


THE UNIVERSITY
*of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.



THE UNIVERSITY
*of* EDINBURGH

# Probing ribosomal RNA structural rearrangements: a time lapse of ribosome assembly dynamics

Elena Burlacu

Thesis submitted for the degree of
Doctor of Philosophy

The University of Edinburgh
January 2016

# DECLARATION

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise. The work has not been submitted for any other degree or professional qualification. The results and their discussion presented in chapter 3 have been published (Hector ED, Burlacu E et al, 2014), which is included at the end of this thesis, the citation is as follows:

Hector RD, Burlacu E, Aitken S, Le Bihan T, Tuijtel M, Zaplatina A, Cook AG & Granneman S (2014) **Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution.** *Nucleic Acids Res. 42: 12138–12154. doi: 10.1093/nar/gku815*

Elena Burlacu, January 2015

# Acknowledgements

First and foremost I would like to thank my supervisor Sander Granneman. Uit de grond van mijn hart wil ik je bedanken voor onzesamenwerking en de tijd in jouw lab. Thank you for all the enthusiasm and energy invested in mentoring and forming the scientist that I am today. It is privilege to be your first PhD student! I deeply appreciate you sharing your scientific wisdom. Your constant encouragement and contagious optimism made me believe I can achieve the impossible. The fact that I can now write Python scripts is indubitable proof. The deep “ribosomal” discussions were a lot of fun. I will thoroughly miss them.

I would like to thank the amazing Granneman lab team members (past and present) Thank you Ralph, Rob, Ira and Erica! Your warm friendship, constant support, humour and loud chatting coloured the way too many Scottish grey and rainy days.

I am grateful to Atlanta Cook and David Tollervey who as part of my thesis committee offered valuable suggestions and provoked me to look at my research from different perspectives. I would also like to thank the Tollervey and Beggs lab members for useful tips and discussions, especially Elisabeth Petfalski (for help with FISH protocol) and David Barrass (for help with the 4-thio-uracil pulse chase protocol). Many thanks to our collaborators: Thierry Le Bihan, Carolina Aguilar and Marlene Oeffinger (mass spectrometry), Stuart Aitken ($TCP^{EM}$Algorithm), Philipp Milkereit (Rps17 constructs), Fredrik Lackman and Lars Wieslander (35S pre-rRNA purification). This study would not have been complete without your help. I am grateful to the Wellcome Trust for funding my PhD.

Thank you Johanna, Nony, Andy, Adina, Gratiela and Gabi for your support, for the long chats and even longer phone calls.

Thank you Mom, Dad, Tudor for being there for me. I would have never made it without you! Mulțumesc mama, tata și Tudor! Daca am ajuns până aici este pentru că m-ați susținut și încurajat la fiecare pas. Labai aciū Regis for going through everything with me, for your patience and constant encouragement and for all the late time point walks to the lab.

# Lay summary

The ribosome is the molecular machine responsible for the synthesis of the building blocks of the cell: proteins. In eukaryotes the ribosome is formed by two subunits: the 60S and the 40S. The biogenesis of these subunits is a very complex and energy costly process that involves over 200 non-ribosomal factors, apart from the ribosomal proteins themselves. The assembly of the ribosome is a very tightly regulated process, as errors lead to cell death. However, in the 40S subunit, several ribosomal proteins “mistake” their assembly place, “correcting” it only after the 40S subunit is fully processed. This study showed that this initial “mistake” buys the 40S time to assemble properly before the ribosomal proteins adopt their final position. Rps17 is one of the ribosomal proteins that is initially misplaced. Further analysis into its function revealed that it might play a role in the transport of the 40S intermediate complexes through the cell. Moreover, this study generated snapshots of early assembly stages of the large 60S subunit intermediates. Put together, these snapshots form a time lapse of early 60S that reveals novel insights in ribosome biogenesis regulation.

# Abstract

Ribosome synthesis is a very complex and energy consuming process in which pre-ribosomal RNA (pre-rRNA) processing and folding events, sequential binding of ribosomal proteins and the input of approximately 200 *trans*-acting ribosome assembly factors need to be tightly coordinated. In the yeast *Saccharomyces cerevisiae*, ribosome assembly starts in the nucleolus with the formation of a very large 90S-sized complex. This ~2.2MDa pre-ribosomal complex is subsequently processed into the 40S and 60S assembly intermediates (pre-40S and pre-60S), which subsequently mature largely independently. Although we have a fairly complete picture of the protein composition of these pre-ribosomes, still very little is known about the rRNA structural rearrangements that take place during the assembly of the 40S and 60S subunits and the role of the ribosome assembly factors in this process. To address this, the Granneman lab developed a method called ChemModSeq, which made it possible to generate nucleotide resolution maps of RNA flexibility in ribonucleoprotein complexes by combining SHAPE chemical probing, high-throughput sequencing and statistical modelling. By applying ChemModSeq to ribosome assembly intermediates, we were able to obtain nucleotide resolution insights into rRNA structural rearrangements during late (cytoplasmic) stages of 40S assembly and for the early (nucleolar) stages of 60S assembly.

The results revealed structurally distinct cytoplasmic pre-40S particles in which rRNA restructuring events coincide with the hierarchical dissociation of assembly factors. These rearrangements are required to trigger stable incorporation of a number of ribosomal proteins and the completion of the head domain. Rps17, one of the ribosomal proteins that fully assembled into pre-40S complexes only at a later assembly stage, was further characterized.

Surprisingly, my ChemModSeq analyses of nucleolar pre-60S complexes indicated that most of the rRNA folding steps take place at a very specific stage of maturation. One of the most striking observations was the stabilization of 5.8S pre-rRNA region, which coincided with the dissociation of the assembly factor Rrp5 and stable incorporation of a number of ribosomal proteins.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . . .I
LAY SUMMARY . . . . . II
ABSTRACT . . . . . III
LIST OF FIGURES . . . . . .VIII
LIST OF TABLES. . . . . XI
ABBREVIATIONS. . . . . XII

**INTRODUCTION . . . . . 1**

1.1. Ribosomal rRNA processing in the yeast *S. cerevisiae* . . . . . *1*
1.2. The pre-40S assembly pathway. . . . . 4
1.3. The pre-60S assembly pathway. . . . . 9
1.4. Nuclear export of pre-40S and pre-60S intermediates . . . . . .18
1.5. Determining RNA secondary structure through structure probing approaches . . . . . .22

**MATERIALS AND METHODS . . . . . 28**

2.1. Yeast media and general solutions . . . . . .28
2.2. Yeast strains . . . . . .28
2.3. Designing the Rps17 mutants . . . . . .30
2.4. Radiolabeling of oligonucleotide probes. . . . . .31
2.5. Northern Blot . . . . . .31
2.6. Immunoprecipitation of FLAG-tagged ribosomal proteins: . . . .32
2.7. Immunoprecipitation and chemical modification of rRNA . . . . .32
2.8. Tandem affinity purifications. . . . . .33
2.9. Primer extension analyses . . . . . .34
2.10. Preparation of ChemModSeq libraries and high-throughput

sequencing . . . . . .34
2.11. Data analysis . . . . . .36
2.12. Western Blot . . . . . .37
2.13. Growth complementation assay . . . . . .37
2.14. 4-thio-uracil pulse chase . . . . . .38
2.15. FISH . . . . . .39
2.16. Fixing cells for fluorescence microscopy . . . . . .39
2.17. *In vitro* binding assay . . . . . .40

**PROBING THE DYNAMICS OF PRE-RRNA FOLDING IN 40S ASSEMBLY INTERMEDIATES . . . . . .41**

3.1. Introduction and aim . . . . . .41
3.2. The 20S and 18S rRNAs can be effectively purified from pre-ribosomes . . . . . .42
3.3. Affinity purified pre-40S particles can be classified into ”early”, “middle” and “late” intermediates . . . . . .44
3.4. Late cytoplasmic pre-40S complexes contain fewer assembly factors . . . . . .47
3.5. Applying ChemModSeq to pre-40S complexes . . . . . .48
3.6. ChemModSeq data reveal a stabilization of the 3’ major domain during late stages of 40S assembly. . . . . .53
3.7. Restructuring of the 3’ major domain coincides with the dissociation of Enp1, Ltv1, Tsr1 and Rio2 and takes place before the association of Rio1 . . . . . .57
3.8. The 3’ major domain of *in vitro* refolded 18S rRNA adopts a similar structure as the 20S pre-rRNA in the late pre-40S intermediates . . . . . .60
3.9. Discussion . . . . . .62

**THE ROLE OF RPS17 IN 40S SUBUNIT MATURATION . . . . 66**

4.1. Introduction and aim . . . . . .66

4.2. The C-terminal region of Rps17 is essential for function and cell viability . . . . . .68

4.3. Rps17 C-terminal Δ100-136 mutant is dominant negative . . . .69

4.4. Rps17 N-terminal domain is essential for ribosome incorporation . . . . . .70

4.5. Is Rps17 misplaced in the pre-40S to accommodate Nob1 binding? . . . . . .71

4.6. The 18S rRNA maturation is affected in Rps17 depletion and Rps17Δ100-136 mutants . . . . . .73

4.7. Rps17 depletion delays the early steps 18S rRNA processing . .76

4.8. Rps17 C-terminal domain has a putative nuclear export signal (NES) . . . . . .78

4.9. The Rps17 NES is a weak NES . . . . . .82

4.10. Does Rps17 NES bind Crm1 *in vitro*? . . . . . .83

4.11. Discussion . . . . . .85

**RIBOSOMAL RNA STRUCTURAL REARRANGEMENTS DURING NUCLEOLAR PRE-60S ASSEMBLY . . . . . .90**

5.1. Introduction and aim . . . . . .90

5.2. 35S, 27SA$_2$ and 27SB pre-rRNA can be efficiently purified from pre-ribosomes . . . . . .92

5.3. Rrp5 and Nsa2 pre-60S intermediates contain a different set of assembly factors . . . . . .93

5.4. Classifying rRNA restructuring events based on ChemModSeq data analysis . . . . . .96

5.5. The 5’ end of the 27S pre-rRNA becomes more structured while

the 3' end becomes more flexible as 60S pre-rRNA processing proceeds . . . . . 101
5.6. The 5.8S undergoes major conformational changes during the conversion of 35S pre-rRNA to 27SB pre-RNA . . . . . 103
5.7. ITS2 undergoes subtle structural changes during $27SA_2$ to 27SB processing . . . . . 107
5.8. The 25S region is more flexible in 27SB compared to $27SA_2$ and 35S pre-rRNAs . . . . . 110
5.9. The 5.8S -25S interaction . . . . . 113
5.10. Examples of rRNA structural rearrangements in the 25S region . . . . . 117
5.11. Discussion . . . . . 121

**CONCLUDING REMARKS . . . . . 127**

**REFERENCES. . . . . 129**

# LIST OF FIGURES

Figure 1.1. The major ribosomal RNA (rRNA) processing pathway in *S .cerevisiae.* . . . . . 3

Figure 1.2. A structure model of the *S. cerevisiae* 18S rRNA. . . . . 4

Figure 1.3. Schematic representation of the pre-40S assembly pathway . . 5

Figure 1.4. A structure model of the S. cerevisiae 25S and 5.8S rRNAs. . 10

Figure 1.5. Schematic representation of the pre-60S assembly pathway. . 11

Figure 1.6. Crm1 mediated export of pre-ribosomal particles. . . . . 19

Figure 1.7. RNA structure probing using SHAPE chemistry. . . . . 27

Figure 3.1 Schematic representation of the pre-40S cytoplasmic assembly . . . . . 42

Figure 3.2. Affinity purification of 20S pre-rRNA precursors and mature 18S rRNA . . . . . 43

Figure 3.3 Pre-40S complexes can be classified into early, mid and late particles. . . . . 45

Figure 3.4 Rio1 associated pre-40S complexes contain fewer 40S assembly factors. . . . . 48

Figure 3.5 ChemModSeq experimental workflow. . . . . 49

Figure 3.6. The pyChemMod data analysis pipeline. . . . . 51

Figure 3.7. The $TCP^{EM}$ algorithm. . . . . 52

Figure 3.8 A digital snapshot of pre-rRNA secondary structures . . . . . 56

Figure 3.9. Ribosomal protein rRNA binding sites are more flexible in early and middle pre-40S complexes. . . . . 59

Figure 3.10. Ribosome assembly factors are required to maintain an open head domain conformation. . . . . .61
Figure 3.11. A model for structural rearrangements during late stages of 40S subunit maturation. . . . . .63
Figure 4.1. Rps17 binds to early pre-40S but it is repositioned in the late pre-40S assembly complexes . . . . . .67
Figure 4.2. Design of Rps17 N-terminal and C-terminal deletion mutants .68
Figure 4.3. The C-terminal Rps17 Δ100-136 deletion mutant is a dominant negative. . . . . .70
Figure 4.4. The first 26 aminoacids in the N-terminal domain are essential for Rps17 incorporation into pre-ribosomes. . . . . .71
Figure 4.5. Nob1 depletion prevents Rio2 dissociation without blocking the maturation pathway. . . . . .73
Figure 4.6. Rps17 is required for efficient 18S rRNA processing . . . . . .75
Figure 4.7. The maturation of 18S rRNA is kinetically delayed in Rps17 depleted cells. . . . . .77
Figure 4.8. Rps17 is not required for cleavage at site D in the cytoplasm. .78
Figure 4.9. Rps17 has a putative NES in the C-terminal domain . . . . . .79
Figure 4.10. The NES sequence in Rps17 is required for 18S rRNA synthesis. . . . . .81
Figure 4.11. The Rps17 NES is a weak NES. . . . . .83
Figure 4.12. Assessing Rps17 Crm1 interaction *in vitro*. . . . . .84
Figure 4.13 The C-terminal domain of Rps17 must remain accessible for optimal nuclear export. . . . . .89

Figure 5.1. Schematic representation of the nucleolar steps of pre-60S assembly . . . . . .91
Figure 5.2. Isolation of 35S, 27SA2 and 27SB pre-rRNAs using specific bait proteins. . . . . .93
Figure 5.3. The distribution of assembly factors in Rrp5 and Nsa2 affinity purified pre-60S intermediates. . . . . .94
Figure 5.4. The ChemModSeq experimental workflow. . . . . .97
Figure 5.5. Classifying rRNA restructuring events. . . . . 100
Figure 5.6. A digital snapshot of the ChemModSeq data. . . . . 102
Figure 5.7. The 5.8S region undergoes a major restructuring event during the conversion of 27SA2 to 27SB pre-rRNA . . . . . 104
Figure 5.8. The ChemModSeq data indicated that ITS2 is more structured than previously thought. . . . . 110
Figure 5.9. Analysis of the ChemModSeq data for each 25S domain. . . 112
Figure 5.10. ChemModSeq reveals changes in 5.8S and 25S regions that form base-pairing interactions. . . . . 115
Figure 5.11. Gradual stabilization of Rpl26, Rpl4, Rpl17 binding. . . . . 116
Figure 5.12 Helixes H35a and H37 are stabilized as the 27SB pre-rRNA is formed. . . . . 119
Figure 5.13 Rpl16, Rpl40 and Rpl20 bring domain II and domain VI together during the conversion of 27SA2 to 27SB pre-rRNA. . . . . 120
Figure 5.14 Model of rRNA structural changes and pre-60S remodelling steps that take place during the nucleolar stages of 60S synthesis. . 126

# LIST OF TABLES

Table 2.1. *Saccharomyces cerevisiae* strains used in this study. . . .29

Table 2.2. Oligonucleotides used for Northern Blot analyses . . . . .31

Table 2.3. Oligonucleotides used for primer extensions . . . . . .34

Table 2.4. Oligonucleotides for library preparation. . . . . .35

# Abbreviations

| | |
|---|---|
| 1M7 | 1-methyl-7-nitroisatoic-anhydride |
| A | Adenine |
| acp-modification | Aminocarboxy propyl modification |
| ATP | Adenosine Triphosphate |
| C | Cytosine |
| cDNA | Complementary DNA |
| CP | Central protuberance |
| Cryo-EM | Cryo-Electron Microscopy |
| CTD | C-terminal domain |
| Da | Dalton |
| DBA | Diamond-Blackfan anemia |
| EtBr | Ethidium Bromide |
| ETS | External Transcribed Spacer |
| FG-rich repeats | Phenylalanine-Glycine rich repeats |
| FISH | Fluorescence *in situ* hybridization |
| G | Guanine |
| GAL | Galactose |
| GFP | Green fluorescent protein |
| GLC | Glucose |
| GTP | Guanosine Triphosphate |
| H89 | Helix 89 |
| HTP tag | His6-TEV-ProtA tag |
| IgG | Immunoglobulin G |
| ITS | Internal Transcribed Spacer |
| mRNA | Messenger RNA |
| MUT | Mutant |
| NAI | 2-methylnicotinic acid imidazolide |
| NES | Nuclear export signal |
| NLS | Nuclear localization signal |
| NMD | Nonsense-mediated decay |
| NPC | Nuclear Pore complex |
| nt | Nucleotide |
| NTD | N-terminal domain |
| PET | Polypeptide exit tunnel |
| PKI | Protein kinase inhibitor |
| Pol | Polymerase |

| | |
|---|---|
| pre-rRNA | Precursor ribosomal RNA |
| R-protein | Ribosomal protein |
| RNP | Ribonucleoprotein particle |
| Rpl | Ribosomal protein of the large subunit |
| Rps | Ribosomal protein of the small subunit |
| rRNA | Ribosomal RNA |
| RT | Reverse transcription |
| SBDS | Shwachman-Bodian-Diamond syndrome |
| SHAPE | Selective 2'-hydroxyl acylation analysed by primer extension |
| SV40 | Simian vacuolating virus 40 |
| T | Thymine |
| TAP tag | Tandem affinity purification tag |
| $TCP^{EM}$ | Two Channel Poisson Expectation Maximization Algorithm |
| TRAMP | Trf4/Trf5/Mtr4 polyadenylation complex |
| U | Uracil |
| WT | Wild type |

# CHAPTER 1

# Introduction

## 1.1. Ribosomal rRNA processing in the yeast *Saccharomyces cerevisiae*

Ribosome synthesis involves the tight coordination of rRNA processing and folding events, sequential binding of r-proteins and the input of hundreds of *trans*-acting factors that direct the assembly process. Many of the maturation steps are spatially separated and take place in both the nucleus and the cytoplasm. All three polymerases are involved in ribosome synthesis: Pol I transcribes the initial polycistronic 35S pre-ribosomal RNA (pre-rRNA), Pol II transcribes r-protein mRNAs and Pol III transcribes the 5S rRNA. The 5S rRNA is transcribed independently of the 35S pre-rRNA precursor. The 5S rRNA binds ribosomal proteins (r-proteins) Rpl5 and Rpl11 forming the 5S RNP. The 5S RNP then associates with the pre-ribosome containing the 35S pre-rRNA (Zhang *et al*, 2007).

In *S. cerevisiae*, ribosome synthesis starts in the nucleolus with the transcription of the 35S pre-rRNA precursor (fig 1.1). The 35S pre-rRNA contains the sequences for the mature 18S, 25S and 5.8S rRNAs, separated by the internal transcribed sequences (ITS1 and ITS2) and flanked by the 5' and 3' external transcribed sequences (ETS) (fig 1.1). This pre-rRNA is packaged into a large complex referred to as the 90S complex (Schäfer *et al*, 2003; Grandi *et al*, 2002) or the SSU processome (Dragon *et al*, 2002), in which most of the processing steps take place. Cleavage of ITS1 at site $A_2$ leads to the formation of the 20S and 27S pre-rRNAs and the formation of pre-40S and pre-60S assembly intermediates. These ribosomal subunit assembly intermediates mature mostly independently before they form the mature 80S ribosome in the cytoplasm.

The processing of the 35S pre-rRNA starts with the co-transcriptional cleavage of the 5' ETS at sites $A_0$ and $A_1$ (unknown endonucleases) and with the cleavage of ITS1 at site $A_2$, which was proposed to be carried out by Rcl1 (Horn *et al*, 2011)(fig 1.1).

Cleavage at $A_2$ generates the 20S pre-rRNA, which is subsequently exported to the cytoplasm where it is cleaved at site D by the Nob1 endonuclease, yielding the mature 18S rRNA (Fatica *et al*, 2003). Before the 20S pre-rRNA is cleaved at site D, late cytoplasmic pre-40S particles are subjected to a quality control step, which involves joining with a mature 60S to form a translation-like 80S or 80S-like complex (Lebaron *et al*, 2012; Strunk *et al*, 2012). It was proposed that successful formation of the 80S-like complex would indicate that the pre-40S is competent for translation. Subsequently, Nob1 cleaves at site D to remove the remainder of ITS1 (fig 1.1), the 80S-like complex dissociates and the matured 40S subunit can engage in translation.

The pre-60S rRNA processing pathway is even more complex. It involves the processing of two mature rRNAs (5.8S and 25S rRNAs) and the association of a third (5S rRNA), it takes place in all three cellular compartments and involves two parallel processing pathways for 5.8S rRNA maturation (fig 1.1). Most of the pre-rRNA processing steps in the pre-60S take place in the nucleolus. The cleavage of 35S at sites $B_0$ (by Rnt1) and $A_2$ (by Rcl1) leads to the formation of the $27SA_2$ pre-rRNA transcript that includes the mature 5.8S and 25S separated by ITS2 (fig 1.1). The $27SA_2$ can be cleaved at site $A_3$ (by RNAse MRP; (Schmitt & Clayton, 1993)) forming $27SA_3$ pre-rRNA or at site $B_L$ (unknown nuclease), thereby generating the $27SB_L$ pre-rRNA. The $27SA_3$ is then processed by the exonuclease Rat1 to generate the $27SB_S$ pre-rRNA (Henry *et al*, 1994). The 27SB pre-rRNAs are subsequently cleaved at site $C_2$ (by Las1 (Gasse *et al*, 2015)), which generates the 7S and a 5' extend 25S precursor (25.5S). After $C_2$ cleavage, pre-60S particles are exported to the nucleus. Here the 25.5S pre-rRNA is trimmed by Rat1 to generate the mature 25S rRNA (Geerlings *et al*, 2000). Processing of the 3' end of the 5.8S starts in the nucleus, which involves a complex of 3'-5' exonucleases called the exosome; (Allmang *et al*, 1999), and is finalized in the cytoplasm by the Ngl2 nuclease (Thomson & Tollervey, 2010).

The 35S pre-rRNA can also be first cleaved co-transcriptionally at site $A_3$. This generates 23S pre-rRNA (5' end of ETS-$A_3$), 22S pre-rRNA ($A_0$-$A_3$) or 21S pre-rRNA ($A_1$-$A_3$) precursors. These alternative rRNAs are generally described as dead-end pre-rRNAs that accumulate in strains defective in ribosome synthesis (Allmang *et al*, 2000). However, there is evidence suggesting that

some of these pre-rRNA species can be processed into functional 18S rRNA (Henras *et al*, 2015). For example, the 21S rRNA that accumulates in Rrp5 mutants can still be processed into the mature 18S rRNA (Torchet & Hermann-Le Denmat, 2000).

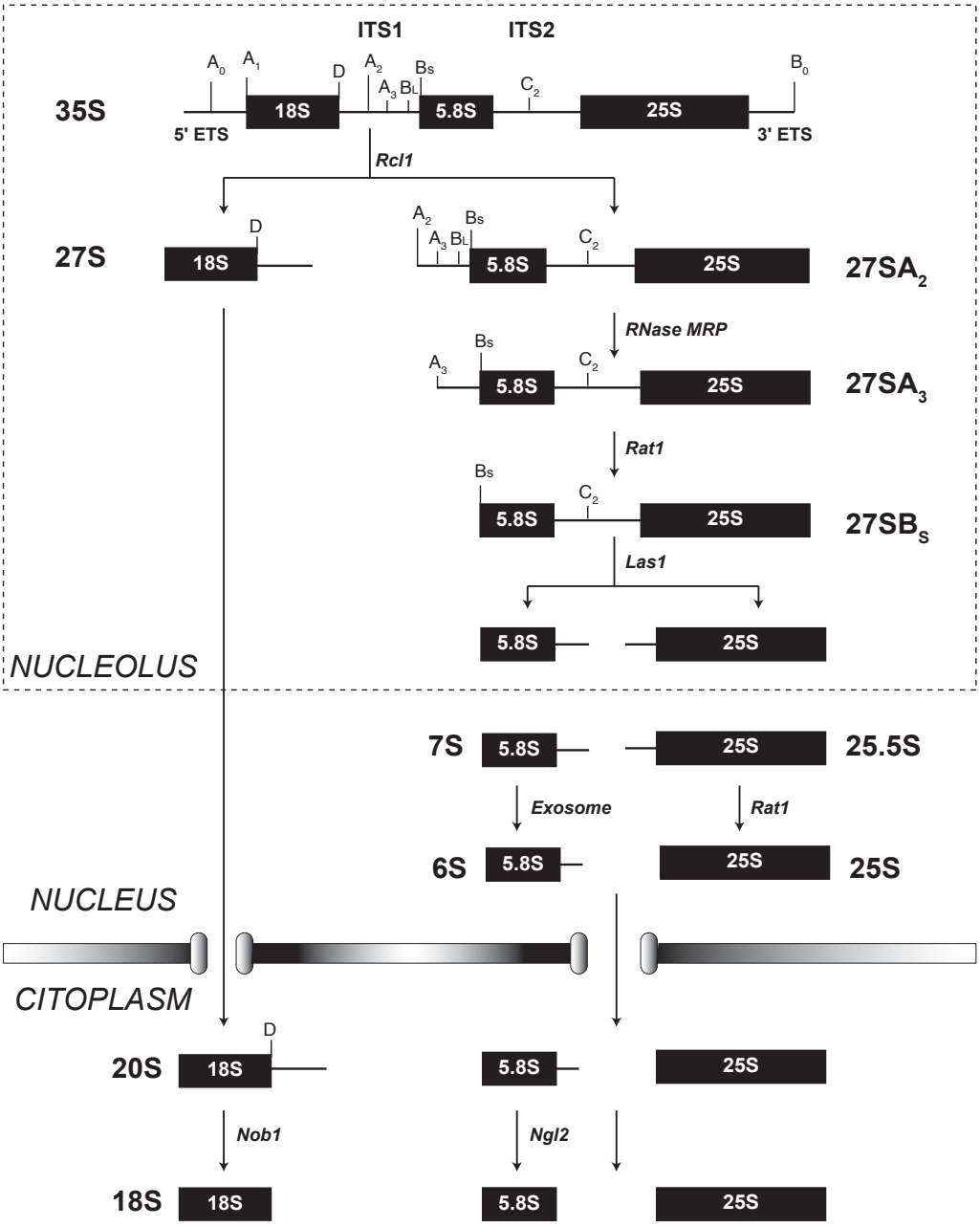



**Figure 1.1. The major ribosomal RNA (rRNA) processing pathway in *S .cerevisiae*.** Shown is a schematic representation of the yeast pre-rRNA processing pathway. Positions of cleavage sites and some of the nucleases involved in pre-rRNA processing are also included.

## 1.2. The pre-40S assembly pathway

The 18S rRNA within the mature 40S complex has a “body”, shoulder”, a “platform” and a “head” with a “beak” (fig 1.2) (Spahn *et al*, 2001). These structures are hierarchically formed in a 5’ to 3’ direction through the binding of r-proteins and the activity of assembly factors. The body is the first to form, which occurs shortly after cleavages at sites $A_0$-$A_2$ in the 35S pre-rRNA. The formation of the head starts in the nucleus and finishes in the cytoplasm with the formation of the beak (Ferreira-Cerca *et al*, 2007; Schäfer *et al*, 2006).

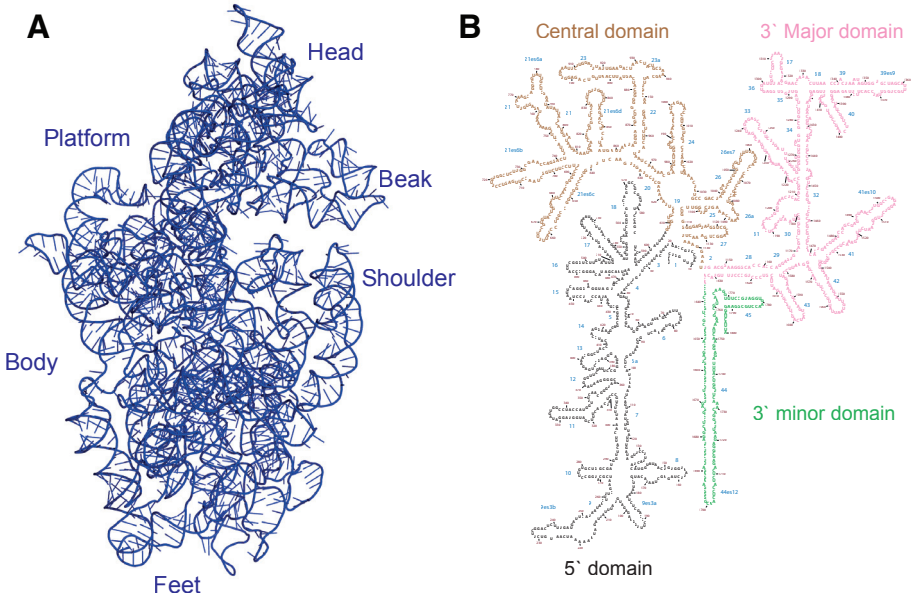



**Figure 1.2. A structure model of the *S. cerevisiae* 18S rRNA.**
(A) The three dimensional structure of the 18S rRNA in which the head, body, platform and beak are highlighted; (B) A secondary structure of the 18S rRNA in which the 3’ major domain (pink), the 3’ minor domain (green), the 5’ domain (black) and the central domain (brown) are highlighted.

Pre-40S complexes contain an almost complete set of r-proteins, however many r-proteins are weakly associated with early pre-40S complexes and they bind more stably as the maturation progresses. Besides playing a role in translation, many r-proteins are also required for (efficient) 40S assembly. For example, depletion of Rps5, Rps18 or Rps9 leads to late 20S pre-rRNA processing defects while Rps15, Rps10, Rps26, Rps3, Rps2, Rps20, Rps0 are needed for nuclear export (Ferreira-Cerca *et al*, 2005). It

was thus proposed that binding of the latter to pre-40S complexes is required for the formation of nuclear export competent complexes. Depletion of Rps5 and Rps15 affects the incorporation of r-proteins required for export, suggesting that assembly of r-proteins occurs in a hierarchical fashion (Ferreira-Cerca *et al*, 2007). R-proteins such as Rps7, Rps30 and Rps31 do not play a role in 40S assembly but are essential for translation. Cells lacking these r-proteins accumulate translation initiation defective 80S ribosomes (Ferreira-Cerca *et al*, 2005).

Many of the pre-40S assembly factors (Enp1, Tsr1, Hrr25, Nob1, Dim1, Dim2, Rio2, Rio1 and Ltv1) assemble during the formation of the 90S complex or shortly after $A_2$ cleavage (fig 1.3) (Schäfer *et al*, 2003). Some of them are required for early $A_0$-$A_2$ rRNA processing steps (Enp1 and Dim1) while others (Hrr25, Nob1, Ltv1, Tsr1, Rio2, Dim2 and Rio1) were shown to be required only for the last processing step, cleavage at site D (Lafontaine *et al*, 1998; Fatica *et al*, 2003; Chen *et al*, 2003; Schäfer *et al*, 2006).

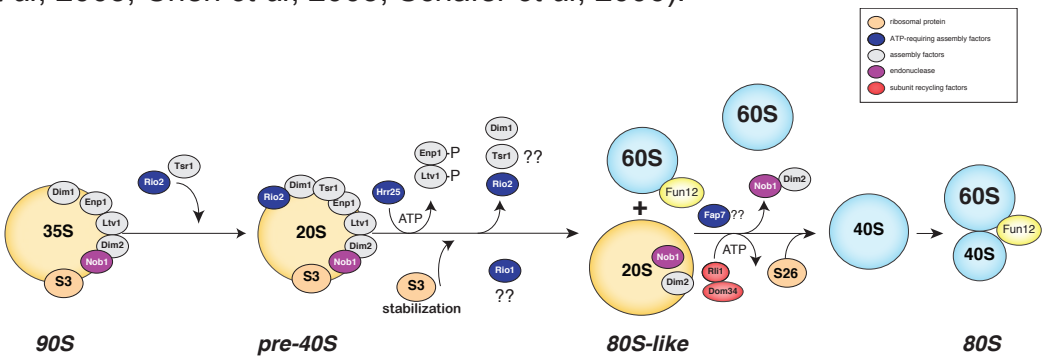



**Figure 1.3. Schematic representation of the pre-40S assembly pathway**
Most of the assembly factors bind to the 90S and accompany the pre-40S maturation until its final (cytoplasmic) steps. In the cytoplasm the pre-40S joins the 60S forming the 80S-like complex, the 20S is cleaved as part of a translation initiation like cycle and then the mature 40S subunit is released.

Recent UV cross-linking and cryo-EM studies have shown that Enp1 binds primarily to the head domain in pre-40S complexes, whereas Ltv1 binds to the head and the body (Granneman *et al*, 2010; Strunk *et al*, 2011). Some Enp1 and Ltv1 binding sites overlap with rRNA regions that are bound by the N-terminus of Rps3, suggesting that binding of these assembly factors blocks some Rps3-rRNA interactions. Consistent with this idea, it was shown that Rps3 is not stably bound to pre-40S complexes, but that it forms a very stable complex with Enp1 and Ltv1 (Schäfer *et al*, 2006). All the available

data suggest that Enp1 and Ltv1 are required for the recruitment of Rps3 to the pre-40S complexes, however, they do not allow Rps3 to adopt its final conformation in the pre-40S particle (Schäfer *et al*, 2006; Strunk *et al*, 2011). Phosphorylation of Ltv1 and Enp1 by the casein kinase Hrr25 triggers the release of these assembly factors and allows Rps3 to stably bind to pre-40S. This early incorporation of Rps3 but delayed stabilization of its binding was shown to be important for pre-40S nuclear export (Schäfer *et al*, 2006).

Rps3 binds the 20S pre-rRNA in close proximity in H33, which in the mature 18S rRNA forms the beak structure (fig 1.1). Its stable incorporation appears to be important for the formation of the beak structure. Premature formation of the beak impairs nuclear export (Schäfer *et al*, 2006; Strunk *et al*, 2011), suggesting that the assembly state of Rps3 is somehow an indicator of export competence. Thus, Ltv1 and Enp1 delay the stable incorporation of Rps3 as well as the formation of the beak structure.

As discussed above, the Hrr25 kinase activity is important for the recycling of Enp1 and Ltv1. Deleting Hrr25 or mutating the active site slows down the recycling of Ltv1 and prevents the formation of the 80S-like complex (Ghalei *et al*, 2015). In conclusion, Ltv1 and Enp1 dissociation is crucial for the formation of the 80S-like complex.

Initially, Ltv1 was thought to regulate pre-40S nuclear export directly through its C-terminal Nuclear Export Signal (NES) (Fassio *et al*, 2010; Seiser *et al*, 2006). Deletion of Ltv1 resulted in nuclear accumulation of a GFP-tagged Rps3 reporter, suggesting that Ltv1 could be an export adapter for pre-40S export (Seiser *et al*, 2006). However, contradicting these results, FISH analyses on strains expressing an Ltv1 protein with an inactive NES, showed that 20S export to the cytoplasm was not noticeably affected (Fassio *et al*, 2010; Merwin *et al*, 2014). This suggests that Ltv1 NES is not required for the pre-40S export but may be required for shuttling of Ltv1 between the nucleus and the cytoplasm. The role of Ltv1 in pre-40S nuclear export is therefore likely indirect, by delaying the stable incorporation of Rps3.

Shortly after being exported into the cytoplasm, the 20S pre-rRNA is dimethylated (A1780-A1781) by Dim1 methyl transferase (Lafontaine *et al*, 1994). Although Dim1 plays an essential role in 40S maturation, its enzymatic activity is not important for pre-rRNA processing (or translation) (Lafontaine *et*

*al*, 1998), suggesting that Dim1 has mainly a structural role in 40S synthesis.

Tsr1 has a GTPase-like domain, but it does not bind or hydrolyse GTP *in vitro* (Lebaron *et al*, 2012). Although Tsr1 is essential for 20S pre-rRNA processing at site D, mutations in the predicted active site had no effect on pre-40S maturation, suggesting that the GTPase-like domain is not essential (Lebaron *et al*, 2012; Léger-Silvestre *et al*, 2004). In cells lacking Tsr1, Dim2 is not incorporated into pre-40S complexes and Ltv1 is not efficiently released (Strunk *et al*, 2011). This implies that Tsr1 has a structural role rather than an enzymatic function in pre-40S assembly.

Rio2 and Rio1 are serine/threonine kinases and belong to RIO family of protein kinases. Rio2 and Rio1 are required for 20S pre-rRNA cleavage at site D and shuttle between the nucleus and the cytoplasm. Rio1 and Rio2 do not co-purify, suggesting that they are part of different pre-40S intermediates (Vanrobays *et al*, 2003; Ferreira-Cerca *et al*, 2014; Vanrobays *et al*, 2001).

Unlike Hrr25, the Rio2 kinase activity does not seem to be required for phosphorylating ribosome assembly factors. Cryo-EM and crystallography studies have indicated that its ATP binding site is masked when bound to the pre-40S. Instead, it was proposed that Rio2 uses its kinase activity to dissociate itself from the pre-40S complexes (Geerlings *et al*, 2003; Ferreira-Cerca *et al*, 2012). It is not yet known precisely what the role of Rio2 is in pre-40S assembly and what triggers its release. As its ATPase activity seems to be important only for its recycling, but not for other assembly steps it could be that, like Tsr1, Rio2 mainly has a structural role in pre-40S assembly. It was proposed that Rio2 release might be linked to conformational changes in the head domain (Ferreira-Cerca *et al*, 2012). It could thus be that its binding or dissociation signals the beginning or the end of specific assembly steps or influences rRNA folding steps.

Rio1 ATP binding and ATPase activity promotes Nob1 mediated cleavage at site D (Turowski & Tollervey, 2014). Dominant negative ATPase mutants block the final maturation step in the 80S-like particles (Ferreira-Cerca *et al*, 2014). Apart from its role in 20S pre-rRNA processing, Rio1 has also a role in rDNA stability and segregation (Iacovella *et al*, 2015).

Dim2/Pno1 also shuttles between the nucleus and the cytoplasm. Cells lacking Dim2 accumulate the 35S pre-rRNA and have reduced levels of 20S

and 27S pre-rRNAs, suggesting that it is required for $A_2$ cleavage (Vanrobays *et al*, 2004). Dim2 directly binds Nob1 in the cytoplasm and influences the stability of Nob1-rRNA interactions and consequently its ability to cleave at site D (Woolls *et al*, 2011) (Fatica *et al*, 2003). Nob1 cleavage at site D takes place shortly after the formation of the 80S-like complex and is enhanced by two alternative mechanisms that rely on GTPase or ATPase activity. Fun12 (or eIF5B) is a GTPase that interacts with the 60S subunit and promotes subunit joining. Fun12 is required for efficient formation of 80S-like complexes, which was shown to enhance Nob1-dependent cleavage at site D. It was proposed that GTP hydrolysis by Fun12 induces a rotation in the head domain that brings the Nob1 PIN domain closer to site D, thereby promoting site D cleavage (Lebaron *et al*, 2012). Rio1 ATP hydrolysis activity has also been shown to enhance the efficiency of Nob1 binding (Turowski *et al*, 2014).

The 20S pre-rRNA is cleaved at site D as part of the 80S-like complex. Soon after the 18S rRNA is matured, the 80S-like complex is dissociated by the Rli1 ATPase and Dom34, two subunit release factors that are also involved in recycling of stalled ribosomal subunits (van den Elzen *et al*, 2014). This allows the newly synthesized 40S subunit and the 60S subunit to start a new round of translation (Strunk *et al*, 2012).

Many of the assembly factors bind to the pre-40S in functionally important regions, such as the decoding centre or sites important for translation (Granneman *et al*, 2010; Strunk *et al*, 2011). For example, binding of Rio2, Tsr1 and Dim2 to pre-40S complexes is incompatible with the binding of the translation initiation factors eIF1 and eIFA1. Moreover, several Enp1, Rio2 and Ltv1 binding sites are located in or close to (and thus block) the mRNA channel.

Nob1 and Pno1 are among the last assembly factors to dissociate from the 80S-like complexes (Turowski *et al*, 2014; Hector *et al*, 2014). Their presence is believed to block the incorporation of translation initiation factor eIF3 that binds the mRNA tunnel near the exit site and the 3' end of 18S (Strunk *et al*, 2012). Similarly, dissociation of Tsr1 is believed to be required to allow Rli1 to bind and recycle the subunits (Strunk *et al*, 2012).

Nob1, Enp1, Ltv1 and Rio2 were shown to UV cross-link in close proximity to the rRNA binding sites of Rps3, Rps0, Rps14, Rps20, Rps15 r-proteins (the head domain). Intriguingly, both these assembly factors and r-proteins were

shown to be important for site D cleavage and/or nuclear export of the pre-40S complex (Strunk *et al*, 2011; Granneman *et al*, 2010; Ferreira-Cerca *et al*, 2007; 2005). Moreover, Rio2 and Nob1 were shown to interact with Rps5 and Rps14 while Rio2 and Ltv1 interact with Rps15 (Campbell & Karbstein, 2011). It might be that Nob1, Enp1, Ltv1, Rio2 could somehow regulate the timing of the binding/stable incorporation of these r-proteins.

Like Rps3, Rps14 also binds to early pre-ribosome complexes but it only becomes stably associated with the 20S pre-rRNA upon completion of the platform region (Loc'h *et al*, 2014). The Fap7 NTPase directly binds Rps14 (Granneman *et al*, 2005) and was proposed to chaperone Rps14 assembly into pre-40S complexes by blocking specific Rps14-rRNA interactions. Subsequent ATP hydrolysis would trigger release of Fap7 and allow Rps14 to adopt its final conformation (Loc'h *et al*, 2014).

### 1.3. The pre-60S assembly pathway

Similar to the 40S assembly, the hierarchical binding of r-proteins plays an important role in pre-60S maturation. Most of the 60S r-proteins are bound rather weakly but become more stably bound as the assembly progresses. Not only do r-proteins play a structural role, but many are also required for pre-rRNA processing. These r-proteins have been divided into three different categories: early (required for $27SA_2$ and $27SA_3$ pre-rRNA processing), middle (required for 27SB pre-rRNA processing) and late (required for 7S pre-rRNA processing). It was proposed that the gradual incorporation of the r-proteins dictates the sequential folding of the 27S pre-rRNA species (de la Cruz *et al*, 2015).

The assembly pathway follows the direction of transcription: the early r-proteins bind mostly to the 5' end of the 25S rRNA while the late r-proteins bind mostly to the 3' end of the 25S rRNA in pre-rRNAs intermediates. The middle r-proteins bind both 25S and 5.8S rRNA regions in pre-rRNA (Gamalinda *et al*, 2014). It was therefore proposed that regions that form the solvent exposed side (parts of domains I and II) would fold first, followed by the formation of the polypeptide exit tunnel (PET); formed by domains I and III) and by the central

protuberance (CP; domains II and V (fig.1.4) (Gamalinda *et al*, 2014).

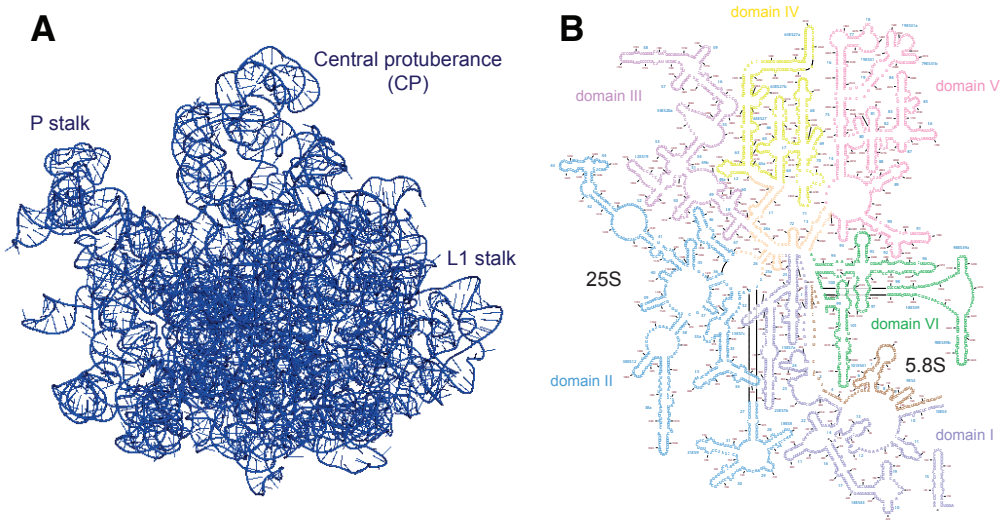



**Figure 1.4. A structure model of the S. cerevisiae 25S and 5.8S rRNAs.**
(A) The three dimensional structure of the 25S and 5.8S rRNAs in which the central protuberance (CP), P stalk and L1 stalk are highlighted; (B) A secondary structure of the 5.8S and 25S rRNAs respectively, in which domains I-IV of 25S are colour coded.

A consequence of the hierarchical binding/stabilization of the r-proteins is that the pre-60S intermediates are destabilized to a greater degree by the depletion of early r-proteins compared to the depletion of later r-proteins (Gamalinda *et al*, 2014). Rpl3 is probably one of the earliest r-proteins incorporated into pre-60S particles. Depletion of Rpl3 leads to a deficit of 60S subunits, blocks 27S pre-rRNA formation and pre-60S nuclear export, quite possibly the most drastic effect on 60S maturation observed (Rosado *et al*, 2007).

Apart from r-proteins a vast number of assembly factors are needed for pre-60S maturation (fig 1.5). These factors are endonucleases (including Rcl1, Las1), and exonucleases (including Rat1, Rrp17, MRP, Rex1, Rrp6, Ngl2), AAA-ATPases (i.e Rea1, Drg1), GTPases (Nog1, Nog2, Lsg1, Elf1), RNA-dependent ATPases (such as Dbp10, Spb4 and Mrt4) and proteins that represent scaffolds for the incorporation of other factors (most of the $A_3$ factors). The maturation of the pre-60S intermediates is more equally distributed among the three cellular compartments (nucleolus, nucleus and cytoplasm) than the pre-40S. As a result, the pre-60S intermediates have different assembly factors bound depending on their cellular localization.

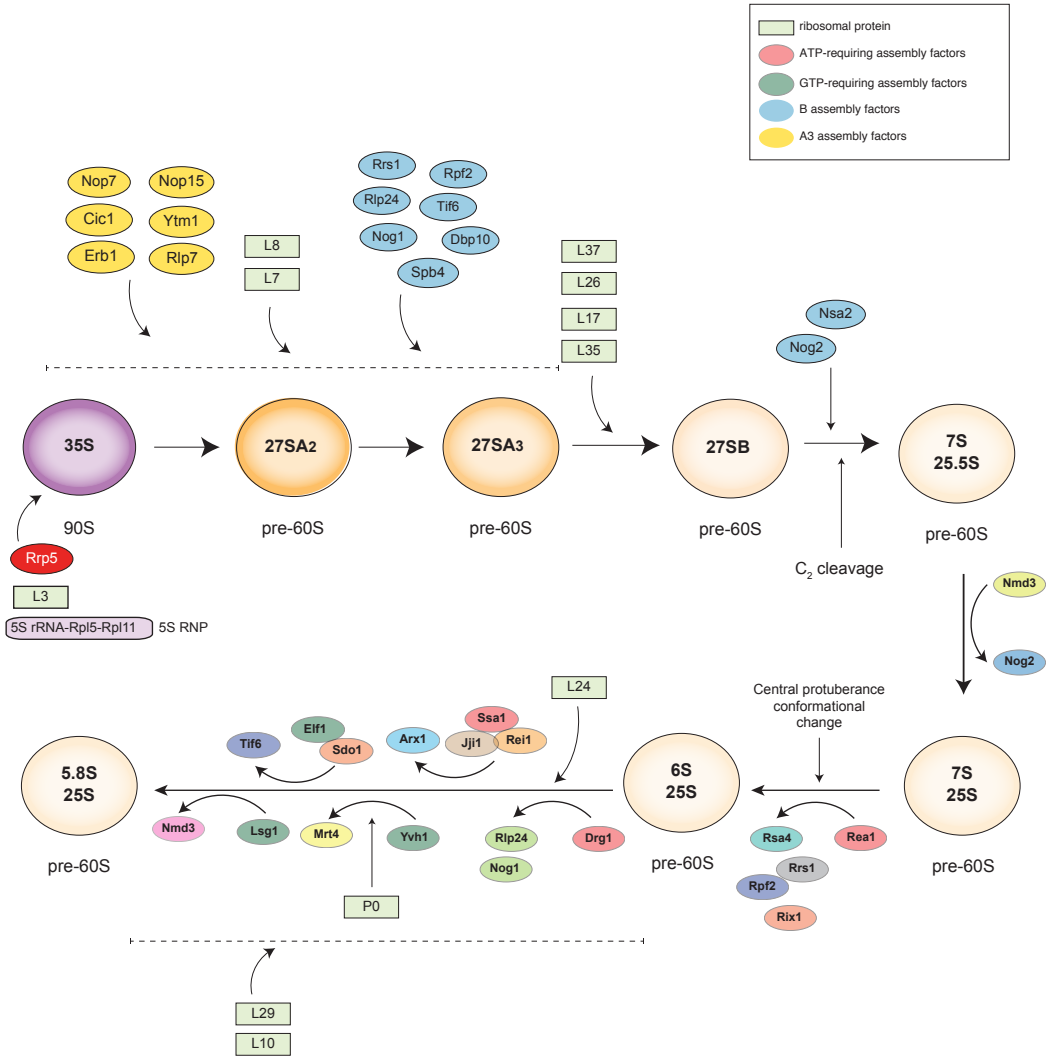



**Figure 1.5. Schematic representation of the pre-60S assembly pathway.**
The assembly factors are represented based on the timing of their association or dissociation. The assembly factors that have enzymatic activity are represented in red (ATPases) and in green (GTPases). Some of the r-proteins that regulate pre-60S assembly are represented in squares.

Rrp5 binds co-transcriptionally to the 35S pre-rRNA, making it one of the earliest assembly factors to be incorporated. Rrp5 was shown to crosslink to the 5' ETS and 18S regions but also to ITS1 between $A_2$ and B sites, 5.8S and domains I, IV and V (Lebaron *et al*, 2013). Rrp5 plays a role in both pre-40S and pre-60S maturation but remains associated with the pre-60S after $A_2$ cleavage (Venema & Tollervey, 1996; Lebaron *et al*, 2013). The current model is that during transcription Rrp5 chaperones the folding of the pre-rRNA by preventing the formation of inefficient structures (Lebaron *et al*, 2013).

Moreover, Rrp5 binding protects the rRNA from aberrant processing events and from degradation (Hierlmeier *et al*, 2013).

The nucleolar pre-60S assembly involves the formation and processing of the 27S pre-rRNA intermediates, the removal of ITS1 and the cleavage of ITS2 at $C_2$ site. The assembly factors that associate with nucleolar pre-60S particles have been divided into several groups. The $A_3$ assembly factors (Nop7, Erb1, Ytm1, Nop15, Cic1, Nop12, Rlp7) are required for the formation of $27SA_3$ pre-rRNA and the B factors (Nop2, Nip7, Rpf2, Rrs1, Mak11, Rlp24, Sbp4, Dbp10, Tif6, Nsa2, Nog2) are required for 27SB pre-rRNA cleavage (fig 1.5). The role of these assembly factors is mostly structural: they provide a scaffold for the binding of other assembly factors. Their assembly is required for the association of the late B-factors, Nsa2 and Nog2. Nog2 is the last assembly factor to associate before 27SB pre-rRNA is cleaved at site $C_2$ by the endonuclease Las1 (Saveanu *et al*, 2001; Gasse *et al*, 2015; Talkish *et al*, 2012).

The A3 factors, B factors together with several r-proteins (Rpl7, Rpl8, Rpl17, Rpl35, Rpl37, Rpl26) are part of precisely coordinated cascade of hierarchical binding steps. It starts with the incorporation of the A3 factors Nop7, Erb1, Ytm1, Nop15 and Cic1 and r-proteins Rpl7 and Rpl8. Their assembly is interdependent as removal of any of these proteins affects the proper incorporation of the others (Sahasranaman *et al*, 2011; Jakovljevic *et al*, 2012). Once all of the A3 factors are associated with pre-60S complexes, Rpl17, Rpl35, Rpl26 and Rpl37 are stably incorporated (Babiano & de la Cruz, 2010; Gamalinda *et al*, 2013). These four r-proteins bind to 5.8S and 25S sequences and are required for the formation of the polypeptide exit tunnel (PET).

The early B factors assemble in a hierarchical fashion via two parallel pathways. In both pathways the assembly starts with binding of Nop2 and Nip7 and finishes with the incorporation of the GTPase Nog2. In one pathway, the binding of Nop2 and Nip7 is followed by the assembly of the Rpf2 complex (Rpf2, Rrs1 and the 5S RNP) which in turn leads to the incorporation of the RNA helicase Spb4 (fig 1.5). In the second pathway, Nop2, Nip7 and a third assembly factor (Mak11) bind first, which triggers the assembly of Rlp24, Nog1 and Nsa2. Nog2 can only associate if Nsa2 and Spb4 are both present in the

pre-60S complex (Talkish *et al*, 2012; Lebreton *et al*, 2006a; Saveanu *et al*, 2003). Nsa2 and Nog2 bind just before 27SB pre-rRNA is cleaved (Saveanu *et al*, 2001).

The successful incorporation of Nog2 signals that all upstream assembly and folding steps have been completed and that the pre-60S particle is ready for cleavage at site $C_2$ in ITS2. It was proposed that before cleavage of $C_2$, the ITS2 region undergoes a number of conformational changes, which include the formation of the ITS2 proximal stem, a helix formed by base-pairing interactions between 5.8S and 25S sequences. Furthermore, ITS2 might switch from an "open" conformation (Joseph *et al*, 1999) to a more stable or "closed" conformation (Peculis & Greer, 1998). It is believed that these rearrangements are essential for efficient $C_2$ cleavage, although why is not yet clear (Côté *et al*, 2002). This raised the question whether the A3 factors, B factors and the r-proteins that are important for $C_2$ cleavage (Rpl17, Rpl35, Rpl26 and Rpl37), may somehow direct the proposed conformational switch in ITS2. To test this idea, several groups have used RNA structure probing methods (explained in section 1.4) to determine if the absence of these proteins altered the rRNA secondary structure. The most promising A3 factor candidates were the Nop15, Cic1, Rlp7 and Nop12 as they directly bind to ITS2 or 5.8S (Granneman *et al*, 2011; Dembowski *et al*, 2013b). Cells lacking Nop15, Cic1 or Rlp7 showed an altered structure in the 5' end of ITS2, while cells lacking Nop12 had an altered/aberrant 5.8S structure (Granneman *et al*, 2011; Dembowski *et al*, 2013b; Talkish *et al*, 2014). Although important for $C_2$ cleavage, Nog2 does not appear to influence ITS2 folding (Gamalinda *et al*, 2013). None of the r-proteins important for assembly of late B factors (Rpl17, Rpl35, Rpl26 and Rpl37) were required for folding of ITS2, although the absence of Rpl26 resulted in accumulation of pre-60S particles that had a slightly different structure in 5.8S and domain I of the 25S (Babiano *et al*, 2012; Gamalinda *et al*, 2013). Collectively, these studies showed that only the proteins that directly bind ITS2 were important for ITS2 folding.

Cryo-EM studies have revealed that the 60S subunit has a core structure from which a central protuberance (CP), a P stalk and an L1 stalk emerge (Spahn *et al*, 2001) (fig 1.4).

The CP is formed by the binding of the 5S RNP (5S rRNA-Rpl5-Rpl11) to

the 90S complexes (Zhang *et al*, 2007) (fig. 1.5). After the pre-60S is exported into the nucleus, the CP undergoes major conformational changes, namely the 180° rotation of the 5S rRNA followed by structural changes in its neighbouring regions: helices H82, H86, H89, H69 and H71 in the 25S (fig 1.5) (Leidig *et al*, 2014). The assembly factors Rpf2, Rrs1, Arx1, Rsa4, Rix1 and Rea1 are some of the key players in the CP conformational change (Leidig *et al*, 2014; Barrio-Garcia *et al*, 2015) as follows. The 5S RNP, which consists of the 5S rRNA bound to Rpl5 and Rpl11, is recruited to the CP region in the 90S pre-ribosome by the Rpf2-Rrs1 complex (Zhang *et al*, 2007). Several crystal structures of the Rpf2-Rrs1 complex were recently published (Asano *et al*, 2015; Kharde *et al*, 2015; Madru *et al*, 2015). These studies suggested that the Rpf2-Rrs1 complex interacts with the 5S RNP by directly binding the 5S rRNA (Rpf2) as well as the r-protein Rpl5. Shortly after the incorporation of the 5S RNP into the 90S pre-ribosomes, the Rpf2-Rrs1 complex binds to several helices in the 25S sequence (H80, H82, H88) that form the base of the CP (Kharde *et al*, 2015). CRAC data showed that Rpf2 binds H80 and the region between H87 and H87 in the 25S rRNA. This binding site appeared to be only accessible before the 5S RNP is rotated (Madru *et al*, 2015). This strongly suggests that the presence of Rpf2-Rrs1 complex into the pre-60S blocks the rotation of the CP and that their dissociation is essential to allow the CP to adopt its final conformation.

In the pre-ribosomes, Rpf2-Rrs1 complex interacts with Rsa4 (Madru *et al*, 2015; Kharde *et al*, 2015). Rsa4 interacts with Rpl5 in the 5S RNP and Nsa2 (Bassler *et al*, 2014) and with the AAA-ATPase Rea1 (Ulbrich *et al*, 2009). The pre-60 containing particles that contained Rpf2-Rrs1, Arx1 and Rsa4 were shown to have the CP its initial pre-rotated conformation (Leidig *et al*, 2014), while pre-60S particles that contained Arx1 and Rsa4 and Rix1 have the mature CP (post 180° rotation) (Barrio-Garcia *et al*, 2015). It was proposed that the ATP hydrolysis of Rea1 (AAA-ATPase) triggers the conformational change in the CP (fig 1.4). Rea1 interacts with pre-60S intermediates by binding the Rix1 complex via its AAA hexameric motor domain and by binding Rsa4 through its C-terminal MIDAS domain (Ulbrich *et al*, 2009; Nissan *et al*, 2004; Galani *et al*, 2004). Upon ATP hydrolysis, Rea1 undergoes a major conformational change, which results in its dissociation and the dissociation of its interacting partners

(Rix1 complex, Rsa4) as well as of the Rpf2-Rrs1 complex. The release of these factors subsequently triggers the rotation of the CP (fig 1.5) (Ulbrich *et al*, 2009). The energy released by Rea1 ATP hydrolysis is transmitted to H89 through the Rsa4-Nsa2 interaction. This results in relocation of H89 closer to the peptidyl transferase center (PTC). The Nog1 GTPase, which binds to H89, was also proposed to play a role in this structural rearrangement (Bassler *et al*, 2014). Recent work, however, suggests that the Rpf2-Rrs1 complex dissociates before Rix1 and Rea1 are incorporated and that Rix1 and Rea1 binding (but not Rea1 mediated ATP hydrolysis) induces the rotation of the CP (Barrio-Garcia *et al*, 2015).

Interestingly, Rea1 was also shown to be required for removal of the Ytm1-Nop7-Erb1 complex *in vitro*. Like Rsa4, Rea1 binds Ytm1 through its MIDAS domain and the removal of the trimeric complex also appeared to require ATP-hydrolysis by Rea1 (Bassler *et al*, 2010). How exactly Rea1 distinguishes between its different interaction partners (Ytm1 and Rsa4) and how the timing of their release is regulated is not known.

