 of pre-rRNA accumulate in Saccharomyces cerevisiae lacking 5'----3' exoribonuclease 1. *Journal of bacteriology* **173**, 7024-7028 (1991).

40 Moy, T. I. & Silver, P. A. Nuclear export of the small ribosomal subunit requires the ran-GTPase cycle and certain nucleoporins. *Genes & development* **13**, 2118-2133 (1999).

41 Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *Embo J* **22**, 1370-1380 (2003).
42 Strunk, B. S. & Karbstein, K. Powering through ribosome assembly. *Rna* **15**, 2083-2104 (2009).
43 Venema, J. & Tollervey, D. Ribosome synthesis in Saccharomyces cerevisiae. *Annual review of genetics* **33**, 261-311, doi:10.1146/annurev.genet.33.1.261 (1999).
44 Fatica, A., Tollervey, D. & Dlakic, M. PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA* **10**, 1698-1701, doi:10.1261/rna.7123504 (2004).
45 Seiser, R. M. *et al.* Ltv1 is required for efficient nuclear export of the ribosomal small subunit in Saccharomyces cerevisiae. *Genetics* **174**, 679-691 (2006).
46 Schafer, T. *et al.* Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441**, 651-655 (2006).
47 Vanrobays, E., Gelugne, J. P., Gleizes, P. E. & Caizergues-Ferrer, M. Late cytoplasmic maturation of the small ribosomal subunit requires RIO proteins in Saccharomyces cerevisiae. *Mol Cell Biol* **23**, 2083-2095 (2003).
48 Gelperin, D., Horton, L., Beckman, J., Hensold, J. & Lemmon, S. K. Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. *Rna* **7**, 1268-1283 (2001).
49 Lafontaine, D., Delcour, J., Glasser, A. L., Desgres, J. & Vandenhaute, J. The DIM1 gene responsible for the conserved m6(2)Am6(2)A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast. *J Mol Biol* **241**, 492-497 (1994).
50 Fatica, A., Oeffinger, M., Dlakic, M. & Tollervey, D. Nob1p is required for cleavage of the 3' end of 18S rRNA. *Molecular and cellular biology* **23**, 1798-1807 (2003).
51 Lamanna, A. C. & Karbstein, K. Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proc Natl Acad Sci U S A* **106**, 14259-14264 (2009).
52 Draptchinskaia, N. *et al.* The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nature genetics* **21**, 169-175, doi:10.1038/5951 (1999).
53 Galili, N., Qasim, S. A. & Raza, A. Defective ribosome biogenesis in myelodysplastic syndromes. *Haematologica* **94**, 1336-1338, doi:10.3324/haematol.2009.012021 (2009).

54 Liu, J. M. & Ellis, S. R. Ribosomes and marrow failure: coincidental association or molecular paradigm? *Blood* **107**, 4583-4588, doi:10.1182/blood-2005-12-4831 (2006).
55 Campbell, M. G. & Karbstein, K. Protein-Protein Interactions within Late Pre-40S Ribosomes. *PLoS One* **6**, e16194, doi:10.1371/journal.pone.0016194.
56 Granneman, S., Petfalski, E., Swiatkowska, A. & Tollervey, D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *Embo J* **29**, 2026-2036.
57 Arcus, V. L., Backbro, K., Roos, A., Daniel, E. L. & Baker, E. N. Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. *The Journal of biological chemistry* **279**, 16471-16478, doi:10.1074/jbc.M313833200 (2004).
58 Levin, I. *et al.* Crystal structure of a PIN (PilT N-terminus) domain (AF0591) from Archaeoglobus fulgidus at 1.90 A resolution. *Proteins* **56**, 404-408, doi:10.1002/prot.20090 (2004).
59 Glavan, F., Behm-Ansmant, I., Izaurralde, E. & Conti, E. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *The EMBO journal* **25**, 5117-5125, doi:10.1038/sj.emboj.7601377 (2006).
60 Jakovljevic, J. *et al.* The carboxy-terminal extension of yeast ribosomal protein S14 is necessary for maturation of 43S preribosomes. *Molecular cell* **14**, 331-342 (2004).
61 LaRonde-LeBlanc, N. & Wlodawer, A. The RIO kinases: an atypical protein kinase family required for ribosome biogenesis and cell cycle progression. *Biochim Biophys Acta* **1754**, 14-24, doi:S1570-9639(05)00298-0 , 10.1016/j.bbapap.2005.07.037 (2005).
62 Laronde-Leblanc, N., Guszczynski, T., Copeland, T. & Wlodawer, A. Structure and activity of the atypical serine kinase Rio1. *FEBS J* **272**, 3698-3713, doi:EJB4796, 10.1111/j.1742-4658.2005.04796.x (2005).
63 Zemp, I. *et al.* Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *The Journal of cell biology* **185**, 1167-1180, doi:10.1083/jcb.200904048 (2009).
64 Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C. & Raue, H. A. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in Saccharomyces cerevisiae. *J Biol Chem* **278**, 22537-22545, doi:10.1074/jbc.M300759200 (2003).
65 Vanrobays, E., Gelugne, J. P., Gleizes, P. E. & Caizergues-Ferrer, M. Late cytoplasmic maturation of the small ribosomal subunit requires RIO proteins in Saccharomyces cerevisiae. *Molecular and cellular biology* **23**, 2083-2095 (2003).

66 Vanrobays, E. *et al.* Processing of 20S pre-rRNA to 18S ribosomal RNA in yeast requires Rrp10p, an essential non-ribosomal cytoplasmic protein. *The EMBO journal* **20**, 4204-4213, doi:10.1093/emboj/20.15.4204 (2001).
67 Rouquette, J., Choesmel, V. & Gleizes, P. E. Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *The EMBO journal* **24**, 2862-2872, doi:10.1038/sj.emboj.7600752 (2005).
68 Hanszen, K. J. *Adv. Opt. Electron Microsc.* **4** (1971).
69 Lafontaine, D. L., Preiss, T. & Tollervey, D. Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Molecular and cellular biology* **18**, 2360-2370 (1998).
70 Lafontaine, D., Vandenhaute, J. & Tollervey, D. The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes & development* **9**, 2470-2481 (1995).
71 O'Farrell, H. C., Xu, Z., Culver, G. M. & Rife, J. P. Sequence and structural evolution of the KsgA/Dim1 methyltransferase family. *BMC Res Notes* **1**, 108, doi:1756-0500-1-108, (2008).
72 Dong, J. *et al.* Genetic identification of yeast 18S rRNA residues required for efficient recruitment of initiator tRNA(Met) and AUG selection. *Genes & development* **22**, 2242-2255, doi:10.1101/gad.1696608 (2008).
73 Vanrobays, E., Gelugne, J. P., Caizergues-Ferrer, M. & Lafontaine, D. L. Dim2p, a KH-domain protein required for small ribosomal subunit synthesis. *RNA* **10**, 645-656 (2004).
74 Woolls, H. A., Lamanna, A. C. & Karbstein, K. Roles of Dim2 in ribosome assembly. *J Biol Chem* **286**, 2578-2586, doi:M110.191494.
75 Senapin, S., Clark-Walker, G. D., Chen, X. J., Seraphin, B. & Daugeron, M. C. RRP20, a component of the 90S preribosome, is required for pre-18S rRNA processing in Saccharomyces cerevisiae. *Nucleic acids research* **31**, 2524-2533 (2003).
76 Vanrobays, E. *et al.* TOR regulates the subcellular distribution of DIM2, a KH domain protein required for cotranscriptional ribosome assembly and pre-40S ribosome export. *RNA* **14**, 2061-2073, doi:10.1261/rna.1176708 (2008).
77 Jia, M. Z., Horita, S., Nagata, K. & Tanokura, M. An archaeal Dim2-like protein, aDim2p, forms a ternary complex with a/eIF2 alpha and the 3' end fragment of 16S rRNA. *J Mol Biol* **398**, 774-785.
78 Ho, J. H., Kallstrom, G. & Johnson, A. W. Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *The Journal of cell biology* **151**, 1057-1066 (2000).
79 Fassio, C. A., Schofield, B. J., Seiser, R. M., Johnson, A. W. & Lycan, D. E. Dominant mutations in the late 40S biogenesis factor Ltv1 affect cytoplasmic maturation of the small ribosomal subunit in Saccharomyces

cerevisiae. *Genetics* **185**, 199-209, doi:10.1534/genetics.110.115584 (2010).

80 Chen, W., Bucaria, J., Band, D. A., Sutton, A. & Sternglanz, R. Enp1, a yeast protein associated with U3 and U14 snoRNAs, is required for pre-rRNA processing and 40S subunit synthesis. *Nucleic acids research* **31**, 690-699 (2003).

81 Hughes, J. M. & Ares, M., Jr. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *The EMBO journal* **10**, 4231-4239 (1991).

82 Li, H. D., Zagorski, J. & Fournier, M. J. Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in Saccharomyces cerevisiae. *Molecular and cellular biology* **10**, 1145-1152 (1990).

83 Lepault, J., Booy, F. P. & Dubochet, J. Electron microscopy of frozen biological suspensions. *Journal of microscopy* **129**, 89-102 (1983).

84 Vanheel, M. Angular Reconstitution - a Posteriori Assignment of Projection Directions for 3-D Reconstruction. *Ultramicroscopy* **21**, 111-123 (1987).

85 Wade, R. H. A Brief Look at Imaging and Contrast Transfer. *Ultramicroscopy* **46**, 145-156 (1992).

86 Typke, D., Hegerl, R. & Kleinz, J. Image-Restoration for Biological Objects Using External Tem Control and Electronic-Image Recording. *Ultramicroscopy* **46**, 157-173 (1992).

87 Zhu, J., Penczek, P. A., Schroder, R. & Frank, J. 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, doi:10.1006/jsbi.1997.3845 (1997).

88 Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97 (1999).

89 Fuller, S. D., Butcher, S. J., Cheng, R. H. & Baker, T. S. Three-dimensional reconstruction of icosahedral particles--the uncommon line. *Journal of structural biology* **116**, 48-55, doi:10.1006/jsbi.1996.0009 (1996).

90 Vanheel, M. & Harauz, G. Resolution Criteria for 3-Dimensional Reconstruction. *Optik* **73**, 119-122 (1986).

91 Menetret, J. F. *et al.* Architecture of the ribosome-channel complex derived from native membranes. *Journal of molecular biology* **348**, 445-457, doi:10.1016/j.jmb.2005.02.053 (2005).

92 Brink, J. *et al.* Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure* **12**, 185-191, doi:10.1016/j.str.2004.01.015 (2004).

93 Armache, J. P. *et al.* Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. *Proc Natl Acad Sci U S A* **107**, 19748-19753, doi:1009999107 (2010).

Chapter 2
Localization of Assembly Factors on 40S Subunit Precursor

2.1 Abstract

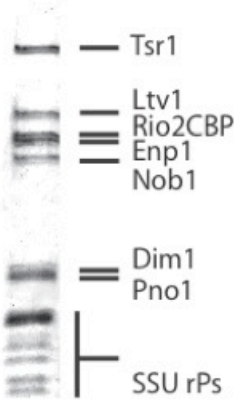
Premature 40S subunits that are erroneously incorporated into 80S complexes are rapidly degraded. The ways by which these premature 40S are normally prevented from joining with 60S is not well understood. The biochemical and structural characterization of pre-40S will elucidate the specific mechanisms preventing subunit fusion. To this end, we have employed cryo-EM to obtain the 3D reconstruction of a yeast late cytoplasmic premature 40S subunit (pre-40S) to a resolution of 18 Å. Substantial differences between premature and mature 40S are present throughout the 60S-binding interface, as well as at the beak and head regions. To further characterize the pre-40S structure, we employed molecular dynamics flexible fitting (MDFF) to model the atomic resolution mature 40S structure into our 3D map. The pre-40S model shows an altered conformation of helix 44 that disrupts the decoding site, and a beak more retracted toward the body of the complex, as compared to the conformation of helix 44 in mature 40S subunits (40S). The cryo-EM 3D map of pre-40S contains

several regions of density not present in mature 40S, attributed to the seven assembly factors associated with this precursor.

2.2 Results

Isolation of wild-type Rio2-TAP pre-40S

Late cytoplasmic 40S subunit precursor particles (pre-40S) were isolated from *Saccharomyces cerevisiae* using a tandem affinity purification (TAP) tag on the assembly factor Rio2, which associates with the premature 40S late in the assembly process[1,2]. Purification was carried out essentially as described[1,3] and details are present in Section 2.3. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis confirms the association of seven AFs on the isolated pre-40S; Pno1 (30.3 kilodaltons-kDa), Dim1 (36.0 kDa), Rio2 (49.1 kDa), Nob1 (51.7 kDa), Ltv1 (53.4 kDa), Enp1 (55.1 kDa) and Tsr1 (90.1 kDa) **(Figure 2.1).** Mass spectrometry (MS) results reveal that 30 of the 32 mature 40S rps are present in our preparation of pre-40S particles (**Table 2.1**). The two rps absent were Rps10 and Rps26. Lastly, Northern blot analysis using a probe against this 20S rRNA transcript, confirmed the presence of the 20S precursor in pre-40S (**Figure 2.1**).



AF	Molecular Weight (kDA)
Tsr1	90.1
Enp1	55.1
Ltv1	53.4
Nob1	51.7
Rio2	49.1(+4 with CBP)
Dim1	36.0
Pno1	30.3

Figure 2.1: SDS-PAGE analysis of AFs in pre-40S. Seven assembly factors are identified in the SDS-PAGE analysis of isolated yeast pre-40S; Tsr1, Ltv1, Rio2 with a C-terminal calmodulin binding peptide (CBP) tag, Enp1, Nob1, Dim1 and Pno1. The molecular weights of each AF are listed to the right.

Rps	Peptide Abudance	Rps	Peptide Abundance
1	(38)	15	(10)
2	(14)	16	(14)
3	(37)	17	(14)
4	(26)	18	(15)
6	(14)	19	(9)
7	(29)	20	(7)
8	(9)	21	(4)
9	(16)	22	(14)
10		23	(8)
11	(11)	24	(18)
12	(11)	25	(9)
13	(19)	26	
14	(18)	27	(12)

Table 2.1: Mass spectrometry analysis of rps in pre-40S. The amount of peptides corresponding to each of the mature rps contained in the pre-40S preparation.

Negative stain EM of pre-40S complexes

Isolated pre-40S were examined initially by negative stain EM **(Figure 2.2A)** and particles imaged were subjected to reference-free classification and alignment (Error! Reference source not found.**2B**). The resulting class averages revealed a complex with overall similarity to mature 40S, and the four characteristic ribosomal 40S sub-structures; the head, platform, beak and foot, were present (**Figure 2.2C)**.

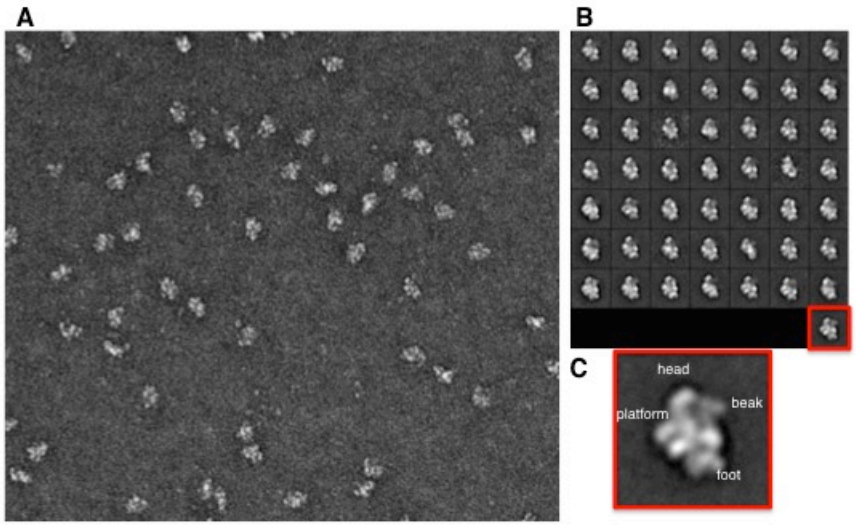


Figure 2.2: Negative stain EM analysis of pre-40S. (A) Raw EM image of pre-40S embedded in negative stain. (B) Class averages of pre-40S particles projections. (C) Zoomed in view of a class average. The characteristic ribosomal substructures are identified.

Cryo-EM 3D structure of pre-40S complexes

To further understand the 3D architecture of the mega-dalton complex in a hydrated, near-native state, as opposed to embedded in negative stain, we employed cryo-EM. Single particle cryo-EM 3D reconstruction generated from WT pre-40S particle projections yielded a final map at a resolution of 18 Å (**Figure 2.3**, **Figure 2.4**). Again, all four characteristic regions of the mature 40S were present, including the beak structure, which was not observed in a previous cryo-EM 3D reconstruction of a similar pre-40S complex[1]. Differences in the

structure of pre-40S and mature 40S are revealed in our 3D reconstruction. First, the beak in pre-40S is retracted closer to the platform, relative to the mature 40S, by a bridge of density connecting the beak to the body. The 3D reconstruction also reveals a kink in helix 44. This kink, not present in the mature 40S, causes the region of the helix near the platform to move away from the complex. Densities not attributed to rRNA or rps are found at the platform, 60S binding interface, head, and the bridge of density connecting the beak to the platform.

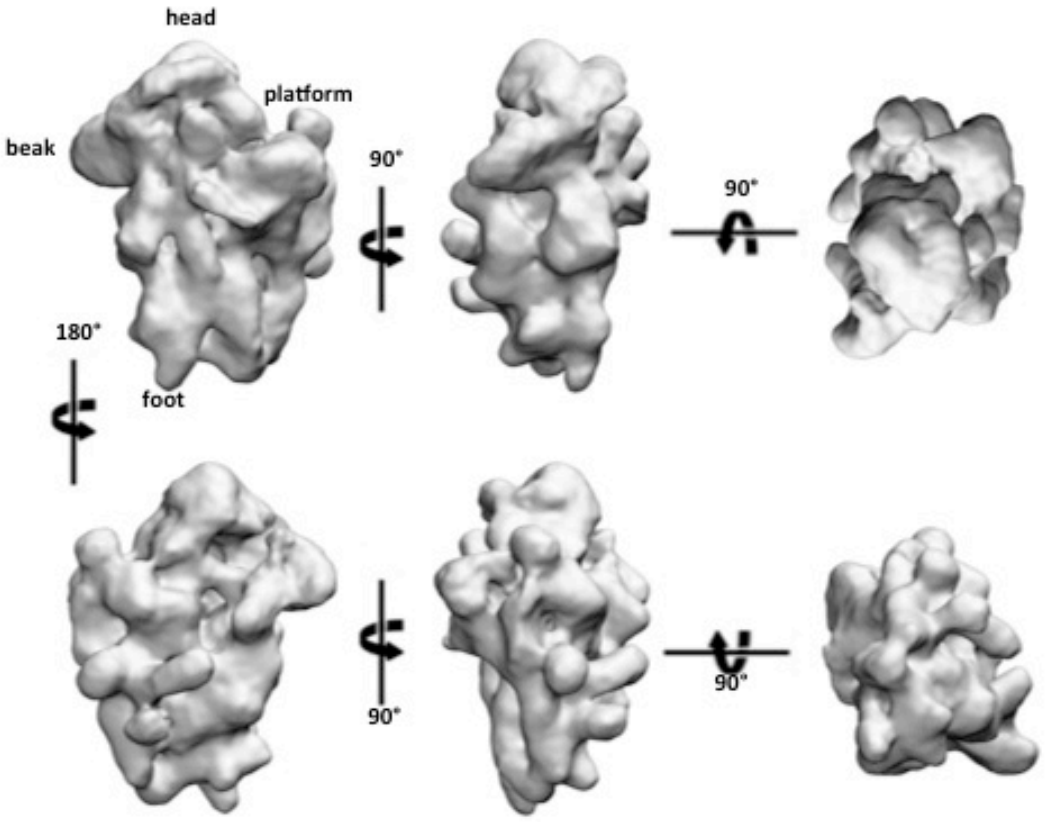


Figure 2.3: Pre-40S cryo-EM 3D reconstruction. 6 views of the yeast pre-40S cryo-EM 3D map calculated to a resolution of 18.2 Å (see **Figure 2.3**D), similar to negative

stain EM class averages, the four characteristic 40S ribosomal subregions are seen and labeled in the first view.

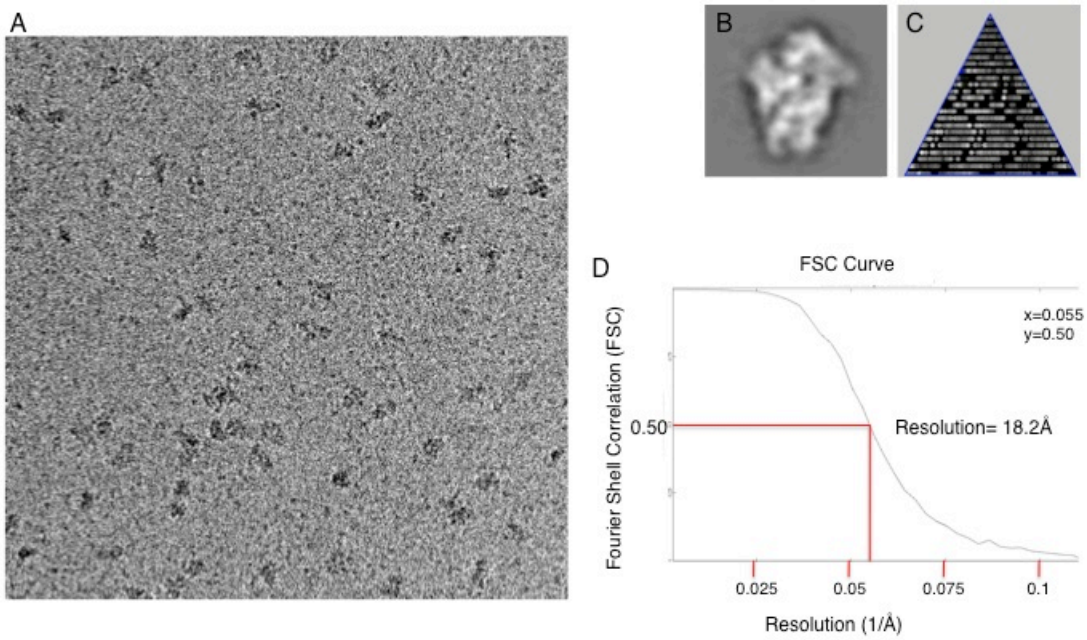


Figure 2.4: Data and analysis of the pre-40S cryo-EM 3D reconstruction. (A) A raw cryo-EM image of yeast pre-40S particles. (B) Cross-section of the final 3D map of pre-40S. (C) Angular distribution profile for particle projections used in the pre-40S angular refinement. The angular distribution profile allows visualization of the amount of particle projections in each class used for the 3D reconstruction; each dot corresponds to a class and the brightness of each dot is proportional to the number of particles in that class that were used for the final 3D reconstruction. (D) Fourier Shell Correlation (FSC) curve for the Rio2-TAP pre-40S, indicated a final resolution of our 3D map at 18.2Å (FSC=0.5).

