

**Structural and Functional Analysis of Eukaryotic  
snoRNP Complexes Catalyzing 2'-O-Ribose  
Methylation of rRNA**

Von der Naturwissenschaftlichen Fakultät der  
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades  
Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation  
von  
Simone Höfler, MSc (Österreich)

2020

Referentin: Prof. Dr. Teresa Carlomagno  
Korreferent: Prof. Dr. rer. nat. Wulf Blankenfeldt  
Weiterer Korreferent: Prof. Dr. rer. nat. Claus-Peter Witte  
Tag der Promotion: 17.12.2020

## Summary

Translating the information encoded in messenger RNAs (mRNAs) into functional proteins is an essential cellular process carried out by large molecular machines termed ribosomes. Ribosomes are large ribonucleoprotein (RNP) particles, whose biogenesis is an energetically demanding and highly regulated process. During the early stages of ribosome biogenesis, the ribosomal RNAs (rRNAs) get covalently modified. One of the most abundant of these covalent modifications is the methylation of the 2' hydroxyl group of the ribose (2'-O-Me) in specific nucleotides of the rRNA. Many of these 2'-O-methylated sites are located in functionally important regions of the matured ribosome, such as the peptidyl transferase center (PTC) or the decoding center. Unsurprisingly, aberrations in 2'-O-Me are associated with pathological developments such as cancer and neurological diseases in human. In archaea and eukaryotes 2'-O-Me modifications on rRNA are transferred by the Box C/D enzymes, which are multi-component RNPs that use guide RNAs to mediate site specific 2'-O-methylation on rRNA. Most of the available structural and functional data on the Box C/D RNP enzymes are based on the archaeal enzyme. Conversely, only little is known on the structural and functional details of the eukaryotic Box C/D enzyme. Therefore, the archaeal system is being used as a structural and functional proxy for the eukaryotic enzyme.

To expand to structural and functional knowledge about the eukaryotic Box C/D small nucleolar ribonucleoprotein (snoRNP) enzymes and examine the validity of the archaeal enzymes as a proxy I used a combination of biochemical, analytical and structural methods to analyze and characterize two subcomplexes of the eukaryotic Box C/D snoRNP from *S. cerevisiae in vitro*. Using fluorescence-based electrophoretic mobility shift assays I could characterize the binding requirements and affinities between the eukaryotic and archaeal Box C/D primary RNA-binding protein Snu13 and L7Ae, respectively, and the lesser conserved of two protein binding motifs on the Box C/D guide RNA. I also present the first high-resolution structure of archaeal L7Ae in complex with a non-standard Box C/D protein binding motif solved by X-ray crystallography. Using a combination of quantitative mass spectrometry, multi-angle light scattering and radioactivity-based enzymatic assays I determined the stoichiometries of *in vitro* reconstituted chimeric Box C/D enzymes based on different guide RNAs, demonstrating a potentially different structural arrangement in eukaryotic Box C/D enzymes as compared to archaeal enzymes. Lastly, the high-resolution structure of the eukaryotic 2'-O-methyltransferase Nop1 in complex with the scaffolding protein Nop56 from the Box C/D enzyme solved by X-ray crystallography highlights significant differences to the archaeal orthologs. The presented data expands the structural and functional information on the eukaryotic Box C/D and suggest together with exiting literature substantial differences between eukaryotic and archaeal Box C/D enzymes.

# Zusammenfassung

Die Übersetzung der in Messenger-RNAs kodierten Informationen in Proteine ist ein wesentlicher zellulärer Prozess, der von Ribosomen ausgeführt wird. Ribosomen sind große Ribonukleoprotein (RNP) -Partikel, deren Biogenese ein energetisch anspruchsvoller und stark regulierter Prozess ist. In den frühen Stadien der Ribosomenbiogenese werden die ribosomalen RNAs kovalent modifiziert. Eine der am häufigsten vorkommenden dieser kovalenten Modifikationen ist die Methylierung der 2'-Hydroxylgruppe der Ribose (2'-O-Me) in spezifischen Nukleotiden der rRNA. Viele dieser 2'-O-methylierten Stellen befinden sich in funktionell wichtigen Regionen des Ribosoms. Aberrationen bei 2'-O-Me sind mit Krebs und neurologischen Erkrankungen beim Menschen verbunden. In Archaeen und Eukaryoten werden 2'-O-Me-Modifikationen auf rRNA durch die Box C/D-Enzyme übertragen, bei denen es sich um Mehrkomponenten-RNPs handelt, die Leit-RNAs verwenden, um positionsspezifische 2'-O-Methylierung auf rRNA zu vermitteln. Der Großteil der verfügbaren Daten zu den Box C/D RNP-Enzymen basiert auf dem archaealen Enzymen und nur wenig ist über die strukturellen und funktionellen Details des eukaryotischen Box C/D-Enzyms bekannt, weswegen das archaeale System als Modell für das eukaryotische Enzym verwendet wird.

Um das Wissen über die eukaryotischen Box C/D Enzyme zu erweitern und die Gültigkeit der archaealen Enzyme als Modell zu untersuchen, verwendete ich eine Kombination aus biochemischen, analytischen und strukturellen Methoden, um zwei Subkomplexe des eukaryotischen Box C/D Enzyms von *S. cerevisiae* *in vitro* zu charakterisieren. Unter Verwendung fluoreszenzbasierter elektrophoretischer Mobilitätsverschiebungstests konnte ich die Bindungsanforderungen und -affinitäten zwischen dem eukaryotischen und dem archaealen Box C/D-RNA-Bindeprotein Snu13 bzw. L7Ae und dem weniger konservierten von zwei Proteinbindungsmotiven auf der Box C/D Leit-RNA charakterisieren. Ich präsentiere die erste hochauflösende Struktur von archaealem L7Ae im Komplex mit einem nicht standardmäßigen Box C/D Proteinbindungsmotiv, gelöst mit Röntgenkristallographie. Unter Verwendung einer Kombination aus quantitativer Massenspektrometrie, Mehrwinkellichtstreuung und enzymatischen Assays bestimmte ich die Proteinstöchiometrien von *in vitro* rekonstituierten chimären Box C/D-Enzymen und zeige eine möglicherweise unterschiedliche strukturelle Anordnung in eukaryotischen Box C/D-Enzymen im Vergleich zu archaealen Enzymen. Die hochauflösende Struktur der 2'-O-Methyltransferase Nop1 im Komplex mit Nop56 aus dem eukaryotischen Box C/D-Enzym zeigt signifikante Unterschiede zu den archaealen Orthologen. Die präsentierten Daten erweitern das strukturelle und funktionelle Wissen über eukaryotische Box C/D Enzyme und legen zusammen mit der vorhandenen Literatur erhebliche Unterschiede zwischen eukaryotischen und archaealen Box C/D-Enzymen da.

Keywords: ribosomal RNA, RNA modification, ribonucleoprotein complexes, ribosome biogenesis, 2'-O-Ribose methylation

Schlagwörter: ribosomale RNA, RNA Modifikationen, Ribonukleoproteinkomplexes, Ribosomenbiogenese, 2'-O-Ribose Methylierung

# Abbreviations and Symbols

|        |   |
|--------|---|
| A      | adenine (nucleobase)                                |
| A-site | aminoacyl site                                      |
| ATP    | adenosine triphosphate                              |
| BME    | beta-mercaptoethanol                                |
| C      | cytosine (nucleobase)                               |
| CPM    | counts per minute                                   |
| CTP    | cytidine triphosphate                               |
| DFC    | dense fibrillar components                          |
| DPM    | decays per minute                                   |
| DTT    | dithiothreitol                                      |
| EDTA   | ethylenediaminetetraacetic acid                     |
| eEF*   | eukaryotic elongation factor *                      |
| eIF*   | eukaryotic initiation factor *                      |
| EMSA   | electrophoretic mobility shift assay                |
| eRF*   | eukaryotic release factor *                         |
| ES     | expansion element                                   |
| E-site | exit site   |
| ETS    | external transcribed spacer sequence                |
| FC     | fibrillar centers                                   |
| G      | guanine (nucleobase)                                |
| GC     | granular components                                 |
| GDP    | guanosine-5'-diphosphate                            |
| GTP    | guanosine-5'-triphosphate                           |
| HEPES  | (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HPLC   | high-pressure liquid chromatography                 |
| IPTG   | Isopropyl $\beta$ -D-1-thiogalactopyranoside        |
| L7Ae   | 50S ribosomal protein L7Ae                          |
| LB     | lysogeny broth                                      |
| LC-MS  | liquid chromatography coupled to mass spectrometry  |
| LSC    | liquid scintillation counting                       |
| LSU    | large ribosomal subunit                             |
| MES    | 2-( <i>N</i> -morpholino)ethanesulfonic acid        |
| mRNA   | messenger RNA                                       |
| Nm     | 2'-O-Ribose methylation                             |

|           |                                     |
|-----------|-------------------------------------|
| NTP       | nucleotide triphosphate             |
| OD600     | optical density at 600nm wavelength |
| PCR       | polymerase Chain Reaction           |
| PEG       | polyethylene glycol                 |
| P-Site    | peptidyl site                       |
| PTC       | peptidyl transferase center         |
| RNP       | ribonucleoprotein                   |
| Rpm       | revolutions per minute              |
| r-protein | ribosomal protein                   |
| rRNA      | ribosomal RNA                       |
| SAM       | S-Adenosyl-methionine               |
| SDS       | sodium dodecyl sulfate              |
| snoRNA    | small nucleolar RNA                 |
| snRNA     | small nuclear RNA                   |
| sRNA      | small RNA                           |
| SSU       | small ribosomal subunit             |
| TBE       | Tris/Borate/EDTA                    |
| TC        | ternary complex                     |
| TEV       | tobacco Etch Virus                  |
| tRNA      | transfer RNA                        |
| U         | uracil (nucleobase)                 |
| UTP       | uridine-5'-triphosphate             |
| Utp       | U three protein                     |

# Table of Contents

|  |           |
|--|-----------|
| <b>Summary</b>   | <b>2</b>  |
| <b>Zusammenfassung</b>   | <b>3</b>  |
| <b>Abbreviations and Symbols</b>   | <b>5</b>  |
| <b>1. Introduction</b>   | <b>10</b> |
| 1.1 <i>The eukaryotic ribosome and ribosome biogenesis</i>                           | 10        |
| 1.1.1 The eukaryotic ribosome  | 10        |
| 1.1.2 Translation  | 13        |
| 1.1.3 Eukaryotic ribosome biogenesis   | 17        |
| 1.2 <i>2'-O-Ribose Methylations and Box C/D snoRNPs</i>                              | 20        |
| 1.2.1 2'-O-Ribose Methylations on rRNA   | 20        |
| 1.2.2 Box C/D snoRNPs  | 23        |
| 1.3 <i>Pathologies related to an aberrant rRNA 2'-O-ribose methylation machinery</i> | 32        |
| 1.3.1 snoRNAs  | 32        |
| 1.3.2 SAM-dependent methyltransferase Fibrillarin                                    | 32        |
| 1.4 <i>Aims of this Thesis Work</i>  | 34        |
| <b>2. Methodological Background</b>  | <b>36</b> |
| 2.1 <i>Size-exclusion Chromatography and Multi-angle Light Scattering</i>            | 36        |
| 2.1.1 Size-exclusion Chromatography (SEC)  | 36        |
| 2.1.2 Multi-angle Light Scattering (MALS)  | 38        |
| 2.2 <i>Electrophoretic Mobility Shift Assays</i>                                     | 41        |
| 2.2.1 Principles   | 41        |
| 2.2.2 Applications   | 42        |
| 2.3 <i>Liquid Scintillation Counting</i>   | 43        |
| 2.4 <i>Quantitative Mass Spectrometry</i>  | 46        |
| 2.4.1 Selected Reaction Monitoring (SRM)   | 47        |
| 2.4 <i>X-ray Crystallography</i>   | 49        |
| 2.5.1 Crystallization of biomolecules and biomolecular complexes                     | 49        |
| 2.5.2 Diffraction and data analysis  | 52        |
| 2.5.3 Electron density map and model building  | 55        |
| <b>3. Results</b>  | <b>61</b> |



|   |            |
|---|------------|
| 3.1 <i>Interaction of eukaryotic Snu13 and archaeal L7Ae with non-canonical kink-turns and kink-loops</i>                       | 61         |
| 3.1.1 Binding specificities of archaeal L7Ae and eukaryotic Snu13 to kink-loop and non-canonical kink-turn motifs               | 61         |
| 3.1.2 Six amino acids fine-tune binding of L7Ae and Snu13 to internal kink-turn structures                                      | 65         |
| 3.1.3 Crystal structure of archaeal L7Ae bound to eukaryotic non-standard internal kink-turn structure                          | 70         |
| 3.1.4 Binding stoichiometry in fully assembled chimeric Box C/D s/snoRNP  | 74         |
| 3.2 <i>Molecular basis of Nop1 and Nop56 interaction</i>  | 80         |
| 3.2.1 Eukaryotic methyltransferase Nop1 cannot be complemented by the archaeal ortholog Fibrillarin                             | 80         |
| 3.2.2 Structural differences between eukaryotic Nop1—Nop56 and archaeal Fibrillarin—Nop5 complex                                | 82         |
| 3.3 <i>Activity of fully assembled chimeric Box C/D s/snoRNPs</i>   | 97         |
| <b>4. Discussion</b>  | <b>100</b> |
| 4.1 <i>L7Ae and Snu13 recognize non-standard Box C'/D' motifs differently</i>   | 100        |
| 4.2 <i>The assembly of active eukaryotic Box C/D snoRNPs is not dependent on the binding of Snu13 to the Box C'/D' motif</i>    | 104        |
| 4.3 <i>Interaction between the Box C/D scaffolding proteins and the methyltransferase differ between archaea and eukaryotes</i> | 106        |
| <b>5. Methods and Materials</b>   | <b>110</b> |
| 5.1 <i>Cloning and Mutagenesis</i>  | 110        |
| 5.1.1 Cloning   | 110        |
| 5.1.2 Mutagenesis   | 112        |
| 5.2 <i>Protein expression and purification</i>  | 113        |
| 5.2.1 <i>P. furiosus</i> proteins   | 113        |
| 5.2.2 <i>S. cerevisiae</i> proteins   | 114        |
| 5.3 <i>RNA in vitro transcription and purification</i>  | 115        |
| 5.4 <i>Reconstitution of chimeric Box C/D s/snoRNP complexes</i>  | 118        |
| 5.5 <i>Electrophoretic mobility shift assays (EMSA)</i>   | 119        |
| 5.6 <i>Size-exclusion Chromatography and Multi-angle Light Scattering</i>   | 120        |
| 5.7 <i>X-Ray Crystallography</i>  | 121        |
| 5.7.1 Crystallization   | 121        |
| 5.7.2 Data collection and processing  | 126        |
| 5.7.3 Structure determination   | 127        |

|  |            |
|--|------------|
| <i>5.8 Quantitative Mass spectrometry</i>      | 131        |
| <i>5.9 Activity assays</i>                     | 133        |
| <i>5.10 Materials</i>                          | 134        |
| <b>Appendix 1 – Mass spectrometry raw data</b> | <b>137</b> |
| <b>Appendix 2 – Enzymatic assay raw data</b>   | <b>149</b> |
| <b>Appendix 3 – Cloning Primers</b>            | <b>158</b> |
| <b>Acknowledgments</b>                         | <b>160</b> |
| <b>Bibliography</b>                            | <b>161</b> |
| <b>Curriculum Vitae</b>                        | <b>178</b> |

# 1.Introduction

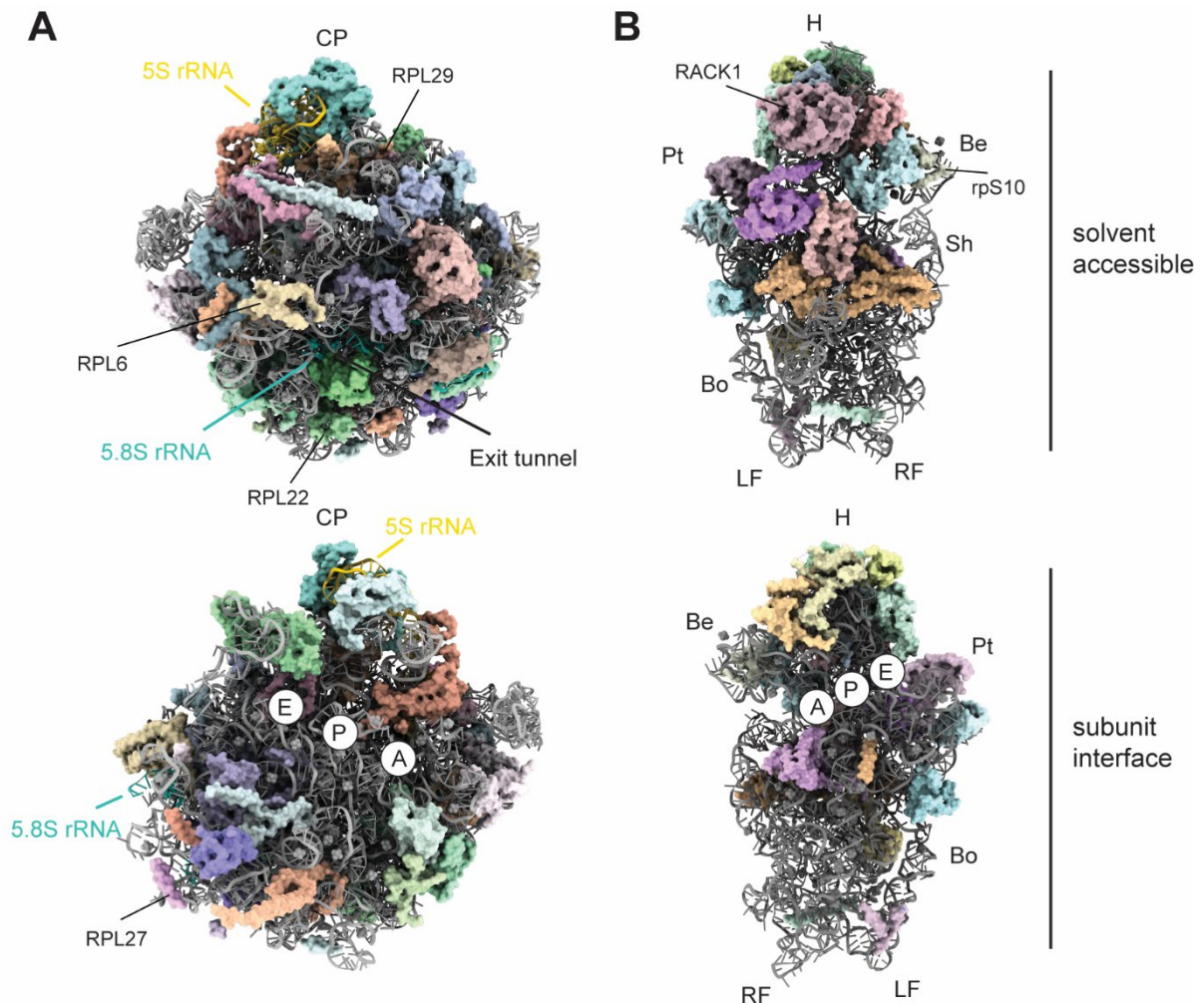
## 1.1 The eukaryotic ribosome and ribosome biogenesis

### 1.1.1 The eukaryotic ribosome

Translation of mRNAs into polypeptide chains (protein synthesis) is a fundamental process in the cellular metabolism of bacteria, archaea, and eukaryotes. This essential function is performed and regulated by large molecular machines termed ribosomes. Ribosomes are ribonucleoprotein complexes that translate the information encoded in the mRNA into a defined polypeptide chain by selecting (decoding) and positioning aminoacyl-tRNAs and catalyzing the formation of a peptide bond between consecutive amino acids (Green and Noller, 1997).

Regardless of the domain of life, all ribosomes consist of two subunits, the small ribosomal subunit (SSU) and the large ribosomal subunit (LSU). In bacteria, the 70S ribosome consists of the 30S subunit (SSU) and the 50S subunit (LSU) (Ramakrishnan, 2002; Yusupov et al., 2001); in eukaryotes, the 80S ribosome consists of the 40S subunit (SSU) and the 60S subunit (LSU) (Klinge et al., 2012a; Wilson and Doudna Cate, 2012). In both bacteria and eukaryotes, the SSU is responsible for binding and decoding the mRNA and monitoring correct base-pairing between the codon in the mRNA and the anti-codon stem-loop in the tRNA. The LSU positions the tRNA acceptor arms and catalyzes the formation of the peptide bond and the transfer of the nascent peptide chain from one tRNA to the next one (Ramakrishnan, 2002). All ribosomes have three sites for tRNA binding to perform an efficient translation: the A (aminoacyl) site, the P (peptidyl) site, and the E (exit) site. New aminoacylated tRNAs enter the ribosome via the A site; the tRNA carrying the nascent peptide chain is located in the P site; the E site holds the deacylated tRNA before it leaves the ribosome. During the entire translation process, which can be divided into initiation, elongation, and termination, both subunits are involved in the precise movement of mRNA and tRNAs through the ribosome, one codon at a time (Ramakrishnan, 2002; Ben-Shem et al., 2010).

Though the fundamental aspects of translation are highly conserved in all kingdoms of life, eukaryotic ribosomes are larger and more complex than their bacterial counterparts (~30 – 40%) (Ramakrishnan, 2002; Klinge et al., 2012b; Thomson et al., 2013). As a consequence, they require a large number of assembly factors during complex interdepartmental biogenesis (Klinge et al., 2012a; Thomson et al., 2013; Woolford and Baserga, 2013).



**Figure 1.1.1.** Architecture of the eukaryotic 80S ribosome from yeast (PDB-ID: 4V7R, Ben-Shem et al., 2010). **(A)** Solvent-accessible and subunit interface view of the yeast 60S subunit (LSU). The 25S rRNA is shown in gray, the 5S rRNA in yellow and the 5.8S rRNA in turquoise. The 46 ribosomal proteins are shown as shapes in various colors. The exit tunnel for nascent peptide, central protuberance (CP) and the A, P and E site are highlighted. **(B)** Solvent-accessible and subunit interface view of the yeast 40S subunit (SSU). The 18S rRNA is shown in gray. The 33 ribosomal proteins in the 40S subunit are shown as shapes in various colors. The head (H), beak (Be), platform (Pt), shoulder (Sh), body (Bo), left foot (LF), right foot (RF) and A-, P- and E- site are highlighted. Eukaryotic specific protein are indicated with their names (RPL6, RPL22, RPL27 and RPL29 in the LSU and rpS10 and RACK1 in the SSU) (Ben-Shem et al., 2010; Klinge et al., 2012b).

As mentioned above, the eukaryotic 80S ribosome consists of two subunits with distinct functions. The large subunit (LSU) or 60S subunit contains three different RNA molecules, the

25S rRNA, the 5.8S rRNA, and the 5S rRNA as well as 46 ribosomal proteins (r-proteins, yeast) (**Figure 1.1.1 A**). The small subunit (SSU) or 40S subunit consists of the 18S rRNA and 33 different r-proteins (yeast) (**Figure 1.1.1 B**) (Ben-Shem et al., 2010; Klinge et al., 2012b).