Once the conformational switch of the CP has been finalized, export adapters Nmd3 and Mex67-Mtr2 bind to the pre-60S. Nmd3 is a Crm1 adapter responsible for the export of the pre-60S particles to the cytoplasm (Ho *et al*, 2000). The presence of Nmd3 in pre-60S particles is mutually exclusive with the presence of Rix1, Rea1 and Rsa4 (Leidig *et al*, 2014) and with Nog2 (Matsuo *et al*, 2014). Interestingly, Nog2 and Nmd3 have overlapping binding sites, suggesting that Nog2 is a placeholder for Nmd3. Moreover, Nog2 promotes the stable association of Rsa4 and Rea1 with Rix1 containing pre-60S particles while at the same time blocking premature Nmd3 binding (Matsuo *et al*, 2014). Thus, the GTPase Nog2 not only regulates the timing of ITS2 cleavage but also the timing of nuclear export of pre-60S complexes.

Once exported into the cytoplasm, the 5.8S rRNA is matured by an 3'-5' exonuclease (Thomson & Tollervey, 2010) after which most of the late assembly factors dissociate. This dissociation takes place in a specific order (fig 1.4): Drg1 binds to the pre-60S after nuclear export and triggers Rlp24 release. Rlp24 release allows Rei1 binding and Rei1 induces Arx1 dissociation. Arx1 and Mrt4 dissociation allows the release of Tif6. Nmd3 release follows (Lo *et al*, 2010), allowing the joining of the matured 60S to the (pre-)40S subunit.

Drg1 is a AAA-ATPase that associates to the pre-60S in the cytoplasm shortly after nuclear export and induces the release of Nog1 and Rlp24 (Pertschy *et al*, 2007). Drg1 binds Rlp24 and this binding stimulates Drg1 ATPase activity, which triggers the removal of Rlp24 and assembly of the r-protein Rpl24 (Kappel *et al*, 2012). Rlp24 is an assembly factor that is a paralog of the r-protein Rpl24 and they occupy the same sites on the ribosome. Thus Rlp24 functions by preventing premature association of Rpl24 (Saveanu *et al*, 2003). Rei1, Jji1 and Ssa1 (ATPase) subsequently interact with pre-60S complexes by binding Rpl24 and trigger the dissociation of Arx1 (Lebreton *et al*, 2006b; Meyer *et al*, 2007; Demoinet *et al*, 2007). Arx1 binds in the nucleus to the pre-60S shortly after the ITS2 cleavage (Greber *et al*, 2012; Bradatsch *et al*, 2012). Arx1 mediates pre-60S export through its interactions with the phenyalalnine-glycine (FG) rich nucleoporins (Bradatsch *et al*, 2007). Arx1 binds before Nmd3 and Mex67-Mtr2 bind and is present in the pre-60S particles before and after the conformation switch of the CP. Arx1 binds the pre-60S directly at the PET. Thus, besides playing an important role in nuclear export, its presence is also incompatible with translation.

In parallel with the release of Arx1, the P stalk structure is formed (fig 1.5). The P stalk plays a key role in translation as it recruits and activates a number of translation factors. The stalk is formed by r-proteins P0, P1 and P2A and P2B. Only P0 remains bound to ribosomes at all times, whereas P1 and P2 are recycled with each translation cycle. In the early pre-60S, the P0 binding site is occupied by Mrt4. P0 and Mrt4 are paralogs (like Rlp24 and Rpl24) and Mrt4 is a placeholder for P0 (Rodríguez-Mateos *et al*, 2009). The release of Mrt4 and binding of P0 is part of a circular process: the presence of Yvh1 is needed for the dissociation of Mrt4, as Yvh1 depleted cells retain Mrt4 bound to the cytoplasmic pre-60S particles (Kemmler *et al*, 2009). At the same time, P0 binding is required for Yvh1 release (fig. 1.5) (Lo *et al*, 2010).

Tif6 is released next. Tif6 is a protein that plays an important role in a tRNA-mRNA translocation-like quality control cycle. Tif6 dissociation is triggered by the GTPase activity of Elf1 and a protein called Sdo1. Elf1 is a homologue of the translation elongation factor eEF2 and its binding is dependent on the formation of the P stalk (Senger *et al*, 2001). Sdo1 couples the activation of GTP hydrolysis of Elf1 with Tif6 release (Finch *et al*, 2011). Tif6 can only be

released once Arx1 and Mrt4 have been removed (Lo *et al*, 2010; Lebreton *et al*, 2006b) Thus, Tif6 release signals the successful completion of both the P stalk and PET (through the absence of Arx1 and Mrt4). Because Elf1 triggers the release of Tif6 in a translocation-like cycle (Bussiere *et al*, 2012), it is believed that the pre-60S particles formed after this step are able to function in tRNA translocation.

In humans, Shwachman-Bodian Diamond syndrome (SBDS) is a ribosomopathy characterized by bone marrow dysfunction and a high risk of myeloid leukaemia. Several mutations in the human ortholog of Sdo1, the SBDS gene, have been shown to cause the disease. The etiology of SBDS lies in the failure to couple the GTPase activity of Elf1 with eIF6 (Tif6) release through SBDS (Sdo1). Failure to release Tif6 also blocks subunit joining, which was also shown to be blocked in SBDS patients (Finch *et al*, 2011).

Tif6 associates with pre-60S particles throughout the whole assembly pathway. Along the way, Tif6 is phosphorylated by Hrr25. Blocking the Hrr25 dependent phosphorylation of Tif6 leads to reduced levels of 25S rRNA (Ray *et al*, 2008). It is not yet clear what the function of Hrr25-dependent phosphorylation of Tif6 is. However, based on the role of Hrr25 in pre-40S export, Ltv1 and Enp1 dissociation and Rps3 stable incorporation, it is tempting to speculate that phosphorylation of Tif6 might also be important for pre-60S export.

Once Tif6 is released the export adapter Nmd3 can bind (Lo *et al*, 2010). Shortly after nuclear export, Nmd3 is removed by the GTPase activity of Lsg1, which also involves the r-protein Rpl10. Mutations in Lsg1 and/or Rpl10 block Nmd3 dissociation and recycling, and this export defect can be reduced by overexpressing Nmd3. Like Tif6, Nmd3 binds on the subunit joining interface of the pre-60S (Sengupta *et al*, 2010) and its presence is therefore incompatible with subunit joining. Dissociation of Tif6 and Nmd3 most likely represents the final quality control step before subunit joining.

The release of the assembly factors in the cytoplasm coincides with the binding of late r-proteins Rpl24, Rpl29, Rlp10 and P0 (fig.1.5) (Gamalinda *et al*, 2014). Rpl40 was initially thought to bind the 60S subunit in the cytoplasm, however a recent study has detected Rpl40 in the nucleolus, suggesting it assembles much earlier (Fernández-Pevida *et al*, 2012). Cells lacking Rpl10, Rpl29 or Rpl40 cannot efficiently join 40S and 60S subunits (Fernández-Pevida

*et al*, 2012; DeLabre *et al*, 2002; Eisinger *et al*, 1997). Therefore, successful incorporation of these r-proteins is a signal that indicates that earlier stages of pre-60S assembly have been completed and that the subunit is ready for translation initiation.

## 1.4. Nuclear export of pre-40S and pre-60S intermediates

The large (>2MDa) pre-40S and pre-60S subunits are actively exported through the nuclear pore complex (NPC) to the cytoplasm. Export of pre-ribosomes involves several redundant export pathways. The advantage of having multiple (parallel) export pathways is that it increases the efficiency of export and ensures that if one of the pathways is blocked (transport factor unavailable/degraded/overloaded) the pre-ribosomal subunits can still be efficiently exported (Nerurkar *et al*, 2015). Due to their size, it might be expected that pre-ribosomes are bound by several export adapters when exported through the NPC (Ribbeck & Görlich, 2002). Several factors have been identified to be involved in export of pre-ribosomes to the cytoplasm. However, it is not clear how all these factors cooperate with each other to promote nuclear export.

The karyopherin Crm1/Xpo1 mediates the nuclear export of the pre-ribosomes. Crm1 interacts with its cargos through their Nuclear Export Sequences (NES). Their export is dependent on a GTPase called Ran in its GTP-bound state (RanGTP). RanGTP interacts with Crm1 bound to the NES of its cargo and is responsible for the directionality of the transport (Cook *et al*, 2007). Once the cargo is exported to the cytoplasm, Ran GTP hydrolysis triggers the dissociation of the cargo, Ran and Crm1 (fig 1.6A). These will then be transported back into the nucleus where the GDP bound to Ran is rapidly substituted with GTP (Pemberton & Paschal, 2005). Several NES containing assembly factors were shown to bind Crm1 and promote export of pre-ribosomes.

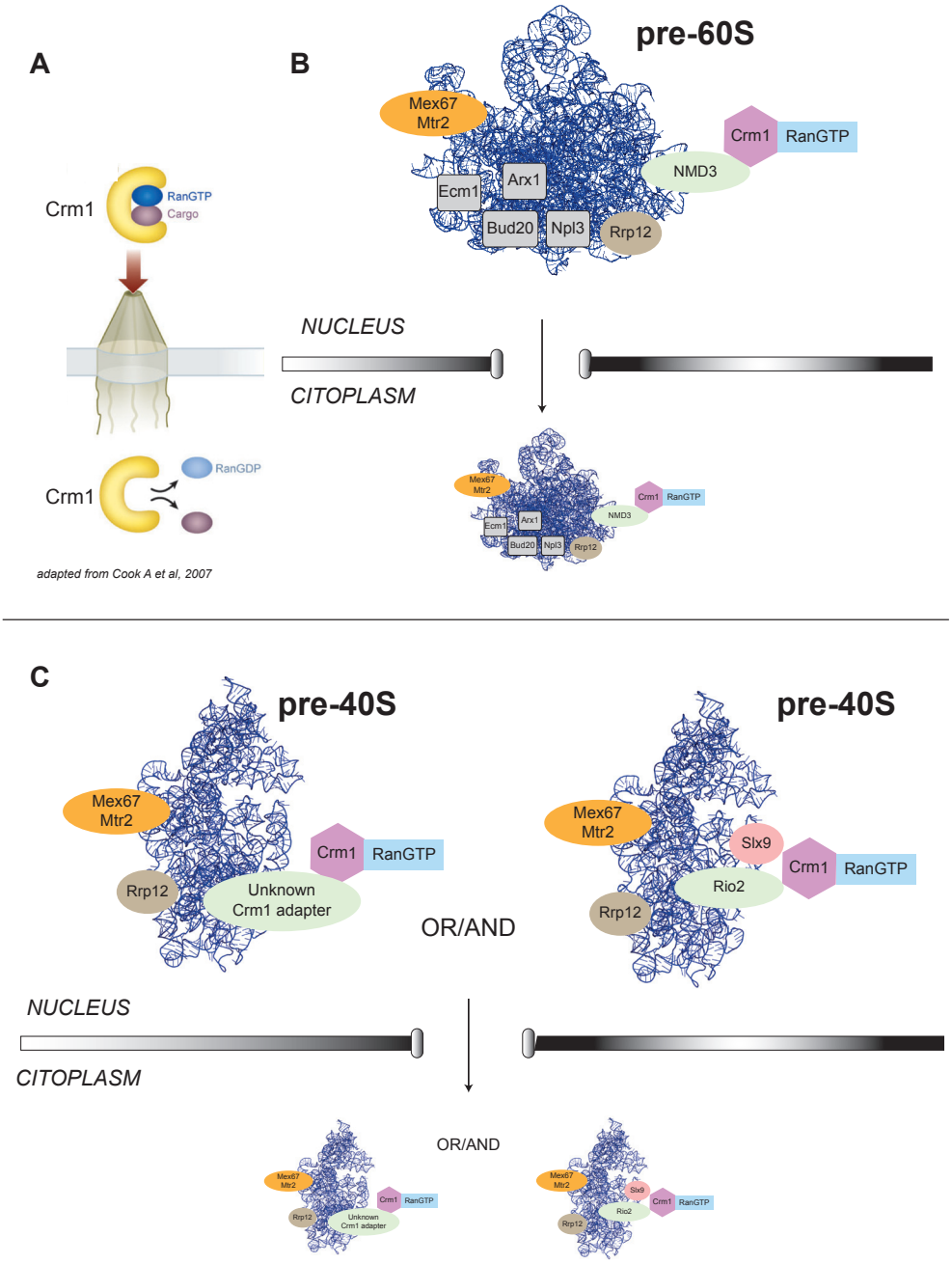



**Figure 1.6. Crm1 mediated export of pre-ribosomal particles.**
(A) Crm1 mediated export is an active export in which Crm1 recognizes and binds to the nuclear export sequence (NES) of its cargo. The gradient of GTP or GDP bound to Ran determines the directionality of the export; (B) Pre-60S export is mainly mediated by the Crm1 adapter Nmd3 but can also involve the Mex67-Mtr2 heterodimer and other auxiliary factors; (C) Rio2 is a weak Crm1 adapter and mediates pre-40S Crm1 dependent export in the presence of Slx9. No other Crm1 adapters have been identified to date. Mex67-Mtr2 export pathway and other auxiliary factors are also involved in pre-40S export.

The best characterized Crm1 adapter is Nmd3 (fig 1.6B). Nmd3 has an NES sequence in the C-terminal domain and is essential for pre-60S export. Deletion of or point mutations in the Nmd3 NES block pre-60S export without affecting Nmd3 incorporation into the pre-ribosomes. Inhibiting export using a drug (leptomycin B) that binds Crm1 also blocks Nmd3-dependent export, indicating that Nmd3 is a Crm1 adapter (Ho *et al*, 2000; Gadal *et al*, 2001; Belk *et al*, 1999).

Although it has been known for many years that Crm1 is required for pre-40S export (Moy & Silver, 1999), the export adapter for pre-40S export remains elusive (fig 1.6C).

Dim2, Ltv1 and Rio2 all contain putative NESs and were considered promising candidates as pre-40S export adapters. However, mutations in the Dim2 NES did not affect nuclear export, suggesting its NES is probably not required for pre-40S export (Vanrobays *et al*, 2008). In addition, although the Ltv1 NES binds Crm1 in yeast two-hybrid assay (Neuber *et al*, 2008), several lines of evidence suggest that Ltv1 is not a pre-40S export adapter. Mutations in the Ltv1 NES did not block the nuclear export of 20S pre-rRNA but affected its cytoplasmic processing (Fassio *et al*, 2010). In addition, Ltv1 mutants lacking the NES can rescue the growth defect induced by Ltv1 depletion suggesting that Ltv1 is not an essential Crm1 adapter for pre-40S export (Merwin *et al*, 2014). Interestingly, Rps3 accumulates in the nucleus in Ltv1 depletion mutants (Seiser *et al*, 2006) while dominant negative NES mutations in Ltv1 can be suppressed by overexpressing Rps3 (Merwin *et al*, 2014). This suggests that Ltv1 role in pre-40S export is most probably indirect.

Human Rio2 was shown to bind to Crm1 in a RanGTP-dependent manner *in vitro* and to act as an export adapter for pre-40S subunits (Zemp *et al*, 2009). Although yeast Rio2 also contains an NES that binds Crm1, it is not essential for pre-40S export. A Rio2 mutant that lacked the NES was able to support growth comparable to wild-type levels (Ferreira-Cerca *et al*, 2012; Vanrobays *et al*, 2003).

Interestingly, the Panse lab showed that Rio2 binds Crm1 in a RanGTP-dependent manner but only in the presence of the RanGTP binding protein Slx9 (Slx9 is found in yeast but not in higher Eukaryotes) (fig 1.6C). Binding of Slx9 to RanGTP is predicted to enhance the binding of Crm1 to the Rio2

NES. These results suggested that binding to Crm1 of adapters with weak NES sequences might be enhanced/stabilized by additional factors (Fischer *et al*, 2015). Although the Rio2 NES is not essential for pre-40S export in yeast, deletion of Slx9 does cause defects in pre-40S complexes (Faza *et al*, 2012). This implies that Slx9 might be important for recruiting Crm1 to other pre-40S adapter proteins that have weak NES signals.

Interestingly, pre-ribosome export defects in Nmd3 and Slx9 mutants could be rescued by overexpressing the Mex67-Mtr2 complex, which was previously shown to play a role in pre-ribosome export (Yao *et al*, 2007). Overexpression of Mex67-Mtr2 in an *slx9Δ* strain rescued the pre-40S nuclear export defect, while mutations in Mex67-Mtr2 aggravated the pre-40S export defect in *slx9Δ* strains. Slx9 incorporation into pre-40S particles requires Mex67-Mtr2 but not the other way around (Faza *et al*, 2012). As Slx9 is involved in the Crm1 pathway as well, it could be that Slx9 participates in the crosstalk between the Crm1 and Mex67-Mtr2 pathways. In yeast, the Mex67-Mtr2 complex was shown to interact with the pre-40S and the pre-60S through a dimerization surface formed by yeast specific loops of Mex67 and Mtr2 (Fribourg & Conti, 2003). This dimerization surface was shown to directly bind to the rRNA.

Collectively, these data suggest that the Mex67-Mtr2 dimer can take over pre-40S export if the Crm1-dependent pathway is impaired (Faza *et al*, 2012; Yao *et al*, 2007). Mechanistic insights into how Mex67-Mtr2 select their substrates are lacking, but it is believed that Mex67-Mtr2 uses different interaction surfaces for mRNA, pre-ribosomes and NPC FG repeats (Strässer *et al*, 2000; Conti & Izaurralde, 2001; Stutz *et al*, 2000).

Mex67-Mtr2 seem to be able to bind the NPC and pre-60S or NPC and mRNA simultaneously, but not pre-ribosomes and mRNA at the same time, ensuring that the export of these molecules remains separated. Mutations in the loops that form the dimerization surface impaired pre-60S export but did not affect mRNA export. *Vice versa*, mutations that impaired mRNA export did not affect pre-60S export. This confirms that recognition of mRNA and pre-ribosomes by Mex67-Mtr2 involves different binding sites (Yao *et al*, 2007).

The positively charged amino acids that are part of the dimerization loop-rRNA interaction surface were shown to form electrostatic bonds with

the negatively charged 5S rRNA in the pre-60S (Yao *et al*, 2007). The amino acids needed for the interaction with the rRNA are on the opposite side from the amino acids predicted to bind the FG rich proteins in the NPC (Fribourg & Conti, 2003). This would allow the simultaneous interaction with the NPC structure and pre-60S complexes. On the other hand, Mex67-Mtr2 5S rRNA binding sites overlap with the binding sites to nucleoporin Nup85 which is essential for mRNA export (Yao *et al*, 2008).

Apart from transport receptors for active export, the negative charge of the rRNA needs to be shielded in pre-ribosomes during passage through the NPC. This is facilitated by the HEAT repeat protein Rrp12 (pre-40S and pre-60S nuclear export) which interacts with the FG repeats and helps overcome the hydrophobic barrier of the NPC (Oeffinger *et al*, 2004). Other such auxiliary factors for the 60S are Npl3, Arx1, Ecm1 and Bud20. Arx1 interacts with the FG repeats (Bradatsch *et al*, 2007; Yao *et al*, 2010; Hackmann *et al*, 2011; Altvater *et al*, 2012). Interestingly, the NES sequence in the N-terminus of Bud20 does not interact with Crm1 but binds FG nucleoporins directly (Altvater *et al*, 2012). Apart from the FG interactions, the pre-60S also uses adapters that interact with nucleoporins in an FG-independent manner, through Gle2. Gle2 interacts specifically with pre-60S through a basic surface while using another binding surface to make non-FG interactions with Nup116 (Occhipinti *et al*, 2013).

## 1.5. Determining RNA secondary structure through structure probing approaches

SHAPE (selective 2`-hydroxyl acylation analysed by primer extension) is a structure probing approach that relies on the ability of SHAPE chemicals to modify the backbone of the RNA (fig 1.7A) (Mortimer & Weeks, 2007). SHAPE chemicals (among which 1M7 and NAI) acylate the 2'-OH of the ribose and generically react with all four bases, although they have a slight preference for adenosines and guanosines (Wilkinson *et al*, 2009).

The sites of chemical modification can be detected using a reverse transcription reaction. The enzyme generally does not go past the modified nucleotide (unless the reaction is performed in the presence of Manganese).

Therefore when the enzyme reaches the modified nucleotide it will fall off, resulting in truncated cDNAs. By running these products alongside a DNA sequencing ladder, one can identify the site of modification, which is always one nucleotide upstream of the reverse transcriptase drop-off site (fig 1.7A) Alternatively, the Weeks lab has shown that exonucleases are also sensitive to SHAPE modifications and therefore can also be used to map the site of modification (Steen *et al*, 2010).

SHAPE reagents generally react with nucleotides that are in a flexible/ single-stranded conformation. Although the mechanism is not fully understood, it seems that some flexible conformations increase the nucleophilicity of the nucleotide making it more prone to react to the electrophilic chemical (Merino *et al* 2005; Mortimer & Weeks, 2007; McGinnis *et al*, 2012). The reactivity to SHAPE chemicals was shown to correlate to molecular motion (the dynamic transition between different RNA structures with different degrees of flexibility) and not to solvent accessibility. An *in crystallo* experiment with in which *E.coli* ribosome crystals were soaked in 1M7 showed that there was a poor correlation between rRNA flexibility and solvent accessibility of the nucleotide. The poor correlation with solvent accessibility could be explained by the fact that nonadjacent, long range interactions stabilize the nucleotides and alter their reactivity to SHAPE chemicals (McGinnis *et al*, 2012). Nucleotides that form base stacking interactions were also found to be less modified by SHAPE chemicals (Bindewald *et al*, 2011). Alongside tertiary interactions, protein binding can also influence a nucleotide's reactivity to SHAPE chemicals by forcing the nucleotide to adopt a conformation that is thermodynamically less favourable. As a result, a drop in SHAPE reactivity upon protein binding is likely due to a change in the environment of a nucleotide, not due to steric hindrance. Thus it is important to remember that SHAPE reactivity is a measure of nucleotide flexibility, not solvent accessibility (Weeks & Mauger, 2011).

The *in crystallo* experiment also revealed that a small number of nucleotides with unusual, but well-defined, conformations were hyper-reactive to SHAPE chemicals. These conformations were incompatible with Watson-Crick interactions suggesting that SHAPE chemicals modify mainly, but not exclusively, single stranded nucleotides (McGinnis *et al*, 2012).

SHAPE chemical probing has been a very useful tool for analysing RNA

secondary structures under physiological conditions. Structural information obtained using SHAPE chemistry is not only extremely valuable for refining existing secondary structure models, but it can also help with modelling of *de novo* RNA structures (Rice *et al*, 2014). Several studies have used SHAPE chemistry to analyse the secondary structure of long non-coding RNAs (lncRNAs)(Novikova *et al*, 2013) and of viral genomes such as HIV-1 genome (Low *et al*, 2014; Wilkinson *et al*, 2008; Watts *et al*, 2009). Moreover, SHAPE chemistry was used to track conformational changes that take place during viroid maturation of the Murine Leukaemia Virus (MuLV) (Grohman *et al*, 2014). In addition, SHAPE reagents have been used to measure the secondary structure of mature rRNAs in a variety of species (16S *in E. coli* (McGinnis & Weeks, 2014; Deigan *et al*, 2009), 18S , 25S and 5S in *S. cerevisiae* (Leshin *et al*, 2011)).

Some SHAPE chemicals can be used both *in vitro* and *in vivo:* NAI (Spitale *et al*, 2014) and 1M7 (McGinnis & Weeks, 2014). SHAPE chemicals generally react very quickly with RNA (milliseconds to minutes) and self-quench in water-based solutions such as buffers or media (Weeks & Mauger, 2011).

ChemModSeq is a RNA structure probing method developed by Sander Granneman and Ralph Hector, that combines SHAPE chemistry with high-throughput sequencing and statistical modelling (Hector *et al*, 2014). The method can be applied both in *in vivo* and *in vitro* and it can provide nucleotide resolution insights into (changes of) RNA flexibility. The biggest advantage of ChemModSeq is that it can be applied to study the structure of very long RNAs in a single reverse transcription reaction. This greatly accelerated our analyses of structural changes on long rRNAs in yeast (Hector *et al.* 2014; Burlacu *et. al*; in preparation). The reason why it is so effective is ChemModSeq uses random priming oligonucleotides (rather than sequence-specific oligonucleotides) to perform the reverse transcription reaction, which results in a collection of cDNA fragments of variable lengths. To enable high-throughput sequencing of these cDNA pools, Illumina specific barcodes are added to each end of the cDNAs. The sequence reads are aligned to the reference sequence (for example a transcriptome, genome or gene-specific sequences, such as rRNAs). The modified nucleotides can then be determined by mapping the nucleotides that were positioned one nucleotide upstream of the reverse transcriptase drop-

off site. Sander Granneman has developed a data-processing pipeline for quantifying the nucleotide reactivity to the SHAPE chemicals (pyChemMod). To identify the reactivity of a particular nucleotide to the SHAPE chemical the data analysis pipeline calculates reverse transcriptase (RT) drop-off rates for each nucleotide, which is defined as total number of reads that stop at a given position (drop-off counts) divided by the total number of reads that cover that position. It then compares the RT drop-off rate of the unmodified (control) to the RT drop-off rate in the modified sample. Stuart Aitken developed a two Channel Poisson Expectation Maximization ($TCP^{EM}$) algorithm that estimates the probability of a nucleotide being modified (detailed in Chapter 3).

A number of similar methods have recently been described, including SHAPE-seq (Lucks *et al*, 2011), Structure-seq (Ding *et al*, 2014), Mod-seq (Talkish *et al*, 2012) and SHAPE-MaP (Siegfried *et al*, 2014) (fig 1.7B). ChemModSeq, SHAPE-seq and Structure-seq use similar approaches for generating the cDNA libraries. As it happens, Structure-seq was published while our ChemModSeq manuscript was under review in Cell. Both ChemModSeq and Structure-seq use random primers for the reverse transcription reaction and ligate adapters to the 3' end of cDNAs using Circligase.

Contrary to Structure-seq, ChemModSeq uses statistical modelling to call modified nucleotides with high accuracy and outperforms Structure-seq in several ways (Hector et al 2014). Structure-seq calculates SHAPE reactivities by calculating the number of stops at a nucleotide relative to the average number of stops of a particular transcript for each nucleotide (dependent on cDNA fragment length), subtract these values of the negative control from the treated sample and then normalizes the values so that they are between 0 and 2 (Deigan *et al*, 2009). A nucleotide is called modified if its SHAPE reactivity is higher than 0.7. ChemModSeq calculates the ratio between the number of stops at one position and the total number of reads covering that position (dependent on sequencing coverage). These values are then used as an input for the $TCP^{EM}$ statistical model, which takes into account the probability of a nucleotide being modified based on the input data. The algorithm is explained in more detail in Chapter 3.

ModSeq uses a very different strategy to detect sites of modification. The modified and untreated RNAs are first randomly fragmented and then

the adapters are ligated to both ends of the RNA. RT is performed using a primer that hybridizes to the 3' adapter. The resulting cDNAs is circularized. If the cDNA contained the 5' adapter sequence, then this means that the RT was able to reverse transcribe the entire RNA fragment and did not run into a modified nucleotide. The cDNAs that did not contain the 5' adapter were most likely derived from RNAs that contained modified nucleotides (either natural modifications or modified by the chemical probe). The next stage of the library preparation involves the removal of circularized cDNA fragments that contained 5' adapter sequences using a hybridization approach. The remaining cDNAs sequenced. The Illumina specific primers are then added by PCR and the cDNA libraries are sequenced. To generate a reactivity value for the chemical, the total number of stops at a nucleotide are divided by the average of the total number of stops in one transcript. A nucleotide is called modified if this ratio in the treated is at least 1.5-fold higher compared to the control sample.

SHAPE-MaP uses the property of a reverse transcriptase to misread and incorporate a non-complementary nucleotide or deletions when encountering obstacles such as a SHAPE modified nucleotide. The number of mutations is dramatically enhanced by performing the primer extension reaction in a buffer containing manganese instead of magnesium. As a result, SHAPE-MaP calculates SHAPE reactivities by analysing mutation rates in cDNA reads. The mutation rate is calculated by dividing the mutation count for a given nucleotide divided by the read count. An advantage of SHAPE-MaP is that it is less prone to sequence biases introduced during the library preparation steps, as it only reads the number of mutations. Some RNA ligases have a strong preference for specific nucleotide combinations, which could bias the results. A disadvantage of using SHAPE-MaP is that the mutation frequency is often low and to generate reliable data it is essential to have very high quality sequencing data, as mutations introduced during the sequencing reactions can sometimes generate high background levels. SHAPE-MaP can also be used to predict protein binding sites (Smola *et al*, 2015)

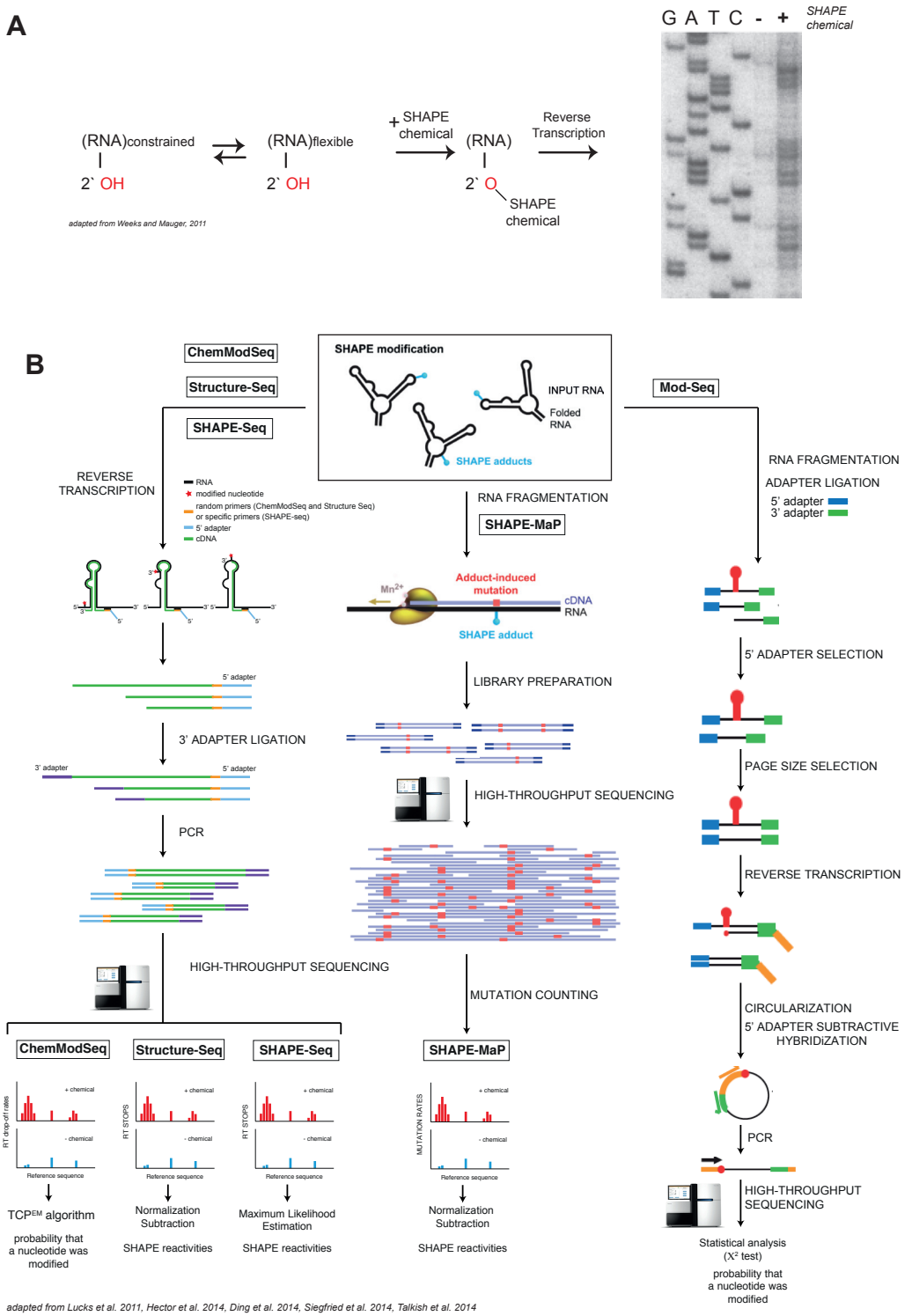



**Figure 1.7. RNA structure probing using SHAPE chemistry.**
(A) SHAPE chemicals react with the 2' hydroxyl of the ribose to form 2'-O-adducts. The primer extension gel shows the stops induced by the chemical modification in a reverse transcription reaction; (B) Comparison of the various high-throughput RNA structure probing methods.

# CHAPTER 2

# Materials and methods

## 2.1. Yeast media and general solutions

YPG/R medium - 2% galactose, 2% raffinose, 1% yeast extract, 2% peptone (Formedium)

YPDA medium - 2% glucose, 1% yeast extract 2% peptone (Formedium)

SGR-/-URA medium - 2% galactose, 2% raffinose, yeast nitrogen base, Complete supplement mixture (CMS) URA drop-out (Formedium)

SD-/-URA medium – 2% glucose, yeast nitrogen base, CMS URA drop-out (Formedium)

20X SSC solution – NaCl 150mM, NaCitrate $2H_2O_2$ 15mM

TMN150 buffer - 50 mM Tris–HCl pH 7.8, 1.5 mM $MgCl_2$, 150 mM NaCl, 0.1% NP-40 and 5 mM β-mercaptoethanol

Tris-EDTA (TE) buffer – 10mM Tris – Cl pH=8, 1mM EDTA

## 2.2. Yeast strains

*Saccharomyces cerevisiae* strain BY4741 (*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*) was used as the parental strain. An overview of the strains used in this study is provided in Table 2.1.

Strains with galactose-inducible genes (GAL7), TAP or HTP-tags or gene deletions were generated using homologous recombination (Oldenburg *et al*, 1997). The *GAL::3HA* cassette (with Kan resistance marker) was amplified by PCR from the pFa6a-kanMX6-pGal1 plasmid. The TAP and HTP tags were amplified from the pBS1539-TAP or HTP plasmid. To generate gene deletions, the nourseothricin (Nat), G418 (Kan) or hygromycin B (hyg) antibiotic resistance genes were amplified by PCR (Takara) from the respective pFa6a-plasmids. The PCR product was purified using the DNA Clean & Concentrator 5 kit (Zymo Research) and then transformed into *S. cerevisiae*. For the yeast transformation, the strains were grown until $OD_{600}$=0.5, pelleted and then resuspended in LiAc/

TE solution (100mM LiAc, 1xTE buffer). Cells were then mixed with the DNA of interest and 5µl carrier DNA (salmon sperm DNA, Ambion) in 50% PEG 4000 solution, incubated at 30°C for 30min, heat shocked at 42°C for 30min and then plated on agar with the appropriate selection marker.

For most experiments the strains were grown in YPD or YPG/R at 30°C. For the Nob1 and Fap7 depletion experiments using the *GAL::3HA-fap7* and *GAL::3HA-nob1* strains, cells were grown in YPG/R to an $OD_{600}$ of ~0.5, shifted to YPDA and grown for four hours to an $OD_{600}$ of ~1.0. For the Rps17 depletion experiments, cells were grown in YPG/R to an $OD_{600}$ of ~0.5, shifted to YPDA and grown for 4h and 8h.

*Table 2.1. Saccharomyces cerevisiae* strains used in this study

| Strain | Genotype | Reference |
|---|---|---|
| BY4741 | MATa*; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0* | (Brachmann et al, 1998) |
| D1092 | as BY4741 but with *enp1-HTP::K.l.URA3* | (Granneman et al, 2010) |
| YSG793 | as BY4741 but with *nob1-TAP::K.l.URA3* | This study |
| D1077 | as BY4741 but with *rio1-HTP::K.l.URA3* | (Granneman et al, 2010) |
| YSG843 | as BY4741 but with *rio1-TAP::K.l.URA3* | This study |
| D1079 | as BY4741 but with *rio2-HTP::K.l.URA3* | (Granneman et al, 2010) |
| YSG839 | as BY4741 but with *rio2-TAP::K.l.URA3* | This study |
| D1081 | as BY4741 but with *ltv1-HTP::K.l.URA3* | (Granneman et al, 2010) |
| D1089 | as BY4741 but with *tsr1-HTP::K.l.URA3* | (Granneman et al, 2010) |
| YSG723 | as BY4741 but with *fun12-TAP::K.l.HIS3* | This study |
| YSG773 | as YSG723 but with *GAL::3HA-fap7::kanMX4* | This study |
| YSG808 | as BY4741 but with *GAL::3HA-nob1::kanMX4 rio2-TAP::K.l.URA3* | This study |
| YSG779 | as BY4741 but with *GAL::3HA-nob1::kanMX4 fun12-TAP ::K.l.HIS3* | This study |
| 509 | as BY4741 but with *rps9B::kanMX4*, *rps9A::HIS3* + *Yplac111-GAL-FLAG-RPS9A (LEU,CEN)* | P. Milkereit |

| | | |
|---|---|---|
| 603 | as BY4741 but with *rps20:kanMX4 + Yplac111-GAL-FLAG-RPS20 (LEU,CEN)* | P. Milkereit |
| 484 | as BY4741 but with *rps10B::kanMX4 + Yplac111-GAL-FLAG-RPS10A (LEU,CEN)* | P. Milkereit |
| 693 | as BY4741 but with *rps3::kanMX4 + Yplac111-GAL-FLAG-RPS3 (LEU,CEN)* | P. Milkereit |
| YSG766 | as BY4741 with *rps17B::hygMX6, GAL::3-HA-rps17A::kanMX6* | This study |
| YSG783 | as YSG766 with *pRPS28-FLAG-RPS17 wild type (URA)* | This study |
| YSG784 | as YSG766 with *pRPS28-FLAG-RPS17 Δ127-136 (URA)* | This study |
| YSG785 | as YSG766 with *pRPS28-FLAG-RPS17 Δ100-136 (URA)* | This study |
| YSG786 | as YSG766 with *pRPS28-FLAG-RPS17 Δ75-136 (URA*) | P. Milkereit |
| YSG797 | as YSG766 with *pRPS28-FLAG-RPS17 Δ1-26 (URA)* | This study |
| YSG798 | as YSG766 with *pRPS28-FLAG-RPS17 Δ1-49 (URA)* | This study |
| YSG848 | as D1092 with *snR35::natMX6* | This study |
| YSG849 | as with D1077 *snR35::natMX6* | This study |
| YSG871 | as BY4741 but with *tsr1-TAP::K.l.URA3* | This study |

### 2.3. Designing the Rps17 mutants

Rps17 truncation mutants: Δ100-136, Δ127-136 were generated using using the QuickChange II site-Directed Mutagenesis Kit (Agilent Technologies) following the manufacturer's instructions. The Rps17 truncation mutants Δ1-26 and Δ1-49 were generated by PCR using the pRPS28-FLAG-Rps17 plasmid as a template and primers containing BamHI and PstI restriction sites. The PCR products were then digested with BamHI and PstI restriction enzymes (NEB) and ligated into the pRPS28-FLAG plasmid (digested with the same enzymes) using T4 ligase (NEB).

The Rps17 NES mutant was generated as follows. Synthetic BamHI, PstI digested Rps17 fragments in which Leu113, Leu115, Leu117 and Leu119 were mutated to Alanines, were cloned into the pRPS27-FLAG-Rps17 plasmid digested with the same enzymes. The pRPS27-FLAG-Rps17 Nmd3 NES

mutant was generated by cloning a synthetic BamHI, PstI digested Rps17 fragment in which the Rps17 NES was replaced with the Nmd3 NES into the pRPS27-FLAG-Rps17 plasmid digested with the same enzymes. All of the fragments were synthesized by GeneArt.

To generate the modified p1372 GFP reporter constructs, the PKI (Protein Kinase Inhibitor) NES in the original p1372 plasmid (Taura et al, 1998) was replaced with PCR amplified wild-type or mutant Rps17 NES and cloned into the p1372 vector using XmaI and EcoRI restriction sites.

## 2.4. Radiolabeling of oligonucleotide probes

The oligonucleotide probes (1μM) were radiolabelled with $\gamma^{32}$P-ATP using PNK (polynucleotide kinase) enzyme (NEB) in the presence of 10 mM DTT. The radiolabeled oligonucleotide was purified by size exclusion chromatography using mini Quick Spin columns (Roche) following the manufacturer's instructions.

## 2.5. Northern Blot

Northern blotting was performed as previously described (Tollervey, 1987). Radiolabeled probes were incubated with membranes at 37ºC o/n, and were washed for 10 minutes with 6xSSC followed by another 10 minute wash with 2xSSC+0.1%SDS. The radioactive signal was visualized on a FLA5100 phosphoimager system.

Table 2.2. Oligonucleotides used for Northern blot analyses

| Primer | Sequence 5'-3' | It hybridizes to: |
|---|---|---|
| 003 | AATGCTCTATCCCCAGCACG | A2-A3 region<br>35S, 32S and 27SA$_2$ pre-rRNAs |
| 004 | CGTCCTTGGCAAATGC | D-A2 region<br>35S, 32S. 23S and 20S pre-rRNAs |
| 007 | ATCGGTACTAGCGACGG | 5` end of 25S<br>35S, 32S, 27S and 25S pre-rRNAs |
| 008 | CCACTATTTAGTAGGTTAAGGTCTC | 5` end of 18S<br>35S, 32S, 23S, 20S and 18S rpre-RNAs |

## 2.6. Immunoprecipitation of FLAG-tagged ribosomal proteins:

To immunoprecipitate FLAG-tagged Rps17, Rps3, Rps9, Rps10 and Rps20, 10 $OD_{600}$ units of cells grown in YPDA were lysed in ice-cold TMN150 by vortexing the cells with 3 volumes of 0.5mm Zirconia/Silica beads (Thistle Scientific). The lysates were incubated with 30µl of anti-FLAG M2 magnetic beads (Sigma) for 1 h at 4°C. After washing the beads three times with 1 ml TMN150, RNA was phenol-chloroform extracted and ethanol precipitated. The RNA was subsequently resolved on a 1.25% agarose gel and visualized by ethidium bromide staining (EtBr) and northern blot analysis using radioactive probes.

## 2.7. Immunoprecipitation and chemical modification of rRNA

Cells were harvested at OD600 ~1.0 and lysed in TMN150 buffer as previously described (Granneman *et al*, 2010). The lysate was incubated with 400µl IgG Sepharose beads (Amersham) for 1h at 4°C. The beads were then washed three times with 10ml lysis buffer and resuspended in 200µl TMN150 buffer. Subsequently, the beads were incubated with 1M7 (dissolved in DMSO, 30mM final) or DMSO (negative control) for 3 min at room temperature. The final concentration of DMSO was 5%. The RNA was phenol-chloroform extracted and ethanol precipitated and resolved on a 1.2% agarose gel. The (pre-)rRNAs were subsequently extracted from the gel using the MinElute kit (Qiagen) according to manufacturers instructions. Purified rRNAs were quantified on a 2100 Bionalyzer (Agilent) using an RNA Nano 6000 assay.

For analyses on *in vitro* refolded rRNA, gel purified 18S rRNA was refolded in TMN150 by heating the RNA at 65°C for 10 min and allowed to gradually cool to room temperature in a heating block. 1M7 was added to a final concentration of 15mM (5% DMSO final) and incubated at room temperature for 3 min. RNA was extracted by phenol-chloroform extraction and ethanol precipitated.

Free 40S subunits (devoid of mRNA, tRNA and initiation factors) were purified as previously described (Leshin *et al*, 2011). Briefly, cells were lysed

in a high salt buffer (50mM Tris pH 7.8, 500mM KAc, 10mM $MgCl_2$, 0.2% Triton X-100 and 2mM DTT) and clarified extracts were incubated with GTP (Guanosine-5'-triphosphate) and puromycin (1mM final each) for half an hour at 37°C. Extracts were then fractionated on linear 10-50% sucrose gradients (SW40-Ti rotor, 8 hours at 39.000 rpm). Fractions highly enriched for free 40S subunits were modified with 1M7 (15mM final) or incubated with DMSO (5% final) for 5 min at room temperature. 18S rRNA was extracted and checked by primer extension.

### 2.8. Tandem affinity purifications

For tandem-affinity purifications (TAP), cells were harvested at $OD_{600}$ ~1.0 and lysed in TMN150 buffer as previously described (Granneman *et al*, 2010). Lysates were incubated with 250µl of IgG-Sepharose beads (Amersham) for one hour at 4ºC. After washing the beads three times with 10 ml of TMN150, complexes were eluted by incubating the beads with 100 units of AcTEV protease (Invitrogen) for 1.5 hours at 18ºC. After adding $CaCl_2$ to a final concentration of 2mM, eluates were incubated with 200µl of calmodulin-coated beads (Agilent) for 1 h at 4°C, with rotation. Beads were washed four times with 1 ml TMN150 buffer containing 2mM $CaCl_2$, and then incubated with elution buffer (50mM Tris pH=7.8, 100mM NaCl, 5mM EGTA, 0.1% NP40) for 10 min at 4°C. Eluted proteins were trichloroacetic acid (TCA) precipitated overnight at 4ºC and resolved on 4-12% NuPAGE BisTris gels (Invitrogen). The membrane was blocked for 1h in blocking buffer (1xPBS, 0.1%Tween, 5% milk). Antibodies recognizing assembly factors were kindly provided by Katrin Karbstein. The blots were incubated over night at 4°C with the antibodies (Enp1, Ltv1, Rio1, Nob1 1:2000; Tsr1 1:1500), washed 3 times 10 minutes with PBST 1xPBS, 0.1%Tween) and incubated 1h at room temperature with the secondary antibody: goat anti-rabbit peroxidase conjugated (1:10.000) (Pierce Goat Anti Rabbit IgG (H+L), ThermoScientific). The membrane was then washed 3 times for 5 minutes with PBST (1xPBS, 0.1%Tween) and incubated with ECL solution according to the manufacturer's instructions (ThermoScientific). To generate the Rio1-TAP western blot results, a single blot was sequentially probed with

all antibodies.

## 2.9. Primer extension analyses

Reverse-transcription was performed using Superscript III (Invitrogen) with 8ng of purified rRNA and $^{32}$P-radiolabelled oligonucleotides. Samples were incubated at 45°C for 45 minutes after which 1µl of Exonuclease I (NEB) was added, followed by a 30 minute incubation at 37°C. Reactions were stopped by the addition of 5µl of RNA loading dye (NEB). Sequencing ladders were prepared using a plasmid containing an rDNA repeat (pBS-rDNA) and Sequenase v2.0 (USB/Affymetrix). Samples were resolved on a 6% PAA/8M Urea gels and visualized using the FLA5100 phosphoimager system.

Table 2.3. Oligonucleotides used for primer extensions

| Primer | Sequence 5'-3' |
|---|---|
| 68 | ATCGGTACTAGCGACGG |
| 69 | CCACTATTTAGTAGGTTAAGGTCTC |
| 72 | GGTTAGACTCGCTGGCT |
| 187 | TACCACAGTTATACCATGTAGT |
| 287 | ACTTGCGCTTACTAGGAATTCC |
| 288 | AAGGGCATCACAGACCTGTT |
| 343 | GATTAGCCGCAGTTGGTAAA |
| 318 | TTCATCGATGCGAGAACC |

## 2.10. Preparation of ChemModSeq libraries and high-throughput sequencing

Reverse-transcription reactions were performed using Superscript III (Invitrogen) in 20µl reactions according to manufacturers instructions using 50ng of purified ribosomal RNA and 2.5 µM random hexamers. After a

30-minute incubation at 45ºC, 1µl of Exonuclease I (NEB) and 1 µl RNaseIf (NEB) was added, followed by another 30 minute incubation at 37°C. The enzymes were inactivated by incubating the samples at 95°C for 5 minutes. The resulting cDNAs were phenol/chloroform purified, ethanol precipitated and resuspended in $H_2O$. Ligation of the 5' adapter (Table 2.5) was performed using CircLigase II (EpiCentre) according to manufacturers instructions. The ligation reactions were incubated at 60°C for 5 h and then purified using the DNA Clean & Concentrator 5 kit (Zymo Research) according to manufacturers instructions. PCR amplification was performed using Pfu polymerase (Promega), primer P5 and a barcoded primer (200 nM, see Table 5.2). Thermal cycling conditions were: 95°C for 2min, followed by 20 cycles of 95°C for 20s, 52°C for 30s and 72°C for 1 min, with a final extension of 72°C for 5min. PCR products were purified using the DNA Clean & Concentrator 5 kit. Libraries were resolved on a 2% MetaPhor agarose gel, 150-800bp products were gel-excised and purified using the MinElute kit (Qiagen) according to manufacturers instructions. All libraries were quantified on a 2100 Bionalyzer (Agilent) using a High-Sensitivity DNA assay. Individual libraries were pooled based on concentration and barcoding, and paired-end sequencing was performed on HiSeq systems by BGI (Hong Kong) and by Genepool (Edinburgh).

Table 2.4. Oligonucleotides for library preparation.

| **RT oligo** | Sequence (5'-3') |
|---|---|
| PE_hexamer | CGTGTGCTCTTCCGATCTNNNNNN |

| **5' adapter** | Sequence (5'-3') |
|---|---|
| IDX1 | 5Phos/NCGTGATNNNNNNAGATCGGAAGAGCGTCGTGTAGGG/SpC3 |
| IDX2 | 5Phos/NACATCGNNNNNNAGATCGGAAGAGCGTCGTGTAGGG/SpC3 |
| IDX3 | 5Phos/NGCCTAANNNNNNAGATCGGAAGAGCGTCGTGTAGGG/SpC3 |
| IDX4 | 5Phos/NTGGTCANNNNNNAGATCGGAAGAGCGTCGTGTAGGG/SpC3 |
| IDX6 | 5Phos/NATTGGCNNNNNNAGATCGGAAGAGCGTCGTGTAGGG/SpC3 |

| **PCR primer** | Sequence (5'-3') |
|---|---|
| P5 | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| BC1 | CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |
| BC2 | CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |
| BC3 | CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |
| BC4 | CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |
| BC5 | CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |
| BC6 | CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |

## 2.11. Data analysis

Raw data processing was carried out using tools described in the pyCRAC software package version 1.1.9 (https://bitbucket.org/sgrann/pycrac). The pyBarcodeFilter tool was used to split reads based on barcoded indices. PyFastqDuplicateRemover was used to remove potential PCR duplicates using random nucleotide information in the 5' adapter sequence. Reads were mapped to the *Saccharomyces cerevisiae* 35S gene (RDN37-1) using Novoalign version 2.07 (www.novocraft.com). Only perfectly paired reads that mapped to a single position and completely overlapped with the 20S or 18S reference sequence were considered for further analyses.

Read counts for the 18S and 20S-coding sequences, generated using pyPileup, were used to calculate reversetranscriptase (RT) drop-off rates, which are defined as the total number of reads that stops at a nucleotide position, divided by the total number of reads that cover that position. For the analyses of the pre-40S ChemModSeq data we used the Two-Channel Poisson Expectation Maximization ($TCP^{EM}$) algorithm, which was developed by our collaborator Stuart Aitken. The ChemModSeq protocol produces two channels of data, one series of read counts per nucleotide position on the rRNA for the chemically modified sample, and another for the control. In both cases, there are nucleotide positions where the polymerase is more likely to drop-off (high), and positions where drop-off is less likely (low). Assuming that high and low drop-off rates are approximately constant along the rRNA, we assigned each position to one of three categories: high drop-off in both channels (no assignment), low drop-off in both channels (unmodified), high drop-off in the modified channel and low in the control (modified). The drop-off rates ($\lambda_1$ $\lambda_2$) and the probabilities of a position belonging to a category (pi1 pi2 pi3 pi4) were used to calculate the probability of the observed number of drop-offs ($d_i$) in modified and control 4 channels at position i given the observed read count ($c_i$) and (inferred) values of $\lambda_1$ $\lambda_2$ using a Poisson model. We calculate the likelihood as the product of these probabilities for all positions, and all categories. For example, for the first category where drop-off rates in both channels are $\lambda_1$ we compute:

$p_{i1}$*Poisson($d^{modified}_i$; $c^{modified}_i\lambda_1$)*Poisson($d^{control}_i$; $c^{control}_i\lambda_1$)

The assignment of positions to categories, and the calculation of $\lambda_1$ and $\lambda_2$ are performed iteratively using an expectation maximization algorithm that maximizes the likelihood of the data. In practice, the $TCP^{EM}$ requires, on average, 22 iterations to converge (between 5 and 52), and on convergence the assignment of positions to categories is crisp: we use thresholds of 1.0 and 0.9 to decide class membership (from the range 0-1).

## 2.12. Western Blot

The equivalent of 5 $OD_{600}$ units of cells were lysed by incubating the cells in 0.2N NaOH at room temperature for 5 minutes. The samples were denatured at 95ºC in 1x SDS loading buffer (NuPAGE) and 10µl was loaded on 4-12% gradient polyacrylamide gels (NuPAGE). The proteins were transferred to nitrocellulose membranes using the iBlot, according to the manufacturers procedures (Invitrogen). The membranes were blocked for 1h in blocking buffer (1xPBS, 0.1%Tween, 5% milk). The FLAG-tagged constructs were detected by incubating the membranes for 1h at room temperature with anti-FLAG HRP conjugated antibodies (Sigma; 1:5000). After 3 rounds of 5 minute washes with PBST (1xPBS, 0.1%Tween) the protein were visualised by Enhanced Chemiluminescence (ECL; ThermoFisher). The incubation times and concentrations used for the antibodies that recognize the various assembly factors are described above in section 2.8.

## 2.13. Growth complementation assay

Cells were grown in SGR-/-URA medium to saturation. Ten-fold serial dilutions were prepared starting from 1 $OD_{600}$ unit of cells and spotted on minimal medium agar plates containing galactose or glucose.

## 2.14. 4-thio-uracil pulse chase

This protocol was developed by David Barrass from Jean Beggs's lab. The parental strain (BY4741) was grown in SD-/-URA until an $OD_{600}$ ~0.6-0.8. The Rps17 depletion strain (YSG766) was grown in SGR-/-URA and then shifted to SD-/-URA for 8h. The cells were then pulsed with 4-thio-uracil (final concentration of 10μM) for 2 minutes, followed by a chase with uridine (final concentration of 500μM) for 5, 30 and 60 minutes. After each time-point the cells were immediately fixed by adding half a volume of methanol. The cells were subsequently pelleted, washed with water and then stored at -80°C. The cells were resuspended in AE buffer (50mM NaOAc pH=5.3, 1mM EDTA, 1% SDS) and lysed by vortexing. An equal volume of tris-buffered phenol was added to the lysate. The sample was then incubated at 65°C for 45 minutes and then the rRNA was phenol-chloroform extracted and precipitated with ethanol. The pellet was then resuspended in 300μl TE buffer (10mM Tris, 1mM EDTA, pH 7). The 4-thio-uracil-labeled RNA transcripts were then biotinylated by incubating the total RNA with EZ-link HPDP-Biotin (N-[6-(Biotinamido)hexyl]-3'-(2'pyridyldithio)propionamide)(ThermoFisher) according to manufacturer's instructions. The biotinylated RNA was purified by passing the samples through a 2ml Zeba Column (ThermoFischer) according to the manufacturer's instructions. The isolated RNA was then incubated for 1h in 80% ethanol to remove the residual HPDP-Biotin and then precipitated by centrifugation. The RNA sample was resuspended in water and NaTMg buffer was added to a 1X concentration (10X NaTMg: 10mM TrisCl pH=7, 100mM NaPi pH=6.8, 200mM NaCl, 25mM $MgCl_2$, 0.4% Triton 100X) and twice the NaTMg volume of 0.5M NaPi buffer pH=6.8. The biotinylated rRNA was then purified using 50 μl of Streptavidin Magnetic Beads Dynabeads MyOne C1 beads per sample (Invitrogen). Before adding the samples to the beads, the beads were washed NaPTMg buffer and then blocked with 20μl glycogen (20mg/ml, Roche) for 20min. The biotinylated RNA was incubated with the streptavidin beads for 2h at 4°C. After the incubation the beads were washed 5 times with 1X NaPTMg and 3 times with TEN1000 (10mM TrisCl pH=7.5, 1mM EDTA pH=8, 1M NaCl). The biotynylated rRNA was eluted from the beads for 5 minutes with 50μl of 0.7M β-mercaptoethanol at room temperature. The eluate was then ethanol

precipitated and then resolved on a 1.25% agarose gel.

## 2.15. FISH

Cells were grown to an OD of 0.5 in YPGR media and then shifted to YPDA media. Cells were fixed at 0h, 4h and 8h after the shift by incubating the cells with formaldehyde (3.7% final) for 10 min at room temperature. Cells were then re-suspended in spheroplasting solution (1.2M sorbitol, 0.1M KPO4, 1 x protease inhibitors tablets (Roche), 68mM beta-mercaptoethanol, 1µl RNase inhibitor (Promega), 1X Zymolyase (Amsbio)) and incubated with 1X (0.2mg/ml) Zymolycase at 30°C for 8min. Cells were then washed with solution A (1.2M sorbitol, 0.1M KPO4), placed in poly-L coated coverslips and dehydrated in 100% ethanol over night at -20°C. Cells were incubated in in 2XSSC media for 5min, then 2XSSC 40% formamide media for 5 minutes. Cells were then incubated in hybridisation solution (2XSSC, 40% formamide, 2.5mg BSA, 1 µl RNaseIn, 100µg fish sperm DNA) and 100ng Cy3-ITS1 oligo. The oligo was incubated with the cells for 4h at 37°C. The cells were then washed twice 15 minutes with 2XSSC 40% formamide 37°C, twice 15 minutes with 1XSSC at room temperature and once for 15 minutes with PBS at room temperature. Coverslips were mounted with Vectashield Mounting Medium (H-1000, Vector Laboratories) after the DNA was stained with DAPI (0.5µg/ml final). Images were taken with a Zeiss Axio ImagerZ1 microscope. Cy3_ITS1 oligo: 5'-Cy3-GCTGGACTCTCCATCTCTTGTCTTCTTGCCCAGTAAAAGCTCTCATGCTCTC-Cy3-3'

## 2.16. Fixing cells for fluorescence microscopy

Cells were grown in SD-/-URA medium to an $OD_{600}$ of ~0.4 and fixed with 3.7% formaldehyde for 30 min at room temperature. DNA was stained using DAPI (0.5µg/ml) and samples were mounted with Vectashield Mounting Medium (H-1000, Vector Laboratories). Images were captured with a Zeiss Axio ImagerZ1 microscope.

### *2.17. In vitro* binding assay

The *in vitro* binding assay was performed essentially as previously described (Grünwald & Bono, 2011). GST proteins fused to the wild-type or mutant Rps17 NESs were incubated with recombinant Crm1 and/or RanGTP (4µg each; purified by Atlanta Cook) in binding buffer (40mM HEPES pH=7.5, 50mM NaCl, 15% Glycerol, 0.02% NP40, 4mM $MgCl_2$, 2mM DTT) to a final volume of 50µl and incubated on ice for 1h. The protein mixes where then immobilized on 25µl pre-washed (20mM HEPES pH=7.5, 150mM NaCl, 0.01% NP40) and blocked (25µl glycogen) Glutathione Sepharose Beads (GE Healthcare) and incubate on a rotating wheel at 4°C for 1h. The resin was washed 3 times with 500µl of binding buffer and eluted with 8ul 2xNuPage, boiled and loaded on a 4-12% gradient polyacrylamide gel. Proteins were visualized by Coomassie brilliant blue staining.

# CHAPTER 3

# Probing the Dynamics of pre-rRNA Folding in 40S Assembly Intermediates

## 3.1. Introduction and aim

Pre-ribosomes are believed to undergo a large number of structural rearrangements during ribosome assembly, many of which involve rRNA folding events. Assembly factors are believed to play a key role in these structural rearrangements, but how they function in this process remains largely unclear. To visualise these dynamic changes and to gain insights into the function of assembly factors, many groups are using Cryo-EM microscopy. However, these studies are quite problematic due to the fact that ribosome assembly intermediates can be quite unstable and isolating homogenous populations of these particles is not trivial. This makes their high-resolution structural characterization very challenging. As an alternative (biochemical) approach, we decided to develop ChemModSeq. This method provides information on changes in rRNA flexibility during ribosome assembly, which provide a read-out of rRNA structural changes. We reasoned that we would be able to get more detailed insights into rRNA structural rearrangements during ribosome synthesis by combining ChemModSeq with SHAPE chemical on isolated particles at various stages of maturation. Because the various pre-40S assembly intermediates have a different set of assembly factors, our data enable us to correlate changes in rRNA structure with binding or release of assembly factors. This would enable us to get mechanistic insights into how assembly factors influence rRNA folding steps.