Modeling of ribosomal proteins and RNA in pre-40S

To obtain more detailed structural information and compare our pre-40S 3D structure with that of the mature 40S, the 4.15 Å crystal structure from *T. thermophila*[4] was docked into the pre-40S 3D map as a rigid body **(Figure 2.5A).**

Most regions of the mature structure fit well in the pre-40S volume with two major exceptions, the beak and helix 44 **(Figure 2.5)**. The shift of the beak in pre-40S causes the rRNA comprising the beak, helix 34, to be outside the map's density. The top of helix 44 also extends away from its density in mature 40S, due to the kink present in middle of the premature helix. To improve the fit of the model into our pre-40S 3D map, solvent explicit molecular dynamics flexible fitting (MDFF) was applied (for detailed explanation of this procedure, see Section 2.3). The MDFF protocol, which was extensively tested (testing is explained in Section 2.3), significantly improved the fit of the rRNA and rps, as indicated by the increase in the cross-correlation values before and after MDFF, and after fitting, the model's atoms were completely enveloped by the density **(Figure 2.5A)**. The pre-40S model shows the rRNA adjacent to the decoding site, located at the top of helix 44, and the rRNA loop between helix 44 and helix 45, becomes single-stranded. The distance between the active site nucleotides in the decoding site is increased, potentially preventing the formation of an active decoding site **(Figure 2.5B)**. Rps and rRNA in the head also undergo shifts after MDFF, but the majority of conformational changes in the remainder of the complex are subtle, as expected by the good overall fit in the initial rigid body docked model (left panel **Figure 2.5**A). In addition to localization of the rps and rRNA in pre-40S and the identification of differences in premature versus mature 40S, the pre-40S

model also revealed the areas of density attributed to the bound assembly factors.

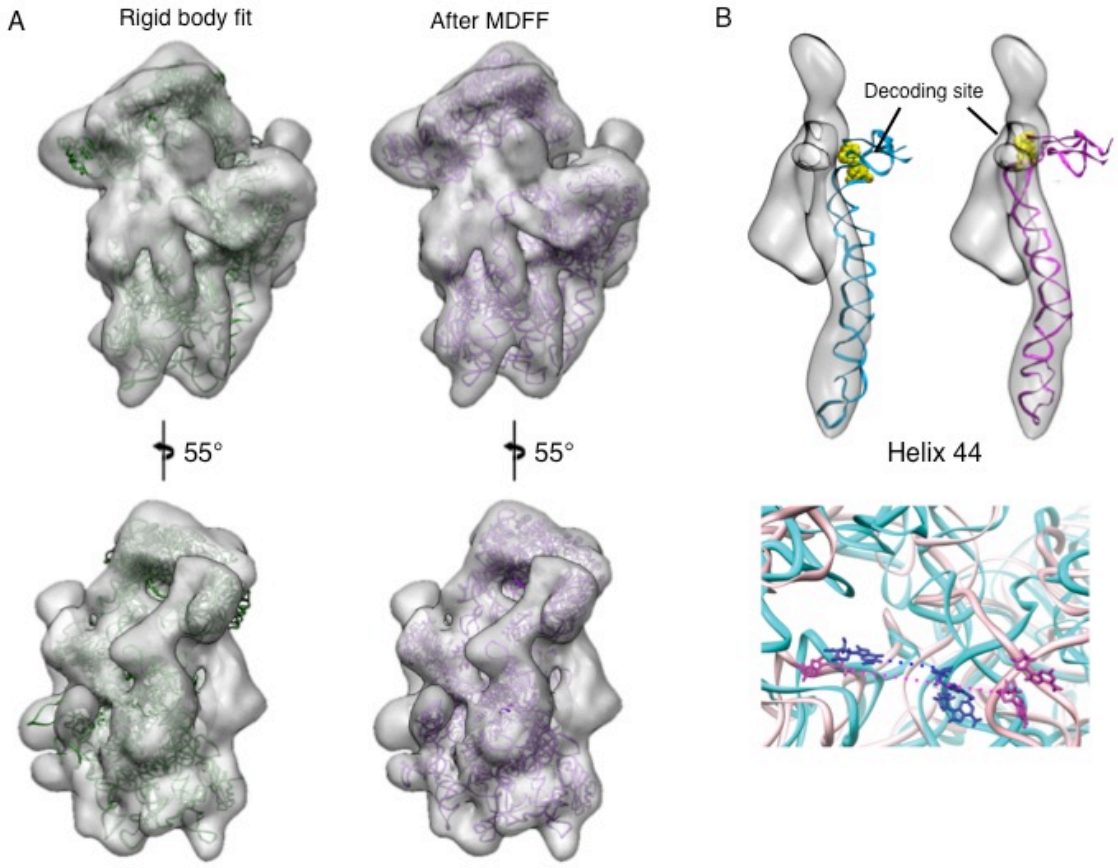


Figure 2.5: Fitting of rps and rRNA into our pre-40S 3D map. (A) Docking of the mature 40S model as a rigid body (left, green) and optimization of the fit using MDFF (right, magenta) into our pre-40S 3D map. The top view shows the solvent side and the movement of helix 44 into the map. The bottom view shows the improvement of the rRNA's fit into the beak region after MDFF. (B) View of the fit helix 44 fit into the map density before (left, cyan) and after MDFF (right, magenta) of the model. The lower panel shows an overlay of pre-MDFF (cyan) and post-MDFF (magenta) models zoomed in at the decoding site, located on the top helix 44. The active site residues of the decoding site are shown in stick representation, showing the increase in distance between the residues in the premature versus mature 40S subunits.

Determination of densities unique to precursor 40S subunits

To determine the protein density on pre-40S (rps and AFs), we subtracted the MDFF model of the rRNA low-pass filtered[5] to 18 Å from our pre-40S 3D map (**Figure 2.6**). Using the data presented by Rabl et al.[4], we localized the density of the rps present in pre-40S and leaving only the densities of AFs associated with the complex (orange in **Figure 2.7**). The AFs were localized to specific areas, which included: two adjacent regions on the back of the platform near site D; two inter-connected perpendicular regions on the platform; a region to the left of helix 44, from the 60S perspective; and densities on the head, and in the bridge connecting the beak to the platform. These densities overlap binding sites on the mature 40S for multiple factors required for initiation of translation, the eukaryotic initation factors: eIF1, eIF1A, and eIF3; the channel in which the mRNA binds; the three tRNA binding sites and the decoding site. The platform densities also overlap regions involved in the bridging of subunits. These data suggests that in addition to their functions in proper 40S maturation, AFs also act to block binding of initiation factors required for 80S complex formation, indicating a mechanism to inhibit the incorporation of premature 40S into 80S complexes.

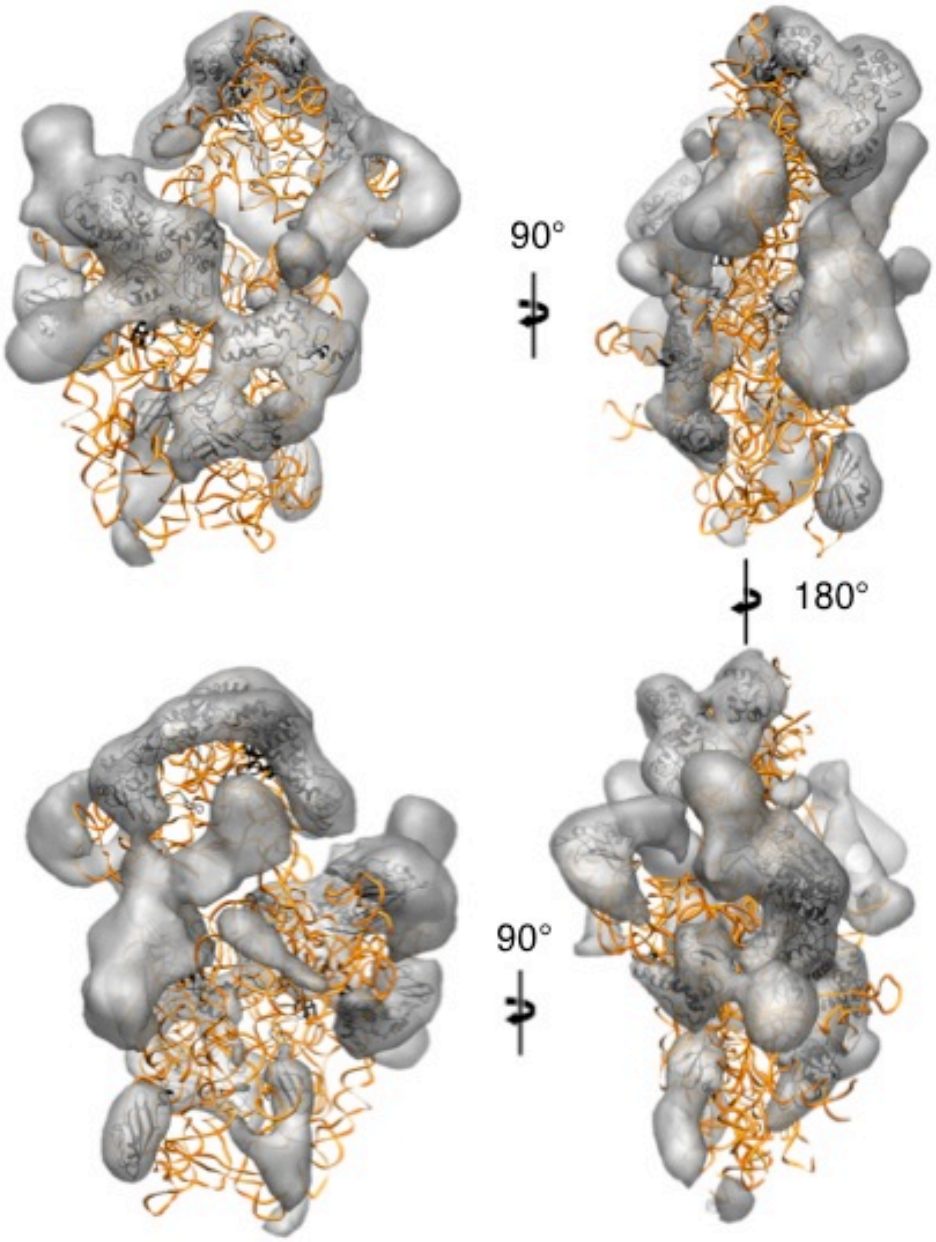


Figure 2.6: Subtraction of the rRNA from the pre-40S density. Subtraction of the MDFF 18S rRNA (orange) filtered to 18 Å from the pre-40S 3D volume reveals the densities attributed to proteins (AFs and rps). The rps present in the model are shown as black ribbons and are contained within the volume density. The remaining densities are attributed to AFs and known rps that were not included in our initial model but are present in the premature particle.

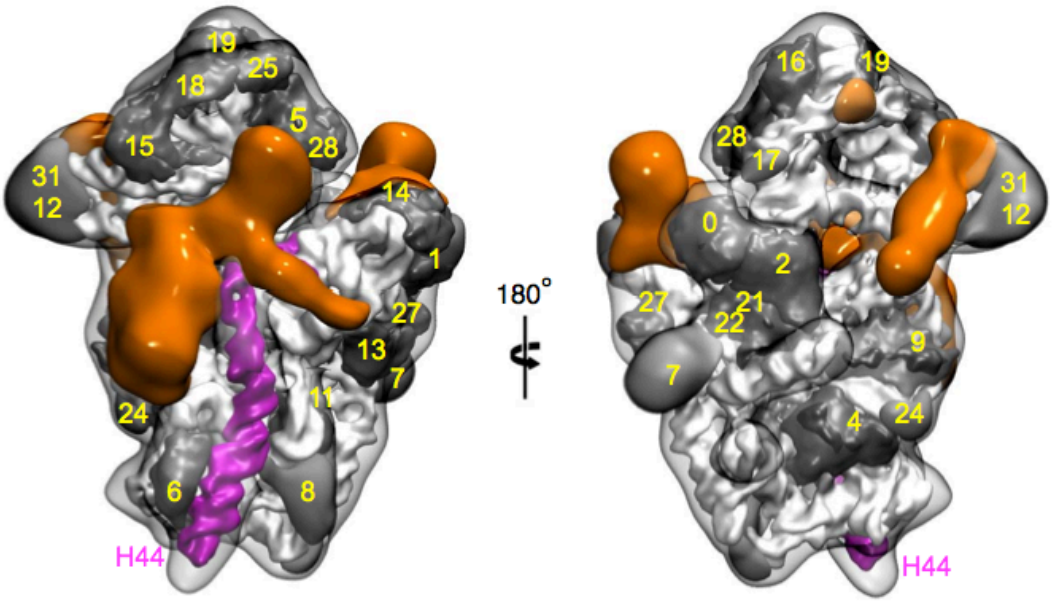


Figure 2.7: Molecular architecture of pre-40S. The rps and rRNA modeled using MDFF of the mature 40S structure are shown in dark gray and white, respectively. The rps are labeled according to the data presented by Rabl et al.[4]. Helices 44 and 45 are colored magenta. The remaining densities are attributed to the associated AFs (orange) and are localized primarily to the density bridging the beak to the platform, the head, platform, the back of the platform and to the left of helix 44 (from the perspective of the 60S subunit).

2.1 Experimental procedures

Purification of pre-40S complexes

Pre-40S complexes were isolated from a *S. cerevisiae* strain containing a tandem affinity purification (TAP) tag on the C-terminus of Rio2, Rio2-TAP (Rio2-TAP strain from Open biosystems). Cells were grown in a 10 L fermenter at 30° C in YP media with 3% dextrose to an optical density at 600 nanometers (OD_{600}) of

1.5. Yeast cells grown in flasks were also grown at 30° C in YP media but with 2% dextrose to an OD_{600} between 0.6 and 0.8. Tandem affinity purifications of pre-40S complexes using TAP-tagged Rio2 as bait were performed essentially as previously described[1,3] **(Figure 2.8)**. Briefly, 20-40 g of yeast cells, grown in the conditions described above, were resuspended in the standard TAP buffer, 100 millimolar (mM) sodium chloride (NaCl), 50 mM tris(hydroxymethyl)aminomethane hydrochloride (TRIS-HCl), 10 mM magnesium chloride ($MgCl_2$) and 0.075% nonyl phenoxypolyethoxylethanol 40 (NP-40) at pH 7.5) and supplemented with DNAse and protease inhibitors. A CryoMill (Retsch) was used to lyse resuspended cells by cryogenic grinding and the lysates were incubated with Immunoglobulin G (IgG) beads to bind the Protein A component of the TAP tag. The unbound impurities were washed from the beads with buffer and the beads were incubated with tobacco etch virus (TEV) protease to remove Protein A by cleaving at the TEV cleavage site, located between Protein A and the calmodulin binding peptide (CBP) in the TAP tag. The eluent was incubated with calmodulin beads to bind the CBP component, and after several washes, complexes were eluted with buffer containing ethylene glycol tetraacetic acid (EGTA) and collected. The eluent was concentrated to a final volume of 20-40 µL using a *Millipore* Biomax-100K NMWL filter device for analysis by EM, SDS-PAGE or MS.

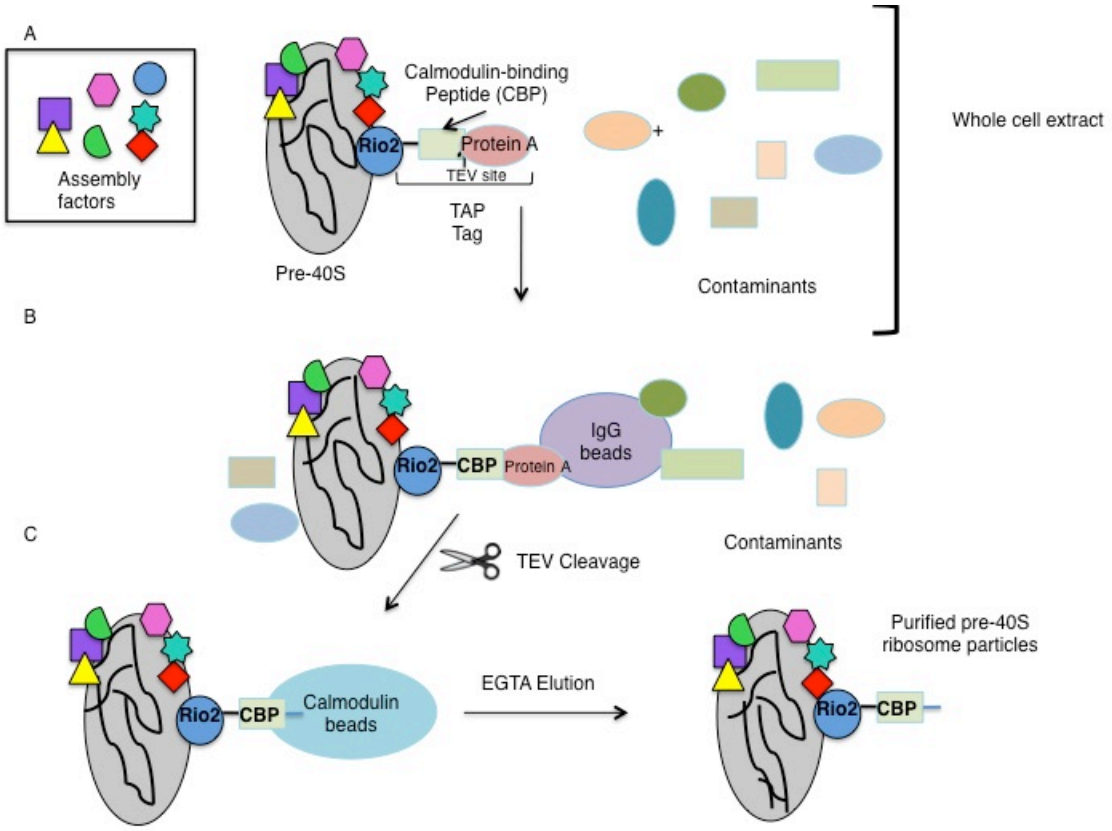


Figure 2.8: Tandem affinity purification of pre-40S. Cell lysates of yeast strains expressing Rio2-TAP are incubated with IgG beads to isolate pre-40S particles containing Rio2-TAP. The Protein A component of the TAP tag is cleaved using TEV protease and the eluent is incubated with calmodulin beads to bind the CBP for a second purification step. Pre-40S are eluted from the beads by washing with buffer containing EGTA.

Mass spectrometry

The proteins associated with the pre-40S complexes were separated using SDS-PAGE, and areas corresponding to proteins of interest were excised from the gel. Gel samples were subjected to proteolytic digestion with trypsin and analyzed

using liquid chromatography (LC) tandem MS (LC/MS/MS). Product ion data were searched against the SwissProt database[6] using the Mascot search engine[7]. Mascot output files were parsed into the Scaffold program for filtering to assess false discovery rates and allow for correct protein identifications. The relative amounts of rps peptides identified in the pre-40S complexes are listed in (**Table 2.1**).

Specimen preparation and EM imaging of pre-40S complexes

3.5 µl of Rio2-TAP isolated pre-40S was applied to glow-discharged continuous carbon grids (Electron Microscopy Sciences). Excess sample was blotted with filter paper and grids were stained with uranyl formate (1%). The samples were imaged at room temperature with a Tecnai T12 electron microscope (FEI) operated at 120 kV using low-dose procedures. Images were recorded at a magnification of 71,138x and a defocus value of approximately -1.5 µm on a Gatan US4000 CCD camera. All images were binned (2 x 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. A total of 3417 pre-40S particle images were manually excised from the micrographs using the Boxer program (EMAN 1.9 software suite[8]). The selected particles were subjected to reference-free alignment and classification into 50 classes using the software SPIDER[9].

Cryo-EM sample preparation and imaging of pre-40S complexes

2 µl of pre-40S sample at an OD_{600} of 10-20 were adsorbed on glow-discharged mesh grid (Quantifoil R2/2 200). The grids were blotted once for 1-2 seconds and plunge frozen in liquid ethane at a temperature lower than -170° C with a Vitrobot (FEI Mark IV) with its temperature maintained at 22° C and humidity greater 85%. Grids were transferred using a Gatan 626 single tilt cryo-transfer system for imagining on a Tecnai F20 transmission electron microscope (FEI) equipped with a field emission electron source operated at 200 kilovolts (kV). The temperature of the grid was maintained throughout transfer and imaging at less than -160° C. Images were recorded at a magnification of 66,964x on a Gatan US4000 CCD camera at defocus values ranging from -1.5 to -4.0 µm. Particles were preferentially located in relatively thick ice and each micrograph contained on average 50-100 particles. The pixel size under these conditions is 2.24 Å at the specimen level.