Besides retaining the fundamental aspects of translation, the eukaryotic ribosome also keeps the general structural elements of the bacterial ribosome. Nevertheless, the 80S ribosome is larger than the 70S ribosome, due to the addition of eukaryotic specific elements.

In the 60S subunit, the highest concentration of eukaryotic specific additions can be found on the solvent-accessible surfaces, whereas the intersubunit interface, the exit region of the nascent peptide chain and the translation factor binding site are mostly conserved from bacteria to eukaryotes (Ben-Shem et al., 2010; Klinge et al., 2011, 2012b; Ramakrishnan, 2002). The 25S of the LSU, which is around 3350 nucleotides long, contains several so-called expansion segments (ES) distinguishing it from the bacterial 23S rRNA. These ES act in mediating intersubunit bridges as well as forming binding and association platforms for eukaryotic specific r-proteins. In addition to 25S rRNA expansion elements, there are six eukaryotic-specific r-proteins in the LSU (RLP6, RLP22, RLP27, RLP28, RLP29, and RLP36, RPL28 and RPL36 are not present in the structure presented in **Figure 1.1.1 A**) and many of the conserved r-proteins carry eukaryotic-specific extension. All eukaryotic-specific proteins are located on the solvent-accessible surface of the subunit (**Figure 1.1.1 A**) (Ben-Shem et al., 2010; Natchiar et al., 2017).

The 40S subunit can be divided into seven different regions: the head, beak, platform, body, shoulder, left foot, and right foot (**Figure 1.1.1 B**). The 18S rRNA consists of ~1750-1800 nucleotides compared to the bacterial 16S rRNA, which is only ~1500 nucleotides long. The additional rRNA nucleotides form the expansion segments ES, as in the 25S rRNA. In the 40S subunit, the ES contribute to the reorganization of the left foot structure and provide a platform for protein binding. The eukaryotic-specific r-proteins rpS7, rpS10, rpS12, and RACK1 are also located on the solvent-accessible surface of the subunit (rpS7 and rpS17 are not present in **Figure 1.1.1 B**). Similarly to the 60S subunit, conserved r-proteins in the 40S also contain eukaryotic-specific extensions (Ben-Shem et al., 2010; Natchiar et al., 2017).

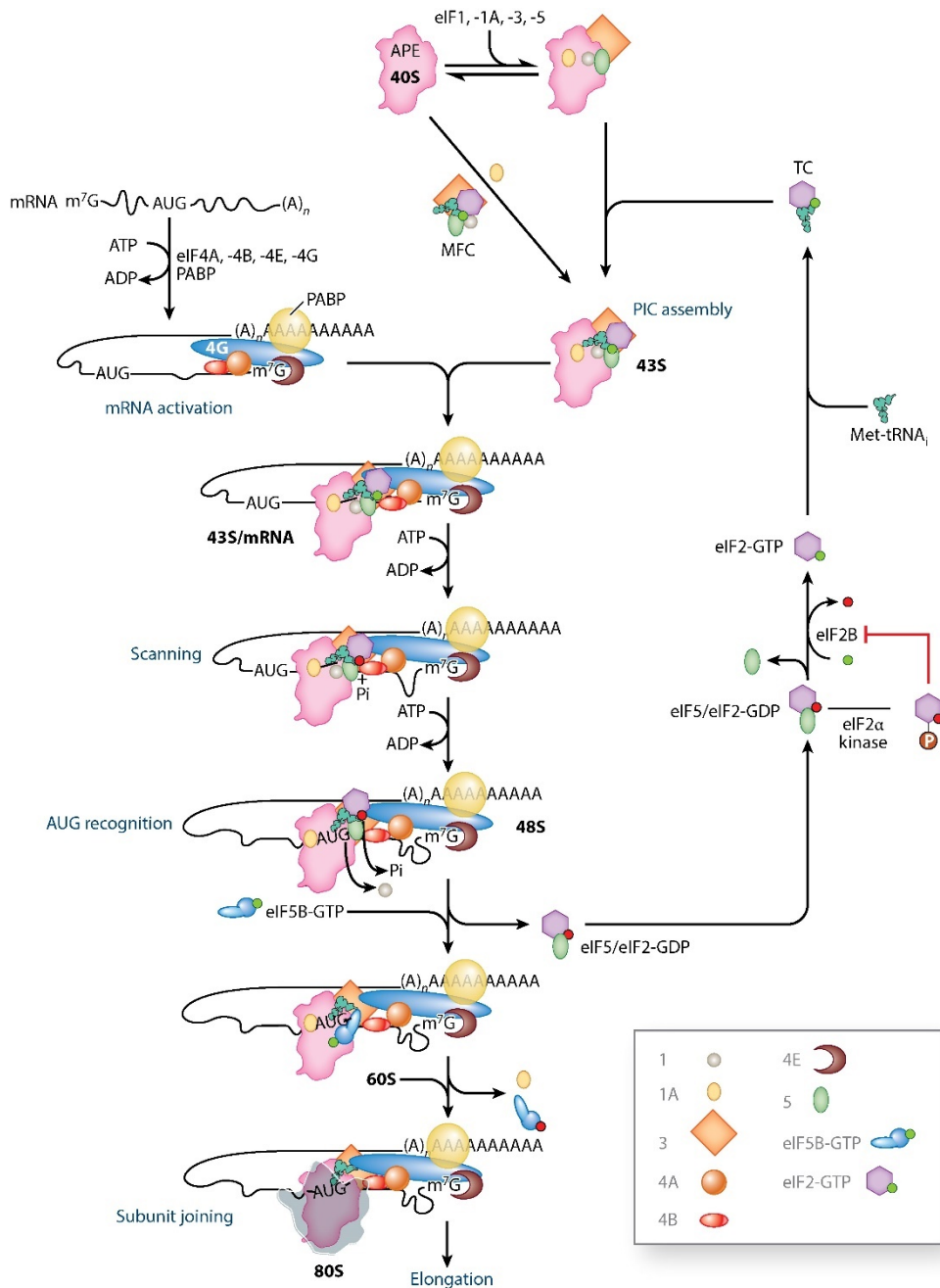
Taken together, the eukaryotic ribosome displays more interconnections between the SSU and the LSU, which are mediated by new r-proteins acquired during evolution, rRNA expansion segments and extensions to conserved r-proteins. This extensive network of connections between different parts of the ribosome is a new feature of the eukaryotic ribosome as compared to archaea and bacteria.

### 1.1.2 Translation

The elucidation of the structural features of the 80S ribosome has also aided the understanding of how the different stages of translation (initiation, elongation, and termination) are regulated.

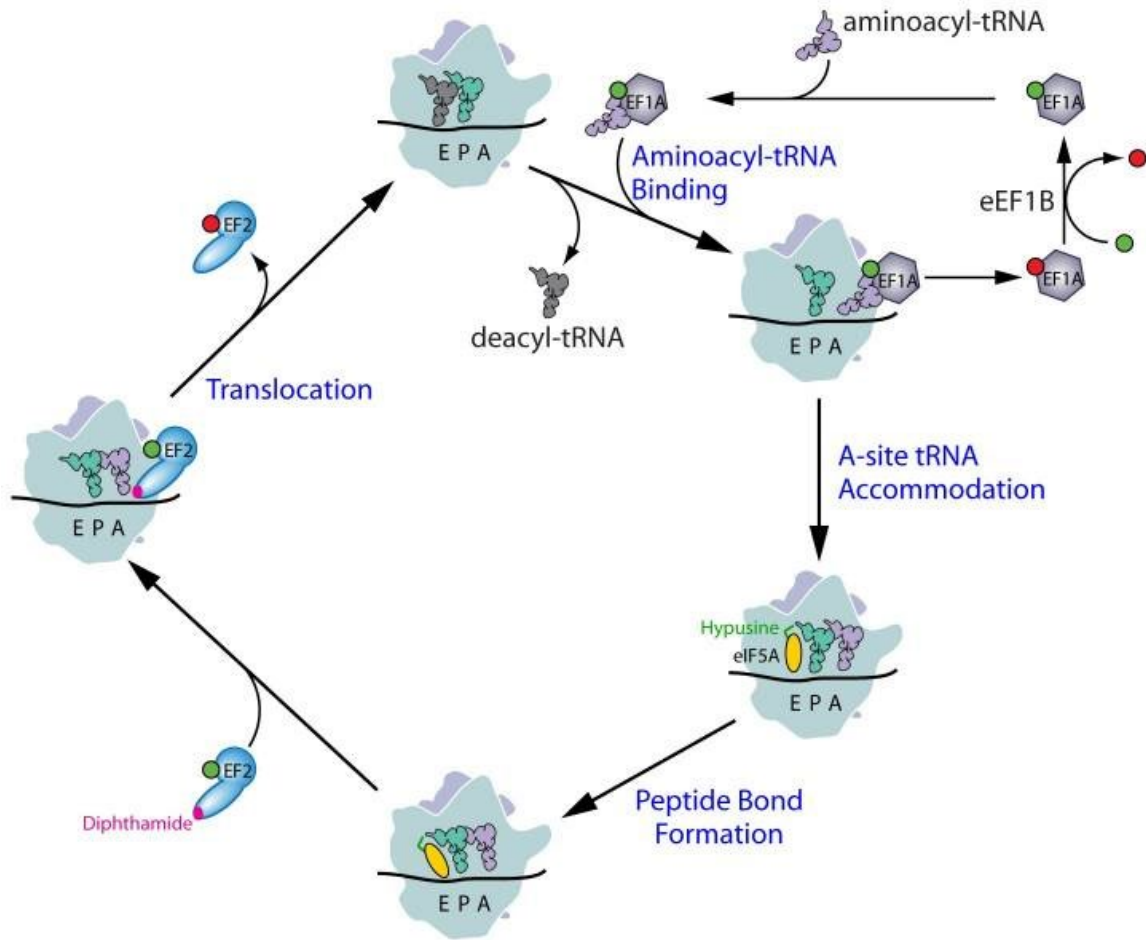
The initiation step differs substantially between bacterial and eukaryotic ribosomes. In bacteria, translation is initiated by the 16S rRNA in the 30S subunit, which base-pairs with the Shine-Dalgarno sequence of the bacterial mRNA, positions the AUG start codon in the P-site of the ribosome and triggers the recruitment of the 50S subunit (Duval et al., 2015). This specific interaction is not present in eukaryotes, where initiation is achieved through a scanning process. The 40S subunit preloaded with the initiator tRNA for methionine (Met-tRNA<sub>i</sub>), the GTP-bound form of initiation factor 2 (eIF2), eIF1, eIF1A, eIF5, and eIF3 form the 43S pre-initiation complex (PIC). The PIC scans the mRNA 5' to 3' codon-for-codon and checks the base-pairing with the Met-tRNA<sub>i</sub> anti-codon until the AUG start codon is reached (**Figure 1.1.2**) (Hinnebusch, 2014). Thus, the first encountered AUG codon is favored for initiation; however, specific sequences directly adjacent to the AUG have been shown to enhance their selection as start codon (Kozak, 1986). After a successful encounter with the start codon, eIF1 leaves the 40S subunit and eIF5, eIF1A, eIF2 $\beta$ , and eIF3c rearrange to accommodate the Met-tRNA<sub>i</sub> in the P site. Lastly, eIF5/eIF2-GDP dissociates, and recruitment of the 60S subunit is catalyzed by eIF5B to form the 80S initiation complex (IC), which can then initiate peptide synthesis from the mRNA template (**Figure 1.1.2**) (Hinnebusch, 2014).

After successful translation initiation, the elongation phase of translation commences. At this point, the Met-tRNA<sub>i</sub> is in the P-site of the ribosome, and the second codon of the mRNA is located in the A-site. The cognate aminoacyl-tRNA binds to the A-site together with the elongation factor eEF1A-GTP; the correct interaction of the anti-codon and the codon induces GTP hydrolysis, eEF1A-GDP leaves the ribosome and the tRNA in positions in the A-site (Dever et al., 2018). Three nucleotides of the 18S rRNA ensure that only cognate tRNAs are recruited during elongation. In *S. cerevisiae* A1755, A1756 and G577 interact with the minor groove of the codon-anticodon helix and stabilize the tRNA in the A-site (Loveland et al., 2017; Ogle et al., 2001; Shao et al., 2016). These three nucleotides also sense the presence of 2'-O-methylations in the mRNA, which leads to translation down-regulation (Choi et al., 2018; Elliott et al., 2019). Once the new tRNA is correctly placed in the A-site, peptide bond formation between the amino acids attached to the Met-tRNA<sub>i</sub> or peptidyl-tRNA and the new tRNA in the A-site proceeds rapidly. This process leaves the nascent peptide chain attached to the tRNA in the A-site while the tRNA at the P-site is no longer loaded.



**AR** Hinnebusch AG. 2014.  
 Annu. Rev. Biochem. 83:779–812

**Figure 1.1.2.** Schematic representation of translation initiation in eukaryotes. The 40S subunit associates with eIF1, -1A, -3, -5 and the multifactor complex (MFC) including the Met-tRNA<sub>i</sub> and form the 43S pre-initiation complex (PIC). The PIC is recruited to the mRNA that has been bound by eIF4A, -4b, -4E, -4G and the poly(A)-binding protein (PABP) in a process termed mRNA activation. The PIC scans the mRNA 5' to 3' for a AUG start codon. Once the start codon is found eIF1 and eIF2-GDP/eIF5 leaves the PIC and eIF5B-GTP catalyzes the recruitment of the 60S subunit to form the 80S initiation complex (IC). eIF5/eIF2-GDP is recycled for the formation of a new ternary complex (TC). This figure is adapted from (Hinnebusch, 2014).



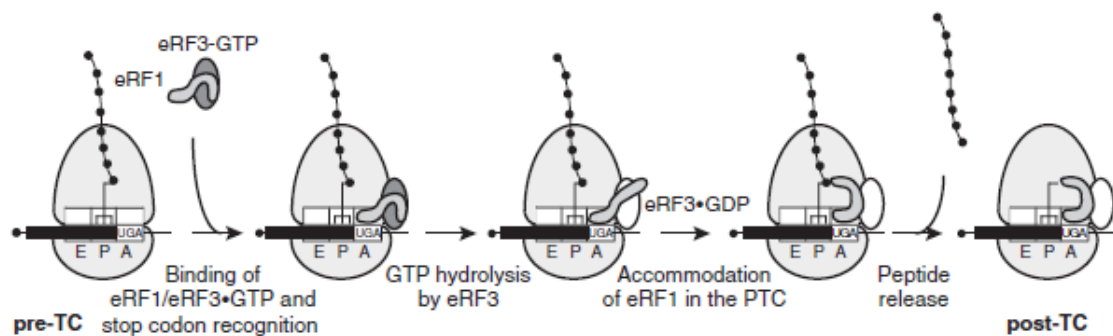
**Figure 1.1.3.** Schematic representation of the translation elongation cycle in eukaryotes. Elongation is initiated by binding of a new aminoacyl-tRNA to the A-site in complex with elongation factor 1A (eEF1A). Dissociation of eEF1A leads to accommodation of the aminoacyl-tRNA in the A-site and peptide bond formation occurs shortly after, leading to the transfer of the nascent peptide chain to the tRNA in the A-site. Translocation of both tRNAs is finalized by elongation factor 2 (eEF2) and the release of the deacylated -tRNA. A new cycle of elongation can begin. This figure is adapted from (Dever et al., 2018).

The reaction occurs in the peptidyl transferase center (PTC) in the 60S subunit, which is composed of highly conserved rRNA elements, which act as a ribozyme (Ben-Shem et al., 2010; Natchiar et al., 2017). This ribozyme functions by positioning the aminoacyl- and peptidyl-tRNAs for catalysis, while eIF5A binds to the E-site and is also thought to promote catalysis (Bs et al., 2017; Gutierrez et al., 2013). Because of the peptide bond formation, both tRNAs are shifted into hybrid positions, and the ribosomal subunits rotate towards each other. Right after peptide bond formation the rotation has not yet happened and the two tRNAs stay in the respective positions in the A and P sites. In the first rotated state, the now unloaded tRNA in the P-site remains base-pairs with the P-site, but the acceptor arm shifts to the E-site; the tRNA carrying the nascent peptide chain (peptidyl-tRNA) stays in the A-site. In the second



rotated state, the unloaded tRNA remains in the same hybrid position between E- and P-site, whereas the acceptor arm of the peptidyl-tRNA shifts in the P-site while the tRNA remains base-paired to the A-site (Moazed and Noller, 1989). Final translocation of the unloaded and peptidyl-tRNAs to the E- and P-sites, respectively, is aided by eEF2. eEF2 binds to the A-site and helps release the base-pairing and decoding interactions of the peptidyl-tRNA (Abeyrathne et al., 2016; Murray et al., 2016). After translocation and release of eEF2, elongation continues with another cycle of the described steps (**Figure 1.1.3**). It is yet not fully understood how the unloaded tRNA is released from the E-site in eukaryotes.

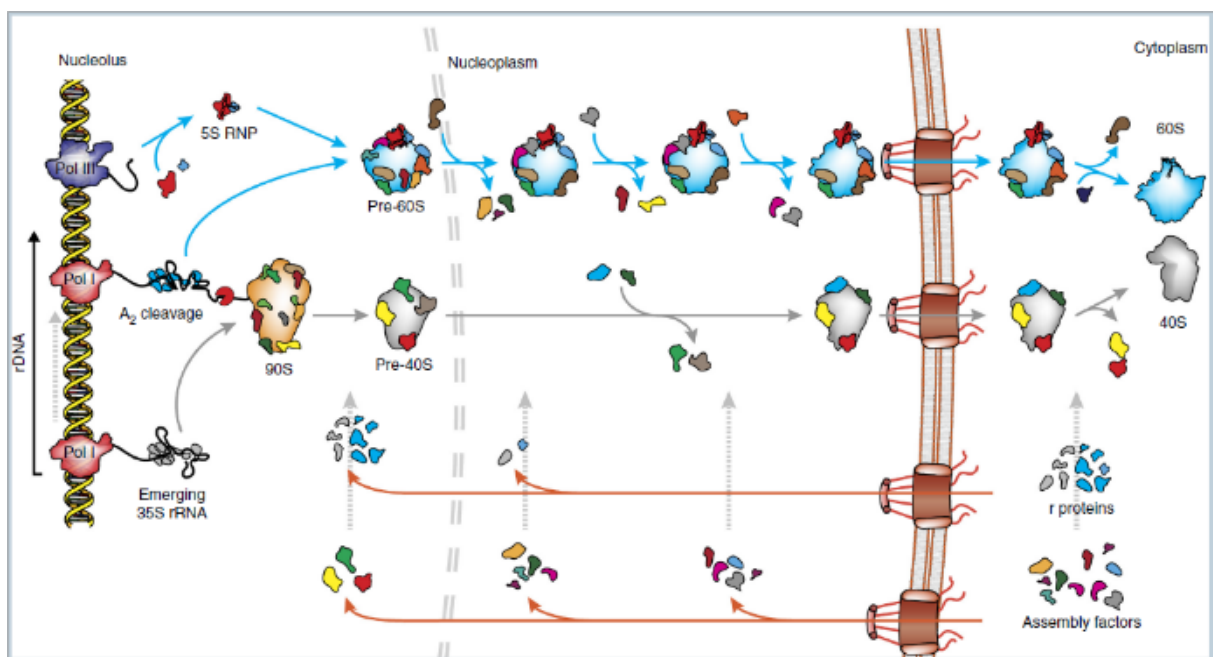
The elongation continues until a stop codon is encountered on the mRNA. Once any of the three stop codons (UAG, UGA, and UAA) occurs, the translation termination is initiated by the release factors eRF1 and eRF3-GTP. eRF1 can recognize all three stop codons and releases the new peptide chain, with the help of the eRF3 GTPase. After the release of the polypeptide from the post-termination complex (post-TC), the ribosome is separated into the 40S and 60S subunits with the help of the ATP-binding cassette sub-family E member 1 (ABCE1) and the unloaded tRNA and mRNA are released (Hellen, 2018) (**Figure 1.1.4**). Both subunits are recycled for translation of a new mRNA transcript.



**Figure 1.1.4.** Schematic representation of translation termination in eukaryotes. Upon entrance of a stop codon in the A-site of the ribosome (pre-termination complex, pre-TC) the termination of translation is initiated by release factors 1 and 3 (eRF1 and eRF3-GTP). eRF1 recognizes any of the three stop codons and releases also the newly synthesized polypeptide chain with the help of eRF3. This results in the post termination complex (post-TC), which is then disassembled into the ribosomal subunits, free deacyl-tRNA and mRNA with the help of ABCE1. This figure is adapted from (Hellen, 2018).

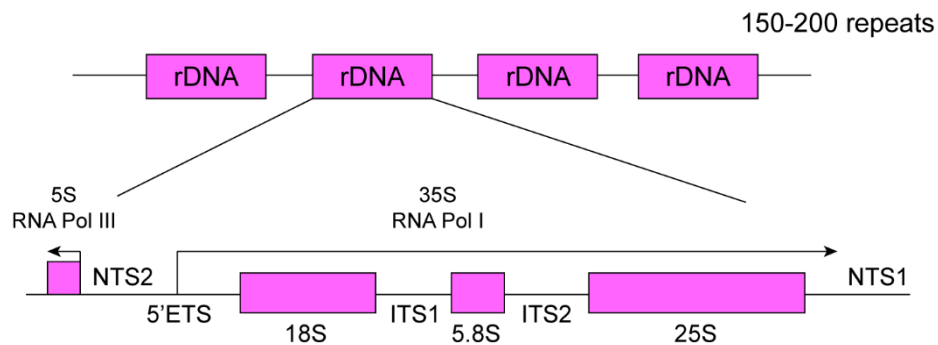
### 1.1.3 Eukaryotic ribosome biogenesis

To perform the essential process of translating information encoded in mRNAs into functional proteins, a high-level of accuracy, coordination, and reaction rate is necessary. Therefore, functional ribosomes require all their components to be positioned correctly within this large molecular machine, both temporally and spatially. To ensure this, a highly coordinated and intricate assembly and maturation process has evolved in eukaryotes, which requires more than 200 assembly factors. Ribosome maturation occurs in three different cellular compartments; the nucleolus, the nucleoplasm, and the cytosol (**Figure 1.1.5**). The nucleolus is a sub-structure of the eukaryotic nucleus, which is mainly dedicated to ribosome biogenesis. It assembles around tandem repeats of ribosomal DNA (rDNA) genes and can be divided into three distinct sub-regions; the fibrillar centers (FCs), the dense fibrillar components (DFCs), and the granular components (GCs). The transcription of the 35S pre-rRNA by RNA polymerase I (Pol I) occurs at the interface between FC and DFC.