This chapter focuses on the rRNA structural rearrangements that take place during the late (cytoplasmic) stages of pre-40S maturation: the conversion of the 20S pre-rRNA into mature 18S rRNA (fig 3.1).

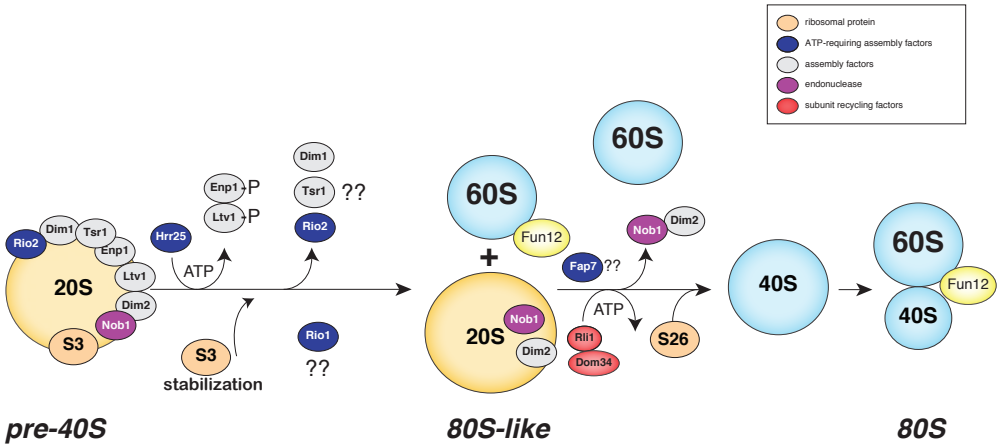



**Figure 3.1 Schematic representation of the pre-40S cytoplasmic assembly**
Phosphorilation of Enp1 and Ltv1 (by Hrr25) leads to their release and allows the stable incorporation of Rps3. Tsr1, Rio2 and Dim1 dissociate before the 80S-like complex is formed but the exact timing is not known. The 20S pre-rRNA is cleaved by Nob1 forming the mature 18S rRNA. Its cleavage mimics a translation initiation like cycle. The mature 40S particle is then released.

## 3.2. The 20S and 18S rRNAs can be effectively purified from pre-ribosomes

To isolate 20S pre-rRNA and 18S rRNA I used HTP or TAP tagged pre-40S assembly factors. Enp1, Ltv1, Tsr1, Rio2, Nob1 and Rio2 were used as baits to purify 20S pre-rRNA species. The translation initiation factor Fun12 was used to purify the 18S rRNA from the 80S mature ribosome. Purifying 20S pre-rRNA from native 80S-like complexes proved to be very difficult due to the fact that it is very low abundant, possibly because the 20S pre-rRNA in this complex is rapidly converted to 18S rRNA. Depletion of Fap7 leads to the accumulation of 20S pre-rRNA in defective 80S-like particles containing Fun12 (Strunk *et al*, 2012; Granneman *et al*, 2005). Therefore, we used Fun12 as bait in Fap7 depleted cells to purify the 20S pre-rRNA from the 80S-like particles (fig 3.2A).

To determine if the TAP or HTP tags interfered with protein function, I performed a northern blot analysis on total RNA purified from the strains expressing the tagged proteins. The mature 25S and 18S rRNAs were detected by ethidium bromide staining. Pre-rRNA precursors were detected by hybridizing the northern blot with oligonucleotides 004 and 003 (fig 3.2B).

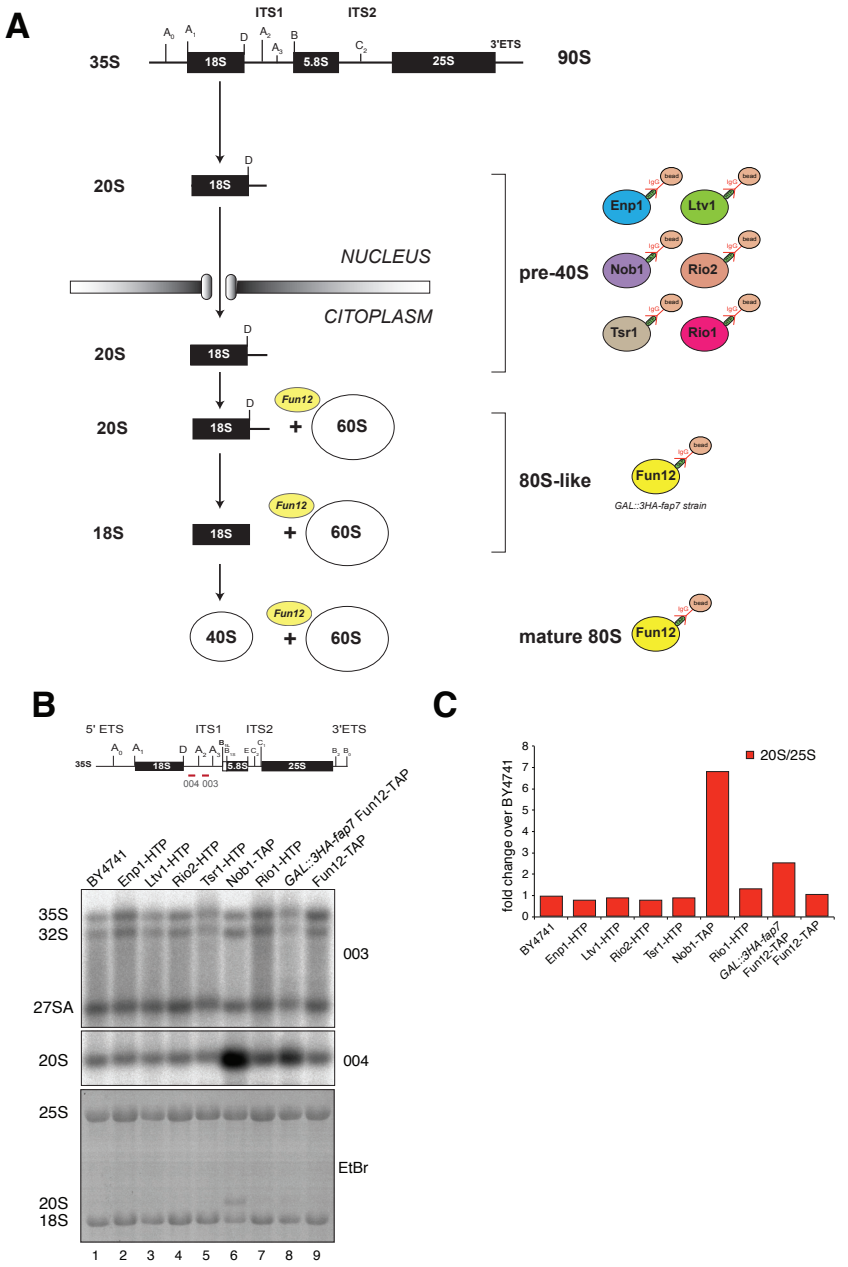



**Figure 3.2. Affinity purification of 20S pre-rRNA precursors and mature 18S rRNA.**
(A) The 40S subunit rRNA processing pathway and the HTP or TAP tagged assembly factors used to affinity purify the 20S pre-rRNA and 18S rRNA. A Fun12-TAP strain was used to purify the 18S rRNA from 80S ribosomes and 20S pre-rRNA from 80S-like complexes in a *GAL::3HA-fap7* strain. (B) Total RNA was extracted from the parental strain (BY4741) and strains expressing HTP or TAP-tagged proteins (indicated above the panel) and resolved on 1.25% agarose gels. The 20S pre-rRNA associated with the 80S-like complexes was purified from a *GAL-3HA::fap7* Fun12-TAP strain that was grown in glucose for four hours (lane 8). Mature rRNAs were detected by ethidium bromide (EtBr) staining. Pre-rRNA species were detected by Northern blotting using oligonucleotides 003 and 004. (C) Quantification of the 20S pre-rRNA levels on the northern blot shown in (B). The 20S levels were normalized to 25S rRNA levels and then divided by the number obtained for the parental strain BY4741.

The quantification of the results (fig 3.2C) showed that HTP tagging Enp1, Ltv1, Tsr1, Rio2, Rio1 and the TAP tagging of Fun12 did not noticeably affect pre-rRNA processing. Surprisingly, although C-terminally TAP-tagging Rio1 was reported to affect the function of the protein (as judged by 20S pre-rRNA accumulation and delay in 18S rRNA formation in this strain; (Soudet *et al*, 2010)), this was not the case for the strain expressing a C-terminally HTP tagged Rio1. TAP or HTP-tagging Nob1 at its C-terminus was previously shown to cause a significant accumulation of 20S pre-rRNA, suggesting that the tags interfere with Nob1-dependent cleavage at site D (Granneman *et al*, 2010; Fatica *et al*, 2003). This strain was therefore used to determine how a delay in cleavage at site D might affect rRNA structure. My northern blot analysis showed that TAP-tagged Nob1 accumulated ~7 times more 20S pre-rRNA compared to the parental strain (fig 3.2C). The slight 20S accumulation in the Fun12-TAP *GAL::3HA-fap7* strain was the result of a short (4h) Fap7 depletion. This slight accumulation of pre-20S rRNA in the 80S-like particles was sufficient to enable us to purify enough material for structural probing studies (fig 3.2B lane 8 and C). In conclusion, apart from Nob1-TAP and Fun12-TAP *GAL::3HA-fap7* strains, the tags did not significantly interfere with protein function.

## 3.3. Affinity purified pre-40S particles can be classified into ”early”, “middle” and “late” intermediates

The fact that Enp1, Ltv1, Rio2, Tsr1, Nob1 are associated with the pre-40S complexes in both the nucleus and the cytoplasm brought into question the homogeneity of the pre-40S complexes that I purified. To address this, I looked for ways in which I could somehow distinguish the various assembly intermediates. Upon nuclear export, Dim1 methylates sites A1780 and A1781 of the 20S pre-rRNA (Lafontaine *et al*, 1994). As a result, the presence of A1780-A1781 methylation can be interpreted as a mark of cytoplasmic pre-40S intermediates. Ralph Hector performed primer extension to assess the levels of A1780-A1781 modification of 20S pre-rRNAs enriched by Enp1, Ltv1, Tsr1 and Fun12 respectively (fig 3.3A).

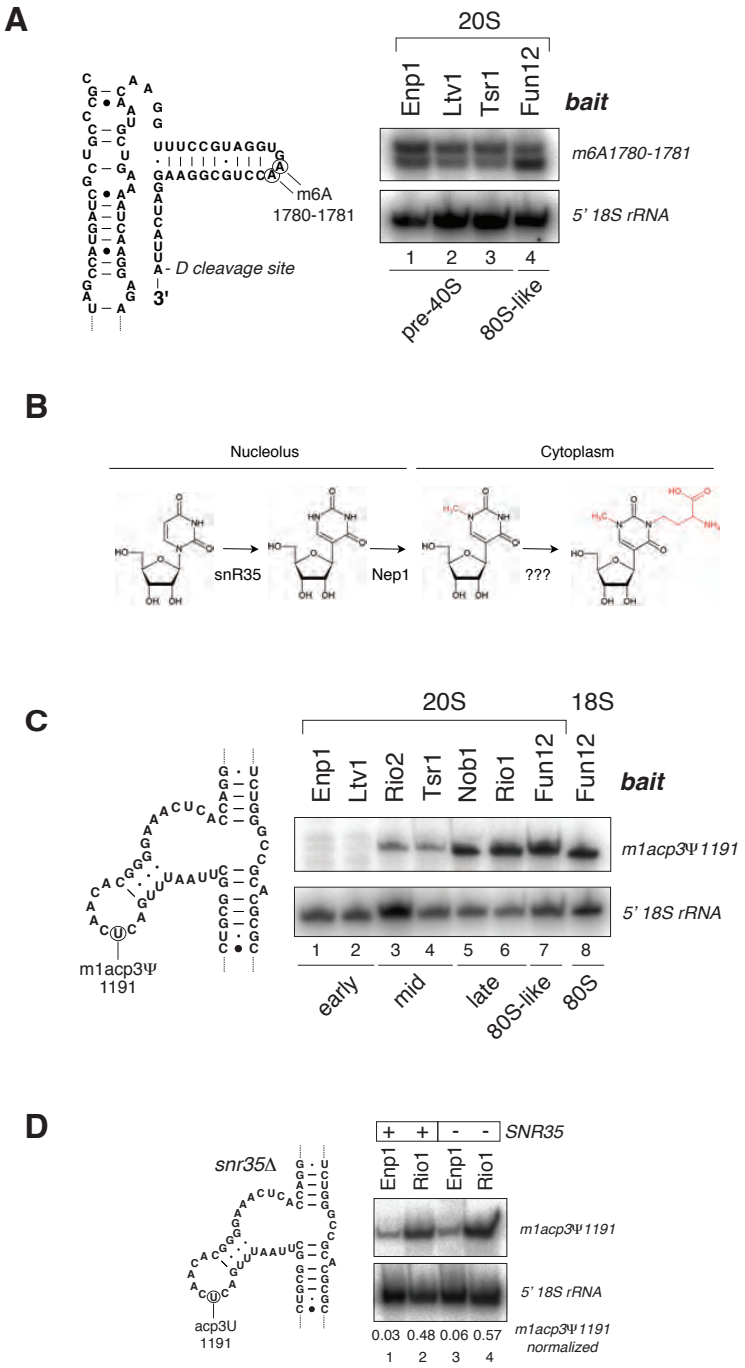



**Figure 3.3 Pre-40S complexes can be classified into early, mid and late particles.**
(A) (performed by Ralph Hector) Detection of cytoplasmic methylation at A1780 and A1781 in 20S pre-rRNAs by primer extension. As loading controls, primer extension was performed to measure the level of cDNAs that terminated at the 5' end of the (pre-) rRNAs (A1 cleavage site). (B) Hypermodification of U1191 in the 18S rRNA. U1191 is first pseudouridinylated by snR35, then N1 methylated by Nep1 in the nucleolus and finally aminocarboxypropyl (acp)-modified in the cytoplasm by an unknown enzyme. (C) Detection of U1191 acp-modification in 20S and 18S rRNAs by primer extension. Baits used to purify the individual pre-rRNA species are indicated above the panel. Loading controls were as in (A). (D) Same as in (C) but compared to strains lacking the *SNR35* gene

The primer extension signal at the 5' end of the 18S rRNA was used as a loading control. The 20S pre-rRNA enriched in all three pre-40S complexes displayed variable levels of A1780 and A1781 methylation. Moreover, the pattern of A1780 and A1781 methylation in the pre-20S from the pre-40S intermediates was different from the pattern observed in pre-20S from the 80S-like complexes. This suggested that the isolated pre-40S complexes were a mixture of cytoplasmic and nuclear intermediates for the pre-40S complexes (fig 3.3A). In conclusion, by measuring the level of A1780-A1781 methylation we were able to determine that the affinity purified pre-40S complexes are mostly, but not exclusively, cytoplasmic assembly intermediates.

Just before cleavage of 20S pre-rRNA at site D, an aminocarboxy propyl (acp) group is added to a pseudouridinylated U at position 1191 in the 18S rRNA (Brand *et al*, 1978) (fig 3.3B). Because this modification completely blocks reverse transcription, Ralph Hector and myself assessed the levels of acp modification of the purified (pre-)rRNAs by primer extension. The results showed that the U1191 acp modification varied significantly, depending on the bait used for 20S pre-rRNA purification. Compared to the 20S purified from 80S-like particles and the 18S rRNA purified from the mature 80S (fig 3.3.C lane 7 and 8), 20S pre-rRNA purified using Ltv1 and Enp1 as baits showed very little acp-modification at U1191 (fig 3.3C lanes 1 and 2), whereas intermediate levels of acp modification could be detected in 20S pre-rRNA isolated from Rio2 and Tsr1 particles (fig 3.3C lanes 3 and 4). The levels of acp-modification of Nob1 and Rio1 associated 20S pre-rRNA (fig 3.3C lane 5, 6) were similar to the levels measured the mature 18S rRNA. Based on these results we classified the purified pre-40S assembly intermediates in "early" (Enp1, Ltv1), "middle" (Tsr1, Rio2), "late" (Nob1, Rio1) particles (fig 3.3D).

The acp-modification of U1191 is the last in a series of modifications on this nucleotide and is preceded by SNR35-dependent pseudouridylation and Nep1-dependent N1 methylation (fig 3.3.B). As a result, it could be argued that the strong primer extension signal for the U1191 might not only be the result of the late acp modification but could be a more general consequence of U1191 hyper-modification. To address this, I assessed the U1191 modification on 20S pre-rRNA purified with Enp1 (early) and Rio1 (late) intermediates from *SNR35* deletion strains. In the absence of *SNR35*, U1191 cannot be pseudouridylated

and methylated at N1 methylation, however, acp modification is not affected (Liang *et al*, 2009). The primer extension signal of U1191 was almost identical in the presence or absence of *SNR35*, confirming that the acp modification is exclusively responsible for the primer extension stop (fig 3.3D). In conclusion, the acp-modification can be considered an accurate mark of “early”, “middle” or “late” pre-40S intermediates.

### 3.4. Late cytoplasmic pre-40S complexes contain fewer assembly factors

We next asked whether there was a difference in the composition of assembly factors between the purified pre-40S complexes. To address this, Sander Granneman purified Rio2-HTP (middle) and Rio1-HTP (late) particles using a single step affinity purification method (Oeffinger *et al*, 2007), followed by label-free quantitative mass spectrometry (in collaboration with Thierry Le Bihan’s group). The results showed that Rio2-HTP efficiently enriched for all assembly factors apart from Rio1. In contrast, Rio1-HTP significantly co-precipitated only Nob1 and Dim2 (fig 3.4A). Moreover, Rio1 and Rio2 presence in the pre-40S complexes was mutually exclusive.

To validate these findings, I performed western blot analysis on tandem-affinity purified middle (Rio2-TAP and Tsr1-TAP) and late (Rio1-TAP) particles. Rio2-TAP and Tsr1-TAP enriched for the same assembly factors: Ltv1, Enp1, Rio2, Tsr1 and Nob1, although Nob1 appeared to be less efficiently enriched (fig 3.4B). In agreement with the mass spectrometry data, Rio1-TAP only significantly co-precipitated Nob1 (fig 3.4C). As Rio1 antibodies were not available and Dim2 antibodies were quite unspecific, their presence could not be validated by western blotting.

In conclusion, the late pre-40S complexes (purified using Rio1 as bait) contain a much smaller number of assembly factors compared to the earlier complexes.

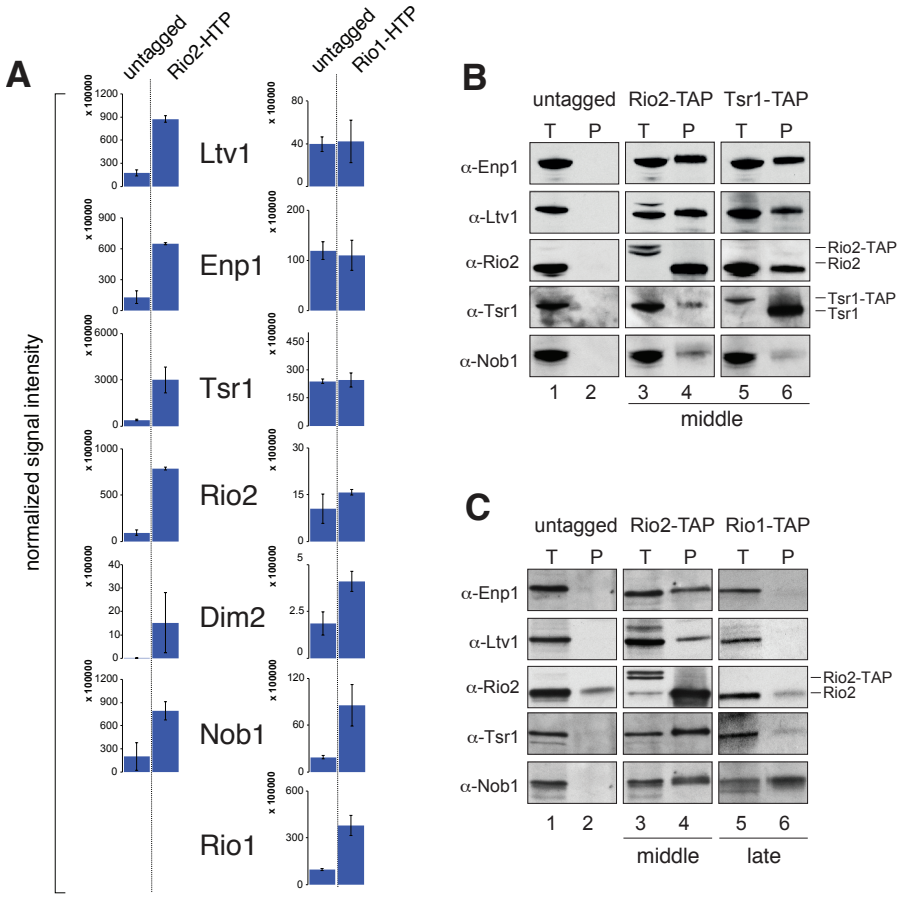



**Figure 3.4 Rio1 associated pre-40S complexes contain fewer 40S assembly factors.** (A) (performed by Sander Granneman) HTP-tagged Rio2 and Rio1 were affinity purified and associated proteins were quantified by label-free LC-MS. Shown are the averages and standard deviations from 2 (Rio2) or 3 (Rio1) replicate experiments. The y-axis indicates the relative abundance of the proteins in the purified complexes. No Rio1 peptides were found in the Rio2 affinity purification. (B and C) Western blot analysis of protein composition of mid (Rio2, Tsr1) and late (Rio1) pre-40S particles purified by Tandem Affinity Purification (TAP) using TAP-tagged Rio1 and Rio2 strains.

## 3.5. Applying ChemModSeq to pre-40S complexes

Having characterized the purified particles in detail, the next step was to perform our structure probing method on the purified pre-ribosomes to measure differences in rRNA flexibility. The rRNA in purified pre-ribosomes, still bound to the IgG-Sepharose resin, was modified *in vitro* using 1M7 for three minutes at room temperature. The 1M7 chemical modifies the 2' hydroxyl

of the RNA backbone if the nucleotide is in a flexible conformation, while nucleotides in constrained conformations are less reactive to the chemical (fig 3.5A). The modified (sample) and unmodified (DMSO control) rRNA samples were resolved on an agarose gel to separate the 20S pre-rRNA species from the 18S and 25S rRNA contaminants. The rRNA of interest was gel extracted and then its integrity checked on a Bioanalyser (fig 3.5B and C). The rRNA was subsequently reverse transcribed using random primers (see Chapter 1 and 2 for details), which generated a library containing cDNAs with a broad size range. ChemModSeq libraries were then made from these cDNA pools by adding Illumina specific barcodes that allow paired-end sequencing (see Chapter 2 for details). Sander Granneman has developed a data-processing pipeline for quantifying the level of nucleotide modification for each position in the rRNA.

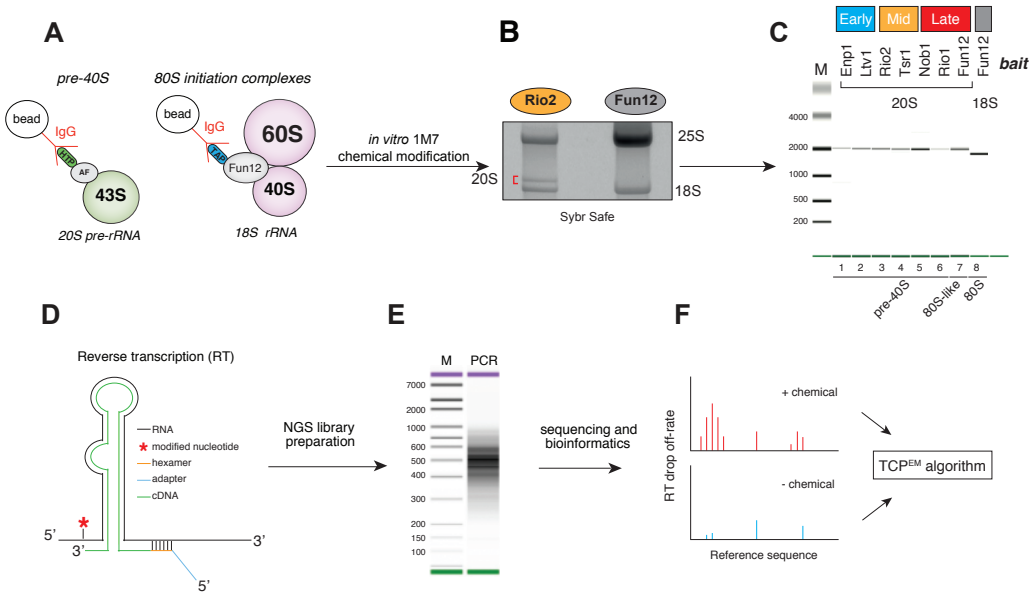



**Figure 3.5 ChemModSeq experimental workflow.**
(A)Pre-40S particles at different stages of assembly are purified using strains expressing TAP or HTP-tagged assembly factors. Particles are chemically modified and RNA is extracted and 20S and 18S rRNA are gel purified (B) and then the RNA integrity is checked on a Bioanalyzer (C). (D) cDNA libraries, generated by random priming, were PCR amplified and (E) sequenced on Illumina HiSeq systems. (F) Chemically modified samples were compared to solvent treated samples using a two-channel Poisson expectation maximization algorithm (TCP$^{EM}$) to identify nucleotides that had the highest likelihood of being modified.

Figure 3.6 schematically represents the ChemModsSeq library design and the data analysis pipeline. Each cDNA formed has two Illumina compatible adapters that have two different barcodes. The 3' barcodes are used to differentiate between replicates and to trace back the details of the individual experiments. The 5' adapters were used to identify the conditions of each experiment included in one experimental repeat (such as the control and 1M7 treated samples). These barcodes are used for demultiplexing in the first two steps of the data analysis. The next step was to remove potential PCR duplicates, which are cDNA products that have been amplified multiple times during the amplification reaction. Some cDNAs could have been amplified more efficiently and therefore it is important to remove such duplicates. In order to do that, paired-end reads that had identical nucleotide sequences as well as identical barcode sequences were collapsed into a single read. The sequences in the resulting Fasta files are then aligned to the genome of interest, in my case to the *S.cerevisiae* rDNA gene (RDN37-1).

Using the mapping coordinates returned by the aligner, I generated an overview of the total number of reads that covered each nucleotide and the total number of reads that stopped at each nucleotide (drop-off counts). This was useful in determining the drop-off rates for each nucleotide in the sequence, which are calculated by dividing the drop-off counts by the coverage. A high drop-off rate value for nucleotide N in the 1M7 treated sample compared to the negative control (DMSO) would indicate that nucleotide N+1 has a high probability of being modified.

What difference in drop-off rates between the 1M7 treated and untreated sample would be considered statistically significant? To answer this question, Stuart Aitken developed an Expectation Maximization (EM) algorithm ($TCP^{EM}$) that estimates from the sequencing data the probability that a nucleotide has been modified and highlights those nucleotides that were most likely modified. Up to this point, the negative control and 1M7 treated sample were treated as separate samples in separate files. Before the $TCP^{EM}$ algorithm is applied, for each experimental condition, the negative control and 1M7 treated samples from all replicates summed respectively, as the algorithm does not know how to deal with biological or technical replicates (fig 3.6 step 7). The $TCP^{EM}$ then determines for each nucleotide in the sample if the drop-off events in the treated

sample are statistically significantly higher than the control (DMSO) sample.

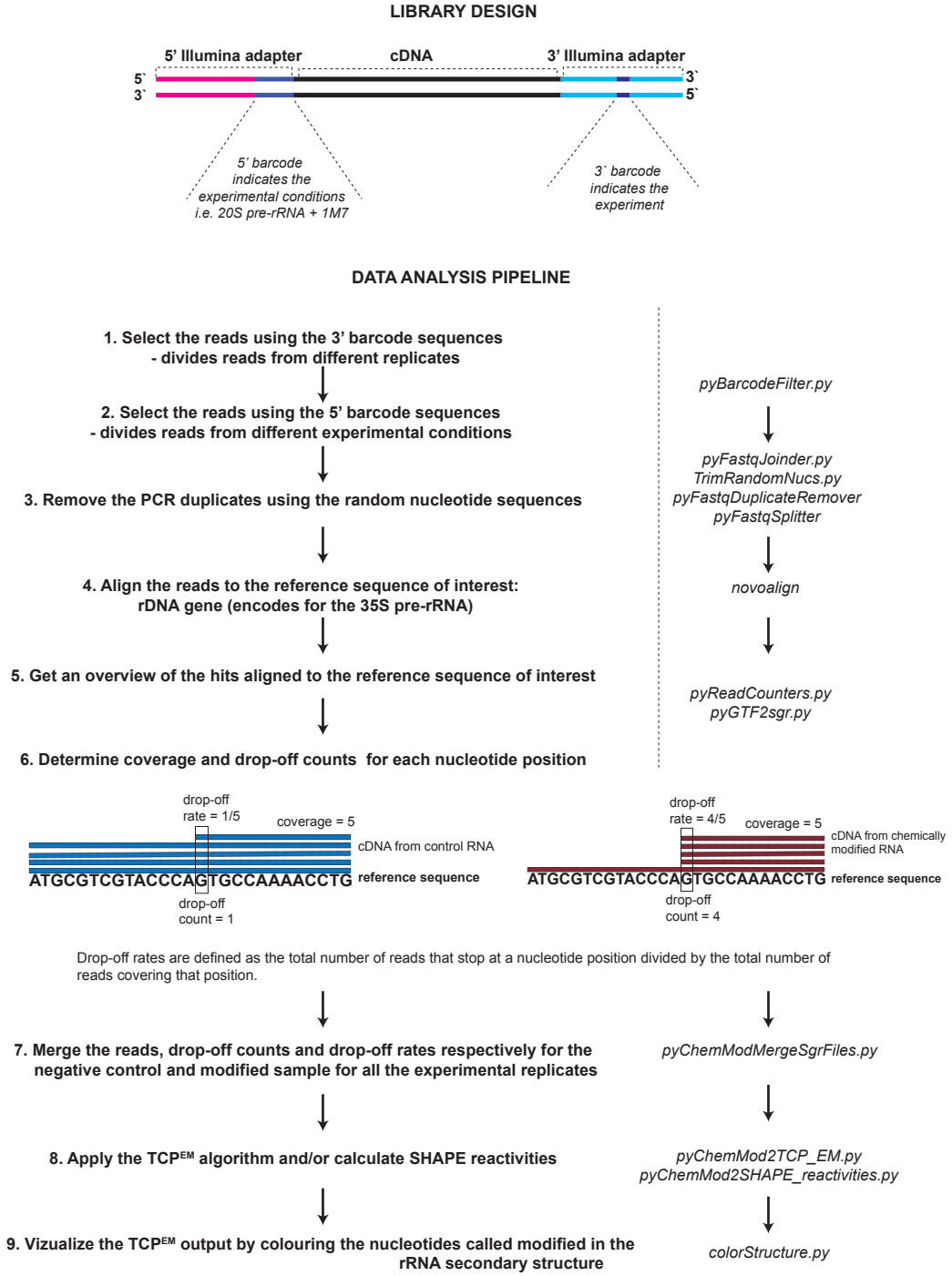



**Figure 3.6. The pyChemMod data analysis pipeline.**
An overview of the library design and an outline of the analysis steps, including the scripts used during the analysis.

The algorithm assumes that the ChemModSeq drop-off counts fit a Poisson distribution: (1) they represent discrete values, (2) the drop-off counts for a nucleotide is independent from the drop-off counts of the neighbouring nucleotides and (3) the probability of having two drop-off events very close to each other is zero. The Expectation Maximization (EM) algorithm is used to calculate a λ value, which is the expected mean of the drop-off rates of all the nucleotides in the sequence. For both the treated and the control sample the EM algorithm generates two clusters, one with a high λ value and one with a low λ value (fig 3.7).

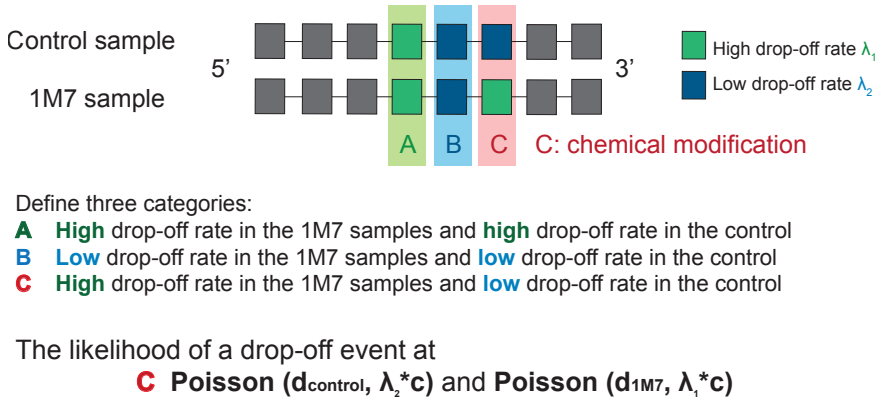



**Figure 3.7. The TCP$^{EM}$ algorithm.**
The 1M7 treated sample and the negative control were fitted to Poisson distributions respectively. The Expectation Maximization (EM) algorithm is used to calculate the mean λ value, which is the expected mean of the drop-off rates "d" for all the nucleotides in the sequence. The EM algorithm generates two clusters (for each Poisson distribution), one with a high λ value and one with a low λ value. As a result, a nucleotide can be assigned to one of 3 categories: (A) high drop-off rates in both (the stop was not caused by the chemical), (B) low drop-off rates in both (the chemical did not react with the nucleotide), (C) high in treated and low in control (the nucleotide is likely modified)

The known drop-off rates for each nucleotide are then compared to the expected high and low drop-off rates values (λ) generated by the EM algorithm in the control and treated sample respectively. As a result, a nucleotide can be assigned to one of 3 categories: (A) high drop-off rates in both (the stop was not caused by the chemical), (B) low drop-off rates in both (the chemical did not react with the nucleotide), and, most importantly, (C) high in treated and low in control (the nucleotide is likely modified (fig 3.7). The TCP$^{EM}$ then determines for each nucleotide the probability that it belongs to any of these

categories. I considered as being chemically modified only the nucleotides that fit category 3 with a probability of 0.9 or higher. The idea behind this was that this would limit the number of false positive nucleotides.

The TCP$^{EM}$ algorithm essentially generates a binary readout (i.e. modified or unmodified), which is very useful if you simply want to know what nucleotides have the highest likelihood of being modified. This enabled us to generate an overview of nucleotides of which we were very confident that they were differentially modified between rRNA samples. However, subtle changes in reactivity to chemical probes can also be informative and because of the high p-value threshold used by the TCP$^{EM}$ algorithm some of these changes might be missed by the algorithm. Therefore, for these analyses we used both the output of the TCP$^{EM}$ algorithm as well as drop-off rates to guide us to regions where there are differences in RNA flexibility.

The ChemModSeq drop-off counts can also be used to calculate SHAPE reactivities, which are the more conventional way to indicate how reactive a nucleotide was to a chemical probe (Deigan *et al*, 2009; Ding *et al*, 2014). This generates a value between 0 and 2, with 2 indicating very high reactivity and 0 indicating unreactive. To calculate such reactivities I used the pyChemMod2SHAPE_reactivities.py script (fig 3.6). SHAPE reactivities were mainly used for analysing the large subunit chemical probing data (see Chapter 5).

## 3.6. ChemModSeq data reveal a stabilization of the 3' major domain during late stages of 40S assembly

The heat map in Figure 3.8A displays the ChemModSeq output for all the samples that I processed and essentially provides an overview of rRNA structural changes during the conversion of early pre-40S particles (lanes 1 and 2) to late pre-40S complexes (lanes 5-7). The first two columns show results from negative control samples (average drop-off rates from untreated 20S pre-rRNA isolated from early pre-40S or late pre-40S particles). Columns 1-7 represent a snapshot of rRNA flexibility for rRNA intermediates analysed

here. Dark grey coloured boxes represent nucleotide positions that had high drop-off rates and therefore reacted strongly with 1M7.

I next compared the ChemModSeq data with the primer extension data that I generated for the methylated adenosines in the 3' end of 18S rRNA (fig 3.8B) and the U1191 acp modification (fig 3.8C). The ChemModSeq results are in good agreement with the primer extension data, indicating that the method can accurately and reliably detect sites of (natural) modification in RNAs.

I next applied the $TCP^{EM}$ algorithm to the data to identify regions that displayed statistically significant differences in 1M7 reactivity between the different pre-rRNAs. Overall, the number of nucleotides called modified by the $TCP^{EM}$ gradually decreased from early to late particles. This trend reflects the overall changes in rRNA flexibility and describes a clear-cut stabilization of rRNA structure during the maturation of the 40S subunit (fig 3.8D).

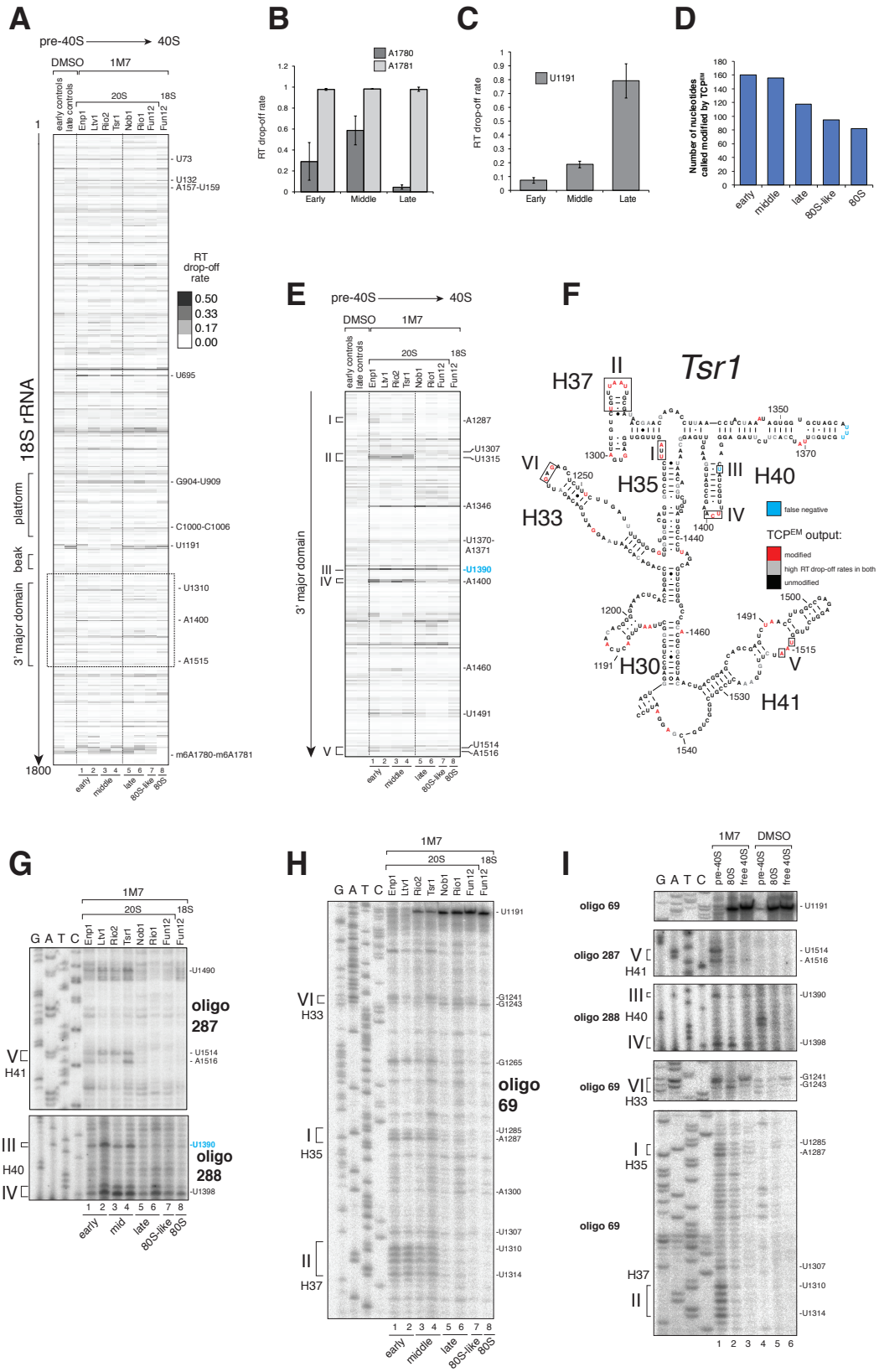

A
pre-40S
40S
DMSO
1M7
early controls
late controls
20S
18S
Enp1
Ltv1
Rio2
Tsr1
Nob1
Rio1
Fun12
Fun12
18S rRNA
platform
beak
3' major domain
U73
U132
A157-U159
U695
G904-U909
C1000-C1006
U1191
U1310
A1400
A1515
m6A1780-m6A1781
RT drop-off rate
0.50
0.33
0.17
0.00
early
middle
late
80S-like
80S
B
A1780
A1781
RT drop-off rate
Early
Middle
Late
C
U1191
RT drop-off rate
Early
Middle
Late
D
Number of nucleotides called modified by TCPEM
early
middle
late
80S-like
80S
E
pre-40S
40S
DMSO
1M7
3' major domain
A1287
U1307
U1315
A1346
U1370-A1371
U1390
A1400
A1460
U1491
U1514
A1516
F
H37
II
Tsr1
VI
H33
H35
III
H40
IV
H30
V
H41
false negative
TCPEM output:
modified
high RT drop-off rates in both
unmodified
G
1M7
20S
18S
G A T C
U1490
oligo 287
U1514
A1516
III
H40
IV
U1390
oligo 288
U1398
early
mid
late
80S-like
80S
H
1M7
20S
18S
G A T C
U1191
VI
H33
G1241
G1243
G1265
oligo 69
I
H35
U1285
A1287
A1300
II
H37
U1307
U1310
U1314
early
middle
late
80S-like
80S
I
1M7
DMSO
G A T C
pre-40S
80S
free 40S
oligo 69
U1191
oligo 287
V
H41
U1514
A1516
III
oligo 288
H40
IV
U1390
U1398
oligo 69
VI
H33
G1241
G1243
I
H35
U1285
A1287
oligo 69
H37
II
U1307
U1310
U1314

**Figure 3.8 A digital snapshot of pre-rRNA secondary structures (A-I)**
ChemModSeq and primer extension results obtained using rRNAs extracted from in vitro 1M7 modified particles. (A) Heat maps of average RT drop-off rates in 18S coding sequences (n >= 2) generated from chemically modified or solvent treated pre-rRNAs. (B) ChemModSeq data for the Dim1-dependent methylations at A1780 and A1781. (C) ChemModSeq data showed that the 20S pre-rRNAs associated with late pre-40S complexes have significantly higher levels of acp-modification at U1191. Average U1191 drop-off rates were generated from at least two ChemModSeq experiments on unmodified 20S pre-rRNA. "Early" includes data from 20S pre-rRNAs isolated using Ltv1-HTP and Enp1-HTP as baits (n=2), "Middle" includes data from 20S pre-rRNAs isolated using Tsr1-HTP as bait (n=3) and "Late" includes data from 20S pre-rRNAs isolated using Fun12-TAP as bait (n=3). (D) Pre-rRNAs in early and mid pre-40S complexes adopt more flexible conformations compared to late particles. Plotted are the average number of nucleotides called modified by the $TCP^{EM}$ algorithm (y-axis) for early, middle, late, 80S-like and mature 80S translation initiation complexes (x-axis). (E) Heat map describing the ChemModSeq output for the the 3' major domain. (F) The 3' major domain secondary structure represents the ChemModSeq results for the middle (Tsr1) particle. Red nucleotides were indicated as modified by the $TCP^{EM}$ algorithm. Blue nucleotides were visibly modified by 1M7 but not called modified by the $TCP^{EM}$ algorithm (false negative). (G-H) Representative primer extension reactions for the 3' major domain. Roman numbers indicate sites were differences in SHAPE reactivity were observed between pre-40S particles. Oligonucleotides used for primer extensions are indicated in bold. (I) (performed by Ralph Hector and Sander Granneman) The 3' major domain of the 18S rRNA in empty and free 40S subunits is structurally more similar to 18S in translation initiation complexes. Middle pre-40S complexes, 80S translation initiation complexes and salt-washed, puromycin treated free 40S subunits were purified and incubated with 1M7 or solvent (DMSO, 5% final) for five minutes at room temperature. RNA was extracted and analysed by primer extension using the oligonucleotides listed on the left side of each panel. Shown are the results for the regions in the 3' major domain where we could consistently detect differences in SHAPE reactivities between early-middle and late pre-40S complexes (indicated by brackets and roman numbers on the left side of each panel). Pre-40S and 80S translation initiation complexes were purified using Rio2 and Fun12 as baits, respectively. The positions of the modified nucleotides are indicated on the right side of each panel.

In particular the 3' major domain (part of the head domain fig 1.2) was much more reactive to 1M7 in early and middle intermediates compared to the late, 80S like and mature ribosomes (fig 3.8E). In this region, nucleotides in the loops of H30, H33, H35, H37, H40 and H41 were generally more reactive to 1M7 in early and middle pre-40S complexes (fig 3.8F, red nucleotides). These observations were validated by primer extension (fig 3.8G and H). Nucleotides G1241-G1243 (box VI) showed an interesting modification pattern reminiscent

of a seesaw switch: G1241 was more flexible in early and middle pre-40S particles, but became constrained in late pre-40S, 80S-like and mature 80S particles. G1243 has the opposite pattern: more flexible in the late particles compared to the early and middle pre-40S complexes.

U1390 (fig 3.8F, box III, blue nucleotide) proved to be a false negative. It was not called modified by the TCP$^{EM}$ algorithm but the primer extension data showed that it was modified.

To determine if some of the observed rRNA structural rearrangements could be the result of subunit joining, Ralph Hector and Sander Granneman performed structure probing experiments on 18S isolated from free 40S subunits, which were generated by incubating extracts with high salt and puromycin. The 1M7 modification patterns from these 18S samples in the 3' major domain was strikingly similar to the 20S pre-rRNAs isolated from late pre-40S, 80S-like and 18S from mature 80S. This indicates that the majority of rRNA restructuring events described above were not the result of subunit joining (figure 3.8I).

I conclude that the 3' major domain in the early and middle pre-40S intermediates is more flexible and becomes more structured as 40S assembly progresses.

## 3.7. Restructuring of the 3' major domain coincides with the dissociation of Enp1, Ltv1, Tsr1 and Rio2 and takes place before the association of Rio1

Recent UV cross-linking experiments have shown that Enp1, Ltv1 and Nob1 bind to the 3' major domain, while Rio2 and Tsr1 cross-linked near the neck and the mRNA channel (fig 3.9B) (Granneman *et al*, 2010). Intriguingly, many of these cross-linking sites were located in close proximity to nucleotides that we showed to undergo structural rearrangements during 40S assembly (fig 3.9A and B). Because most of the assembly factors that bind the 3' major domain (Enp1, Ltv1 and Rio2) are released during late stages of 40S assembly (as discussed above), I speculated that the observed rRNA structural rearrangement could coincide with the release of these factors. In addition,

many of the restructured nucleotides were in the vicinity of r-protein binding sites. For example, in the crystal structure of the 3' major domain, nucleotides in the internal loop of H37 and the stem of H40 are bound by Rps17, nucleotides U1514 and U1516 located in the internal loop of H41 are bound by Rps20 and N-terminal region of Rps3 respectively. In the beak region, Rps15 binds nucleotides in the loop of H33 (fig 3.9D-F, fig 3.8 regions II, V, VI). All of these r-proteins are known to interact with pre-ribosomes during very early stages of assembly (Ferreira-Cerca *et al*, 2007), however, our structure probing data suggests that these r-proteins fully assemble into the head domain only during late stages of maturation.

To determine whether these head-domain r-proteins are indeed less stably bound pre-40S complexes, Sander Granneman performed immunoprecipitation experiments using FLAG-tagged Rps3, Rps17, Rps20, Rps10 and Rps9 as baits under low and high salt conditions (fig 3.9.G and H). Rps3 was used as a positive control as it was previously shown to dissociate from the pre-40S complexes in high salt conditions (Schäfer *et al*, 2006; Strunk *et al*, 2011). Rps9 binds to the body of the pre-40S. The body region folds before the head domain region and therefore was probably already completed in our particles (Ferreira-Cerca *et al*, 2007). Therefore, we assumed that Rps9 would bind stably to both pre-40S and 40S subunits. The data show (fig 3.9G) that high salt conditions dissociated both Rps3 and Rps17 from 20S pre-rRNA but not mature 18S rRNA. This indeed demonstrates that these proteins are less stably associated with pre-40S complexes compared to mature 40S subunit. In contrast, binding of the negative control Rps9 and two other head domain r-proteins (Rps10 and Rps20) to 20S or 18S was not noticeably affected by high salt conditions (fig 3.9G and H).

Collectively, the result suggest that the presence of Enp1, Ltv1, Rio2 and Tsr1 is required to maintain the rRNA flexible conformation of the 3' major domain, a conformation that is incompatible with stable Rps3 and Rps17 incorporation into the pre-40S complexes.

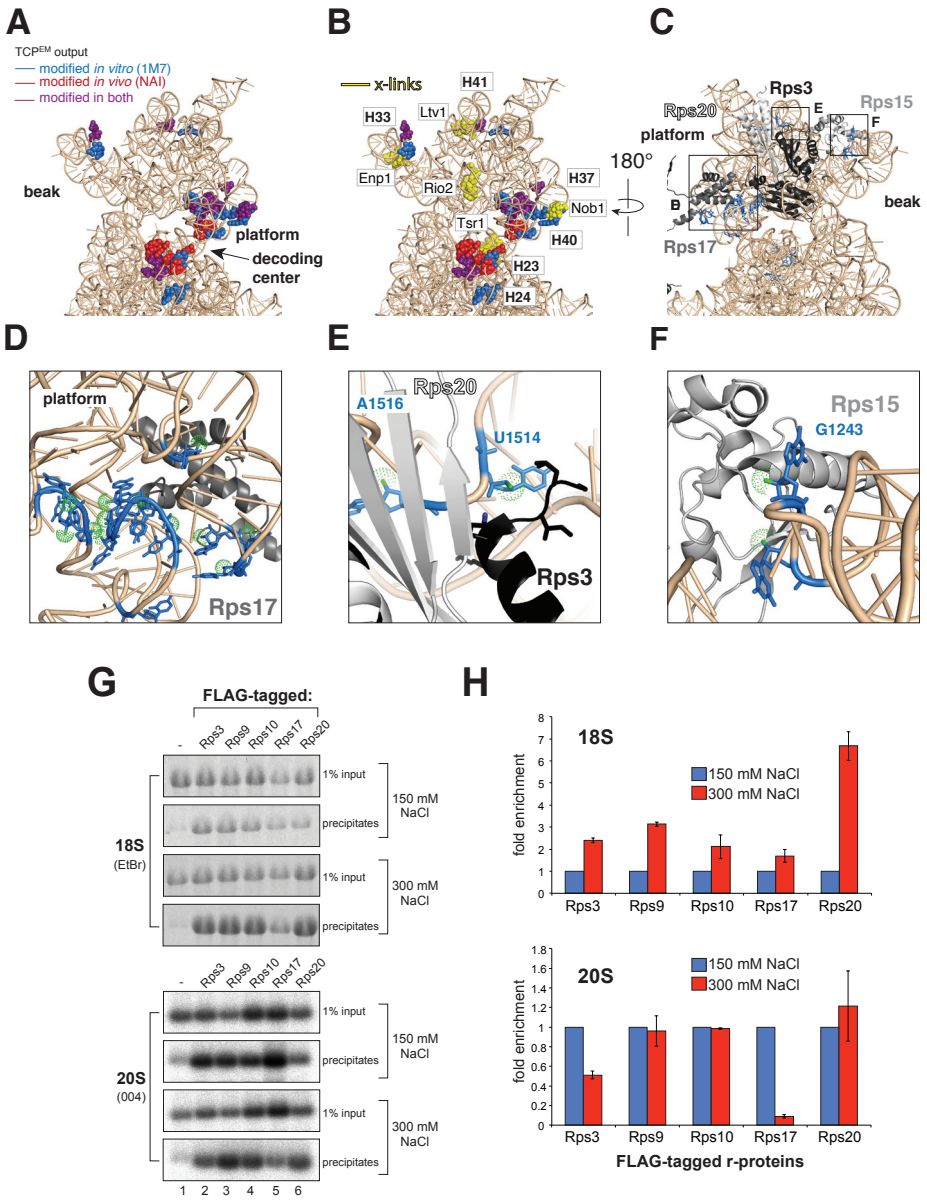



**Figure 3.9. Ribosomal protein rRNA binding sites are more flexible in early and middle pre-40S complexes.**

(A-B) Nucleotides in the head domain reactive to SHAPE chemicals (A) cluster near assembly factor UV cross-linking sites (B). (C-F) SHAPE-modified nucleotides (blue) in the head domain coincide with Rps3, Rps15, Rps17 and Rps20 binding sites. Relevant nucleotide positions are indicated. Green dots highlight the 2'-OH. (G-H) (performed by Sander Granneman) Ribosomal proteins Rps3 and Rps17 form salt-labile interactions with pre-40S complexes. Immunoprecipitations were performed using strains expressing FLAG-tagged r-proteins (indicated above the panels). Co-precipitated 20S was detected by Northern hybridization with oligo 004, 18S rRNA was detected by ethidium bromide staining (EtBr). Input indicates 1% of total RNA extracted from cell lysates. (H) Quantification of results shown in (G). Error bars indicate standard deviations obtained from two biological replicates.

### 3.8. The 3' major domain of *in vitro* refolded 18S rRNA adopts a similar structure as the 20S pre-rRNA in the late pre-40S intermediates

In order to further determine if the observed rRNA restructuring of the 3' major domain was due to protein binding or rRNA folding events, Maarten Tuijtel (Erasmus summer student) together with Ralph Hector probed the structure of 18S rRNA that was refolded in the same buffer used for chemical probing. The primer extension results revealed that H35(I), H37(II), H40 (III and IV), H41(V) and H33(VI) showed a very similar 1M7 modification pattern in 18S rRNA sample compared to the data of late pre-40S particles (fig 3.8A compare lane 7 and 8). These regions were generally less flexible compared to 20S pre-rRNA isolated from early and middle pre-40S intermediates (fig 3.10A and B, compare lane 7 with lanes 2-5). This enforces the notion that the presence of the assembly factors keeps the rRNA in a more flexible conformation, although we cannot exclude the possibility that some regions in the refolded rRNA adopt alternative or slightly different structures. Thus, in the absence of proteins, the 3' major domain appears to adopt a structure similar to that of the mature 18S rRNA. Many of the strongly 1M7 modified nucleotides found in the early and middle pre-40S intermediates did not show the same degree of flexibility in the *in vitro* refolded 18S rRNA (fig 3.10C, purple nucleotides). This suggests that under the conditions used these regions can fold into what appear to be relatively stable structures, independent of proteins *in vitro*. In the crystal structure of the 80S ribosome (Ben-Shem *et al*, 2011), the nucleotides in the loops of H37, H40 and H41 are stabilized through short and long range base stacking interactions. Many of these stacking interactions are s53 interactions (fig 3.10C green lines) which are generally less reactive to SHAPE chemicals (Bindewald *et al*, 2011). The *in vitro* refolded data suggests that at least some of these base stacking interactions could be formed in the absence of proteins.

All in all, our data suggest that assembly factors provide the energy necessary to maintain a more flexible conformation in the 3' major domain of the 20S pre-rRNA.

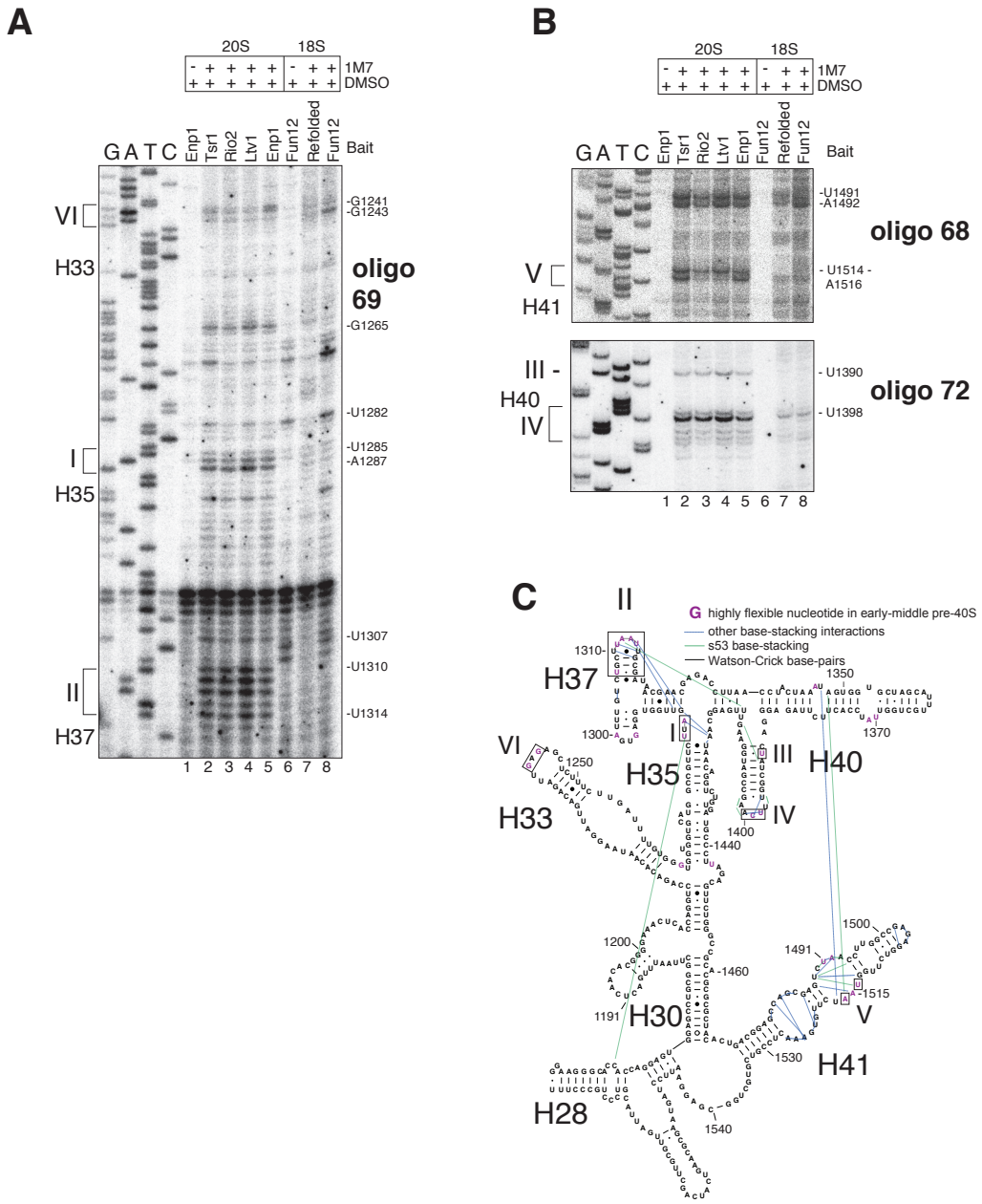



**Figure 3.10. Ribosome assembly factors are required to maintain an open head domain conformation.**

(A-B) (performed by Ralph Hector) Comparison of primer extension data obtained from 1M7 modified early (Ltv1, Enp1), middle (Tsr1, Rio2) pre-40S complexes (lanes 2-5), mature 40S subunits (Fun12) (lane 8) and in vitro refolded 18S rRNA (lane 7, Refolded). Unmodified 20S and 18S rRNAs were used as control samples (lanes 1 and 6). The positions of the modified nucleotides are indicated on the right side of each panel. (C) The TCPEM output generated from in vitro refolded 18S rRNA ChemModSeq data for the 3' major domain. Purple letters indicate nucleotides that were modified in the early-middle intermediates but not in the in vitro refolded 18S rRNA.