Single-particle reconstruction of pre-40S complexes

Pre-40S particles from cryo-EM images were manually excised from micrographs using Boxer with a box size of 160 pixels. The CTF parameters were determined for each micrograph using the *ctfit* software and CTF phase correction was applied to each micrograph in the dataset using *Applyctf* (*ctfit* and *Applyctf* are

contained in the EMAN 1.9 package[8]). An *ab initio* model was generated by randomly selecting 100 pre-40S particles from the total dataset, and randomly assigning an orientation value for each projection resulting in a noisy, sphere-like initial reference volume for refinement without introducing model bias (data not shown). The 3D reconstruction of pre-40S was produced using model-based projection matching and iterative refinement ('refine' command in EMAN 1.9) until the FSC curves of successive 3D maps converge. The approximate refinement parameters used for the pre-40S 3D were as follows: an angular spacing between reprojections of 5-7^{o}, a mask with a radius 80 pixels, a maximum phase error for class averages permitted to be incorporated into the reconstruction of 60, symmetry of c1 (no symmetry), and a maximum value of 0.5σ (0.5 times the value of the standard deviation) for individual particles in a given class to be used in the generation of the final 3D reconstruction. 11,604 pre-40S projections were refined to obtain a 3D reconstruction at a resolution of 18 Å. To determination the value of the resolution of the final 3D map, particles were randomly split into two datasets and 3D reconstructions were generated for each subset ('eotest' command in EMAN 1.9) and the value when their FSC = 0.5 is the resolution stated. Prior to visualization, the 3D map was low pass filtered at a resolution of 18 Å with a Gaussian filter using the software EM-Bfactor[10].

MDFF: Initial structure preparation

MDFF simulations were used to improve the fit of mature 40S docked as a rigid body into the pre-40S 3D map. The recently published crystal structure of the 4.15 Å *S. cerevisiae* 80S ribosome from Yusupov and colleagues[11] (PDB ID: 3O2Z) was used to generate the initial model. This crystal structure contains the 18S rRNA nucleotides numbered 1 to 1800 with the exception of 668 to 678 and most ribosomal proteins present in the 80S complex except Rps6, Rps7, Rps8, Rps10, Rps12, Rps21 and Rps26. The coordinates of the following proteins in this crystal structure were used to generate our initial model: Rps0, Rps2, Rps4, Rps5, Rps9, Rps11, Rps13, Rps14, Rps15, Rps16, Rps18, Rps19, Rps22, Rps23, Rps24, Rps25, Rps27, and Rps28. Alternatively, coordinates for Rps5, Rps18, and Rps19 were obtained from the recently published model from a 5.5 Å cryo-EM structure of the yeast 80S ribosome from Beckman and colleagues[12] (PDB ID: 31ZB), and Rps1 and Rps14 were used from the crystallographic coordinates from Ban's group *Tetrahymena thermophila* (thermophilic yeast) mature 40S structure[4] (PDB ID: 2XZM) **(Table 2.2).** The basis for using the coordinates for Rps5, Rps14, Rps18, and Rps19 from these structures, as opposed to those from the structure deposited by Yusupov and colleagues[11], is that each of these rps structures contain a more complete amino acid sequence than those solved by the Yusupov group[11]. This approach is justifiable as the

rps from the Rabl et al.[4] and Armache et al.[12,13] maintain similar secondary structure elements and three-dimensional folds to those present in the 4.15 Å crystal structure[4]. The coordinates for Rps1 were obtained from the 3.9 Å 40S crystal structure[4] because this protein is not identified in the Rabl et al. structure. Rps3, Rps9 (residues 6-21), Rps20, Rps29, RACK1, and P/E-tRNA were omitted from this initial structure on the basis of having zero (Rps20, RACK, P/E-tRNA) or decreased (Rps3 and Rps2) occupancies or no corresponding density in our pre-40S 3D map, presumably due to a shift (Rps9) of the protein in the premature versus mature 40S complexes. The missing rRNA nucleotides from 668 to 678 were added to the model, as well as all hydrogens and other missing atoms. The initial model was solvated, contained in a simulation box, neutralized and ionic concentration was maintained using the Visual Molecular Dynamics (VMD) program[14].

rps	PDB	Omitted
Rps0	3O2Z	
Rps1	2XZM	
Rps2	3O2Z	
Rps3		X
Rps4	3O2Z	
Rps5	3IZB	
Rps6		
Rps7		
Rps8		
Rps9	3O2Z	
Rps10		
Rps11	3O2Z	
Rps12		
Rps13	3O2Z	
Rps14	3IZB	
Rps15	3O2Z	
Rps16	3O2Z	
Rps17		X
Rps18	3IZB	
Rps19	3IZB	
Rps20		X
Rps21		
Rps22	3O2Z	
Rps23	3O2Z	
Rps24	3O2Z	
Rps25	3O2Z	
Rps26		
Rps27	3O2Z	
Rps28	3O2Z	
Rps29		X
Rps30		X
Rps31		X

Table 2.2: Rps structures used in our pre-40S model. As described in the text, rps coordinates were used from structures deposited by Rabl et al.[4] (PDB ID: 3O2Z0, Ben-Shem et al.[11] (PDB ID: 2XZM) and Armache et. al[12,13] (PDB ID: 31ZB). The origin of the

PDB file for each rps in the model is listed, as well as denoting those omitted. The remaining rps were not identified in the Rabl et al. structure[4].

Prior to publication of the 5.5 Å 80S cryo-EM model[12,13], the 4.15 Å 80S[11] and the 3.9 Å 40S[4] crystal structures, the highest resolution model for the structure of the eukaryotic ribosome 40S subunit was from the 8.9 Å cryo-EM 3D reconstruction of the 80S ribosome bound to RACK1 and the P/E-tRNA from the thermophilic yeast *Thermomyces lanuginosus*[15,16] (PDB IDs: 3JYV and 3JYW, 40S and 60S coordinates, respectively). The sequence of the *T. lanuginosus* ribosomal components shares greater than 85% identity with *S. cerevisiae*[17]. All the initial validations of the MDFF protocol were carried out with the coordinates for the 40S subunit from this model. It contains the mature 18S rRNA and 17 of the 32 SSU rps. The Rps included in 3JYV are: Rps0, Rps2, Rps3, Rps5, Rps9, Rps11, Rps13, Rps14, Rps15, Rps16, Rps18, Rps19, Rps20, Rps22, Rps23 and Rps29. On the basis described for the 3O2Z initial model, Rps3, Rps9 (residues 6-21), Rps20, and Rps29, as well as the mature 40S protein RACK1, and P/E-tRNA, were excluded from this model. The complete sequence of the *S. cerevisiae* 18S rRNA (containing 1800 residues) was obtained from Gen Bank (Z75578.1) and aligned to the *T. Thermophila* 18S rRNA sequence with ClustalW[18]. Based on the alignment, the following changes were made to rRNA sequence of the model (chain A in 3JYV): the nucleotides previously at positions 898, 1041, 1042, 1182,

1287, and 1444, are now at 899, 1042, 1043, 1183, 1286, and 1443, respectively (the shift in position of these residues is concurrent with remaining nucleotides), and the sequence GUGUC (884, 1001, 1003, 1284 and 1285) was substituted to AAACU. The Cartesian coordinates were adjusted accordingly, and, in the case of substitutions, the nucleotide type was corrected. The system was then solvated and ionized using the same protocol described above.

MDFF: Initial equilibration phase

The initial equilibration phase for both models of the 40S subunit (3O2Z and 3JYV) included an energy-minimization of the final systems for 5,000-10,000 cycles of conjugate-gradient optimization followed by a short 200 ps molecular dynamics (MD) simulation in the NVT (moles, volume, temperature) ensemble. The temperature was held constant at 310 K using the Langevin thermostat with a damping coefficient of 5 ps^{-1}, and pressure was held constant using the Nose-Hoover barostat. A 2 fs time-step with bonds held rigid was used in the initial run. Non-bonded interactions were cut-off beyond 12 Å with smooth switching taking effect at 10 Å. Long-range electrostatic interactions were handled using the particle mesh Ewald (PME) method[19]. All MD trajectories were generated with NAMDv2.7[20] and the CHARMM force field[21,22]. The system was created and

visualized using VMDv1.8.7[14], and VMDv1.8.8.a17 was used to generate the input files used in MDFF simulations (see below).

MDFF simulation details

The protocol of Trabuco et al. was followed for MDFF simulations[23-25]. The increase in the cross-correlation value before and after flexible fitting reveals that MDFF significantly improves the fit of the model**,** and multiple repeats of the same simulation give the same result **(Figure 2.10)**, this consistency allows confidence in the MDFF protocol used. To further validate the MDFF results, a procedure was repeated with different initial structures (PDB ID: 3O2Z and 3JYV) fit into the pre-40S 3D map (**Figure 2.9**, **Figure 2.10).** The MDFF protocol is as follows: first, the initial atomic-model is docked as a rigid body into the pre-40S 3D map using the software SITUS[26]. Next, the rRNA undergoes MDFF using a scaling factor of 0.5 kilocalorie (kcal)/mol (a value of 0.3 kcal/mol results in forces on the order of 10-15 pN per atom for a carbon atom), a tolerated value as the MDFF experiments were carried out in explicit solvent[25,27]. The angles or distances in the rRNA secondary structure were restrained with force constant values of 400 kcal mol^{-1} rad^{-2} or 400 kcal mol^{-1} $Å^{-2}$. In this step, the molecular bonds in the protein structures were restrained with a force constant of 400 kcal mol^{-1} rad^{-2} (secondary-structure elements) and 200 kcal mol^{-1} rad^{-2} (non-

secondary-structure elements), with these values for the secondary structure elements of rRNA and rps maintained during fitting. This step was carried out for approximately 3 nanoseconds repeatedly until the root mean squares deviation (RMSD) converged. Finally, the structures of the rps were refined in the pre-40S 3D map by reducing the restraints of the bonds to a force constant of 200 kcal mol^{-1} rad^{-2}, allowing flexibility of secondary structures. Three separate runs were carried out for the pre-40S 3D map using 3JYV as an initial atomic model **(Figure 2.10)**, and once using 3O2Z as an initial atomic model **(Figure 2.9).** The CC values at each step of MDFF are listed in **Table 2.3**.

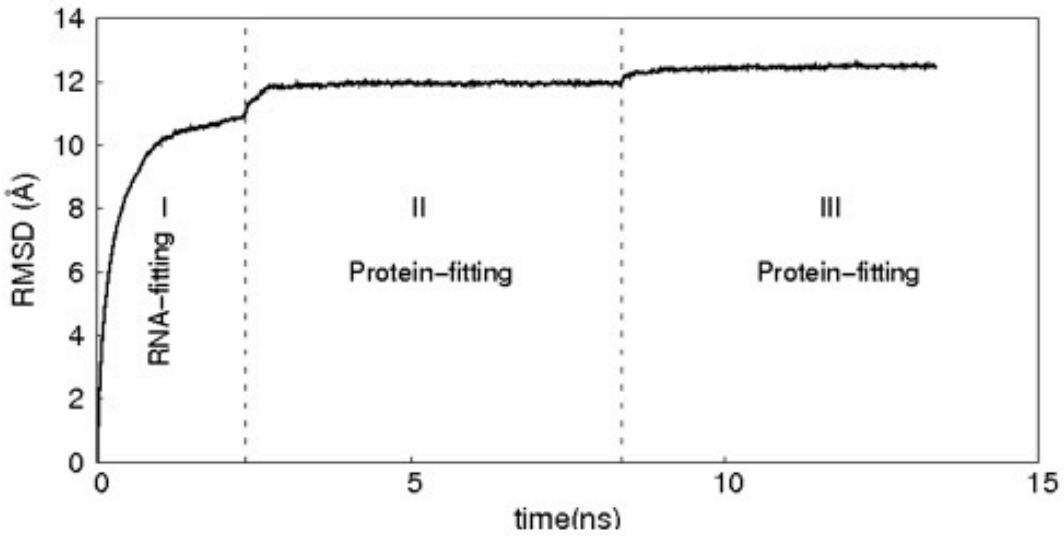


Figure 2.9: MDFF results for the fitting of 3O2Z into the pre-40S 3D map. (A) The RMSD value compared to the rigid body docked model is plot as a function of simulation time (ns). The three stages of the plot correspond to the fitting of the rRNA alone, the fitting of rRNA and rps before and after relaxing constraints on the secondary structure elements.

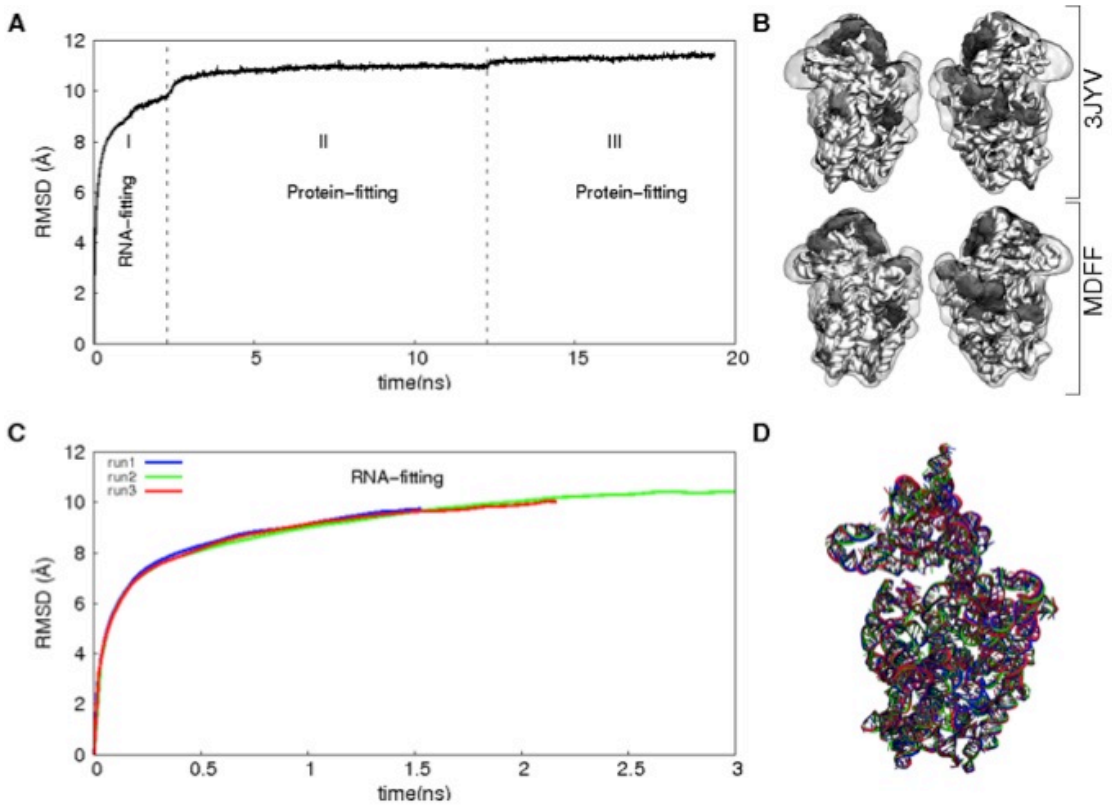


Figure 2.10: MDFF results for the fitting of 3JYV into the pre-40S 3D map. (A) The RMSD value (Å) of the overall model compared to the rigid body docked model is plot as a function of simulation time (ns). The three stages of the plot correspond to the fitting of the rRNA alone, the fitting of rRNA and rps before and after relaxing constraints on the secondary structure elements. (B) Pre- and post-MDFF side-view snapshots of initial atomic model (PDB code 3JYV) positioned into the pre-40S 3D map, rRNA is silver and rps in dark gray. (C) RMSD value (Å) vs. simulation time (ns) of rRNA fit into the pre-40S 3D map in three independent MDFF runs. (D) Overlay of rRNA structures at the end of three independent MDFF runs. RNA backbones are in the same color as RMSD traces in panel C.

Map	PDB	Res. (Å)	CCC			
			Initial	Step1	Step2	step3
Pre-40S	3JYV	~18	0.672	0.781	0.786	0.787
Pre-40S	3O2Z	~18	0.673	0.795	0.802	0.805

Table 2.3: Cross-correlation (CCC) values for MDFF runs. CCC values for the fitting of 3O2Z and 3JYV into the pre-40S 3D map for each step of MDFF.

Difference mapping

The flexibly fitted rRNA was simulated to a density map, filtered to resolution of 18 Å, and subtracted from the pre-40S 3D volume using the SITUS software[26]. The resulting densities are those attributed to only proteins (rps and AFs) in the pre-40S complex.

2.2 Discussion

Investigation of the 18 Å pre-40S 3D structure reveals conformational differences and additional density in the premature 40S relative to the mature. The differences in their conformations are more apparent in the pre-40S model generated by fitting the mature 40S structure into the pre-40S 3D map using MDFF. The structure of helix 44, which contains the nucleotides of the decoding site, is disrupted in pre-40S, resulting in a greater distance between the active site nucleotides (**Figure 2.5C**). Helix 44 is located at the region in which the small

subunit binds the large, near structural elements necessary for bridging the two subunits. The other prominent difference in the premature 40S is its beak structure. The beak is retracted toward the platform in pre-40S, altering the surface of the subunit binding interface and potentially the structure of the mRNA channel.

To determine which densities are attributed to AFs, difference mapping allowed the removal of components present in the mature 40S from the pre-40S 3D map. The densities unique to the premature complex correspond to the seven associated assembly factors, which are located on the platform, at the back of the platform, the bridge and the head regions (**Figure 2.6**). These densities overlap the binding sites for the initiation factors eIF1, eIF1A, eIF2 and eIF3. The mRNA channel and locations of subunit bridging are also blocked. These structural observations reveal that multiple mechanisms are present to prevent the joining of precursor 40S subunits with 60S. To determine which AFs are responsible for blocking the binding of the various factors required for translation initiation and to allow a greater understanding of the AF's functions in 40S biogenesis, each were localized as described in the following chapter.

2.3 Concluding remarks

Cryo-EM is a powerful technique to study ribosomal precursor particles, as they are relatively large, flexible and preparations of these complexes can be heterogeneous in composition. Determining the structure of this late cytoplasmic 40S precursor to a moderate resolution and using MDFF to fit the atomic-resolution mature 40S crystal structure into the 3D map allows characterization of the structural features of pre-40S. This analysis of the pre-40S structure reveals prominent structural differences relative to mature 40S and areas of unique density. These data will contribute significant knowledge to understanding ribosome biogenesis and the mechanism by which pre-40S subunits are prevented from incorporation into 80S complexes.

2.4 Acknowledgements

Data from this chapter is contained in the manuscript "Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates" which, as of 5/13/2011, is in review in for publication in *Science.* Bethany Strunk isolated pre-40S particles and performed SDS-PAGE analysis on the particles (Katrin Karbstein's lab, Scripps Florida). Justin Schilling (Georgios Skiniotis's lab) analyzed the pre-40S complexes with negative stain EM. Sample preparation, image acquisition and processing/analysis of cryo-EM data was done by Cherisse Loucks with guidance and technical expertise from Min Su

(Georgios Skiniotis lab). MDFF modeling and difference mapping was carried out by Harish Vashisth and Shanshan Cheng (Charles Brooks' lab, University of Michigan). Mass spectrometry analysis performed at the Protein Structure Facility at the University of Michigan.

2.5 References

1 Schafer, T. *et al.* Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441**, 651-655, doi:nature04840, 10.1038/nature04840 (2006).

2 Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J* **22**, 1370-1380, doi:10.1093/emboj/cdg121 (2003).

3 Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218-229, doi:10.1006/meth.2001.1183, S1046-2023(01)91183-1 (2001).

4 Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730-736, doi:science.1198308, 10.1126/science.1198308.

5 Zhang, C. M., Perona, J. J., Ryu, K., Francklyn, C. & Hou, Y. M. Distinct kinetic mechanisms of the two classes of Aminoacyl-tRNA synthetases. *Journal of molecular biology* **361**, 300-311, doi:10.1016/j.jmb.2006.06.015 (2006).

6 Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research* **31**, 365-370 (2003).

7 <http://www.matrixscience.com> (1999).

8 Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97, doi:10.1006/jsbi.1999.4174, S1047-8477(99)94174-6 (1999).

9 Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-199, doi:S1047-8477(96)90030-1, 10.1006/jsbi.1996.0030 (1996).

10 Fernandez, J. J., Luque, D., Caston, J. R. & Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of structural biology* **164**, 170-175, doi:10.1016/j.jsb.2008.05.010 (2008).

11 Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203-1209, doi:330/6008/1203, 10.1126/science.1194294.

12 Armache, J. P. *et al.* Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. *Proc Natl Acad Sci U S A* **107**, 19748-19753, doi:1009999107, 10.1073/pnas.1009999107.

13 Armache, J. P. *et al.* Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. *Proc Natl Acad Sci U S A* **107**, 19754-19759, doi:1010005107,10.1073/pnas.1010005107.

14 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38, 27-38 (1996).

15 Taylor, D. J. *et al.* Comprehensive molecular structure of the eukaryotic ribosome. *Structure* **17**, 1591-1604, doi:S0969-2126(09)00416-X, 10.1016/j.str.2009.09.015 (2009).

16 Taylor, D. J. *et al.* Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J* **26**, 2421-2431, doi:7601677, 10.1038/sj.emboj.7601677 (2007).

17 Nilsson, J., Sengupta, J., Gursky, R., Nissen, P. & Frank, J. Comparison of fungal 80 S ribosomes by cryo-EM reveals diversity in structure and conformation of rRNA expansion segments. *J Mol Biol* **369**, 429-438, doi:S0022-2836(07)00377-4, 10.1016/j.jmb.2007.03.035 (2007).