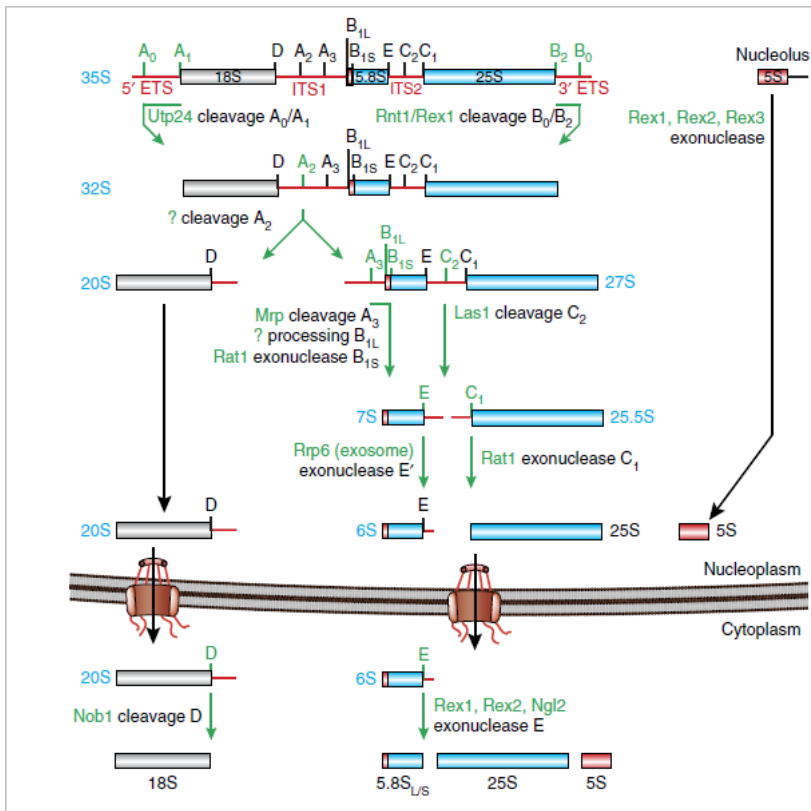
**Figure 1.1.5.** Schematic representation of the assembly and maturation pathway of the eukaryotic ribosome. The initial step of eukaryotic ribosome biogenesis is the transcription of the 35S pre-rRNA from the rDNA locus in the nucleolus by RNA polymerase I (Pol I). The nascent 35S pre-rRNA associates with several snoRNAs and 40S-specific proteins to form the 90S pre-ribosome. Cleavage at the A<sub>2</sub> site separates then the 40S from the 60S subunit. The 60S associates also with the 5S rRNA, transcribed by RNA polymerase III (Pol III) in the nucleolus and continues its maturation process in the nucleoplasm before being exported the cytoplasm. In contrast, the 40S is rapidly exported to the cytoplasm from the nucleolus and matures there. This figure is adapted from (Peña et al., 2017).

The 35S pre-rRNA is then processed in the DFC, where the small nucleolar ribonucleoprotein particles (snoRNPs) that mediate the cleavage and modification of rRNAs are located. These processing steps, together with maturation processes in the nucleoplasm and the cytoplasm, yield the mature 25S, 5.8S, and 18S rRNAs (**Figure 1.1.6**) (Boisvert et al., 2007; Woolford and Baserga, 2013; Peña et al., 2017).



**Figure 1.1.6.** Genomic organization of ribosomal DNA genes and resulting rRNA transcripts. Ribosomal DNA genes are arranged in tandem repeats, which are cluster in a nuclear sub-structure termed the nucleolus. The 35S pre-rRNA is transcribed by dedicated RNA polymerase I (RNA Pol I), whereas the 5S rRNA is transcribed by RNA polymerase III (RNA Pol III). The 35S pre-rRNA and the 5S rRNA are separated by the non-transcribed spacer 2 (NTS2). The 18S rRNA is preceded by the 5' external transcribed spacer sequence (5' ETS). The 18S and 5.8S rRNAs are separated by the internal transcribed spacer sequence (ITS1), whereas the 5.8S and 25S rRNAs of the 60S subunit are separated by the ITS2.

The nascent 5' external transcribed spacer sequence (5' ETS) of the 35S pre-rRNA has been shown to first associate with the UtpA protein complex, which prepares the recruitment of the U3 snoRNP and the UtpB complex (Hunziker et al., 2016). Followed by the recruitment of a plethora of additional assembly factors, r-proteins, and pre-rRNA rearrangements, the intermediate 90S pre-ribosome or small subunit processome is formed (Barandun et al., 2017). This structural arrangement also contains the nuclease Utp24, which has been proposed to catalyze the cleavage at the  $A_0$  and  $A_1$  site (**Figure 1.1.7**) (Bleichert et al., 2006). The cleavage at the  $A_2$  site that separates the 20S pre-rRNA, which at this stage is already embedded into the pre-40S subunit, happens co-transcriptionally in the eukaryotic model system *S. cerevisiae*. In parallel, the 35S pre-rRNA and cleavage products are being modified by numerous snoRNPs present in the DFC co- and post-transcriptionally, predominantly Box C/D snoRNPs and Box H/ACA snoRNPs. The two most abundant covalent rRNA modifications transferred by snoRNPs are the methylation of the hydroxyl group in the 2' position of the ribose (2'-O-Me), by Box C/D snoRNPs, and the isomerization of uridine to pseudouridine, by H/ACA snoRNPs (Penzo and Montanaro, 2018; Ayadi et al., 2019). After its release from the 90S pre-ribosome, the pre-40S subunit, gains nuclear export competence and is exported into the cytoplasm to undergo the final maturation steps.



**Figure 1.1.7.** Ribosomal RNA processing pathways in eukaryotes. The 35S pre-rRNA is cleaved and processed in several consecutive steps. In the first processing steps the 5' and 3' ETS (external transcribed spacer) are removed by cleavage at A<sub>1</sub> and B<sub>2</sub>. Consecutively the 18S rRNA is separated from the 5.8S and 25S rRNA by cleavage at the A<sub>2</sub> site. The 5.8S and 25S rRNA get separated by cleavage at the C<sub>2</sub> site. Final maturation happens after nuclear export in the cytoplasm. Cleavage steps occurring at specific stages and enzymes are indicated in green. Pre-rRNAs are indicated in blue. This figure is adapted from (Peña et al., 2017).

The release of the pre-40S particle after A<sub>2</sub> cleavage allows the remaining 27S pre-rRNA to start assembling in the pre-60S subunit. The exact pathway of pre-60S assembly is yet not well understood, but experimental evidence suggests that the recruitment of r-proteins and assembly factors happens both co-transcriptionally and post-transcriptionally (Chen et al., 2017). Likewise, a multitude of snoRNPs, RNA helicases, and RNA-processing and -modifying enzymes, as well as binding of the 5S RNP, are needed for pre-60S assembly and nuclear export.

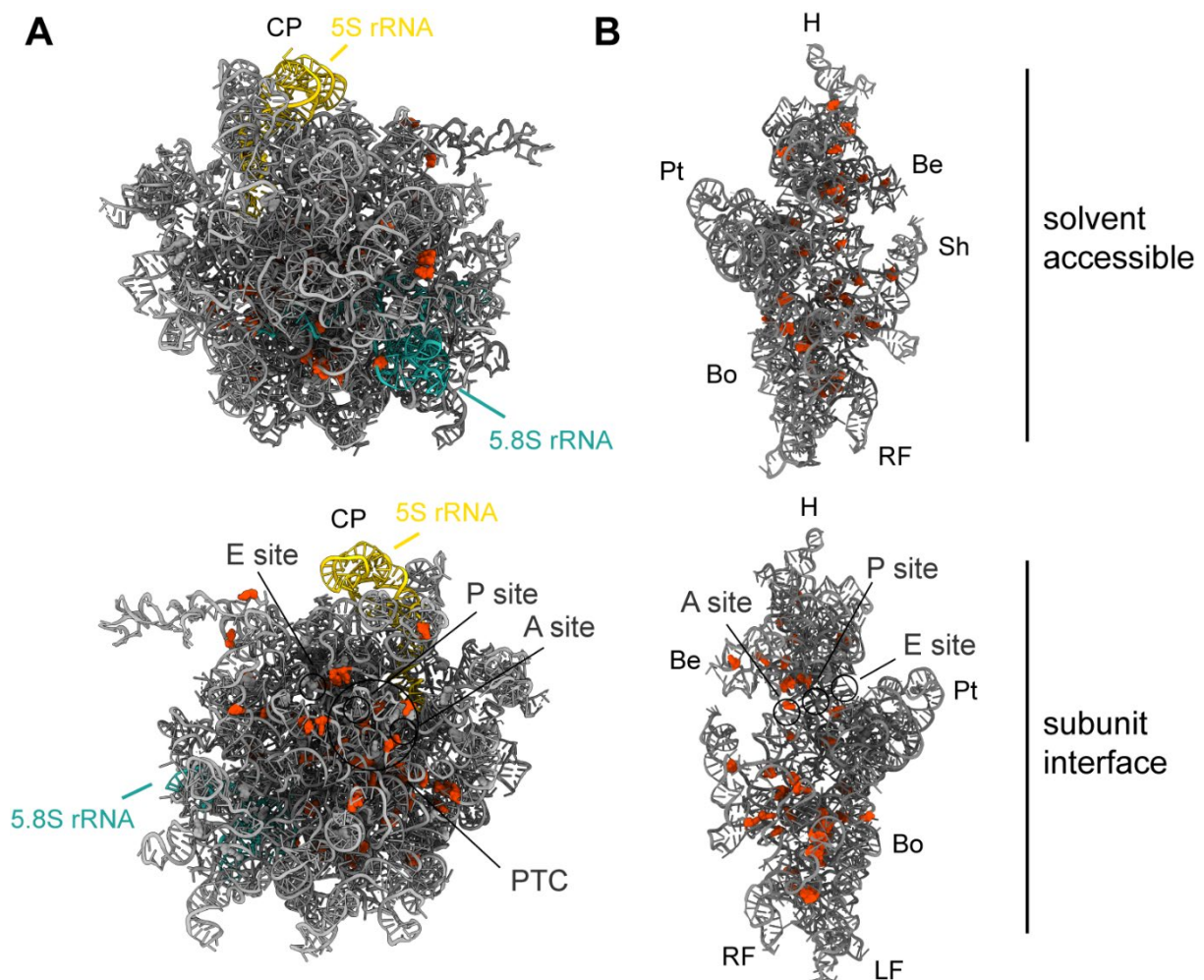
After nuclear export, both pre-subunits undergo final maturation steps. These include the association with additional r-protein, release of assembly and transport factors, and final rRNA cleavage (**Figure 1.1.5 and 1.1.7**). Before reaching translation competence, functional proofreading is performed in the cytoplasm (Konikkat and Woolford, 2017).

Once both subunits are entirely matured, they can assemble around an mRNA template to form the functional and translation-competent 80S ribosome.

## 1.2 2'-O-Ribose Methylations and Box C/D snoRNPs

### 1.2.1 2'-O-Ribose Methylations on rRNA

As briefly mentioned in the previous section, during the early stages of ribosome biogenesis in the nucleolus, the 18S, 5.8S, and 25S rRNAs are covalently modified both co- and post-transcriptionally by specific snoRNPs. One of the most abundant modifications installed at this stage is the methylation of the 2' hydroxyl group of the ribose sugar (2'-O-Me, Nm). In *S. cerevisiae*, 55 2'-O-Me sites are currently characterized, 18 (8 Am, 3 Cm, 6 Gm, and 1 Um)



**Figure 1.2.1.** 2'-O-Ribose methylated nucleotides in the 5.8S, 18S and 25S rRNAs of the human ribosome (PDB-ID: 6EK0, Natchiar et al., 2017). **(A)** The 28S (gray), 5.8S (turquoise) and 5S (yellow) rRNAs of the 60S subunit are displayed from the solvent-accessible and intersubunit surface. Identified 2'-O-Me sites are highlighted in red. **(B)** The 18S (gray) rRNA of the 40S subunit is displayed from the solvent-accessible and intersubunit surface. Identified 2'-O-Me sites are highlighted in red. A-, P- and E-site as well as the PTC are indicated in both panels. Central protrusion (CP) head (H), beak (Be), platform (Pt), shoulder (Sh), body (Bo), light foot (LF) and right foot (RF) are indicated.

of which are located in the 18S rRNA in the 40S subunit and 37 (12 Am, 7 Cm, 10 Gm, and 8 Um) in the 25S rRNA in the 60S subunit (Birkedal et al., 2015; Marchand et al., 2016; Yang et al., 2016a). No 2'-O-Me sites are reported for the 5.8S and 5S rRNAs.

In human ribosomes, 110 potential 2'-O-Me sites have been proposed. The 18S, 28S and 5.8S rRNAs carry 41, 67 and 2 characterized 2'-O-Me sites, respectively (Nazar et al., 1980; Krogh et al., 2016; Eroles et al., 2017; Incarnato et al., 2017; Sharma et al., 2017a; Zhu et al., 2017; Taoka et al., 2018).

In both yeast and human, the majority of the 2'-O-Me modifications are located in functionally crucial areas and the intersubunit interface (**Figure 1.2.1**). These include the decoding center (A-site and P-site) in the 18S rRNA of the 40S subunit, and the PTC in the 25S or 28S rRNA in the 60S subunit (Natchiar et al., 2017).

Strikingly, two or three 2'-O-Me sites have been reported in yeast and human mitochondrial rRNAs. These modifications (Gm2270 and Um2791 or Gm1145, Um1369, and Gm1370 in yeast or human, respectively) are all located in the functionally crucial PTC (Baer and Dubin, 1981; Sirum-Connolly et al., 1995; Lee et al., 2013; Lee and Bogenhagen, 2014).

The detection of the 2'-O-Me modification sites has been made possible by new technological advances that also allowed the detection and quantification of the 2'-O-Me modification.

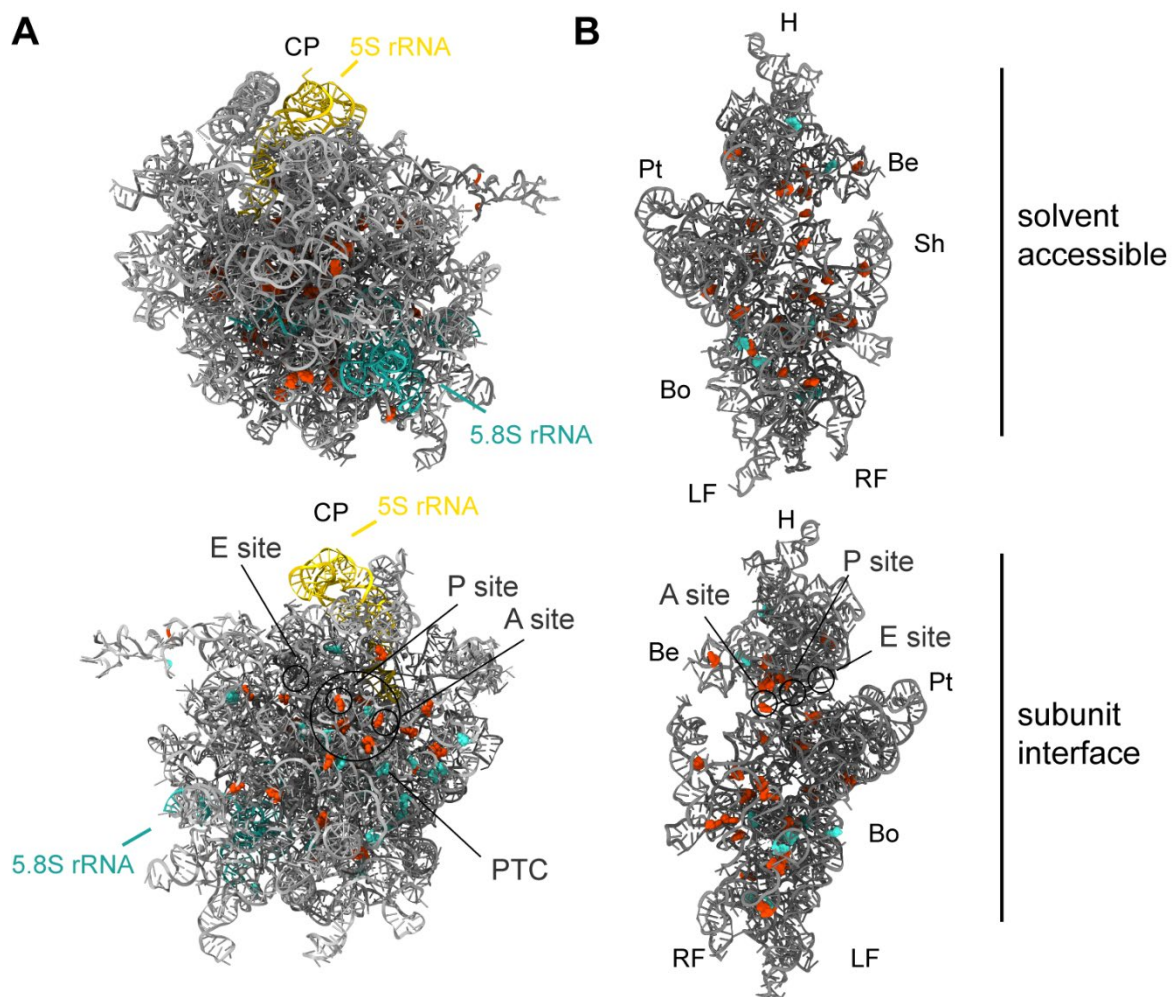
RiboMethSeq is a high-throughput method exploiting the increased resistance of 2'-O-methylated RNA towards alkaline treatment. It has been used to map and quantify 2'-O-methylation of human rRNA (Krogh et al., 2016; Eroles et al., 2017; Sharma et al., 2017a).

2'-O-Me-Seq exploits the different processivity of the reverse transcriptase through 2'-O-methylated nucleotides in the presence of a high or low concentration of nucleotide triphosphates. 2'-O-Me-Seq is more precise than RiboMethSeq in that it does not generate unspecific stops at pseudouridine positions. The method, combined with massive parallel sequencing, can be used in high-throughput experiments for rRNA (Incarnato et al., 2017).

From a chemical perspective, the methyl group at the 2'-position abolishes the ability of the ribose to engage in hydrogen bonding interactions at this position. However, it also removes the nucleophilic character, which increases the chemical stability of the RNA against alkaline hydrolysis (Ayadi et al., 2019). This fact is exploited in the above mentioned RiboMethSeq detection method. Furthermore, the 2'-O-Me stabilizes the 3' endo conformation of the ribose sugar, which in turn favors the formation of an A-type RNA helix (Prusiner et al., 1974).

2'-O-Me modification levels are not equal at all positions in the rRNA and also vary between different cell types. Recent studies showed that the methylation levels of one-third of all identified sites vary between different cell types in humans (Krogh et al., 2016; Eroles et al.,

2017). In addition, it was demonstrated that methylation sites are differently affected by the downregulation of the methylation machinery. Interestingly, sites that have previously been shown to be hypomethylated under physiological conditions are also heavily affected by the downregulation of the methyltransferase (**Figure 1.2.2**) (Sharma et al., 2017a). These findings indicate that ribosomes are not homogeneous entities, as it had been previously thought, and support the hypothesis of the existence of “specialized” ribosomes optimized for the expression of specific genes. Altering the 2'-O-Me methylation levels during ribosome biogenesis in response to cell internal or external stimuli may be a means to generate such “specialized” ribosomes.



**Figure 1.2.2.** Differentially modified 2'-O-methylated nucleotides in the 5.8S, 18S and 25S rRNAs of the human ribosome (PDB-ID: 6EK0, Natchiar et al., 2017). **(A)** The 28S (gray), 5.8S (turquoise) and 5S (yellow) rRNAs of the 60S subunit are displayed from the solvent-accessible and intersubunit surface. **(B)** The 18S (gray) rRNA of the 40S subunit is displayed from the solvent-accessible and intersubunit surface. Nucleotides showing 100% methylation are shown in red, whereas nucleotides showing variable methylation levels are shown in cyan. 2'-O-Me levels are taken from (Erales et al., 2017). A-, P- and E-site as well as the PTC are indicated in both panels. Central protrusion (CP) head (H), beak (Be), platform (Pt), shoulder (Sh), body (Bo), light foot (LF) and right foot (RF) are indicated.

Within eukaryotes, the positions that carry 2'-O-Me modifications are well conserved, especially in functionally essential areas, such as the decoding center and the PTC. The level of conservation also suggests that these 2'-O-Me play a functionally important role in ribosome biogenesis and translation fidelity (Ayadi et al., 2019). Nevertheless, the exact biological role of 2'-O-Me is still mostly unknown, with a few exceptions. For the yeast decoding center (A- and P-site), the removal of several 2'-O-Me modifications, together with some pseudouridines, was shown to impair the growth rates, impact the translation, and reduce the amount of free 40S subunit (Liang et al., 2007, 2009; Baudin-Baillieu et al., 2009). In zebrafish, the deletion of single 2'-O-Me modification caused severe morphological defects and embryonic lethality (Higa-Nakamine et al., 2012). In bacteria, it was shown that 2'-O-Me modifications alter the folding of rRNA and help recruit r-proteins during biogenesis (Arai et al., 2015). Furthermore, some 2'-O-Me modifications on rRNA have been associated with leukemia (Zhou et al., 2017; Pauli et al., 2020).

### 1.2.2 Box C/D snoRNPs

In eukaryotes and archaea, the Box C/D snoRNP and sRNP complexes, respectively, are the enzymes that transfer the methyl group to well-defined 2'-O-positions of rRNAs (Kiss-László et al., 1996; Omer et al., 2000). In bacteria, this function is carried out by stand-alone specialized protein enzymes.

While it is possible to reconstitute active archaeal Box C/D sRNPs from various species *in vitro*, the reconstitution of a functional and homogeneous eukaryotic Box C/D snoRNP enzyme has failed so far. Thus, multiple high-resolution structures and functional studies are available for the archaeal enzymes (Bleichert et al., 2009; Ye et al., 2009; Xue et al., 2010; Lin et al., 2011; Lapinaite et al., 2013; Yang et al., 2016b; Yip et al., 2016), whereas for eukaryotic Box C/D snoRNPs only the structure of the non-methylating U3 snoRNP (*S. cerevisiae*) is available as well as crosslinking data from the U25 snoRNP (*Xenopus leavis*) (Cahill et al., 2002; Hunziker et al., 2016; Barandun et al., 2017).

In eukaryotes Box C/D guide snoRNP complexes consist of a Box C/D guide snoRNA, the primary RNA-binding protein Snu13, two scaffolding proteins Nop56 and Nop58 and the SAM-dependent methyltransferase Nop1 (yeast) or Fibrillarin (humans) (Tollervey et al., 1991; Kiss-László et al., 1996; Gautier et al., 1997; Cahill et al., 2002). Similarly, in archaea the Box C/D sRNP is comprised of a guide Box C/D sRNA, the primary RNA-binding protein L7Ae, the scaffolding protein Nop5 and the SAM-dependent methyltransferase Fibrillarin (Omer et al., 2000).