## 3.9. Discussion

Although the structure of the mature ribosome has been elucidated in detail (Ben-Shem *et al*, 2011), determining the detailed structure of native pre-ribosomal particles is a major challenge. The heterogeneity of purified particles and the dynamic nature of the assembly pathway make the analysis of native intermediates by crystallography or cryo-EM studies very challenging. Although with cryo-EM it is now possible to reach near atomic resolution (Passos & Lyumkis, 2015), structures of flexible regions in rRNA still generally remain unresolved. Therefore, as an alternative approach to direct structural studies, we decided to use a biochemical approach that involved RNA structure probing with high-throughput sequencing. The method that our group developed, dubbed ChemModSeq, enabled me to generate snapshots of changes in rRNA flexibility during the assembly of the 40S subunit. The results presented in this Chapter demonstrate that ChemModSeq is a very powerful method for rRNA structure determination and together with protein-RNA interaction studies (such as CRAC) could nicely complement cryo-EM studies. ChemModSeq is a fast and precise tool for obtaining nucleotide resolution insights into changes in flexibility of the RNAs even if they are part of RNP complexes.

This study significantly advanced our understanding of the rRNA structural rearrangements that take place during the maturation of the 40S subunit. We demonstrate, for the first time, that the 3' major domain undergoes a restructuring event before the formation of the 80S-like complex, a major quality control step in 40S maturation. I show that these rRNA structural rearrangements coincide with the release of the assembly factors Enp1, Ltv1, Rio2 and Tsr1. Consequently, it seems that the binding of the assembly factors maintains the rRNA in a more flexible conformation and prevent the completion of the 3' major domain, which in turn blocks the stable incorporation of r-proteins Rps3 and Rps17 (fig 3.11). It could be that their presence is needed for the formation of functional sites required for translation (Thomson *et al*, 2013) or/and that their presence blocks the premature formation of translational competent complexes (Strunk *et al*, 2011).

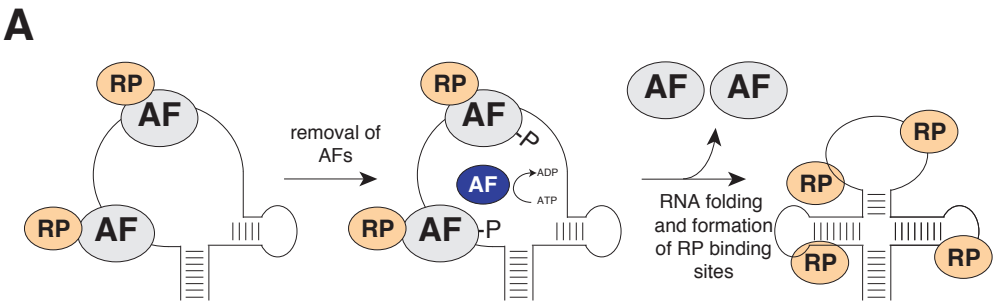



**Figure 3.11. A model for structural rearrangements during late stages of 40S subunit maturation.**
(A) Proposed model for the function of ribosome assembly factors in 40S assembly. Binding of assembly factors (AF) to RNA is required to maintain a more flexible conformation. Their presence allows the assembly of ribosomal proteins (RP) but prevent that these RPs adopt their final conformation. At some point during the maturation pathway, the system receives a signal that certain ribosome assembly / RNA folding steps have been completed and AFs are no longer needed. Their release could be triggered through phosphorylation (such as Hrr25-dependent phosphorylation of Enp1 and Ltv1). This allows completion of RNA folding steps and binding of late RPs.

## Restructuring of the 3' major domain in the late pre-40S complexes

The 3' major domain (in the head region) undergoes the most dramatic changes in flexibility as the pre-40S maturation progresses. Enp1 and Ltv1 directly bind to the head domain and I predict that Enp1 and Ltv1 act as RNA chaperones by delaying the folding of the 3' major domain until earlier steps have been completed. The binding of Ltv1 and Enp1 to the pre-40S blocks the formation of the beak structure. Upon export into the cytoplasm Ltv1 and Enp1 dissociate in a Hrr25 phosphorylation dependent manner and the beak, represented by a protrusion of H33, is formed (Schäfer *et al*, 2006).

Very little is known about the roles of Tsr1 and Rio2 in rRNA folding or 40S maturation. Tsr1 depletion was shown to block Ltv1 dissociation (Strunk *et al*, 2011), which suggests that Tsr1 also plays a role in the formation of the 3' major domain, even though it does not directly bind to this region (Granneman *et al*, 2010). Failure to release Ltv1 blocks the formation of the 80S-like complex (Ghalei *et al*, 2015), suggesting that the flexible structure of the 3' domain in early and middle pre-40S particles is incompatible with 60S

subunit joining. It was proposed that Rio2 positioning between the head and the body allowed it to “sense” the changes in the conformation of the head domain (Ferreira-Cerca *et al*, 2012). It is conceivable that dissociation of Rio2 somehow influences the rRNA structure of the head domain or that completion of the head domain triggers its release, making the mRNA channel available for translation initiation.

The ChemModSeq data revealed that in the 3’ domain, nucleotides that correspond to binding sites for r-proteins adopt a more constrained conformation in the late pre-40S compared to early and middle particles. Among these proteins are Rps3, Rps17 and Rps15. The data indicate that these r-proteins do not make all their contacts in the head domain during early stages of maturation. In accordance, Rps3 and Rps17 incorporation into the pre-40S complexes is salt labile, while their incorporation into mature 40S complexes is not. Interestingly, Rps10 assembles in the cytoplasm and its binding site overlaps with Enp1 binding sites. It is plausible that Rps10 can only fully assemble once Enp1 is released.

Formation of the head domain starts in the nucleus but is finalized in the cytoplasm. Premature completion of the head domain was proposed to block pre-40S nuclear export, which might explain why it is important to keep the head domain in an alternative state before nuclear export (Ferreira-Cerca *et al*, 2007; Schäfer *et al*, 2006). Consistent with this idea, many r-proteins initially only weakly interact with pre-40S complexes but their binding gradually becomes more stable during later steps of 40S assembly. Rps3 and Rps15 are both important for efficient nuclear export of pre-40S particles. Consequently, it could be that Rps3 and Rps15 low affinity binding is initially needed for pre-40S export, while the high affinity binding of Rps3, Rps15 and Rps17 is subsequently needed for the completion of the structure of the head domain. It would be interesting to see if blocking their stabilization affects subunit joining. As none of the three r-proteins binds to the subunit interaction side of the pre-40S, if they indeed influence subunit joining I would expect their effect to be indirect possibly by influencing the folding of the rRNA.

The data suggest that assembly factors are not only required to delay the complete assembly of several r-proteins, but also to keep the rRNA in a flexible conformation. I speculate that the assembly state of the head domain

is somehow sensed by the nuclear export machinery, but detailed mechanistic insights are still lacking.

Late pre-40S complexes contain only Rio1, Nob1 and Dim2

This study shows that Rio1 only interacts with pre-40S complexes after most of the assembly factors have dissociated and the rRNA restructuring events in the 3' major domain have been completed. Besides Rio1, these late pre-40S complexes also contain Nob1 (the endonuclease that cleaves at site D) and Dim2/Pno1. These results were recently confirmed by the Tollervey group (Turowski & Tollervey, 2014). This late pre-40S complex is most likely the intermediate that joins with 60S subunits to form the 80S-like complex. The nucleotide flexibility pattern of the rRNA in the late pre-40S is almost identical to the 20S pre-rRNA found in the 80S-like complexes and both pre-40S particles have a similar protein composition. The current model is that Nob1 cleaves shortly after 80S-like formation. Using an *in vitro* processing system, the Tollervey lab demonstrated that the enzymatic activity of Nob1 is enhanced by both Dim2 and the ATPase activity of Rio1 (Turowski & Tollervey, 2014; Vanrobays *et al*, 2003). It remains unclear, however exactly how Rio1 enhances cleavage at site D. It is possible that Nob1 exonuclease activity is enhanced by phosphorylation, which would explain why the Rio1 kinase is present. We often detected multiple Nob1 species in our western blots, suggesting that the protein is indeed post-translationally modified (Hector *et al*, 2014). Mass-spectrometry analyses of Nob1 should shed some light on this. The role of Dim2 is also not clear but it does bind the pre-40S close to the 3' end of 18S (Woolls *et al*, 2011). It is possible that Dim2 somehow stabilizes the interaction between Nob1 to the 3' end of 18S, which could be important for efficient cleavage at site D.

# CHAPTER 4

# The Role of Rps17 in 40S Subunit Maturation

## 4.1. Introduction and aim

The ChemModSeq data indicated that some r-proteins bind early pre-40S ribosome intermediates but are prevented from adopting their final position until the late stages of pre-40S assembly (fig 4.1). The r-proteins Rps3, Rps15, Rps17 bind to the head domain of the 40S subunit during early stages of assembly but are repositioned upon completion of rRNA folding events in the 3' major domain and dissociation of assembly factors. We and others have demonstrated that Rps3 and in particular Rps17 association to pre-40S are salt labile (fig 3.9A).

Rps17 is one of the r-proteins that was not previously characterized (de la Cruz *et al*, 2015). The gene encoding this protein is often disrupted in patients with Diamond-Blackfan anaemia (DBA), suggesting that it plays an important role in ribosome function. Diamond-Blackfan anaemia is characterized by a defect in red blood cell maturation, congenital defects and a predisposition to cancer (Danilova & Gazda, 2015). At the genetic level, DBA is triggered by haploinsufficiency of specific r-proteins. This haploinsufficiency leads to a decrease in the levels of mature ribosomal subunits and RNA processing or export defects. Mutations in r-proteins that cause DBA are generally frameshift mutations that target the mRNA for degradation in the cytoplasm by non-sense mediated decay (NMD) (Gazda *et al*, 2004). Rps19 is the most frequently affected r-protein in DBA. Frameshift mutations of Rps19 account for 25% of DBA cases (Gazda *et al*, 2004). Intriguingly, unlike Rps19, the Rps17 DBA mutations were localized in the start codon. As a result, the next start codon is recognized resulting in a short 4 amino acid peptide that replaces the full length Rps17 protein (Cmejla *et al*, 2007; Song *et al*, 2010). Although the levels of functional ribosomes are affected in all cell types, DBA mainly affects cell types that rely on high ribosomal synthesis levels, such as developing red cells, explaining why these patients are severely anaemic (Robledo *et al*,

2008).

Despite its importance in ribosome assembly and its role in DBA, Rps17 has not been functionally characterized. Intrigued by the ChemModSeq results, as well as the links with DBA, I proceeded to further investigate the function of Rps17 in ribosome biogenesis.

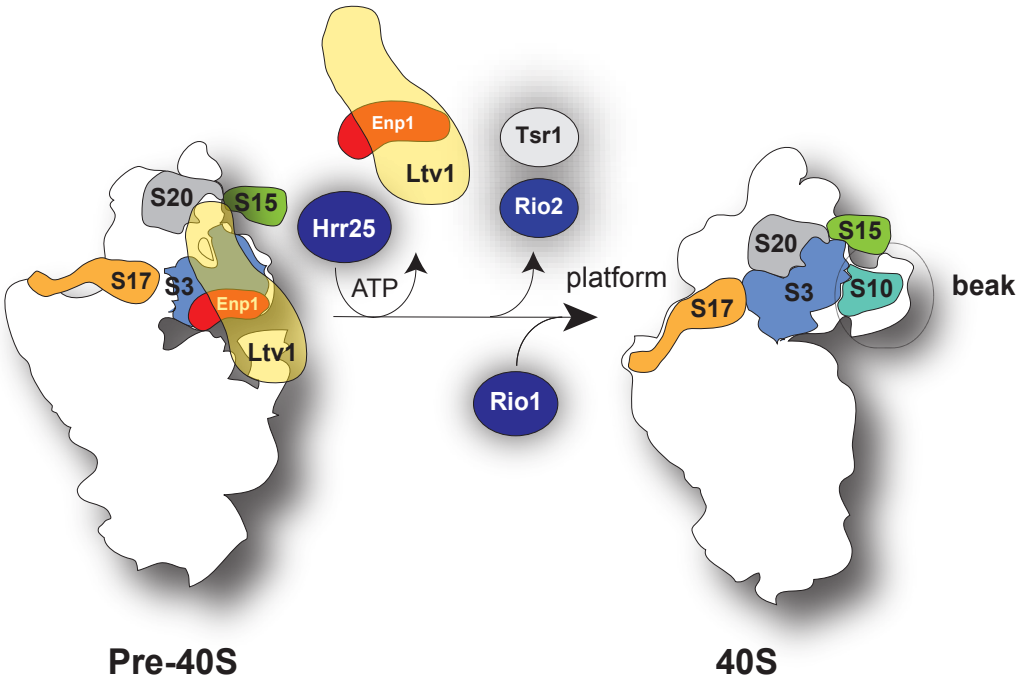



**Figure 4.1. Rps17 binds to early pre-40S but it is repositioned in the late pre-40S assembly complexes**

Ribosomal proteins Rps3, Rps17, Rps15 bind to the head of the pre-40S but are allowed to adopt their final conformation only at a later stage, subsequent to the completion of certain RNA folding steps and the release of assembly factors. In early and middle pre-40S complexes, binding of Ltv1 and Enp1 to the head domain is required to maintain a more flexible conformation in the 3' major domain. Hrr25 phosphorylation of Ltv1 and Enp1 triggers their release, allowing late RNA folding steps and assembly of head domain r-proteins to be completed. We propose that Rio2 and Tsr1 are released shortly afterwards after which Rio1 enters the assembly pathway.

## 4.2. The C-terminal region of Rps17 is essential for function and cell viability

My first aim was to determine which amino acid domains of Rps17 are important for its function. There is a high degree of conservation between the Rps17 sequence of an Archaea (*T. acidophillum*) and Rps17A from *S. cerevisiae*. Remarkably, the *T. acidophillum* Rps17 is much shorter and lacks the C-terminal domain found in its yeast counterpart (represented in blue in fig 4.2A). The crystal structure of Rps17 bound to the yeast 80S mature ribosome (Ben-Shem *et al*, 2011) revealed that the Rps17 N-terminal domain (NTD) contacts the terminal loop of H37 of the 18S rRNA and is more structured than the C-terminal domain (CTD) (fig 4.2B).

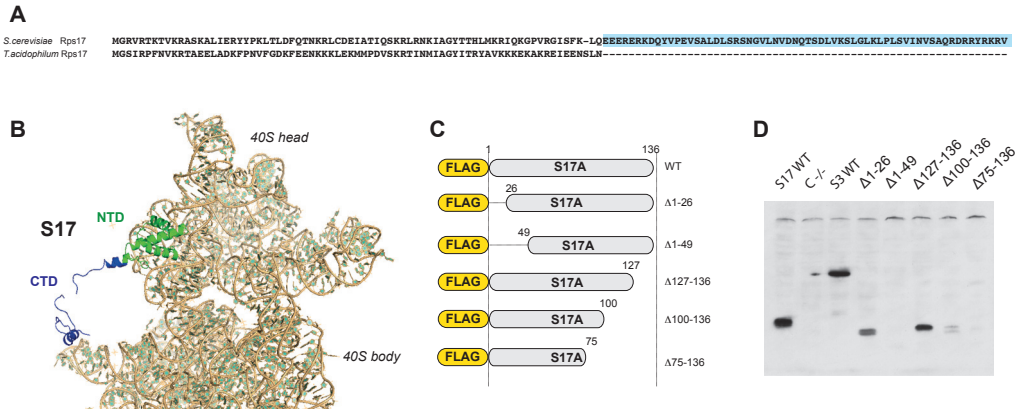



**Figure 4.2. Design of Rps17 N-terminal and C-terminal deletion mutants**
(A) Alignment of Rps17 from *S.cerevisiae* and *T. acidophilum*. The eukaryotic specific Rps17 C-terminal sequence (amino acids 75-136) is highlighted in blue. (B) Crystal structure of Rps17 bound to the mature 80S ribosome. Highlighted in blue is the eukaryotic specific Rps17 sequence. (C) Design of Rps17 N-terminal (NTD) and C-terminal (CTD) deletion mutants. (D) Western blot determining the expression of the FLAG-tagged Rps17 WT and mutants. WT-Rps17 wild type, -/- - negative control *GAL::3HA-rps17AΔrps17B* strain, S3 - *GAL::3HA-rps17AΔrps17B* with FLAG-Rps3.

Based on these observations, I decided to make deletion mutants in an attempt to determine the role of these domains. For this purpose, I used a strain in which the *RPS17B* allele was deleted and *RPS17A* allele was placed under the control of a galactose inducible promoter. This enabled me to deplete the endogenous Rps17A protein by shifting the cells from galactose to glucose containing medium and assess the effect of Rps17 mutations on cell growth and

ribosome synthesis. The plasmids containing FLAG-tagged Rps17 and FLAG-tagged Rps17Δ75-136 were kindly provided by Philipp Milkereit (University of Regensburg, Germany). The Rps17Δ75-136 deletion excluded the eukaryotic specific C-terminal sequence (blue in fig 4.2A and 4.2B). However, as this truncation removed almost 50% of the protein, I decided to also generate two shorter C-terminal deletions mutants: Rps17Δ127-136 and Rps17Δ100-136. In addition, I also made two N-terminal deletion mutants: Rps17Δ1-26 and Rps17Δ1-49 (fig 4.2C). The N-terminal deletions included amino acids that, according to the crystal structure (Ben-Shem *et al*, 2011), are involved in the protein-rRNA interactions. I predicted that these deletions would prevent the (stable) incorporation of Rps17 into pre-ribosomes.

I first analysed whether the Rps17 mutants were stably expressed by western blotting. The results in Figure 4.2D show that, except for the mutants with long truncations (Rps17Δ1-49 (lane 5) and Rps17Δ75-136 (lane 8)), all mutants were detectably expressed.

### 4.3. Rps17 C-terminal Δ100-136 mutant is dominant negative

The next step was to assess the effect of the Rps17 truncations on cell growth. This was performed using a growth complementation assay in the strain where the genomically encoded Rps17A could be conditionally depleted. Because the expression of the endogenous Rps17A was galactose dependent, by spotting the cells on glucose plates I could assess the effect of the truncations on cells growth (fig. 4.3A and B).

As expected, the wild-type Rps17 expressed from the plasmid was able to support growth both on galactose and glucose containing plates. The results furthermore showed that only Rps1Δ127-136 mutant was able to support growth comparable to the wild-type protein (fig 4.3B). This suggests that the last 10 amino acids are not essential for Rps17 function. The inability of the Rps17Δ1-49 and Rps17Δ75-136 to support growth was expected, as these mutants were not stably expressed (fig 4.2C).

The Rps17 NTDΔ1-26 protein did not support growth, suggesting that the first 26 aminoacids are critical for Rps17 function. Interestingly the

Rps17Δ100-136 mutant affected growth on both galactose and glucose plates, suggesting that this mutant is dominant-negative.

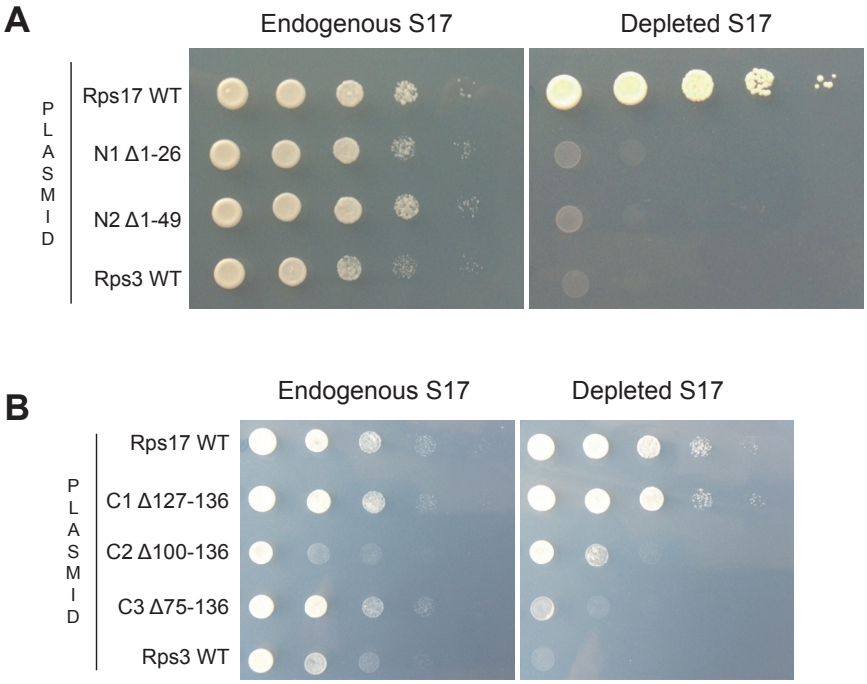



**Figure 4.3. The C-terminal Rps17 Δ100-136 deletion mutant is a dominant negative.**
Serial dilutions of yeast cells expressing N-terminal (A) and C-terminal (B) Rps17 deletion mutants on galactose (to express endogenous Rps17) and glucose plates (to deplete endogenous Rps17). WT-Rps17 wild type, Rps3 - *GAL::3HA-rps17AΔrps17B* with FLAG-Rps3 used as negative control.

## 4.4. Rps17 N-terminal domain is essential for ribosome incorporation

I next determined which regions are important for Rps17 binding to the (pre-)40S subunit. If the mutants were stably incorporated into pre-40S and 40S subunits, I would expect to be able to co-immunoprecipite both 20S pre-rRNA and mature 18S rRNA with the FLAG-tagged proteins. If the deletion mutants would fail to co-precipitate these rRNA species or would co-precipitate them less efficiently, then that would indicate that the truncations affected Rps17 incorporation into pre-ribosomes. Figure 4.4A shows that the Rps17Δ1-26 deletion mutant failed to co-precipitate any of the analysed rRNA species,

indicating that the first 26 amino acids in the N-terminal domain are essential for Rps17 incorporation into pre-40S complexes. This was not unexpected, since the crystal structure showed that this region directly contacts the 18S rRNA (Ben-Shem *et al*, 2011). On the other hand, the Rps17Δ127-136 and Rps17Δ100-136 truncation mutants co-precipitated both 20S pre-rRNA and 18S rRNA (fig 4.4B). This indicates that although the Rps17 C-terminal truncations disrupted cell growth, these regions were not required for Rps17 incorporation into pre-40S complexes.

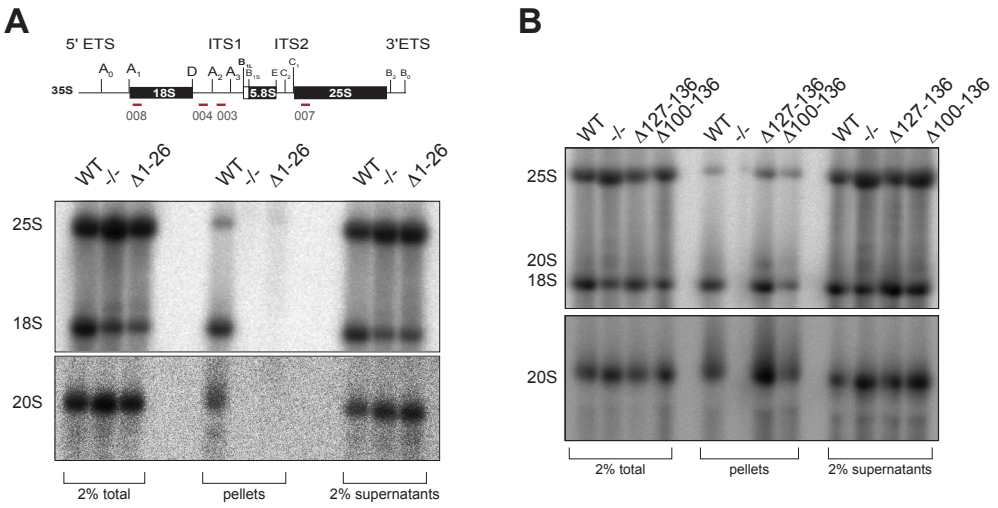



**Figure 4.4. The first 26 aminoacids in the N-terminal domain are essential for Rps17 incorporation into pre-ribosomes.**
Anti-FLAG immunoprecipitations on strains expressing FLAG-tagged Rps17 WT, NTD (A) and CTD (B) mutants and detection of the co-precipitated rRNAs by Northern Blot hybridization (20S pre-rRNA (oligo 004), 25S rRNA (oligo 007), and 18S rRNA (oligo 008)). WT-Rps17 wild type, -/- - negative control *GAL::3HA-rps17AΔrps17B* strain

## 4.5. Is Rps17 misplaced in the pre-40S to accommodate Nob1 binding?

Previous work had shown that Nob1 cross-links to the pre-rRNA in close proximity to the Rps17 NTD-rRNA binding sites (Granneman et al 2010) (fig. 4.5A). I hypothesized that Nob1 might interfere with the stable binding of Rps17 to pre-ribosomes during early stages of assembly through sterical hindrance. If this hypothesis were correct, then I would expect that in the absence of Nob1 Rps17 would be able to already contact these sites during

the early stages of assembly. We have previously shown that the Rps17 H37 binding sites in the 20S pre-rRNA are flexible in early and middle pre-40S complexes but become less flexible in late pre-40S and mature 40S subunits (Hector *et al*, 2014). If the absence of Nob1 allowed Rps17 to bind the H37 in 20S pre-rRNA during early pre-40S assembly steps, I would expect those nucleotides to become less accessible to 1M7 in these particles. To test this, I compared the 1M7 reactivity pattern from complete middle (Rio2) particles and middle particles that lacked Nob1 (fig 4.5B). At first glance, Rps17 binding sites in H37 (bottom of the gel) did indeed appear to be less flexible in the absence of Nob1, suggesting that the presence of Nob1 blocks early binding of Rps17 to this region. However, the level of acp-modification at U1191 in 20S pre-rRNA isolated from particles lacking Nob1 was much higher compared to the control sample (fig 4.5B, compare lane 1 with lane 2). This indicates that, structurally, the Rio2 particles that lack Nob1 more closely resemble late (rather than middle) pre-40S complexes. Furthermore, the region where Rps10 binds in the crystal structure (fig 4.5B, U1285-A1287) also became less modified, indicating Rps10 was also incorporated into these particles. This suggests that although the depletion of Nob1 leads to a complete block in D-site cleavage, parallel processing steps such as restructuring of the head domain can still continue. I conclude that the differences in accessibility we observed in the H37 loop is not a consequence of Nob1 binding at H40 but to the purification of a later pre-40S complex that is structurally more similar to mature 40S subunits.

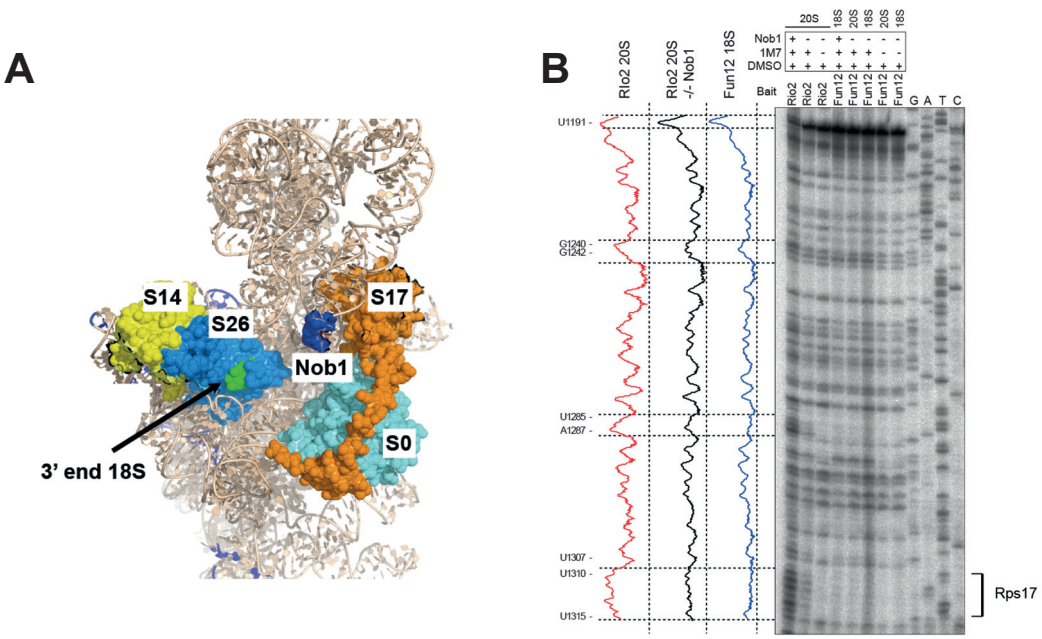



**Figure 4.5. Nob1 depletion prevents Rio2 dissociation without blocking the maturation pathway.**

(A) Nob1 crosslinking site (blue) are close to Rps17 N-terminal domain rRNA binding site (orange). (B) Primer extension (oligo 69) of 1M7-modified rRNA from Rio2-TAP (mid) or Fun12-TAP (mature) ribosomal complexes in the presence or absence of Nob1. Lane 1 represents the positive control: modified 20S pre-rRNA from Rio2-TAP particle in the presence of Nob1, lane 2 represents the modified 20S pre-rRNA from Rio2-TAP particle upon 8h Nob1 depletion, lane 4 represents the modified 18S rRNA from Fun12-TAP particles in the presence of Nob1, lane 5 and 6 represent the modified 20S pre-rRNA and 18S rRNA from Fun12-TAP particles upon 8h Nob1 depletion. Lanes 3, 7, 8 represent the negative solvent (DMSO) controls.

## 4.6. The 18S rRNA maturation is affected in Rps17 depletion and Rps17Δ100-136 mutants

To study the effect of Rps17 mutants on rRNA processing I measured the effect of the mutations on pre-rRNA processing. I transformed the plasmids containing the various Rps17 deletion mutants to the *GAL::rps17AΔrps17B* strain. For controls I used the same strain carrying an empty plasmid (negative control) or the wild-type Rps17 (positive control). All strains were grown in galactose to exponential phase (0 time-point) and then shifted to glucose for 4 and 8 hours (fig 4.6B). In glucose medium the cells completely rely on the

Rps17 expressed from the plasmids. After 4h and 8h of depletion, the rRNA processing defects in cells expressing the Rps17Δ1-26 mutant was almost identical to cells completely lacking Rps17 (fig 4.6B, compare lanes 7 and 8, 12 and 13). This was expected, as the Rps17Δ1-26 mutant was not detectably incorporated into pre-40S complexes (fig 4.4A). Similarly, the dominant negative Rps17Δ100-136 mutant showed decreased levels of 18S rRNA (fig 4.6B, lanes 10 and 15). I also noticed a mild accumulation of 20S and 35S pre-rRNA at the 8h time point (fig 4.6B, lane 15). This suggests that both early and late processing steps might be affected/delayed in the absence of Rps17.

The mild increase in 20S pre-rRNA levels suggest that side D cleavage is affected. However, if only site D cleavage was affected I would have expected to see a much higher level of 20S pre-rRNA accumulation, since the 18S rRNA levels were dramatically reduced in these mutants. I considered the possibility that the 20S pre-rRNA was not exported to the cytoplasm. Many RNAs that are (for whatever reason) retained in the nucleus are often rapidly degraded by the nuclear TRAMP-exosome complex (Dez *et al*, 2006). Therefore, I next determined if deleting TRAMP (Trf4 and Trf5) and exosome components (Rrp6) would stabilize the 20S pre-rRNA in cells lacking Rps17. For this purpose, I individually deleted Trf4, Trf5 and Rrp6 in the *GAL::rps17AΔrps17B* strain. Total RNA was extracted before the shift to glucose and after 4 and 8 hours after the shift (fig 4.6C). If the nuclear 20S pre-rRNA had been rapidly degraded by the TRAMP-exosome complex then I would have expected that knocking out these factors would lead to an increase in 20S pre-rRNA steady state levels. However, the quantification of the northern blot results illustrated in figure 4.6D showed that this was not the case.

Thus, the relatively mild accumulation of 20S pre-rRNA in Rps17Δ100-136 mutant is not because most of the accumulated 20S is rapidly degraded by the nuclear surveillance machinery.

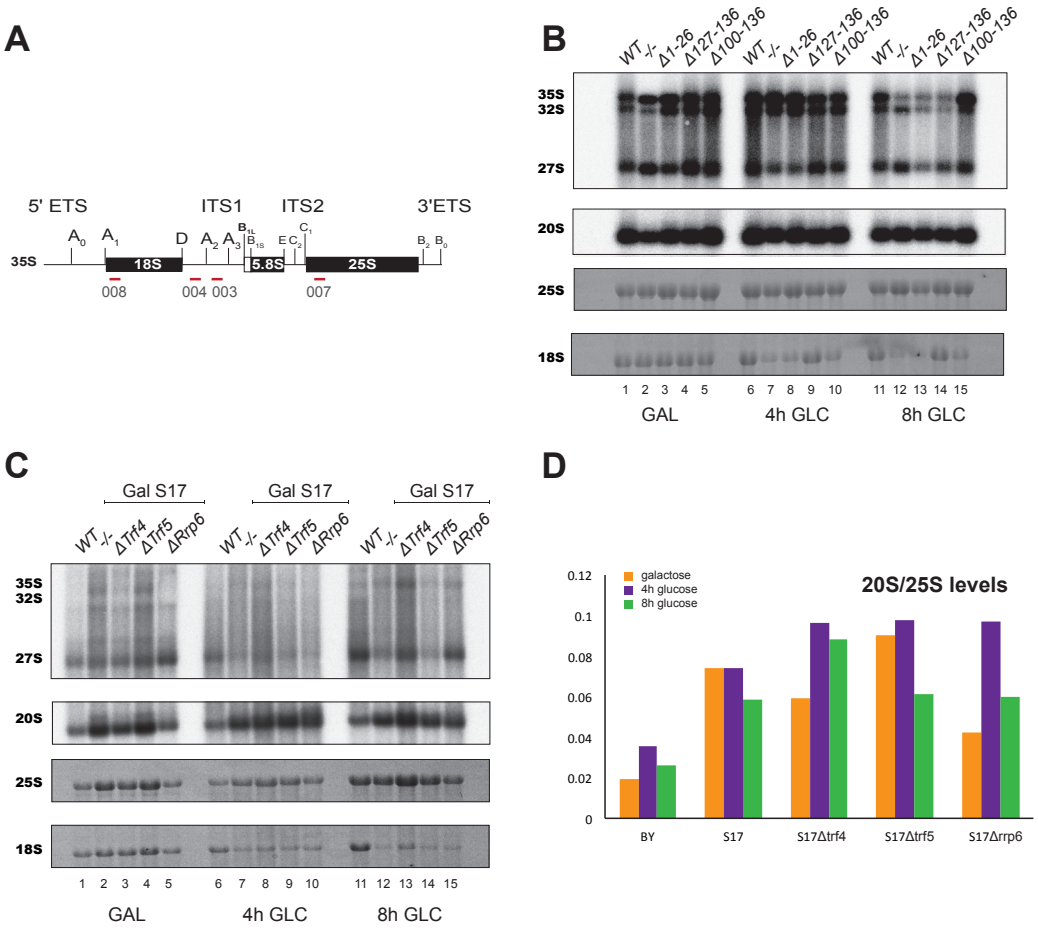



**Figure 4.6. Rps17 is required for efficient 18S rRNA processing**

(A) Representation of the oligonucleotide hybridization sites used for Northern blot analyses. (B) The effect of Rps17 mutations on pre-rRNA processing. Northern Blot analyses of pre-rRNA precursors and ethidium bromide (EtBr) staining of mature rRNAs extracted from strains expressing Rps17 WT and mutants in the presence (GAL) or absence (4h or 8h in glucose; GLC)) of endogenous Rps17. (C) The 20S that accumulates in Rps17 depleted strains is not rapidly degraded by TRAMP/Exosome components. Shown are the Northern blot results for Rps17 depletion strains in which individual TRAMP components (Trf4, Trf5) or the exosome component (Rrp6) were deleted. (D) Quantification of 20S levels from the Northern blot shown in (C)

## 4.7. Rps17 depletion delays the early steps 18S rRNA processing

In order to more precisely assess the effect of Rps17 depletion on rRNA processing, I performed a 4-thio-uracil pulse chase experiment using a protocol developed by David Barrass from Jean Beggs's lab. Rps17 was depleted for 8h in glucose medium without uracil. The cells were subsequently pulsed with 4-thio-uracil for 2 minutes, followed by a chase with an excess of uridine (see Chapter 2 Materials and Methods) for 5, 30 and 60 minutes. The thiolated transcripts were subsequently biotinylated and isolated using streptavidin beads. The nascent rRNA transcripts were then precipitated and resolved on agarose gels. The pre-rRNA intermediates were subsequently detected by northern blotting (fig 4.7). The mature rRNA and pre-rRNA intermediates were visualized using the oligonucleotide probes shown in Figure 4.7A. Compared to the northern blot analyses shown in Figure 4.6, this gave a much clearer picture of the pre-rRNA processing defects in Rps17 depleted cells. The results show that cells lacking Rps17 generally processed pre-rRNA much slower. Compared to the parental strain, Rps17 depleted cells showed a marked accumulation of both 35S and 32S pre-rRNA intermediates, suggesting processing defects at $A_0$, $A_1$ and $A_2$. Moreover, I also observed a strong accumulation of 27S pre-rRNA and a reduction in the levels of the mature 25S rRNA. This suggests that large subunit synthesis is also affected, however, not to the same extent as 18S rRNA synthesis. It is likely that the 27S pre-rRNA species that accumulate in Rps17-depleted cells represent $27SA_3$ pre-rRNAs as $A_2$ cleavage in these cells is kinetically delayed. The slow processing of these 27SA species suggests that $A_3$ processing is also slower than usual. The delay in $A_2$ processing also means that less 20S pre-rRNA is being made compared to the parental strain. This explains the lack of a massive 20S pre-rRNA accumulation upon depletion of Rps17 (fig 4.6) . Although 20S levels were lower in the Rps17 depleted cells, some 18S rRNA was still made (fig 4.7).

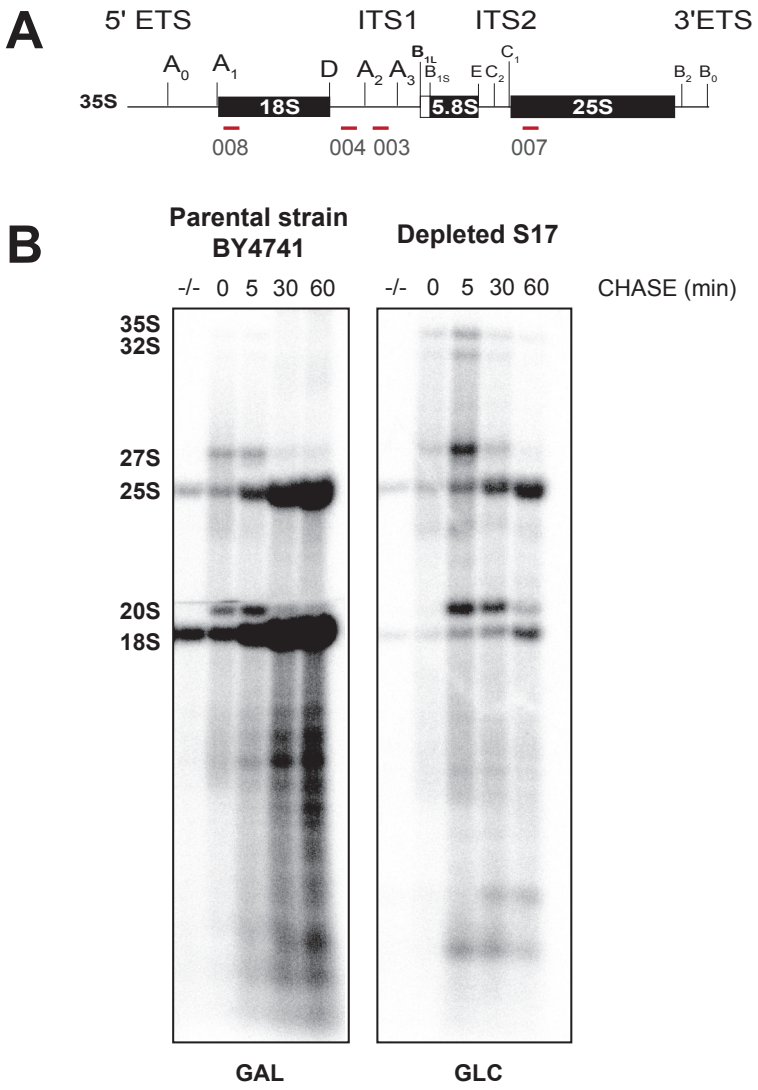



**Figure 4.7. The maturation of 18S rRNA is kinetically delayed in Rps17 depleted cells.** (A) Representation of the oligonucleotide hybridization sites used for Northern blot analyses. (B) Analyses of pre-rRNA processing kinetics using 4-thio-uracil pulse chase. Cells from the parental strain and the depleted Rps17 strain were grown in media without uracil (URA) to exponential phase, pulsed with 4-thio-uracil for 2 minutes and then chased with excess of uridine for 5, 30, 60 minutes. The thiolated, newly synthesized RNAs were extracted and (pre-)rRNAs were detected by Northern blotting using oligos 004, 003, 007, 008. The 4-thio-uracil pulse-chase protocol was developed by David Barrass in Jean Begg's lab.

To determine if 20S pre-rRNA in Rps17 depleted cells accumulates in the nucleus or the cytoplasm, I used fluorescent *in situ* hybridization (FISH) to determine the localization of 20S pre-rRNA, using an oligonucleotide complementary to ITS1 (fig 4.8). If cleavage at site D were affected then I would

expect to see 20S pre-rRNA accumulation in the cytoplasm (Granneman *et al*, 2005). However, the staining pattern in the parental and Rps17-depleted strain was very similar. This suggests that the 20S that is produced in the Rps17 depleted strain does not significantly accumulate in the cytoplasm and that Rps17 is probably not involved in cleavage at site D.

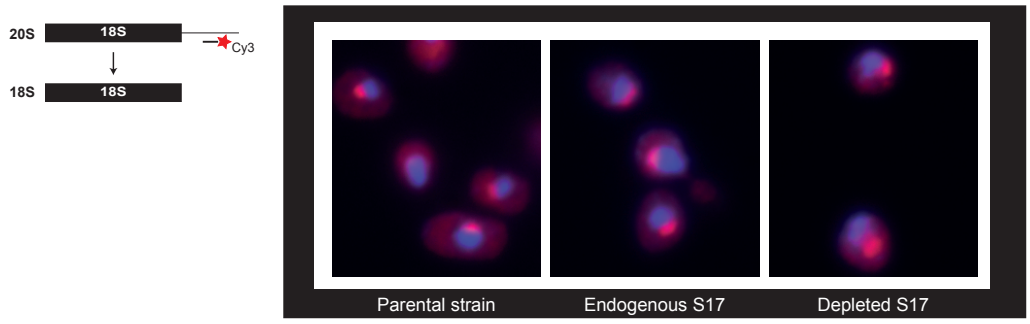



**Figure 4.8. Rps17 is not required for cleavage at site D in the cytoplasm.**
Fluorescent *in situ* hybridization (FISH) was performed using a Cy3 labelled oligonucleotide that hybridizes between D and A2 cleavage sites in ITS1. The FISH analyses were performed on cells expressing Rps17 (BY4741 and *GAL::3HA-rps17AΔrps17B* grown in galactose) and on Rps17 depleted cells (*GAL::3HA-rps17AΔrps17B* strain grown in glucose for 8h).

The low levels of 18S rRNA and lack of 20S pre-rRNA accumulation are highly reminiscent of the processing defects observed in cells depleted of Rps5 and Rps15. Cells lacking Rps5 or Rps15 accumulate 20S pre-rRNA in the nucleus, suggesting that these cells are defective in nuclear export of pre-40S complexes (Ferreira-Cerca 2005; Leger-Silvestre 2004). It could thus be that Rps17 is involved in pre-40S export.

## 4.8. Rps17 C-terminal domain has a putative nuclear export signal (NES)

Interestingly, closer inspection of the Rps17 amino acid sequence revealed that part of the CTD contains an NES consensus sequence: DLVKSLGLKLPLS (fig 4.9A). Moreover, this NES was deleted in the Rps17Δ100-136 mutant (fig 4.3A and B). It was enticing to speculate that this NES plays a role as export adapter in Crm1-dependent export of pre-40S, although it would not be an essential adapter as deletion of the NES does not completely block 18S rRNA synthesis. To my knowledge, no ribosomal

protein was previously described to be involved in Crm1 dependent export of ribosomal subunits. This therefore warranted further investigation.

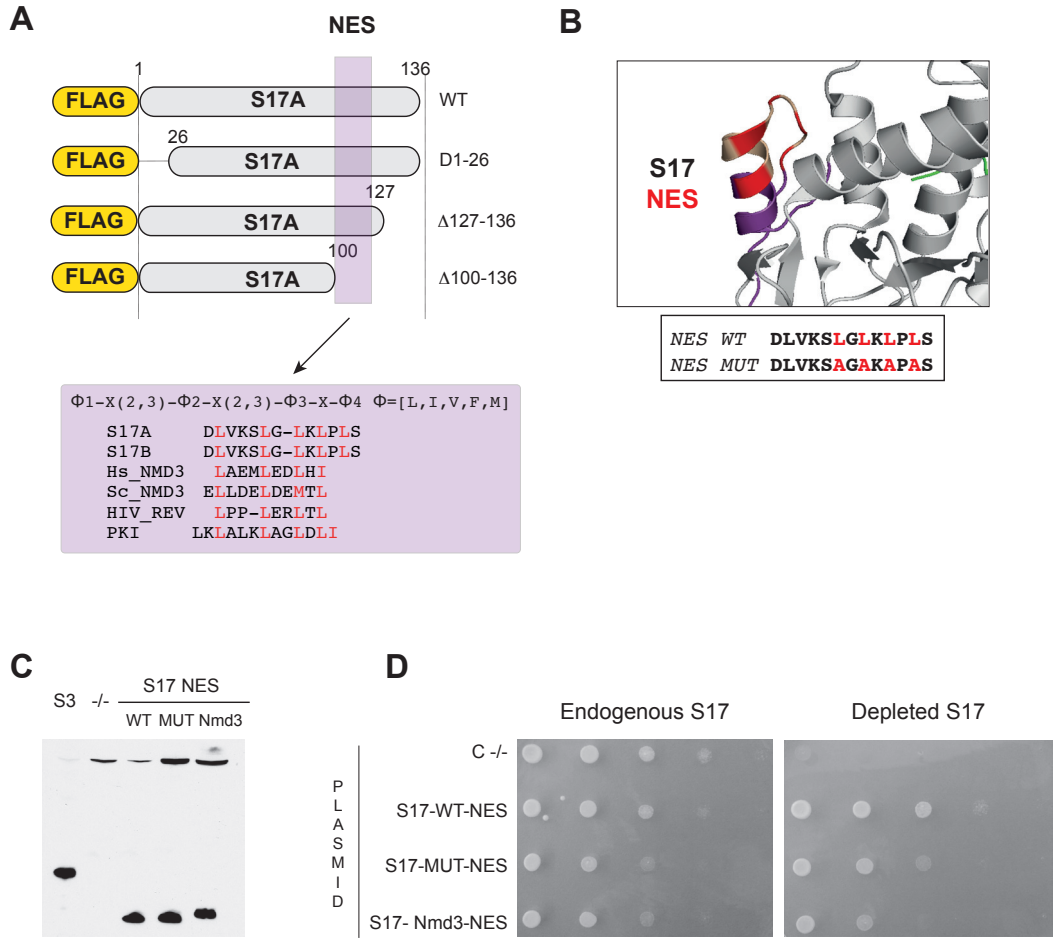



**Figure 4.9. Rps17 has a putative NES in the C-terminal domain**
(A) Localization of the Rps17 NES and comparison of Rps17 NES to the consensus sequence and to other well studied NES sequences: PKI, Nmd3 and HIV-REV NES, Φ – hydrophobic amino acid. (B) Secondary structure of the Rps17 NES. The amino acids represented in red are the essential hydrophobic Leucine (L) residues. To generate the Rps17 NES mutant 4 of the 5 leucine residues were mutated to Alanines (A). (C) Western blot showing the stable expression of the S17 constructs: S17 WT NES – wild type S17, S17 MUT NES – S17 with the 4 NES leucine residues to alanines substitutions, S17 Nmd3 NES – S17 NES replaced with Nmd3 NES, -/- - negative control *GAL::3HA-rps17AΔrps17B* strain, S3-*GAL::3HA-rps17AΔrps17B* expressing FLAG-Rps3. (D) The Rps17 NES mutant is dominant negative. Serial dilutions of cells expressing the wild-type Rps17 or Rps17 NES mutants on galactose (to express endogenous S17) and glucose (to deplete endogenous S17). Wild-type Rrp17 (S17-WT-NES), the 4 Leu residues to Ala Rrp17 NES mutant (S17-MUT-NES) and Rps17 containing the NES from Nmd3). C-/- is the negative control *GAL::3HA-rps17AΔrps17B* strain.

Most NES sequences have several characteristics in common: length of approximately 8-15 amino acids, intrinsic structural disorder, high accessibility and flexibility (are usually located either at the N-terminal or C-terminal extremities), 4-5 key hydrophobic residues (generally, but not exclusively leucine rich). Most of these NESs also have a similar secondary structure: 3 amino acids are part of an alpha helix followed by 2 residues part of a floppy strand that ensures the interaction with Crm1 (la Cour et al, 2004; Fu et al, 2011; Dong et al, 2009; Monecke et al, 2009; Güttler et al, 2010; Guettler & Goerlich, 2011; Xu et al, 2012). The NES can either be in a plus or minus orientation, depending on the localization of the $\Phi_3$-X-$\Phi_4$ strand motif (fig 4.9A): plus orientation if present at the C-terminal (PKI, HIV REV, SNUPN, Nmd3, Ltv1) and minus orientation if present at the N-terminal (Rio2, CPEB4) (Fung et al, 2015).

The putative NES sequence of Rps17 fits the consensus sequence reasonably well. It is also located at the C-terminus in an accessible region that appears to be highly flexible (fig 4.9B). Some of the hydrophobic leucine residues are part of an alpha helix while the remainder are part of a largely unstructured region (fig 4.9B). The Rps17 NES is also conserved in higher eukaryotes (data not shown). Altogether, this indicates that the Rps17 NES could be a real NES.

Detailed mutational analysis of the Nmd3, PKI and HIV REV NESs have shown that when the residues that formed the hydrophobic side were mutated the NES is no longer functional (Hedges et al, 2006; Güttler et al, 2010). Guided by these findings, I designed an Rps17 NES mutant (Rps17NESmut) in which I replaced 4 of the 5 hydrophobic leucine residues to alanines. These mutations included residues that are part of the structured alpha helix (and found on the same hydrophobic side) and of the flexible region (fig 4.9B). By mutating these conserved leucine residues, I wanted to determine if Rps17 NES is an active NES and if it is required for pre-40S export to the cytoplasm in a Crm1-dependent fashion.

Nmd3 is the best characterized Crm1 adapter involved in pre-ribosomal export. Therefore, as a positive control, I designed a construct in which I replaced the Rps17 NES with the Nmd3. I will refer to this construct as Rps17nmd3NES. Both the Rps17NESmut and the Rps17nmd3NES were stably expressed (fig

4.9C).

I next assessed the effect of these mutations on cell growth. I performed a complementation assay as described previously for Figure 4.2. Figure 4.9D shows that Rps17NESmut could partially support growth in Rps17 depleted cells, however, like the Rps17Δ100-136 mutant the Rps17NESmut also appears to affect growth on galactose, indicating it is dominant-negative.

Surprisingly, the Rps17nmd3NES mutant was unable to support growth on glucose medium. This demonstrates that the Nmd3 NES sequence is not functional in Rps17.

I next analysed the pre-RNA processing defects in cells dependent on the Rps17 NES mutant. The northern blot results (fig 4.10) showed that the Rps17NESmut causes pre-rRNA processing defects very similar to the longer C-terminal deletion mutant (Rps17Δ100-136). However, I did not observe accumulation of the 35S and 32S pre-rRNA precursors with the NES mutant. This suggests that these NES mutations do not significantly affect $A_0$-$A_2$ pre-rRNA processing, although this should be confirmed using a pulse-chase analysis.

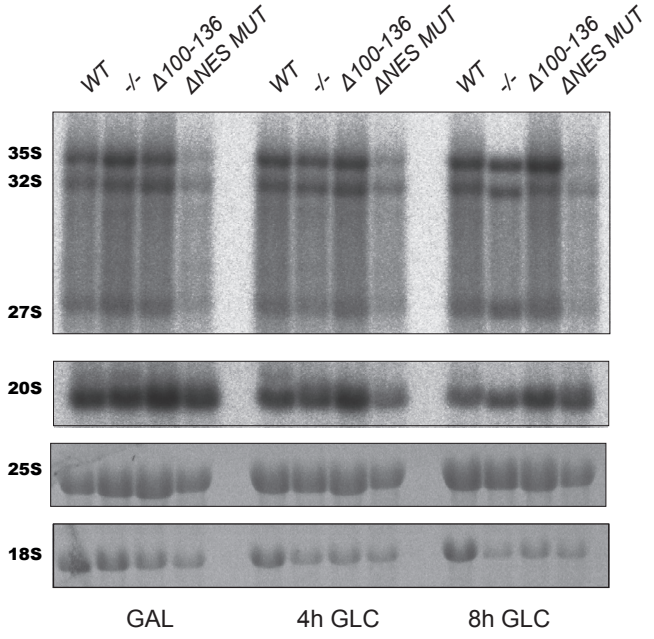



**Figure 4.10. The NES sequence in Rps17 is required for 18S rRNA synthesis.** Shown is a Northern Blot analysis of precursors and ethidium bromide stained mature rRNAs. Total RNA was extracted from strains expressing Rps17 WT and NES mutants (Rps17Δ100-136 and Rps17NESmut) in the presence of endogenous Rps17 (GAL) or absence of Rps17 (4h and 8h glucose depletion (4h, 8h GLC)).

## 4.9. The Rps17 NES is a weak NES

The next step was to test if Rps17 NES is an active NES. A previous study used a NLS-2xGFP-NES reporter construct to analyse NES sequences (Taura *et al*, 1998). I reasoned that I could use this reporter to test if the Rps17 NES sequence can export the GFP construct into the cytoplasm. The initial construct had an SV40 NLS and a PKI NES. I used the construct with the PKI NES as a positive control. I then replaced the PKI NES with the Rps17 NES and the mutated NES (leucine to alanine mutant). I then visualized the GFP signal using fluorescence microscopy (fig 4.11A). The PKI NES is clearly stronger than the SV40 NLS, as the GFP signal in the cytoplasm is higher than in the nucleus. If the Rps17 NES had been an active NES then I would have expected to see a similar trend as with the positive control (a higher GFP fluorescence signal in the cytoplasm compared to the nucleus). However the construct expressing the NESmut of Rps17 displayed a higher GFP fluorescence signal in the nucleus compared to the cytoplasm. I quantified the GFP signal intensity for each of the three constructs (20 cells each). I used ImageJ to measure the fluorescence intensity of the nucleus and then of the cytoplasm and calculated the nucleus/cytoplasm ratio for each construct (fig 4.11B). The lower the ratio, the higher the cytoplasmic signal compared to the nuclear signal. The nucleus/cytoplasm fluorescence intensity ratio for Rps17 NES is higher than for the PKI NES suggesting that the PKI NES is much stronger than the Rps17 NES. Moreover, the NESmut of Rps17 showed only a slight increase in nucleus/cytoplasm ratio, indicating that the mutations did slightly decreased the export of the GFP reporter protein but the results were not statistically significant. This indicates that if the Rps17 NES sequence is indeed functional, it is probably a weak NES.

In conclusion, the Rps17 NES fused to an exogenous reporter behaves like a weak NES. It is, however, important to note that previous studies have shown that NES can behave differently when part of a full length protein compared to being fused to an exogenous reporter (Hedges *et al*, 2006). Therefore, I will next be analysing how efficiently the FLAG-tagged Rps17 and FLAG-tagged Rps17NESmut are exported using immunofluorescence microscopy.

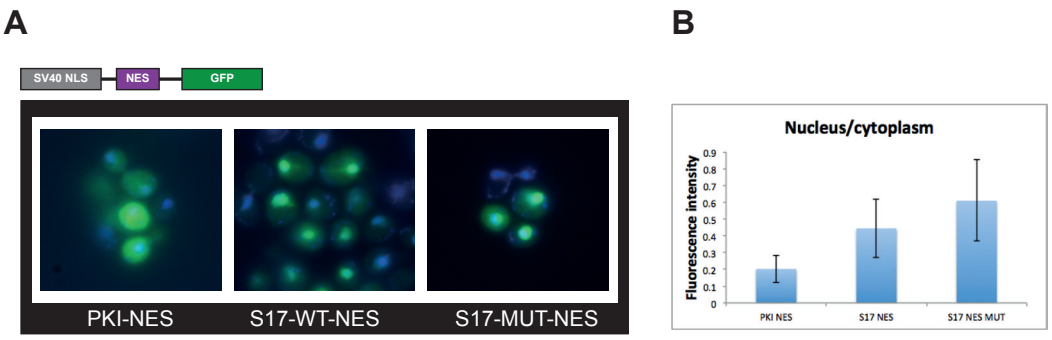



**Figure 4.11. The Rps17 NES is a weak NES.**
(A) A construct containing SV40 NLS-NES-GFP was used to asses how efficiently the wild-type (WT) and mutant (MUT) Rps17 NES can export GFP to the cytoplasm. The PKI NES was used as a positive control. (B) Quantification of the nucleus/cytoplasm fluorescence intensity ratio, calculated using ImageJ.

## 4.10. Does Rps17 NES bind Crm1 *in vitro*?

If the Rps17 NES sequence is indeed a Crm1 adapter, then I would expect the Rps17 NES to bind to Crm1 in the presence of RanGTP. Sander Granneman expressed and purified recombinant GST protein that was fused to the wild-type or mutated Rps17 NES from *E. coli* and I tested whether these proteins bound Crm1 (provided by Yu Min) in a RanGTP-dependent manner *in vitro* (RanGTP expressed by Atlanta Cook). The three proteins were mixed in roughly equimolar amounts with Crm1 and RanGTP and after a one hour incubation on ice, the samples were mixed with glutathione-sepharose beads (fig 4.12), which allowed me to determine if the GST-NES fusions bound to Crm1 and RanGTP. RanGTP appeared to bind the Rps17 NES mutant non-specifically (fig 4.12 lanes 5 and 8, input and GST pull-down). If Rps17 NES had interacted directly with Crm1 then we would have expected GST-Rps17 NES WT to pull down both Crm1 and RanGTP but not the NES mutant. Unfortunately, this was not the case (fig.4.12 lanes 6 and 9, GST-pull down). However, at this stage the results are inconclusive as I did not have a positive control that would allow me to determine if the protein preparations were biologically active. As mentioned above, it is also possible that the Rps17 NES amino acid sequence, when attached to GST, is simply not functional or accessible, or perhaps it

does not fold properly. These experiments will be repeated using full-length Rps17 protein. I finally managed to generate soluble Rps17 (wild type and mutant) by fusing it to the maltose binding protein (MBP) and therefore I will next repeat the *in vitro* binding experiments with the full-length Rps17 proteins.