18 Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25**, 4876-4882 (1997).

19 Norberto de Souza, O. & Ornstein, R. L. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *Journal of biomolecular structure & dynamics* **16**, 1205-1218 (1999).

20 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).

21 MacKerell, A. D. Developments in the CHARMM all-atom empirical energy function for biological molecules. *Abstracts of Papers of the American Chemical Society* **216**, U696-U696 (1998).

22 Mackerell, A. D., Jr., Feig, M. & Brooks, C. L., 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **25**, 1400-1415, doi:10.1002/jcc.20065 (2004).

23 Trabuco, L. G., Harrison, C. B., Schreiner, E. & Schulten, K. Recognition of the regulatory nascent chain TnaC by the ribosome. *Structure* **18**, 627-637, doi:S0969-2126(10)00104-8, 10.1016/j.str.2010.02.011.

24 Trabuco, L. G. *et al.* Applications of the molecular dynamics flexible fitting method. *J Struct Biol*, doi:S1047-8477(10)00293-5,10.1016/j.jsb.2010.09.024.

25 Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174-180, doi:S1046-2023(09)00088-7, 10.1016/j.ymeth.2009.04.005 (2009).

26 Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys Rev* **2**, 21-27, doi:10.1007/s12551-009-0026-3.

27 Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673-683, doi:S0969-2126(08)00133-0, 10.1016/j.str.2008.03.005 (2008).

Chapter 3
Localization of Assembly Factors on 40S Subunit Precursor

3.1 Abstract

Eukaryotic ribosome assembly is a step-wise process requiring the coordinated actions of approximately 200 assembly factors. In the previous chapter, I have shown the structure of a precursor 40S subunit and the densities in this 3D map unique to the premature versus mature particles. In this chapter, each of the seven assembly factors are localized on pre-40S complexes by obtaining cryo-EM 3D maps of pre-40S with depleted assembly factor(s), immuno-labeling of AFs on pre-40S, and previously published AF interaction and functional data. The locations of Tsr1 and Enp1 indicate their roles in the maintenance of helix 44 and the beak, respectively, in their premature conformations. The localization of Rio2, Dim1, Pno1 and Nob1 indicate a network of interactions between the AFs, as well as rps, predicted by each of their requirement for cleavage by Nob1 and site D. Localization results for all seven assembly factors also reveal their role in preventing pre-40S from joining 60S by blocking the binding sites of one or multiple factors required for the formation of the 80S complex.

3.2 Results

To localize binding sites of AFs on pre-40S, particles were isolated from yeast strains with specific AFs genetically deleted or depleted under the control of a galactose-inducible promoter, resulting in the loss of one or more AFs on these particles. The loss of additional AFs may be attributed to cooperative binding of the AFs. Cryo-EM 3D maps were generated for each of these particles and compared to WT pre-40S 3D maps to reveal the density attributed to the missing factor(s).

The 3D structure of recombinant Tsr1

Crystal structures of partial or full-length proteins homologous to yeast Rio2, Nob1, Dim1 and Pno1 are available, but structural information is lacking for Tsr1, Ltv1 and Enp1. These three factors were expressed recombinantly in *E. coli,* purified and analyzed by negative stain EM. Characterization of Ltv1 and Enp1 was unsuccessful due to flexibility and aggregation issues, respectively. Preparations of the recombinant *S. cerevisiae* Tsr1 (rTsr1) were amenable to analysis by single particle negative stain EM. 2D projection averages of the 91 kDa protein revealed a preferred orientation of an elongated particle with a hook-like appearance, displaying a handle that is formed by a globular domain (**Figure 3.1**). To locate the protein's N-terminus, rTSr1 with a maltose binding protein (MBP) fused at its N-terminal end (MBP-Tsr1) were isolated and examined by

EM. 2D class averages of MBP-Tsr1 reveal an increase in density at the end of the protruding globular domain, indicating that the protein's N-terminus is located in this region (**Figure 3.2**). Next, to understand the 3D structure of rTsr1, the random conical tilt method (Radermacher, 1987) was used on negative stain EM images. Tilt-pair particles from rTSr1 were used to calculate a 3D reconstruction of rTsr1 to a resolution of ~30 Å (FSC=0.5) (**Figure 3.3**). The 3D reconstruction reveals a smaller globular domain located at the protein's N-terminus and larger more extended domain at its C-terminus.

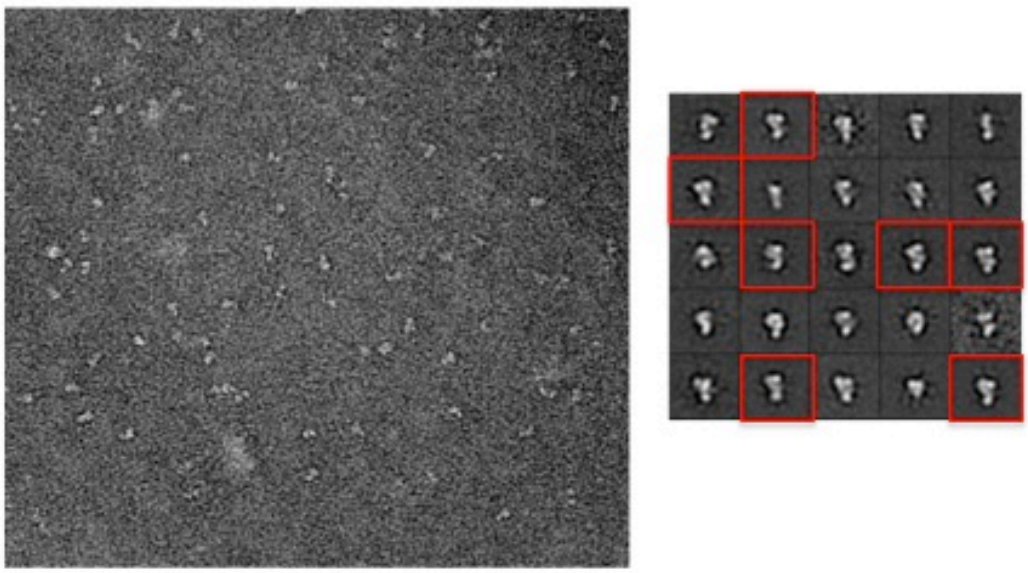


Figure 3.1: Negative stain EM analysis of purified recombinant Tsr1. (A) Raw EM image of rTsr1 embedded in negative stain. (B) Class averages of rTsr1 show a hook-like appearance and globular domain. Particles belonging to the class averages outlined in red boxes were used for 3D reconstruction rTsr1 shown in **Figure 3.3**.

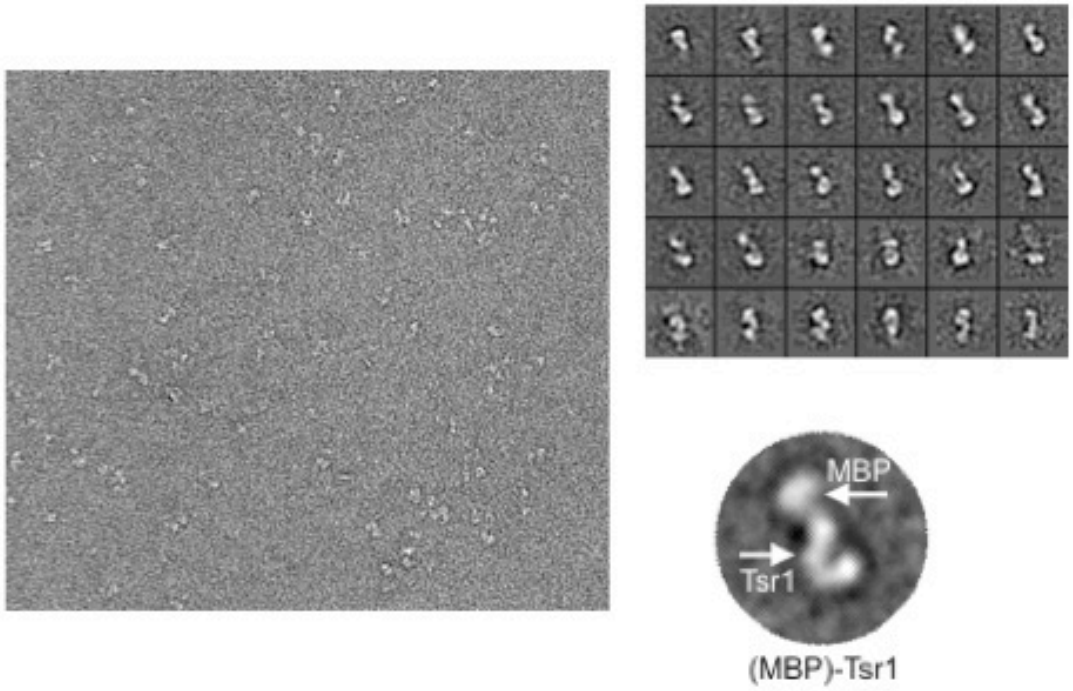


Figure 3.2: An MBP-tag on rTsr1 allows for the localization of its N-terminus. (A) Raw EM image of rTsr1-MBP embedded in negative stain. (B) Class averages of rTsr1-MBP. An increase in density is seen relative to the rTsr1 alone. (C) Zoomed in view of a rTsr1-MBP class average. The MBP tag density is labeled and is opposite the hook-like region of the protein.

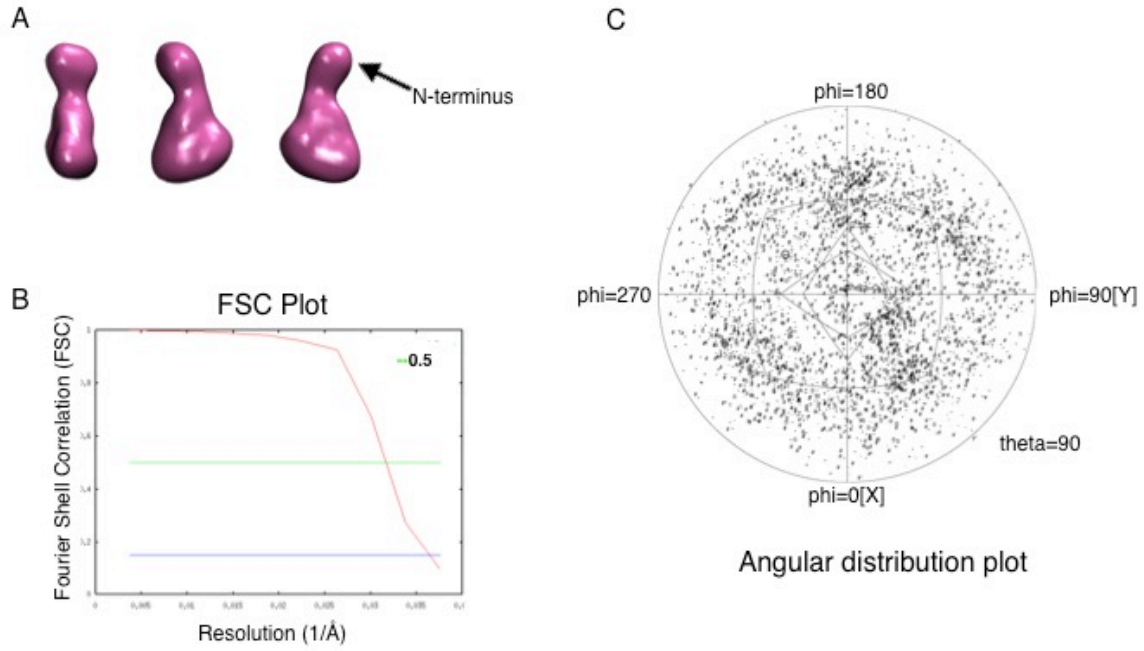


Figure 3.3: 3D reconstruction of Tsr1. (A) 3D volume of rTsr1 was obtained using RCT on the particles belonging to the classes outlined in red in Figure 3.1. The 3D reconstruction of Tsr1 shows a smaller N-terminal globular domain, the N-terminal end was identified as shown in Figure 3.2. (B) FSC curve of the Tsr1 3D reconstruction shows a resolution of approximately 30Å at FSC=0.5. (C) Angular distribution of the angles of the particle projections used in the final 3D reconstruction of Tsr1.

Localization of Nob1 on pre-40S

Nob1 expression was put under the control of a galactose-inducible promoter in the Rio2-TAP yeast strain. After galactose-starvation, these pre-40S particles (Nob1-depl) were isolated via a TAP-tag on Rio2. SDS-PAGE analysis (**Figure 3.4**A) shows that Nob1 is the only AF factor depleted in these particles. Nob1-depl particles were examined with cryo-EM and a 3D reconstruction was calculated to a resolution of 20 Å (at FSC=0.5) (**Figure 3.4**B). Subtraction of the

Nob1-depl 3D map from the pre-40S 3D map, reveals a density located on the back of the platform near site D (**Figure 3.4**C). This placement of Nob1 allows its interaction with Rps5 and Rps14, which are required for site D cleavage[1,2], consistent with direct protein-protein interactions between Nob1 and these proteins[3].

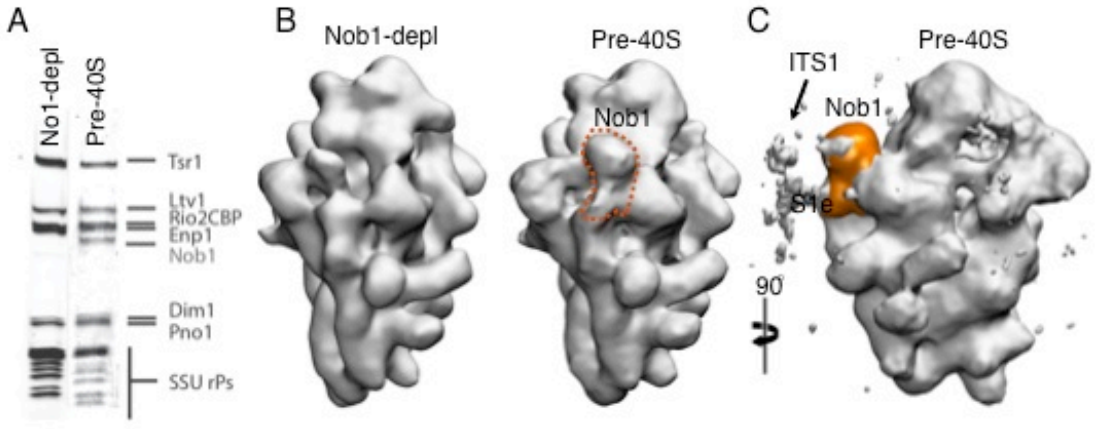


Figure 3.4: Localization of Nob1 on pre-40S. (A) SDS-PAGE results from Nob1-depl particles reveals only Nob1 is has decreased occupancy in the Nob1-depl particles as compared to WT. (B) 20 Å cryo-EM 3D structure of the Nob1-depl. The cryo-EM 3D map of Nob1-depl is compared with the WT pre-40S 3D map, the density decreased in the depletion particles at the back of the platform, revealing Nob1 binds to pre-40S in that region. (C) Visualization of the WT pre-40S particle at lower resolution threshold shows a cloud of density extending from the region corresponding to Nob1 which may be the flexible nucleotides making up ITS1.

Localization of Rio2 and Dim1 on pre-40S

To localize Rio2 on the pre-40S subunit, Rio2 expression was put under the control of a galactose inducible promoter in the Ltv1-TAP yeast strain. Particles

from this strain were isolated using a TAP tag on Ltv1 as bait. Comparison of particles isolated via Rio2-TAP and Ltv1-TAP is possible because native preparations using either AF as bait are essentially identical in composition[4], a result we confirmed by purification of these particles and analysis with SDS-PAGE (**Figure 3.5**A). Isolated Rio2-depletion pre-40S particles (Rio2-depl) were analyzed with SDS-PAGE and MS, and in addition to Rio2; Nob1 and Dim1 levels were also decreased significantly. The loss of both Dim1 and Nob1, in the absence of Rio2, is consistent with direct protein-protein interactions between these AFs and Rio2[3]. A cryo-EM 3D reconstruction of Rio2-depl was calculated to a resolution of 22 Å (**Figure 3.5**B). Comparison of the Rio2-depl and WT pre-40S 3D map reveals missing densities at the previously identified position of Nob1, and two interconnected perpendicular regions on the platform, at the subunit interface.

Data from rRNA footprinting experiments with the conserved bacterial ortholog of Dim1, KsgA[5,6], shows it interacts with helix 44 and helix 27 in premature 30S ribosomal subunits[7]. Docking the crystal structure of hDim1 in the lower density, with its N-terminal domain closer to the beak (PDB ID: 1ZQ9)[8] (green in **Figure 3.5**C), satisfies these biochemical constraints. The docking of the crystal structure results in a very good fit with a cross-correlation (CC) value of 0.854. In

this position, the active site of hDim1 is adjacent to its tandem adenosine substrates, A1781/A1782 (yellow space-fill in **Figure 3.5**), near the 3'-end of 18S rRNA.

Rio2 interacts directly with Dim1 (located on the platform), Rps15 (located in the head region), and Rps5 and Rps14[3] (found on the platform near the mRNA channel). Attributing the last unidentified density in Rio2-depl to Rio2 allows the kinase to interact with each of these proteins. We docked the crystal structure of aRio2[9] (PDB ID: 1ZAO), which contains approximately 80% of the amino acids present in the yeast protein sequence, into our pre-40S 3D map with the N-terminus of Rio2 downward, contacting the density of Dim1 (**Figure 3.5**B). At this position, the N-terminus of Rio2 is buried behind the top of helix 44 and the C-terminus is exposed to solvent, consistent with our ability to isolate pre-40S particles via a TAP-tag on the C-terminus of Rio2. The CC value for the docking of the Rio2 crystal structure into the 3D map is 0.904.

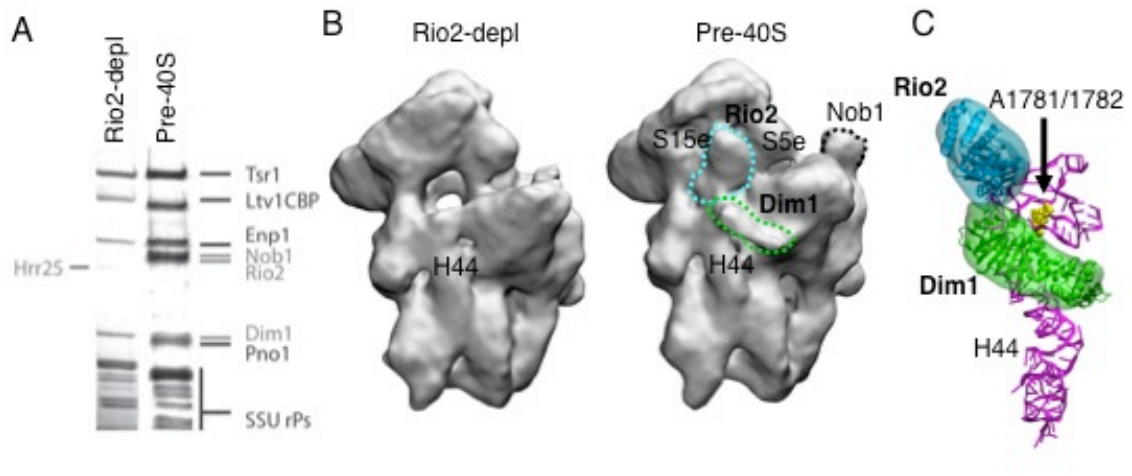


Figure 3.5: Localization of Rio2 and Dim1 on pre-40S. (A) SDS-PAGE of Rio2-depl shows decreased occupancy of both Rio2 and Dim1 in Rio2-depl particles compared with WT pre-40S,. (B) 22 Å 3D map of Rio2-depl compared to the WT pre-40S 3D map, the densities lost in the depletion 3D map (in addition to that already assigned to Nob1) are two interconnected regions located on the platform. (C) Docking of the crystal structures of archaeal Rio2 (PDB ID: 1ZAO) and human Dim1 (PDB ID: 1ZQ9) into their respective densities following constraints explained in the text. Helix 44 is colored magenta and the adenosine nucleotide substrates of Dim1 are in yellow space-fill.

Localization of Tsr1 and Pno1

To localize Tsr1, Rio2-TAP pre-40S particles were isolated from a yeast strain with Tsr1 expression under control of a galactose inducible promoter. These particles were analyzed for quality with negative stain EM and the images indicated the particles were not intact (as compared to the previous pre-40S particles appearance). Rio2 incorporation into pre-40S may require the presence of Tsr1, preventing isolation of this particle by using Rio2-TAP as bait. Alternatively, Tsr1 was put under the control of a galactose inducible promoter in the Ltv1-TAP strain. These particles appeared to be intact when visualized by

negative stain EM. As expected by the inability to isolate the particles with a Rio2 TAP-tag, SDS-PAGE reveals that in addition to Tsr1, the Ltv1-TAP Tsr1-depleted complexes (Tsr1-depl) have less occupancy of Rio2, as well as Dim1, Pno1 and Nob1 as compared to the WT pre-40S (Fig. 2.17). A 3D cryo-EM map was calculated with these particles to a resolution of 26 Å (**Figure 3.6**A). Decreased densities are those previously assigned to Rio2, Dim1, and Nob1, and two additional unassigned densities. The first is a fairly large density located to the left of H44, on the binding interface from the perspective of 60S, and the second is a smaller density located on the back of the platform near the binding site of Nob1. Another surprising feature of the Tsr1-depl, but not the WT, Rio2-depl or Nob1-depl pre-40S 3D maps, is the lack of the pronounced kink in helix 44 (**Figure 3.6**B), characteristic of the premature subunit. This observation unique to this 3D map indicates that one of the AFs depleted uniquely in the Tsr1-depl particles may stabilize the kinked conformation of helix 44.