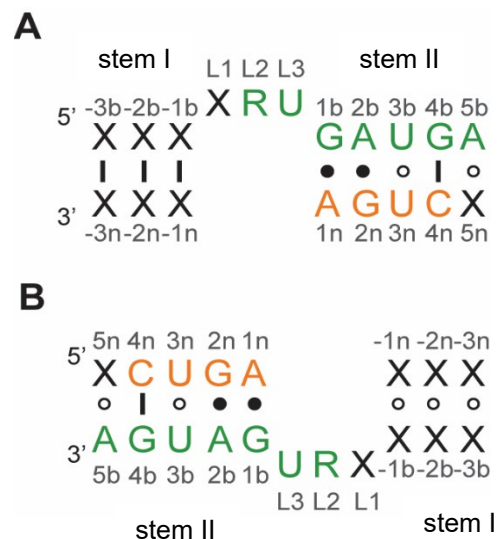


The primary function of the Box C/D guide RNAs within the snoRNP is to determine the methylation site. This function is mediated by specific complementary regions within the guide snoRNA that base pair with the corresponding sequences of the rRNA both co- and post-transcriptionally (Cavaillé et al., 1996; Kiss-László et al., 1996).

Not all snoRNAs classified as Box C/D guide RNAs are responsible for guiding RNA 2'-O-methylation. For example, the U3 Box C/D snoRNA aids the formation of the 90S pre-ribosome by base-pairing with the 5'ETS of the nascent 18S rRNA (Hunziker et al., 2016; Barandun et al., 2017). Other Box C/D snoRNA, like yeast snR4 and snR45, have been shown to guide

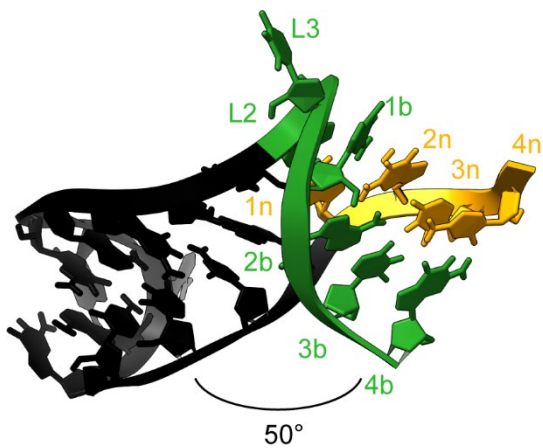
RNA acetylation rather than methylation (Sharma et al., 2017b). All eukaryotic, as well as archaeal, Box C/D guide RNAs share the same conserved motifs, regardless of their function. These are the Box C (5'RUGAUGA, where R is either A or G) and the Box D (5'CUGA) motif (Figure 1.2.3). The Box C and D motif are located at the 5' and 3' end of the guide RNA, respectively and

**Figure 1.2.3.** Secondary structure representation of the Box C and D motif (A) forming a k-turn structure and (B) the internal Box C' and D' motif forming a kinked structure. Full lines between bases indicate Watson-Crick base pairs, full circles indicate sheared base pairs and empty circles indicates other base pairs or non-defined interactions. R is an A or G and X is any nucleotide.



together form the Box C/D motif. Both motifs have also been shown to be required for nucleolar localization of the RNAs (Samarsky et al., 1998). The Box C/D motif in complex with protein or in the presence of Mg<sup>2+</sup> folds in a structural element termed the kink-turn (k-turn) (Figure 1.2.3 A and 1.2.4). This structural element, which is characterized by a typical 50° kink of the RNA backbone is found in almost all RNA classes (Huang and Lilley, 2018).

Copies of Box C and D motifs are located in the central part of the guide RNA, termed Box C' and Box D' (Figure 1.2.3 B) (Kiss-László et al., 1996; Omer et al., 2000). The archaeal Box C' and D' motifs interact similarly to the Box C and D motifs and hence also form a kinked RNA structure (also termed k-turn). In archaea, the stem I next to Box C'/D' is often substituted by a loop; thus, in archaea the Box C'/D' element folds in a so-called kink-loop (k-loop) structure (Figure 1.2.3 B).



**Figure 1.2.4.** Three-dimensional structure of the Box C/D k-turn from *P. furiosus* guide RNA sR26 (PDB-ID: 6TPH, Ahmed et al., 2020). The Box C and D motifs are in green and orange, respectively.

The primary RNA-binding protein Snu13 binds the Box C/D guide snoRNA at the k-turn motif formed by Box C and D (Marmier-Gourrier et al., 2003; Gagnon et al., 2010). This binding event is thought to be the initial step of the Box C/D snoRNP assembly process (Massenet et al., 2017).

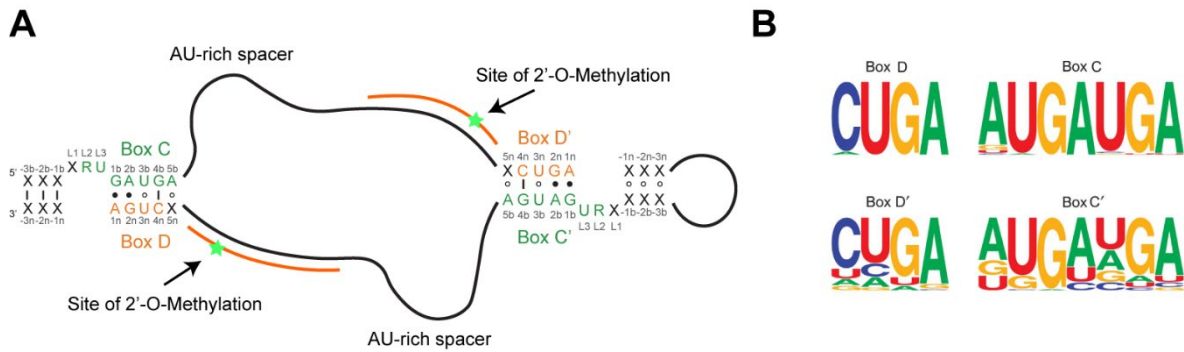
The two scaffolding proteins Nop56 and Nop58 are modular proteins consisting of an N- and C-terminal domain and a central coiled-coil domain. The two proteins form a heterodimer through their central coiled-coil domains. The C-terminal domain (of Nop58, at least) interacts with the

composite surface formed by Snu13 and the guide RNA (Barandun et al., 2017) and the N-terminal domains of both proteins recruit the methyltransferase Nop1 (yeast) or Fibrillarin (human) to the snoRNP (Gautier et al., 1997; Cahill et al., 2002). In this way, Nop56 and Nop58 link the guide RNA, which specifies the methylation target, to the methyltransferase Nop1 or Fibrillarin.

In the cell the assembly of the eukaryotic Box C/D snoRNPs is assisted by several assembly factors, whose exact mechanism of action is still poorly understood. Involved factors include the HSP90/R2TP chaperone-cochaperone system, Rvb1/2, Rsa1, Naf1, Hit1, Bcd1 and others (Massenet et al., 2017).

### Box C/D guide s/snoRNAs

The guide RNA is the central component of the Box C/D methylation machinery. In both archaea and eukaryotes the guide RNA is responsible for specifying the methylation site of the rRNA by base-pairing with the rRNA sequence through a complementary guide sequence located between Box D and Box C' or Box D' and Box C (Tran et al., 2003, 2005; Appel and Maxwell, 2007) (**Figure 1.2.5 A**). Usually, 10 or more bases of rRNA are recognized by the guide sequences (Yang et al., 2016b). In both archaea and in eukaryotes, the nucleotide of the rRNA that base-pairs with the fifth nucleotide upstream of Box D or Box D' is the methylation target of the Box C/D RNP enzyme.



**Figure 1.2.5.** Schematic representation of a Box C/D guide RNA and motif conservation. **(A)** Schematic representation of Box C/D guide RNA with Box C (green) and D (orange) and C' (green) and D' (orange) explicitly represented. Guide sequences are indicated by base-pairing with the target RNA (orange), whereas the site for 2'-O-methylation is indicated with a green star. AU-rich spacers are only present in eukaryotic Box C/D guide RNAs. The structure connecting Box D' and C' can be either a loop or a helical structure depending on the individual Box C/D RNA. R is either an A or a G, X is any nucleotide. **(B)** Conservation of Box C, D, C' and D' in *S. cerevisiae*. A is represented in green, U in red, G in yellow and C in blue. The level of conservation at each position is indicated by the height of the respective nucleotide. Part **(B)** of the figure is adapted from (van Nues et al., 2011).

Both Box C and D are highly conserved in archaea and eukaryotes (van Nues et al., 2011; Omer et al., 2000) (**Figure 1.2.5 B**). The canonical Box C/D sequence consists in an asymmetric three-nucleotide bulge (**Figure 1.2.3**) flanked by a canonical stem I and a non-canonical stem II. The two G•A base pairs in position 1n-1b and 2n-2b are conserved in k-turn structures.

In archaeal guide sRNAs the Box C' and D' motifs are similarly well conserved as their counterparts Box C and D (Omer et al., 2000).

Conversely, in eukaryotes Box C' and D' show a significantly decreased level of conservation (**Figure 1.2.5 B**) (van Nues et al., 2011). The higher divergence of the internal Box C'/D' motif from the canonical sequence questions the ability of Box C' and D' to form kinked structures.

Another difference between archaeal and eukaryotic Box C/D guide RNAs is the presence of so-called AU-rich spacer regions in eukaryotic guide RNAs with still unknown function. These are stretches of various length that are located upstream of Box C as well as downstream of Box C' (Piekna-Przybylska et al., 2007). In archaea, where these AU-rich spacers are missing, the guide sequence alone separates Box C from Box D' or Box C' from Box D; consequently, archaeal Box C/D sRNAs are notably shorter than their eukaryotic counterparts.

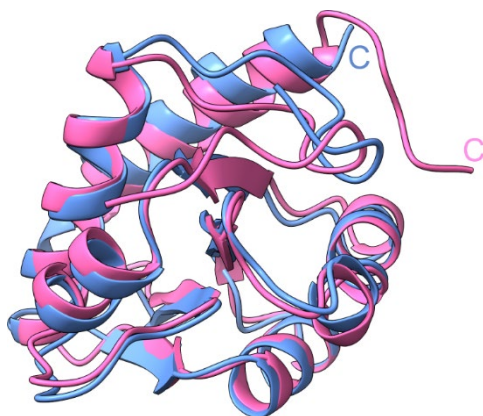
Another distinguishing feature of eukaryotic Box C/D guide RNAs is that they do not always have two guide sequences as most of their archaeal counterparts do (Piekna-Przybylska et al., 2007). In *S. cerevisiae* only about 50% of all Box C/D snoRNAs have two guide sequences, including snR51 and snR41, whereas the other half has one guide sequence that is either

upstream of Box D' or D, including snR54. In guide RNAs with only one guide sequence the space between Box C and D' or Box D' and C is occupied by an AU-rich spacer of various length. In humans almost all identified Box C/D guide RNAs carry one guide sequence only (Yoshihama et al., 2013).

### Eukaryotic Snu13 and archaeal L7Ae

The initial assembly step of the Box C/D RNP is the recognition of the Box C/D guide RNA by the primary RNA-binding protein. In both archaea and eukaryotes, the primary RNA-binding protein is a member of the L7Ae/L30 protein family. One of the defining features of this family of proteins is their ability to bind to k-turns (**Figure 1.2.4**).

In archaea, the ~13 kDa L7Ae protein binds the Box C/D guide sRNAs by recognizing the k-turn structure formed by the Box C and D motifs and the k-turn or k-loop structure formed by Box C' and D' (Moore et al., 2004; Huang and Lilley, 2013).



**Figure 1.2.6.** Structural comparison of Snu13 from *S. cerevisiae* (pink) (PDB-ID: 2ALE, Dobbyn et al., 2007) and L7Ae from *P. furiosus* (blue) (PDB-ID: 3NVI, Xue et al., 2010). The RMSD for all aligned C<sub>α</sub> atoms is 1.99 Å. Structural alignment was done with matchmaker in ChimeraX 1.0 (Goddard et al. 2018). C indicates the C-terminus of the respective protein.

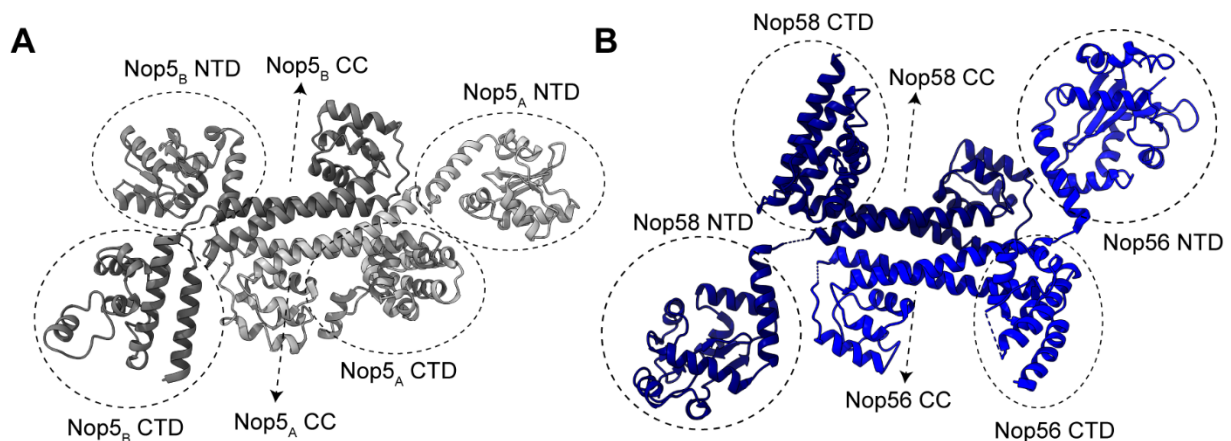
In eukaryotes the L7Ae homolog Snu13 (~13 - 15 kDa) binds the k-turn structure formed by the conserved Box C and D motifs (Cahill et al., 2002; Marmier-Gourrier et al., 2003). However, it is unclear whether Snu13 also binds to the less conserved internal

Box C' and D' motifs of eukaryotic guide RNAs. Despite having essentially the same structure (**Figure 1.2.6**), L7Ae and Snu13 show a different binding behavior towards k-turns and k-loops (Goddard et al., 2018; Gagnon et al., 2010; Dobbyn et al., 2007).

Extensive structural and functional studies on archaeal Box C/D sRNPs from different species have demonstrated that the stable interaction of L7Ae to both kinked structures on the guide RNA is essential for the correct positioning of the scaffolding protein Nop5 (Aittaleb et al., 2003; Bleichert et al., 2009; Ye et al., 2009; Lapinaite et al., 2013). In eukaryotes, the presence of the second binding site along the Box C' and D' is still under debate; thus, it remains unclear whether assembly and positioning of the enzyme components follows similar pathways as in archaea.

### Eukaryotic Nop56, Nop58 and archaeal Nop5

The next components in the assembly process are the scaffolding proteins Nop56 and Nop58 in eukaryotes or Nop5 in archaea (Gautier et al., 1997; Omer et al., 2000). All three proteins have a modular architecture, with a globular N- and C-terminal domain and a central elongated coiled-coil domain. In archaea, Nop5 forms a homodimer via the coiled-coil domain (Nop5<sub>2</sub>), which leads to an antiparallel orientation of the two proteins (**Figure 1.2.7 A**). Similarly, Nop56 and Nop58 form a heterodimer, also through the central coiled-coil domain, leading to an antiparallel orientation (**Figure 1.2.7 B**).



**Figure 1.2.7.** Structure of archaeal Nop5 homodimer and eukaryotic Nop56–Nop58 heterodimer. **(A)** Structure of the Nop5 homodimer from *P. furiosus* (PDB-ID: 4BY9, Lapinaite et al., 2013). The three domains of each monomer are indicated: The N-terminal domain (NTD), the central coiled-coil domain (CC) and the C-terminal domain (CTD). The homodimer forms via interactions of the CC domains, whereas the NTDs and the CTDs are accessible to interact with other partners. **(B)** Structure of the Nop56–Nop58 heterodimer from *S. cerevisiae* in the U3 snoRNP within the 90S pre-ribosome (PDB-ID: 5WLC, Barandun et al., 2017). The three domains of Nop56 (light blue) and Nop58 (dark blue) are indicated: The N-terminal domain (NTD), the central coiled-coil domain (CC) and the C-terminal domain (CTD). Similarly, the heterodimer forms by interactions of the CC domains of the two proteins, which leaves the NTDs and CTDs free for other interactions partners.

Each C-terminal domain of Nop5<sub>2</sub> interacts with one of two composite surfaces formed by L7Ae and the guide RNA (**Figure 1.2.8**). This interaction leads to the integration of Nop5<sub>2</sub> into the Box C/D enzyme. For archaeal Box C/D sRNPs two different active assembly states have been proposed: the mono-RNP and the di-RNP (Bleichert et al., 2009; Ye et al., 2009; Lapinaite et al., 2013). In the mono-RNP state, one Nop5<sub>2</sub> interacts with one L7Ae–guide RNA complex, with one C-terminal domain recognizing the L7Ae–Box C/D motif interface and the other C-terminal domain recognizing the L7Ae–Box C'/D' motif interface (**Figure 1.2.9**, mono-RNP model) (Ye et al., 2009; Lin et al., 2011). Contrary, in the di-RNP state, one Nop5<sub>2</sub> interacts with the L7Ae–Box C/D and L7Ae–Box C'/D' interfaces of two different L7Ae–guide RNA complexes and two copies of guide RNA together with two copies of Nop5<sub>2</sub> and 4 copies of

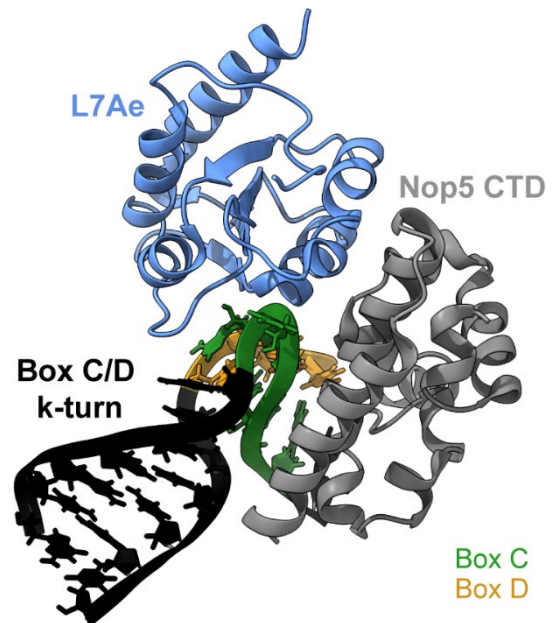
L7Ae are integrated into the RNP (**Figure 1.2.9**, di-RNP model) (Bleichert et al., 2009; Lapinaite et al., 2013).

Cross-linking data available for eukaryotic Box C/D snoRNPs suggests that the C-terminal domain of Nop58 interacts with the composite surface formed by Snu13 and the Box C/D k-turn structure, whereas the C-terminal domain of Nop56 interacts with the guide RNA next to the Box C'/D' motif (Cahill et al., 2002). Whether Snu13 binds the Box C'/D' motif is still a matter of debate. Definite evidence of such interaction is provided by the U3 snoRNP structure in the context the 90S pre-ribosome (Barandun et al., 2017). However, the U3 snoRNP is not active in 2'-O-methylation, but functions as a folding chaperone during early stages of the ribosome assembly and consequently might differ structurally from methylation-guiding snoRNPs. In addition, the Box C'/D' motif of the U3 snoRNA has a canonical Box C/D sequence, unlike most other snoRNAs. Nevertheless, the available data supports a mono-

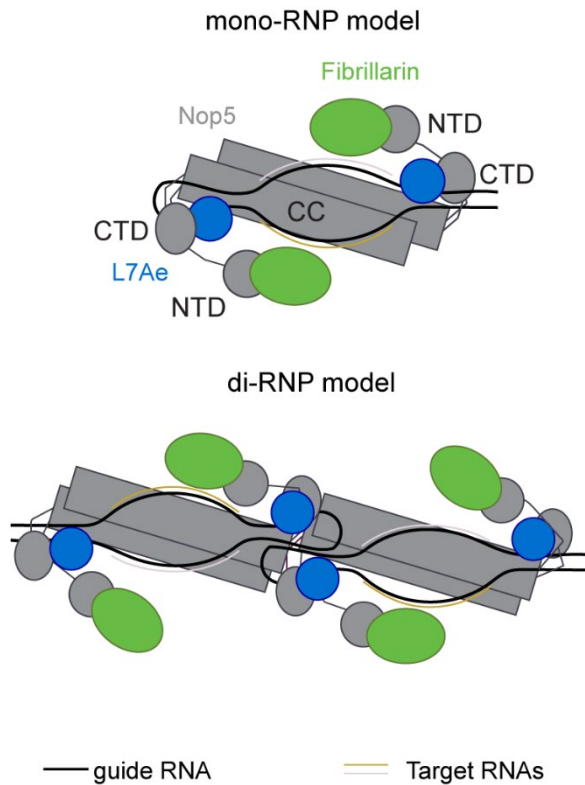
RNP model for the eukaryotic Box C/D snoRNP, in which the guide RNA is (probably) oriented along the coiled-coil domains of the Nop56 and Nop58 heterodimer.

The N-terminal domain of archaeal Nop5 forms a complex with the methyltransferase Fibrillarin; similarly, the N-terminal domains of Nop56 and Nop58 interact with the eukaryotic methyltransferase Nop1 (yeast) or Fibrillarin (human). Through this interaction the catalytically active component is integrated into the Box C/D s/snoRNP enzyme (**Figure 1.2.10**).

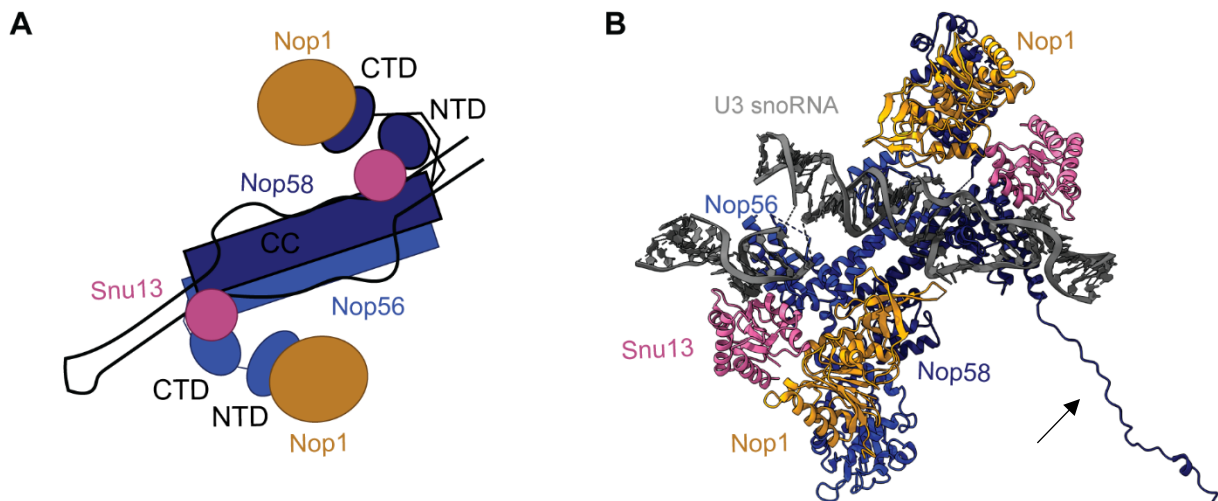
Finally, Nop56 and Nop58 both have extended KDE-rich C-terminal tails, whose function is unclear (Gautier et al., 1997).