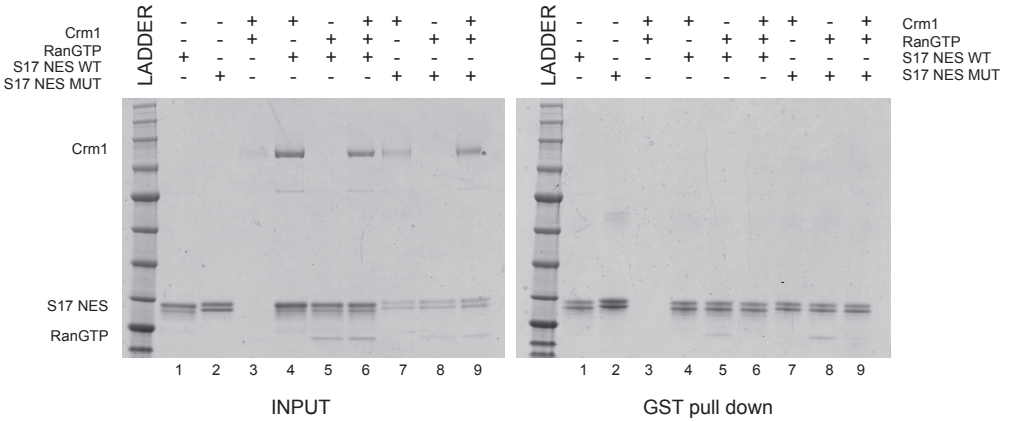



**Figure 4.12. Assessing Rps17 Crm1 interaction *in vitro*.**
*In vitro* binding assays with GST fused to the wild type Rps17 NES (WT) or mutant Rps17 NES (MUT) (constructs expressed by Sander Granneman) and recombinant Crm1 (provided by Yu Min) and RanGTP (provided by Atlanta Cook). The proteins were mixed in various different combinations.

## 4.11. Discussion

### What is the role of Rps17 in 40S maturation?

This study provides the first detailed characterization of the role of Rps17 in ribosome maturation. Rps17 is one the few 40S r-proteins that remained uncharacterized (de la Cruz *et al*, 2015). Here I show that Rps17 depletion affects early pre-rRNA processing steps as well as pre-40S nuclear export, a phenotype highly reminiscent to what was observed Rps5 and Rps15 mutants (Léger-Silvestre *et al*, 2004; Ferreira-Cerca *et al*, 2005).

My mutational analyses of Rps17 showed that (1) the first 26 amino acids are important for the incorporation of Rps17 into pre-ribosomes and (2) that the last 36 amino acids are essential for cell growth but not association with pre-40S particles. Although the Rps17Δ100-136 mutant caused almost identical processing defects as the absence of Rps17, the truncation mutant was dominant negative. This is most likely because the mutant is still incorporated into pre-ribosomes and therefore competes with the wild-type Rps17 for incorporation into pre-40S.

When I analysed the kinetics of rRNA processing in Rrp17 depleted cells using a pulse-chase method, I found that Rps17 was required for efficient processing at sites $A_0$-$A_3$. Genetic depletion of most of the small subunit r-proteins affected the processing at $A_0$-$A_2$ and also showed an accumulation of 23S pre-rRNA (Ferreira-Cerca *et al*, 2005). The presence of the 23S pre-rRNA is a sign that the processing at $A_0$-$A_2$ is bypassed through cleavage of $A_3$. I could not detect any accumulation of 23S pre-rRNA in Rps17 depleted cells, indicating that the processing at sites $A_0$-$A_3$ was affected to a similar extent. It seems that depletion of Rps17 leads to a general delay in the early rRNA processing steps.

Slow $A_0$-$A_3$ processing would explain why, at steady state, I only detected low levels of 20S pre-rRNA accumulation in Rps17 depleted cells although 18S rRNA levels were dramatically reduced. Defects in site D processing generally leads to a dramatic accumulation of 20S pre-rRNA in the cytoplasm (Granneman *et al*, 2005; Fatica *et al*, 2003). However, my FISH experiment did not show an accumulation of 20S pre-rRNA in the cytoplasm, indicating that

the 20S is mostly blocked in the nucleus. Cells lacking Rps5 or Rps15 show almost identical pre-rRNA processing defects as Rps17-depleted cells and are defective in nuclear export (Ferreira-Cerca *et al*, 2005). However, unlike these r-proteins, Rps17 actually has a putative NES in its C-terminal domain. Although I have not yet been able to demonstrate this conclusively, the data do suggest that the Rps17 NES could play a role in pre-40S export.

But if Rps17 depletion blocks pre-40S export, then why do the FISH results show no nuclear 20S pre-rRNA accumulation? The fact that the mature 18S rRNA is still made in the absence of Rps17 implies that the pre-40S nuclear export would be slowed down but not completely blocked. Previous work has shown that it is not always possible to measure changes in 20S pre-rRNA nucleus/cytoplasm distribution in r-protein mutants that cause moderate pre-40S export defects. In these mutants the 20S pre-rRNA export, albeit impaired, was able to balance the FISH nuclear to cytoplasmic fluorescence intensity after only 2h of r-protein depletion (Ferreira-Cerca *et al*, 2005).

In conclusion, I believe that Rps17 is involved in an (partially redundant) export pathway. However, more experiments are required to confirm this.

Is Rps17 a Crm1 adapter and does it mediate pre-40S export?

Pre-ribosomal subunits are exported into the cytoplasm through a Crm1 dependent pathway. Crm1 mediated export of pre-ribosomes is dependent on the transport adapters containing nuclear export signals (NES). Nmd3, an NES containing protein, was shown to be an export adapter for Crm1 and mediate pre-60S export (Ho *et al*, 2000). To date, several 40S-specific assembly factors (Dim2, Ltv1 and Rio1) were shown to have NES sequences. However, only Rio2 plays a (redundant) role in pre-40S export (Fischer *et al*, 2015). It was generally assumed that pre-40S and pre-60S export would involve ribosome assembly factors as export adapters. I was therefore surprised to find a ribosomal protein with a putative NES.

The NES of Rps17 is localized in the C-terminal domain. The Rps17 NES mutants (Rps17Δ100-136 and Rps17NESmut) cannot fully support growth and are dominant negative. I show that these mutants are still incorporated into pre-40S complexes, however, these mutants do not efficiently produce

18S rRNA. My fluorescent microscopy data suggest (although more evidence is needed) that Rps17 NES is an active but a weak NES. Most NES adapters generally bind Crm1 with very low affinity to allow dissociation once exported to the cytoplasm (Engelsma *et al*, 2004; Kutay & Guttinger, 2005). Therefore, these interactions generally are difficult to detect *in vitro*. The export of human 40S and 60S pre-ribosomes was reduced by 38% and 4%, respectively, when HeLa cells were transfected with PKI NES peptides (which competed with the pre-ribosome specific Crm1 adapters) (Thomas & Kutay, 2003). Therefore it is very likely that the Crm1 adapter for pre-40S export has a weak NES.

The *in vitro* binding assays with Rps17 NES peptides, Crm1 and RanGTP were inconclusive. At the time of writing I did not have a positive control to test whether the recombinant proteins were biologically active. However, it is also possible that the interaction between the Rps17 and Crm1 requires additional factors. This is based on the observation that the interaction between the Rio2 NES and Crm1, in addition to RanGTP required Slx9, a protein that was previously shown to be required for pre-40S export (Fischer *et al*, 2015). Rio2 also has a weak NES and it needs Slx9 to bind to Crm1 *in vitro*. Slx9 does not promote cooperative binding between Rio2, RanGTP and Crm1, but instead helps overcome the low affinity of Rio2 NES (Fischer *et al*, 2015). It could be that, similar to Rio2, the weak Rps17 NES might need Slx9 to mediate Crm1 binding.

Slx9 does not bind strong NESs like the one in Nmd3 (Fischer *et al*, 2015). The Rps17 mutant that contained the Nmd3 NES failed to complement growth in Rps17 depleted cells. Adding the Nmd3 NES to other r-proteins was previously attempted and was successful only in some cases: fusing the NES of Nmd3 to Rpl3 restored pre-60S export in export defective strains, while fusing it to Rpl11 did not (Lo & Johnson, 2009). It is unclear why fusing Nmd3 NES to r-proteins can have different outcomes but it may be linked to how the NES folds when it is fused to a different protein, to whether the fusion protein is efficiently incorporated into pre-ribosomes as well as to the accessibility of the NES for Crm1 binding (Lo & Johnson, 2009). This may explain why the Rps17nmd3NES did not support growth.

I show that Rps17 NES has all the structural characteristics of a Crm1 export adapter for pre-40S subunit export. Clearly, several experiments need

to be performed to prove this conclusively and this will be done once this thesis has been submitted. The reason why it is important to continue this project is that Rps17 could be the first example of an r-protein that mediates pre-ribosome export in a Crm1 dependent manner. This is why it is important to continue studying Rps17 as it could lead to a significant breakthrough in the ribosome synthesis field.

Why is Rps17 repositioned in the late pre-40S complexes?

I initially thought the N-terminal region of Rps17 simply could not bind to H37 due to steric hindrance, as Nob1 binds closely to this region in early and middle pre-40S complexes. This led to speculations regarding a possible role for Nob1 in regulating the assembly of r-proteins, apart from its endonuclease function. I tested this hypothesis by depleting Nob1 and asking whether I could detect a change in the chemical probing pattern in H37, however, these experiments were inconclusive. What I did learn from these experiments is that particles that lack Nob1 can continue to mature into late pre-40S complexes. Similarly, cryo-EM data of Nob1 depleted intermediates showed that Nob1 absence did not block the assembly of the other assembly factors (Strunk *et al*, 2011).

These results also motivated us to only perform chemical probing experiments on native pre-ribosomes and not in assembly factors depletion mutants as the results could be misleading, unless purifying pre-rRNA intermediates from native particles was technically extremely hard to achieve.

The second hypothesis, which is also my current working model, revolves around the NES sequence of Rps17. In the mature 80S ribosome the CTD of Rps17 (including the NES) binds to r-protein Rps0 and it is possible that this makes the NES inaccessible to Crm1. This could explain why Rps17 must not be able to prematurely assume its final conformation in the pre-40S particles as this would prevent efficient pre-40S export (fig 4.13).

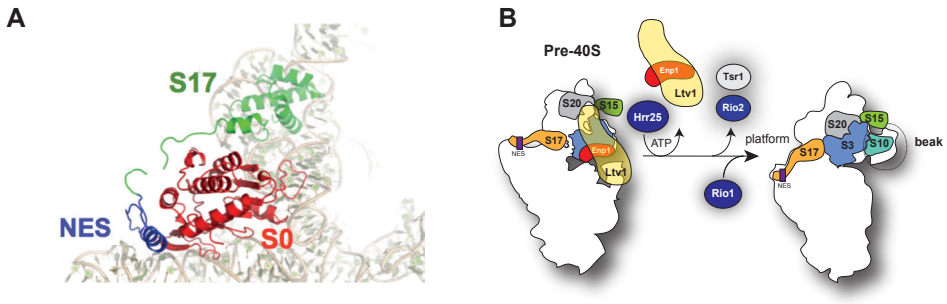



**Figure 4.13 The C-terminal domain of Rps17 must remain accessible for optimal nuclear export.**
(A) In the crystal structure the C-terminal domain of Rps17 contacts Rps0. (B) Working model: the delay in Rps17 incorporation into the pre-40S might be required for the Rps17 C-terminal domain (containing the NES) to be accessible to Crm1. Binding of Rps17 to Rps0 might shield the NES.

## Future plans

Several key experiments are needed in order to directly prove that Rps17 is Crm1 adaptor and that it mediates pre-40S export. Future plans include testing by immunofluorescence the ability of the WT and MUT NES to export the entire Rps17 protein. Previous attempts to determine the distribution of full length Rps17 containing the WT or MUT NES were problematic due to the fact that the FLAG tag was hard to detect in the cells by immunofluorescence.

The key experiment that needs to be done is to perform the MBP-Rps17-Crm1-RanGTP pull-down assay and also to perform it in the presence of Slx9. I believe that due to the weak NES, Rps17 might need Slx9 to bind Crm1 in the presence of RanGTP. As Rio2 binds Crm1 in the presence of Slx9 (Fischer *et al*, 2015), Rio2 will be used as a positive control. The outcome of this experiment will determine the impact of this story.

# CHAPTER 5

# Ribosomal RNA Structural Rearrangements During Nucleolar pre-60S Assembly

## 5.1. Introduction and aim

This chapter discusses my work on the dynamics of RNA structural rearrangements that take place during nucleolar pre-60S maturation steps and the interplay between rRNA restructuring and protein binding events.

Many of the pre-rRNA processing steps in pre-60S formation take place in the nucleolus. Within this subnuclear compartment, the 35S pre-rRNA is cleaved, ITS1 and ITS2 sequences are removed and the precursors of 5.8S and 25S rRNAs are formed as a result of 27SB pre-rRNA cleavage (at site $C_2$). The rRNA processing is regulated by the binding/release of assembly factors and the gradual stabilization of r-proteins, which bind to the 90S initially with low affinity. These assembly factors and r-proteins form a complex network in which their binding/stabilization is hierarchical and inter-dependent (fig 5.1). Firstly, the binding of $A_3$ factors and the stable incorporation of r-proteins such as Rpl3, Rpl4, Rpl7 and Rpl8 are essential for $27SA_3$ formation. All these proteins need to bind to the pre-60S for r-proteins Rpl17, Rpl37, Rpl26, Rpl35 to become stably incorporated (Jakovljevic *et al*, 2012; Sahasranaman *et al*, 2011). Furthermore, Rpl17, Rpl35 and Rpl37 are required for the stable binding of B factors such as Nsa2 and Nog2 (Gamalinda *et al*, 2013). Nog2 is the last B factor to bind before $C_2$ cleavage takes place and before the 7S is separated from the 25.5S pre-rRNAs (Saveanu *et al*, 2001). Apart from Rpl17, Rpl35, Rpl37, the B factors Nsa2 and Spb4 need to be incorporated before Nog2 binds to the pre-60S intermediates. In turn, Nsa2 and Sbp4 are recruited through two independent pathways: Spb4 through the Rpf2 complex and Nsa2 through Mak11, Rlp24, Dbp10 pathway (Talkish *et al*, 2012).

Consequently, depleting any of these assembly factors or r-proteins impairs the binding of downstream proteins and ultimately the cleavage of 27SB pre-rRNA. This is because these proteins not only function as protein-

protein interaction scaffolds but also because they are believed to be required for the formation of a 27S pre-rRNA conformation that is compatible with $C_2$ cleavage.

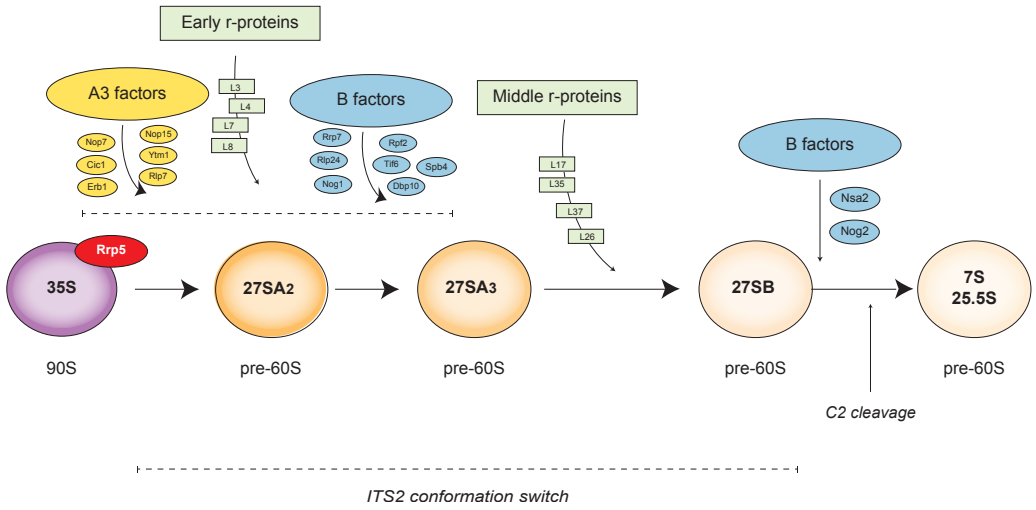



**Figure 5.1. Schematic representation of the nucleolar steps of pre-60S assembly**
The nucleolar rRNA processing steps. Cleavage of 27SB pre-rRNA is dependent on the binding of A3 factors, B factors, early and middle r-proteins.

It was previously reported that a major rRNA structural change might act as a checkpoint for 27SB pre-rRNA cleavage. The combined results of rRNA structure predictions, enzymatic and chemical structure probing experiments and genetic analyses (Joseph *et al*, 1999; Yeh & Lee, 1990; Côté *et al*, 2002; Granneman *et al*, 2011; Peculis & Greer, 1998) led to the idea that ITS2 undergoes a conformational switch that would allow 27SB pre-rRNA cleavage at site $C_2$. All of these ITS2 chemical probing studies were performed using total RNA and these analyses were all done in depletion or deletion strains. Because total RNA contains a mixture of different pre-rRNA species, it was not always possible to dissect the differences in rRNA secondary structure and linking them to the individual pre-rRNA processing steps. In order to effectively separate the distinct nucleolar pre-60S processing steps and to obtain a step by step picture of the changes in rRNA folding during 60S maturation, my aim was to isolate the 35S, $27SA_2$, 27SB pre-rRNA intermediates from native particles and compare their secondary structure using SHAPE reagents and ChemModSeq.

The results presented in this chapter represent the first nucleotide

resolution snapshots of the RNA flexibility of the entire 35S, $27SA_2$ and 27SB pre-rRNAs. Using these snapshots we have obtained detailed insights into the dynamics of rRNA structural rearrangements of pre-60S nucleolar assembly.

## 5.2. 35S, $27SA_2$ and 27SB pre-rRNA can be efficiently purified from pre-ribosomes

The first goal was to purify the pre-ribosomes containing the 35S, $27SA_2$ and 27SB pre-rRNA intermediates. For this purpose I used as bait assembly factors that allowed me to specifically affinity purify these pre-rRNA species. TAP-tagged Rrp5 was used to enrich for $27SA_2$ while TAP-tagged Nsa2 was used to enrich for 27SB pre-rRNAs (fig 5.2A). Rrp5 and Nsa2 were shown to be part of different pre-60S intermediates: Rrp5 binds to the 90S and remains associated to early pre-60S intermediates containing 35S, $27SA_2$ and $27SA_3$ pre-rRNAs while Nsa2 binds to 27SB pre-rRNA shortly before $C_2$ cleavage (Lebreton *et al*, 2008; 2006a). To obtain sufficient amounts of 35S pre-rRNA, Fredrik Lackmann and myself used an Mrd1 mutant that accumulated 35S pre-rRNA. The 35S accumulates in this strain because processing at $A_0$, $A_1$ and $A_2$ is delayed. As processing at site $A_3$ can still occur, the mutation does not affect pre-60S subunit maturation.

After the IgG affinity purification, the co-precipitated rRNAs were resolved on an agarose gel and the pre-rRNAs of interest were excised from the gel and purified (fig 5.2B). To assess the homogeneity of the purified pre-rRNA species, I performed primer extension analyses. Consistent with our expectations, the pre-rRNA isolated using Rrp5 as bait was predominantly $27SA_2$ pre-rRNA. Only very low levels of $27SA_3$ and 27SB were present in the purified material (fig 5.2C, lane 1). The rRNAs that co-precipitated with Nsa2 were almost exclusively 27SB pre-rRNAs (fig 5.2C, lane 2). Because the $27SA_3$ pre-rRNAs were very low abundant in my purified 27SA samples, I will refer to the 27SA species as $27SA_2$ pre-rRNA.

In conclusion, I was able to isolate pure and sufficient quantities of 35S pre-rRNA, $27SA_2$ pre-rRNA and 27SB pre-rRNA for structure probing analyses.

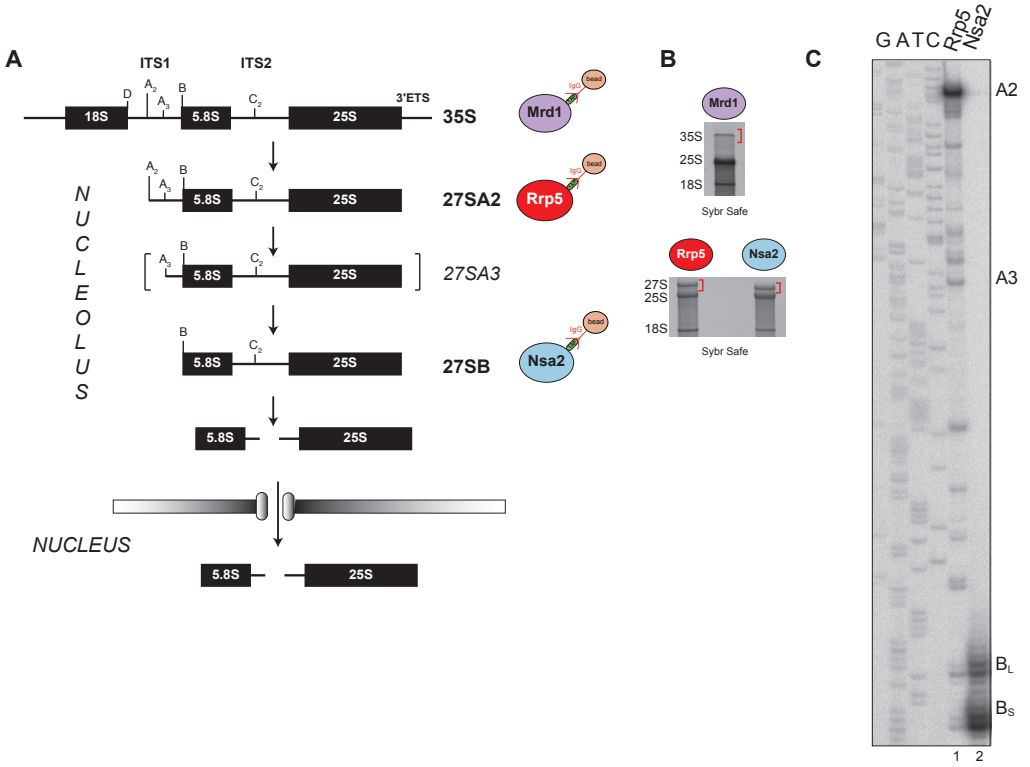



**Figure 5.2. Isolation of 35S, 27SA$_2$ and 27SB pre-rRNAs using specific bait proteins.** (A) An overview of the nucleolar pre-rRNA processing steps and of the baits used to affinity purify 35S pre-rRNA (Mrd1), 27SA2 pre-rRNA (Rrp5) and 27SB pre-rRNA (Nsa2): (B) 35S, 27SA2 and 27SB pre-rRNAs were immunoprecipitated using the baits indicated above each lane, resolved by gel electrophoresis and gel purified; (C) the 27S pre-rRNA species can be individually isolated using specific baits. Shown are the results of primer on rRNAs isolated from the Rrp5 and Nsa2 pre-60S particles.

## 5.3. Rrp5 and Nsa2 pre-60S intermediates contain a different set of assembly factors

As protein binding can affect rRNA folding, I also wanted to determine whether the protein composition of the Rrp5 and Nsa2 particles differed. This was done in collaboration with Carolina Aguilar from Marlene Oeffinger's lab who performed label free quantitative mass spectrometry on the purified pre-ribosomes. The results are largely consistent with previous mass spectrometry studies on pre-60S particles (Babiano & de la Cruz, 2010; Gamalinda *et al*, 2014; Ohmayer *et al*, 2013; Jakovljevic *et al*, 2012). Firstly, in accordance

with the clear-cut separation of the $27SA_2$ and 27SB pre-rRNAs, Rrp5 was not detected in the Nsa2 particles. *Vice versa*, Nsa2 was absent in Rrp5 particles. Moreover, early 90S-associated assembly factors only co-precipitated with Rrp5 and not Nsa2. This is consistent with the fact that Rrp5 associates with both the 90S pre-ribosomes and early pre-60S intermediates (Venema & Tollervey, 1996) while Nsa2 associates later with the 27SB pre-rRNA containing pre-60S intermediates (Lebreton *et al*, 2006a; 2008).

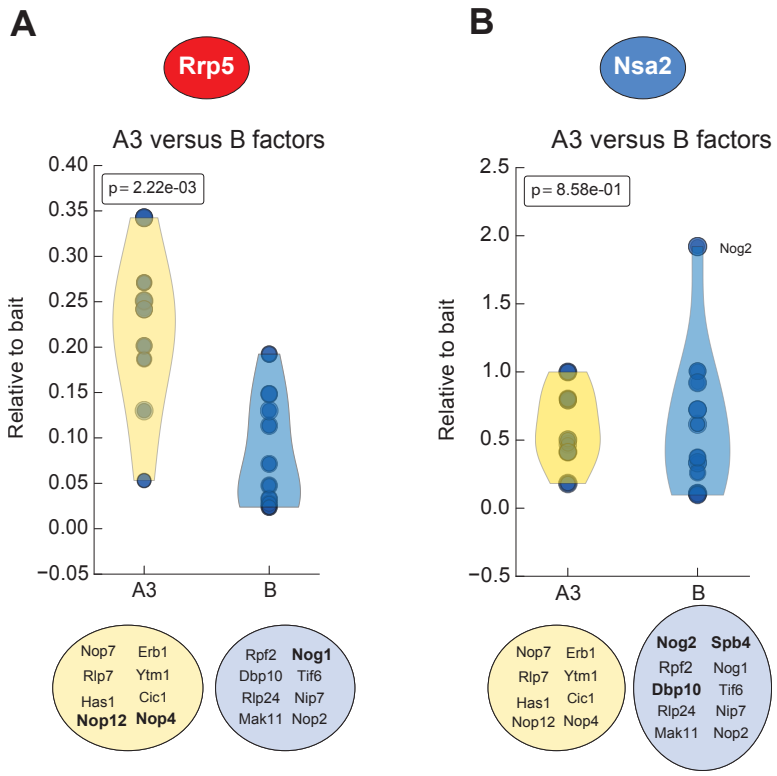



**Figure 5.3. The distribution of assembly factors in Rrp5 and Nsa2 affinity purified pre-60S intermediates.**
(A-B) Violin plots showing the mass spectrometry results for the ribosome assembly factors. The plot shows the enrichment of A3 factors and B factors in the isolated pre-60S complexes relative to the bait protein (Rrp5 (A) or Nsa2 (B)).

The $A_3$ factors Nop12 and Nop4 were significantly enriched in the Rrp5 (fig 5.3A). This was not surprising since Nop12 deletion blocks $27SA_3$ formation (Granneman *et al*, 2011) and Nop4 was also previously found in Rrp5 containing particles (Saveanu *et al*, 2003). $A_3$ factors Nop7, Erb1, Ytm1, Rlp7, Cic1, Nop15 and Has1 bind to the $27SA_2$ containing pre-60S intermediates and their binding is interdependent (Sahasranaman *et al*, 2011). Except for Nop15,

all $A_3$ factors could be detected in the Rrp5 and Nsa2 particles. Although they are needed for $A_3$ processing, they dissociate at a later step of maturation. Nop7-Erb1-Ytm1 dissociate from the pre-60S shortly before nucleolar export (Bassler *et al*, 2010) and Has1 dissociates before Nog2 binds (Dembowski *et al*, 2013a).

Although all B factors are required for Nog2 incorporation the actual assembly of these factors takes place in two stages (Talkish *et al*, 2012). Some B factors bind to the early 35S and $27SA_2$ containing pre-ribosomes (Nop2, Nip7, Rpf2, Rrs1, Rlp24, Nog1, Tif6, Mak11, Spb4, Dpb10), while others bind only just before $C_2$ is cleaved (Nsa2, Nog2). The hierarchical binding of the early B factors starts with Nop2-Nip7 complex that binds to the 90S complexes. This allows the association of Rpf2 and Rlp24, Nog1, Tif6. Rlp24, Nog1, Tif6 association is interdependent but independent of Rpf2. Mak11, another B factor that binds early, most likely dissociates upon Nog1 binding. (Zhang *et al*, 2007; Saveanu *et al*, 2003; Talkish *et al*, 2012; Lebreton *et al*, 2008). Consistent with this, the early B factors were detected in both the Rrp5 and Nsa2 pre-60S intermediates, whereas the later B factors were predominantly present in the Nsa2 particles. The Dpb10 and Spb4 B factors were found to be enriched in the Nsa2 particles, while Nog2 was identified only in the Nsa2 particles where it was almost twice as abundant as the bait Nsa2 (fig 5.3B)

The mass-spectrometry results are in excellent agreement with previously published work. This confirmed that the rRNA intermediates analysed here are part of pre-ribosomal particles that have all the expected assembly factors.

Interestingly, in the Rrp5 associated pre-ribosomes, the $A_3$ factors were significantly enriched over the B factors (fig 5.3A) while in the Nsa2 particles the $A_3$ and B factors were present in comparable amounts (fig 5.3B). Also, $A_3$ factors such as Cic1 and Rlp7 were detected in the Nsa2 particles as well as the Rrp5 particles. Cic1 and Rlp7, together with Nop15, cross-link to ITS2 and are thought to play a role $C_2$ cleavage by inducing a cleavage competent conformation for the ITS2. It seems that whatever rRNA structural changes take place during the nucleolar rRNA processing steps the $A_3$ factors are present throughout. As a result, it is unlikely that Cic1 and Rlp7 dissociation is needed for a possible ITS2 conformational change, nor for the binding of the

late B factors, as previously proposed (Granneman *et al*, 2011; Dembowski *et al*, 2013b). However, the data do suggest that the release of Rrp5 might be important for the subsequent assembly of late assembly factors such as Nsa2 and Nog2.

Unfortunately, most of the r-proteins were underrepresented in the mass-spectrometry data. A possible explanation could be that many of the trypsin digested fragments were not detected because they were too small or did not ionize properly. At present, the Oeffinger lab is testing different proteases for digestion of the proteins, which is expected to increase the detection of ribosomal proteins (r-proteins).

## 5.4. Classifying rRNA restructuring events based on ChemModSeq data analysis

The next step was to apply ChemModSeq to my purified pre-ribosomes (fig 5.4). I initially probed the rRNA both *in vivo* using NAI and *in vitro* (on purified pre-ribosomal particles) using 1M7 SHAPE chemicals. However when using NAI, far less 27S and 35S pre-rRNA was recovered after the IgG purification step, indicating that the conditions needed for efficient NAI modification (in PBS after pelleting the cells) and possibly the longer incubation time (15 min *in vivo* for NAI compared to 3 min *in vitro* for 1M7) might lead to disruption of the pre-60S particles. As a result, I decided to only perform the *in vitro* 1M7 modification. I also wanted to determine the flexibility of the pre-rRNAs in the absence of proteins. To generate these rRNAs, I treated the particles with proteinase K prior to 1M7 modification (fig 5.4A). This enabled me to determine what changes in rRNA flexibility could have been the result of protein binding events. The expectation was that those regions that were bound by proteins in the pre-ribosomes would become more reactive to SHAPE in the deproteinized particles, in particular the regions where proteins contact single-stranded rRNA.

The purity and the integrity of the 1M7 modified rRNA samples obtained were analysed on a Bioanalyser (fig 5.4B) and primer extension was performed to assess whether the samples were indeed modified. Notably, I

found that when I added SYBRsafe to the agarose to visualise the rRNA before gel extraction, I obtained many additional reverse transcription stops when checking the purified pre-rRNAs by primer extension. This suggested that the SYBRsafe interfered with reverse transcription, even though the rRNA was purified using silica resin (Qiagen), which should have removed it. I managed to solve this problem by post-staining the agarose gels with low concentrations of dye.

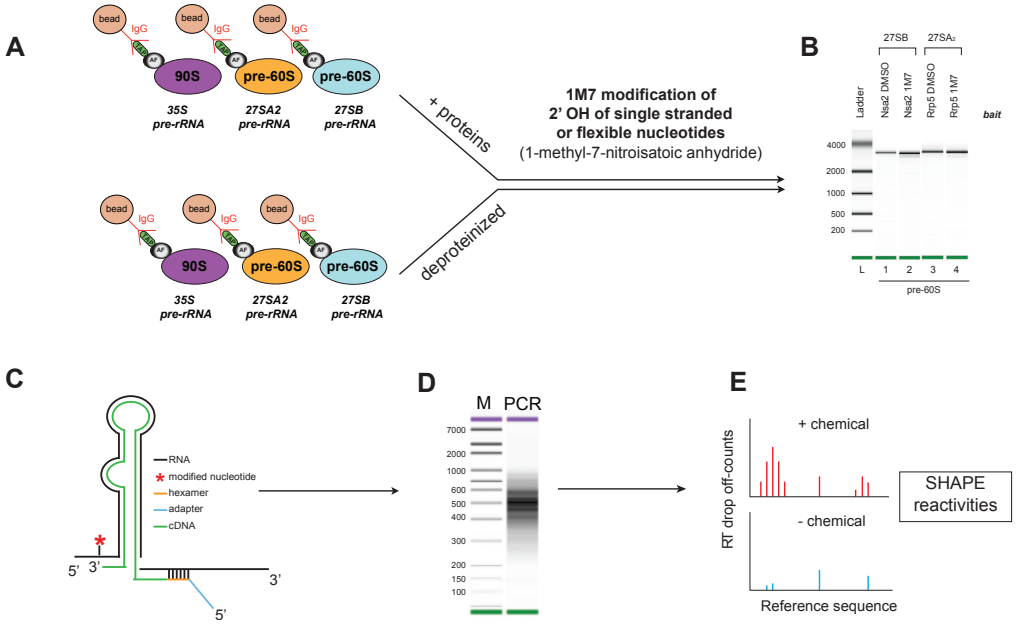



**Figure 5.4. The ChemModSeq experimental workflow.**
(A) The 35S, 27SA$_2$ and 27SB pre-rRNAs were individually purified using TAP-tagged Mrd1, Rrp5 or Nsa2 as baits, respectively. Analyses were performed on pre-RNA isolated from intact and deproteinized pre-ribosomal complexes. The pre-rRNAs were modified *in vitro* using 1M7 SHAPE chemical or treated with solvent (DMSO) (B) The modified rRNAs were extracted, gel purified and their integrity checked on a Bioanalyzer; (C) cDNA libraries, generated by random priming, were PCR amplified and (D) sequenced on Illumina HiSeq systems. (E) SHAPE reactivities were calculated for each sample: treated/untreated with 1M7, native/deproteinized.

Next I applied the ChemModSeq protocol to the purified pre-rRNAs (fig 5.4C-E). As a measure of RNA flexibility, I calculated SHAPE reactivities (according to (Ding *et al*, 2014)). The reason why I calculated SHAPE reactivities and did not use the TCP$^{EM}$ algorithm for these data (see Chapter 3) was that during the course of this work I noticed that many subtle (but reproducible) changes in rRNA flexibility were not always detected by the TCP$^{EM}$ algorithm. Using SHAPE reactivities therefore increased the richness of the data.

The SHAPE reactivities for each nucleotide were generated by summing the read counts of the two replicates and then normalizing to values between 0 and 2 (Ding *et al*, 2014). Several nucleotides have SHAPE reactivities higher than 2. These hyper-reactive nucleotides generally adopt unusual conformations that favour SHAPE chemical modification (McGinnis *et al*, 2012).

Because there was some variability in the level of 1M7 modification between experiments, I defined reactivity thresholds by calculating quartiles for each dataset. The nucleotides with SHAPE reactivities below the first quartile were considered unmodified. Nucleotides with SHAPE reactivities above the first quartile were considered modified, while the nucleotides with SHAPE reactivities above the third quartile were considered highly modified.

When analysing the data, I made several assumptions (fig 5.5): (1) If the SHAPE reactivity signal in deproteinized samples was *the same* as observed in pre-ribosomes containing 27SB pre-rRNAs, then I assumed that any changes observed in SHAPE reactivity between 35S and 27SB pre-rRNA samples was the result of *rRNA (re)folding events* (fig. 5.5C-D). It is possible that nucleotides that belong to this category become more stabilized or more flexible due to proteins binding to that region, however the structure that is formed does not require the presence of the proteins to remain stable. Therefore it is likely that not all protein-RNA binding events are detected. (2) If the SHAPE reactivity signal in deproteinized samples was *the opposite* of what I observed in pre-ribosomes containing 27SB pre-rRNAs (fig 5.5D-F), then I assumed that any changes observed in SHAPE reactivity between 35S and 27SB pre-rRNA samples must be the result of *a protein binding event*. The presence of the protein is also required to keep the new rRNA structure stable as removing the protein led to an increase in flexibility of the nucleotide. As a result, the data interpretation does not determine what triggered the change in rRNA structure in the first place, but it can accurately distinguish if an adopted rRNA secondary structure is dependent on the presence of proteins or not.

There is a set of scenarios that can describe the SHAPE data for each nucleotide based on the assumptions mentioned above. These scenarios are schematically presented in Figure 5.5 and are further described below. For the purpose of this analysis I took into consideration only the nucleotides for which the SHAPE pattern in the deproteinized $27SA_2$ and 27SB samples was the

same: either unmodified or modified.

The simplest scenario is where I see no changes in reactivity (either no reactivity or high reactivity in all the samples), as shown in Figure 5.5B. Here I assumed that the overall structure of these nucleotides was probably not influenced by the presence or absence of proteins. Figure 5.5C shows an example where nucleotides are only reactive in the earliest pre-rRNA precursors, but not in the late 27SB pre-rRNA or the deproteinized samples. In this case I would conclude that an rRNA folding event had taken place during the conversion of 35S to 27SB pre-rRNA, which resulted in the formation of a stable structure, such as an RNA duplex. Moreover, this stable structure once formed was not destabilized by removing the proteins (low SHAPE reactivities in the deproteinized data). In Figure 5.5D I show the opposite scenario: low reactivity in early particles and high reactivity in late (27SB pre-rRNA) and deproteinized particles. In this situation I would conclude that an rRNA refolding step had occurred that resulted in the nucleotide becoming more flexible. In figure 5.5E-G I show a couple of scenarios that are most likely the result of a protein contacting the nucleotide. Figure 5.5E represents a scenario in which there is no change in reactivity as the 35S is converted to 27SB pre-rRNA, but the reactivity profile changes in the deproteinized sample indicating that the presence of a protein is needed for the initial rRNA structure to be maintained. If I would see a reactivity profile shown in figure 5.5F, I would conclude that a protein bound directly to or in proximity of this nucleotide during the conversion of 35S to $27SA_2$ or $27SA_2$ to 27SB, which resulted in an increase in flexibility. The scenario in figure 5.5G is likely to occur when a protein bound this nucleotide shortly after the processing of the 35S pre-rRNA, resulting in the nucleotide becoming less flexible. The SHAPE reactivity is (at least partly) restored when the protein is removed by proteinase K.

The main reason for making this classification was that it helped to simplify the data analyses. Although this classification is not exhaustive (it excludes the nucleotides for which the modification pattern is different in the $27SA_2$ and 27SB pre-rRNA samples) it aids the interpretation of the results by focussing on a specific set of (main) scenarios.

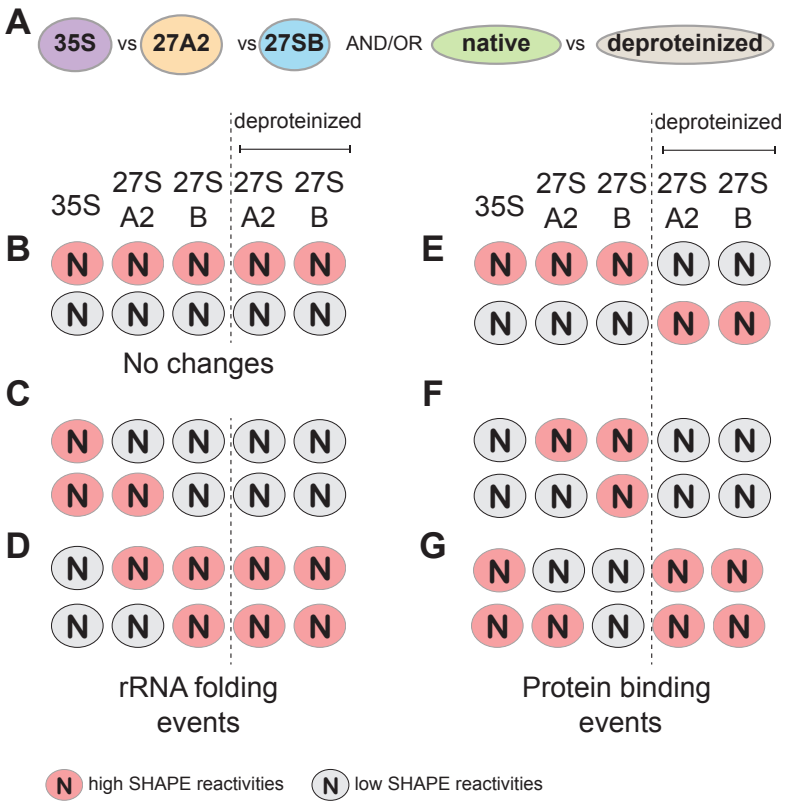



**Figure 5.5 Classifying rRNA restructuring events.**

(A) SHAPE reactivities for the three pre-rRNA intermediates from the particles are compared to those of the deproteinized samples. (B-G) Catagorization of SHAPE reactivity profiles (B) No changes in reactivity (either no reactivity or high reactivity in all the samples). The overall structure of these nucleotides was probably not influenced by the presence or absence of proteins. (C) An example where nucleotides are only reactive in the earliest pre-rRNA precursors, but not in the late 27SB pre-rRNA or the deproteinized samples. Here we would conclude that an rRNA folding event had taken place during the conversion of 35S to 27SB pre-rRNA, which resulted in the formation of a stable structure, such as an RNA duplex. Moreover, this stable structure once formed was not destabilized by removing the proteins (low SHAPE reactivities in the deproteinized data). (D) The opposite scenario: low reactivity in early particles and high reactivity in late (27SB pre-rRNA) and deproteinized particles. In this situation an rRNA folding step had likely occurred that resulted in the nucleotide becoming more flexible. (E-G) Three scenarios that are most likely the result of a protein contacting the nucleotide. (E) represents a scenario in which there is no change in reactivity as the 35S is converted to 27SB pre-rRNA, but the reactivity profile changes in the deproteinized sample indicating that the presence of a protein is needed for the initial structure to be maintained. (F) A protein bound directly to or in proximity of this nucleotide during the conversion of 35S to 27SA2 or 27SA2 to 27SB, resulting in an increase in flexibility. The scenario in (G) is likely to occur when a protein bound this nucleotide shortly after the processing of the 35S, resulting in the nucleotide becoming less flexible. The SHAPE reactivity is (at least partly) restored when the protein is removed by proteinase K.

## 5.5. The 5' end of the 27S pre-rRNA becomes more structured while the 3' end becomes more flexible as 60S pre-rRNA processing proceeds

The heat map in Figure 5.6B shows a digital snapshot of changes in rRNA flexibility during nucleolar stages of 60S maturation. Here I only focus on the SHAPE reactivates for the 27SB corresponding region in all three 35S, $27SA_2$, 27SB pre-rRNAs (Figure 5.6A). Each block indicates a single nucleotide. The darker blue the nucleotide, the higher the SHAPE reactivity. The last column shows the binding sites for ribosomal proteins. If a nucleotide is coloured blue in this column then, according to the yeast 80S crystal structure (Ben-Shem *et al*, 2010), it is contacted by one or more r-proteins. The data from the heat map revealed many rRNA restructuring events, which are summarized in figure 5.6C. Thirty percent of the total 3396 nucleotides analysed were not taken into consideration as their coverage was too low (the nucleotide was covered by less than five reads) or their reactivity pattern did not fit in the categories described above (fig. 5.5; fig 5.6C, grey coloured sections). Interestingly, roughly 40% of all analysed nucleotides did not show any changes in SHAPE reactivities in all of the pre-rRNA samples. Of these, 75% of the nucleotides were constrained (have low SHAPE reactivities and therefore are likely base-paired) (fig 5.6C, dark green) and 25% were flexible nucleotides (have high SHAPE reactivities and therefore were likely in single stranded regions) (fig 5.6.C, light green). This indicates that a significant fraction of the 27S pre-rRNA probably folds co-transcriptionally. The remaining 30% of all the analysed nucleotides appeared to undergo post-transcriptional rRNA structural rearrangements during the conversion of 35S to 27SB pre-rRNA. Of all the nucleotides that showed a clear change in reactivity, two thirds showed SHAPE reactivity patterns outlined in Figure 5.5F and G, being categorized as being part of protein binding events (fig 5.6C light orange). The rest showed a reactivity pattern identical to what is shown in figure 5.5C and D and were considered rRNA (re)folding events (fig 5.6C dark orange). Moreover, 1940 of all the nucleotides in the 5.8S and 25S rRNAs sequences were reactive to 1M7, suggesting they are in single-stranded regions. This is only marginally higher than the 1863 nucleotides reported to be single stranded in the crystal structure (Bernier *et al*, 2014).

In conclusion, using ChemModSeq I was able to generate a comprehensive map of changes in rRNA flexibility during nucleolar stages of pre-60S pre-rRNA processing events.

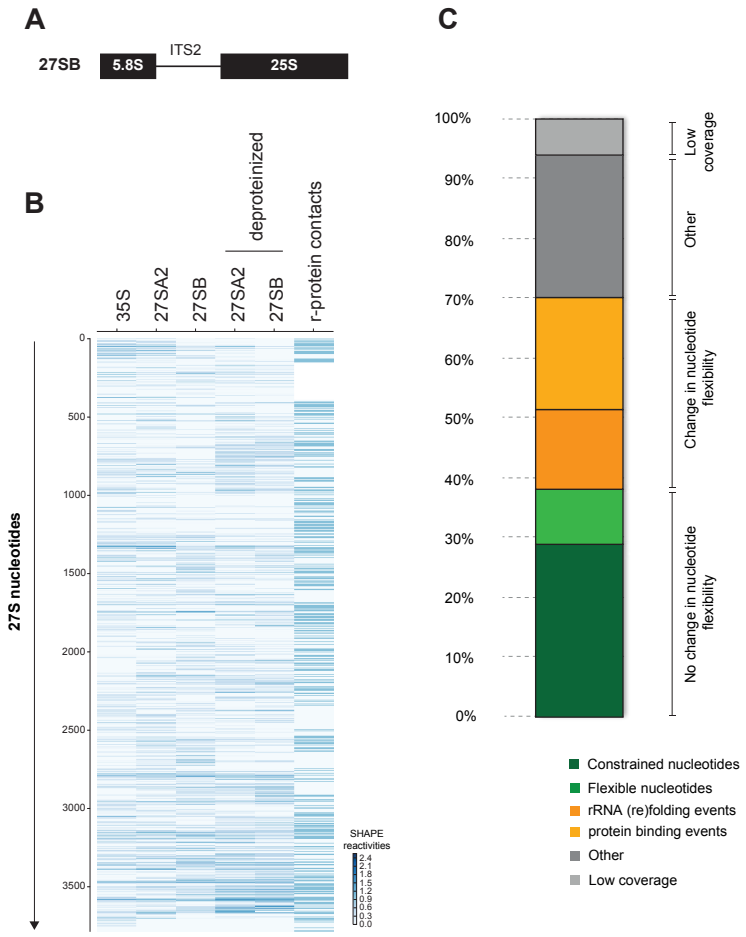



**Figure 5.6. A digital snapshot of the ChemModSeq data.**
(A) The 27SB pre-rRNA sequence contains the mature 5.8S and 25S corresponding sequences separated by ITS2; (B) A heat map represents a digital snapshot of changes in SHAPE reactivity during the conversion of 35S (left), to 27SB pre-rRNA. The data for the deproteinized 27SA2 and 27SB pre-rRNA samples was also included. Each block indicates a single nucleotide. The darker the blue color, the higher the SHAPE reactivity. The last column of the heat map represents known interactions with ribosomal proteins. (C) Classifying the rRNA restructuring events based on the categories described in fig 5.5: a third of the total number of nucleotides was not taken into consideration as it had either low coverage or did not fit any of the categories previously described (grey), a third underwent rRNA structural changes as the 35S pre-rRNA was processed to 27SB pre-rRNA (orange: dark orange nucleotides that fitted one of the rRNA folding events categories and light orange nucleotides that fitted one of the protein binding events categories) and the remaining third did not undergo any change in rRNA flexibility (green: dark green nucleotides that were constrained, low SHAPE reactivities, and light green nucleotides that were flexible, high SHAPE reactivities).

## 5.6. The 5.8S undergoes major conformational changes during the conversion of 35S pre-rRNA to 27SB pre-RNA

The heat map in figure 5.7A represents the SHAPE reactivities for each nucleotide in the 5.8S region. Intriguingly, the ChemModSeq data suggested that the 5.8S region only becomes properly folded during the conversion of 27SA$_2$ to 27SB pre-rRNA (fig 5.7A, compare lane 2 and 3).

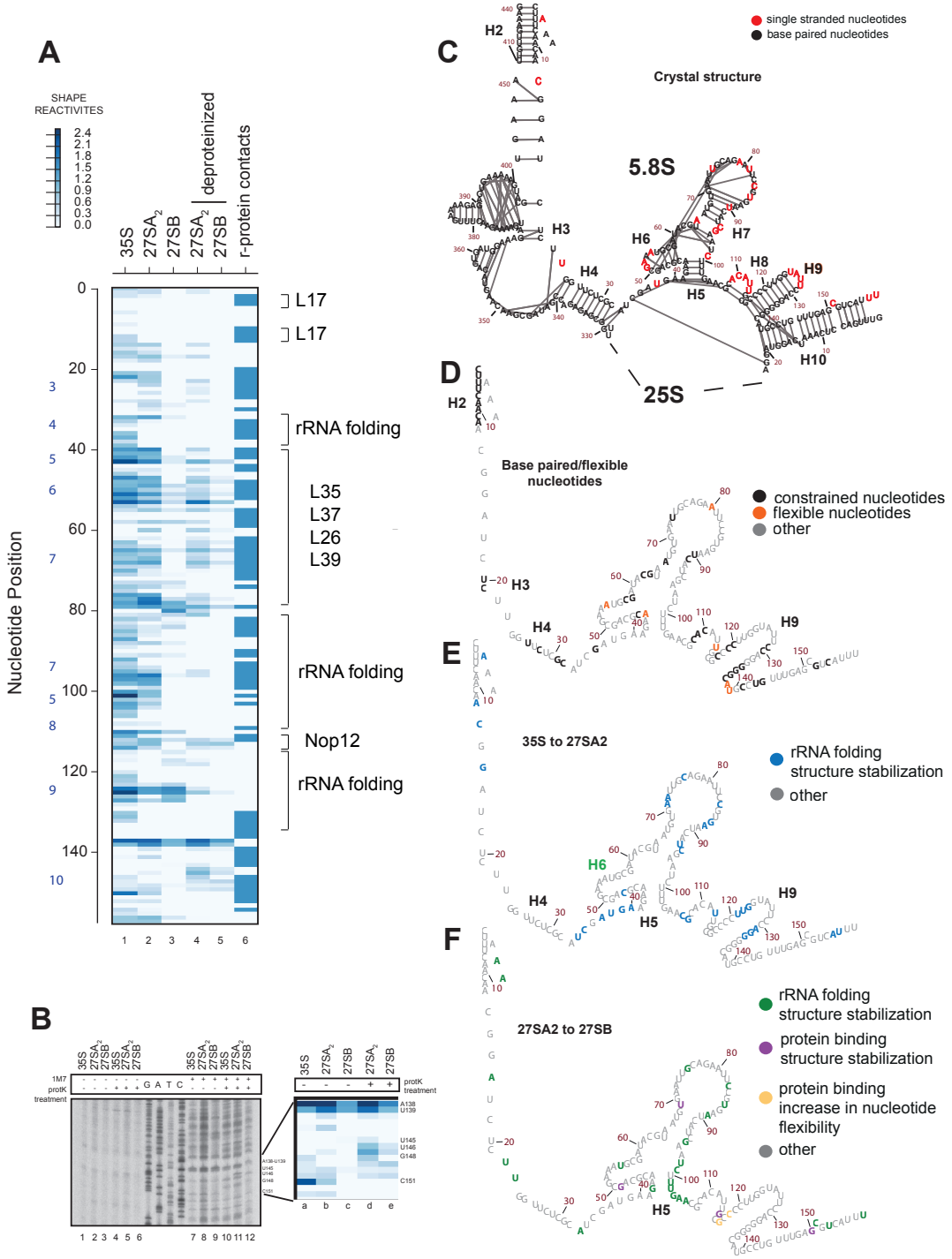


**Figure 5.7. The 5.8S region undergoes a major restructuring event during the conversion of 27SA2 to 27SB pre-rRNA**

(A) Heat map representing the ChemModSea SHAPE reactivities. Each block indicates a single nucleotide. The darker the block the higher the SHAPE reactivity. (B) Validation of the ChemModSeq data by primer extension. The gel shows the results for the 3' end region of the 5.8S. (C) Secondary structure of the 5.8S and base-pairing interactions (lines) found in the yeast 80S crystal structure. The nucleotides in red represent nucleotides that are single stranded in the mature 5.8S according to the crystal structure. (D) Nucleotides that do not undergo changes in nucleotide flexibility (black-constrained; orange-flexible) as the 35S is processed to the 27SB pre-rRNA. (E) The changes in nucleotide flexibility as the 35S is converted to 27SA2 pre-rRNA lead to the stabilization of 5.8S secondary structure (blue nucleotides). (F) The changes in nucleotide flexibility as the 27SA2 is converted to 27SB pre-rRNA fit mostly the rRNA folding category (green nucleotides), with a few nucleotides undergoing changes in flexibility dependent on the presence of proteins (purple and orange nucleotides).

To validate some of these findings I performed primer extension reactions on the same rRNA samples to determine how well the ChemModSeq data agree with the conventional gel-based approach. Figure 5.7B shows a comparison between primer extension data and ChemModSeq data for the 3' end of the 5.8S region. Lanes 1-6 in the primer extension represent the negative control while lanes 7-12 represent the 1M7 treated samples. Note that the primer extension gel contains one extra sample: the deproteinized 35S pre-rRNA (lanes 4 and 10). Here I specifically focused on a region where I could see clear differences between the samples. In general, the primer extension analyses showed a similar modification pattern as the ChemModSeq output: nucleotides A138 and U139 are detected as being modified in all samples; nucleotide U145 is modified only in the $27SA_2$ deproteinized sample (lane 11 and d); nucleotides U146 and G148 were unreactive to 1M7 in the samples from the native particles (lane 7-9 and a-c) while being modified in the 35S and $27SA_2$ deproteinized (lanes 10-11 and d) and to a smaller extent in 27SB pre-rRNA (lanes 12 and e). The modification pattern for C151 is similar in the primer extension and in the ChemModSeq data. Moreover, in the ChemModSeq data the differences in SHAPE reactivities between the rRNA precursors from the native particles was often more obvious than in the primer extension: the C151 nucleotide was more strongly modified in the 35S pre-rRNA than in the $27SA_2$ pre-rRNA, whereas I generally observed quite weak modification in the 27SB

pre-rRNA.

Figure 5.7C shows the predicted secondary structure of the 5.8S rRNA based on phylogenetic comparisons (Bernier *et al*, 2014). The red nucleotides in this structure represent the nucleotides that are single-stranded in the 80S ribosome (Bernier *et al*, 2014). Thus, although the secondary structure drawing suggests that there are many single stranded regions in the mature 5.8S rRNA, in the crystal structure most of these nucleotides are actually base-paired (indicated by black lines). Several nucleotides also form non-canonical base-pairing interactions such as base-stacking interactions. Also some of these nucleotides are involved in long-range rRNA interactions with domains of the 25S (not shown in the figure).

Figure 5.7D-F summarizes the results of my data analyses for the secondary structure of the 5.8S region. These revealed that the 5.8S rRNA undergoes an rRNA restructuring event as the $27SA_2$ is processed to 27SB pre-rRNAs. Figure 5.7D shows an overview of nucleotides that did not undergo any changes during the conversion of 35S to 27SB pre-rRNA. The orange nucleotides were consistently modified by 1M7 (and thus flexible). The black nucleotides in the secondary structure show unreactive or weakly reactive nucleotides. Helices H2 and H3 and parts of H4, which are formed by base-paring interactions between 5.8S and the 25S (fig. 5.7D) appeared to be structured in all three precursors, indicating that these base-pairing interactions form very early, possibly co-transcriptionally (fig 5.7D, black nucleotides). The same can be concluded for the base of (helix) H9 (first four base-pairing interactions). The data furthermore suggest that the region between H4 and H5 as well as parts of H9 become stabilized during the conversion of 35S to $27SA_2$ (fig 5.7E, blue nucleotides). A number of nucleotides, in particular nucleotides that form H5, appear to become more stable during the conversion of $27SA_2$ to 27SB (fig 5.7F; green nucleotides).

In conclusion, the data suggest that the 5.8S rRNA undergoes a major compaction before the 27SB pre-rRNA is formed. Moreover, in most regions, this stabilization fits the category describing rRNA folding events. The SHAPE profiles of the 27SB pre-rRNA from the native and the deproteinized samples are very similar indicating that the absence of the proteins does not destabilize the secondary structure that was formed during the synthesis of 27SB (fig 5.7A

compare lanes 3 and 5). Only a few nucleotides appeared to fit the category of protein binding events and being thus dependent on protein presence (fig 5.7F purple and yellow nucleotides). Many of these nucleotides include Nop12 binding sites.

The data do not allow me to determine what initially induced the change in rRNA structure: if this change in rRNA folding was the result of protein binding or if the rRNA stabilization promoted protein binding. What I can conclude, however, is that once the 5.8S has folded, its structure does not require proteins to maintain its secondary structure.

The conclusion that the major rRNA restructuring event takes place as the $27SA_2$ is converted to 27SB pre-rRNA (represented by a clear drop in SHAPE reactivities between the native $27SA_2$ and 27SB pre-rRNAs) is best supported when the data are presented as a heat map of SHAPE reactivities (fig 5.7A, compare lanes 2 and 3) rather than dividing the nucleotides in categories (fig 5.7E and F). This is because, when I defined the categories I only included the nucleotides for which the modification pattern in the deproteinized $27SA_2$ and 27SB pre-rRNA was the same, whereas it is clear from the heat map that the SHAPE reactivities in the deproteinized samples vary (fig 5.7A compare lanes 4 and 5). As a result, several nucleotides in the 5.8S region did not fit any of the categories. The 5.8S region is only 158 nucleotides long, which makes its manual analysis possible. However when analysing rRNA thousands of nucleotides long (27S, 25S rRNAs) using the categories described in Figure 5.5 becomes very useful, even if they represent only the main and not all the possible modification outcomes.

## 5.7. ITS2 undergoes subtle structural changes during $27SA_2$ to 27SB processing

Three ITS2 secondary structure models were proposed: the “ring” model (fig 5.8A), which was based on phylogenetic analysis (Joseph *et al*, 1999), the “hairpin” model (fig 5.8B), which was largely based on enzymatic structure probing (Yeh & Lee, 1990) and the Coleman model (fig 5.8C) (Coleman, 2015). The latter is mostly based on phylogenetic data and is essentially a hybrid of the ring and hairpin models (which we refer to as the ring-pin model). It was previously proposed that ITS2 undergoes a conformational change from an open to a more helical structure before the 27SB pre-rRNA is cleaved at site $C_2$ (Joseph *et al*, 1999; Yeh & Lee, 1990; Côté *et al*, 2002). Hence, I was very curious whether the ChemModSeq data supported such a conformational change. The heat map in Figure 5.8D represents the ChemModSeq data for the ITS2 region from the different pre-rRNA intermediates. Although we did notice several regions that seemed to undergo rRNA structural rearrangements during 27SB pre-rRNA formation, these changes were generally minor and did not support such a dramatic conformational change. These and other regions of interest (numbered I-X) are highlighted in the ChemModSeq data (fig 5.8D) and in the three ITS2 secondary structure models (fig 5.8A-C).

The ChemModSeq data indicated an increase in nucleotide flexibility as the $27SA_2$ is processed to 27SB pre-rRNA in regions numbered II (fits the rRNA fold category), V (fits the protein binding category), IV (fits the rRNA folding category), VII (fits the protein binding category) and a stabilization of the rRNA structure in the region numbered III (fits the rRNA folding event category). Only a few nucleotides underwent changes in flexibility as the 35S pre-rRNA was processed to $27SA_2$ pre-rRNA (fig 5.E, blue nucleotides). Region III is part of a stem that gradually becomes stabilized during the conversion of 35S to 27SB pre-rRNA.