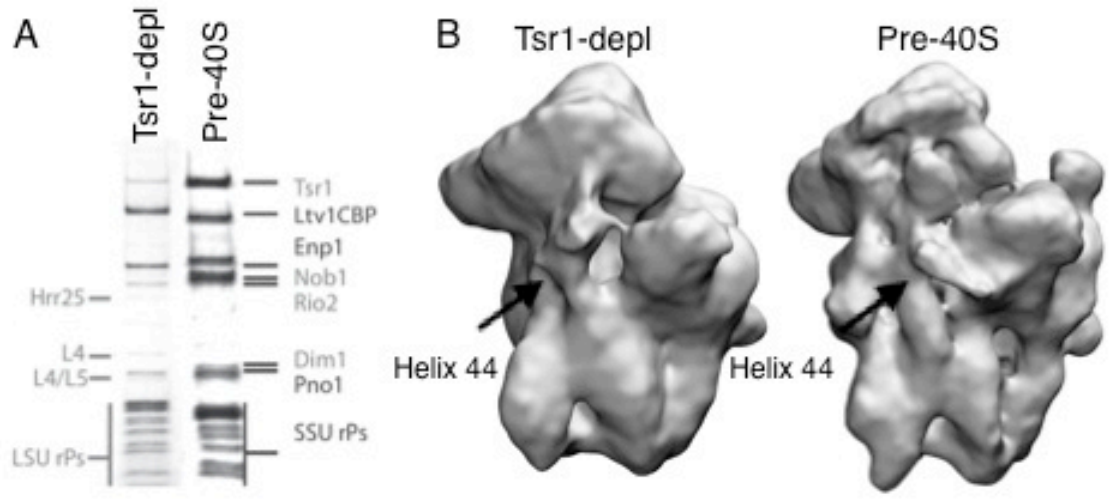


Figure 3.6: The conformation of helix 44 is altered in Tsr1-depl pre-40S. (A) SDS-PAGE of Tsr1-depl particles compared to WT pre-40S showing decreased occupancy of Rio2, Dim1, Pno1 and Tsr1. (B) The 26 Å 3D map of Tsr1-depl, compared to the WT pre-40S 3D map, shows a distorted conformation of helix 44.

The 3D reconstruction of recombinant Tsr1 possesses a overall shape and size similar to the region of density lost in Tsr1-depl adjacent to helix 44 on the subunit interface of pre-40S. The Tsr1 volume from our 3D reconstruction fits well within the pre-40S 3D map in the region to the left of helix 44, docked with the N-terminus toward the beak (**Figure 3.7**). At this position, the N-terminus of Tsr1 can interact with Rio2's N-terminus, consistent with direct binding between the proteins that is sensitive to a N-terminal tag on Rio2[3].

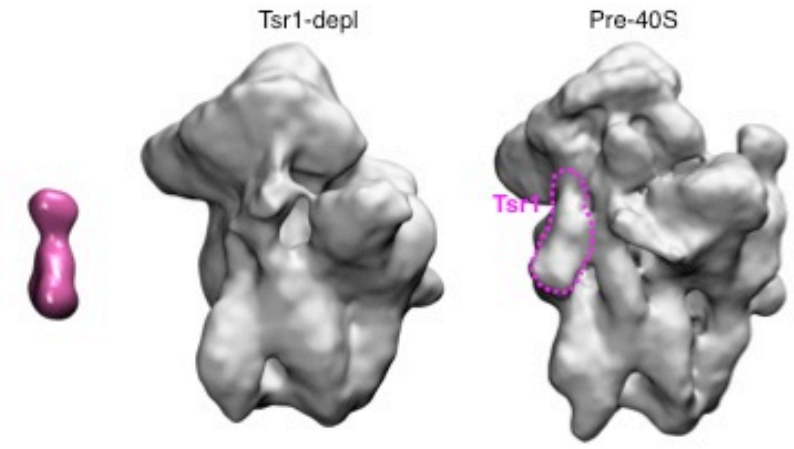


Figure 3.7: Localization of Tsr1 on pre-40S. The 3D reconstruction of rTsr1 (left panel, pink) possesses a similar structure to the larger density adjacent to helix 44 lost in the Tsr1-depl 3D reconstruction (middle panel) compared to WT pre-40S (right panel). The density lost is outlined in pink

Helix 44 in all pre-40S complexes presented in this thesis possess a kink in the center of the helix, with the exception of the Tsr1-depl pre-40s. Out of the five AFs depleted in Tsr1-depl, Tsr1 and Pno1 are the only factors not depleted in any of the previous pre-40S 3D reconstructions investigated, indicating that either one or both of these factors is responsible for stabilizing the helix in its premature conformation. The location of Tsr1 to the region near helix 44 allows interaction between the protein and helix. Tsr1 may function to maintain helix 44 in its premature state and keeping the decoding site disrupted as well as preventing pre-40S from joining the 60S subunit.

The second unidentified density lost in the Tsr1-depl 3D map is located on the back of the platform, between Nob1 and Rps14. This position is near the 3'-end of the 18S rRNA. The highly conserved C-terminal KH domain of the archaeal Pno1 binds directly the universally conserved GGAUC sequence at the 3' end of the bacterial 16S rRNA, as well as eIF2α, forming a trimeric complex, which has been crystallized[10]. This indicates that Pno1 binds the rRNA near site D. In addition, Pno1's central KH domain binds and regulates Nob1[11]. Docking of the archaeal *Pyrococcus horikoshii* Pno1 crystal structure (PDB ID: 2E3U), containing the two C-terminal KH domains but lacking the N-terminal KH domain, on the back of the platform allows it to bind both Nob1 and the GGAUC sequence (**Figure 3.8**). Pno1's binding to the 3'-end of 18S rRNA and localization near Nob1, may indicate its function to stabilize and/or pass the 3'-end rRNA to Nob1, thereby regulating site D cleavage.

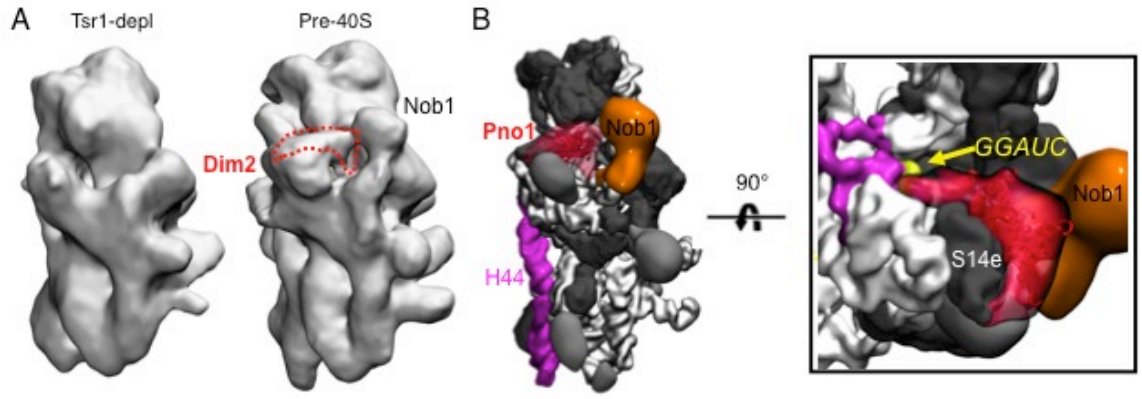


Figure 3.8: Localization of Pno1 on pre-40S. (A) Side views of Tsr1-depl and WT pre-40S 3D maps. The density attributed to Pno1 is on the back of the platform (outlined in red) next to Nob1 and near the end of the 18S rRNA. (B) Docking of the crystal structure of the archaeal Pno1 ortholog[12] (red ribbon) (PDB ID: 2E3U) into this density allows its interaction with Nob1 (orange) and the universally conserved GGAUC sequence of nucleotides (yellow) that it binds with high affinity[10].

Localization of Ltv1 and Enp1

To localize Ltv1, the only non-essential AF[13], Ltv1 was genetically deleted in the Rio2-TAP strain. SDS-PAGE and MS show particles purified from this strain (ΔLtv1) also lack Rps3 and have reduced levels of Enp1, relative to the WT pre-40S (**Figure 3.9**A, **Table 3.1**). This finding is consistent with previous data indicating Ltv1, Enp1 and Rps3 form trimeric subcomplex[4]. Furthermore, Enp1 and Ltv1 also interact directly according to protein-protein interaction studies[3]. The cryo-EM 3D reconstruction of ΔLtv1 particles was calculated at a resolution of 20 Å (**Figure 3.9**A). The position of the beak in this 3D reconstruction is more similar to the mature than premature 40S beak. Investigation of the ΔLtv1 3D

map reveals that the density connecting the beak and helix 16 (in the body of the particle) is lost. This density appears to holds the bridge in its retracted, premature conformation in pre-40S. In addition to the loss of the bridge density, ΔLtv1 also have decreased density at the top of the beak, and on the head region near the beak.

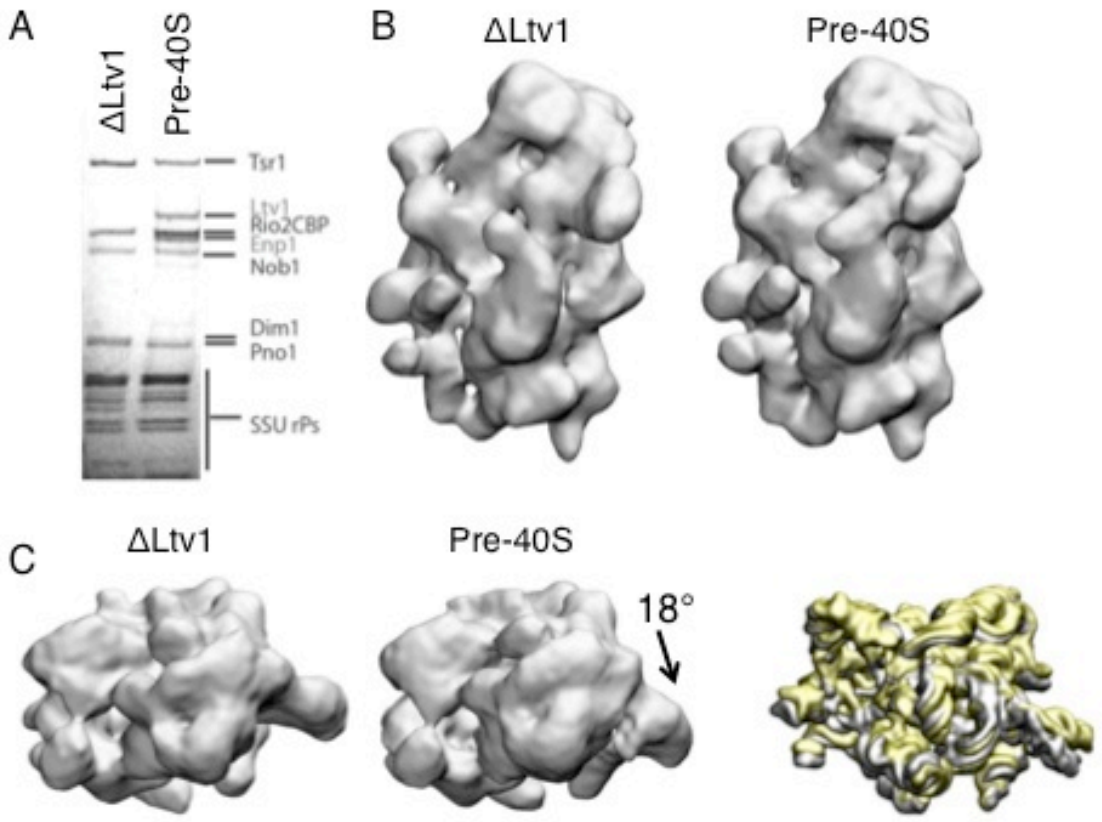


Figure 3.9: Localization of Ltv1 and Enp1 on pre-40S. (A) SDS-PAGE of ΔLtv1, comparison with the WT pre-40S gel shows that Ltv1 is absent from the particles and Enp1 is at decreased occupancy, compared to WT. (B) Side-view of ΔLtv1 (left) and WT (right) pre-40S 3D maps shows the bridge of density connecting the beak to the platform is lost, as well as density in the head. (C) Top view of ΔLtv1 (left) and WT (center) pre-40S show the beak is shifted 18° away from the body in the ΔLtv1 3D reconstruction compared to WT. Far right panel shows an overlay of the rRNA fit into the ΔLtv1 (gold) and WT pre-40S (silver), showing the rRNA in the beak is shifted in the ΔLtv1 particles.

The structural differences observed in the ΔLtv1 map are attributed to the loss or reduction in occupancy of all three proteins. The observation that the deletion of Ltv1 results in the absence of Rps3, and heavy reduction in Enp1 levels, indicate

that a small portion of the ΔLtv1 particle projections dataset will contain Enp1. This density corresponding to Enp1 should be present in the ΔLtv1 3D reconstruction. To quantitatively separate and examine the 3D structure of this subpopulation of particles, we employed multiple reference-supervised alignment[14,15] on our original ΔLtv1 cryo-EM dataset, with the WT pre-40S and ΔLtv1 pre-40S 3D maps as initial volume references. After a single iteration, approximately 80% of particles were assigned (based on CC values) to the ΔLtv1 reference subset, and these particles were refined with the ΔLtv1 3D map as the initial model. The 3D reconstruction of this subset shows the same features (shifted beak, lack of connection to the beak, loss of density at head and beak regions) as the initially calculated ΔLtv1 3D map (**Figure 3.10**). The remaining particle projections were assigned to the WT pre-40S reference and refined using WT pre-40S as an initial model. This 3D reconstruction shows a loss of density in the head, consistent with the single-reference 3D reconstruction, a retracted beak structure, and contains the connection between the beak and helix 16, but the bridge is much smaller relative to the one in WT pre-40S. These results suggest that a subset of particles in the ΔLtv1 dataset contain the protein responsible for the bridge density. The protein responsible for part of the bridge must be Enp1, as it present at low levels on ΔLtv1, but not completely absent, as the case for Ltv1 and Rps3. The observation that the retracted beak is present in

the subset of particles assigned to the pre-40S reference, indicates that Enp1 is sufficient for maintaining the retracted beak. In contrast, density differences on the top of the head and beak, are attributed to Ltv1, due to the consistent loss of these densities in the 3D reconstructions from both particle subsets. In mature 40S, Rps3 is located where the bridge meets the body of the complex, leading to the hypothesis that Rps3 is responsible for the portion of the bridge not occupied by Enp1.

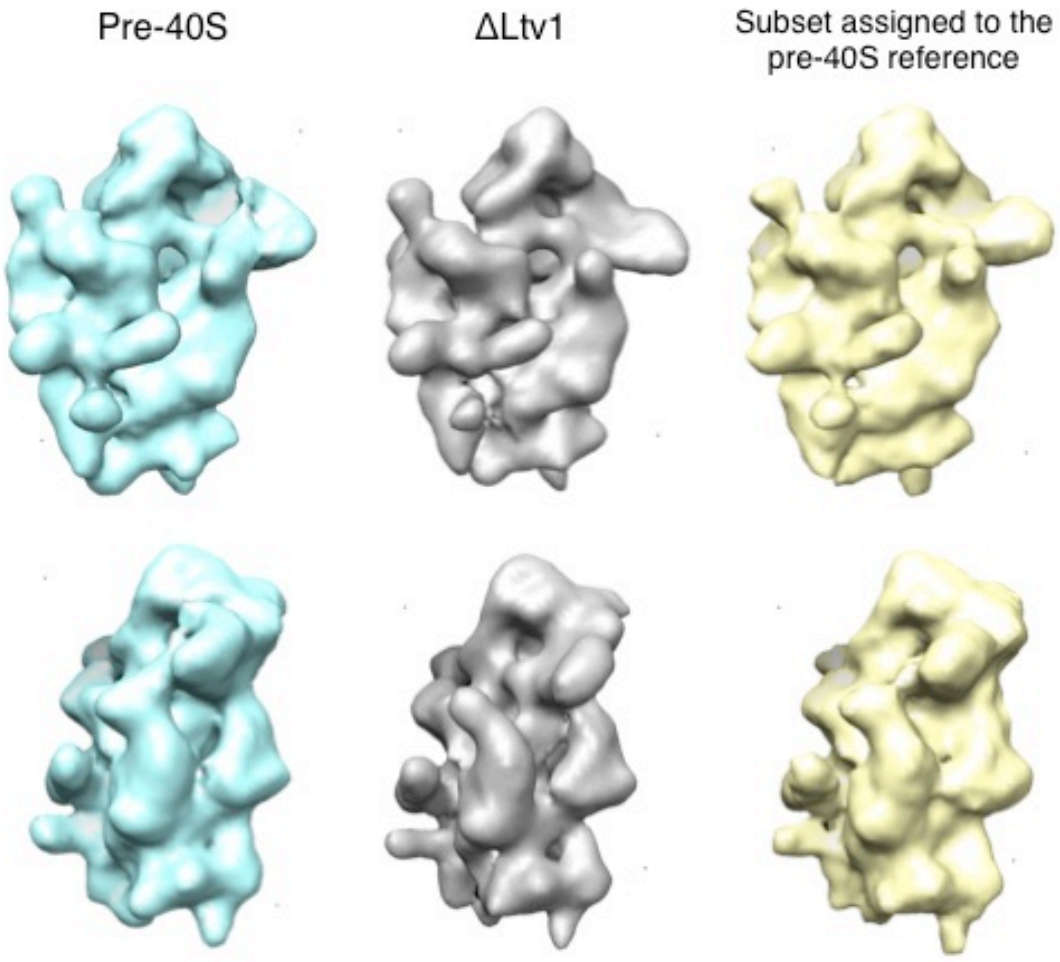


Figure 3.10: Multi-reference alignment of the ΔLtv1 projection dataset. Front and side views of 3D reconstructions from the particle projections assigned to the ΔLtv1 reference (gray, center) and the particle projections aligned to the WT pre-40S reference (gold, right), the 18 Å WT pre-40S 3D map is shown on the right in cyan for comparison. The 3D reconstruction of the particle projections assigned to the WT pre-40S reference contain the bridge density, similar to WT pre-40S.

To provide additional evidence for the localization of Enp1 and Ltv1 at the bridge and head, respectively, we examined pre-40S particles separately labeled with

an antibody binding a Hemaggluttin (HA) tag fused to Enp1 or a CBP tag fused to Ltv1 by negative stain EM. 2D class averages of pre-40S with HA 1°-antibody bound to HA-Enp1 reveal the additional antibody density is at the beak structure (**Figure 3.11**A,B), further confirming that Enp1 is located in the bridge. 2D class averages of pre-40S with CBP 1°-antibody bound to CBP-Ltv1 reveal the additional density attributed to the antibody extend from the top of the particle's head, confirming Ltv1's location on the head (**Figure 3.11**C,D). Together, these data indicate that Enp1 and Rps3 make the connection between the beak and helix 16, while Ltv1 binds at the top of the head and extends toward the beak.

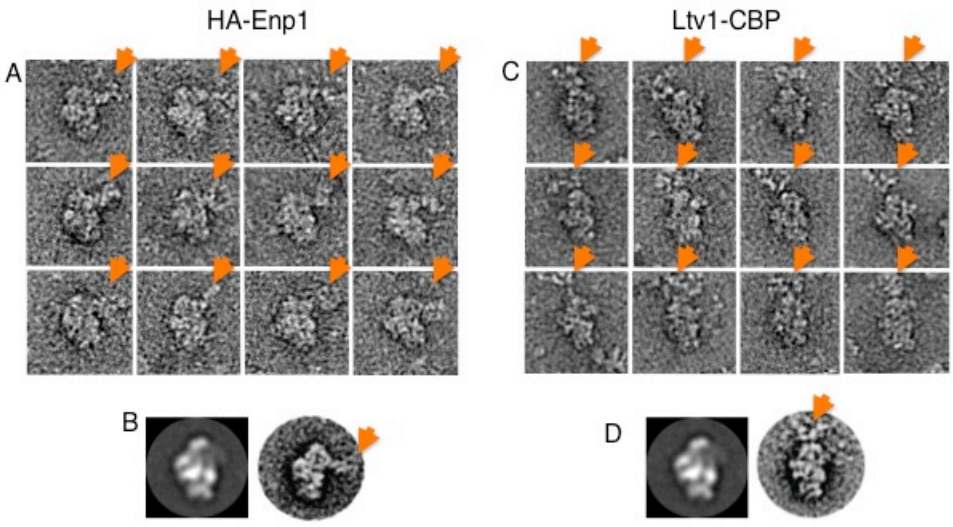


Figure 3.11: Antibody-labeling of Enp1 and Ltv1 on pre-40S complexes. Raw EM images of negative stained pre-40S complexes with HA 1^{o}-antibody bound to Enp1-HA (A) and CBP 1^{o}-antibody bound to Ltv1-CBP (C). (B) Comparison of the class average of Rio2-TAP pre-40S particles (left) and Enp1-HA pre-40S with HA 1^{o}-antibody to Enp1 (right) showing increased density directly on the beak. (D) Comparison of the class average of Rio2-TAP pre-40S particles (left) and Ltv1-CBP pre-40S with CBP 1^{o}-antibody to Ltv1 (right) showing increased density on the head.

3.3 Materials and methods

Isolation of Nob1-depl, Rio2-depl, Tsr1-depl and ΔLtv1

The assembly factor targeted for depletion was put under the control of a galactose inducible promoter in the background of Rio2-TAP or Ltv1-TAP yeast strains (Open Biosystems). In the case of ΔLtv1, the protein was genetically deleted in the Rio2-TAP strain (obtained from Open Biosystems). Initial growth conditions were essentially the same for these strains as Rio2-TAP WT cells. At mid log phase, depletion strains were switched from YP galactose to YP glucose for 12 h of growth (or 20h for the *GAL1::Tsr1* strain) to deplete the specific assembly factors. Particles were isolated using TAP purifications as described for WT pre-40S and were concentrated to a final volume of 20-40 µL using a *Millipore* Biomax-100K NMWL filter device.