**Figure 1.2.8.** Interaction of the L7Ae–RNA composite surface with the C-terminal domain (CTD) of Nop5 in Box C/D sRNP of *P.furiosus* (PDB-ID: 4BY9, Lapinaite et al., 2013). L7Ae is depicted in blue, Nop5 CTD in gray, the Box C/D k-turn RNA in black, Box C in green and Box D in orange.



**Figure 1.2.9.** Schematic models of the archaeal mono-RNP and di-RNP conformation. In the mono-RNP model two copies of each of the three core proteins (L7Ae, Nop5, Fibrillarin) are integrated into the complex and assemble around one copy of a Box C/D sRNA. In the di-RNP model four copies of each of the three core proteins assemble around two copies of Box C/D guide RNA. In both cases two target RNAs (gray and gold) can base-pair to the guide sequences upstream of Box D and Box D'. L7Ae is depicted in blue, Nop5 in gray, Fibrillarin in green and the guide RNA in black. The N-terminal (NTD) and C-terminal (CTD) domains of Nop5 are indicated.

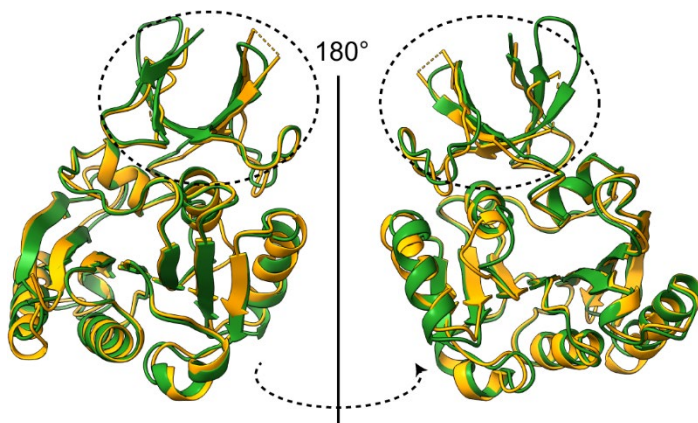


**Figure 1.2.10.** Schematic model of eukaryotic Box C/D snoRNP and structure of U3 snoRNP in the 90S pre-ribosome from *S. cerevisiae*. **(A)** Schematic representation of the mono-RNP model of the eukaryotic Box C/D snoRNP enzyme, based on the structure of the U3 snoRNP. Here, Snu13 (pink) binds to both Box C/D and Box C'/D' motifs. Nop56 (light blue) and Nop58 (dark blue) form a heterodimer via interactions of their coiled-coil domains (CC). The C-terminal domains (CTD) of Nop56 and Nop58 interact with the Snu13–RNA complex and the N-terminal domains (NTD) recruit the methyltransferase Nop1 (orange). **(B)** Structure of the U3 snoRNP in the 90S pre-ribosome structure of *S. cerevisiae* solved by cryo-electron microscopy (EM) (PDB-ID: 5WLC, Barandun et al., 2017). Color and interactions are equal to does described in 1.2.10 A. The arrow indicates the extended KDE-rich C-terminal tail of Nop58 that was resolved in this structure.

### Nop1 and archaeal Fibrillarin

Archaeal or eukaryotic Nop1 (yeast) or Fibrillarin (human) is the catalytic subunit of the Box C/D enzyme (Tollervey et al., 1991; McKeegan et al., 2007; Quinternet et al., 2015). Archaeal and eukaryotic Fibrillarins are SAM-dependent methyltransferases. They share a common core fold comprised of a small N-terminal  $\beta$ -sheet domain and a central nucleotide binding Rossmann-fold consisting of a seven-stranded  $\beta$ -sheet with three  $\alpha$ -helices on either side (**Figure 1.2.11**) (Deng et al., 2004; Martin and McMillan, 2002). The central domain binds the SAM cofactor, which provides the methyl group for the enzymatic reaction and leaves the protein bound to S-adenosyl-homocysteine (SAH) after catalysis. Additionally, eukaryotic Fibrillarins also have a  $\sim 80$  amino acid long N-terminal domain rich in Arginines (R) and Glycines (G), which is therefore termed the RGG- or Glycine-Arginine-rich (GAR) domain (Rodriguez-Corona et al., 2015). The RGG-domain is not involved in interactions with Nop56 and Nop58 and its functions remains mostly elusive. However, similar domains are found in more than 1000 eukaryotic nuclear proteins, where they have been shown to be involved in nucleolar localization and protein-protein interactions (Thandapani et al., 2013).

In yeast, Nop1 has been shown to associate with both Nop56 and Nop58 *in vivo*, but seems to be stronger associated to Nop56 as compared to Nop58 in pull-down assays (Gautier et al., 1997).



**Figure 1.2.11.** Structural comparison of archaeal Fibrillarin from *P. furiosus* (green) (PDB-ID: 1PRY, Deng et al., 2004) and eukaryotic Fibrillarin from *H. sapiens* (orange) (PDB-ID: 2IPX, to be published). The N-terminal RGG—rich domain has been truncated in the construct used to crystallize human Fibrillarin. The smaller  $\beta$ -sheet based small N-terminal domain is circled with a dashed line. The RMSD for all aligned  $C_{\alpha}$  atoms is 1.88 Å. Structural alignment was done with matchmaker in ChimeraX 1.0 (Goddard et al. 2018).



### 1.3 Pathologies related to an aberrant rRNA 2'-O-ribose methylation machinery

Dysregulations of 2'-O-methylation and the components of the enzymatic machinery including the guide snoRNAs, have been associated with various diseases, such as cancer and neuropathologies (Dimitrova et al., 2019). As 2'-O-methylation is involved in ribosome biogenesis and functions to stabilize rRNA, this finding is not surprising (Polikanov et al., 2015). Pathologies with a characterized link to 2'-O-Me machinery are either related to the guide snRNAs or the SAM-dependent methyltransferase Fibrillarin.

#### 1.3.1 snoRNAs

The importance of Box C/D snoRNAs guiding 2'-O-methylations was shown in zebrafish. In this eukaryotic model system loss-of-function mutations of SNORD26, SNORD44, and SNORD78 guide RNAs were shown to lead to severe morphological defects as well as embryonic lethality (Higa-Nakamine et al., 2012).

In humans, the deficiency or microdeletion of one or more paternally expressed imprinted transcripts at genome locus 15q11-q13 causes the Prader-Willie syndrome (PWS). PWS is a complex human neurological disease that causes mental retardation, low stature, obesity and muscle hypotonia (Cavaillé et al., 2002; Sahoo et al., 2008; Doe et al., 2009; Bortolin-Cavaillé and Cavaillé, 2012; Cavaillé, 2017). The locus 15q11-q13 contains numerous copies of the Box C/D guide RNAs SNORD115 (HBII-52) and SNORD116 (HBII-85). Loss of SNORD116 is associated with the etiology of PWS (Sahoo et al., 2008).

Interestingly, a duplication of the 15q11-q13 locus has been associated with autism spectrum disorder (ASD) (Cook Jr and Scherer, 2008). In both PWS and ASD the exact mechanism by which the Box C/D snoRNAs are involved in the disease etiology is still unclear.

#### 1.3.2 SAM-dependent methyltransferase Fibrillarin

Fibrillarin (human) or Nop1 (yeast) is the evolutionary highly conserved 2'-O-ribose methyltransferase, which gets its name from its localization in the dense fibrillar components of the nucleolus, where it was first discovered (Tollervey et al., 1991). The strongly conserved methyltransferase can be found in archaea, *Giardia* up to higher mammals, which underlines the importance of this enzyme for cellular metabolism (Amiri, 1994; Shubina et al., 2018). Indeed, Nop1 has been shown to be involved in pre-rRNA processing, pre-rRNA 2'-O-

methylation and ribosome assembly, with mutants impacting ribosome synthesis in different aspects (Tollervey et al., 1993). Furthermore, Fibrillarin is involved in regulating the transcription of rRNA by methylating a glutamine residue in histone H2A in yeast, plants and humans (Tessarz et al., 2014; Loza-Muller et al., 2015).

Through its involvement in ribosome biogenesis, downregulation of Fibrillarin also leads to the genesis of impaired ribosomes, which have an altered ability to initiate translation from internal ribosome entry site (IRES) elements independent of initiation factors (Erales et al., 2017). Additionally, mice which have mutated copies of Fibrillarin are non-viable as they die during early development due to massive cellular apoptosis (Newton et al., 2003). In zebrafish, deletion of Fibrillarin heavily effects the neural differentiation, the optic tectum and the eye (Bouffard et al., 2018). Another study showed that Fibrillarin regulates stem cell pluripotency in mice embryonic stem cells (Watanabe-Susaki et al., 2014).

Due to its involvement in the essential cellular process of ribosome biogenesis it comes as no surprise that the loss or unregulated expression of Fibrillarin is involved in different pathologies, including cancer, autoimmune diseases, and aging.

One of the hallmarks of cancer cells is an unregulated rate of cell proliferation, which in turn requires increased ribosome production to meet the higher translation demands. As Fibrillarin is highly involved in ribosome biogenesis its upregulated expression is often observed in a variety of different cancer types (Derenzini et al., 1998; Koh et al., 2011; Marcel et al., 2013; Su et al., 2014). The tumor suppressor p53 has been shown to bind to the *FBL* gene to suppress its expression, whereas the oncogene C-MYC increases the expression of the *FBL* gene by binding to the 5' upstream region of the gene (Koh et al., 2011; Marcel et al., 2013). In cancer cells p53 is often suppressed, which leads to an increased Fibrillarin expression that alters the 2'-O-methylation patterns of rRNA at a single-nucleotide level. This leads to an increased IRES-dependent translation of oncogenes, including IGF-1R, C-MYC, VEGF-A, and FGF1/2 (Marcel et al., 2013; Erales et al., 2017). For these reasons Fibrillarin is also being recently investigated as a potential target for cancer therapeutics (El Hassouni et al., 2019).

Anti-fibrillarin antibodies have been detected in several diseases of connective tissues, such as mixed connective tissue diseases (60%), CREST syndrome (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia syndrome, 58%), systemic lupus erythematosus (39%), rheumatoid arthritis (60%), and Sjogern's syndrome (84%), and SSc patients (58%) (Kasturi et al., 1995; Bae and Lee, 2018). Despite the fact that in autoimmune diseases the production of anti-fibrillarin antibodies is thought to be symptomatic rather than causative, the presence of such antibodies functions to decrease the amount of protein: the impact of decreased Fibrillarin levels in these diseases is still nebulous.

Lastly, elevated levels of ribosome biogenesis and increased nucleolar size are correlated with increased age in healthy humans and patients with pre-mature aging, such as Hutchinson-Gilford syndrome (Buchwalter and Hetzer, 2017). The size of the nucleolus has been shown to be regulated by the *let-7/ncl-1/fib-1* pathway in *Caenorhabditis elegans* in a recent study (Tiku et al., 2017). The protein NCL-1 represses the expression of Fibrillarin and thereby controls the size of the nucleolus, while NCL-1 itself is regulated by the miRNA let-7. The same study shows that Fibrillarin knock-down leads to a decreased nucleolar size and an increased lifespan (Tiku et al., 2017). This mechanism, which still needs to be verified in higher mammals, suggests a key role of Fibrillarin in the cellular aging process.

## 1.4 Aims of this Thesis Work

The existing body of research provides a plethora of functional and structural information on the assembly, conformation and mechanism of action of the archaeal Box C/D sRNP enzyme. The wealth of studies on the archaeal system is due to the simplicity of the system that allows its *in vitro* reconstitution in an active and homogeneous form. Conversely, data on the eukaryotic Box C/D snoRNP is limited to structural information on the non-methylating U3 snoRNP from *S. cerevisiae*, and *in vivo* crosslinking on the U25 snoRNP from *Xenopus leavis*, as a successful *in vitro* reconstitution of an active and homogeneous methylating snoRNP has so far been hindered by the complexity of the system and the assembly process itself.

The archaeal system has so far been used as a proxy for the eukaryotic Box C/D snoRNP, but the translatability of the results from archaea to eukaryotes is still unclear. Thus, information on the molecular details that govern intermolecular interactions in the eukaryotic Box C/D snoRNP enzyme as well as its architecture and functional mechanism would not just contribute to increase our knowledge of the eukaryotic methylation enzyme but also help evaluating the suitability of the archaeal Box C/D sRNP as a structural and functional proxy.

Here, we use the Box C/D snoRNP components from the eukaryotic model system *S. cerevisiae* and the archaeum *Pyrococcus furiosus* to achieve the following objectives:

1. Understand the molecular basis of the recognition of Box C'/D' motifs by Snu13 *in vitro* in isolation and in the context of a chimeric Box C/D s/snoRNP by:
  - Assessing the binding affinities of Snu13 and L7Ae towards isolated Box C'/D' motifs derived from archaeal sRNA sR26, eukaryotic snoRNAs snR41, snR51 and snR54 and their mutants.
  - Mutational analysis of Snu13 to determine the residues responsible for affinity differences between Snu13 and L7Ae.

- Determining high-resolution structures of Snu13 and L7Ae bound to archaeal and eukaryotic Box C'/D' motifs.
  - Reconstituting chimeric Box C/D complexes with archaeal sRNA sR26, eukaryotic snoRNAs snR41, snR51, snR54, eukaryotic Snu13, archaeal Nop5 and archaeal Fibrillarin and determining the protein stoichiometry.
  - Assessing the activity of the chimeric Box C/D complexes.
2. Understand the molecular basis of the Nop1–Nop56 and Nop1–Nop58 interaction by
- solving high-resolution structures of Nop1 in complex with the N-terminal domains of Nop56 and Nop58.

## 2. Methodological Background

To gain insights into the structural details of the Box C/D snoRNPs from *S. cerevisiae*, we applied a set of biochemical, biophysical, and structural methods. The theoretical basis for each of these techniques is presented in this chapter.

### 2.1 Size-exclusion Chromatography and Multi-angle Light Scattering

#### 2.1.1 Size-exclusion Chromatography (SEC)

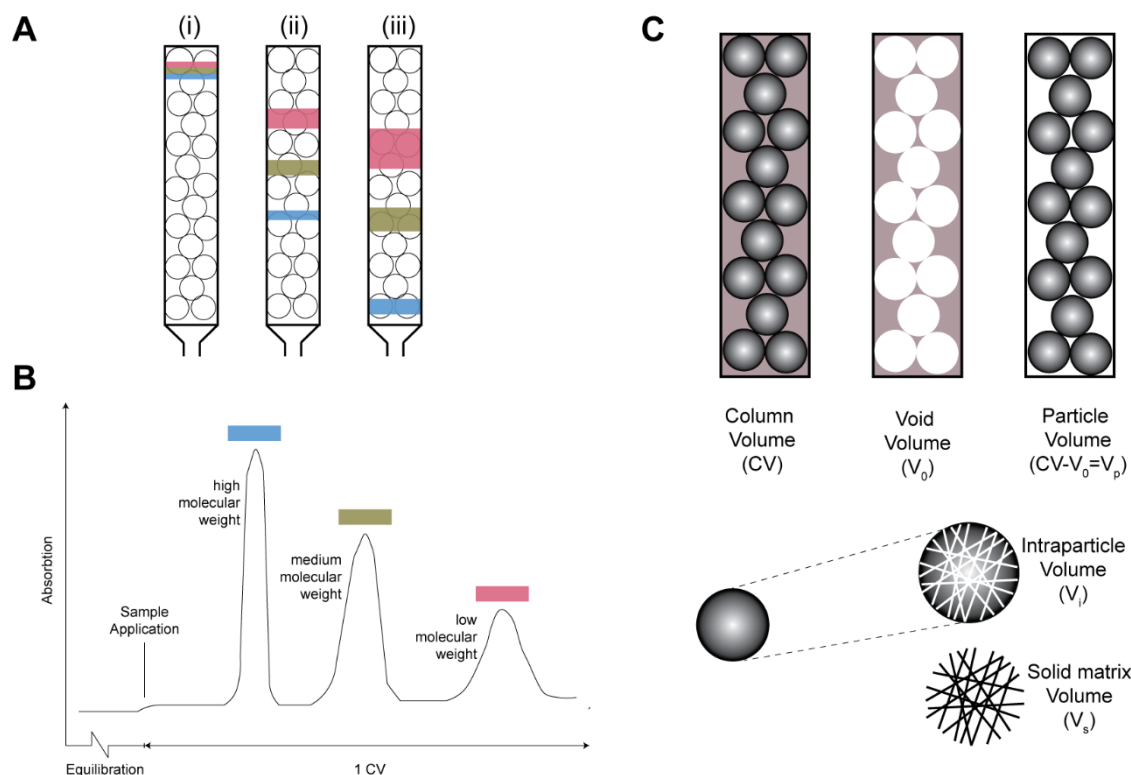
Size-exclusion chromatography (SEC) is a biophysical method that separates molecules based on their apparent size. The molecular weight, the shape, and the hydration shell of the molecule influence its apparent size. SEC is frequently used in the preparation and analytical characterization of biomolecules, such as proteins and nucleic acids (DNA and RNA) and has the advantage of preserving the molecular structure and the biological activity.

In SEC applications for biomolecules, the stationary phase is made of different types of micro-scale polymers, whereas the mobile phase is an aqueous buffer solution that solubilizes the molecules to be separated or analyzed. The molecules should ideally have no interactions with the stationary phase, contrary to other types of biomolecular chromatography such as affinity chromatography. The stationary phase consists of a polymer matrix with different pore size distributions, usually packed into a column. The molecules solubilized in the mobile phase migrate through the pores of the polymer matrix once an eluent flow is applied. Molecules with an apparent size smaller than a particular pore size can enter these pores, whereas larger molecules can only access pores of bigger sizes. Therefore, smaller molecules can access a greater pore volume than larger molecules. Consequently, molecules with a larger apparent size elute earlier than smaller molecules (**Figure 2.1.1 A-B**).

The pore volume a molecule can access when separated by SEC equals the volume of the mobile phase that has to pass through the stationary phase for the molecule to elute from the column. This volume is specific to each molecule and is referred to as the elution or retention volume ( $V_e$ ).  $V_e$  is a characteristic parameter that can be determined by SEC for each molecule; besides the specific properties of the molecule,  $V_e$  depends on the type of stationary phase (pore size distribution), the flow rate of the mobile phase, and column volume (CV) and is not transferable if any of these parameters is changed.

In SEC applications, the total volume of a column (CV) can be divided into separate partial volumes (**Figure 2.1.1 C**). The void volume ( $V_0$ ) is the fraction of the CV that is not occupied

by the stationary phase matrix. Molecules that are too big to enter the pores of the stationary phase matrix can only access the  $V_0$ , which therefore equals their  $V_e$ . Notably, no information on the molecule's apparent size (besides a rough lower limit) can be derived if the molecule elutes in the  $V_0$ . In such cases, a different stationary phase matrix with larger pore sizes should be used. The  $V_0$  of SEC columns is around 30-35 % of the CV. The particle volume ( $V_p$ ) is the volume occupied by the stationary phase matrix. The  $V_p$  itself is the sum of the intraparticle volume ( $V_i$ ) and the solid matrix volume ( $V_s$ ). To achieve the best possible separation for any given molecule, meaning well-resolved elution peaks with minimal peak broadening, the sample volume should range from 0.5 to 4 % of the CV.



**Figure 2.1.1.** Schematic representation of the basic principles of size-exclusion chromatography. **(A)** Schematic representation of the size dependent migration of molecules of different apparent sizes. The smallest molecule (pink) has access to a wider range of pore sizes as compared to the medium size (green) or the large (blue) molecule and therefore needs a larger volume of mobile phase to elute from the column. **(B)** Schematic elution profile of the three molecules from panel A. Due to their size the large (blue) molecule elutes early, followed by the medium size (green) and the small (pink) molecule. **(C)** Simplified representation of the different volume types in size-exclusion chromatography.

As mentioned above,  $V_e$  is a molecule specific parameter obtained by SEC that can be used to separate biomolecules of different apparent sizes as well as to monitor changes in the biomolecules that affect the apparent size. Such changes include the unfolding of proteins, substantial conformational changes in large RNA molecules, and complex formation between

two or more biomolecules. In order to gain meaningful results when comparing the  $V_e$  of the same biomolecule in different SEC runs or when trying to detect the formation or disassembly of a biomolecular complex, it is necessary to use the same column, including the stationary phase material and the CV, as well as to apply the same mobile phase flow rate, as  $V_e$  is dependent on these factors. An example of how to use SEC to analyze the formation of biomolecular complexes is given in chapter 3.2, **Figure 3.2.1**.

A parameter better suited to characterize a biomolecule or biomolecular complex than the  $V_e$  is the distribution coefficient  $K_d$ . The  $K_d$  is independent of the CV and can therefore be used for comparison if the same matrix material and flow rate are used. The  $K_d$  is defined in equation 2.1.

$$K_d = \frac{V_e - V_0}{CV - V_0} = \frac{V_e - V_0}{V_i} = \frac{\text{Concentration of solute in stationary phase}}{\text{Concentration of solute in mobile phase}} \quad \text{Eq. (2.1)}$$

The  $V_e$  or  $K_d$  value determined by SEC can also give information on the approximate molecular weight of a biomolecule. To extract size information from  $V_e$  or  $K_d$ , it is necessary to generate a calibration curve with biomolecules of known size. By doing so, it is possible to assign  $V_e$  or  $K_d$  values to different molecular weights, which allows estimating the molecular weight of an unknown biomolecule based on its  $V_e$  or  $K_d$ . Notably, any calibration curve is only valid for analysis done on the same column with the same flow rate.

In general, the eluted biomolecule is detected by measuring the absorption at 280 nm for proteins that contain at least one tryptophane, 254 nm for nucleic acids, or 215 nm for proteins without any tryptophans. The absorption intensity, which is often given in absorption units (AU), is also related to the concentration of the eluting molecule.

In this work, SEC was used for preparative (size-based purification) and analytical purposes. SEC was used to analyze the formation of stable protein-protein and protein-RNA complexes (see chapters 3.2.1, 5.4, and 5.7.1).