The changes in nucleotide flexibility for sequences marked with V and VIII were the only ones that appeared to be dependent on the presence of proteins, becoming more flexible in the 27SB pre-rRNA. Interestingly, Nop15 and Rlp7 were shown to cross-link close to this region (fig 5.8E) (Granneman *et al*, 2011; Dembowski *et al*, 2013b). As Nop15 and Rlp7 were identified by

mass spectrometry in both the Rrp5 ($27SA_2$ pre-rRNA) particles and in the Nsa2 (27SB pre-RNA) particles it could be either that they might not be directly involved in the rRNA restructuring event, or that they could still be directly involved but no longer contact these regions in the 27SB pre-rRNA.

When comparing the ChemModSeq data to the three proposed ITS2 structural models we noticed that the ring-pin model best fits the data, but the ChemModSeq data did not provide any support for the open loop structure. Sander Granneman and myself subsequently generated an alternative ITS2 structure that best fits the chemical probing data. The ChemModSeq data suggest that regions VII and IX (35S to 27SB pre-rRNA) and region VIII (35S and $27SA_2$ but not 27SB pre-rRNA) form stable stem structures (fig 5.8D and E). Region IX was modeled as a stem in all three previous ITS2 structures (fig 5.8A and B), but regions VII and VIII form a stem only in the ring-pin model (fig 5.8C). Similarly, the ChemModSeq data suggested that the nucleotides in region VI are flexible which is also consistent with the ring-pin model (fig 5.8 compare A-C).

On the other hand, regions I and X are single stranded in the ring-pin model while the ChemModSeq data strongly suggests that these nucleotides are actually base-paired (early, probably co-transcriptionally) (fig 5.8D and E). In our alternative ITS2 structure both regions are drawn part of stems (fig 5.8E). In our structure region I base-pairs to region III, reminiscent (but not identical) to the hairpin model (fig 5.8, regions I and III, compare B and E). Interestingly, there is a difference in the timing of the stabilization of regions I and III: region I is already stabilized in the 35S pre-rRNA, while region III is stabilized gradually, as the rRNA processing pathway progresses.

What all the proposed ITS2 structures have in common is the elongated stem structure that includes regions IV-X (fig 5.8E). Most of the differences are in regions I and III. The next step is to determine, with the help of Rob van Nues, if there is any phylogenetic support for our new ITS2 model.

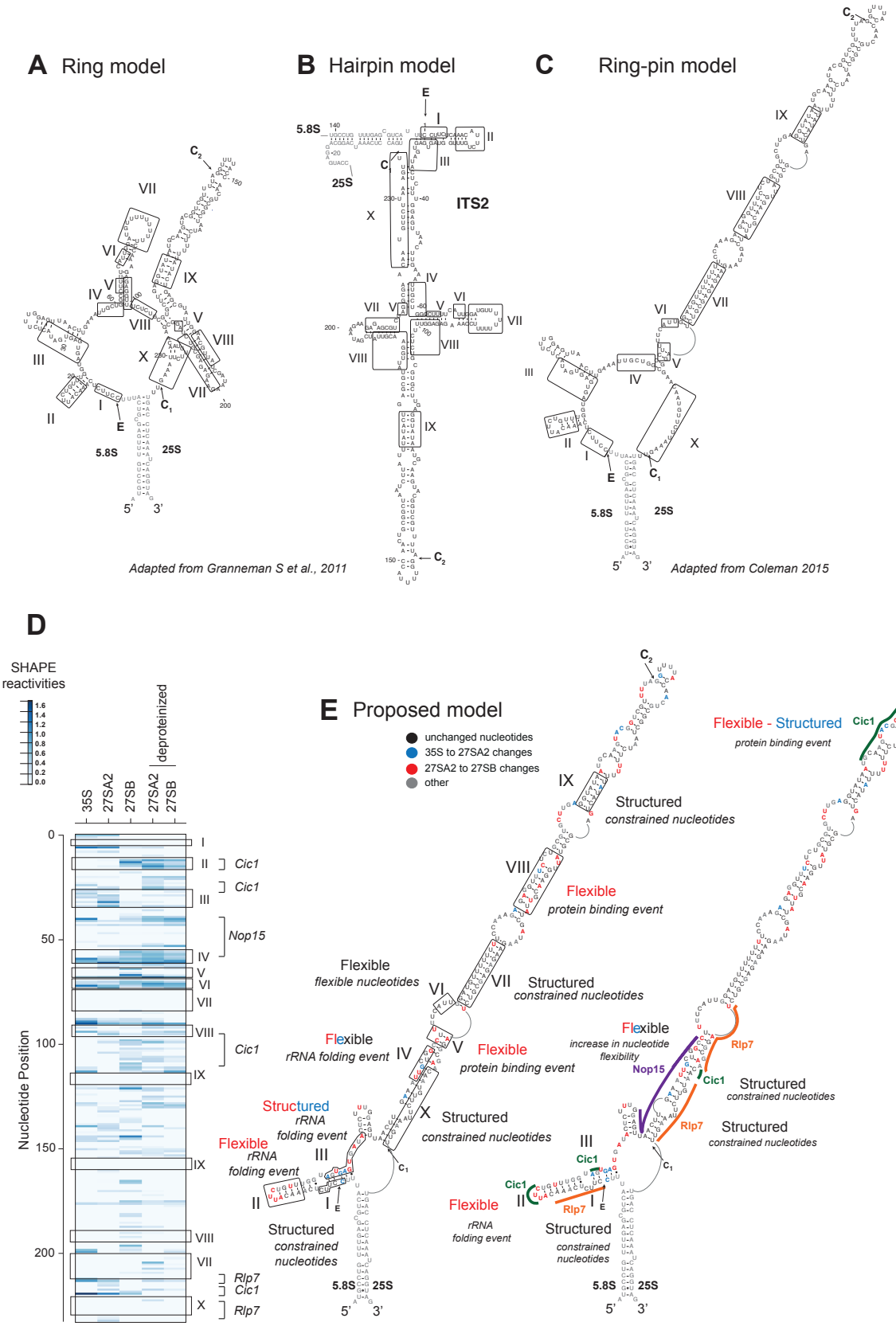

A Ring model
B Hairpin model
C Ring-pin model
Adapted from Granneman S et al., 2011
Adapted from Coleman 2015
D
SHAPE reactivities
35S
27SA2
27SB
deproteinized
Nucleotide Position
Cic1
Nop15
Rlp7
E Proposed model
unchanged nucleotides
35S to 27SA2 changes
27SA2 to 27SB changes
other
Structured constrained nucleotides
Flexible protein binding event
Flexible flexible nucleotides
Flexible rRNA folding event
Structured rRNA folding event
Flexible rRNA folding event
Structured constrained nucleotides
Flexible - Structured protein binding event
Flexible increase in nucleotide flexibility
Flexible rRNA folding event
Structured constrained nucleotides

**Figure 5.8. The ChemModSeq data indicated that ITS2 is more structured than previously thought.**
(A-C) Three proposed ITS2 models. (A) The ring structure of the ITS2, (B) The hairpin structure of ITS2, (C) The ring-pin structure of ITS2. (D) Heat map showing the changes in nucleotide flexibility in ITS2 as the 35S pre-rRNA is processed to 27SB pre-rRNA. (E) Our ITS2 secondary structure model based on the ChemModSeq data. The regions of interest are numbered I-X and are highlighted in all structures (A-E). The nucleotides that undergo changes in nucleotide flexibility between the 35S and 27SA2 pre-rRNAs (blue nucleotides), between 27SA2 and 27SB pre-rRNAs (red nucleotides) or that display no change in nucleotide flexibility (black) are highlighted in the proposed ITS2 structure (E) and so are the known binding sites for proteins that interact with ITS2.

## 5.8. The 25S region is more flexible in 27SB compared to 27SA$_2$ and 35S pre-rRNAs

Much to my surprise, I could detect a relatively low number of rRNA structural rearrangements in pre-rRNA sequences in the 25S region. It was proposed that the hierarchical stabilization of ribosomal protein binding dictates the timely folding of 25S domains (Gamalinda *et al*, 2014). Surprisingly, the ChemModSeq data does not indicate a specific hierarchy in which the structures of the 25S domains are stabilized (fig 5.9). To generate the data shown in Figure 5.9 I used the same criteria and thresholds as described above (section 5.5). For each domain around a third of the nucleotides was not taken into account as they had either low coverage or did not fit into the defined categories. Apart from domain VI for which the number of nucleotides with low coverage (less than 5 reads per nucleotide) was higher, domains I-V the nucleotides with low coverage represented less than 10% of the total number of nucleotides per domain (fig 5.9A).

Surprisingly, around half of the nucleotides analysed (a third of all nucleotides) in each 25S domain did not undergo any major rRNA structural rearrangements during pre-60S nucleolar maturation, which suggests that much of the pre-rRNA already folds co-transcriptionally. These nucleotides form stable secondary structures that do not appear to be affected by rRNA

processing events or the binding/dissociation of proteins. Most of these nucleotides did not or reacted poorly with 1M7 (fig 5.9A dark blue). The remaining third of the nucleotides in each domain undergo rRNA structural changes (fig 5.9A), as part of both rRNA (re)folding and the protein binding events as described in above in section 5.4.

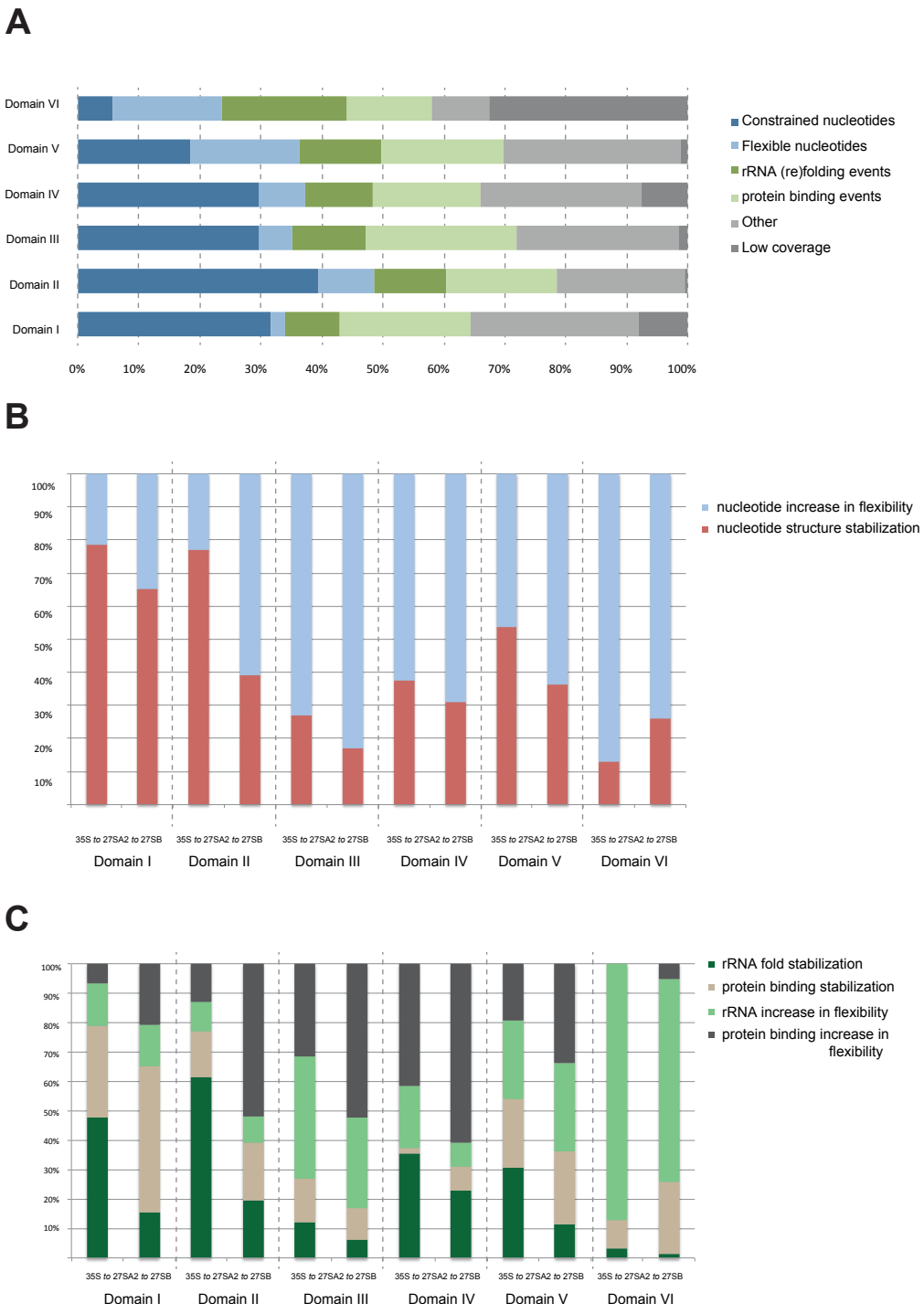


**Figure 5.9. Analysis of the ChemModSeq data for each 25S domain.**
(A) The barplot shows the distribution of nucleotides that fit to the categories we defined. (B) Overview of nucleotides that became more flexible or more constrained as the 35S was processed to the 27SB pre-rRNA. (C) Overview of nucleotides that became more flexible or more constrained when comparing the ChemModSeq data from the individual samples.

I next calculated the fraction of nucleotides that became more flexible and the fraction that became more constrained (fig 5.9B). I also assessed how many of these rRNA restructuring events fit in the rRNA folding category and how many in the protein binding category (fig 5.9C). Overall, the data suggest that during the processing of the 35S pre-rRNA to $27SA_2$ pre-rRNA domains I, II and V adopted a more stable conformation, whereas domains III, IV and VI became more flexible. As the $27SA_2$ was converted to 27SB pre-rRNA domains II-VI (but not domain I) became even more flexible. The observed changes in nucleotide flexibility (regardless of the trend) during the conversion of the $27SA_2$ to 27SB pre-rRNA were mostly dependent on the presence of proteins. Thus these structural changes events could be explained by r-proteins or assembly factors binding to 27SB pre-rRNA.

In the case of domain I, the vast majority of the nucleotides are likely already folded co-transcriptionally (fig 5.9A). The ones that did show changes in 1M7 reactivity appeared to gradually fold in more stable rRNA structures (fig 5.9B). Moreover, the ChemModSeq data revealed that the initial step of structure stabilization (35S to $27SA_2$) fits the category rRNA folding events, while the nucleotides that become more stabilized during the conversion of $27SA_2$ to 27SB require the presence of proteins to maintain their conformation (fig 5.9C).

During the conversion of 35S to $27SA_2$ about 60% of the nucleotides in domain II became more stable. Intriguingly, during the conversion of $27SA_2$ to 27SB pre-rRNA, more than half of the nucleotides actually became more flexible due to the binding of proteins (fig 5.9B and C).

Many nucleotides in domains III and IV appear to become more flexible as the 35S is processed to 27SB pre-rRNA (fig 5.9B). In the case of domain IV the increase in flexibility mostly involves protein-binding events (fig 5.9C).

Domain V is probably the most balanced of the six domains: the

percentage of stabilized nucleotides is similar to the percentage of nucleotides that had an increase in nucleotide flexibility during nucleolar pre-rRNA processing. Moreover, these changes are both due to rRNA folding events and to protein binding events (fig 5.9B and C).

Strikingly, domain VI seems to have the highest percentage of nucleotides in a flexible conformation. Unlike domains I-V, the nucleotides that fold co-transcriptionally are mostly part of flexible (rather than constrained) rRNA structures (fig 5.9A). Moreover, the increase in nucleotide flexibility represents rRNA folding events, as it is maintained in the absence of proteins (fig 5.9B).

In summary, a third of all the analysed nucleotides appear to fold co-transcriptionally, while another third undergo rRNA structural rearrangements during the conversion of 35S to 27SB pre-rRNA. The data suggest that proteins binding to 25S sequences in 27SB pre-rRNAs play an important role in maintaining the overall structure compared to earlier rRNA precursors. Moreover, the majority of these proteins have binding sites on the solvent side of the pre-60S particles.

## 5.9. The 5.8S - 25S interaction

The 5' end of 25S base pairs with the 3' end of the 5.8S to form helices H2, H4 and H10. Figure 5.10A shows the 3D structure of these helices and figure 5.10B-D highlights the changes in nucleotide flexibility as the rRNA is processed. Several important conclusions can be drawn from these results. Most of the nucleotides appear to form a constrained structure during very early stages, possibly co-transcriptionally (fig 5.10B, black nucleotides). Interestingly, these constrained nucleotides are part of the 25S in H10 and part of the 5.8S in H2 and H4. These helices are further stabilized as the 35S pre-rRNA is processed to 27SB pre-rRNA (fig 5.10B and C, red and blue nucleotides). In the case of H2 and H10 this stabilization is a result of rRNA folding events, independent of protein presence (fig 5.10.B and C, red nucleotides) while H4, H23 and H24 are stabilized as the 27SB pre-rRNA is formed, mostly as a result of protein presence (fig 5.10B and C, blue nucleotides).

Interestingly, many of the nucleotides in H2, H4, H23 and H24 which were catalogued as stabilized due to protein binding events (fig 5.10B and C, blue nucleotides) were binding sites for r-proteins Rpl26, Rpl17, Rpl4. The ChemModSeq data very nicely show at what stage of pre-rRNA maturation specific r-protein-rRNA contacts are being made (fig 5.11). For example, the r-protein Rpl26 contacts sites in H4 in the $27SA_2$ pre-rRNA and binds sites in H24 when the 27SB pre-rRNA is formed (fig 5.11B). Rpl17 seems to be more stably bound to H24 in the 27SB pre-rRNA compared to the earlier precursors (fig 5.11B). We predict that Rpl4 contacts many sites in H23 and H4 already in the 35S pre-rRNA but several additional interactions in H23 and H4 could also be detected in $27SA_2$ and 27SB pre-rRNAs (fig 5.11C). These results beautifully demonstrate the gradual stabilization of r-protein binding to pre-ribosomes during pre-rRNA processing and rRNA folding.

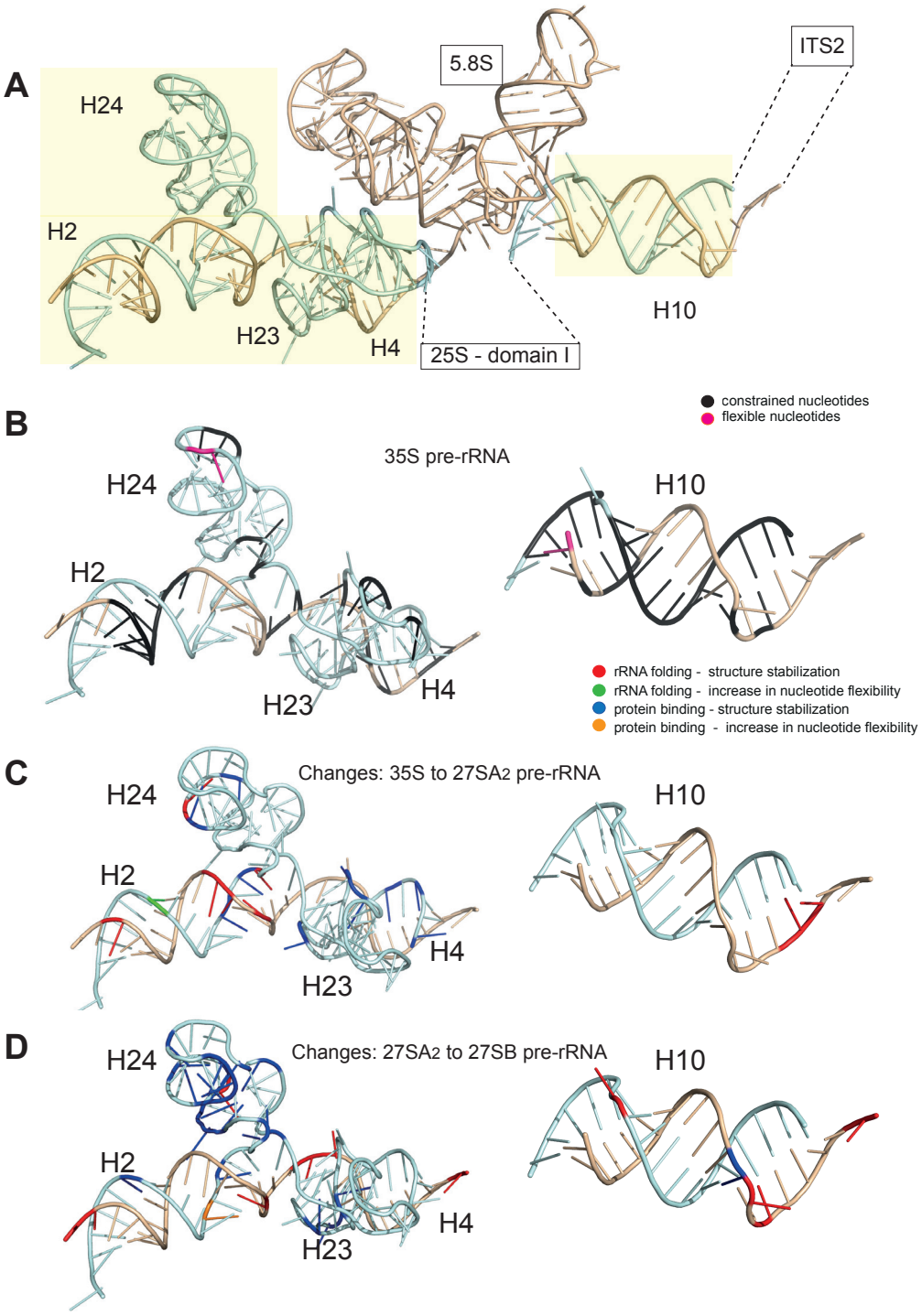



**Figure 5.10. ChemModSeq reveals changes in 5.8S and 25S regions that form base-pairing interactions.**

(A) 3D structure of the 5.8S and 25S regions that form base-pairing interactions. (B) Overview of nucleotides in the 5.8S-25S interaction sites that were constrained (black) or flexible (pink) as early as the 35S pre-rRNA. (C-D) Overview of changes in nucleotide flexibility during the conversion of 35S pre-rRNA to 27SB pre-rRNA.

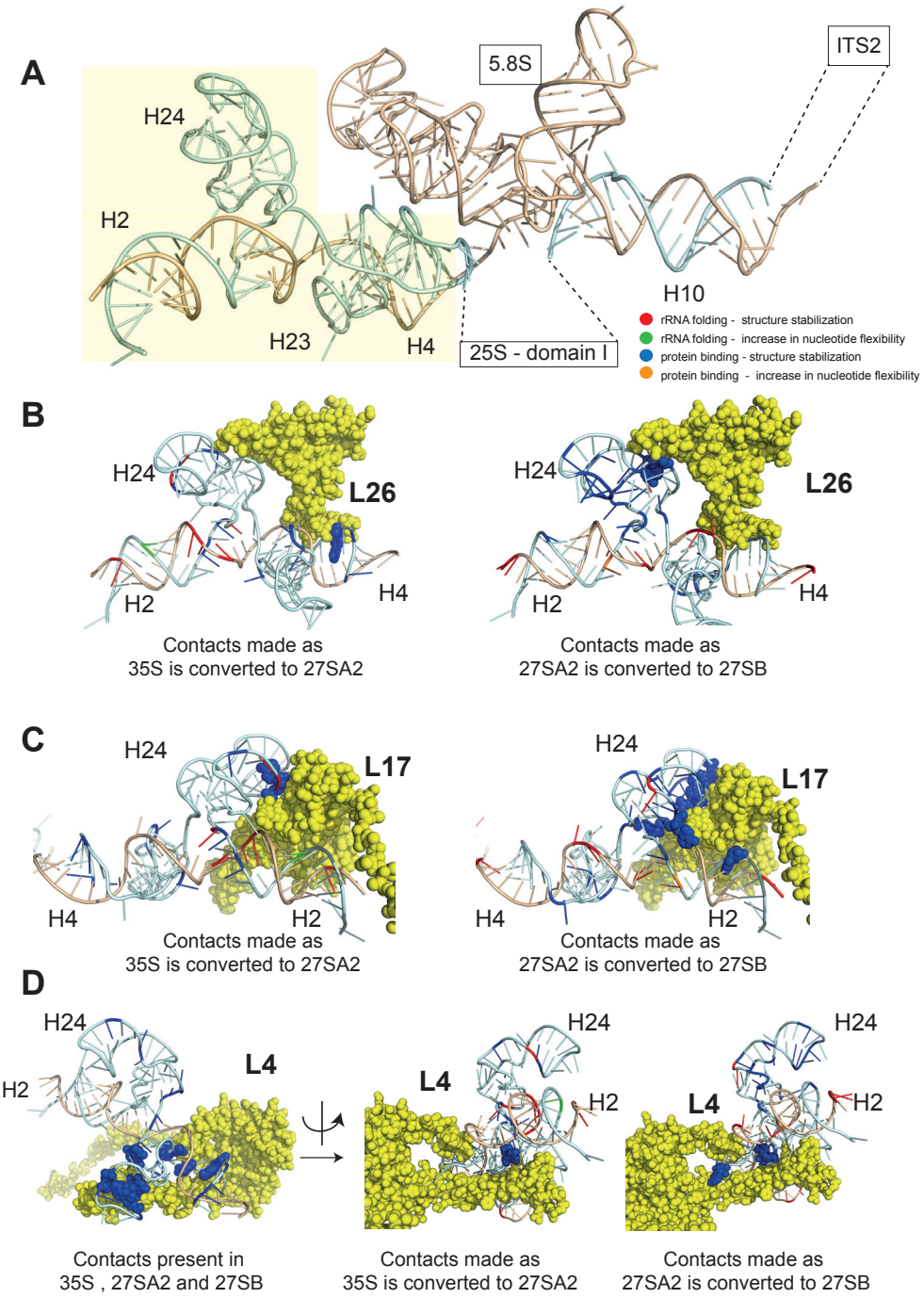



**Figure 5.11. Gradual stabilization of Rpl26, Rpl4, Rpl17 binding.**
(A) 3D structure of the 5.8S and 25S regions that form base-pairing interactions. (B-D) Gradual assembly of Rpl26, Rpl17 and Rpl4 into pre-60S pre-ribosomes. The blue spheres represent r-protein-rRNA interaction sites that the ChemModSeq data suggest form during specific stages of rRNA processing. Rpl4 makes most of its contacts in the 5.8S-25S interaction region early, in the 35S rRNA. A few other contacts become stabilized as the 35S is processed to 27SB pre-rRNA (blue spheres).

## 5.10. Examples of rRNA structural rearrangements in the 25S region

Figure 5.12 shows a number of rRNA restructuring events for H35a-H37 in domain II of the 25S. Both H35a and H37 are gradually stabilized as 35S is processed to 27SB pre-rRNA (fig 5.12A and B). The stabilization of H35a is mostly due to rRNA folding (fig 5.12A, red nucleotides), whereas stabilization of H37 involves both rRNA folding (fig 5.12B, red nucleotides) and r-protein binding (fig 5.12B, blue nucleotides) events. Both Rpl28 and Rpl29 form contacts with H37 in the 27SB pre-rRNA (fig 5.12C and E). In the 80S crystal structure, Rpl4 binds a to a single-stranded region between H35a and H37 (fig 5.12D, blue). Interestingly, this interaction decreases the flexibility of the nucleotide it binds to but increases the flexibility of a neighbouring nucleotide (fig 5.12D, orange).

Figure 5.13 provides an overview of the changes in nucleotide flexibility in helices H41-H44 from domain II and helices H94-H98 from domain VI. In the crystal structure of the mature ribosome these two regions come together and form the P-stalk structure responsible for the recruitment of GTPases involved in many stages of translation (fig 1.4A).

The changes in nucleotide flexibility for helices H41-H44 appeared to be mostly the result of protein-rRNA interactions that form once the 27SB pre-rRNA is made (fig 5.13A, blue and orange nucleotides). Blue nucleotides indicate nucleotides that become more constrained, whereas orange nucleotides indicate nucleotides that became more flexible. Interestingly, many nucleotides in H43 and H44 of the stalk region (GTPase centre) showed changes in flexibility that were dependent on protein binding (fig 5.13B, blue and orange nucleotides) that is dependent on protein interactions, although no r-proteins are known to bind to this region. Mrt4, which is a placeholder for P0, binds at the base of H43-H44. It could be that the binding of Mrt4 (and then P0) might induce the increase in flexibility for H41-H44.

Most regions in the 25S sequence that I analysed became more compact during the conversion of 35S to 27SB pre-rRNA. However, several helices showed just the opposite. A large number of nucleotides in H94-H98 actually became more flexible, which appeared to be independent of protein presence or absence (fig 5.13B, green nucleotides).

What was stinking is that binding sites for Rpl16, Rpl20 and Rpl40 localized in the H41-H44 and H94-H98 region seem to become more stabilized in the 27SB pre-rRNA (fig 5.13C, blue nucleotides). This suggests that the binding or stable incorporation of these r-proteins brings these two regions of domain II and domain VI in close proximity in the particles containing the 27SB pre-rRNA. Rpl16 bridges H98ES39 and the loop between H41 and H42, Rpl40 bridges H42 and H97 and Rpl20 bridges H42 and H98ES39. In the case of Rpl16 and Rpl20 some of the binding sites display an increase in nucleotide flexibility (fig 5.13C, orange nucleotides). In conclusion, the data suggest that parts of domain II and domain VI are brought in close proximity through r-protein binding during the conversion of $27SA_2$ to 27SB pre-rRNAs.

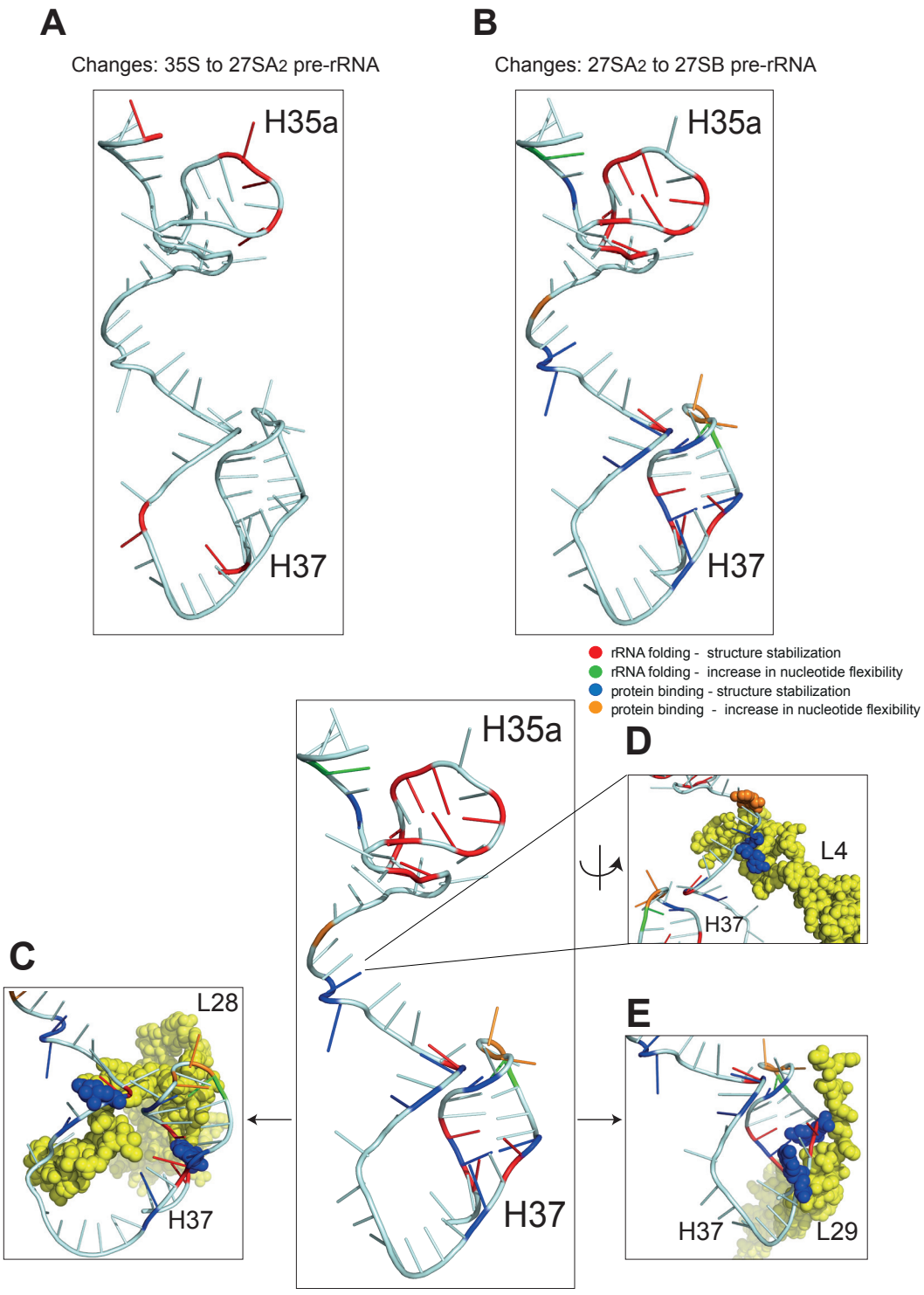



**Figure 5.12 Helixes H35a and H37 are stabilized as the 27SB pre-rRNA is formed.**
(A-B) The conversion of the 35S pre-rRNA to 27SB pre-rRNA leads to the stabilization of H35a as part of an rRNA folding events and of H37 as part of both rRNA folding and protein binding events. (C-E) The blue spheres show r-protein-rRNA contacts made by Rpl28, Rpl4 and Rpl29 to H37 that lead to the stabilization of the rRNA structure.

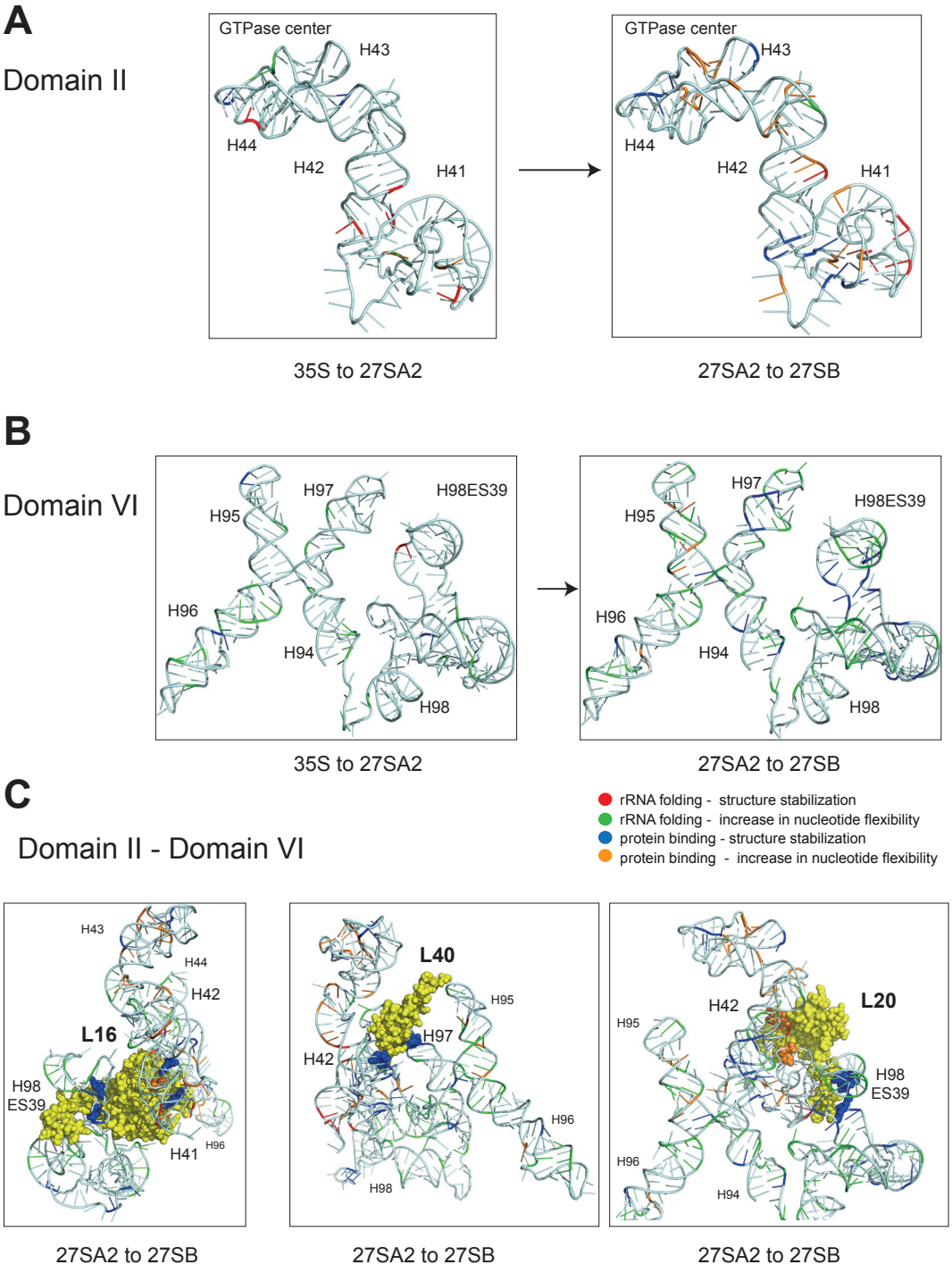



**Figure 5.13 Rpl16, Rpl40 and Rpl20 bring domain II and domain VI together during the conversion of 27SA2 to 27SB pre-rRNA.**

(A) The changes in rRNA flexibility in H41-H43 (domain II) as the 35S pre-rRNA is processed to 27SB pre-rRNA. (B) The changes in rRNA flexibility in H94-H98 (domain VI) as the 35S pre-rRNA is processed to 27SB pre-rRNA. (C) The blue spheres show r-protein-rRNA contacts made by Rpl16, Rpl40 and Rpl20 that led to the stabilization of the rRNA structure.

## 5.11. Discussion

This study provides for the first time a global overview of the rRNA structural rearrangements that take place during the early nucleolar stages of pre-60S maturation. Using specific protein baits I was able to obtain very pure and fairly homogenous populations of particles containing 35S, $27SA_2$ and 27SB pre-rRNAs for structure probing analyses. I also tried very hard to isolate high-quality 25S rRNA from free 60S subunits, but was not successful. To make a lot of free 60S subunits, I disrupted 80S ribosomes by lysing the cells in high salt and incubated them with puromycin and GTP to remove tRNAs and mRNAs. The 60S subunits were subsequently separated from the 40S subunits by sucrose centrifugation (8 hours) and the fractions containing free 60S subunits were modified with 1M7. The biggest problems I had were that the 25S rRNAs in these fractions were often slightly degraded and, for reasons unknown, the 1M7 modification in the sucrose fractions was inefficient. Therefore, I decided not to continue with this but focus on generating solid data for the pre-rRNA precursors. Regardless, the study of the pre-rRNAs generated an enormous amount of very exciting data (fig 5.14).

One general conclusion that one can draw from this study is that the 5' half of the 27SB pre-rRNA (5.8S-ITS2-domain I of 25S rRNA) mostly forms stable structures, whereas the 3' half (domains II-VI of the 25S rRNA) is generally quite flexible. Interestingly, only some of the nucleotides that are part of constrained rRNA structures in the 27SB pre-rRNA become structured as the rRNA processing progresses (most of the changes taking place during the conversion of $27SA_2$ to 27SB pre-rRNA). The vast majority of the constrained nucleotides appear to fold very early, possibly co-transcriptionally as we could not detect any significant changes in SHAPE reactivity in the three pre-rRNAs.

Unexpectedly, almost the entire domain I of the 25S appears to be completely folded in the 35S pre-rRNA as well as several helices in ITS2. This suggests that the 5.8S and 25S regions that base-pair are brought together very early. But unlike domain I the 5.8S and ITS2 regions undergo several additional rearrangements before 27SB pre-rRNA is made.

The 5.8S appears largely unfolded in the 35S pre-rRNA and seems to adopt its final structure during the conversion of $27SA_2$ to 27SB (although I

did not do a comparison with mature 5.8S yet). Interestingly, the only change in nucleotide flexibility in the deproteinized sample was observed for Nop12 cross-linking site. Nop12 is the only A3 assembly factor that was shown to cross-link to the 5.8S sequence (Granneman *et al*, 2011). Deletion of Nop12 was reported to lead to an increase in flexibility in the Nop12 cross-linking sites and in H5 and the loops adjacent to it (Talkish *et al*, 2014). However, the rRNA from deproteinized samples in this study showed a change in reactivity only for the Nop12 crosslinking site and not for H5 and its adjacent loops. This suggests that Nop12 might be needed to initiate the folding of H5 as the 27SB is formed, but it is not needed to maintain its structure.

Ribosomal proteins Rpl17, Rpl35 and Rpl37 bind to the 5.8S region. It was previously shown that depletion of these r-proteins blocks $C_2$ cleavage. It was proposed that this could be due to their role in recruiting other B factors essential for $C_2$ cleavage and also because they may play a role in the formation/maintenance of an ITS2 structure that is competent for $C_2$ cleavage (Gamalinda *et al*, 2013; Babiano & de la Cruz, 2010). The ChemModSeq data suggests that most of the structures in the 5.8S region that form during the nucleolar pre-rRNA processing steps do not require r-proteins to maintain their structure. Structure probing analyses on total rRNA extracted from cells in which these r-proteins were depleted did not reveal any major changes in rRNA flexibility (Gamalinda *et al*, 2013). This suggests that Rpl17, Rpl35, Rpl37 are not involved in the folding of the 5.8S region but may perhaps recognize the 5.8S structure that is formed in the 27SB pre-rRNA. This suggests that the effect on $C_2$ cleavage seen when these r-proteins were depleted was due to their role in recruiting the late B factors rather than folding or stabilizing the 5.8S secondary structure.

The structural changes I observed in ITS2 mainly appeared to take place once the $27SA_2$ is converted to 27SB pre-rRNA and are both due to rRNA folding events and to protein binding events (as these categories were defined in section 5.5). It was proposed that during 60S assembly ITS2 undergoes a conformational change from an open (ring structure) to a more closed (hairpin structure) conformation prior to $C_2$ cleavage (Côté *et al*, 2002). However, the ChemModSeq data suggests that if this happens, it must occur co-transcriptionally as we could not detect such major rearrangements

during the processing of 35S to 27SB. Actually, our ChemModSeq data are more consistent with the a hybrid of the “ring” and “hairpin” (ring-pin) model proposed by Coleman (Coleman, 2015), However, our data suggests that several regions that were drawn as single-stranded loops in the ring-pin model are actually structured. Using the ChemModSeq data several alternative secondary structures of ITS2 can be drawn and in this Chapter I presented one of the most likely structures. We are currently using phylogenetic data to further refine the structure of ITS2.

Much to my surprise, only two regions of ITS2 displayed somewhat dramatic changes in nucleotide flexibility that seemed to be dependent on protein binding/dissociation. Nop15, Cic1, Nop12 and Rlp7 cross-link to ITS2 and their depletion results in changes in ITS2 structure (Granneman *et al*, 2011; Talkish *et al*, 2014; Dembowski *et al*, 2013b). Unfortunately, our data do not give us any clues to whether they have a direct role in remodelling of ITS2. The main difference between our studies and the analyses performed by other groups is that we looked (mostly) at native particles whereas the other studies depleted assembly factors or r-proteins to look at structural changes (Granneman *et al*, 2011; Dembowski *et al*, 2013b). Depletion of one A3 factor results in the dissociation of many A3 proteins and affects the binding of several r-proteins. It is therefore difficult to assess whether the observed changes in pre-rRNA structure are specifically due to the absence of your protein of interest. Moreover, all of these analyses were performed on total RNA, which means that the probing data showed an average of 27SA and 27SB conformations. Clearly, more detailed analyses are required to unravel this.

These rRNA restructuring events (in the 5.8S and ITS2) appear to coincide with the release of Rrp5 and binding of the late B factors Nsa2 and Nog2 (fig 5.14). Rrp5 binds to the rRNA co-transcriptionally and it has multiple cross-linking sites in the pre-rRNA and was proposed to chaperone the proper folding of the 35S pre-rRNA as it is being transcribed (Lebaron *et al*, 2013). As a result, it is perhaps not surprising that Rrp5 needs to be released before the rRNA can undergo these conformational changes. Similarly, Nsa2 and Nog2 are the last factors to bind before the 27SB pre-rRNA is cleaved at site $C_2$ and

the pre-60S is exported (Saveanu *et al*, 2001; Lebreton *et al*, 2008). The data suggest that $C_2$ cleavage signals that 5.8S is now properly folded and that r-proteins that bind 5.8S (i.e. Rpl17, Rpl35, Rpl26 and Rpl37) are assembled before Nsa2 and Nog2 bind. Several A3 factors have been suggested to be required for folding of ITS2. It was proposed that release of these proteins would trigger these events and that the A3 factors would therefore regulate ITS2 folding steps that are necessary for optimal $C_2$ cleavage (Granneman *et al*, 2011; Dembowski *et al*, 2013b). Our results do not provide strong evidence for this. Most of the observed structural changes occurred while A3 factors were still bound to the pre-60S particles. However, although our mass-spectrometry clearly detected the presence the A3 factors in both 27SA2 and 27SB particles, the data do not allow us to assess whether all or only a few pre-60S particles contained these proteins. Therefore, it could still be that the presence of the proteins triggered the change in rRNA structure, but regardless of initial impulse, their presence is not needed later on to maintain the rRNA structure.

Regarding structural changes in the 25S region, we did not see many changes in domain I when comparing the different samples, while the nucleotide flexibility of domains II-VI generally increased. The 5.8S interacts with regions in domain I of the 25S (ITS2 proximal stem or H10, H2 and H4) and our data suggests that the proximal stem (H10) had most probably already been formed in the 35S pre-rRNA. The r-proteins Rpl8, Rpl15 and assembly factor Nop12 bind to the ITS2 proximal stem but from our data it is not clear if they also play a role in the formation of the ITS2 proximal stem. H2 and H4 are stabilized by rRNA folding events and by protein binding events that (also) involve the gradual assembly of Rpl26, Rpl17 and Rpl4. The data very nicely show which nucleotides become constrained due to protein binding. Thus, such ChemModSeq analyses can provide a very detailed overview of the gradual stabilization of r-proteins into pre-ribosomes.

Another interesting observation was that we could detect at which stage certain 25S domains were brought together by r-proteins. The data suggest that Rpl40, Rpl20 and Rpl16 bring together domain II and VI during the conversion of $27SA_2$ to 27SB pre-rRNAs. Rpl40 is thought to assemble

mainly in the cytoplasm and to be important for subunit joining, however, Rpl40 depletion was later shown to cause a slight delay in cleavage at $C_2$ and to block the pre-60S export to the cytoplasm, which indicates that Rpl40 assembles in the nucleolus (but perhaps binds only with a lower affinity than in the cytoplasm) (Fernández-Pevida *et al*, 2012). Thus, I propose that Rpl40 brings together domains II-VI in the 25S region. This would indicate that apart from the role of Rpl40 in subunit joining, it also plays a role in pre-rRNA folding steps in the nucleolus.

The functional sites of the pre-60S subunit are formed in the nucleus and cytoplasm. It is in the nucleus that the CP adopts its final conformation and the H89 is repositioned and brought closer to the PTC (Leidig *et al*, 2014; Bassler *et al*, 2014). The PET is thought to form mostly before the CP conformational change and finalized in the cytoplasm once Arx1 is released. The P stalk is also finalized in the cytoplasm.

The data from this study did not detect major changes in the rRNA secondary structure of regions forming the catalytic active sites: PET, CP, PTC or P stalk. The helices that form these regions adopt their secondary structure mostly co-transcriptionally. This suggests that the secondary structures of these regions (involved in translation) are stabilized very early (co-transcriptionally). The changes that follow along the maturation pathway to ensure their correct positioning are mostly changes in their 3D structure (such as the rotation of the 5S rRNA (Leidig *et al*, 2014; Barrio-Garcia *et al*, 2015)). As the SHAPE chemicals measure nucleotide flexibility, it is possible that many 3D movements remain undetected.

It was previously shown that the CP undergoes a major conformational change that involves the 180° rotation of the 5S rRNA and of the neighbouring helices (Leidig *et al*, 2014). Moreover, Arx1 and Rsa4 are bound to the pre-60S before the 5S rRNA rotates (Leidig *et al*, 2014), while Rix1 and Rea1 are bound to the pre-60S post 5S rRNA rotation (Barrio-Garcia *et al*, 2015). Interestingly, the mass spectrometry data detected Rix1, Rsa4 and Arx1 only in the Nsa2 particles (containing mostly 27SB pre-rRNA) but not in the Rrp5 particle (containing $27SA_2$ pre-rRNA). This suggests that the pre-60S intermediates containing Nsa2 in this study have the CP already repositioned.

It was proposed that the convex solvent side of the pre-60S folds first, followed by the PET, the flat subunit interface and the CP (Gamalinda *et al*, 2014). In agreement with this, our data suggests that most of the r-proteins induced rRNA restructuring events bind to the convex side of the pre-60S. It is possible that nucleolar compaction of the solvent side of the rRNA might play a role in generating export competent pre-60S particles. Binding to the rRNA these r-proteins shield the negative charges of the rRNA facilitating the passage through the NPC.

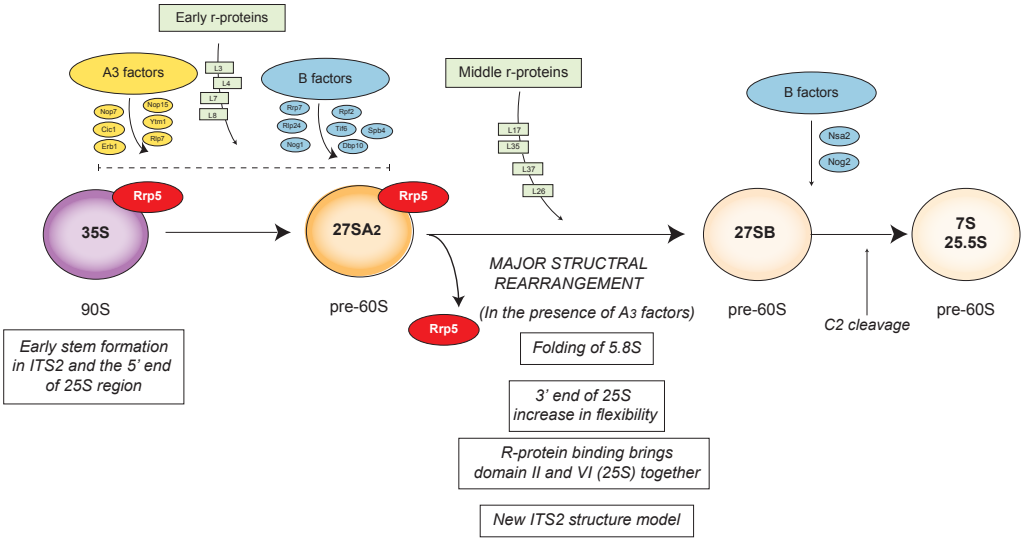



**Figure 5.14 Model of rRNA structural changes and pre-60S remodelling steps that take place during the nucleolar stages of 60S synthesis.**
Regions in ITS2 and domain I of the 25S are stabilized in the 35S pre-rRNA, probably co-transcriptionally. There is a major rRNA restructuring event as $27SA_2$ pre-rRNA is processed to 27SB pre-rRNA which includes: the folding of the 5.8S region, subtle changes in the ITS2 structure, an increase in the 3' end of the 25S, the protein mediated compaction of domain II and VI of the 25S. All these changes occur after Rrp5 dissociates and in the presence of the A3 factors.

# Concluding remarks

ChemModSeq is a very powerful tool in analysing native ribosomal assembly intermediates and the changes that occur between them at different assembly stages. Apart from the direct information on rRNA secondary structure, it can also provide, in combination with known r-protein binding sites, valuable information on the exact timing of r-protein incorporation. This study has shown that the r-proteins Rps17, Rps20 and Rps3 are initially misplaced, adopting all their contacts only in late pre-40S particles (Chapter 1). In the case of the 60S the gradual stabilization of r-proteins Rpl40, Rpl20, Rpl16 contact domains II and VI in the 27SB pre-rRNA, bringing those domains together. Moreover, the ChemModSeq data together with phylogenetic alignments can be used to refine secondary structure predictions. Using the ChemModSeq data for the 27S pre-rRNA intermediates several improved ITS2 secondary structures can be drawn and we proposed one of them in Chapter 5.

Another interesting observation is that assembly factors can chaperone the rRNA secondary structure. In the case of the pre-40S, the presence of the assembly factors Enp1, Ltv1, Rio2 and Tsr1 keeps the rRNA in a more open conformation. Their dissociation regulates the timing of the folding of the 3' major domain, allowing the stable incorporation of Rps17, Rps3 and Rps20. In the case of the pre-60S particles we noticed that the major rRNA restructuring that takes place coincides with the release of Rrp5, which was previously proposed to chaperone the 35S rRNA during transcription (Lebaron *et al*., 2013). On the other hand, the A3 factors were thought to chaperone the 5.8S-ITS2 region in the 27SB pre-rRNA (Granneman *et al.*, 2011; Dembowski *et al.*, 2013b ; Talkish *et al.*, 2014). Interestingly, the rRNA restructuring event that took place as the $27SA_2$ was converted to 27SB pre-rRNAs took place in the presence of the A3 assembly factors, suggesting that they might not play a (major) role in the nucleolar rRNA restructuring events.

Analysing the SHAPE reactivities for the 35S, $27SA_2$ and 27SB pre-rRNAs it became apparent that around a third of the nucleotides in the 27SB pre-rRNA corresponding sequence were in a constrained conformation low SHAPE reactivities) in all rRNA precursors, independent of the presence/

absence of proteins (native vs deproteinized samples). This suggests that these nucleotides fold early, most probably co-transcriptionally. Moreover, another third of the nucleotides underwent changes in rRNA flexibility, and only half of those were dependent on the presence of proteins. As a result, only a small proportion of the entire 27S pre-rRNA length is possibly chaperoned by assembly factors.

Chapter 4 focuses on the role of Rps17 in ribosome assembly. Initially interesting due to its initial misplacement in the head domain, Rps17 quickly became interesting for another reason: Rps17 has a putative nuclear export sequence in its C-terminal domain. The results obtained so far suggest that Rps17 could be a Crm1 adapter and could mediate (albeit through a redundant pathway) pre-40S export. Although I still need to directly confirm Rps17 interaction with Crm1, it is possibly the first case of a ribosomal protein being directly involved in the active pre-40S export.

To conclude, this study generated nucleotide resolution snapshots of various pre-rRNA intermediates that provide novel mechanistic insights in pre-ribosome assembly and regulation.