Yeast strains

Galactose-inducible strains were created by polymerase chain reaction (PCR) based recombination[16] into the Rio2-TAP or Ltv1-TAP strains from Open Biosystems. Similarly, the entire TAP-tagged Ltv1 open reading frame and approximately 450 nucleotide flanking sequence was amplified by PCR from genomic DNA in the Ltv1-TAP strain from Open Biosystems. This DNA was used to genomically integrate Ltv1-TAP into the Rio2 depletion strain generously provided by J.P. Gélugne[17].

Mass spectrometry

Mass spectrometry was carried out similar to WT pre-40S. The relative peptide amounts for each rps in the different preparations are listed in below.

Rps	Nob1-depl	Rio2-depl	Tsr1-depl	ΔLtv1
0	17	52	64	64
1	46	87	76	76
2	15	12	49	14
3	19	44	117	
4	26	36	79	26
5	27	46	61	17
6	43	63	53	14
7	19	55	122	29
8	4	5	47	9
9	13	59	81	16
10				
11	2	29	35	11
12	19	33	29	11
13	27	53	58	19
14	17	55	44	18
15	16	50	34	10
16	12	46	62	14
17	34	60	51	14
18	22	92	56	15
19	29	70	46	9
20	14	54	41	
21	17	11	12	4
22	17	31	49	14
23	2	7	19	8
24	18	34	53	18
25	13	18	30	9
26		4	3	
27	16	13	8	12

Table 3.1: Mass spectrometry analysis of rps in pre-40S particles. Peptide numbers for the rps in Nob1-depl, Rio2-depl, Tsr1-depl and delta-Ltv1.

Immuno-labeling of pre-40S particles

For negative stain EM imaging of immuno-labeled HA-Enp1 or Ltv1-CBP pre-40S, particles were incubated with the corresponding primary antibody targeting the C-terminal tag on the AF (1^{o}-CBP or 1^{o}-HA (Covance)). Incubations were carried out on ice for 15-30 minutes at a molar ratio of approximately 1:1 of pre-40S:antibody (as assessed by EM images of antibodies or particles alone). The mix was diluted in standard TAP buffer and applied to glow-discharged continuous carbon grids for negative stain embedding.

Purification of rTsr1 and MBP-rTsr1

Recombinant Tsr1 and MBP-Tsr1 (rTsr1 and MBP-rTsr1, respectively) were purified according to Campbell and Karbstein[3]. In this protocol, the two yeast proteins were overexpressed in Rosetta (Novagen) *E. coli* cells. Overnight cultures of the cells were grown in Luria Broth (LB) Miller medium supplemented 0.05 mM kanamycin and 0.05 mM chloramphenicol at 37°C to an OD_{600} of approximately 0.5, at this density protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Proteins were expressed for 5 h at 30° C followed by centrifugation of cells to remove excess media. Cell pellets were resuspended in standard TAP buffer supplemented with 0.1 mM of the protease inhibitor phenylmethanesulfonylfluoride (PMSF) and 5 mM of the

DNAse inhibitor benzamidine, according to the Qiagen protein-purification handbook. The resuspended cells were lysed by sonication and the first step of purification was passing the lysate over nickel nitrilotriacetic acid (Ni-NTA) resin (Qiagen) to bind the 6X histinde tag on the N-terminus. Proteins were eluted with 100 mM imidazole, imidazole was removed by over-night dialysis in standard buffer and cleavage of the tag was carried out by incubation with TEV protease (omitted for MBP-Tsr1) in a buffer containing 50 mM TRS, pH 8, 100 mM potassium chloride (KCl), and 1 mM dithiothreitol (DTT). The sample was purified further by passing the eluate over a MonoQ ion-exchange and gel-filtration columns.

Specimen preparation and negative stain EM imaging

3.5 µl of antibody-labeled pre-40S complexes, rTsr1 and MBP-rTsr1 samples were applied to glow-discharged continuous carbon grids (Electron Microscopy Sciences). Excess sample was blotted from the grid with filter paper and coated with 1% weight-to-volume (w/v) uranyl formate, excess stain was also blotted with filter paper. Grids were imaged at room temperature with a Tecnai T12 electron microscope (FEI) operated at 120 kV using low-dose procedures. Images were recorded at a magnification of 71,138x and a defocus value of approximately -1.5 µm on a Gatan US4000 charge-coupled device (CCD)

camera. All images were binned (2 x 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. 250 Enp1-HA/1°-HA, 503 Ltv1-CBP/1°-CBP and 759 MBP-rTsr1 particle images were excised manually and subjected to reference-free alignment and classification using the SPIDER software[18] into 10-30 classes. For 2D analysis and 3D reconstructions of rTsr1, 4927 particle pairs were interactively selected from 106 60°/0° image pairs for random conical tilt analysis (described in the next section).

3D reconstruction of rTsr1

The particles excised from untilted rTsr1 images were subjected to reference-free alignment and classification into 25 classes. Based on the results of the classification, seven independent 3D reconstructions were derived from particles belonging to similar classes: 2, 6, 12, 14, 22 and 25 (red boxes in **Figure 3.1**). The random conical tilt technique[19] was used to calculate a first back projection 3D map using the images of the tilted specimen followed by angular refinement. Next, the corresponding particles from the untilted images were added to the dataset, and the new dataset was subjected to iterative refinement. The resulting 3D maps were used as 3D reference models, and final 3D reconstructions were calculated using the program FREALIGN[20]. FREALIGN was again used for further refinement of the orientation parameters and CTF correction.

Reconstructions revealed volumes with similar features, therefore 3450 60°/0° projections were merged from all seven classes and used to calculate a final density map. The resolution of the final 3D map was approximately 30 Å according to Fourier shell correlation (FSC) using the FSC = 0.5 criterion (**Figure 3.3**B).

Cryo-EM sample preparation and imaging of depletion/deletion pre-40S

2 µl of Nob1-depl, Rio2-depl, Tsr1-depl or ΔLtv1 pre-40S samples, at OD_{600} of 10-20, were adsorbed on glow-discharged Quantifoil R2/2 200 mesh grid. The grids were blotted once for 1-2 seconds and plunge frozen in liquid ethane using a Vitrobot (FEI Mark IV) with the temperature maintained at 22° C and humidity greater than 85%. Using a Gatan 626 single tilt cryotransfer system, specimens were imaged on a Tecnai F20 transmission electron microscope (FEI) equipped with a field emission electron source operated at 200 kV. Images were recorded at a magnification of 66,964x on a Gatan US4000 CCD camera at defocus values ranging from -1.5 to -4.0 µm. Particles were preferentially found in relatively thick ice, lowering the signal-to-noise ratio of the particle projections. The micrographs were not binned and pixel size under these conditions is 2.24 Å on the specimen level.

Single-particle reconstruction of depletion/deletion pre-40S

The particles were excised from cryo-EM images, CTF parameters were determined, CTF phases were flipped, the 3D reconstructions were calculated, and analyzed as described for WT pre-40S in Chapter 2. The WT Pre-40S 3D reconstruction was used as an initial reference to refine 9132 Nob1-depletion projections to a resolution of 20 Å and 7069 Rio2-depletion projections to a resolution 22 Å (**Figure 3.12**).

After separation of 80S-like particles from the Tsr1-depl dataset (for an explanation and justification of this procedure, see Chapter 4), the resulting 9176 particle projections were manually screened for quality to a final dataset of 5127 projections. The WT pre-40S density was used as the initial model to calculate a 3D reconstruction with a resolution of 26 Å (FSC=0.5) (**Figure 3.12**). The lowered resolution relative to the other 3D maps described may be explained by particle heterogeneity, due to the multiple combinations of the five AFs at decreased occupancy in these particles, or flexibility of the particles, due to a less stabilized state of helix 44.

A total of 10,235 ΔLtv1 projections were refined using the pre-40S map as the initial model to calculate a 3D reconstruction at a resolution of 20 Å (FSC=0.5)

(**Figure 3.12**). To examine the presence of Enp1 in a subpopulation of the particle projections, as observed in SDS-PAGE data, we employed multi-reference supervised alignment[14,15] using our WT pre-40S and ΔLtv1 3D maps as initial reference volumes. After the first iteration of refinement, 1641 projections were assigned to the WT pre-40S subset, and 7507 projections were assigned to the ΔLtv1 subset. 3D reconstructions of the two resulting subpopulations were iteratively refined against the initial model to which they were assigned, using refinement parameters previously described for WT pre-40S.

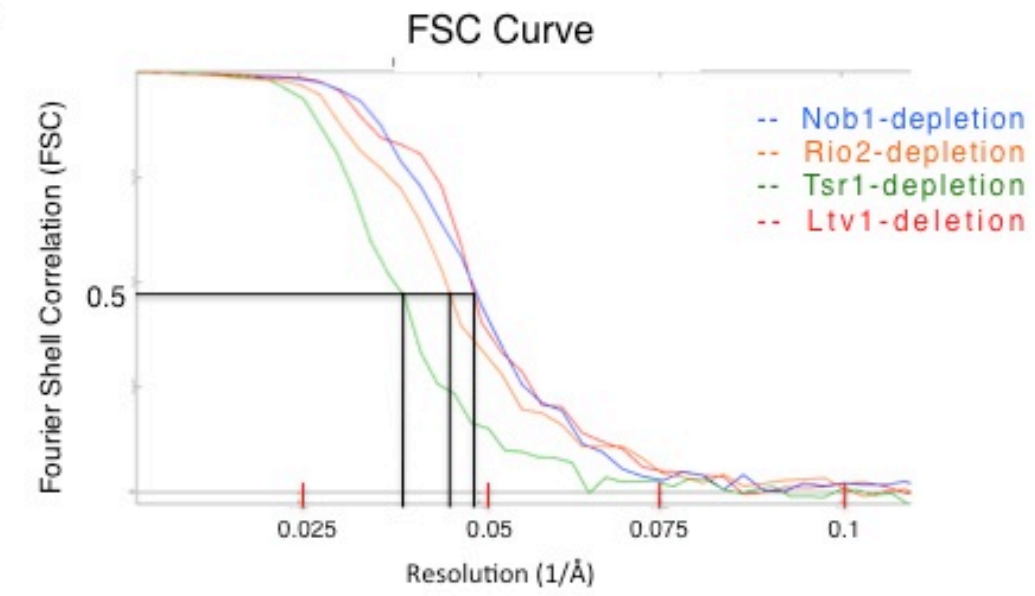


Figure 3.12: FSC curves for depletion and deletion pre-40S 3D maps. According to the FSC at 0.5 criterion, the FSC curves for the pre-40S 3D map indicate a resolution of ~20Å for Nob1-depletion (blue), ~22Å for Rio2-depletion (orange), ~26Å for Tsr1-depletion (green), and ~20Å for Ltv1-deletion (red).

Docking of assembly factors

Crystal structures of human Dim1 (hDim1; PDB ID: 2E3U), archaeal Rio2 (aRio2; PDB ID: 1ZAO), and archaeal Pno1 (aDim2, PDB ID: 1ZQ9) were docked into the pre-40S density corresponding to the locations and following specifications described in Section 2.2. Crystal structures were fit as rigid bodies using SITUS[21]. The final cross-correlation values for docking were 0.854 for aDim1 and 0.904 for aRio2,

3.4 Discussion

The results presented in this chapter and in Chapter 2, as well as previously published biochemical data, allowed us to generate a model of the overall molecular architecture of a late cytoplasmic 40S ribosomal subunit maturation intermediate (**Figure 3.13**). Localization of the seven associated AFs agrees with RNA-protein cross-linking[22] and protein pull-down[3,11,23] experiments. I have also presented a novel observation that the binding of assembly factors on the pre-40S complex blocks the binding sites of factors required for the initiation of protein translation by the ribosome.

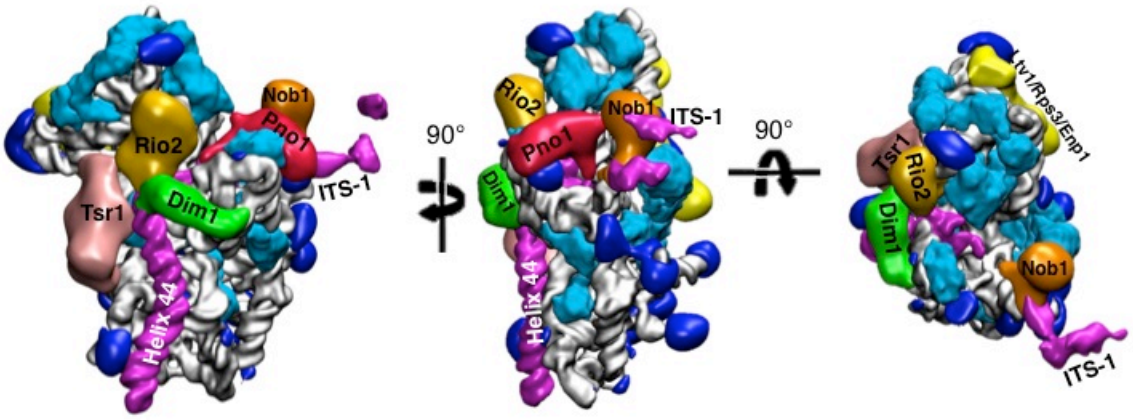


Figure 3.13: Structural model of a 40S subunit precursor. The pre-40S model of rps and rRNA, generated by MDFF, are shown in silver, densities corresponding to rps not in our model are royal blue and helix 44 in magenta. Using the data presented, we have localized the seven AFs on pre-40S. Rio2 (gold) and Dim1 (green) are located on the platform, Tsr1 (light pink) is adjacent to helix 44, Pno1 (red) and Nob1 (orange) are interacing on the back of the platform, and the Ltv1-Rps1-Enp1 sub-complex (yellow) forms the density connecting the beak to the platform, as well as density on the head.

Localization of AFs agrees with their interaction data

Confidence in our localization of the AFs on pre-40S is increased by its agreement with a large body of published data. Cross-linking experiments were performed with each of the AFs localized in this thesis (with the exception of Pno1)[22] and the results agree with our placement of Dim1, Rio2 and Enp1 (**Figure 3.14**). However, only a subset of the cross-linking hits for Nob1, Tsr1, and Ltv1 agree with our localization of the proteins on pre-40S. The sites of RNA interaction for these three proteins are dispersed on the premature 40S subunit. The discrepancy of our localization data and the results from the cross-linking

experiments may be attributed to the use of whole cell lysates in the crosslinking experiments, contrary to the use of isolated complexes that were used for our structural characterization of pre-40S. The lysates presumably contain multiple premature 40S species at different stages of maturation with various AF populations and/or altered locations/conformations of the AFs on pre-40S.

Data from AF protein-protein interaction experiments by Campbell and Karbstein[3] were consistent with our localization of the seven pre-40S AFs (**Figure 3.15**A).

The work of Campbell and Karbstein revealed Nob1 interacts with Pno1, Rps5 and Rps14 and our localization of Nob1 on the back of the platform allows its direct interaction with each of these proteins. Placement of Rio2 in the center of the pre-40S platform permits its binding to both Dim1 and Tsr1, interactions also observed in the protein-pulldowns[3]. This thesis has localized Ltv1 to the head and top of the beak on pre-40S, which allows contact between Ltv1 and Rps5, Rps14, Rps15, and Enp1, consistent with results in Campbell and Karbstein study[3].

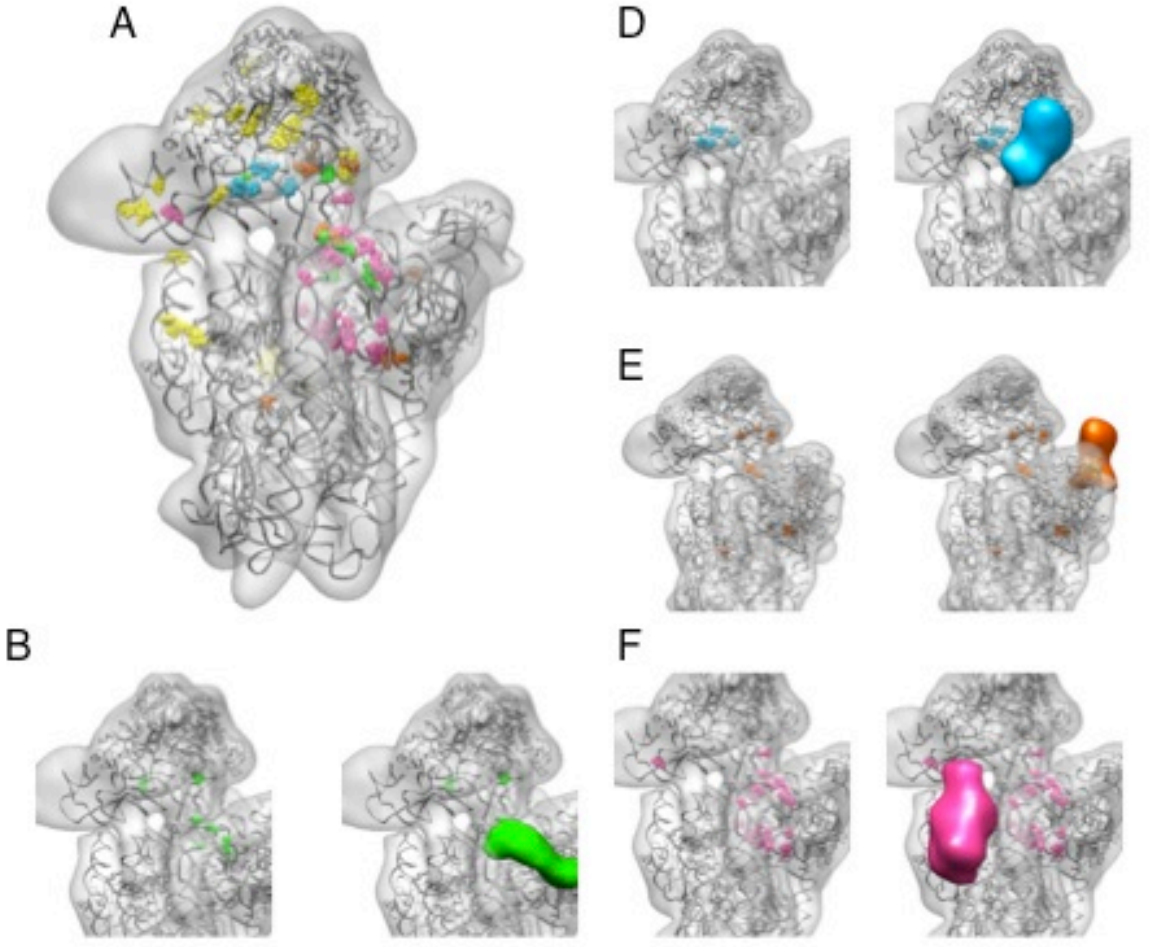


Figure 3.14: Comparing AF locations with RNA interaction data. (A) Sites of crosslinks[22] to specific AFs are shown in space-fill: Dim1(green); Enp1(yellow); Rio2(blue); Nob1(orange;) Tsr1(pink) and Ltv1(yellow). B-G. Comparison of sites of individual crosslinks with the density we localized each protein to in our pre-40S 3D map, (B) Dim1 (C) Enp1 (D) Rio2 (E) Nob1 (F) Tsr1 (G) Ltv1.

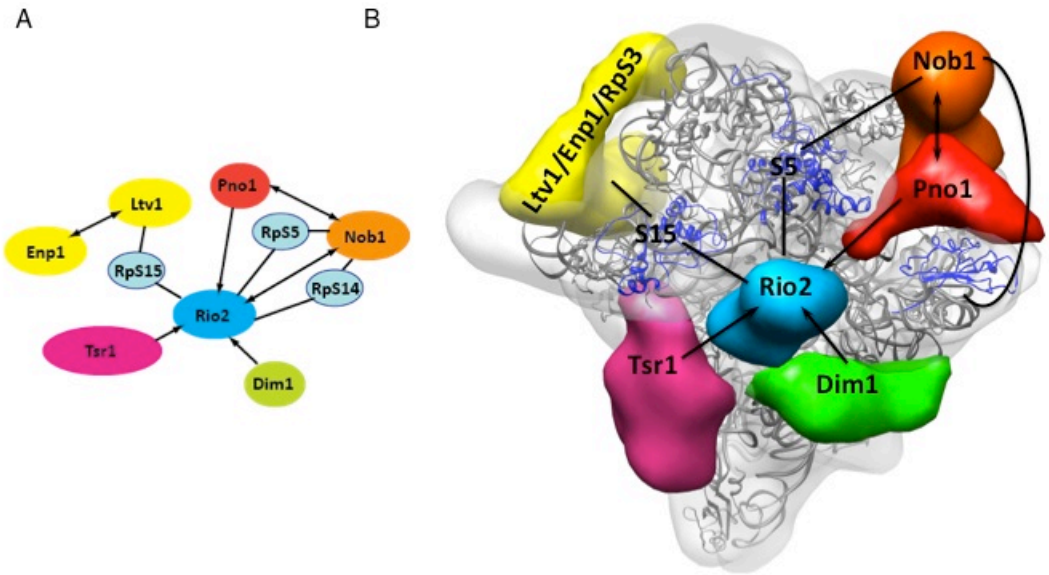


Figure 3.15: Comparing AF locations with protein interaction data. (A) Summary of previously published protein-protein interactions (arrows between proteins) from *in vitro* protein pulldowns[3,11,24], and yeast 2-hybrid interactions[25]. (B) Interactions of those described that are consistent with the placement of the AFs in our pre-40S 3D map. Ribosomal proteins are shown as blue ribbons.