### 2.1.2 Multi-angle Light Scattering (MALS)

Size-exclusion chromatography can provide useful information on changes in the apparent size of biomolecules and inform on the formation or disassembly of biomolecular complexes, but it has only limited applicability in the determination of molecular weights or masses.

Multi-angle static light scattering (MALS), or Rayleigh scattering techniques provide a fast and non-invasive way to determine the molar mass ( $M$ ), the radius of gyration ( $r_g$ ), and the second virial coefficient ( $A_2$ ) of biomolecules and biomolecular complexes in solution. MALS can be

used in batch measurements or on-line with SEC (SEC-MALS) or other separation techniques. In this work, an SEC-MALS application was used to determine the molar mass of biomolecular complexes (see chapter 3.1, **Figure 3.1.11**).

In the light scattering experiment, the molecule is irradiated by a laser beam ( $\gamma \sim 660$  nm) either in batch or on-line after being separated by SEC or another technique. The light scattered by the solution, including the molecule, is simultaneously detected by multiple detectors installed at different angles  $\Theta$ . The scattering biomolecule is not damaged during the measurement and can be used in downstream applications.

The determination of the parameters mentioned above by MALS is based on the polarizability of matter through interaction with the oscillating electric field component of light. The higher the polarizability of a given material, the more light will be scattered. The polarizability of a specific type of matter is directly related to its index of refraction  $n$ . The refractive index of a material is related to the speed with which light transverses through it (equation 2.2).

$$v_{material} = \frac{v_{vacuum}}{n} \quad \text{Eq. (2.2)}$$

Where  $v_{material}$  is the speed of light through the a given material and  $v_{vacuum}$  the speed of light in vacuum. The polarizability or refractive index of a molecule in solution is usually quantified as the refractive index increment, denoted as  $dn/dc$  ( $n$  - refractive index,  $c$  - molecule concentration), to account for the polarization difference between solute and solvent in a concentration-dependent manner. The relationship between the refractive index increment  $dn/dc$  and the scattered light intensity at angle  $\Theta$   $I(\Theta)_{scattered}$  is given in equation 2.3.

$$I(\Theta)_{scattered} \propto Mc \left( \frac{dn}{dc} \right)^2 \quad \text{Eq. (2.3)}$$

The light scattering intensity is proportional to the molar mass  $M$  of a given molecule, the concentration of the molecule  $c$ , and the refractive index increment  $dn/dc$ . Therefore, the molar mass  $M$  of molecule can be determined by measuring the intensity of the scattered light at any given angle  $\Theta$  if the concentration  $c$  and the refractive index increment  $dn/dc$  for the molecule are known.

Molecules that are much smaller than the wavelength of the incident beam are called isotropic scatterers. This means that they show no angular variation in the intensity of the scattered light. For a beam with  $\lambda = 660$  nm, molecules  $< 10$  nm are considered isotropic scatterers. Bigger particles, termed anisotropic scatterers, show an angular variation in the intensity of scattered light, with the scattering intensity decreasing with increasing angles. The angular variation of the scattered intensity is related to the size of the solute expressed as the or radius of gyration



( $r_g$ ), which is the root mean square distance from the center of mass of the scattering particle (equation 2.4).

$$\frac{1}{P(\theta)} = 1 + \frac{16\pi^2 n_0^2 r_g^2}{3\lambda_0^2} \sin^2\left(\frac{\theta}{2}\right) \quad \text{Eq. (2.4)}$$

$P(\Theta)$  is called the form factor or particle scattering function,  $n_0$  is the refractive index of the solvent,  $r_g$  the radius of gyration of the particle or molecule,  $\lambda_0$  the wavelength of the incident beam in vacuum, and  $\Theta$  the angle of measurement. Thus, provided that the scattered light intensity can be measured at different angles and the molecule is an anisotropic scatterer ( $> 10$  nm at  $\lambda = 660$  nm), the  $r_g$  of the molecule can be determined without knowing its concentration or the refractive index increment.

Relations represented by equations 2.3 and 2.4 are combined in the Zimm equation (equation 2.5), which relates  $I(\Theta)$  and  $P(\Theta)$  (Zimm, 1948b, 1948a).

$$\left(\frac{Kc}{I(\theta)_{scattered}}\right) = \left(\frac{1}{MP(\theta)}\right) + 2A_2c \quad \text{Eq. (2.5)}$$

Where  $c$  is the molecule concentration,  $I(\Theta)$  the intensity of scattered light at angle  $\Theta$ ,  $K$  is a constant defined in equation 2.6,  $P(\Theta)$  the form factor, and  $A_2$  the second virial coefficient. The second virial coefficient is a thermodynamic term, that gives information on non-specific solvent-molecule interactions and can inform on the quality of a given solvent for a molecule.

$$K = \frac{4\pi^2 n_0}{N_A \lambda_0^4} \left(\frac{dn}{dc}\right)^2 \quad \text{Eq. (2.6)}$$

Where  $n_0$  is the refractive index of the solvent,  $N_A$  the Avogadro's number and  $\lambda_0$  the wavelength of the incident light in vacuum. In equation 2.5,  $I(\Theta)_{scattered}$  can be substituted with the excess Rayleigh ratio  $R(\Theta)$ , yielding equation 2.7.

$$\left(\frac{Kc}{R(\theta)}\right) = \left(\frac{1}{MP(\theta)}\right) + 2A_2c \quad \text{Eq. (2.7)}$$

The relationship described in equation 2.7 is based on the Rayleigh-Gans-Debye approximation, which is only valid if the following prerequisites are met:

- The analyzed particles must be smaller than the wavelength of the incident light.
- The ratio between the refractive index of the particle  $n$  and the solvent  $n_0$  must be 1, meaning the particle is invisible in the solvent.
- The particle cannot disturb the phase of the incident light.

These criteria are generally met with biomolecular solutions. Therefore, MALS provides a non-invasive, non-destructive way to determine the molar mass and size of biomolecules or biomolecular complexes in solution (Patel et al., 2018; Wyatt, 1993). To determine the molar

mass, it is only necessary to know the molecular concentration and the refractive index increment  $dn/dc$ , which is very similar for most biomolecules, which a standard value of 0.185 ml/g (Zhao et al., 2011). To determine the molecule size, the molecule needs to be an anisotropic scatterer at the given incident beam wavelength.

## 2.2 Electrophoretic Mobility Shift Assays

### 2.2.1 Principles

Native gel electrophoretic mobility shift assays (EMSA) or gel retardation assays are one of the most sensitive techniques to obtain qualitative and quantitative information on specific protein-DNA and protein-RNA interactions (Fried, 1989; Hellman and Fried, 2007). The assay is based on the fact that during electrophoretic separation in a non-denaturing gel matrix unbound nucleic acids migrate faster than if they are in complex with proteins (Garner and Revzin, 1981). During native gel-electrophoresis, the rate at which the molecules migrate through the gel matrix in the electrical field depends on their shape, size, and net charge; larger molecules migrate slower, and molecules with a higher net negative charge migrate faster. Thus, nucleic acids migrate according to their shape and size as their charge is directly proportional to the size. On the other hand, in proteins the net charge is not correlated to the size and all three parameters (shape, size and charge) in combination determine the migration rate (Arndt et al., 2019). When a protein binds a nucleic acid, the resulting complex has a larger size than the unbound nucleic acid and a reduced negative net charge; as a consequence, the complex migrates slower through the gel matrix than the free nucleic acid and will yield a band higher in the gel. This behavior is referred to as a gel or mobility shift.

In this work, native gel EMSAs were used to reveal whether L7Ae, Snu13 or mutants thereof form a complex with a set of short RNA constructs and, in case of interaction, to determine the dissociation constant ( $K_D$ ) of the complex (see chapter 3.1).

In a classical mobility shift assay, a purified protein solution is mixed with a solution of a defined DNA or RNA molecule; after an equilibration time the resulting mixture is subjected to gel electrophoresis. Cell extracts can also be used instead of purified proteins to identify new nucleic acid-binding protein for specific DNA or RNA sequences. To gain useful information on potential protein-nucleic acid interactions, the band corresponding to the free nucleic acid must be well-separated from that corresponding to the protein-bound form. Thus, the composition of the gel matrix must be adjusted to the size of the individual components and the potential complex. For large DNA and RNA molecules (> 500-1000 nts), agarose gels offer a good resolution, whereas for smaller nucleic acids native polyacrylamide gels should be used.

Besides the composition of the gel matrix, several factors need to be considered for successful mobility shift assays: the buffer composition, the temperature, and the electrophoretic dead time (Fried, 1989). These factors influence the stability of the protein-nucleic acid complex before and during the separation and should be optimized for each interaction individually. The conditions used in this work are described in chapter 5.5.

After electrophoretic separation, the nucleic acid is detected in the gel. There are different methods to detect nucleic acids after electrophoretic separation, the most common of which require either the labeling of the nucleic acid or an intercalating dye. Frequently used labels are either based on the incorporation of radioactive  $^{32}\text{P}$  in the nucleic acid backbone or on the covalent coupling of the nucleic acid with a fluorescent dye (Hellman and Fried, 2007; Ruscher et al., 2000). Labels require special equipment for detection by either autoradiography or a fluorescence imager; they are used for both DNA and RNA and allow accurate quantification of the detected signal. Label-free detection is possible by staining the gel with ethidium bromide or other intercalating dyes; however, in this case accurate quantification is impossible. In this work, I used a Cy5 fluorescent dye coupled to the 5'-end of the RNA for detection and quantification.

## 2.2.2 Applications

In most cases, mobility shift assays are used to obtain qualitative information on protein-nucleic acid interactions, as described above. When accurate signal quantification is possible, mobility shift assays can provide information on binding constants, binding stoichiometries and binding cooperativity in the complex. In this work, the mobility shift assays were used to determine approximate dissociation constants ( $K_D$ ) and binding stoichiometries (see chapter 3.1 and 5.5).

Association rates ( $k_{on}$ ) can be determined by mixing the protein and the nucleic acid at known concentrations and then loading the mixture onto the gel after precise time intervals of equilibration (Spinner et al., 2002). To determine the dissociation rates ( $k_{off}$ ), a protein-nucleic acid complex is mixed and equilibrated, and then a competing nucleic acid is added. Samples are taken at different time points during the equilibration process and are separated electrophoretically using a mobility shift assay (Fried and Crothers, 1981). The binding constant ( $K$ ) or the dissociation constant ( $K_D$ ) can be determined from the ratios of free nucleic acid  $[N]$  and bound nucleic acid  $[NP]$  as a function of the protein concentration or as the ratio of the association and dissociations rates (Eq. 2.8 and 2.9). Equation 2.8 is recapitulated in chapter 5.5, equation 5.1.

$$K = \frac{k_{on}}{k_{off}} = \frac{[NP]}{[N][P]} \quad \text{Eq. (2.8)}$$

$$K_D = \frac{[N][P]}{[NP]} \quad \text{Eq. (2.9)}$$

If more than one copy of a given protein binds to a nucleic acid, the copy number of the protein bound to the nucleic acid in the saturated state can be determined by applying mobility shift assays. Similarly, to the determination of the binding constant, a constant amount (concentration) of the nucleic acid is mixed with increasing amounts of protein in small increments and the resulting complexes resolved on a native gel by electrophoresis (Watanabe et al., 2013; Yin et al., 2015). The number of proteins bound to given nucleic acids can then be inferred by the number of shifts induced by increasing protein concentration provided that the  $K_D$  of the two protein binding events are different (Fried, 1989).

If two or more different proteins bind a given nucleic acid, their binding could be cooperative, meaning that the binding of one protein influences the binding of the second one. In this case, the binding order can be inferred by performing mobility shift assays (Narasimhan et al., 2015). For example, if nucleic acid X is bound first by protein A and then by protein B, no shift will be detected if X and B are mixed in the absence of A, whereas X and A will form a complex as well as X-A-B.

In conclusion, electrophoretic mobility shift assays are a classical and sensitive method to obtain qualitative and quantitative information on protein-nucleic acid interactions. If a quantifiable detection method is applied, it also allows determining bindings constants, binding stoichiometries and cooperativity.

## 2.3 Liquid Scintillation Counting

Most elements of the periodic table have isotopes, which share the same number of protons and electrons but differ in the number of neutrons. When an isotope has too many or too few neutrons, it becomes unstable and undergoes spontaneous nuclear changes to reach a stable isotope form. These nuclear changes, or decays, result in the emission of particles or electromagnetic radiation. Such isotopes are radioactive. Three main types of radioactive emissions can be distinguished: the  $\alpha$  emission, which is the release of an  $\alpha$  particle (two protons and two neutrons); the  $\beta$  emission, which is the release of a high energy electron that results from the conversion of a neutron to a proton and a neutrino, and the  $\gamma$  rays, which is a type of high-energy radiation that gets emitted as a byproduct of  $\alpha$  or  $\beta$  decays. Radioactive isotopes used in biological sciences, such as  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ , and  $^{32}\text{P}$ , are all  $\beta$  emitters. The rate of decay from a radioactive to a stable is characteristic for each isotope and reported as half-life. The half-life of a radioactive isotope is the time it takes until half of the original radioactive activity is reached. Radioactive activity is measured in becquerel (Bq) (SI unit) or curie (Ci). 1

Bq is equal to one decay per second, which equals  $2.703 \times 10^{-11}$  Ci. Another common unit is decays per minute (DPM).

In the activity assays performed in this work, the successful enzymatic transfer of a  $^3\text{H}$ -labeled methyl group from the S-Adenosyl-Methionine (SAM) cofactor to the 2' hydroxyl group of an RNA molecule was detected by measuring the  $\beta$  emission from the  $^3\text{H}$  radionuclide in the purified target RNAs. The intensity of  $\beta$  emission and thus the amount of radioactive material in each sample was quantified by liquid scintillation counting (LSC) using either a Pharmacia Wallac 1410, a Tri-Carb (both Perkin Elmer) or a Hidex 300SL liquid scintillation counter (chapter 3.3 and 5.9).

LSC is a radiometric technique for the detection and quantification of  $\beta$ -emitting radionuclides. Before the measurement, the radioactive sample is mixed with a scintillation cocktail consisting of a solvent and scintillator molecules. The solvent is an efficient energy collector for  $\beta$  particles and conducts this energy to the scintillator molecules. Popular solvents are toluene, phenyl xylene, and pseudocumene. Scintillator molecules transform the captured energy into light emissions that can be detected by photomultiplier tubes in the counter. Molecules that are used as scintillators include Butyl PBD (2-[4-biphenyl]-5-[4-*tert*-butylphenyl]-1,3,4-oxadiazole), PPO (2,5-diphenyloxazole) and Bis-MSB(1,4-bis[2-methylstyryl]-benzene).

In the case of a  $\beta$  emission event during the measurement, the emitted high energy electron encounters the solvent, which transfers the energy of this electron to a scintillator molecule. This molecule absorbs the energy moving to an excited state and decays to the ground state by emitting a photon, which is detected by photomultiplier tubes. One  $\beta$  emission event leads to the excitation of multiple scintillator molecules, and the number of excited scintillator molecules is dependent on the emission energy of the  $\beta$  particle. Consequently, the intensity of each detected light pulse is proportional to the emission energy, and the number of light pulses per second equals the number of decays or emission events per second. The duration of the counting experiment depends on the emission energy of the analyzed isotope. Low energy emitters such as  $^3\text{H}$  require longer counting times, as compared to high energy emitters such as  $^{32}\text{P}$ .

The number of detected light pulses per time unit is measured in counts per minute (CPM), which are proportional to the amount of radioactive material in the sample. This assumption is only valid in scintillation measurements with 100% counting efficiency, which means that every  $\beta$  emission produces a detected light pulse. Practically, scintillation cocktails are less than 100 % efficient. The amount of emission that is converted into detectable light pulses is expressed as counting efficiency ( $E$ ). Reduced counting efficiency ( $E$ ) is correlated to signal quenching. Quenching can be caused by low emission energy to light conversion (scintillation efficiency) or by photon loss due to specific sample characteristics. Equation 2.10 describes the definition

of the counting efficiency ( $E$ ) as being the ratio between the counts per minute (CPM) and the actual number of decays per minute (DPM).

$$E = \frac{CPM}{DPM} \quad \text{Eq. (2.10)}$$

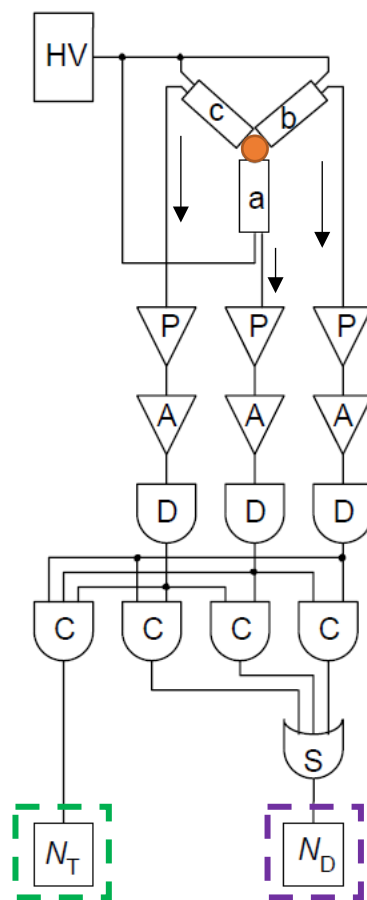
To determine the amount of radioactive material in each sample, the CPM need to be converted into DPM (chapter 5.8, equation 5.3, and 5.4). For this conversion, it is necessary to determine the counting efficiency ( $E$ ). The counting efficiency ( $E$ ) can be determined either by using external standards of known radionuclide concentrations with the same sample composition as in the unknown samples or by applying the triple-to-double coincidence ratio (TDCR) method. In this work, the TDCR method was used to determine the counting efficiency. The TDCR method is implemented in all modern liquid scintillation counters and is an absolute method to determine radionuclide activity (Broda, 2003; Broda and Pochwalski, 1992; Cassette and Bouchard, 2003; Hou, 2018; Pochwalski et al., 1981).

The TDCR method detects the triple coincidences ( $N_T$ ) and the sum of double coincidences ( $N_D$ , including the triple coincidences) simultaneously with a set of three photomultiplier tubes. **Figure 2.3.1** shows to detection set-up according to (Pochwalski et al., 1981). The TDCR derived parameter  $K$  is defined in equation 2.11 and equals the triple-to-double coincidence ratio.

$$K = N_T/N_D \quad \text{Eq. (2.11)}$$

Since  $N_T$  is always lower than  $N_D$ , the parameter  $K$  will always be  $0 \leq K \leq 1$ . If the counting efficiency ( $E$ ) approaches 1,  $N_T$  and  $N_D$  will both approach the number of actual decays per minute (DPM) ( $N_0$ ), resulting in  $K = N_0/N_0 \approx 1$ . Therefore the TDCR parameter  $K$  is a valuable experimental indicator for the counting efficiency ( $E$ ) (Broda, 2003; Pochwalski et al., 1981). The TDCR method has also been improved over the past years yielding special applications for high energy  $\beta$  emitters or the enhanced TDCR method (Broda and Pochwalski, 1992; Kossert, 2010).

In conclusion, liquid scintillation counting using the TDCR method is a fast, reliable, and straightforward method to detect radionuclides in biological experiments or industrial settings.



**Figure 2.3.1.** Schematic and simplified representation of a TDCR counter using three photomultiplier tubes (Pochwalski's model). The three photomultiplier tubes are oriented around the sample position (orange circle), denoted a, b and c. The high voltage source for the photomultipliers is denoted as HV. The signals detected by the photomultiplier tubes are first transferred to the preamplifiers (P), then to the amplifiers (A) and the discriminators (D). The signals get sorted at the coincidence gates (C). Triple coincidences ( $N_T$ ) get counted separately and the sum of all coincidences ( $N_D$ ) gets summed up at the summing gate (S). Signal direction is indicated by arrows. This figure is adapted from (Broda 2003).

## 2.4 Quantitative Mass Spectrometry

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio ( $m/z$ ) of ions. It is a widespread technique to address a variety of biological questions, including protein and nucleic acid identification (e.g., proteomics), analysis of post-transcriptional and post-translational modifications, and protein quantification.

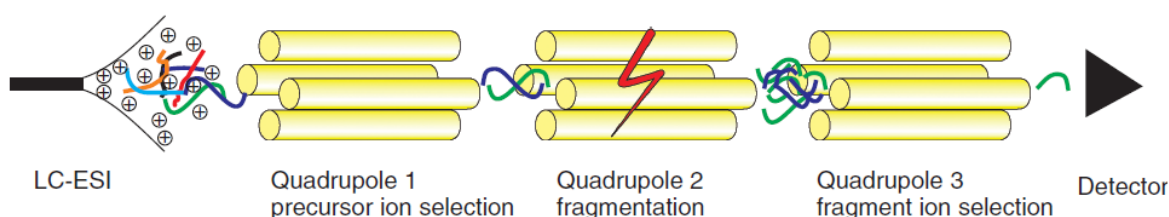
To perform quantitative mass spectrometric analysis, purified protein samples or cell extracts are proteolytically digested (e.g., with trypsin) and separated by liquid chromatography (LC). The liquid chromatography system is usually coupled to the mass spectrometer (LC-MS): the peptides eluting from the chromatography column are ionized and transferred to the mass spectrometer by electron spray ionization (ESI). ESI is considered a soft ionization, as it causes less fragmentation during the ionization, which makes it well suited for macromolecules such as proteins or peptides (Ho et al., 2003).

To use mass spectrometry for relative quantification of protein amounts, several experiment types are available. The simplest method is based on extracted ion chromatograms (XIC or

EIC). Here the entire mass spectrum of the sample is recorded, and after the acquisition, the mass-to-charge ( $m/z$ ) ratios of the desired peptides are extracted, and peak heights or areas are used for relative quantification (Ruse et al., 2013). The disadvantage is the loss of sensitivity due to the lack of selectivity during the measurement. The selected ion monitoring (SIM) assay offers a similar approach to the XIC, with the difference that only pre-selected ions are detected. Due to the increased selectivity, SIM assays are more sensitive than XIC and can be performed on LC-MS or LC-MS/MS systems (Chen, 1979; Murray et al., 2013). Lastly, selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) assays are similar to SIM assays but are performed on triple-quadrupole (QQQ) instruments and provide higher sensitivity compared to SIM (Lange et al., 2008).

### 2.4.1 Selected Reaction Monitoring (SRM)

In this work, selected reaction monitoring (SRM) assays on a triple quadrupole LC-MS system were used to determine the absolute quantities and relative stoichiometries of three proteins in an *in vitro* assembled biomolecular complex (chapter 3.1.4 and 5.8).