# References


Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E & Tollervey D (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18:** 5399–5410

Allmang C, Mitchell P, Petfalski E & Tollervey D (2000) Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res.* **28:** 1684–1691

Altvater M, Chang Y, Melnik A, Occhipinti L, Schütz S, Rothenbusch U, Picotti P & Panse VG (2012) Targeted proteomics reveals compositional dynamics of 60S pre-ribosomes after nuclear export. *Mol. Syst. Biol.* **8:** 628

Asano N, Kato K, Nakamura A, Komoda K, Tanaka I & Yao M (2015) Structural and functional analysis of the Rpf2-Rrs1 complex in ribosome biogenesis. *Nucleic Acids Res.* **43:** 4746–4757

Babiano R & de la Cruz de J (2010) Ribosomal protein L35 is required for 27SB pre-rRNA processing in Saccharomyces cerevisiae. *Nucleic Acids Res.* **38:** 5177–5192

Babiano R, Gamalinda M, Woolford JL & de la Cruz de J (2012) Saccharomyces cerevisiae ribosomal protein L26 is not essential for ribosome assembly and function. *Mol. Cell. Biol.* **32:** 3228–3241

Barrio-Garcia C, Thoms M, Flemming D, Kater L, Berninghausen O, Bassler J, Beckmann R & Hurt E (2015) Architecture of the Rix1-Rea1 checkpoint machinery during pre-60S-ribosome remodeling. *Nat. Struct. Mol. Biol.*

Bassler J, Kallas M, Pertschy B, Ulbrich C, Thoms M & Hurt E (2010) The AAA-ATPase Rea1 drives removal of biogenesis factors during multiple stages of 60S ribosome assembly. *Mol. Cell* **38:** 712–721

Bassler J, Paternoga H, Holdermann I, Thoms M, Granneman S, Barrio-Garcia C, Nyarko A, Stier G, Clark SA, Schraivogel D, Kallas M, Beckmann R, Tollervey D, Barbar E, Sinning I & Hurt E (2014) A network of assembly factors is involved in remodeling rRNA elements during preribosome maturation. *J. Cell Biol.* **207:** 481–498

Belk JP, He F & Jacobson A (1999) Overexpression of truncated Nmd3p inhibits protein synthesis in yeast. *RNA* **5:** 1055–1070

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L,

Yusupova G & Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334:** 1524–1529

Ben-Shem A, Jenner L, Yusupova G & Yusupov M (2010) Crystal structure of the eukaryotic ribosome. *Science* **330:** 1203–1209

Bernier CR, Petrov AS, Waterbury CC, Jett J, Li F, Freil LE, Xiong X, Wang L, Migliozzi BLR, Hershkovits E, Xue Y, Hsiao C, Bowman JC, Harvey SC, Grover MA, Wartell ZJ & Williams LD (2014) RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss.* **169:** 195–207

Bindewald E, Wendeler M, Legiewicz M, Bona MK, Wang Y, Pritt MJ, Le Grice SFJ & Shapiro BA (2011) Correlating SHAPE signatures with three-dimensional RNA structures. *RNA* **17:** 1688–1696

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14:** 115–132

Bradatsch B, Katahira J, Kowalinski E, Bange G, Yao W, Sekimoto T, Baumgärtel V, Boese G, Bassler J, Wild K, Peters R, Yoneda Y, Sinning I & Hurt E (2007) Arx1 functions as an unorthodox nuclear export receptor for the 60S preribosomal subunit. *Mol. Cell* **27:** 767–779

Bradatsch B, Leidig C, Granneman S, Gnädig M, Tollervey D, Böttcher B, Beckmann R & Hurt E (2012) Structure of the pre-60S ribosomal subunit with nuclear export factor Arx1 bound at the exit tunnel. *Nat. Struct. Mol. Biol.* **19:** 1234–1241

Bussiere C, Hashem Y, Arora S, Frank J & Johnson AW (2012) Integrity of the P-site is probed during maturation of the 60S ribosomal subunit. *J. Cell Biol.* **197:** 747–759

Campbell MG & Karbstein K (2011) Protein-protein interactions within late pre-40S ribosomes. *PLoS ONE* **6:** e16194

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

Cmejla R, Cmejlova J, Handrkova H, Petrak J & Pospisilova D (2007) Ribosomal protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. *Hum. Mutat.* **28:** 1178–1182

Coleman AW (2015) Nuclear rRNA transcript processing versus internal transcribed spacer secondary structure. *Trends Genet.* **31:** 157–163

Conti E & Izaurralde E (2001) Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.* **13:** 310–319

Cook A, Bono F, Jinek M & Conti E (2007) Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* **76:** 647–671

Côté CA, Greer CL & Peculis BA (2002) Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* **8:** 786–797

Danilova N & Gazda HT (2015) Ribosomopathies: how a common root can cause a tree of pathologies. *Dis Model Mech* **8:** 1013–1026

de la Cruz J, Karbstein K & Woolford JL (2015) Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu. Rev. Biochem.* **84:** 93–129

Deigan KE, Li TW, Mathews DH & Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* **106:** 97–102

DeLabre ML, Kessl J, Karamanou S & Trumpower BL (2002) RPL29 codes for a non-essential protein of the 60S ribosomal subunit in Saccharomyces cerevisiae and exhibits synthetic lethality with mutations in genes for proteins required for subunit coupling. *Biochim. Biophys. Acta* **1574:** 255–261

Dembowski JA, Kuo B & Woolford JL (2013a) Has1 regulates consecutive maturation and processing steps for assembly of 60S ribosomal subunits. *Nucleic Acids Res.* **41:** 7889–7904

Dembowski JA, Ramesh M, McManus CJ & Woolford JL (2013b) Identification of the binding site of Rlp7 on assembling 60S ribosomal subunits in Saccharomyces cerevisiae. *RNA* **19:** 1639–1647

Demoinet E, Jacquier A, Lutfalla G & Fromont-Racine M (2007) The Hsp40 chaperone Jjj1 is required for the nucleo-cytoplasmic recycling of preribosomal factors in Saccharomyces cerevisiae. *RNA* **13:** 1570–1581

Dez C, Houseley J & Tollervey D (2006) Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of Saccharomyces cerevisiae. *EMBO J.* **25:** 1534–1546

Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC & Assmann SM (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505:** 696–700

Dong X, Biswas A, Sueel KE, Jackson LK, Martinez R, Gu H & Chook YM (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* **458:** 1136–U71

Dragon F, Gallagher JEG, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlage RE, Shabanowitz J, Osheim Y, Beyer AL, Hunt DF & Baserga SJ (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417:** 967–970

Eisinger DP, Dick FA & Trumpower BL (1997) Qsr1p, a 60S ribosomal subunit protein, is required for joining of 40S and 60S subunits. *Mol. Cell. Biol.* **17:** 5136–5145

Engelsma D, Bernad R, Calafat J & Fornerod M (2004) Supraphysiological nuclear export signals bind CRM1 independently of RanGTP and arrest at Nup358. *EMBO J.* **23:** 3643–3652

Fassio CA, Schofield BJ, Seiser RM, Johnson AW & Lycan DE (2010) Dominant mutations in the late 40S biogenesis factor Ltv1 affect cytoplasmic maturation of the small ribosomal subunit in Saccharomyces cerevisiae. *Genetics* **185:** 199–209

Fatica A, Oeffinger M, Dlakić M & Tollervey D (2003) Nob1p is required for cleavage of the 3' end of 18S rRNA. *Mol. Cell. Biol.* **23:** 1798–1807

Faza MB, Chang Y, Occhipinti L, Kemmler S & Panse VG (2012) Role of Mex67-Mtr2 in the nuclear export of 40S pre-ribosomes. *PLoS Genet.* **8:** e1002915

Fernández-Pevida A, Rodríguez-Galán O, Díaz-Quintana A, Kressler D & de la Cruz J (2012) Yeast ribosomal protein L40 assembles late into precursor 60 S ribosomes and is required for their cytoplasmic maturation. *J. Biol. Chem.* **287:** 38390–38407

Ferreira-Cerca S, Kiburu I, Thomson E, LaRonde N & Hurt E (2014) Dominant Rio1 kinase/ATPase catalytic mutant induces trapping of late pre-40S biogenesis factors in 80S-like ribosomes. *Nucleic Acids Res.* **42:** 8635–8647

Ferreira-Cerca S, Pöll G, Gleizes P-E, Tschochner H & Milkereit P

(2005) Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol. Cell* **20:** 263–275

Ferreira-Cerca S, Pöll G, Kühn H, Neueder A, Jakob S, Tschochner H & Milkereit P (2007) Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol. Cell* **28:** 446–457

Ferreira-Cerca S, Sagar V, Schäfer T, Diop M, Wesseling A-M, Lu H, Chai E, Hurt E & LaRonde-LeBlanc N (2012) ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nat. Struct. Mol. Biol.* **19:** 1316–1323

Finch AJ, Hilcenko C, Basse N, Drynan LF, Goyenechea B, Menne TF, González Fernández A, Simpson P, D'Santos CS, Arends MJ, Donadieu J, Bellanné-Chantelot C, Costanzo M, Boone C, McKenzie AN, Freund SMV & Warren AJ (2011) Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. *Genes Dev.* **25:** 917–929

Fischer U, Schäuble N, Schütz S, Altvater M, Chang Y, Boulos Faza M & Panse VG (2015) A non-canonical mechanism for Crm1-export cargo complex assembly. *Elife* **4:**

Fribourg S & Conti E (2003) Structural similarity in the absence of sequence homology of the messenger RNA export factors Mtr2 and p15. *EMBO Rep.* **4:** 699–703

Fu S-C, Imai K & Horton P (2011) Prediction of leucine-rich nuclear export signal containing proteins with NESsential. *Nucleic Acids Res.* **39:**

Fung HYJ, Fu S-C, Brautigam CA & Chook YM (2015) Structural determinants of nuclear export signal orientation in binding CRM1. *Elife* **4:**

Gadal O, Strauss D, Kessl J, Trumpower B, Tollervey D & Hurt E (2001) Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol. Cell. Biol.* **21:** 3405–3415

Galani K, Nissan TA, Petfalski E, Tollervey D & Hurt E (2004) Rea1, a dynein-related nuclear AAA-ATPase, is involved in late rRNA processing and nuclear export of 60 S subunits. *J. Biol. Chem.* **279:** 55411–55418

Gamalinda M, Jakovljevic J, Babiano R, Talkish J, de la Cruz de J & Woolford JL (2013) Yeast polypeptide exit tunnel ribosomal proteins L17, L35 and L37 are necessary to recruit late-assembling factors required for

27SB pre-rRNA processing. *Nucleic Acids Res.* **41:** 1965–1983

Gamalinda M, Ohmayer U, Jakovljevic J, Kumcuoglu B, Woolford J, Mbom B, Lin L & Woolford JL (2014) A hierarchical model for assembly of eukaryotic 60S ribosomal subunit domains. *Genes Dev.* **28:** 198–210

Gasse L, Flemming D & Hurt E (2015) Coordinated Ribosomal ITS2 RNA Processing by the Las1 Complex Integrating Endonuclease, Polynucleotide Kinase, and Exonuclease Activities. *Mol. Cell* **60:** 808–815

Gazda HT, Zhong R, Long L, Niewiadomska E, Lipton JM, Ploszynska A, Zaucha JM, Vlachos A, Atsidaftos E, Viskochil DH, Niemeyer CM, Meerpohl JJ, Rokicka-Milewska R, Pospisilova D, Wiktor-Jedrzejczak W, Nathan DG, Beggs AH & Sieff CA (2004) RNA and protein evidence for haplo-insufficiency in Diamond-Blackfan anaemia patients with RPS19 mutations. *Br. J. Haematol.* **127:** 105–113

Geerlings TH, Faber AW, Bister MD, Vos JC & Raué HA (2003) 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

Geerlings TH, Vos JC & Raué HA (2000) The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5‘-->3’ exonucleases. *RNA* **6:** 1698–1703

Ghalei H, Schaub FX, Doherty JR, Noguchi Y, Roush WR, Cleveland JL, Stroupe ME & Karbstein K (2015) Hrr25/CK1ð-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *J. Cell Biol.* **208:** 745–759

Grandi P, Rybin V, Bassler J, Petfalski E, Strauß D, Marzioch M, Schäfer T, Kuster B, Tschochner H, Tollervey D, Gavin AC & Hurt E (2002) 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol. Cell* **10:** 105–115

Granneman S, Nandineni MR & Baserga SJ (2005) The putative NTPase Fap7 mediates cytoplasmic 20S pre-rRNA processing through a direct interaction with Rps14. *Mol. Cell. Biol.* **25:** 10352–10364

Granneman S, Petfalski E & Tollervey D (2011) A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO J.* **30:** 4006–4019

Granneman S, Petfalski E, Swiatkowska A & Tollervey D (2010) Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO J.* **29:** 2026–2036

Greber BJ, Boehringer D, Montellese C & Ban N (2012) Cryo-EM structures of Arx1 and maturation factors Rei1 and Jjj1 bound to the 60S ribosomal subunit. *Nat. Struct. Mol. Biol.* **19:** 1228–1233

Grohman JK, Gorelick RJ, Kottegoda S, Allbritton NL, Rein A & Weeks KM (2014) An immature retroviral RNA genome resembles a kinetically trapped intermediate state. *J. Virol.* **88:** 6061–6068

Grünwald M & Bono F (2011) Structure of Importin13-Ubc9 complex: nuclear import and release of a key regulator of sumoylation. *EMBO J.* **30:** 427–438

Guettler T & Goerlich D (2011) Ran-dependent nuclear export mediators: a structural perspective. *EMBO J.* **30:** 3457–3474

Güttler T, Madl T, Neumann P, Deichsel D, Corsini L, Monecke T, Ficner R, Sattler M & Görlich D (2010) NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. *Nat. Struct. Mol. Biol.* **17:** 1367–1376

Hackmann A, Gross T, Baierlein C & Krebber H (2011) The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits. *EMBO Rep.* **12:** 1024–1031

Hector RD, Burlacu E, Aitken S, Le Bihan T, Tuijtel M, Zaplatina A, Cook AG & Granneman S (2014) Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. *Nucleic Acids Res.,* **42:** *12138–12154.*

Hedges J, Chen Y-I, West M, Bussiere C & Johnson AW (2006) Mapping the functional domains of yeast NMD3, the nuclear export adapter for the 60 S ribosomal subunit. *J. Biol. Chem.* **281:** 36579–36587

Henras AK, Plisson-Chastang C, O'Donohue M-F, Chakraborty A & Gleizes P-E (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA* **6:** 225–242

Henry Y, Wood H, Morrissey JP, Petfalski E, Kearsey S & Tollervey D (1994) The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J.* **13:** 2452–2463

Hierlmeier T, Merl J, Sauert M, Perez-Fernandez J, Schultz P, Bruckmann A, Hamperl S, Ohmayer U, Rachel R, Jacob A, Hergert K, Deutzmann R, Griesenbeck J, Hurt E, Milkereit P, Bassler J & Tschochner H (2013) Rrp5p, Noc1p and Noc2p form a protein module which is part of early large ribosomal subunit precursors in S. cerevisiae. *Nucleic Acids Res.* **41:** 1191–1210

Ho JH, Kallstrom G & Johnson AW (2000) Nmd3p is a Crm1p-dependent adapter protein for nuclear export of the large ribosomal subunit. *J. Cell Biol.* **151:** 1057–1066

Horn DM, Mason SL & Karbstein K (2011) Rcl1 protein, a novel nuclease for 18 S ribosomal RNA production. *J. Biol. Chem.* **286:** 34082–34087

Iacovella MG, Golfieri C, Massari LF, Busnelli S, Pagliuca C, Dal Maschio M, Infantino V, Visintin R, Mechtler K, Ferreira-Cerca S & De Wulf P (2015) Rio1 promotes rDNA stability and downregulates RNA polymerase I to ensure rDNA segregation. *Nat Commun* **6:** 6643

Jakovljevic J, Ohmayer U, Gamalinda M, Talkish J, Alexander L, Linnemann J, Milkereit P & Woolford JL (2012) Ribosomal proteins L7 and L8 function in concert with six $A_3$ assembly factors to propagate assembly of domains I and II of 25S rRNA in yeast 60S ribosomal subunits. *RNA* **18:** 1805–1822

Joseph N, Krauskopf E, Vera MI & Michot B (1999) Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res.* **27:** 4533–4540

Kappel L, Loibl M, Zisser G, Klein I, Fruhmann G, Gruber C, Unterweger S, Rechberger G, Pertschy B & Bergler H (2012) Rlp24 activates the AAA-ATPase Drg1 to initiate cytoplasmic pre-60S maturation. *J. Cell Biol.* **199:** 771–782

Kemmler S, Occhipinti L, Veisu M & Panse VG (2009) Yvh1 is required for a late maturation step in the 60S biogenesis pathway. *J. Cell Biol.* **186:** 863–880

Kharde S, Calviño FR, Gumiero A, Wild K & Sinning I (2015) The structure of Rpf2-Rrs1 explains its role in ribosome biogenesis. *Nucleic Acids Res.* **43:** 7083–7095

Kutay U & Guttinger S (2005) Leucine-rich nuclear-export signals: born to be weak. *Trends Cell Biol.* **15:** 121–124

la Cour T, Kiemer L, Molgaard A, Gupta R, Skriver K & Brunak S (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.* **17:** 527–536

Lafontaine D, Delcour J, Glasser AL, Desgrès J & Vandenhaute J (1994) 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

Lafontaine DL, Preiss T & Tollervey D (1998) Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? *Mol. Cell. Biol.* **18:** 2360–2370

Lebaron S, Schneider C, van Nues RW, Swiatkowska A, Walsh D, Böttcher B, Granneman S, Watkins NJ & Tollervey D (2012) Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nat. Struct. Mol. Biol.* **19:** 744–753

Lebaron S, Segerstolpe A, French SL, Dudnakova T, de Lima Alves F, Granneman S, Rappsilber J, Beyer AL, Wieslander L & Tollervey D (2013) Rrp5 binding at multiple sites coordinates pre-rRNA processing and assembly. *Mol. Cell* **52:** 707–719

Lebreton A, Rousselle J-C, Lenormand P, Namane A, Jacquier A, Fromont-Racine M & Saveanu C (2008) 60S ribosomal subunit assembly dynamics defined by semi-quantitative mass spectrometry of purified complexes. *Nucleic Acids Res.* **36:** 4988–4999

Lebreton A, Saveanu C, Decourty L, Jacquier A & Fromont-Racine M (2006a) Nsa2 is an unstable, conserved factor required for the maturation of 27 SB pre-rRNAs. *J. Biol. Chem.* **281:** 27099–27108

Lebreton A, Saveanu C, Decourty L, Rain J-C, Jacquier A & Fromont-Racine M (2006b) A functional network involved in the recycling of nucleocytoplasmic pre-60S factors. *J. Cell Biol.* **173:** 349–360

Leidig C, Thoms M, Holdermann I, Bradatsch B, Berninghausen O, Bange G, Sinning I, Hurt E & Beckmann R (2014) 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nat Commun* **5:** 3491

Leshin JA, Heselpoth R, Belew AT & Dinman J (2011) High throughput structural analysis of yeast ribosomes using hSHAPE. *RNA Biol* **8:** 478–487

Léger-Silvestre I, Milkereit P, Ferreira-Cerca S, Saveanu C, Rousselle J-C, Choesmel V, Guinefoleau C, Gas N & Gleizes P-E (2004) The ribosomal

protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *EMBO J.* **23:** 2336–2347

Liang X-H, Liu Q & Fournier MJ (2009) Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* **15:** 1716–1728

Lo K-Y & Johnson AW (2009) Reengineering ribosome export. *Mol. Biol. Cell* **20:** 1545–1554

Lo K-Y, Li Z, Bussiere C, Bresson S, Marcotte EM & Johnson AW (2010) Defining the pathway of cytoplasmic maturation of the 60S ribosomal subunit. *Mol. Cell* **39:** 196–208

Loc'h J, Blaud M, Réty S, Lebaron S, Deschamps P, Bareille J, Jombart J, Robert-Paganin J, Delbos L, Chardon F, Zhang E, Charenton C, Tollervey D & Leulliot N (2014) RNA mimicry by the fap7 adenylate kinase in ribosome biogenesis. *PLoS Biol.* **12:** e1001860

Low JT, Garcia-Miranda P, Mouzakis KD, Gorelick RJ, Butcher SE & Weeks KM (2014) Structure and dynamics of the HIV-1 frameshift element RNA. *Biochemistry* **53:** 4282–4291

Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA & Arkin AP (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.* **108:** 11063–11068

Madru C, Lebaron S, Blaud M, Delbos L, Pipoli J, Pasmant E, Réty S & Leulliot N (2015) Chaperoning 5S RNA assembly. *Genes Dev.* **29:** 1432–1446

Matsuo Y, Granneman S, Thoms M, Manikas R-G, Tollervey D & Hurt E (2014) Coupled GTPase and remodelling ATPase activities form a checkpoint for ribosome export. *Nature* **505:** 112–116

McGinnis JL & Weeks KM (2014) Ribosome RNA assembly intermediates visualized in living cells. *Biochemistry* **53:** 3237–3247

McGinnis JL, Dunkle JA, Cate JHD & Weeks KM (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.* **134:** 6617–6624

Merino EJ, Wilkinson KA, Coughlan JL & Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl

acylation and primer extension (SHAPE). J. Am. Chem. Soc. 127: 4223–4231

Merwin JR, Bogar LB, Poggi SB, Fitch RM, Johnson AW & Lycan DE (2014) Genetic analysis of the ribosome biogenesis factor Ltv1 of Saccharomyces cerevisiae. *Genetics* **198:** 1071–1085

Meyer AE, Hung N-J, Yang P, Johnson AW & Craig EA (2007) The specialized cytosolic J-protein, Jjj1, functions in 60S ribosomal subunit biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **104:** 1558–1563

Monecke T, Guettler T, Neumann P, Dickmanns A, Goerlich D & Ficner R (2009) Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP. *Science* **324:** 1087–1091

Mortimer SA & Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129:** 4144–4145

Moy TI & Silver PA (1999) Nuclear export of the small ribosomal subunit requires the ran-GTPase cycle and certain nucleoporins. *Genes Dev.* **13:** 2118–2133

Nerurkar P, Altvater M, Gerhardy S, Schütz S, Fischer U, Weirich C & Panse VG (2015) Eukaryotic Ribosome Assembly and Nuclear Export. *Int Rev Cell Mol Biol* **319:** 107–140

Neuber A, Franke J, Wittstruck A, Schlenstedt G, Sommer T & Stade K (2008) Nuclear export receptor Xpo1/Crm1 is physically and functionally linked to the spindle pole body in budding yeast. Mol. Cell. Biol. 28: 5348–5358

Nissan TA, Galani K, Maco B, Tollervey D, Aebi U & Hurt E (2004) A pre-ribosome with a tadpole-like structure functions in ATP-dependent maturation of 60S subunits. *Mol. Cell* **15:** 295–301

Novikova IV, Hennelly SP, Tung C-S & Sanbonmatsu KY (2013) Rise of the RNA machines: exploring the structure of long non-coding RNAs. *J. Mol. Biol.* **425:** 3731–3746

Occhipinti L, Chang Y, Altvater M, Menet AM, Kemmler S & Panse VG (2013) Non-FG mediated transport of the large pre-ribosomal subunit through the nuclear pore complex by the mRNA export factor Gle2. *Nucleic Acids Res.* **41:** 8266–8279

Oeffinger M, Dlakić M & Tollervey D (2004) A pre-ribosome-associated HEAT-repeat protein is required for export of both ribosomal subunits. *Genes Dev.* **18:** 196–209

Oeffinger M, Wei KE, Rogers R, DeGrasse JA, Chait BT, Aitchison JD & Rout MP (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods* **4:** 951–956

Ohmayer U, Gamalinda M, Sauert M, Ossowski J, Pöll G, Linnemann J, Hierlmeier T, Perez-Fernandez J, Kumcuoglu B, Léger-Silvestre I, Faubladier M, Griesenbeck J, Woolford J, Tschochner H & Milkereit P (2013) Studies on the assembly characteristics of large subunit ribosomal proteins in S. cerevisae. *PLoS ONE* **8:** e68412

Oldenburg KR, Vo KT, Michaelis S & Paddon C (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* **25:** 451–452

Passos DO & Lyumkis D (2015) Single-particle cryoEM analysis at near-atomic resolution from several thousand asymmetric subunits. *J. Struct. Biol.* **192:** 235–244

Peculis BA & Greer CL (1998) The structure of the ITS2-proximal stem is required for pre-rRNA processing in yeast. *RNA* **4:** 1610–1622

Pemberton LF & Paschal BM (2005) Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6:** 187–198

Pertschy B, Saveanu C, Zisser G, Lebreton A, Tengg M, Jacquier A, Liebminger E, Nobis B, Kappel L, van der Klei I, Högenauer G, Fromont-Racine M & Bergler H (2007) Cytoplasmic recycling of 60S preribosomal factors depends on the AAA protein Drg1. *Mol. Cell. Biol.* **27:** 6581–6592

Ray P, Basu U, Ray A, Majumdar R, Deng H & Maitra U (2008) The Saccharomyces cerevisiae 60 S ribosome biogenesis factor Tif6p is regulated by Hrr25p-mediated phosphorylation. *J. Biol. Chem.* **283:** 9681–9691

Ribbeck K & Görlich D (2002) The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J.* **21:** 2664–2671

Rice GM, Leonard CW & Weeks KM (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* **20:** 846–854

Robledo S, Idol RA, Crimmins DL, Ladenson JH, Mason PJ & Bessler M (2008) The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA* **14:** 1918–1929

Rodríguez-Mateos M, Abia D, García-Gómez JJ, Morreale A, de la Cruz de J, Santos C, Remacha M & Ballesta JPG (2009) The amino terminal domain from Mrt4 protein can functionally replace the RNA binding domain of the ribosomal P0 protein. *Nucleic Acids Res.* **37:** 3514–3521

Rosado IV, Kressler D & de la Cruz J (2007) Functional analysis of Saccharomyces cerevisiae ribosomal protein Rpl3p in ribosome synthesis. *Nucleic Acids Res.* **35:** 4203–4213

Sahasranaman A, Dembowski J, Strahler J, Andrews P, Maddock J & Woolford JL (2011) Assembly of Saccharomyces cerevisiae 60S ribosomal subunits: role of factors required for 27S pre-rRNA processing. *EMBO J.* **30:** 4020–4032

Saveanu C, Bienvenu D, Namane A, Gleizes PE, Gas N, Jacquier A & Fromont-Racine M (2001) Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. *EMBO J.* **20:** 6475–6484

Saveanu C, Namane A, Gleizes P-E, Lebreton A, Rousselle J-C, Noaillac-Depeyre J, Gas N, Jacquier A & Fromont-Racine M (2003) Sequential protein association with nascent 60S ribosomal particles. *Mol. Cell. Biol.* **23:** 4449–4460

Schäfer T, Maco B, Petfalski E, Tollervey D, Böttcher B, Aebi U & Hurt E (2006) Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441:** 651–655

Schäfer T, Strauß D, Petfalski E, Tollervey D & Hurt E (2003) The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J.* **22:** 1370–1380

Schmitt ME & Clayton DA (1993) Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **13:** 7935–7941

Seiser RM, Sundberg AE, Wollam BJ, Zobel-Thropp P, Baldwin K, Spector MD & Lycan DE (2006) Ltv1 is required for efficient nuclear export of the ribosomal small subunit in Saccharomyces cerevisiae. *Genetics* **174:** 679–691

Senger B, Lafontaine DL, Graindorge JS, Gadal O, Camasses A, Sanni A, Garnier JM, Breitenbach M, Hurt E & Fasiolo F (2001) The nucle(ol) ar Tif6p and Efl1p are required for a late cytoplasmic step of ribosome synthesis. *Mol. Cell* **8:** 1363–1373

Sengupta J, Bussiere C, Pallesen J, West M, Johnson AW & Frank J (2010) Characterization of the nuclear export adaptor protein Nmd3 in association with the 60S ribosomal subunit. *J. Cell Biol.* **189:** 1079–1086

Siegfried NA, Busan S, Rice GM, Nelson JAE & Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11:** 959–965

Smola MJ, Calabrese JM & Weeks KM (2015) Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54:** 6867–6875

Song M-J, Yoo E-H, Lee K-O, Kim G-N, Kim H-J, Kim S-Y & Kim S-H (2010) A novel initiation codon mutation in the ribosomal protein S17 gene (RPS17) in a patient with Diamond-Blackfan anemia. *Pediatr Blood Cancer* **54:** 629–631

Soudet J, Gélugne J-P, Belhabich-Baumas K, Caizergues-Ferrer M & Mougin A (2010) Immature small ribosomal subunits can engage in translation initiation in Saccharomyces cerevisiae. *EMBO J.* **29:** 80–92

Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G & Frank J (2001) Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. *Cell* **107:** 373–386

Spitale RC, Flynn RA, Torre EA, Kool ET & Chang HY (2014) RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip Rev RNA* **5:** 867–881

Steen K-A, Malhotra A & Weeks KM (2010) Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease. *J. Am. Chem. Soc.* **132:** 9940–9943

Steen K-A, Siegfried NA & Weeks KM (2011) Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease (RNase-detected SHAPE) for direct analysis of covalent adducts and of nucleotide flexibility in RNA. *Nat Protoc* **6:** 1683–1694

Strässer K, Bassler J & Hurt E (2000) Binding of the Mex67p/Mtr2p

heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. *J. Cell Biol.* **150:** 695–706

Strunk BS, Loucks CR, Su M, Vashisth H, Cheng S, Schilling J, Brooks CL, Karbstein K & Skiniotis G (2011) Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science* **333:** 1449–1453

Strunk BS, Novak MN, Young CL & Karbstein K (2012) A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* **150:** 111–121

Stutz F, Bachi A, Doerks T, Braun IC, Séraphin B, Wilm M, Bork P & Izaurralde E (2000) REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6:** 638–650

Talkish J, Campbell IW, Sahasranaman A, Jakovljevic J & Woolford JL (2014) Ribosome assembly factors Pwp1 and Nop12 are important for folding of 5.8S rRNA during ribosome biogenesis in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **34:** 1863–1877

Talkish J, Zhang J, Jakovljevic J, Horsey EW & Woolford JL (2012) Hierarchical recruitment into nascent ribosomes of assembly factors required for 27SB pre-rRNA processing in Saccharomyces cerevisiae. *Nucleic Acids Res.* **40:** 8646–8661

Taura T, Krebber H & Silver PA (1998) A member of the Ran-binding protein family, Yrb2p, is involved in nuclear protein export. *Proc. Natl. Acad. Sci. U.S.A.* **95:** 7427–7432

Thomas F & Kutay U (2003) Biogenesis and nuclear export of ribosomal subunits in higher eukaryotes depend on the CRM1 export pathway. *J. Cell. Sci.* **116:** 2409–2419

Thomson E & Tollervey D (2010) The final step in 5.8S rRNA processing is cytoplasmic in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **30:** 976–984

Thomson E, Ferreira-Cerca S & Hurt E (2013) Eukaryotic ribosome biogenesis at a glance. *J. Cell. Sci.* **126:** 4815–4821

Torchet C & Hermann-Le Denmat S (2000) Bypassing the rRNA processing endonucleolytic cleavage at site A2 in Saccharomyces cerevisiae. *RNA* **6:** 1498–1508

Turowski TW & Tollervey D (2014) Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley Interdiscip Rev RNA*

Turowski TW, Lebaron S, Zhang E, Peil L, Dudnakova T, Petfalski E, Granneman S, Rappsilber J & Tollervey D (2014) Rio1 mediates ATP-dependent final maturation of 40S ribosomal subunits. *Nucleic Acids Res.* **42:** 12189–12199

Ulbrich C, Diepholz M, Bassler J, Kressler D, Pertschy B, Galani K, Böttcher B & Hurt E (2009) Mechanochemical removal of ribosome biogenesis factors from nascent 60S ribosomal subunits. *Cell* **138:** 911–922

van den Elzen AMG, Schuller A, Green R & Séraphin B (2014) Dom34-Hbs1 mediated dissociation of inactive 80S ribosomes promotes restart of translation after stress. *EMBO J.* **33:** 265–276

Vanrobays E, Gélugne J-P, Caizergues-Ferrer M & Lafontaine DLJ (2004) Dim2p, a KH-domain protein required for small ribosomal subunit synthesis. *RNA* **10:** 645–656

Vanrobays E, Gélugne J-P, Gleizes P-E & Caizergues-Ferrer M (2003) Late cytoplasmic maturation of the small ribosomal subunit requires RIO proteins in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **23:** 2083–2095

Vanrobays E, Gleizes PE, Bousquet-Antonelli C, Noaillac-Depeyre J, Caizergues-Ferrer M & Gélugne JP (2001) Processing of 20S pre-rRNA to 18S ribosomal RNA in yeast requires Rrp10p, an essential non-ribosomal cytoplasmic protein. *EMBO J.* **20:** 4204–4213

Vanrobays E, Leplus A, Osheim YN, Beyer AL, Wacheul L & Lafontaine DLJ (2008) 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

Venema J & Tollervey D (1996) RRP5 is required for formation of both 18S and 5.8S rRNA in yeast. *EMBO J.* **15:** 5701–5714

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL & Weeks KM (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460:** 711–716

Weeks KM & Mauger DM (2011) Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* **44:** 1280–1291

Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews

DH, Giddings MC & Weeks KM (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **6:** e96

Wilkinson KA, Vasa SM, Deigan KE, Mortimer SA, Giddings MC & Weeks KM (2009) Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* **15:** 1314–1321

Woolls HA, Lamanna AC & Karbstein K (2011) Roles of Dim2 in ribosome assembly. *J. Biol. Chem.* **286:** 2578–2586

Xu D, Farmer A, Collett G, Grishin NV & Chook YM (2012) Sequence and structural analyses of nuclear export signals in the NESdb database. *Mol. Biol. Cell* **23:** 3677–3693

Yao W, Lutzmann M & Hurt E (2008) A versatile interaction platform on the Mex67-Mtr2 receptor creates an overlap between mRNA and ribosome export. *EMBO J.* **27:** 6–16

Yao W, Roser D, Köhler A, Bradatsch B, Bassler J & Hurt E (2007) Nuclear export of ribosomal 60S subunits by the general mRNA export receptor Mex67-Mtr2. *Mol. Cell* **26:** 51–62

Yao Y, Demoinet E, Saveanu C, Lenormand P, Jacquier A & Fromont-Racine M (2010) Ecm1 is a new pre-ribosomal factor involved in pre-60S particle export. *RNA* **16:** 1007–1017

Yeh LC & Lee JC (1990) Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from Saccharomyces cerevisiae. *J. Mol. Biol.* **211:** 699–712

Zemp I, Wild T, O'Donohue M-F, Wandrey F, Widmann B, Gleizes P-E & Kutay U (2009) Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *J. Cell Biol.* **185:** 1167–1180

Zhang J, Harnpicharnchai P, Jakovljevic J, Tang L, Guo Y, Oeffinger M, Rout MP, Hiley SL, Hughes T & Woolford JL (2007) Assembly factors Rpf2 and Rrs1 recruit 5S rRNA and ribosomal proteins rpL5 and rpL11 into nascent ribosomes. *Genes Dev.* **21:** 2580–2592

Nucleic Acids Research Advance Access published September 8, 2014

*Nucleic Acids Research, 2014* **1**
*doi: 10.1093/nar/gku815*

# Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution

**Ralph D. Hector[1,†], Elena Burlacu[1,2,†], Stuart Aitken[3], Thierry Le Bihan[1], Maarten Tuijtel[1], Alina Zaplatina[1], Atlanta G. Cook[2] and Sander Granneman[1,*]**

[1]Centre for Synthetic and Systems Biology (SynthSys), University of Edinburgh, Edinburgh, EH9 3JD, UK, [2]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, EH9 3JR, UK and [3]MRC Human Genetics Unit, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK

Received May 27, 2014; Revised August 25, 2014; Accepted August 28, 2014

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

## ABSTRACT

**Ribosome assembly in eukaryotes involves the activity of hundreds of assembly factors that direct the hierarchical assembly of ribosomal proteins and numerous ribosomal RNA folding steps. However, detailed insights into the function of assembly factors and ribosomal RNA folding events are lacking. To address this, we have developed ChemModSeq, a method that combines structure probing, high-throughput sequencing and statistical modeling, to quantitatively measure RNA structural rearrangements during the assembly of macromolecular complexes. By applying ChemModSeq to purified 40S assembly intermediates we obtained nucleotide-resolution maps of ribosomal RNA flexibility revealing structurally distinct assembly intermediates and mechanistic insights into assembly dynamics not readily observed in cryo-electron microscopy reconstructions. We show that RNA restructuring events coincide with the release of assembly factors and predict that completion of the head domain is required before the Rio1 kinase enters the assembly pathway. Collectively, our results suggest that 40S assembly factors regulate the timely incorporation of ribosomal proteins by delaying specific folding steps in the 3′ major domain of the 20S pre-ribosomal RNA.**

## INTRODUCTION

Ribosome synthesis in eukaryotes is an incredibly complex process that requires around 200 ribosome assembly factors to facilitate the modification, folding and processing of rRNA precursors (pre-rRNA) and their ordered assembly with r-proteins (1). In the yeast *Saccharomyces cerevisiae*, ribosome assembly starts in the nucleolus, where a 35S rRNA precursor is cleaved at sites $A_0$, $A_1$ and $A_2$ within a large, 90S-sized complex. This results in the formation of a pre-60S particle that contains the $27SA_2$ pre-rRNA and a pre-40S particle that contains a 3′ extended 18S molecule called the 20S pre-rRNA. The pre-40S complexes are rapidly exported to the cytoplasm where cleavage at site D generates the mature 18S rRNA. During maturation, several nucleotides in the 20S pre-rRNA are enzymatically modified. Shortly after export to the cytoplasm, Dim1 methylates two adenosines near the D-site (2). Just before D-site cleavage, an aminocarboxypropyl (acp) group is added to a hypermodified uridine at position 1191 in the P-site of the 18S rRNA decoding center (3). This hypermodification is important for efficient translation and D-site cleavage (4).

Figure 1A shows a schematic representation of our current knowledge about the cytoplasmic steps of yeast 40S maturation. As in bacteria, small subunit assembly in yeast involves many restructuring and/or remodeling steps; however, these events are still poorly understood. The earliest pre-40S complexes contain an almost complete set of r-proteins, a handful of assembly factors and the endonuclease Nob1 that performs the final cleavage event at site D (Figure 1A). The majority of assembly factors associate with pre-40S complexes in the nucleus and, upon entry into the cytoplasm, are believed to progressively dissociate from intermediates. However, it is not entirely clear in which order these factors are released (Figure 1A). CRAC UV cross-linking and cryo-electron microscopy (cryo-EM) experiments have localized binding sites for many assembly factors (5,6). Several assembly factors cross-linked near r-protein binding sites and it was proposed that this could interfere with stable binding of r-proteins to pre-40S (5). Their presence is also incompatible with binding of mRNA, tRNAs and translation initiation factors, indicating that they block premature translation initiation (5,6).

*To whom correspondence should be addressed. Tel: +44 131 651 9082; Fax: +44 131 651 9068; Email: sgrannem@staffmail.ed.ac.uk
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2014. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

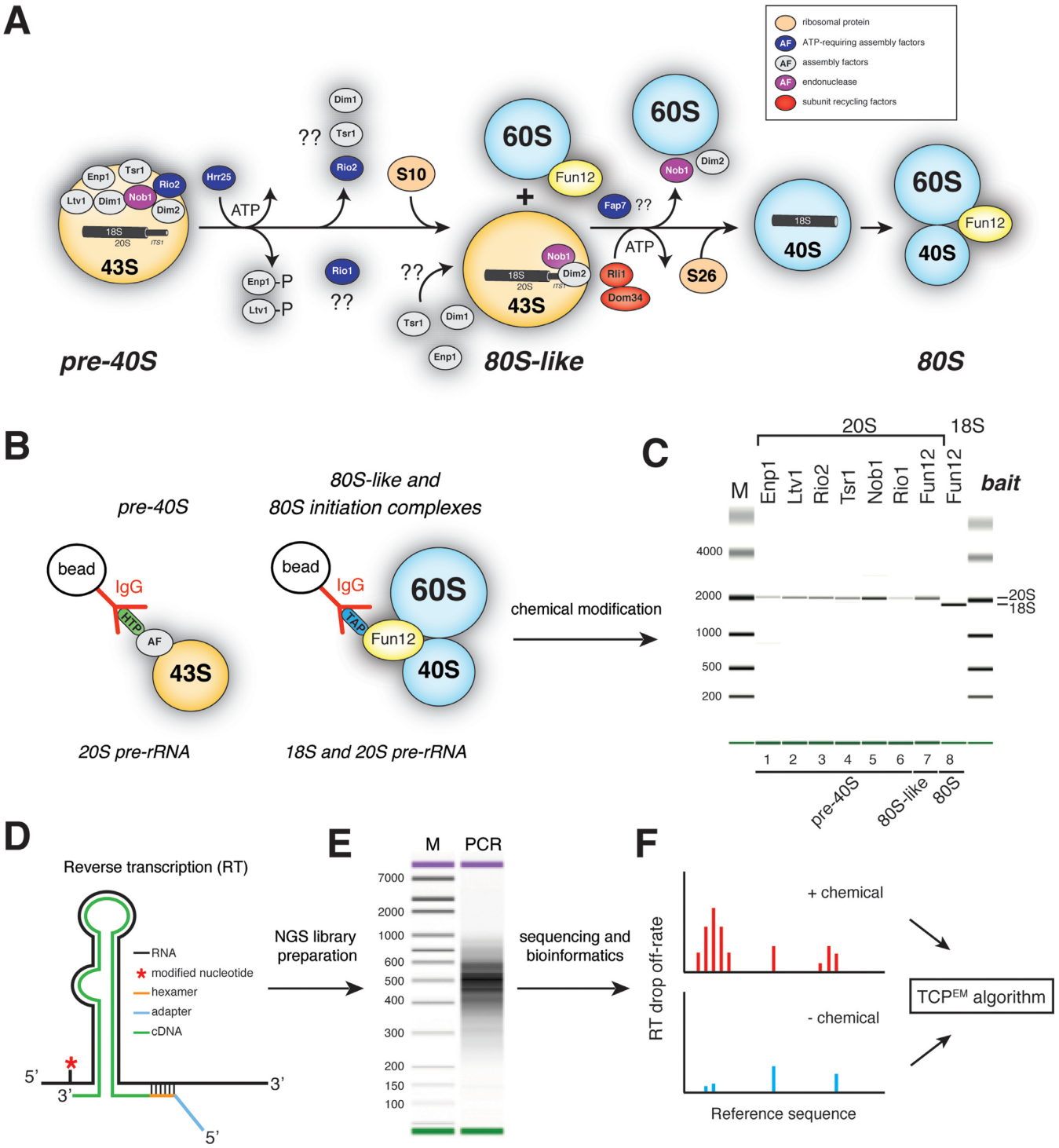



**Figure 1.** 40S subunit maturation in *Saccharomyces cerevisiae*. (**A**) Schematic representation of the cytoplasmic 40S synthesis pathway. (**B** and **C**) Pre-40S particles at different stages of assembly are purified using strains expressing TAP-or HTP-tagged assembly factors. Particles are chemically modified (B), RNA is extracted and gel purified 20S and 18S rRNAs are analyzed on a Bioanalyzer (C). (**D** and **E**) cDNA libraries, generated by random priming, were PCR amplified and (**E**) sequenced on a HiSeq. (**F**) Chemically modified samples were compared to solvent treated samples using a two-channel Poisson expectation maximization algorithm (TCMP$^{EM}$) to identify nucleotides that had the highest likelihood of being modified.

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

Hrr25 phosphorylation of Ltv1 and Enp1 was shown to trigger a cascade of events that results in their release from pre-40S complexes (7), and this is likely the first cytoplasmic remodeling step (8) (Figure 1A). Whether Rio1 interacts with pre-40S complexes in the nucleus or the cytoplasm is still under debate. Although Rio1 is predominantly cytoplasmic, it shuttles between the nucleus and the cytoplasm (9). Translation initiation factor eIF5B/Fun12 stimulates processing at site D by promoting the joining of a pre-40S particle to a 60S subunit, forming an 80S-like complex (10–12) (Figure 1A). In the current model this step ensures that Nob1 only cleaves in pre-40S particles that are able to adopt a translation-competent conformation. Precisely when the other assembly factors dissociate is not fully understood. Yeast Enp1, Tsr1, Nob1, Dim1 and Dim2 co-sedimented with stalled 80S-like complexes in density gradients (10), suggesting that these proteins remain associated with pre-40S complexes until the very final stages. In contrast, work in human cells indicated that only hNob1 and hDim2 are present in late particles (13).

Despite tremendous progress over the years, there is still much to be learned about the function of ribosome assembly factors, in particular how they influence rRNA folding steps. Using cryo-electron microscopy, the Böttcher and Hurt laboratories demonstrated a role for Hrr25 in the formation of the 18S rRNA 'beak' structure, which is formed by a protruding helix in the 18S rRNA (7). CRAC UV cross-linking experiments have also indicated that ribosome assembly factors could contribute to the formation of the functional center of the ribosome (5). Unfortunately, high-resolution insights into rRNA folding events are lacking, making it difficult to obtain mechanistic insights into how assembly factors regulate this process. Atomic resolution structures of assembly intermediates would undoubtedly provide the most valuable insights into ribosome biogenesis, however, the instability and heterogeneity of purified ribosome assembly intermediates makes high-resolution structural studies extremely challenging.

As an alternative to direct structural approaches, we decided to establish protocols that would enable us to measure the flexibility of pre-rRNAs in ribosome assembly intermediates. This would provide high-resolution insights into rRNA restructuring events. SHAPE (Selective 2′-hydroxyl acylation analyzed by primer extension) is currently the gold standard method for mapping dynamic regions within RNA. SHAPE chemicals generically react with 2′-OH of all four nucleotides in single-stranded and/or flexible regions (14). Sites of 2′-O-adduct formation are identified as stops in reverse transcription (RT) reactions. SHAPE analysis of rRNA in crystallized bacterial ribosomes showed a strong correlation between nucleotide flexibility and SHAPE reactivity, independent of solvent or molecular accessibility (15). Because SHAPE provides a read-out of RNA dynamics (rather than solvent accessibility or protein association), it is ideally suited to measure RNA conformational changes during the assembly of macromolecular complexes, such as pre-ribosomes. We developed a high-throughput method called ChemModSeq that combines SHAPE with next-generation sequencing and statistical modeling. This quantitative approach allowed us to rapidly generate an overview of RNA restructuring events during 40S synthesis in yeast. Our results indicate that assembly factors regulate the timely assembly of ribosomal proteins by temporarily maintaining more flexible pre-rRNA conformations. We predict that this activity is required to prevent premature interactions that could lead to kinetic folding traps. We show that late restructuring in the 3′ major domain coincides with the release of assembly factors Ltv1, Enp1, Rio2 and Tsr1 in the cytoplasm. We propose that once this has been completed Rio1 enters the assembly pathway and 80S-like complexes can be formed.

Collectively, our results reveal new mechanistic insights into how ribosome assembly factors regulate the formation of 40S subunits in a eukaryote.

## MATERIALS AND METHODS

### Yeast strains and media

*Saccharomyces cerevisiae* strain BY4741 (MATa; *his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*) was used as the parental strain (16). An overview of the strains used in this study is provided in Supplementary Table S3. The HTP (HIS6-TEV-2xProtA) and TAP carboxyl-tagged strains (Calmodulin binding peptide-TEV-2xProtA) were generated by polymerase chain reaction as described (17) (see Supplementary Table S4). Strains were grown in YPD (1% yeast extract, 2% peptone, 2% dextrose) or YPG/R (YP with 2% galactose and 2% raffinose) at 30°C. For the Nob1 and Fap7-depletion experiments, cells were grown in YPG/R to and $OD_{600}$ of ∼0.5, shifted to YPD and grown for four hours to an $OD_{600}$ of ∼1.0.

### Immunoprecipitation and chemical modification of rRNA

Immunoprecipitations performed with HTP-and TAP-tagged proteins were performed with TNM150 buffer (50 mM Tris–HCl pH 7.8, 1.5 mM $MgCl_2$, 150 mM NaCl, 0.1% NP-40 and 5 mM β-mercaptoethanol), as previously described (18), using 1–5 g of cells as starting material. To immunoprecipitate FLAG-tagged r-proteins, TMN150 lysates prepared from 10 $OD_{600}$ units of cells were incubated with 30 μl of anti-FLAG M2 magnetic beads (Sigma) for 1 h at 4°C. After washing the beads three times with 1 ml TMN150, RNA was extracted as described below. For *in vitro* modification reactions, pre-40S particles immobilized on 250 μl of IgG beads were resuspended in 200 μl of TMN150 buffer (without β-mercaptoethanol; final volume ∼400 μl). One-methyl-7-nitroisatoic anhydride (1M7; 15 mM final concentration) or DMSO (Dimethyl sulfoxide) (control; 5% final) was subsequently added to the beads and incubated for 3 min at room temperature. *In vitro* dimethyl sulfate (DMS) modification reactions were performed as previously described (15,19,20). For *in vivo* modification reactions with 2-methylnicotinic acid imidazolide (NAI), 1 l of cells was grown in YPD to an $OD_{600}$ of ∼1.0. Cells were harvested, washed with phosphate-buffered saline (PBS) and resuspended in 4 ml of PBS per gram of cells. NAI was added to a final concentration of 100 mM and incubated for exactly 10 min at room temperature. Cells were washed with 50 ml of ice-cold PBS and pellets were snap frozen in liquid nitrogen. NAI was synthesized as previously described (21). RNA was extracted (see below).

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

20S and 18S pre-rRNAs were gel purified as previously described (19).

For analyses on *in vitro* refolded rRNA, gel purified 18S rRNA was refolded in TMN150 by heating the RNA at 65°C for 10 min and allowed to gradually cool to room temperature in a heating block. 1M7 was added to a final concentration of 15 mM (5% DMSO final) and incubated at room temperature for 3 min. RNA was extracted by phenol-chloroform extraction and precipitated by ethanol precipitation.

Free 40S subunits (devoid of mRNA, tRNA and initiation factors) were purified using a combination of published protocols (22–24). Briefly, cells were lysed in a high salt buffer (50 mM Tris pH 7.8, 500 mM KAc, 10 mM $MgCl_2$, 0.2% Triton X-100 and 2 mM DTT (Dithiothreitol) ) and clarified extracts were incubated with GTP (Guanosine-5'-triphosphate) and puromycin (1 mM final each) for half an hour at 37°C. Extracts were then fractionated on 10–50% sucrose gradients. Fractions highly enriched for free 40S subunits were modified with 1M7 (15 mM final) or incubated with DMSO (5% final) for 5 min at room temperature. 18S RNA was extracted and analyzed by primer extension (for details, see Supplementary Data).

We designed the chemical probing conditions so that 1 in every 100–300 nucleotides was modified (25). We only selected RNA samples where, by primer extension, we could detect high levels of chemical modification but little signal decay over a stretch of ∼200 nucleotides. The quality and integrity of isolated rRNAs were analyzed on a 2100 Bionalyzer (Agilent) using an RNA Nano 6000 kit.

Additional methods can be found in the Supplementary Data online.

## RESULTS

### Purification and chemical modification of pre-40S intermediates

For the rRNA structure probing experiments, 40S assembly intermediates at various stages of maturation were affinity purified using yeast strains expressing HTP or TAP-tagged assembly factors (Figure 1A–C). Intriguingly, unlike the Rio1-TAP strain (26), the Rio1-HTP strain used for our studies did not show noticeable growth or pre-rRNA processing defects (Supplementary Figure S1A–C). TAP-tagged Fun12 was used as bait for 80S translation initiation complexes, a source of mature 40S subunits (Figure 1B). Using Fun12-TAP as bait, we also purified the recently described 80S-like complexes that accumulated in Fap7 depleted strains (10) (Supplementary Figure S1A, lane 8). The Nob1-TAP strain was included as it is severely delayed in D-site cleavage (5) (Supplementary Figure S1A and B) allowing us to determine whether stalled pre-40S complexes had any unusual structural features.

Bioanalyzer analyses showed that our purification protocol generated intact and highly pure pre-rRNA samples (Figure 1C).

### Development of ChemModSeq, an ultra-high-throughput RNA structure probing and analysis method

Recent technological advances have made it possible to obtain RNA secondary structure information by combining chemical or enzymatic probing with high-throughput sequencing (27–33). We developed a protocol, dubbed ChemModSeq that uses RT with oligonucleotides that randomly hybridize to the RNA template (Figure 1D), allowing us to quickly generate an overview of chemically modified sites in rRNA molecules in a single RT reaction. While this work was in progress, a similar method, called structure-seq, was described to measure mRNA structure transcriptome-wide in *Arabidopsis thaliana in vivo* (31). The main difference between our protocol and structure-seq is that we apply a probabilistic model to map sites of chemical modification with high accuracy. In addition, we paired-end sequenced the cDNAs to obtain RT drop-off positions and RT oligonucleotide annealing sites. For each nucleotide we then calculate RT drop-off rates, which we define as the total number of reads that stop at a nucleotide divided by the total number of reads covering that nucleotide. This provided an indication of how reactive to a chemical a particular nucleotide was. Data generated using the ChemModSeq and structure-seq protocols can be used to calculate chemical (DMS/SHAPE) reactivity values for each nucleotide (31). However this approach does not consider whether the observed drop-off counts in chemically modified RNA are statistically significantly over background. To address this we developed a two-channel Poisson expectation maximization (EM) algorithm ($TCP^{EM}$; Figure 1F, see Supplementary Data) that calculates for each nucleotide the likelihood that it was chemically modified. In sequencing data from both unmodified and modified RNA there are nucleotide positions where the polymerase is more likely to drop-off (high), and positions where drop-off is less likely (low). The $TCP^{EM}$ algorithm assumes that these high and low drop-off rates are approximately constant along the RNA and assigns each nucleotide position to one of three categories: high drop-off in both channels (stops not caused by chemical), low drop-off in both channels (unmodified), high drop-off in the modified channel and low in the control (modified). A Poisson model is then used to assess the likelihood of a drop-off count $d_x$ for an inferred RNA-wide drop-off rate of $\lambda$. An EM algorithm then assigns each nucleotide position to one of the three categories.

To test the reliability of the ChemModSeq protocol we applied it to purified 80S translation initiation complexes, focusing on the 18S rRNA. This allowed us to compare the results with available crystal structures of yeast ribosomal subunits. To test the specificity of our approach, DMS was used as a chemical probe as it preferentially reacts with the N1 of adenosine and the N3 of cytosine. Biological replicates showed a high Pearson correlation for RT drop-off rates in the 18S rRNA (Supplementary Figure S2), demonstrating that the ChemModSeq protocol generates highly reproducible results.

We next compared ChemModSeq RT drop-off rates to primer extension results quantified using the semi-automated footprinting analysis software (SAFA) (34), focusing on the 5′ end of the 18S rRNA (Figure 2). To al-

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

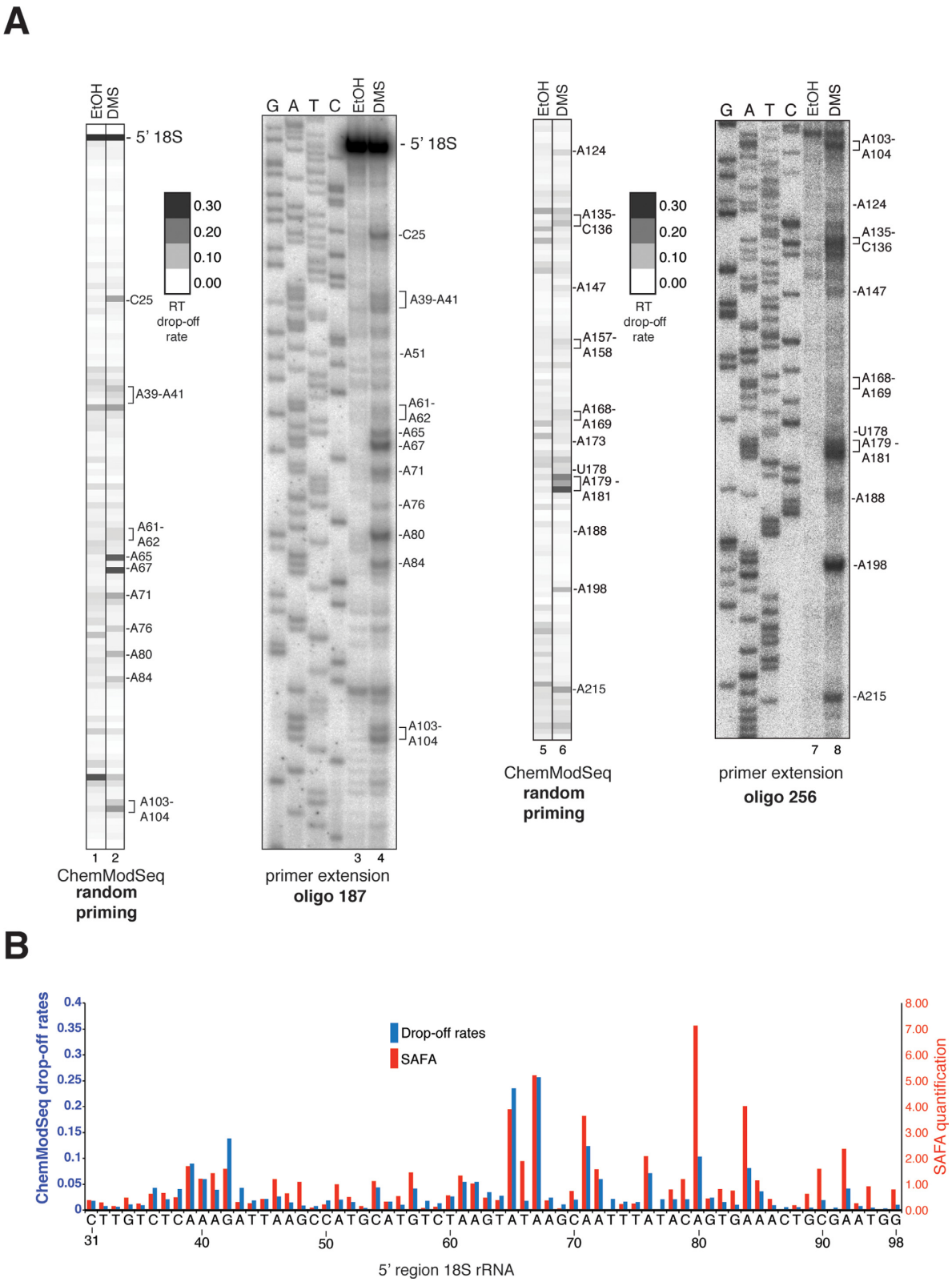



**Figure 2.** ChemModSeq accurately and quantitatively measures RNA secondary structure. (**A**) Primer extension results and ChemModSeq drop-off rates for DMS modified and unmodified 18S rRNA, focusing on the first ∼100 nucleotides. To compare ChemModSeq with primer extension data, drop-off rates were converted to gray scale heat maps in which each block represents a single nucleotide. The darker the color of the block, the higher the drop-off rate. Oligos used for primer extensions are indicated in bold (see Supplementary Table S4). The positions of the modified nucleotides are indicated on the right side of each panel. (**B**) ChemModSeq data agree well with primer extension results. Direct comparison of the primer extension results shown in (A), quantified using SAFA (34) (red bars) and ChemModSeq RT drop-off rates (blue bars) for the same region (nucleotides 31–98).

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

low direct visual comparison, RT drop-off rates were represented as grayscale heat maps (Figure 2A). This revealed that the two different approaches produced remarkably similar results; with statistical analysis confirming a strong correlation between the DMS reactivity values generated by SAFA and the ChemModSeq RT drop-off rates (Figure 2B; $r = 0.77$).

For each nucleotide in the 18S rRNA we then calculated DMS reactivities (as described by (31); Figure 3A and D). Although adenosines generally had higher DMS reactivities, the data also indicated modification of a higher than expected number of guanines and uracils and the pattern was similar to the distribution of nucleotides in single-stranded regions (Figure 3C and D). We then applied the TCP$^{EM}$ algorithm to the same data at a setting in which a nucleotide was only considered 'modified' if its probability of being modified was higher than 0.90 (Figure 3B and E). This greatly reduced the noise: 90% of the nucleotides called modified by the TCP$^{EM}$ algorithm were adenosines and cytosines (compare Figure 3F with D), which is consistent with the known chemical reactivity of DMS (35). Contrary to DMS data generated using the Mod-seq protocol (33), our ChemModSeq data did not show significant modification of guanine residues (Figure 3B and E).

The majority of the nucleotides identified as modified in the 5′ end of the 18S rRNA were located in single-stranded regions (Figure 3B). Modification of base-paired nucleotides (Watson–Crick and non-Watson–Crick) at the base of stem structures indicated that these regions were flexible under our purification conditions. Comparison of the primer extension data (Figure 2A) with the TCP$^{EM}$ output revealed that the number of false negatives was low (Figure 3B, blue colored nucleotides). Notably, the RT drop-off rates of the nucleotides called modified by the TCP$^{EM}$ algorithm were uncorrelated with solvent accessibility of the DMS reactive groups in single-stranded regions (Supplementary Figure S3A). This emphasizes that DMS reactivity also does not simply provide a read-out of solvent accessibility (36).

We conclude that ChemModSeq quantitatively and accurately detect sites of chemical modification and, in combination with the TCP$^{EM}$ algorithm, produces better signal-to-noise ratios than existing methods.

### Composition of pre-40S particles allows a classification into early, middle and late particles

We hypothesized that pre-40S complexes purified with the various baits (Figure 1C) might represent a mixture of particles at various maturation stages. To get an impression of their heterogeneity, we measured the level of Dim1-dependent cytoplasmic methylation at A1780 and A1781 in purified rRNAs by primer extension (Figure 4A). All 20S pre-rRNAs analyzed were methylated at these sites. We could, however, reproducibly see a stronger signal at A1781 in the 20S pre-rRNA associated with Fun12 in the late 80S-like complexes (also see below). This suggests that Dim1-dependent methylation at A1781 was not complete in the purified 20S pre-rRNAs associated with Ltv1, Enp1 and Tsr1. Hence, these baits likely purified a mixture of nuclear and cytoplasmic complexes. U1191 is acp-modified by an unknown enzyme just before D-site cleavage (3) (Figure 4B). Because this modification blocks RT (4), we measured the level of acp modification in purified rRNAs by primer extension. Intriguingly, this revealed that the level of acp modification at U1191 depended on the bait used for purifying the 20S pre-rRNA (Figure 4C). High RT drop-off signals at U1191 were found primarily in 20S pre-rRNA associated with Nob1, Rio1 and 80S-like particles, at levels comparable to the mature 18S rRNA (Figure 4C, lanes 5–8). In contrast, 20S pre-rRNAs associated with Ltv1 and Enp1 contained very low levels of U1191 acp modification (Figure 4C, lanes 1 and 2), whereas intermediate levels were detected in 20S associated with Rio2 and Tsr1 particles (Figure 4C, lanes 3 and 4). To rule out that other U1191 modifications significantly contributed to the primer extension signal, we performed primer extension analyses on 20S pre-rRNAs isolated from strains in which *SNR35* was deleted (Figure 4D). This deletion blocks U1191 pseudouridylation and Nep1-dependent N1 methylation but not acp modification (4,37) (Figure 4B). These experiments confirmed that the acp modification is primarily responsible for the strong stop observed in primer extension reactions. Based on these observations we categorized the pre-40S intermediates into early (Ltv1, Enp1), middle (Rio2, Tsr1) and late (Nob1, Rio1, 80S-like) particles.

### Late Rio1 pre-40S particles lack many 40S assembly factors

The remarkable variation in the level of U1191 acp modification in the purified 20S pre-rRNAs suggests that Enp1 and Ltv1 are probably largely absent in late pre-40S intermediates, whereas Rio2 and Tsr1 are present in substoichiometric amounts. To test this hypothesis, we compared the protein composition of middle (Rio2, Tsr1) and late (Rio1) pre-40S particles. The HTP-tagged Rio1 protein, which we used as bait for all our structure probing experiments, only transiently interacted with pre-40S complexes (Supplementary Figure S4). Therefore, to detect Rio1 associated proteins, we had to use a fast IgG affinity purification protocol (38) followed by label-free quantitative mass spectrometry. These results were then compared to mass spectrometry data obtained using a Rio2-HTP strain. This revealed that Rio1-HTP only significantly enriched for Nob1 and Dim2 (Figure 5A). Rio2 efficiently co-precipitated all of the assembly factors analyzed, except for Rio1, which was not detected in Rio2-HTP purifications (Figure 5A). This suggests that these two proteins are most likely not present in the same complex. To confirm that Rio1 interacts with Nob1 in pre-40S complexes, we performed tandem affinity purifications with strains expressing TAP-tagged versions of Rio2, Tsr1 and Rio1 (Figure 5B and C). Contrary to Rio1-HTP, the TAP-tagged Rio1 protein co-sedimented almost exclusively with 40S-sized particles in sucrose gradients, suggesting it is stably bound to pre-40S complexes (Supplementary Figure S4). Western blot analyses showed that TAP-tagged Rio2 and Tsr1 efficiently co-precipitated all of the tested factors (Figure 5B, lanes 4 and 6), indicating that Rio2 and Tsr1 'middle' particles share most of the assembly factors. Consistent with our mass spectrometry analyses, TAP-tagged Rio1 only significantly enriched for Nob1 (Figure 5C, lane 6).