AFs located on the platform of pre-40S form a network of interactions

Nob1, the endonuclease responsible for site D cleavage, is found in premature 40S complexes containing the 20S precursor to the mature 18S rRNA. The presence of Nob1 and its substrate on pre-40S complexes indicates an external signal may be required for activation of Nob1. Dim1, Pno1[11] and Rio2[4], as well as Rps5 and Rps14, are all required for site D cleavage by Nob1. Our pre-40S model does not indicate a direct interaction of Rio2 and Nob1, however the kinase activity of Rio2 is required for site D cleavage in human cells[26], indicating

that the phosphorylation signal is passed to Nob1 through other AFs and/or rps on pre-40S. Candidate proteins, based on their location, include Pno1, Rps5 and Rps14. Potential experiments designed to test the roles these proteins play in Nob1 activation are described in Chapter 5.

As with Rio2, activity of Dim1 is required for site D cleavage but the methylase does not interact directly with Nob1. Specifically, the methylation of adenosines 1780 and 1781 by Dim1 is required for generation of the 18S rRNA. The addition of two methyl groups on the rRNA nucleotides may cause a conformational change in the rRNA, altering its interaction with Pno1 and/or Nob1. Approximately 20 nucleotides downstream of the adenosine substrates, the universally conserved 3' GGAUC sequence is found, which the bacterial homology of Pno1 binds[10] (**Figure 3.8**). The alteration of the rRNA path and subsequent interaction with Pno1 and Nob1, found at the back of the platform, may be the mechanism that Dim1 activates Nob1, without direct physical interaction of the two proteins.

The presence of Tsr1 holds Helix 44 in its premature conformation

Of the five pre-40S 3D reconstructions calculated in this thesis, only the Tsr1-depl pre-40S 3D map shows an altered conformation of helix 44. The pre-40S

complexes from the yeast strain with Tsr1 depleted also have a lowered occupancy of additional AFs, but Tsr1 and Pno1 are the only AFs depleted exclusively in the Tsr1-depl and not in the additional particles investigated. I have localized Tsr1 to the density adjacent to helix 44 and under the beak, and Pno1 to the density near Nob1 on the back of the platform. The location of Tsr1 on pre-40S allows for the direction interaction of the protein with helix 44. This observation and the previously published data showing that Tsr1 is not found in polysomes[17], while Pno1 is found in small amounts, indicates the presence of Tsr1 prevents the joining of pre-mature 40S subunits with 60S subunits. This regulation presumably occurs by maintaining pre-40S complexes in a premature state when Tsr1 is bound. This hypothesis is further investigated in Chapter 4.

Assembly factors block binding sites of translation initiation factors

The results in this chapter show Rio2, Tsr1 and Dim1 are located on the 60S subunit-binding interface and line the mRNA channel. The sites of Rio2 and Dim1 overlap those of translation initiation factor eIF1, responsible for mRNA recruitment[27], and the P and E-sites of tRNA (**Figure 3.16**). Similarly, the binding of the bacterial homology of Dim1 overlaps the binding of the bacterial initiation factor IF3[7], and Nob1 and Pno1 overlap the binding site of eIF3[28] on the back of the platform.

In addition to recruiting the mRNA 5'-end to mature 40S, cooperative binding of eIF1 and eIF1A also opens the mRNA channel[27]. Opening of the mRNA channel requires stabilization by a connection formed between the shoulder and head (the 'latch') on the solvent side of 40S. The density for the Ltv1-Enp1-Rps3 subcomplex on pre-40S overlaps the location of the latch formation, indicating Ltv1 and/or Enp1 prevent the opening of the mRNA channel (**Figure 3.17**).

The decoding site nucleotides, A1755, A1756 and G577, are on the top of helix 44, near the platform, in 40S subunits. The generation of the pre-40S 3D model with MDFF reveals the kink characteristic of premature subunits may disrupt the formation of the decoding site by increasing the distance between the decoding site nucleotides. During translation initiation, all three nucleotides interrogate the mRNA/tRNA duplex, which may be prevented by the pulling of the nucleotides away from one another in premature 40S subunits. In precursor 40S, Tsr1 stabilizes the premature, kinked state of helix 44, potentially preventing the formation of an active decoding site in premature 40S subunits.

In this thesis I have shown that, in addition to their functions in proper assembly of mature 40S subunits, AFs also prevent the incorporation of precursor 40S

subunits into 80S complexes. I have also shown that the binding of Dim1, Rio2, Tsr1, and Nob1 prevent the binding of initiation factors: eIF1, eIF1A and eIF3. Dim1, Rio2 and Tsr1 block residues responsible for bridging interactions with 60S and the mRNA channel. Additionally, I have revealed the presence of the Ltv1-Rps3-Enp1sub-complex prevents the opening of the mRNA channel. Lastly, we hypothesized that the stabilization of helix 44 by Tsr1 prevents the inappropriate incorporation of pre-40S into 80S complexes, which results in rapid degradation of the aberrant 80S ribosomes.

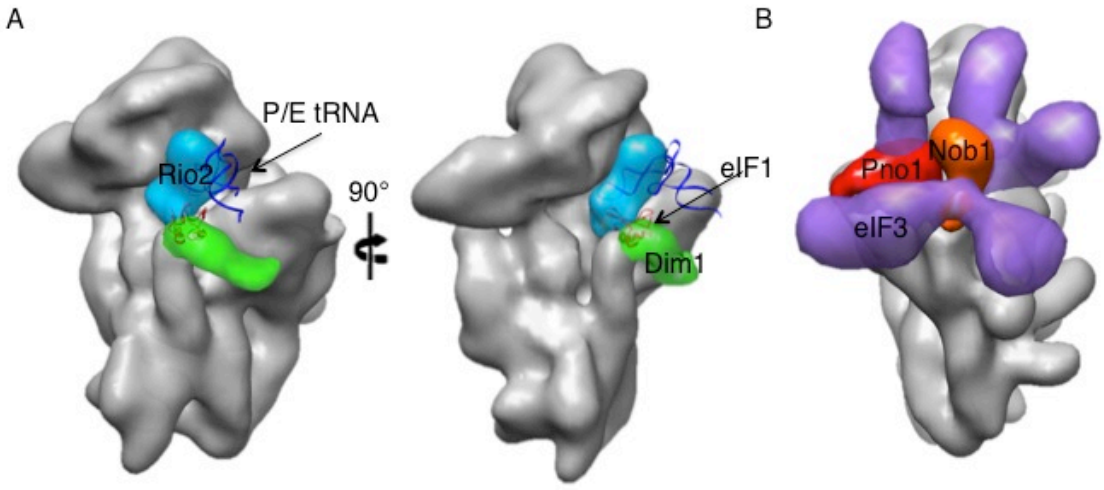


Figure 3.16: AFs obstruct binding sites for factors required for translation. (A) Rio2 (blue) and Dim1 (green) block binding of eIF1 (red ribbon) and P/E tRNA (dark blue ribbon) on pre-40S. (B) Nob1 (red) and Pno1 (orange) block the binding of eIF3 (purple) on pre-40S).

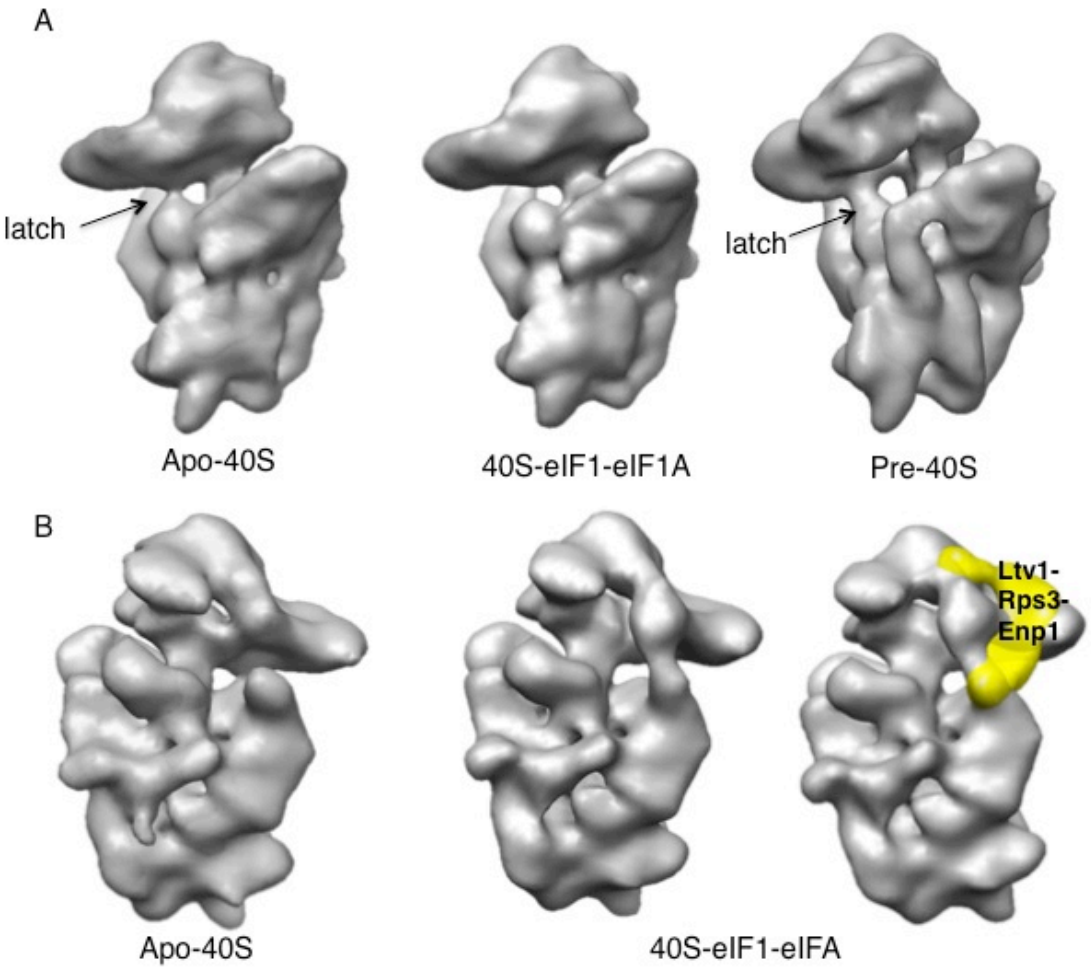


Figure 3.17: The binding of AFs inhibits the opening of the mRNA channel. (A) The latch between helix 18 and helix 34 closes the mRNA channel in mature apo-40S (right) until eIF1 and eIF1A bind cooperatively and open the latch (center). In pre-40S the latch is present, obstructing the mRNA channel (right). (B) On the solvent interface, a bridge of density not present in the apo-40S (left) is formed upon opening of the mRNA channel by eIF1 and eIF1A binding (center), docking of Ltv1-Rps3-Enp1 density onto the 40S-eIF1-eIF1A map shows the sub-complex blocks the formation of this density (right).

3.5 Concluding remarks

Although a vast increase in data will be required to fully understand the process of 40S biogenesis, the results and conclusions of this chapter provide a good starting point for future investigations. Prior to this study, AF protein interaction data was limited to data from investigations using isolated recombinant proteins. The binding observed in purified proteins may lack significance in the context of the pre-40S due to changes in conformation, binding surfaces and solvent exposed regions that occur when the proteins are associated with pre-40S. By understanding their 3D interactions with cryo-EM 3D studies, we can focus on interactions between proteins or specific regions of proteins relevant on pre-40S. This data show a network of spatial AF interactions, in the area extending from the center to the back of the platform. The study shows interactions of Rio2 with Dim1, and Pno1 with Nob1, and indicates the potential domains of these AFs responsible for interactions by docking of their crystal structures into our pre-40S 3D map. This information can be used to design biochemical experiments to further understand how the activities of Rio2 and Dim1, required for site D cleavage, are transmitted as signals to activate Nob1. The determination of the function of Tsr1 to stabilize helix 44 in its premature state requires further studies of the specific sites of interactions between the AF and rRNA. Also, further

understanding of the events leading to Tsr1 dissociation from pre-40S will provide clues about the terminating step(s) of 40S maturation. Investigation of the role of Tsr1 in preventing the joining of premature 40S with 60S will be discussed more thoroughly in Chapter 4.

3.6 Acknowledgements

Data from this chapter is contained in the manuscript "Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates" which, as of 5/13/2011, is in review in for publication in *Science.* Nob1-depl, Rio2-depl, Tsr1-depl and *ΔLtv1* pre-40S particles were isolated by Bethany Strunk (Karbstein Lab) who also carried out SDS-PAGE analysis of the particles. Justin Schilling carried out the antibody labeling experiments and data analysis of pre-40S labeled complexes. Melody Campbell (Karbstein lab) provided the purified rTsr1 and MBP-rTsr1 and Justin Schilling performed negative stain EM analysis on these proteins. Sample preparation, image acquisition and data analysis for cryo-EM structures of depletion/deletion particles were carried out primarily by Cherisse Loucks with technical assistance from Min Su. Mass spectrometry analysis for Nob1-depl and delta-Ltv1 was carried out by Dr. Henriette Remmer and Amber Peariso at the Protein Structure Facility at the

University of Michigan. MS analysis of Rio2-depl and Tsr1-depl was done by Dr. Michael Chalmers and Caitlin Steckler at the Proteomics Core at Scripps Florida.

3.7 References

1 Jakovljevic, J. *et al.* The carboxy-terminal extension of yeast ribosomal protein S14 is necessary for maturation of 43S preribosomes. *Molecular cell* **14**, 331-342 (2004).

2 Mulder, A. M. *et al.* Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science* **330**, 673-677, doi:10.1126/science.1193220 (2010).

3 Campbell, M. G. & Karbstein, K. Protein-Protein Interactions within Late Pre-40S Ribosomes. *PLoS One* **6**, e16194, doi:10.1371/journal.pone.0016194.

4 Schafer, T. *et al.* Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441**, 651-655, doi:nature04840, 10.1038/nature04840 (2006).

5 Pulicherla, N. *et al.* Structural and functional divergence within the Dim1/KsgA family of rRNA methyltransferases. *Journal of molecular biology* **391**, 884-893, doi:10.1016/j.jmb.2009.06.015 (2009).

6 Connolly, K., Rife, J. P. & Culver, G. Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Molecular microbiology* **70**, 1062-1075, doi:10.1111/j.1365-2958.2008.06485.x (2008).

7 Xu, Z., O'Farrell, H. C., Rife, J. P. & Culver, G. M. A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nat Struct Mol Biol* **15**, 534-536, doi:nsmb.1408, 10.1038/nsmb.1408 (2008).

8 Wu, H., Dong, A., Zeng, H., Loppnau, P., Sundstrom, M., Arrowsmith, C., Edwards, A., Bochkarev, A., Plotnikov, A. unpublished data.

9 LaRonde-LeBlanc, N. & Wlodawer, A. Crystal structure of A. fulgidus Rio2 defines a new family of serine protein kinases. *Structure* **12**, 1585-1594, doi:10.1016/j.str.2004.06.016, S0969212604002539 (2004).

10 Jia, M. Z., Horita, S., Nagata, K. & Tanokura, M. An archaeal Dim2-like protein, aDim2p, forms a ternary complex with a/eIF2 alpha and the 3' end fragment of 16S rRNA. *J Mol Biol* **398**, 774-785, doi:S0022-2836(10)00330-X, 10.1016/j.jmb.2010.03.055.

11 Woolls, H. A., Lamanna, A. C. & Karbstein, K. Roles of Dim2 in ribosome assembly. *J Biol Chem* **286**, 2578-2586, doi:M110.191494, 10.1074/jbc.M110.191494.

12 Jia, M. Z., Ohtsuka, J., Lee, W. C., Nagata, K. & Tanokura, M. Crystal structure of Dim2p: a preribosomal RNA processing factor, from Pyrococcus horikoshii OT3 at 2.30 A. *Proteins* **69**, 428-432, doi:10.1002/prot.21381 (2007).

13 Seiser, R. M. *et al.* Ltv1 is required for efficient nuclear export of the ribosomal small subunit in Saccharomyces cerevisiae. *Genetics* **174**, 679-691, doi:genetics.106.062117, 10.1534/genetics.106.062117 (2006).
14 Brink, J. *et al.* Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure* **12**, 185-191, doi:10.1016/j.str.2004.01.015 (2004).
15 Menetret, J. F. *et al.* Architecture of the ribosome-channel complex derived from native membranes. *Journal of molecular biology* **348**, 445-457, doi:10.1016/j.jmb.2005.02.053 (2005).
16 Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953-961, doi:10.1002, (1998).
17 Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J* **22**, 1370-1380, doi:10.1093/emboj/cdg121 (2003).
18 Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-199, doi:S1047-8477(96)90030-1, 10.1006/jsbi.1996.0030 (1996).
19 Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc* **146**, 113-136 (1987).
20 Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J Struct Biol* **157**, 117-125, doi:S1047-8477(06)00169-9, 10.1016/j.jsb.2006.05.004 (2007).
21 Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys Rev* **2**, 21-27, doi:10.1007/s12551-009-0026-3.
22 Granneman, S., Petfalski, E., Swiatkowska, A. & Tollervey, D. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO J* **29**, 2026-2036, doi:emboj201086, 10.1038/emboj.2010.86.
23 Venema, J. & Tollervey, D. Ribosome synthesis in Saccharomyces cerevisiae. *Annual review of genetics* **33**, 261-311, doi:10.1146/annurev.genet.33.1.261 (1999).
24 Lamanna, A. C. & Karbstein, K. Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proc Natl Acad Sci U S A* **106**, 14259-14264, doi:0905403106, 10.1073/pnas.0905403106 (2009).
25 Loar, J. W. *et al.* Genetic and biochemical interactions among Yar1, Ltv1 and Rps3 define novel links between environmental stress and ribosome

biogenesis in Saccharomyces cerevisiae. *Genetics* **168**, 1877-1889, doi:10.1534/genetics.104.032656 (2004).

26 Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C. & Raue, H. A. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in Saccharomyces cerevisiae. *J Biol Chem* **278**, 22537-22545, doi:10.1074/jbc.M300759200, M300759200 (2003).

27 Passmore, L. A. *et al.* The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Molecular cell* **26**, 41-50, doi:10.1016/j.molcel.2007.03.018 (2007).

28 Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A. & Nogales, E. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310**, 1513-1515, doi:10.1126/science.1118977 (2005).

29 Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897-902, doi:10.1126/science.1060612 (2001).

Chapter 4
Tsr1 Prevents Precursor 40S Subunits from binding 60S Subunits

4.1 Abstract

Tsr1 is a 91-kilodalton protein with a hypothesized GTPase function, based on a weak sequence homology to the bacterial GTPase elongation factor Tu (EF-Tu). Tsr1 associates with pre-40S and is required for generation of the mature 18S rRNA found in 40S subunits. In Chapter 3, we investigated the location of Tsr1 on pre-40S by determining the 3D structure of Tsr1-depl pre-40S particles using cryo-EM. Interestingly, the 3D map of these particles was the only, of the five presented in this work, with a conformation of helix 44 different than the other 3D maps. The second interesting observation concerning Tsr1-depl is the presence of 60S rps and rRNA in the preparation at levels much greater than the other four pre-40S preparations described in this thesis. The detection of 60S peptides and rRNA indicates that Tsr1-depl are capable of joining with 60S subunits. Particles similar in size and shape to 80S complexes in the Tsr-depl preparation were observed in negative stain and cryo-EM images. To further characterize these 80S-like particles, we calculated a 3D reconstruction of these particles. Although the low resolution of this 3D map limits detailed structural analysis of these aberrant 80S holoenzymes, the confirmation of their presence has allowed the

proposal of a novel function of Tsr1 on pre-40S, to inhibit the binding of precursor 40S subunits with 60S.

4.2 Results

Northern blotting with a probe to detect the 25S rRNA transcript found in mature 60S confirmed that Tsr1-depl preparations contain 25S rRNA (**Figure 4.1**). Additionally, MS of pre-40S and Rio2-depl preparations contain an 8-10 ratio of small to large ribosomal proteins, while the Tsr1-depl contains nearly equal amounts of the protein populations **(Table 4.1).** Negative stain and cryo-EM micrographs of Tsr1-depl preparations show particles similar in size and shape to 80S complexes (will refer to these particles as Tsr1-depl 80S-like) **(Figure 4.2).** Although depletion of both Pno1 and Tsr1 are exclusive to Tsr1-depl, the observation that Pno1, but not Tsr1, is present in polysomes, indicates Tsr1 is the factor responsible for inhibition of 80S formation by precursor 40S subunits.

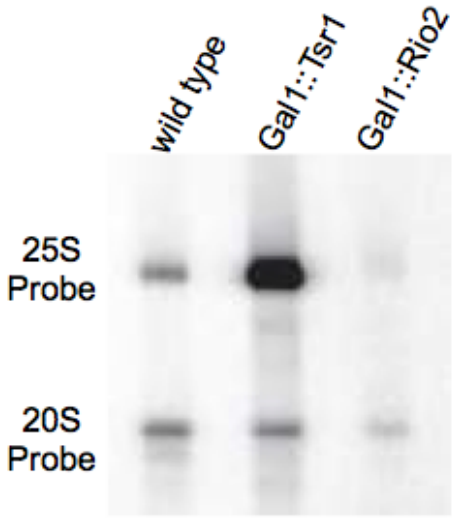


Figure 4.1: Northern blot analysis of Tsr1-depl pre-40S. The Tsr1-depl particles contain a much greater amount of the 25S rRNA found in mature 60S as compared to the WT and Rio2-depl pre-40S particles.

	40S rps peptide #	60S rps peptide #	Ratio
Pre-40S	3583	347	**11.1**
Rio2-depl	3747	467	**8.0**
Tsr1-depl	2239	1526	**1.5**

Table 4.1: Tsr1-depl pre-40S contain significant amounts of 60S rps. WT and Rio2-depl contain an 8 to 10 fold greater amount of small subunit rps compared to large subunit rps while Tsr1 pre-40S contain a nearly 1:1 ratio.