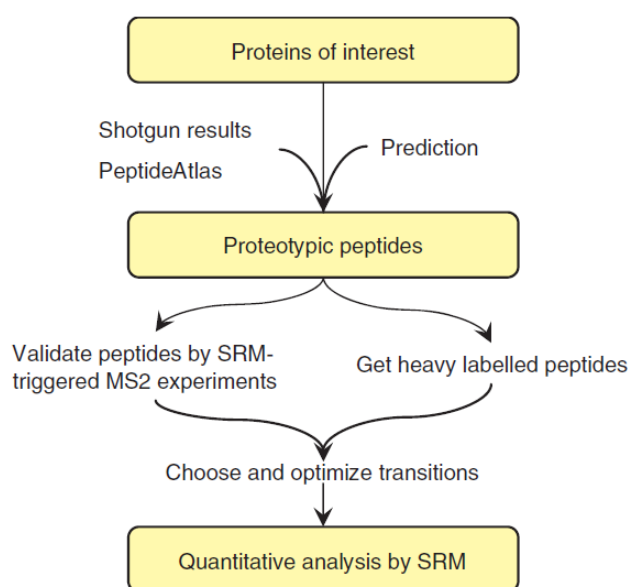
**Figure 2.4.1.** Schematic representation of a selected reaction monitoring experiment performed on a triple-quadrupole electron spray ionization mass spectrometer coupled to a liquid chromatography system (LC-ESI-MS). After the peptides (blue, green, red, light blue and black) elute from the chromatography column they are ionized by electron spray ionization (ESI) and transferred into the mass spectrometer. In quadrupole 1 (Q1) the pre-selected peptide ions are selected (blue and green) and are permitted to enter quadrupole 2 (Q2). Q2 functions as a collision cell and produces fragment ions of the pre-selected peptide ions. These fragment ions enter quadrupole 3 (Q3), which filters pre-selected fragment ions (green) and allows them to enter the detector. This figure is adapted from (Lange et al., 2008).

SRM is a tandem mass spectrometric method performed on a triple-quadrupole instrument (Chen, 1979). The first quadrupole mass analyzer (Q1) is used to filter pre-selected peptide/precursor ions; these are fragmented in the second quadrupole (Q2) that is used as a collision cell. After fragmentation, pre-selected fragment ions are filtered by the third quadrupole (Q3) and transmitted to the detector (**Figure 2.4.1**) (Murray et al., 2013). As Q1 and Q3 act as mass analyzers and permit only the transfer of pre-selected precursor ions (Q1) or fragment ions (Q3), the  $m/z$  ratio of the ions needs to be known and determined before the



set-up of the SRM assay. For this reason, it is necessary to acquire the complete mass spectrum of all proteolytically digested target proteins and select one or more peptide and fragment ions to be detected during the SRM assay (Prakash et al., 2009). Only a few representative peptide ions are chosen per target protein to infer the presence and the quantity of the protein. The correct selection of these peptides is essential for the successful outcome of the SRM assay. During the selection of the peptides several factors need to be considered; these include mass spectrometric properties (detectability), uniqueness, post-translational modifications, and cleavage sites (internal trypsin cleavage sites) (Lange et al., 2008). In this work, two peptides per target protein were chosen.

After the successful set-up of an SRM assay for a given sample, the integrated peak areas of the detected fragment ions can be used for relative quantification of the precursor ions and the intact protein in the sample (Kondrat et al., 1978; Lange et al., 2008; Rauniyar, 2015).



**Figure 2.4.2** Schematized selected reaction monitoring (SRM) assay workflow.

In the first step, proteins of interest need to be identified. Suitable peptides resulting from the trypsin digest of the target proteins are selected from shotgun results, predictions, or databases. The suitability of the selected peptides is then verified, and isotope-labeled peptides are produced or purchased. After choosing and optimizing the fragment ions the quantitative analysis by SRM can be performed. The figure is adapted from (Lange et al., 2008).

Absolute quantification of proteins requires the addition of chemically synthesized isotope-labeled peptides as internal standards during the mass spectrometric measurement. The isotopic-labeled standards need to be added in known and exact amounts so that the integrated peak areas of the labeled standard peptide fragment ions correspond to a known amount of peptide. This information can then be used to determine the amount of non-labeled peptide fragment ions based on the peak area and obtain the absolute amount of target protein present in the sample. The general workflow for SRM assays for absolute quantification is depicted in **Figure 2.4.2**.

A similar set-up to SRM assays is the parallel reaction monitoring (PRM) assay, which follows the same principles but is performed on quadrupole-equipped Orbitrap or time-of-flight (TOF) mass spectrometers. Practically this means that the third quadrupole in the SRM assays is substituted by an Orbitrap ion trap or TOF mass analyzer (Peterson et al., 2012).

## 2.4 X-ray Crystallography

*The theories presented in this section have been mostly adapted from (Blow, 2002). If other sources were used, they are indicated in the respective section.*

X-ray crystallography is one of the most routinely used methods to determine the three-dimensional atomic structures of biomolecules, including proteins, nucleic acids, and complexes thereof. The structure determination by X-ray diffraction relies on the ability of biomolecules to form repetitive ordered objects, namely crystals, that diffract or scatter an incident beam of monochromatic electromagnetic waves, in this case, x-rays, in specific directions. The diffracted waves keep the same wavelength as the incident beam but change the amplitude and phase, dependent on the distribution of the diffracting matter in the crystal. Therefore, the amplitudes and phases of the diffracted spots in the resulting diffraction data contain information on the spatial distribution of the scattering subunits, which can be used to determine the three-dimensional (3D) structure of biomolecules (Blow, 2002).

In this work, x-ray crystallography was used to obtain high-resolution three-dimensional structures of a protein-RNA complex (chapters 3.1.3 and 5.7) and a protein-protein complex (chapters 3.2.2 and 5.7).

### 2.5.1 Crystallization of biomolecules and biomolecular complexes

The prerequisite to obtain the atomic coordinates of a biomolecules by x-ray diffraction is the ability of the respective biomolecule to form well-ordered and diffracting crystals.

Crystallization requires large quantities of the biomolecule or biomolecular complex. Usually, proteins need to be overexpressed recombinantly, as not enough material can be obtained by endogenous purification. Various expression systems may be tested to obtain reasonable amounts of soluble protein. Commonly used expression systems include different strains of *E.coli* and yeast, insect, and mammalian cell lines (Fernandez and Hoeffler, 1999).

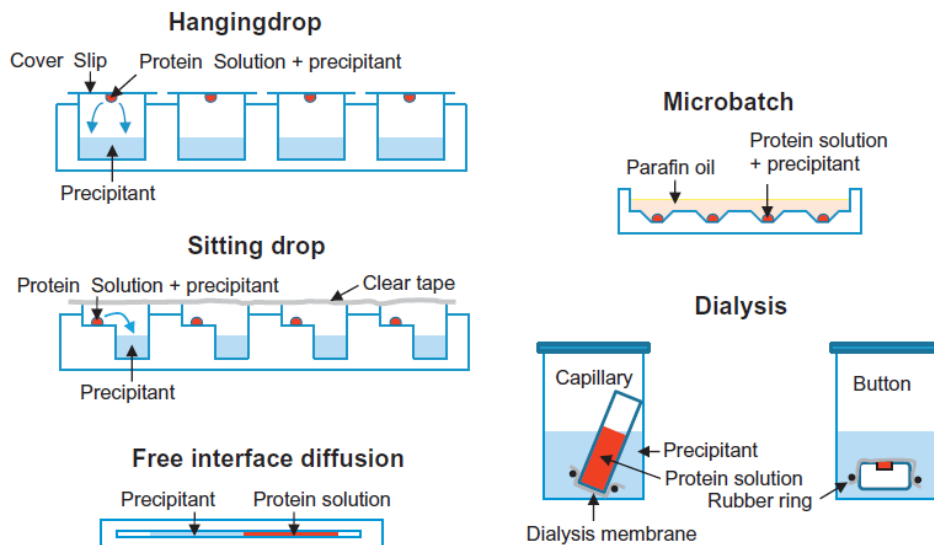
The production of soluble nucleic acids is more straightforward. DNA molecules can be replicated by polymerase chain reaction (PCR) from a plasmid template or double-stranded (ds) DNA oligos; the procedure requires short DNA primers, desoxy-NTPs, and a DNA polymerase, all of which are commercially available. RNA can be produced by *in vitro* transcription from a DNA plasmid template or dsDNA oligos. Solubility is usually not a problem for nucleic acids.

Multi-subunit complexes for crystallization are usually reconstituted *in vitro*. The components are produced separately and are afterward mixed following a specific protocol to yield the final complex.

After production of the material and, if necessary, assembly of the complex, the purity and homogeneity of the sample need to be evaluated, as impurities and inhomogeneities can hinder the formation of an ordered crystal lattice. Depending on the type of biomolecule, different experiments are used to evaluate the purity as well as the chemical and conformational homogeneity. For proteins, SEC (chapter 2.1.1), static and dynamic light scattering (chapter 2.1.2) or gel electrophoresis are the most commonly used techniques (Oliveira and Domingues, 2018). Chemical heterogeneity in proteins can be caused by post-translation modifications, whereas conformational heterogeneity can be an issue in protein complexes or in proteins containing flexible parts. For DNA, agarose gel electrophoresis is often sufficient to check the purity. For RNA, on the other hand, purity as well as chemical and conformational homogeneity should be tested. Chemical inhomogeneities are caused by heterogeneous 5' and 3' end, resulting from the transcription process, primarily if bacteriophage T7 RNA polymerase is used (Pleiss et al., 1998; Triana-Alonso et al., 1995). Conformational inhomogeneity arises from differential RNA folding during the sample preparation. This point needs special care, especially when RNA is crystallized in isolation, whereas the interaction with proteins can help inducing the correct RNA folding (Golden and Kundrot, 2003).

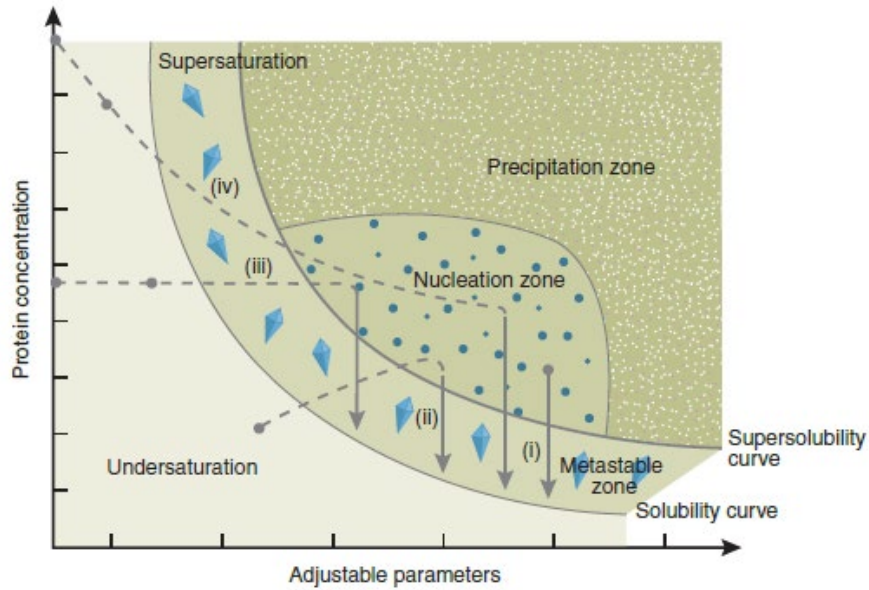
After obtaining pure and homogeneous material and before attempting crystallization, the sample must be concentrated to 2 – 20 mg/ml. Thus, the maximum concentration at which the molecule of interest is stable over an extended time should also be determined before starting crystallization trials.

Crystallization can be attempted with several different methods; the most commonly used ones are microbatch crystallization, vapor diffusion, dialysis, and free interface diffusion (FID) (Chayen and Saridakis, 2008). The set-up of all four crystallization methods is depicted in **Figure 2.5.1** (Rondeau and Schreuder, 2015).



**Figure 2.5.1** Schematic set-up of different crystallization methods. The crystallization methods vapor diffusion by hanging drop and sitting drop, free interface diffusion, microbatch and capillary and button dialysis are depicted. The sample (protein or nucleic acid) solution is in red and the precipitant or reservoir solution is in light blue. In vapor diffusion (hanging or sitting drop) the sample is mixed at different ratios with the reservoir solution and then concentrated by vapor diffusion against a bigger volume of reservoir. In free interface diffusion the reservoir solution and the sample are placed in juxtaposition and mixed by diffusion. In the microbatch experiment the sample-reservoir solution is placed under paraffin oil and in dialysis the initial sample buffer is slowly exchanged against the reservoir solution by diffusion through a semipermeable membrane. This figure is adapted from (Rondeau and Schreuder 2015).

All crystallization methods aim at shifting the sample phase equilibrium to the nucleation zone and from that towards the metastable zone, where crystal growth happens (**Figure 2.5.2**). In crystallization trials different conditions are tested, including the crystallization technique, the initial sample concentration, the temperature and the composition of the precipitant or reservoir solution (pH, type and concentration of precipitant, buffer, and the ionic strength). Because there is still no rationale as to which conditions produce crystals, the initial crystallization trials are essentially based on trial and error. These trials are usually done using commercial crystallization screens that test up to 96 reservoir conditions per screen. If microcrystals or small crystals are obtained at this stage, these conditions can be used to design crystallization fine screens and obtain improved and bigger crystals. Crystals obtained from commercial screens can either be used for data collection or for further crystal refinement (Smyth and Martin, 2000). Suitable crystals are fished and soaked in a cryo-protectant solution (e.g., ethylene glycol, glycerol, or (2*R*, 3*R*)-2,3-butanediol) and flash-frozen in liquid nitrogen. Biomolecular crystals can be stored in liquid nitrogen until they are being measured.



**Figure 2.5.2** Biomolecular crystallization phase diagram. The biomolecule state is depicted as a function of protein (or nucleic acid) concentration (y-axis) and adjustable parameters (x-axis), which include precipitant and additive concentrations, ionic strength, pH and temperature. The dashed lines show the path from the respective starting points (full circles) to the nucleation zone using either (i) Microbatch, (ii) vapor diffusion, (iii) dialysis or (iv) FID crystallization. The full line arrows show the shifts from the nucleation zone to the metastable zone. All four crystallization methods aim at shifting the phase equilibrium first to the nucleation zone and then to the metastable zone. This figure is adapted from (Chayen and Saridakis 2008).

## 2.5.2 Diffraction and data analysis

If suitable biomolecular crystals are available, X-ray diffraction data can be collected. Diffraction data can be collected at a laboratory "in house" source or a synchrotron. Due to the increased availability of synchrotron radiation sources, most data sets are now collected at synchrotron facilities. Synchrotron sources offer highly intense monochromatic and collimated X-ray beams and high-quality optics. Furthermore, charged coupled device (CCD) detectors at synchrotrons have much faster read-out times as compared to x-ray film or imaging plates and, therefore, allow for much shorter exposure times, faster collection of data sets, and potentially reduced radiation damage (Gruner and Ealick, 1995).

When a three-dimensional repetitive object, like a crystal, scatters x-rays, a reflection or diffraction is only produced if all the points in the crystal lattice scatter in phase. This only happens if the crystal is oriented in the beam at a specific angle. This behavior is described by Bragg's law, published by William Lawrence Bragg in 1913 (Bragg and Bragg, 1913). Other proposed models are the Laue equations and the Ewald construction, all three of which lead to the same results (Ewald, 1921; Laue, 1913).

In Bragg's law, the three-dimensional crystal lattice is viewed as a set of parallel lattice planes (Bragg planes), described by the integers  $h$ ,  $k$ , and  $l$ , referred to as the Miller indices (Königlich Preussische Akademie der Wissenschaften zu Berlin, 1816). A lattice is an ordered array of points describing the arrangement of particles that form a crystal. A set of indices describes a particular set of lattice planes, with the index values defining the points at which the planes cut the unit cell edges in terms of fractions of the unit cell dimension, say  $1/h$ ,  $1/k$ , and  $1/l$  (**Figure 2.5.3 A**). This results in lattice planes with equal spacing if the same indices define them. These lattice planes are considered mirrors that produce a diffracted beam or reflection, meaning the scattered waves are in phase. A set of lattice planes, defined by the indices  $h,k$ , and  $l$ , scatter an incoming wave in phase if the difference in the pathlength between the scattered wave of one plane and the scattered wave the second plane equals an integer number of wavelengths ( $n\lambda$ ) (equation 2.12) (Bragg and Bragg, 1913).

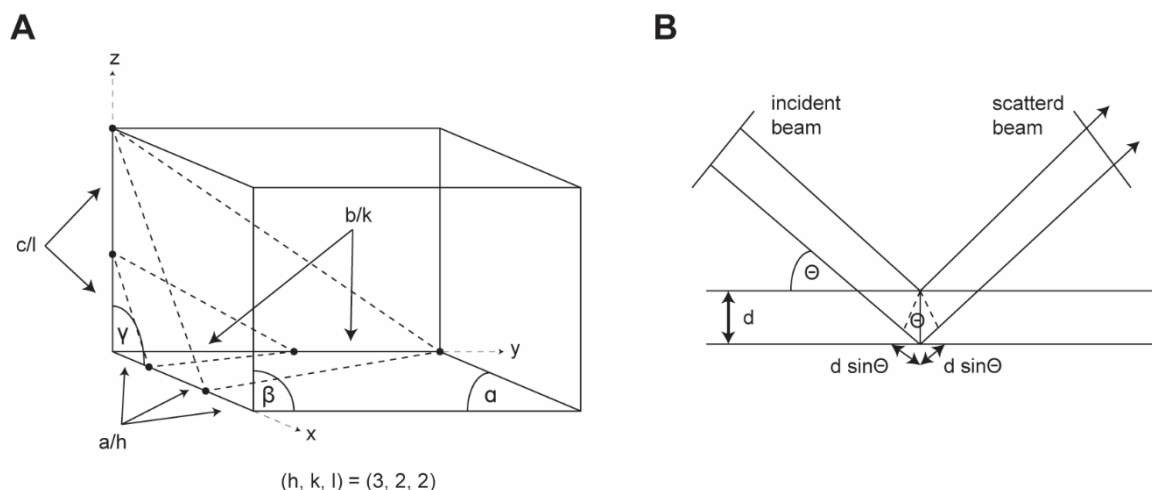
$$n\lambda = 2d_{hkl}\sin\theta \quad \text{Eq. (2.12)}$$

Where  $\lambda$  is the wavelength of the incident beam,  $d$  the distance between the planes and  $\Theta$  the angle between the incoming beam and the plane. Following equation 2.12, the difference in pathlength for a wave scattered by one set of lattice planes equals  $2d_{hkl}\sin\Theta$  (**Figure 2.5.3 B**). For  $n$  being any integer number, the indices can be different; therefore, a more general version of Bragg's law can be written:

$$\lambda = 2d_{hkl}\sin\theta \quad \text{Eq. (2.13)}$$

A smaller spacing between the lattice planes,  $d_{hkl}$ , equals finer details and higher resolution. At a constant wavelength, which is the case for monochromatic X-ray sources, this results in an increase of the incident (or diffraction) angle  $\Theta$ . In other words, diffractions containing high-resolution data scatter at larger angles, and diffractions at smaller angles contain only low-resolution information.

The distances between the diffractions or diffractions spots are related to the unit cell dimension, defined by three lengths ( $a$ ,  $b$ , and  $c$ ) and three angles ( $\alpha$ ,  $\beta$ , and  $\gamma$ ). The unit cell is the smallest repetitive unit that builds the crystal and is represented by lattice points in the crystal lattice. The shape of the diffraction spots is determined by the crystal system (**Table 2.5.1**) and the space group by the symmetry of the diffraction pattern. Space groups define symmetry patterns in three dimensions; for chiral molecules, 230 space groups are defined.



**Figure 2.5.1.** Schematic representation of lattice planes and Bragg diffraction by two lattice planes. **(A)** Simplified representation of unit cell with the cell dimensions  $a$ ,  $b$ ,  $c$ ,  $\alpha$ ,  $\beta$  and  $\gamma$ . Lattice planes defined by the Miller indices  $h$ ,  $k$  and  $l$  equal to 3, 2 and 2 are indicated in dashed lines. The Miller indices define at which points the lattice planes cut the unit cell edges as fractions of the cell dimensions ( $a/h$ ,  $b/k$  and  $c/l$ ). **(B)** Graphical explanation of Bragg's law described in equation 2.13. The incident beam with the wavelength  $\lambda$  hits the lattice planes at the angle  $\Theta$ . The distance between the lattices planes equals  $d$ . The difference in pathlength between the wave scattered from the first lattice plane and the second lattice plane is  $2d\sin\Theta$ .

**Table 2.5.1.** Overview of crystal systems. This table is adapted from (Blow, 2002).

| Crystal system | Minimum symmetry requirement    | Constraints on interaxial angles and axial length   |
|----------------|---------------------------------|---|
| Triclinic      | None                            | None  |
| Monoclinic     | One 2-fold axis                 | $\alpha$ and $\gamma = 90^\circ$  |
| Orthorhombic   | Three perpendicular 2-fold axes | $\alpha$ , $\beta$ and $\gamma$ all $90^\circ$  |
| Trigonal       | One 3-fold axis                 | $\beta = 120^\circ$ , $\alpha$ and $\gamma = 90^\circ$<br><b>a</b> and <b>b</b> equal length    |
| Tetragonal     | One 4-fold axis                 | $\alpha$ , $\beta$ and $\gamma$ all $90^\circ$<br><b>a</b> and <b>b</b> equal length            |
| Hexagonal      | One 6-fold axis                 | $\beta = 120^\circ$ , $\alpha$ and $\gamma = 90^\circ$<br><b>a</b> and <b>b</b> equal length    |
| Cubic          | Four 3-fold axis                | $\alpha$ , $\beta$ and $\gamma$ all $90^\circ$<br><b>a</b> , <b>b</b> and <b>c</b> equal length |

At a fixed wavelength, the crystal needs to be rotated through the incident beam to find incident angles at which diffractions are produced. How much data has to be collected from a crystal for structure determination depends on the level of crystallographic symmetry in the crystal, the level of non-crystallographic symmetry, meaning the amount of symmetry in the asymmetric unit, and the upper-resolution limit that can be obtained with the crystal (Smyth and Martin, 2000).

After sufficient data is collected and the crystal system, the unit cell dimension, and the crystal orientation in the beam are known, diffraction spots are picked and assigned to the corresponding lattice planes and Miller indices. For each diffraction spot, a position and an intensity are recorded during the experiment. The position is determined by the crystal system and the unit cell dimension, which are already known. The intensity results from the amplitude and the phase of the diffracted waves and contains information on both parameters. Each diffracted wave keeps the wavelength  $\lambda$  of the incident beam but changes its phase and amplitude, dependent on the distribution of the diffraction matter in the crystal. Therefore, the phase and the amplitude of the diffracted wave contain information on the spatial distribution of the scattering matter. Once each diffraction spot has been indexed, the intensities can be determined. Due to factors such as radiation damage and imperfections at the x-ray source, the detector, and the crystal conditions during the measurement, the measured intensities are also subject to errors. To evaluate the quality of the obtained or observed intensities, the R or reliability factor is used, which is defined as

$$R = \frac{\text{mean difference between values which should be the same}}{\text{mean value of measured values}} \quad \text{Eq. (2.14)}$$

and measures the accuracy of the measured intensities. A small R value indicates accurate intensities.