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

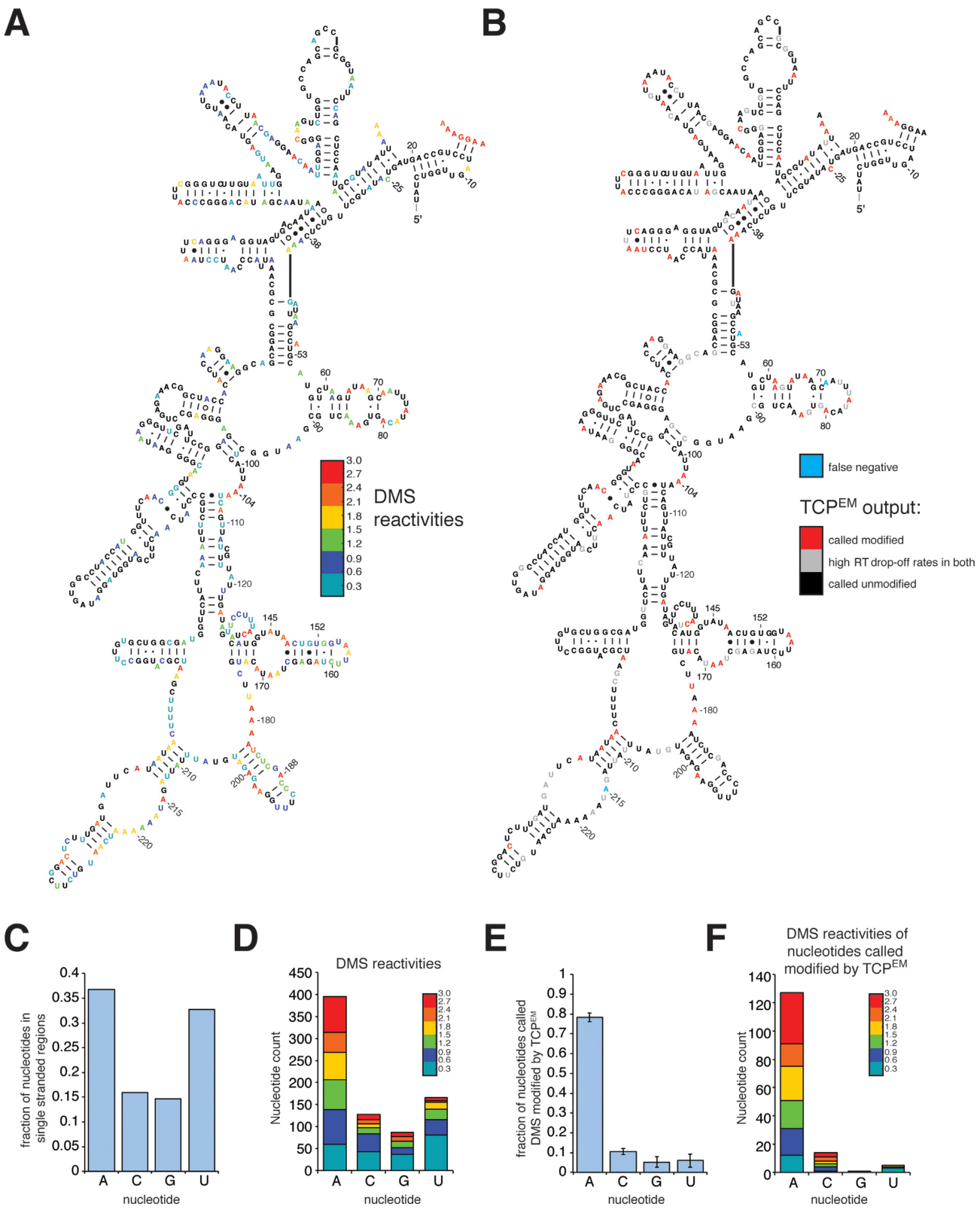



**Figure 3.** ChemModSeq accurately and quantitatively measures RNA secondary structure. (**A** and **B**) DMS reactivities (A) and TCP$^{EM}$ outputs (B) for nucleotides in the 18S rRNA depicted in the secondary structure of the 5′ region of the yeast 18S rRNA. DMS reactivities were calculated as described (31) and capped at 3. Red colored letters in (B) indicate nucleotides called modified by the TCP$^{EM}$ algorithm. Blue letters in (B) indicate nucleotides visually identified as modified by DMS in primer extension reactions (see Figure 2A) but not called modified by the TCP$^{EM}$ algorithm. (**C** and **D**) Distribution of nucleotides located in single-stranded regions in the 18S rRNA (C) and the distribution of nucleotides with DMS reactivities higher than zero (D). (**E** and **F**) Fraction of nucleotides called DMS-modified by the TCP$^{EM}$ algorithm (E; average and standard deviations from three independent experiments) and the DMS reactivities of these nucleotides (F). DMS reactivities were calculated by summing drop-off counts from three independent experiments.

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

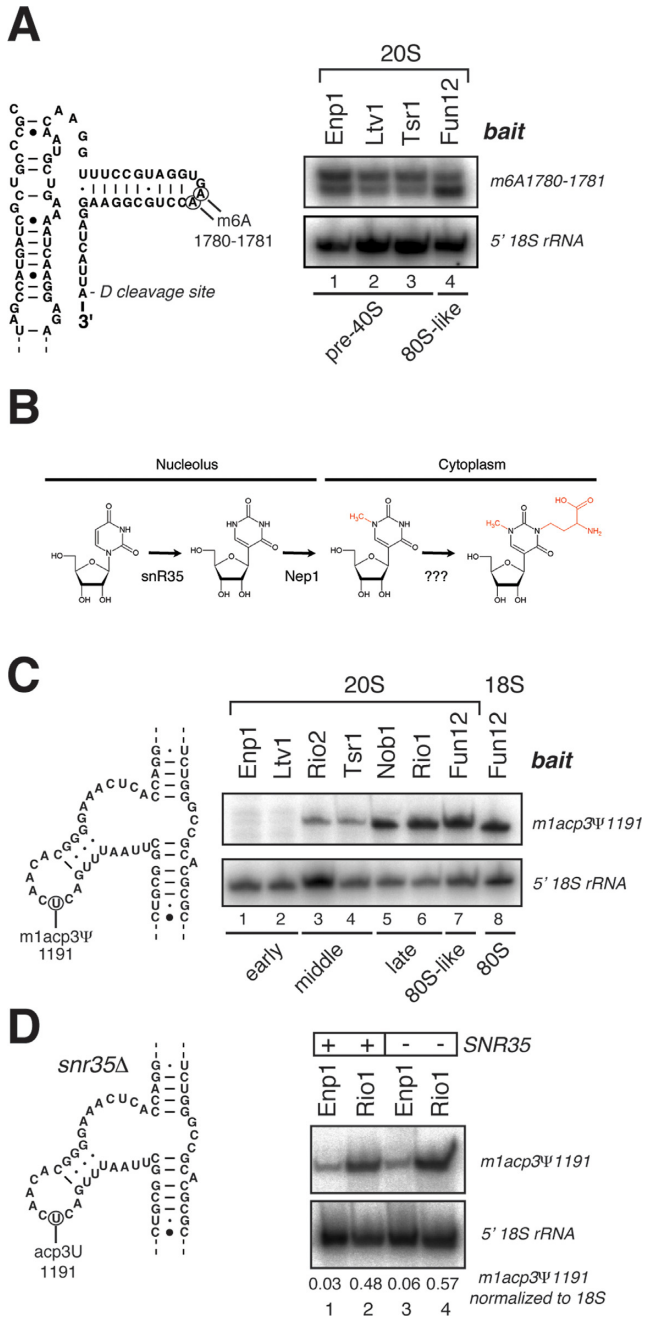



**Figure 4.** Pre-40S complexes can be classified into early, middle and late particles. (**A**) Detection of cytoplasmic methylation at A1780 and A1781 by primer extension in 20S pre-rRNAs. As loading controls, we performed primer extensions to detect the A1 cleavage site at the 5′ end of the pre-rRNAs. (**B**) Hypermodification of U1191 in the 18S rRNA. U1191 is first pseudouridinylated and N1-methylated by snR35 and Nep1 in the nucleolus, respectively, and then aminocarboxypropyl (acp)-modified in the cytoplasm by an unknown enzyme. (**C**) Detection of U1191 acp modification in 20S and 18S rRNAs by primer extension. Baits used to purify the individual pre-rRNA species are indicated above the panel. Loading controls were as in (A). (**D**) Same as in (C) but compared to strains lacking the *SNR35* gene.

We conclude that, as in human cells (13), late Rio1 pre-40S complexes lack the majority of 40S assembly factors.

## ChemModSeq reveals RNA structural differences between early-middle and late pre-40S particles

We next applied our ChemModSeq protocol to modified rRNAs purified from 40S assembly intermediates. The SHAPE reagent 1M7 was the preferred chemical probe for these experiments as it generically reacts with 2′-OH in all four nucleotides in single-stranded and flexible regions (14) and therefore would yield more structural information than DMS. Figure 6A and B show heat map representations of average RT drop-off rates for all samples for the nucleotides in the 18S rRNA and the 3′ major domain, respectively. Drop-off rates for the 18S rRNA for all samples are shown in Supplementary Table S2. Results for Internal Transcribed Spacer 1 (ITS1) will be described elsewhere (manuscript in preparation).

In agreement with the primer extension results shown in Figure 4, average drop-off rates at U1191 in unmodified 20S pre-rRNA samples were consistently higher in late particles, and high drop-off rates could be detected at the Dim1-dependent methylation sites in all samples (Supplementary Figure S5A and B).

The heat maps shown in Figure 6A and B revealed clear differences in nucleotide reactivity to 1M7 between early, middle and late particles (see regions highlighted with roman numbers). We then applied the TCP$^{EM}$ algorithm to identify regions where statistically significant differences in chemical reactivity could be detected between RNA samples. The TCP$^{EM}$ algorithm greatly simplified the interpretation of the data as it allowed us to identify easily rRNA regions that had the highest likelihood of being restructured during 40S assembly (Figure 6C; highlighted with boxes and roman numbers).

To validate the TCP$^{EM}$ results, we performed primer extension analyses (Figure 6D–F). This confirmed the observed differences in SHAPE reactivity between the various purified RNAs (see regions highlighted with roman numbers in Figure 6D–F). Bioinformatics analyses of the 18S rRNA SHAPE ChemModSeq sequencing data revealed that RT drop-off rates of nucleotides identified as modified by the TCP$^{EM}$ algorithm was uncorrelated with 2′OH solvent accessibility of single-stranded regions (Supplementary Figure S3B and C), reinforcing the observation that SHAPE chemicals provide a read-out of RNA dynamics, not RNA accessibility (15,20).

Generally, a larger number of nucleotides were modified in early and middle assembly intermediates suggesting that these adopt more flexible rRNA conformations (Figure 6G). The 3′ major domain in particular was much more reactive to 1M7 in early and middle particles. These included nucleotides in H33, H35, H37, H40 and H41 (Figure 6B–F, boxed regions indicated with roman numbers). 1M7 modified 18S isolated from empty and salt-washed free 40S particles (see Supplementary Methods), did not show the same degree of flexibility in H35, H37, H40 and H41 (Supplementary Figure S6). This indicated that many of the observed changes in SHAPE reactivity in the 3′ major domain were

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

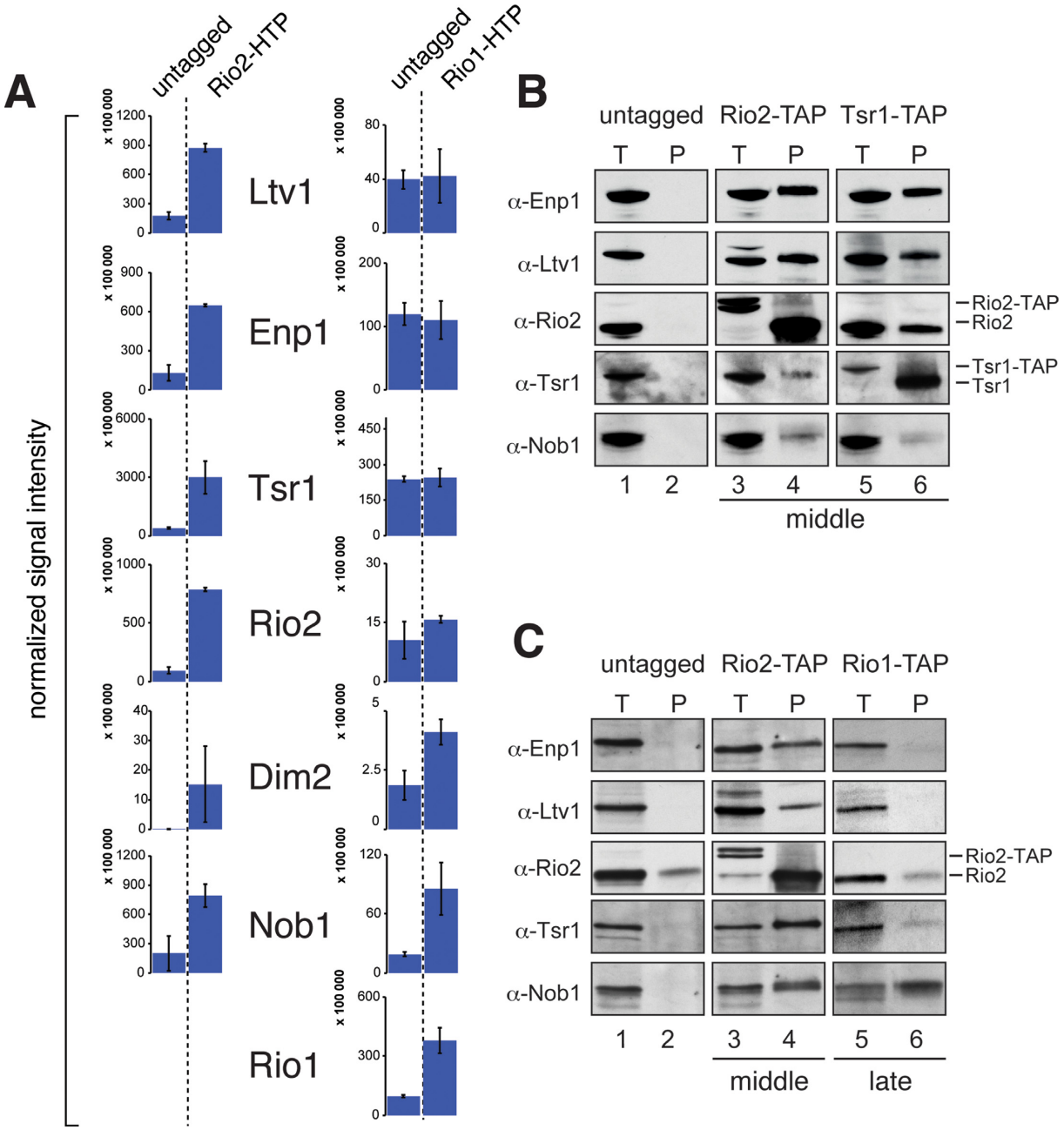



**Figure 5.** Rio1 associated pre-40S complexes contain fewer 40S assembly factors. (**A**) HTP-tagged Rio2 and Rio1 were affinity purified and associated proteins were quantified by label-free LC-MS. Shown are the averages and standard deviations from two (Rio2) or three (Rio1) replicate experiments. The y-axis indicates the relative abundance of the proteins in the purified complexes. The number of peptides used to calculate these abundances is listed in Supplementary Table S1. No Rio1 peptides were found in the Rio2 affinity purification. (**B** and **C**) Western blot analysis of protein composition of middle (Rio2, Tsr1) and late (Rio1) pre-40S particles purified by Tandem Affinity Purification (TAP) using TAP-tagged strains.

not the result of pre-40S complexes joining with 60S subunits or binding of tRNA to the decoding center.

Just before we completed the *in vitro* 1M7 experiments, a new SHAPE reagent NAI for probing of RNA secondary structures in living cells was reported (21). To substantiate our results we compared middle (Tsr1) and late (Rio1) with mature (Fun12) particles modified *in vivo* using NAI (Supplementary Figure S7). Importantly, the same regions in the 3′ major domain were also modified by NAI in early and middle pre-40S complexes *in vivo*, indicating that the observed differences in 1M7 reactivity were not the result of r-proteins dissociating during the purification (Supplementary Figure S7, boxed regions in gels and secondary structure model). We note that some nucleotides reacted differentially to NAI and 1M7. Our *in vitro* 1M7 probing experiments were performed using $MgCl_2$ concentrations that support translation. Under these conditions ribosomal subunits are known to have higher conformational dynamics, which may account for some of the differences observed (39). We can also not rule out the possibility that pre-

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

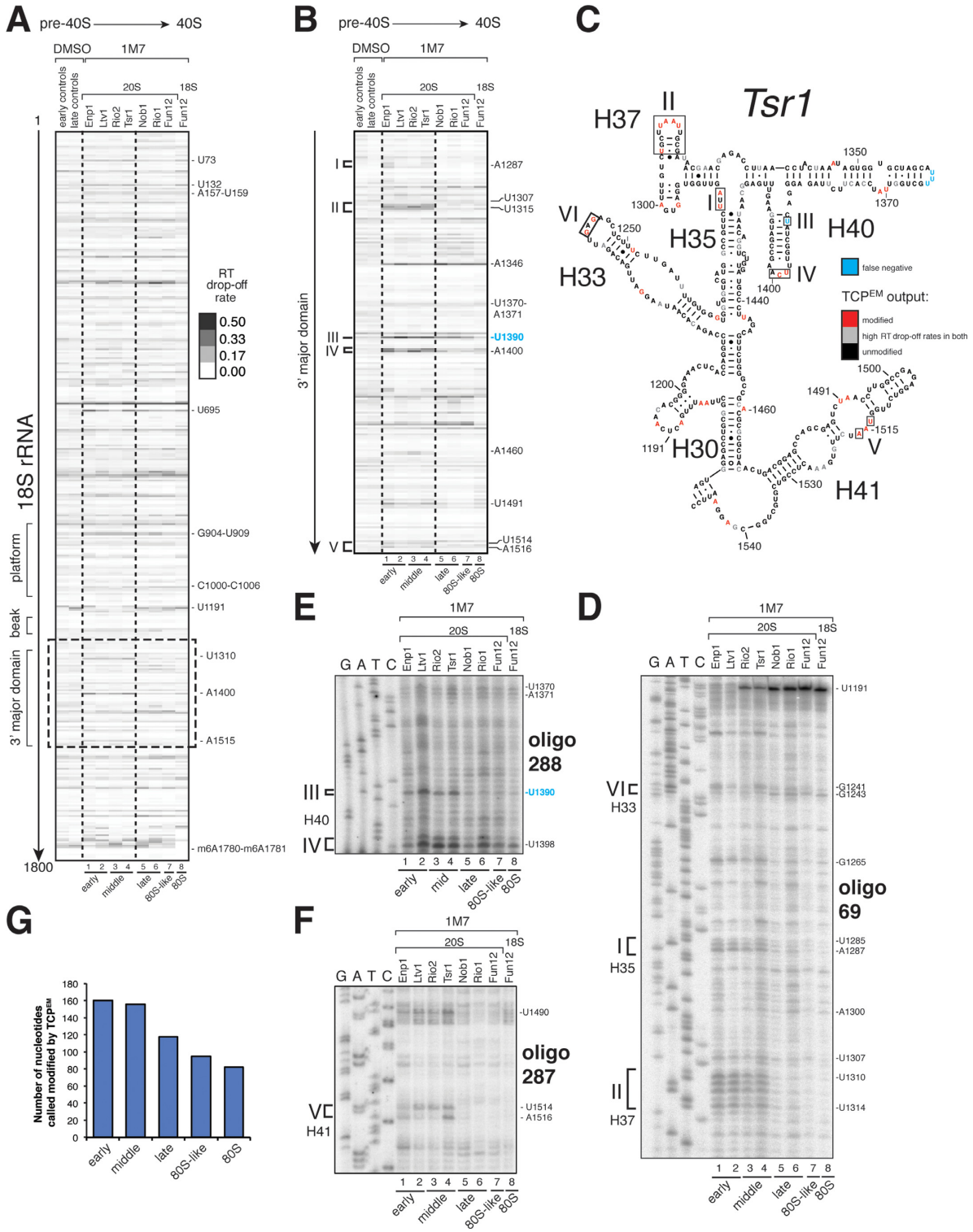



**Figure 6.** A digital snapshot of pre-rRNA secondary structures. (**A** and **B**) ChemModSeq and primer extension results obtained using rRNAs extracted from *in vitro* 1M7 modified particles. Heat maps of average RT drop-off rates in 18S coding sequences ($n \geq 2$) (A) or the 3′ major domain (B) of purified rRNAs. The positions of the modified nucleotides are indicated on the right side of each panel. (**C**) Early-middle and late pre-40S particles are structurally distinct. The 3′ major domain secondary structure represents the ChemModSeq results for the middle (Tsr1) particle. Red nucleotides were indicated as modified by the TCP$^{EM}$ algorithm. Blue nucleotides were visibly modified by 1M7 but not called modified by the TCP$^{EM}$ algorithm (false negative). (**D**–**F**) Representative primer extension reactions for the 3′ major domain. Roman numbers indicate sites where differences in SHAPE reactivity were observed between pre-40S particles. Oligonucleotides used for primer extensions are indicated in bold. The positions of the modified nucleotides are indicated on the right side of each panel. (**G**) Pre-rRNAs in early and middle pre-40S complexes adopt more flexible conformations compared to late particles. Plotted are the average number of nucleotides called modified by the TCP$^{EM}$ algorithm (y-axis) for early, middle, late, 80S-like and mature 80S translation initiation complexes (x-axis).

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

rRNAs undergo slight structural changes during the purification procedure. Finally, some SHAPE chemicals also have different preferences for nucleotide conformations (40).

We conclude that the 3′ major domain in the 20S pre-rRNA is restructured during late stages of maturation.

### Restructuring of the 3′ major domain coincides with the release of assembly factors

Many of the highly flexible nucleotides in early-middle pre-40S complexes clustered near previously identified assembly factor UV cross-linking sites (see (5)) (Figure 7A and B). This indicates that these factors play a role in maintaining a more flexible rRNA conformation. In late pre-40S particles, including the Rio1 particle, the 3′ major domain was significantly less reactive to SHAPE chemicals (Figure 6B–D, Supplementary Figure S7A–E, regions indicated with roman numbers). The observed structural rearrangements in the 3′ major domain therefore presumably occur before Rio1 enters the assembly pathway and may coincide with the release of assembly factors Rio2, Tsr1, Ltv1 and Enp1. Although Nob1 assembles into pre-ribosomes at a very early stage (41), and cross-links to the 3′ major domain (5), it does not appear to contribute to these RNA restructuring events, as pre-40S complexes that accumulated in cells lacking Nob1 had all the hallmarks of late pre-40S particles (Supplementary Figure S8: see regions indicated with roman numbers).

### Ribosome assembly factors are required to maintain an rRNA conformation that is incompatible with stable binding of Rps17 and Rps3

Several of the highly flexible nucleotides detected in early-middle pre-40S particles are contacted by r-proteins in the mature subunit (42). For example, the terminal loop of H37 and the stem of H40 are contacted by Rps17 (Figure 7D). Nucleotides U1514 and A1516, located in an internal loop in H41 in close proximity to an Ltv1 cross-linking site, contact Rps20 (A1516) and the N-terminus of Rps3 (U1514), (Figure 7E). Thus, although these r-proteins are present in pre-40S particles (43) they might not contact these sites in early and middle pre-40S complexes. If so, then it might be expected that these r-proteins are less stably associated with pre-40S particles compared to mature subunits. Consistent with this idea, Rps3 dissociates when pre-40S particles are subjected to high salt conditions (7). To test whether this is also the case for other r-proteins, we performed immunoprecipitation experiments with strains expressing FLAG-tagged r-proteins (Figure 7G and H). Under high salt conditions the binding of both Rps3 and Rps17 to pre-40S complexes was reduced, whereas their recovery with mature 40S subunits was more efficient than under low salt conditions. This indicates that these r-proteins are recruited to pre-ribosomes, but are not in their final conformation in early-middle pre-ribosomes.

Notably, our data also revealed rRNA structural rearrangements in the beak. In mature 40S subunits, Rps15 binds to G1241 in the tip of H33 (Figure 6C, box VI) close to an Enp1 cross-linking site (Figure 7B). G1243, which is exposed to the surface in the crystal structure (Figure 7F), reacted more strongly with SHAPE chemicals in late pre-40S and mature 40S particles. G1241, however, was more reactive to 1M7 in early intermediates (Figure 6D). This indicates that Rps15 may only interact with the tip of H33 once the beak has been formed.

To address whether the observed changes in nucleotide flexibility between early-middle and late pre-40S particles could be the result of ribosomal protein binding events, we performed 1M7 probing experiments on 18S rRNA that was *in vitro* refolded under the same conditions used to purify pre-40S complexes (Figure 8A–C, compare lanes 7 and 8). Comparison of the SHAPE reactivity profiles revealed that large regions of the 3′ major domain in the refolded rRNA adopted secondary structures similar to 18S and 20S rRNA molecules in purified particles (Figure 8A–C). Surprisingly, the majority of the highly SHAPE reactive nucleotides in early and middle pre-40S particles that clustered near assembly factor binding sites (H33, H35, H37, H40 and H41; see Figure 7) did not show the same degree of flexibility in *in vitro* refolded RNA (Figure 8D, red colored nucleotides in regions indicated with roman numbers). This suggests that under the conditions used, these regions can fold into, what appear to be, relatively stable structures independently of proteins *in vitro*. In the yeast 80S crystal structure, the nucleotides in the H37 terminal loop and the internal loop in H41 (U1514–U1517) that do not form Watson–Crick base-pairs, form a large network of base-stacking interactions, including many long-range interactions (Figure 8D, indicated with dotted lines). These presumably help stabilize the structure of the rRNA. Many of these stacking interactions are s53 interactions (5′ side of one base with 3′ side of other base) that are generally less reactive to SHAPE reagents (44), The relatively low SHAPE reactivity of these nucleotides in the *in vitro* refolded 18S rRNA suggests that at least some of these base-stacking interactions form in the absence of proteins. Based on these results, we hypothesize that assembly factors that associate with early and middle pre-40S particles function by providing the energy necessary to maintain a flexible 3′ major domain conformation.

## DISCUSSION

Ribosome synthesis in eukaryotes is a highly complex and dynamic process that involves ordered assembly of r-proteins and numerous RNA structural rearrangements. Hundreds of assembly factors are also involved in the process, many of which have the capacity to extract energy from nucleotide tri-phosphates and presumably play key roles in these restructuring events. Major challenges in the ribosome synthesis field have been to find methods that reveal detailed mechanistic insights into the assembly pathway and to assign functions to assembly factors. Because the assembly intermediates are heterogeneous and dynamic, they are difficult to characterize structurally and other complementary approaches are needed to address these mechanistic questions. Inspired by impressive chemical probing work done by many groups on bacterial ribosomes, we have developed protocols for purification and chemical modification of specific yeast ribosome assembly intermediates. By combining SHAPE chemical probing with high-throughput sequencing and statistical modeling, we were able to rapidly and

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

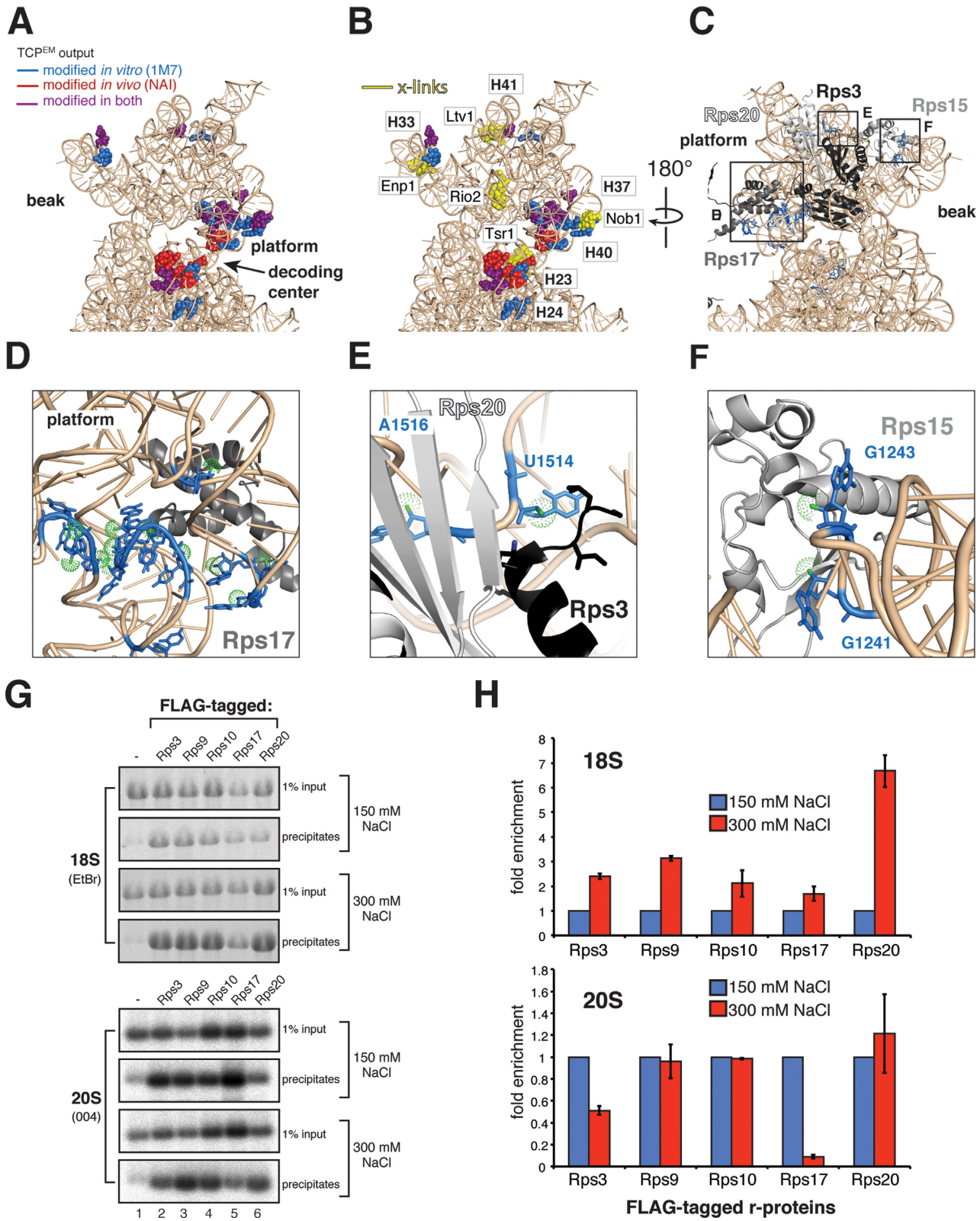



**Figure 7.** Ribosomal protein rRNA binding sites are highly flexible in early and middle pre-40S complexes. (**A** and **B**) Nucleotides in the head domain reactive to SHAPE chemicals (A) cluster near assembly factor UV cross-linking sites (B). (**C**–**F**) SHAPE-modified nucleotides (blue) in the head domain coincide with Rps3, Rps15, Rps17 and Rps20 binding sites. Relevant nucleotide positions are indicated. Green dots highlight the 2′-OH. (**G** and **H**) Ribosomal proteins Rps3 and Rps17 form salt-labile interactions with pre-40S complexes. Immunoprecipitations were performed using strains expressing FLAG-tagged r-proteins (indicated above the panels). Co-precipitated 20S was detected by northern hybridization with oligo 004, 18S rRNA was detected by ethidium bromide staining (EtBr). Input indicates 1% of total RNA extracted from cell lysates. (H) Quantification of results shown in (G). Error bars indicate standard deviations obtained from two biological replicates.

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

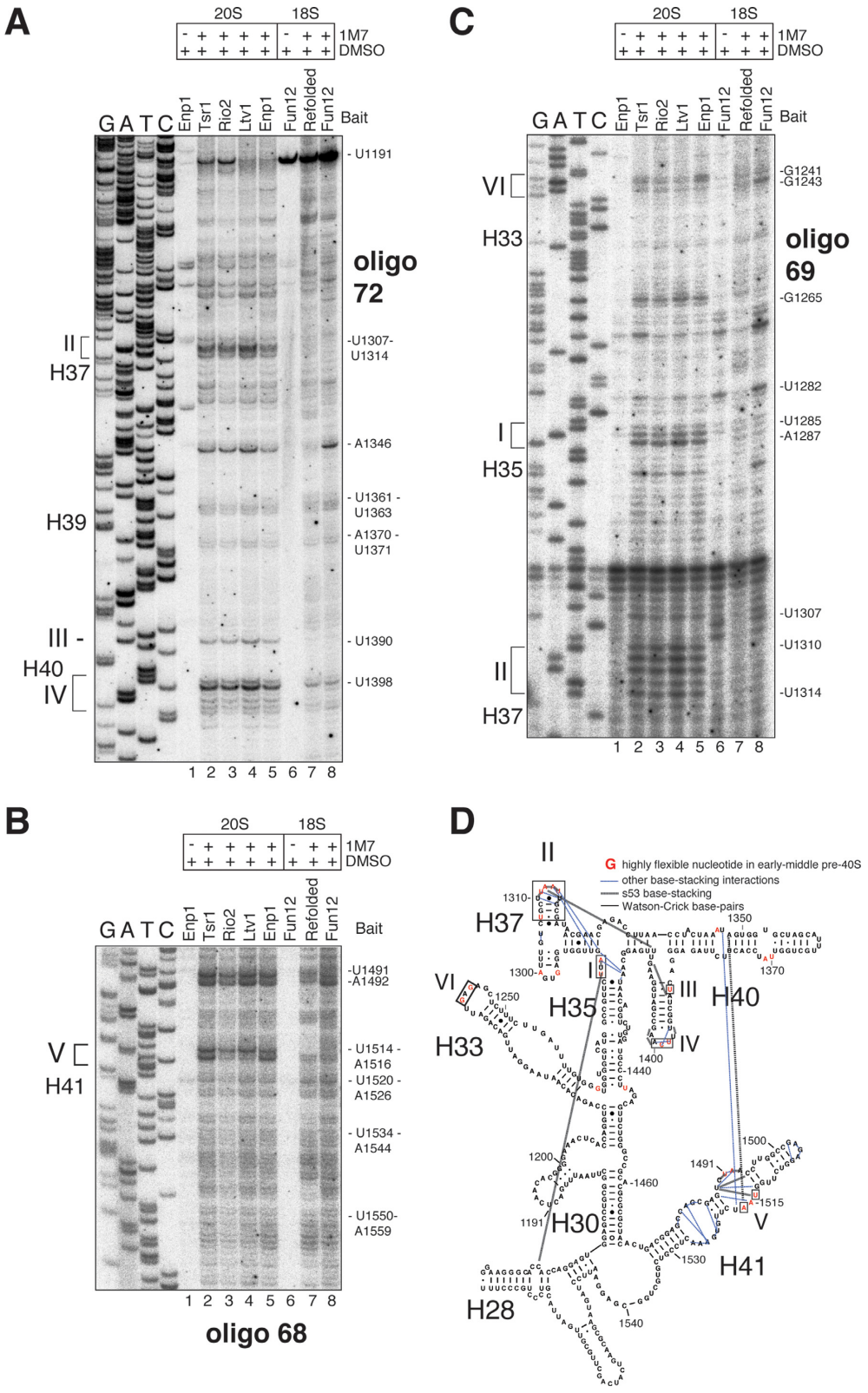



**Figure 8.** Ribosome assembly factors are required to maintain an open head domain conformation. (**A**–**C**) Comparison of primer extension data obtained from 1M7 modified early (Ltv1, Enp1), middle (Tsr1, Rio2) pre-40S complexes (lanes 2–5), mature 40S subunits (Fun12) (lane 8) and *in vitro* refolded 18S rRNA (lane 7, Refolded). Unmodified 20S and 18S rRNAs were used as control samples (lanes 1 and 6). The positions of the modified nucleotides are indicated on the right side of each panel. (**D**) Overview of $TCP^{EM}$ output generated from *in vitro* refolded 18S rRNA ChemModSeq data. Shown are the results for the 3′ major domain. Yellow letters indicate nucleotides that the algorithm predicted were most likely 1M7 modified.

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

quantitatively measure RNA structural changes that take place during 40S ribosome synthesis in yeast. Our method, dubbed ChemModSeq, provides the first digital snapshots of pre-rRNA nucleotide dynamics during 40S assembly, revealing new mechanistic insights into pre-rRNA restructuring events and the function of 40S assembly factors. To detect sites of modification with high accuracy, we developed a two-channel Poisson EM algorithm ($TCP^{EM}$). This algorithm calculates for each nucleotide the likelihood that it is chemically modified, allowing the detection of sites where statistically significant differences in chemical reactivity between various RNA samples could be identified. We demonstrate that the algorithm can greatly reduce noise in structure probing sequencing data, increasing the reliability of the results. $TCP^{EM}$ also simplified the interpretation of the data as it allowed us to quickly identify regions that were most likely restructured during 40S assembly.

The ChemModSeq protocol and the statistical methods described here are applicable to any ribonucleoprotein complex.

### assembly factors chaperone pre-rRNA folding steps during 40S assembly

Previous work had indicated that the binding sites of several 40S ribosome assembly factors is incompatible with translation initiation, as many occupied tRNA and translation initiation factor binding sites (5,6). It was proposed that these factors play an important role in quality control by blocking premature formation of translation initiation complexes (6). An alternative, but not mutually exclusive, interpretation of these results is that binding of ribosome assembly factors to functionally important regions is required for the formation of the functional center of the ribosome (5,45). In agreement with this model, our results suggest that ribosome assembly factors play an important role in regulating RNA folding events that take place during late stages of 40S assembly.

The ChemModSeq analyses revealed distinct differences in the pattern of nucleotide SHAPE reactivity between 40S assembly intermediates. Pre-40S particles that contained Rio2, Tsr1, Ltv1 and Enp1 were generally more flexible compared to late pre-40S complexes that lacked these proteins. Many of these flexible nucleotides in early-middle pre-40S complexes clustered near assembly factor binding sites, including Ltv1 and Enp1 binding sites in the 3′ major domain. Interestingly, our ChemModSeq analyses on *in vitro* refolded 18S rRNA (Figure 8) demonstrated that many of these highly flexible regions could spontaneously fold into, what appeared to be, relatively stable structures. We propose that assembly factors function by providing the necessary energy to maintain a more flexible conformation in the pre-rRNA (Figure 9A). More specifically, we hypothesize that Ltv1 and Enp1, which bind to the head domain, act as RNA chaperones by delaying folding steps in the 3′ major domain (Figure 9A). This would be required to allow earlier pre-rRNA folding steps to be completed and/or to safeguard correct folding by preventing the accumulation of kinetically trapped intermediates. A similar function has been proposed for bacterial RimM, which regulates 3′ major domain formation by holding the rRNA in a conformation that allows proper folding of the beak structure and H43 (46). Similarly, the U3 and snR30 snoRNAs regulate the formation of long-range rRNA interactions in the 18S rRNA. The U3-rRNA-5′18S rRNA base-pairing interaction blocks the formation of the universally conserved central pseudoknot (47), whereas snR30-rRNA base-pairing interferes with long-range interactions between two rRNA expansion segments (48). At some point during the maturation pathway, certain assembly factors are no longer required and need to be released. This can be achieved through phosphorylation of proteins (such as Hrr25-dependent phosphorylation of Enp1 and Ltv1, Figure 9B; (7)), which presumably reduces their affinity for RNA, or through the activity of other energy-dependent remodelers, such as RNA helicases and GTPases (e.g. see (49,50)). Release of these factors would allow completion of RNA folding steps, trigger binding of late assembling r-proteins (our work) or trigger remodeling steps (7,51).

We and others have shown that the presence of assembly factors affects stable binding of Rps17 and Rps3 (Figure 7 and (7)). Several of the more flexible nucleotides in H37 and H40 detected in early and middle pre-40S particles are contacted by Rps17 in the mature subunit (42). Rps17 is present in pre-40S particles (43); however, our data suggest that Rps17 may not contact these sites during early stages of 40S maturation. Our results also indicate that the interaction between Rps15 and the tip of H33 may only take place at late stages of assembly once the beak has been formed (Figure 9B).

It is not entirely clear why some r-proteins do not adopt their final conformation in early-middle pre-40S complexes. Plausible explanations are that this is necessary to maintain the more flexible conformations in 20S pre-rRNA observed in pre-40S complexes, to prevent non-specific protein–RNA and/or ensure that binding sites for late assembling r-proteins remain accessible. It has also been suggested that a flexible head domain conformation is important for efficient nuclear export (7). We note that both Rps3 and Rps15 are required for efficient nuclear export of pre-40S complexes (52,53) and it is possible that their assembly state is a measure of export competence. Similarly, 60S subunit assembly also seems to involve gradual strengthening of r-protein binding to pre-ribosomes (54,55). The binding strength of many 60S r-proteins was shown to increase once internal transcribed spacer 2 (ITS2) is been removed (55). It is becoming clear that a major function of ribosome assembly factors is to direct the timely incorporation of r-proteins by modulating the rRNA folding state.

Several lines of evidence indicate that Ltv1 and Enp1 also play a role in recruiting Rps3, Rps15 and Rps20 to pre-ribosomes: Ltv1, Enp1 and Rps3 form a salt-resistant complex *in vitro* (7), Ltv1 binds Rps3 in the yeast two-hybrid assay (56) and binds Rps15 *in vitro* (57). Furthermore, Rps3 and Rps20 could not be detected in pre-40S complexes purified from cells lacking Ltv1 (6).

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

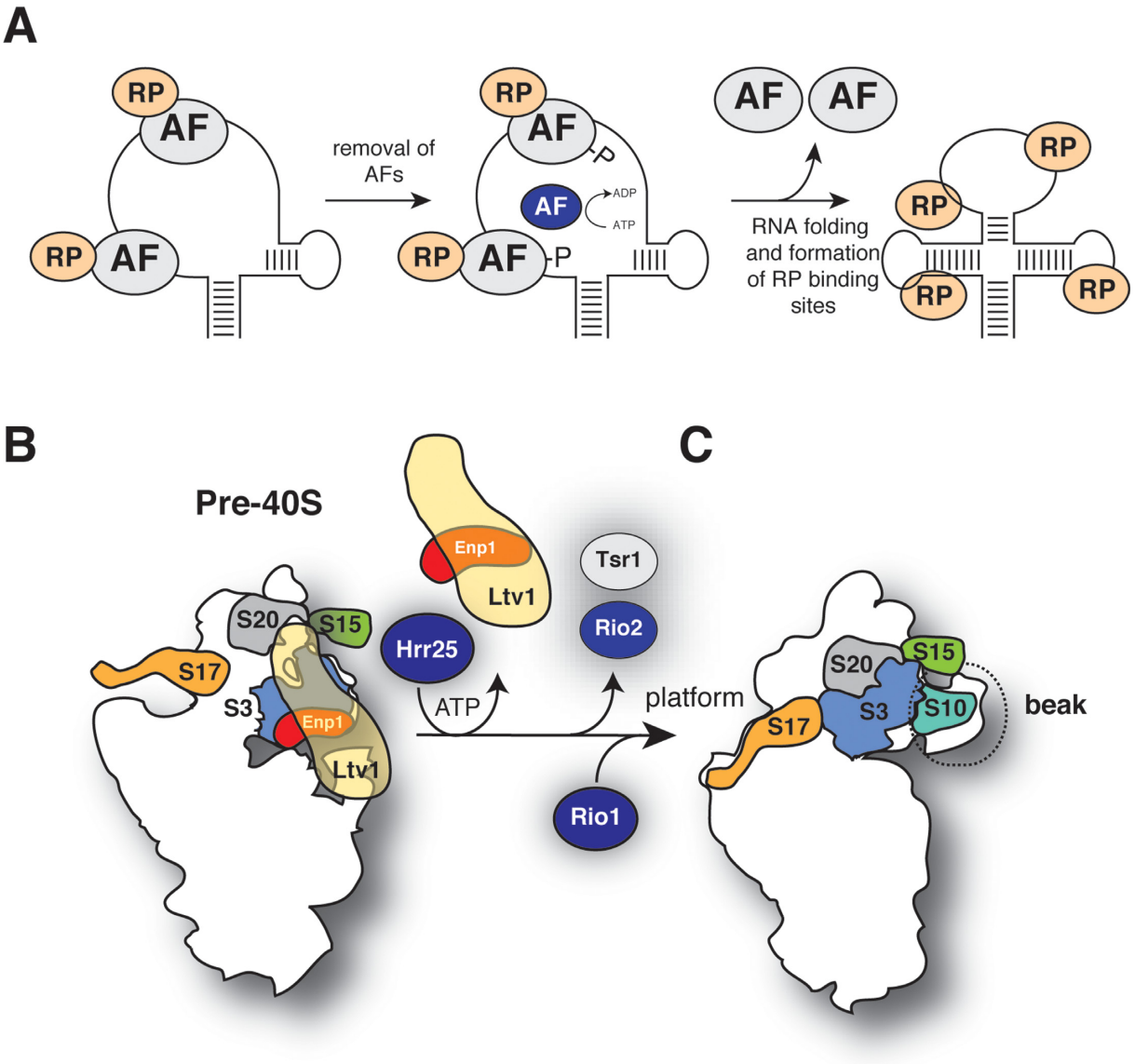



**Figure 9.** A model for structural rearrangements during late stages of 40S subunit maturation. (**A**) Proposed model for the function of ribosome assembly factors in 40S assembly. Binding of assembly factors (AF) to RNA is required to maintain a more flexible conformation. Their presence allows the assembly of ribosomal proteins (RP) but prevent that these RPs adopt their final conformation. At some point during the maturation pathway, the system receives a signal that certain ribosome assembly/RNA folding steps have been completed and AFs are no longer needed. Their release could be triggered through phosphorylation (such as Hrr25-dependent phosphorylation of Enp1 and Ltv1). This allows completion of RNA folding steps and binding of late RPs. (**B** and **C**) Model for rearrangements that take place in the head domain during late stages of 40S assembly. In early and middle pre-40S complexes, binding of Ltv1 and Enp1 to the head domain is required to maintain a more flexible conformation in the 3′ major domain. Rps3, Rps17, Rps20 and Rps17 assemble into pre-ribosomes but do not adopt their final conformation. Hrr25 phosphorylation of Ltv1 and Enp1 triggers their release, allowing late RNA folding steps and assembly of head domain r-proteins to be completed. We propose that Rio2 and Tsr1 are released shortly afterward (B) and Rio1 enters the assembly pathway. The exact function of Rio1, however, remains unclear.

### Restructuring of the 3′ major domain coincides with the release of the majority of pre-40S assembly factors and takes place before Rio1 enters the assembly pathway

Our results suggest that restructuring of the 3′ major domain coincides with the release of Rio2, Tsr1, Ltv1 and Enp1, before Rio1 enters the assembly pathway (Figure 9B and C). This is based on the following observations: (i) The SHAPE reactivity profile of Rio1 associated 20S pre-rRNA and (ii) the level of acp modification were both very similar to what we observed with mature 18S rRNA (Figure 6). This implies that Rio1 primarily associates with late pre-40S complexes. (iii) Rio1 was not significantly enriched in Rio2-associated particles, indicating they are not present in the same complex (Figure 5) and (iv) immunoprecipitation experiments with epitope-tagged Rio1 showed only significant enrichment of Nob1 (Figure 5), consistent with recently published work (58). (v) Finally, data from the Tollervey lab confirmed that Rio1-associated pre-40S particles and the later 80S-like complexes mostly contain Dim2 and Nob1 (David Tollervey, personal communication). Hence,

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

all the available data suggest that Rio1 associates with pre-40S particles that are in their final stages of maturation. We speculate that once Rio2, Tsr1, Ltv1 and Enp1 are released, r-proteins in the head domain can adopt their final conformation and late assembling r-proteins (such as Rps10) can assemble into pre-ribosomes (Figure 9B and C). The resulting pre-40S, containing Nob1 and Dim2 and possibly also Rio1, then joins with a 60S subunit to form an 80S-like complex. Successful joining may trigger Nob1-dependent cleavage at site D (10–12).

## ACCESSION NUMBER

Raw (fastq) and processed sequencing data can be downloaded from the NCBI Gene Expression Omnibus repository (http:// www.ncbi.nlm.nih.gov/geo, accession number GSE52878).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGMENTS

We are grateful to Agata Swiatkowska for help with setting up the chemical probing method in the lab. We thank Sergey Melnikov and Adam Ben-Shem for providing solvent accessibility data, Katrin Karbstein for providing antibodies, Philipp Milkereit for providing yeast strains expressing FLAG-tagged ribosomal proteins, Rob van Nues for critically reading the manuscript and Lisa Muniz for expert technical assistance. We thank the Hi5RNA group for fruitful discussions and helpful comments on the manuscript.

## FUNDING

ERASMUS student fellowships ( to A.Z. and M.T.); Wellcome Trust PhD fellowship [096997 to E.B.]: Medical Research Council Career Development Grant [G1000520 to A.G.C.]; Wellcome Trust Research and Career Development Grant [091549 to S.G.]; Wellcome Trust Centre for Cell Biology core grant [092076]. Funding for open access charge: Wellcome Trust.
*Conflict of interest statement.* None declared.

## REFERENCES

1. Henras,A.K., Soudet,J., Gerus,M., Lebaron,S., Caizergues-Ferrer,M., Mougin,A. and Henry,Y. (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell. Mol. Life Sci.*, **65**, 2334–2359.
2. Lafontaine,D., Delcour,J., Glasser,A.L., Desgres,J. and Vandenhaute,J. (1994) 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.
3. Brand,R.C., Klootwijk,J., Planta,R.J. and Maden,B.E. (1978) Biosynthesis of a hypermodified nucleotide in Saccharomyces carlsbergensis 17S and HeLa-cell 18S ribosomal ribonucleic acid. *Biochem. J.*, **169**, 71–77.
4. Liang,X.H., Liu,Q. and Fournier,M.J. (2009) Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA*, **15**, 1716–1728.
5. Granneman,S., Petfalski,E., Swiatkowska,A. and Tollervey,D. (2010) Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO J.*, **29**, 2026–2036.
6. Strunk,B.S., Loucks,C.R., Su,M., Vashisth,H., Cheng,S., Schilling,J., Brooks,C.L., Karbstein,K. and Skiniotis,G. (2011) Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science*, **333**, 1449–1453.
7. Schäfer,T., Maco,B., Petfalski,E., Tollervey,D., Böttcher,B., Aebi,U. and Hurt,E. (2006) Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature*, **441**, 651–655.
8. Ferreira-Cerca,S., Sagar,V., Schäfer,T., Diop,M., Wesseling,A.-M., Lu,H., Chai,E., Hurt,E. and Laronde-Leblanc,N. (2012) ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nat. Struct. Mol. Biol.*, **19**, 1316–1323.
9. Vanrobays,E., Gelugne,J.-P., Gleizes,P.-E. and Caizergues-Ferrer,M. (2003) Late cytoplasmic maturation of the small ribosomal subunit requires RIO proteins in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **23**, 2083–2095.
10. Strunk,B.S., Novak,M.N., Young,C.L. and Karbstein,K. (2012) A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell*, **150**, 111–121.
11. Lebaron,S., Schneider,C., van Nues,R.W., Swiatkowska,A., Walsh,D., Böttcher,B., Granneman,S., Watkins,N.J. and Tollervey,D. (2012) Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nat. Struct. Mol. Biol.*, **19**, 744–753.
12. García-Gómez,J.J., Fernández-Pevida,A., Lebaron,S., Rosado,I.V., Tollervey,D., Kressler,D. and la Cruz,de, J. (2014) Final pre-40S maturation depends on the functional integrity of the 60S subunit ribosomal protein L3. *PLoS Genet.*, **10**, e1004205.
13. Widmann,B., Wandrey,F., Badertscher,L., Wyler,E., Pfannstiel,J., Zemp,I. and Kutay,U. (2012) The kinase activity of human Rio1 is required for final steps of cytoplasmic maturation of 40S subunits. *Mol. Biol. Cell*, **23**, 22–35.
14. Mortimer,S.A. and Weeks,K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
15. McGinnis,J.L., Dunkle,J.A., Cate,J.H.D. and Weeks,K.M. (2012) The Mechanisms of RNA SHAPE Chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.
16. Brachmann,C.B., Davies,A., Cost,G.J., Caputo,E., Li,J., Hieter,P. and Boeke,J.D. (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, **14**, 115–132.
17. Rigaut,G., Shevchenko,A., Rutz,B., Wilm,M., Mann,M. and Séraphin,B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.*, **17**, 1030–1032.
18. Granneman,S., Kudla,G., Petfalski,E. and Tollervey,D. (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9613–9618.
19. Swiatkowska,A., Wlotzka,W., Tuck,A., Barrass,J.D., Beggs,J.D. and Tollervey,D. (2012) Kinetic analysis of pre-ribosome structure in vivo. *RNA*, **18**, 2187–2200.
20. Weeks,K.M. and Mauger,D.M. (2011) Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.*, **44**, 1280–1291.
21. Spitale,R.C., Crisalli,P., Flynn,R.A., Torre,E.A., Kool,E.T. and Chang,H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.
22. Fleischer,T.C., Weaver,C.M., McAfee,K.J., Jennings,J.L. and Link,A.J. (2006) Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Dev.*, **20**, 1294–1307.
23. Meskauskas,A., Leshin,J.A. and Dinman,J.D. (2011) Chromatographic purification of highly active yeast ribosomes. *J. Vis. Exp.*, **56**, e3214.
24. Leshin,J.A., Heselpoth,R., Belew,A.T. and Dinman,J. (2011) High throughput structural analysis of yeast ribosomes using hSHAPE. *RNA Biol.*, **8**, 478–487.
25. Mortimer,S.A. and Weeks,K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014

snapshots and at single-nucleotide resolution. *Nat. Protoc.*, **4**, 1413–1421.
26. Soudet,J., Gelugne,J.P., Belhabich-Baumas,K., Caizergues-Ferrer,M. and Mougin,A. (2009) Immature small ribosomal subunits can engage in translation initiation in Saccharomyces cerevisiae. *EMBO J.*, **29**, 80–92.
27. Underwood,J.G., Uzilov,A.V., Katzman,S., Onodera,C.S., Mainzer,J.E., Mathews,D.H., Lowe,T.M., Salama,S.R. and Haussler,D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*, **7**, 995–1001.
28. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
29. Lucks,J.B., Mortimer,S.A., Trapnell,C., Luo,S., Aviran,S., Schroth,G.P., Pachter,L., Doudna,J.A. and Arkin,A.P. (2011) Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11063–11068.
30. Seetin,M.G., Kladwang,W., Bida,J.P. and Das,R. (2014) Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol. Biol.*, **1086**, 95–117.
31. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., Bevilacqua,P.C. and Assmann,S.M. (2013) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.
32. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2013) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.
33. Talkish,J., May,G., Lin,Y., Woolford,J.L. and McManus,C.J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA*, **20**, 713–720.
34. Das,R., Laederach,A., Pearlman,S.M., Herschlag,D. and Altman,R.B. (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA*, **11**, 344–354.
35. Lawley,P.D. and Brookes,P. (1963) Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.*, **89**, 127–138.
36. Lavery,R. and Pullman,A. (1984) A new theoretical index of biochemical reactivity combining steric and electrostatic factors. An application to yeast tRNAPhe. *Biophys. Chem.*, **19**, 171–181.
37. Meyer,B., Wurm,J.P., Kötter,P., Leisegang,M.S., Schilling,V., Buchhaupt,M., Held,M., Bahr,U., Karas,M., Heckel,A. *et al.* (2011) The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Ψ1191 in yeast 18S rRNA. *Nucleic Acids Res.*, **39**, 1526–1537.
38. Oeffinger,M., Wei,K.E., Rogers,R., DeGrasse,J.A., Chait,B.T., Aitchison,J.D. and Rout,M.P. (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat. Methods*, **4**, 951–956.
39. Shenvi,C.L., Dong,K.C., Friedman,E.M., Hanson,J.A. and Cate,J.H.D. (2005) Accessibility of 18S rRNA in human 40S subunits and 80S ribosomes at physiological magnesium ion concentrations–implications for the study of ribosome dynamics. *RNA*, **11**, 1898–1908.
40. Steen,K.-A., Rice,G.M. and Weeks,K.M. (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. *J. Am. Chem. Soc.*, **134**, 13160–13163.
41. Fatica,A., Oeffinger,M., Dlakic,M. and Tollervey,D. (2003) Nob1p is required for cleavage of the 3′ end of 18S rRNA. *Mol. Cell. Biol.*, **23**, 1798–1807.
42. Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
43. Ferreira-Cerca,S., Poll,G., Kuhn,H., Neueder,A., Jakob,S., Tschochner,H. and Milkereit,P. (2007) Analysis of the *in vivo* assembly pathway of eukaryotic 40S ribosomal proteins. *Mol. Cell*, **28**, 446–457.
44. Bindewald,E., Wendeler,M., Legiewicz,M., Bona,M.K., Wang,Y., Pritt,M.J., Le Grice,S.F.J. and Shapiro,B.A. (2011) Correlating SHAPE signatures with three-dimensional RNA structures. *RNA*, **17**, 1688–1696.
45. Thomson,E., Ferreira-Cerca,S. and Hurt,E. (2013) Eukaryotic ribosome biogenesis at a glance. *J. Cell Sci.*, **126**, 4815–4821.
46. Clatterbuck Soper,S.F., Dator,R.P., Limbach,P.A. and Woodson,S.A. (2013) In vivo X-ray footprinting of pre-30S ribosomes reveals chaperone-dependent remodeling of late assembly intermediates. *Mol. Cell*, **52**, 506–516.
47. Hughes,J.M. (1996) Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J. Mol. Biol.*, **259**, 645–654.
48. Fayet-Lebaron,E., Atzorn,V., Henry,Y. and Kiss,T. (2009) 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *EMBO J.*, **28**, 1260–1270.
49. Ulbrich,C., Diepholz,M., Bassler,J., Kressler,D., Pertschy,B., Galani,K., Bottcher,B. and Hurt,E. (2009) Mechanochemical removal of ribosome biogenesis factors from nascent 60S ribosomal subunits. *Cell*, **138**, 911–922.
50. Matsuo,Y., Granneman,S., Thoms,M., Manikas,R.-G., Tollervey,D. and Hurt,E. (2013) Coupled GTPase and remodelling ATPase activities form a checkpoint for ribosome export. *Nature*, **505**, 112–116.
51. Leidig,C., Thoms,M., Holdermann,I., Bradatsch,B., Berninghausen,O., Bange,G., Sinning,I., Hurt,E. and Beckmann,R. (2014) 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nat. Commun.*, **5**, 3491.
52. Ferreira-Cerca,S., Pöll,G., Gleizes,P.-E., Tschochner,H. and Milkereit,P. (2005) Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol. Cell*, **20**, 263–275.
53. Leger-Silvestre,I., Milkereit,P., Ferreira-Cerca,S., Saveanu,C., Rousselle,J.C., Choesmel,V., Guinefoleau,C., Gas,N. and Gleizes,P.E. (2004) The ribosomal protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *EMBO J.*, **23**, 2336–2347.
54. Gamalinda,M., Ohmayer,U., Jakovljevic,J., Kumcuoglu,B., Woolford,J., Mbom,B., Lin,L. and Woolford,J.L. (2014) A hierarchical model for assembly of eukaryotic 60S ribosomal subunit domains. *Genes Dev.*, **28**, 198–210.
55. Ohmayer,U., Gamalinda,M., Sauert,M., Ossowski,J., Pöll,G., Linnemann,J., Hierlmeier,T., Perez-Fernandez,J., Kumcuoglu,B., Léger-Silvestre,I. *et al.* (2013) Studies on the assembly characteristics of large subunit ribosomal proteins in S. cerevisae. *PLoS One*, **8**, e68412.
56. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4569–4574.
57. Campbell,M.G. and Karbstein,K. (2011) Protein-protein interactions within late pre-40S ribosomes. *PLoS One*, **6**, e16194.
58. Ferreira-Cerca,S., Kiburu,I., Thomson,E., LaRonde,N. and Hurt,E. (2014) Dominant Rio1 kinase/ATPase catalytic mutant induces trapping of late pre-40S biogenesis factors in 80S-like ribosomes. *Nucleic Acids Res.*, **42**, 8635–8647.

Downloaded from http://nar.oxfordjournals.org/ at Edinburgh University on December 11, 2014