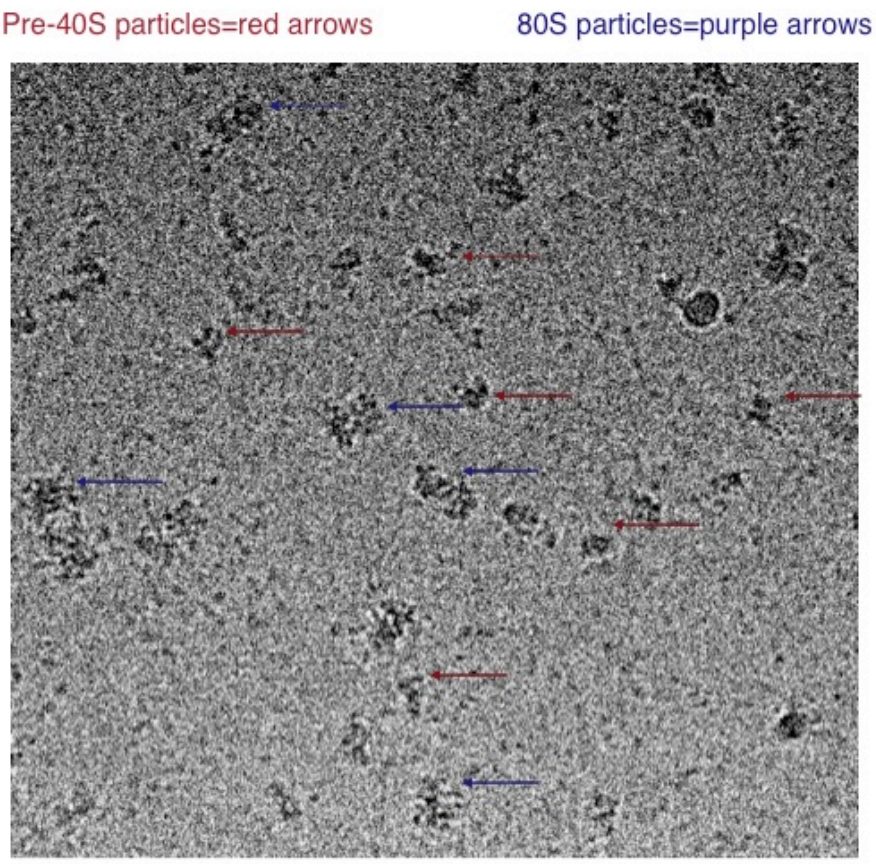


Figure 4.2: Raw cryo-EM micrograph of Tsr1-depl shows 80S-like particles. The raw cryo-EM image reveals particles similar to the pre-40S found in all pre-40S preparation investigated (red arrows) as well as larger particles similar in size and shape to 80S complexes (purple arrows).

To further characterize Tsr1-depl 80S-like, Tsr1-depl particle projection dataset was quantitatively separated using multiple-reference supervised alignment[1,2]. The initial models for separation were our pre-40S 3D map and the published 8.9 Å yeast 80S 3D map[3] (EMDB: 1345). After the first iteration, the subpopulation of projections assigned to the 80S reference (based on CC values) was

approximately 10% of the total dataset. This was consistent with our observation that Tsr1-depl micrographs contained about 10% of particle projections appearing similar to 80S complexes, with the remainder of the particles containing characteristic size and features of pre-40S. The subpopulation of projections assigned to the 80S initial reference was refined iteratively with the 80S 3D map as a single initial reference. Due to the small number of projections (less than 800), and potential flexibility and conformation heterogeneity, the resolution of this 3D map is low, estimated at lower than 30 Å**.** Despite its low resolution, the Tsr1-depl 80S-like 3D map appears similar to WT 80S **(Figure 4.3)**. In the Tsr1-depl 80S-like 3D map the solvent side of the 40S reveals the characteristic 40S head, beak, and foot**.**

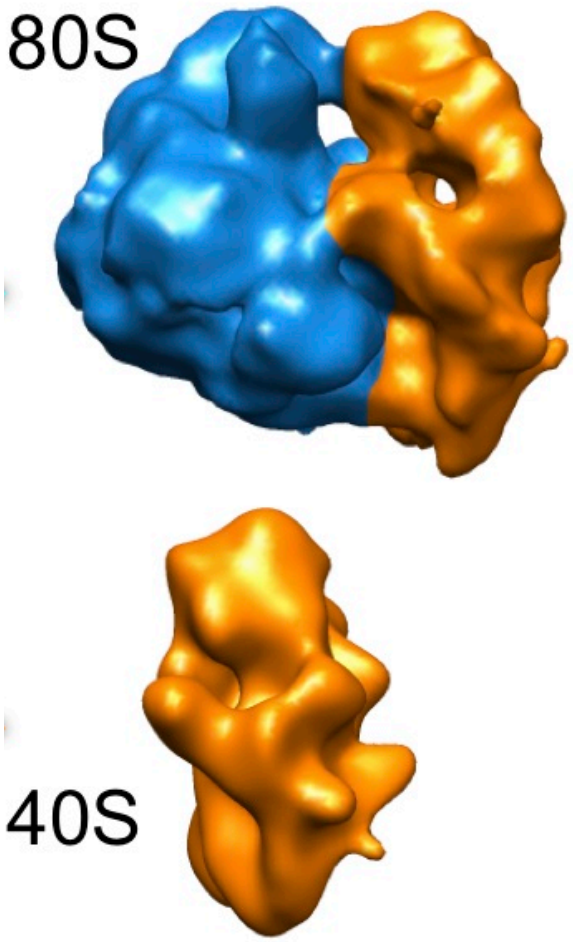


Figure 4.3: Cryo-EM 3D map of 80S-like particles from Tsr1-depl pre-40S. After multiple-reference refinement, the 3D reconstruction of the particle projections assigned to the 80S reference model, the portion in blue corresponds to the bound 60S and orange is the Tsr1-depl 40S. The Tsr1-depl 3D map is shown alone for reference.

4.3 Materials and Methods

Tsr1-depl 80S 3D reconstruction

Tsr1-depl Ltv1-TAP particles were isolated and analyzed via SDS-PAGE and MS as described in Section 3.3. Multiple reference-supervised alignment[1,2] separated the particle projections in Tsr1-depl based on similarity (using CC values) to pre-40S or 80S, using our 18 Å pre-40S 3D map and the 8.9 Å 3D map of the 80S complex from *Thermomyces lanuginosus*[3] (EMDB ID: 1346), after filtering the 3D map to a resolution of 18 Å, as initial volume references. Initial models must be filtered to similar resolution values to prevent a bias toward the higher resolution model. The low-pass Gaussian filtering was carried out using EM-Bfactor software[4]. The initial Tsr1-depl dataset contained 9484 projections, and 759 of these projections were assigned to the 80S reference after the first iteration. To confirm that Tsr1-depl pre-40S are capable of binding 60S, and investigate the morphology of Tsr1-depl-80S-like particles, we calculated a 3D reconstruction of this projection subset using the 8.9 Å 80S 3D map[3] as the initial model. Parameters were similar to those used for the WT pre-40S reconstruction, described in Chapter 2. The final 3D reconstruction was estimated, based on structural features, to be lower than 30 Å in resolution, but the resolution was not determined quantitatively because the low number of projections from this

asymmetric particle prevent any confidence in 3D reconstructions from only half the number of particles.

4.4 Discussion

I believe that the most surprising and novel finding presented in this thesis is the ability of Tsr1-depl particles to join with 60S and form 80S-like complexes. This strongly suggests that the presence of Tsr1 on pre-40S is sufficient to prevent pre-40S incorporation into 80S complexes. Tsr1 depletion in yeast cells results in 20S rRNA accumulation[5], but additional functional information about the GTPase-like protein has not been obtained. This work shows that the location of Tsr1 allows direct interaction between the elongated structure of the protein and helix 44. The premature kinked conformation of helix 44 in pre-40S is not present when Tsr1 is depleted, and pre-40S without Tsr1 are able to bind 60S. These data, along with the observation that Tsr1 is occluded from polysomes[5], underscores the function of Tsr1 as a key indicator for determination of the maturation state of pre-40S by mature 60S. Complex formation between 60S and 40S requires the dissociation of Tsr1, which likely happens after full 40S maturation. The formation of Tsr1-depl-80S like particles also indicates that the loss of Tsr1 occurs after site D cleavage, or that cleavage is not required to signal for the particles' competence to join 60S, as Tsr1-depl contain the 20S

rRNA precursor. This indicates that the dissociation of Tsr1 is one of the final steps in 40S biogenesis. Further understanding the signals leading to Tsr1 dissociation on pre-40S will indicate step(s) at the termination of 40S biogenesis.

The function of an AF holding helix 44 in its premature conformation in ribosomal small subunits may be conserved in bacteria. A candidate protein is the *E. coli* GTPAse YjeQ, which has sequence homology to the C-terminal end of Tsr1 and is required for proper 30S assembly[6-9]. Deletion of YjeQ in bacterial cells results in the accumulation of the precursor to the mature 16S rRNA[7], homologous to the result of depleting Tsr1 in yeast cells. YjeQ associates with premature 30S but studies localizing it on the premature small subunit are contradictory. However, two studies agree with its localization near the A-site located at the top of helix 44[7,10]. These studies also show that its association with 30S disrupts the decoding site. A model proposed by Jomaa *et al.*[11], using premature 30S subunit structural data, indicates that YjeQ is one of the final factors to dissociate from the complex before the 30S and 50S can join. All of these findings reveal a striking similarity in the functions of YjeQ and Tsr1 on the premature small subunit in prokaryotes and eukaryotes, respectively. I hypothesize that YjeQ and Tsr1 are functional homologs and the function of an assembly factor to act as an indicator of the maturation status of the small subunit to the large subunit, is

conserved in ribosome biology from bacteria to eukaryotes. An ortholog to Tsr1 is likely present in human cells, as most AFs are conserved throughout eukaryotes. Preventing the binding or action of YjeQ on premature 30S subunits with chemical inhibitors may result in aberrant ribosome formation or rapid breakdown of 70S complexes, due to the improper incorporation premature 30S complexes into bacterial ribosomes, causing a rapid decrease in growth of bacterial cells. Identifying the potential Tsr1 human ortholog and understanding the differences between YjeQ and this human protein may provide information for the design of inhibitors to be specifically designed to target only the bacterial AF, which could lead to the development of novel antibiotic drugs.

4.5 Acknowledgements

Using the Tsr1-depl dataset described in Chapter 3, data analysis and processing for the Tsr1-depl-80S-like 3D reconstruction was performed by Cherisse Loucks. Northern blotting and SDS-PAGE analysis was done by Beth Strunk.

4.6 References

1 Brink, J. *et al.* Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure* **12**, 185-191, doi:10.1016/j.str.2004.01.015 (2004).

2 Menetret, J. F. *et al.* Architecture of the ribosome-channel complex derived from native membranes. *Journal of molecular biology* **348**, 445-457, doi:10.1016/j.jmb.2005.02.053 (2005).

3 Taylor, D. J. *et al.* Comprehensive molecular structure of the eukaryotic ribosome. *Structure* **17**, 1591-1604, doi:S0969-2126(09)00416-X, 10.1016/j.str.2009.09.015 (2009).

4 Fernandez, J. J., Luque, D., Caston, J. R. & Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of structural biology* **164**, 170-175, doi:10.1016/j.jsb.2008.05.010 (2008).

5 Schafer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J* **22**, 1370-1380, doi:10.1093/emboj/cdg121 (2003).

6 Daigle, D. M. & Brown, E. D. Studies of the interaction of Escherichia coli YjeQ with the ribosome in vitro. *Journal of bacteriology* **186**, 1381-1387 (2004).

7 Himeno, H. *et al.* A novel GTPase activated by the small subunit of ribosome. *Nucleic acids research* **32**, 5303-5309, doi:10.1093/nar/gkh861 (2004).

8 Campbell, T. L., Daigle, D. M. & Brown, E. D. Characterization of the Bacillus subtilis GTPase YloQ and its role in ribosome function. *The Biochemical journal* **389**, 843-852, doi:10.1042/BJ20041873 (2005).

9 Campbell, T. L. & Brown, E. D. Genetic interaction screens with ordered overexpression and deletion clone sets implicate the Escherichia coli GTPase YjeQ in late ribosome biogenesis. *Journal of bacteriology* **190**, 2537-2545, doi:10.1128/JB.01744-07 (2008).

10 Kimura, T. *et al.* Ribosome-small-subunit-dependent GTPase interacts with tRNA-binding sites on the ribosome. *Journal of molecular biology* **381**, 467-477, doi:10.1016/j.jmb.2008.06.023 (2008).

11 Jomaa, A. *et al.* Understanding ribosome assembly: the structure of in vivo assembled immature 30S subunits revealed by cryo-electron microscopy. *RNA* **17**, 697-709, doi:10.1261/rna.2509811 (2011).

Chapter 5
Future Investigations into 40S Biogenesis

5.1 Abstract

In this body of work, I have presented results that will lead to both a better and novel understandings of eukaryotic ribosome biogenesis. The cryo-EM 3D structure of the native pre-40S, the localization of each assembly factor associated at this stage, and the structural and functional characterization of Tsr1, sets the framework for understanding the late cytoplasmic stage of 40S biogenesis and designing the next generation of more detailed experiments to delineate this process. In this chapter, future experiments designed to investigate various structural features of pre-40S and its components in greater detail, define the specific regions of AFs' interaction with themselves, rRNA and rps, and understand signal transduction of signals for site D cleavage, will be discussed.

5.2 Structural investigation of the Ltv1-Enp1-Rps3 subcomplex

Rationale

In this thesis I have shown the pre-40S density attributed to the Ltv1-Enp1-Rps3 subcomplex. Ltv1 was localized to the head and top of the beak, and Enp1 to the bridge of density connecting the beak to the body of the complex. The lack of 3D structures for isolated Ltv1 and Enp1 has inhibited our understanding of their specific locations, regions of interactions and solvent-exposed interfaces of these proteins on pre-40S. By docking the 3D structure(s) of Ltv1 and/or Enp1, or the 3D structure of the entire subcomplex, the interactions of Ltv1 and Enp1 with other components of pre-40S can be determined. The structural characterization of isolated Ltv1 or Enp1 by negative stain EM was unsuccessful in our lab due to flexibility and/or aggregation in preparations of the factors (data not shown). The formation of a trimeric subcomplex that dissociates from pre-40S treated with high-salt concentrations[1] indicates the subcomplex may be stable, and incorporation into the subcomplex may be required for stabilization of the structures of Ltv1 and Enp1, a phenomenon observed with the folding of p27 upon binding to cyclin-dependent kinase/cyclin complexes[2]. To this end, I propose studying the 3D structure of the overall subcomplex, as opposed to Ltv1 and/or Enp1 individually.

Proposed investigation

The 3D structure of the Ltv1-Enp1-Rps3 will be solved using X-ray crystallography on subcomplexes formed *in vitro.* First, yeast proteins of all three components will be recombinantly expressed in *E. coli* cells individually and isolated using affinity purification. The isolated proteins will be incubated in conditions optimized for the formation of the trimeric complex. The proteins will then be passed through a size-exclusion column to remove any unbound proteins or dimeric complexes. This sample will be subjected to crystallographic screens to determine optimal conditions for crystallization of the complex. Upon successful crystallization, the structure of the complex will be solved, ideally, to atomic resolution. The 3D structure of the subcomplex can then be docked into the pre-40S 3D map density attributed to the proteins. This docking will provide further evidence for the locations of the AFs in this density, and will also allow us to determine interactions between the factors and ribosomal proteins and/or RNA. More detailed localization data will also indicated whether Enp1 alone is responsible for the retracted beak conformation, or if Ltv1 and/or Rps3 also play a role in the premature beak structure.

5.3 Structural investigation into Tsr1 and helix 44 interactions

Rationale

In this thesis I have indicated a novel function of Tsr1 in maintaining helix 44 in its premature conformation on pre-40S. Without a high-resolution structure of isolated Tsr1 to dock into its density in the pre-40S 3D map, we are unable to determine the specific domain(s) and residue(s) involved in this stabilization. Information about these interactions may allow us to determine if the region of Tsr1 which binds to helix 44. Additionally, Tsr1 contains a domain with sequence homology to GTPase proteins. An atomic resolution structure of Tsr1 will also allow us to determine if the protein contains a contains a fold similar to known GTPases. If an active GTPase domain exists, the docking of the 3D structure of Tsr1 into the 3D map may provide clues to the GTPase substrate of Tsr1.

Proposed investigation

In this thesis, I have shown that yeast Tsr1 expressed recombinantly in *E. coli* cells can be successfully isolated from cells by His-tag affinity purification. The optimal conditions for the crystallization of the protein will be assayed using crystallographic screens. If crystallization and high-resolution structural determination are successful, the 3D structure will be docked as a rigid body into our pre-40S 3D map. The interaction of the docked Tsr1 and helix 44 from our MDFF model will reveal regions responsible for stabilization of helix 44 by Tsr1.

Pre-40S complexes isolated from yeast strains with specific regions of Tsr1 expected to interact with helix 44 will be mutated or deleted. Cryo-EM 3D reconstructions of these pre-40S subunits will reveal whether the helix is in its premature conformation, indicating the specific regions on Tsr1 that stabilize kinked helix conformation.

5.4 Investigating of signal transduction between Rio2 and Nob1

Rationale

Protein-protein interaction studies show Nob1 binds Pno1, Rps5, and Rps14[3] and all three proteins are required for site D cleavage by Nob1[4-6]. Nob1 interacts with each of these factors on pre-40S. The activity of Rio2[7] is required for cleavage by Nob1, but our model does not allow indicate interaction of Rio2 with Nob1. Determination of how the signal is passed from Rio2 to Nob1 will implicate proteins, such as Pno1, Rps5 and Rps14, that are responsible for the signal transduction. This knowledge will also provide clues to the order of maturation events and the steps leading to the cleavage at site D, one of the final steps in the maturation of 40S subunits.

Proposed investigations

Determining the phosphorylation substrate of Rio2 will allow an understanding of the activity of Rio2 signals to Nob1. Phosphorylation states of proteins in cells

from yeast strains expressing WT and kinase-dead Rio2 could be compared. Protein bands present in the WT but not kinase-dead strain could be analyzed with mass spectrometry to determine potential targets of Rio2 phosphorylation. Rps4, Rps14 and Rps15 are likely candidates, as they are each required for cleavage by Nob1. Proteins that are phosphorylated only in the presence of active Rio2 can be analyzed with mass spectrometry for identification and determination of the serine residue that is phosphorylated. Yeast strains can be generated with the serine residues in the potential target(s) conservatively mutated to cysteines. Western blotting can be carried out as described above and if the band corresponding to the protein with the mutated serine is lost in the WT Western blot, we can predict this is a substrate of Rio2. Northern blotting can then be used to determine whether the cells containing the mutated serine residues accumulate 20S rRNA, indicating an inhibition of cleavage by Nob1. If the activity of Nob1 is decreased in these cells, it can be predicted the mutated protein is located in the signaling pathway upstream of Nob1 and downstream of Rio2.

5.5 Discussion

Determination of the 3D structure of the pre-40S and localization of each associated assembly factor has answered many questions that would be difficult

or impossible to understand without structural biology techniques. This body of works has shown that merging of results from cryo-EM, X-ray crystallographic, MDFF and biochemical experiments allows a more complete understanding of the structure of these large macromolecular intermediates to mature 40S. Determining the 3D maps of pre-40S with one or more AFs depleted in the particles allowed us to determine the binding sites all seven associated AFs on pre-40S. 3D maps of pre-40S particles with Tsr1 depleted and Ltv1 deleted also revealed structural changes, indicating AFs roles in the maintenance of premature conformation of precursor 40S. Conducting additional structural and biochemical experiments will improve our understanding of the biogenesis of 40S, potentially elucidating the interactions, functions and substrates of AFs. Investigations proposed in this chapter provide a minute example of possible experiments that could be designed based on the findings presented in this work. Results of future experiment will, ideally, allow in an expansion of the knowledge of the overall process of ribosome assembly.

5.6 Acknowledgements

I'd like to acknowledge Georgios Skiniotis, Katrin Karbstein, Bethany Strunk, Justin Schilling, Min Su, Gerwin Westfield and Lily Mancour for insightful conversation and advice throughout my work on this project.

5.7 References

1 Schafer, T. *et al.* Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441**, 651-655, doi:nature04840, 10.1038/nature04840 (2006).

2 Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J. & Pavletich, N. P. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**, 325-331, doi:10.1038/382325a0 (1996).

3 Campbell, M. G. & Karbstein, K. Protein-Protein Interactions within Late Pre-40S Ribosomes. *PLoS One* **6**, e16194, doi:10.1371/journal.pone.0016194.

4 Jakovljevic, J. *et al.* The carboxy-terminal extension of yeast ribosomal protein S14 is necessary for maturation of 43S preribosomes. *Molecular cell* **14**, 331-342 (2004).

5 Mulder, A. M. *et al.* Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science* **330**, 673-677, doi:10.1126/science.1193220 (2010).

6 Vanrobays, E., Gelugne, J. P., Caizergues-Ferrer, M. & Lafontaine, D. L. Dim2p, a KH-domain protein required for small ribosomal subunit synthesis. *RNA* **10**, 645-656 (2004).

7 Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C. & Raue, H. A. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in Saccharomyces cerevisiae. *J Biol Chem* **278**, 22537-22545, doi:10.1074/jbc.M300759200, (2003).

8 Lafontaine, D., Vandenhaute, J. & Tollervey, D. The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. *Genes & development* **9**, 2470-2481 (1995).