The crystal disorder and the disorder caused by thermal vibrations leads to a decrease in the intensities. This effect is accounted for by the Debye-Waller factor or B factor. According to the definition of the B factor, the ideal intensity ( $I_0$ ) is reduced by the factor  $\exp[2B(\sin\theta/\lambda)^2]$  to obtain the observed intensity  $I$ , with B being the value of the B factor expressed in  $\text{\AA}^2$  (Blow, 2002; Debye, 1913; Waller, 1923).

### 2.5.3 Electron density map and model building

To gain the positions of atoms in the unit cell from the obtained intensities, the crystallographic structure factor needs to be determined. To determine the structure factor  $F_{hkl}$ , the structure factor amplitude  $|F_{hkl}|$  and the phase angles  $\varphi_{hkl}$  need to be known unambiguously.

$$F_{hkl} = |F_{hkl}| \exp[i\varphi_{hkl}] = \sum_i^N f_i \exp [2\pi i(hx_i + ky_i + lz_i)] \quad \text{Eq. (2.15)}$$

The scattering factor  $f_i$  is a measure for the amplitude of a diffracted wave which depends on the nature of the incident radiation (e.g., X-rays, electrons or neutrons) and the atomic number (Z) the scattering atom. The expression  $2\pi(hx+ky+lz)$  is the phase factor in the three dimensions x, y, and z. The summation is over all atoms  $N$  in the unit cell, which can account



for multiple copies of the scattering biomolecule. To account for the disorder caused by thermal vibration and crystal disorder, the Debye-Waller factor added to equation 2.16.

$$F_{hkl} = \sum_i^N f_i \exp[2\pi i(hx_i + ky_i + lz_i)] \exp[-B_i \sin^2 \theta / \lambda^2] \quad \text{Eq. (2.16)}$$

The intensity is proportional to  $|F_{hkl}|^2$ ; therefore, the amplitudes can be directly derived from the intensities. The phase angles  $\varphi_{hkl}$ , on the other hand, cannot be directly determined from the intensities. The three most frequently used methods to obtain the phase angles in biomolecular X-ray crystallography are multiple isomorphous replacement (MIR), multi-wavelength anomalous dispersion (MAD) and molecular replacement (MR). In this work, molecular replacement was used for both data sets to obtain the phase angles.

Molecular replacement is the fastest method to obtain phase angles but requires the availability of a closely related structure. The existing structure is placed into the unit cell in the same orientation as the molecule of unknown structure. The structure factors back calculated from the template structure are used to derive the phase angles of the  $F_{hkl}$  factors of the unknown structure. Once the phase angles are known, the structure factors for the new unknown structure can be calculated. The process is repeated iteratively, until convergence is reached. More details can be found in (Blow, 2002; Evans and McCoy, 2008).

If no closely related structure is available, the phase angles need to be obtained by multiple isomorphous replacement (MIR) or multi-wavelength anomalous dispersion (MAD). In MIR one native or parent crystal and at least two derivatized crystals need to be measured to obtain unambiguous phase angles. Derivatized crystals can be obtained by soaking a parent crystal in a heavy metal salt solution (e.g., mercury, platinum, or gold) or by co-crystallizing the biomolecule with the heavy metal salts. The soaking or co-crystallization should yield an isomorphous crystal to the parent crystal, meaning it has the same unit cell dimension and symmetry, differing only in the presence of a few heavy metal atoms. As the presence of additional atoms usually disturbs the structure, the achieved isomorphism is not perfect but can be tolerated to a certain degree. If the degree of isomorphism is within the tolerated range, the hypothesis behind the isomorphous replacement method is that the diffraction of the derivatized crystal  $F_{PH}$  is the sum of the diffraction of the parent or native crystal  $F_P$  and the added heavy atoms  $F_H$ . This hypothesis can be expressed in the isomorphous replacement equation (equation 2.17).

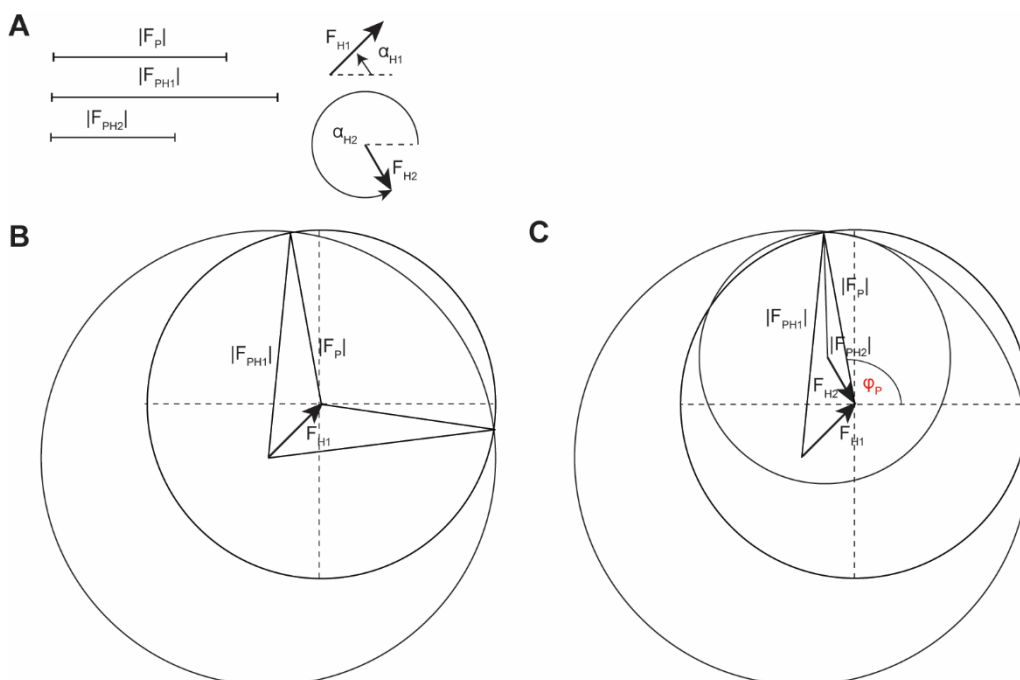
$$F_{PH}(hkl) = \sum_N f_i \exp[2\pi i(hx_i + ky_i + lz_i)] \exp\left[-\frac{B_i \sin^2 \theta}{\lambda^2}\right] + \sum_M f_{H,i} \exp[2\pi i(hx_{H,i} + ky_{H,i} + lz_{H,i})] \exp\left[-\frac{B_{H,i} \sin^2 \theta}{\lambda^2}\right]$$

Eq. (2.17)

Where  $f_i$  is the scattering factor of the native or parent crystal and  $f_{H,i}$  is the scattering factors of the heavy atoms. The sum of the scattering factors  $f_i$ , the phase factors and the Debye-Waller factors over all atoms  $N$  in the parent or native crystal yields the structure factor of the parent crystal  $F_P$  and the sum of the scattering factors  $f_{H,i}$ , the phase factors and the Debye-Waller factors over all heavy atoms  $M$  yields the structure factor of the heavy atoms  $F_H$ . Therefore, equation 2.17 can be simplified to:

$$F_{PH}(hkl) = F_P(hkl) + F_H(hkl) \quad \text{Eq. (2.18)}$$

From the collected diffraction data of the parent and the derivatized crystals the structure factor amplitudes  $|F_P|$  and  $|F_{PH}|$  can be determined. To apply isomorphous replacement the structure factor  $F_H$  for each heavy needs to be calculated. The structure factor amplitude  $|F_H|$  can be obtained as the difference between  $|F_{PH}|$  and  $|F_P|$ , these can be applied in the difference Patterson function to determine the position of the heavy atoms. The atom coordinates can be used to calculate the structure factor  $F_H$  from equation 2.16 and obtain the phase angles  $\varphi_H$ . If  $F_H$ ,  $|F_{PH}|$  and  $|F_P|$  of at the parent at least two derivatives are known the phase can be unambiguously determined as described in **Figure 2.5.2**. More detailed description can be found in (Ke, 1997; Terwilliger and Eisenberg, 1987).



**Figure 2.5.2.** Harker diagram for unambiguous phase determination using multiple isomorphous replacement. **(A)** Geometric representation of the experimentally determined structure factor amplitudes  $|F_P|$ ,  $|F_{PH1}|$  and  $|F_{PH2}|$  and the calculated structure factors  $F_{H1}$  and  $F_{H2}$ , with amplitude and phase angle. **(B)** The Harker diagram using only  $|F_P|$ ,  $|F_{PH1}|$  and  $F_{H1}$  resulting in two possible phase angles for  $F_P$ . **(C)** The Harker diagram using structure factor amplitudes  $|F_P|$ ,  $|F_{PH1}|$ ,  $|F_{PH2}|$  and calculated structure factors  $F_{H1}$  and  $F_{H2}$  results in the determination of one unambiguous phase angle  $\varphi_P$  (red) for the calculation of the structure factor  $F_P$ . In the Harker diagram the center of the circle with radius  $|F_{PH}|$  is used as the origin of vector  $F_{H1}$ . The end point of  $F_{H1}$  is used as the center of the circle with the radius  $|F_P|$ . The end point of vector  $F_{H2}$  is the end of vector  $F_{H1}$ . The angle of the vector amplitudes  $|F_P|$ ,  $|F_{PH1}|$  and  $|F_{PH2}|$  is determined by the circle intersection points.

Another method to obtain phases is multi-wavelength anomalous dispersion (MAD) method. To obtain the phases from multi-wavelength anomalous dispersion a crystal containing anomalous scattering atoms is produced and measured at different wavelength of the incident X-ray beam. An element becomes an anomalous scatterer if the frequency of oscillation of the incident beam is close to the oscillation frequency of the electron shell of the element. Anomalous scattering results in a shift in the amplitude and the phase of the scattered beam. This behavior leads to a violation of Friedel's law, which states that diffractions described by the structure factor  $F_{hkl}$  and its inverse  $F_{-h,-k,-l}$ , which are referred to as Friedel mates, have the same amplitude and the same phase angle but with opposite signs ( $+\varphi$  and  $-\varphi$ ) and therefore equal intensity. Anomalous scattered diffractions described by  $F_A(h,k,l)$  and its inverse  $F_A(-h,-k,-l)$  have the same amplitude but not opposite phase angles. These differences resulting from anomalous scattering can be observed in the diffraction data and used to determine the position of the anomalous scatterer and determine the phase angles applying a similar approach as in isomorphous replacement. As anomalous scattering is only observed at certain wavelength the same crystal can be used to obtain a normal scattering data set and an anomalous scattering data set by shifting the wavelength of the incident beam, which can be done at synchrotron radiation sources. To obtain phase angles for protein crystals, methionines in the primary structure of the protein can be substituted with selenomethionines to introduce selenium as an anomalous scatterer. More detailed information can be found in (Hendrickson, 1991; Smith, 1998).

If both the amplitudes  $|F_{hkl}|$  and the phase angles  $\varphi_{hkl}$  are known, the structure factors can be determined according to equation 2.15. Equation 2.15 can also be reformulated to equation 2.19.

$$F_{hkl} = V \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) \exp[2\pi i(hx + ky + lz)] dx dy dz \quad \text{Eq. (2.19)}$$

$V$  represents the volume of the unit cell, which arises because  $F_{hkl}$  is expressed as electrons per unit cell, whereas the electron density  $\rho(x,y,z)$  is expressed in electrons per unit volume. Three integrals are calculated from 0 to 1 in  $x$ ,  $y$ , and  $z$  to cover the whole unit cell, and

$2\pi(hx+ky+lz)$  is again the phase factor. The electron density equation (equation 2.17) is the inverse Fourier transform of the structure factor equation (equation 2.19). It describes the electron density (the scattering power) of all atoms in the unit cell.

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} F_{hkl} \exp[-2\pi i(hx + ky + lz)] \quad \text{Eq. (2.20)}$$

Solving equation 2.20 for every set of three-dimensional coordinates in the unit cell gives the electron density for each position within the unit cell; all electron densities combined together yield the electron density map. The electron density map determines the contours of the scattering matter in the unit cell as the X-rays are scattered by the electron cloud of each atom in the biomolecule. The contours mapped out by the electron density are used to build the atomic model of the biomolecule. The higher the upper resolution-limit of the acquired diffraction data, the easier it is to build the atoms of the biomolecule into the electron density map.

Once an initial structural model has been built, this model can be refined in a process called structural refinement. To evaluate improvements in the structural model introduced during the refinement process, different parameters can be measured.

To measure the discrepancy between the observed structure factor amplitudes  $|F_{obs}|$  and the calculated structure factor amplitudes  $|F_{calc}|$  from the current model, the R factor, as defined in equation 2.21, is calculated.

$$R = \frac{\sum_h ||F_{obs}| - |F_{calc}||}{\sum_h |F_{obs}|} \quad \text{Eq. (2.21)}$$

To overcome the bias introduced by using the same data for active refinement and monitoring, the "unbiased" R factor or  $R_{free}$  has been introduced (equation 2.22). For that, a set of randomly chosen diffractions (test set) are selected and excluded from the refinement process. Thus, these observed structure factor amplitudes are not used during refinement, and improvements in the  $R_{free}$  better reflect real improvements made through the refinement process.

$$R_{free} = \frac{\sum_{test\ set} ||F_{obs}| - |F_{calc}||}{\sum_{test\ set} |F_{obs}|} \quad \text{Eq. (2.22)}$$

Similar to the R factor used to measure the accuracy of measured intensities, the smaller the value of the R factor or the  $R_{free}$  factor, the more accurate is the model. Therefore, the strategies used in structural refinement aim at reducing the value of the R factor.

Another way to assess the quality of the structural model is by examining the electron difference ( $F_o - F_c$ ) map. The  $F_o$  map results from using structure factors calculated from  $|F_{obs}|$  and phases calculated from the model  $\varphi_{calc}$ . The  $F_c$  map results from structure factors calculated with  $|F_{calc}|$  and  $\varphi_{calc}$ . The difference ( $F_o - F_c$ ) approaches zero in areas where the

model reflects the experimental data, is positive in areas where the model predicts no density but there is some experimentally, and is negative in areas where the model predicts density but there is none experimentally.

Model building and structural refinement result in the three-dimensional coordinates (x,y, and z) of every atom of the biomolecule, which define the three-dimensional structure.

## 3. Results

### 3.1 Interaction of eukaryotic Snu13 and archaeal L7Ae with non-canonical kink-turns and kink-loops

#### 3.1.1 Binding specificities of archaeal L7Ae and eukaryotic Snu13 to kink-loop and non-canonical kink-turn motifs

For the guide sRNA to be positioned correctly with respect to the methyltransferase within the archaeal Box C/D enzyme, two copies of the RNA-binding protein L7Ae need each to bind a kink-turn and/or kink-loop structural motif formed by the guide sRNA (Aittaleb et al., 2003; Ye et al., 2009; Lapinaite et al., 2013). These structural motifs are formed by the Box C, D, C' and D' sequence motifs, which are all evolutionary conserved (Huang and Lilley, 2018).

In eukaryotic Box C/D guide snoRNAs the Box C and D motifs are well conserved, while the C' and D' motifs have experienced less evolutionary pressure and are therefore much less conserved (Watkins et al., 2002; van Nues et al., 2011). This has often led to the loss of one of the two crucial G•A base pairs, which has a potential impact on the ability of the sequence to successfully form kink-turn structures. These deviations from the consensus sequence can therefore also have an impact on the ability of eukaryotic Snu13 to bind to this site. So far binding of Snu13 to a guide RNA has been shown for U3 and U24 snoRNAs from *S. cerevisiae*, which bind two copies of Snu13 with two canonical kinked structures (Qu et al., 2011; Barandun et al., 2017).

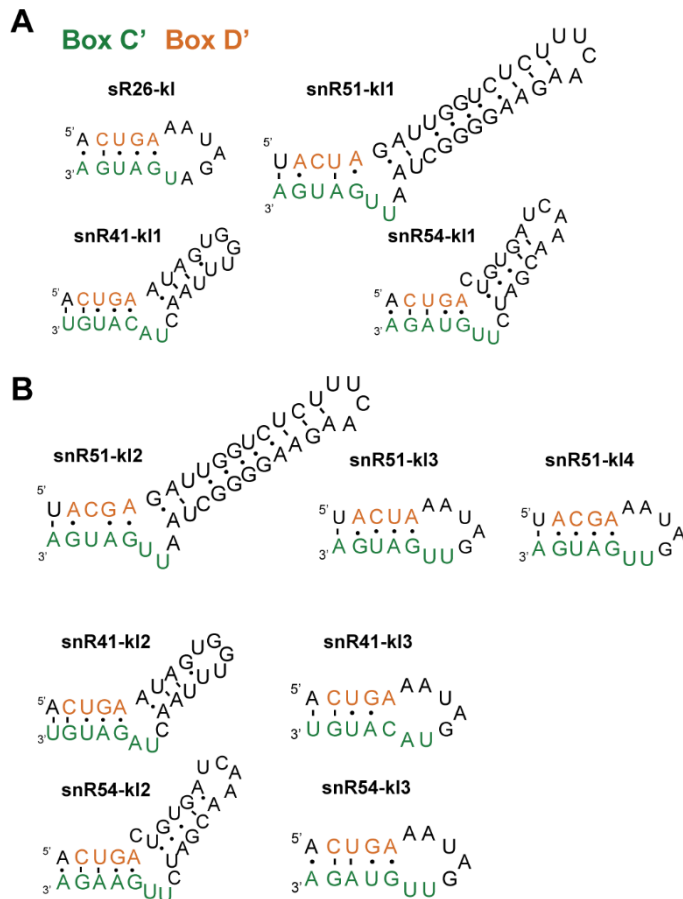
To study the differences between the k-turn binding properties of archaeal and eukaryotic proteins, we selected the archaeal L7Ae protein from the hyperthermophilic archaeon *Pyrococcus furiosus*, with which this lab has previously worked, and Snu13 from the eukaryotic model organism *Saccharomyces cerevisiae* that shares 83.6% primary sequence similarity and 68.0% identity with human Snu13 (Uniprot-ID: P55769) \*.

To gain a better understanding of which RNA sequence elements in the kink-turn or kink-loop are required for binding of L7Ae or Snu13, we selected a set of naturally occurring archaeal and eukaryotic Box C/D guide RNAs and tested the affinity of both proteins towards the kinked structures formed by their C' and D' motifs. For each of these RNAs, we isolated the internal

---

\* All similarity and identity values are derived from pair-wise primary sequence alignments calculated using the Needleman-Wunsch algorithm implemented in the EMBL alignment tool NEEDLE (Madeira et al., 2019).

kink-turn or kink-loop sequence (see **Table 5.3**) and used these sequences to evaluate protein binding. The tested sequences were derived from archaeal guide RNA sR26 from *P. furiosus* (sR26-kl), and eukaryotic guide RNAs snR51 (snR51-kl1), snR41 (snR41-kl1), snR54 (snR54-kl1) from *S. cerevisiae* (**Figure 3.1.1 A**).



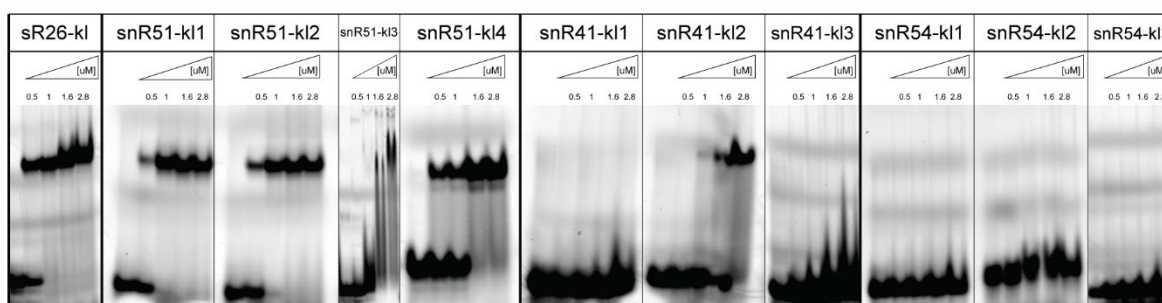
**Figure 3.1.1.** Native and mutant internal kink-turns and kink-loops. The Box C' sequence motif is highlighted in green, the Box D' sequence is highlighted in orange. **(A)** Natively occurring internal kink-turns and kink-loops from archaeal guide RNA sR26 (sR26-kl) and eukaryotic guide RNAs snR51 (snR51-kl1), snR41 (snR41-kl1) and snR54 (snR54-kl1). **(B)** Mutant variants of snR51-kl1, snR41-kl1 and snR54-kl1. Mutants -kl2 reintroduce the second G•A base pairs before the kink. Mutants -kl3 replace the helix after the kink with the loop structure from sR26. Mutant -kl4 reintroduces the second G•A base pair and replaces the helix after the kink with the loop structure of sR26. Note that the RNAs depicted here do not show the three additional nucleotides at the 5' end and the Cy5 label (see Table 3.3).

For each of the eukaryotic RNAs, we designed two to three mutants: the first mutant reintroduced the second G•A base-pair if not present in the wild type; the second mutant substitutes the helix in stem I with the loop structure from sR26-kl; the third mutant combines the previous two mutations (**Figure 3.1.1 B**)

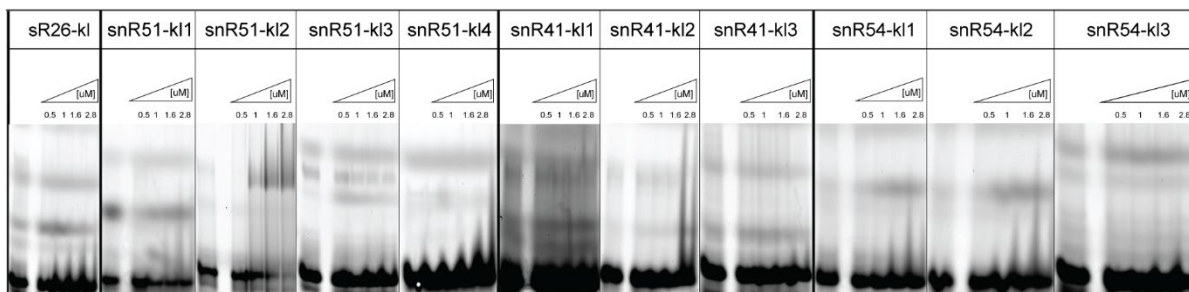
To evaluate the affinities and binding behavior of both the archaeal and eukaryotic protein we performed electrophoretic mobility shift assays (EMSA) with 5' Cy5-labeled RNAs (**Table 5.3**) and unlabeled proteins (described in sections 2.2 and 5.5). To confirm that the Cy5-label and the linker nucleotides are not influencing the folding of the RNA we also tested a subset of unlabeled RNA without linker nucleotides and detected the RNA with ethidium bromide (**Figure 3.1.3**) (see section 5.5). Notably, due to a loss of sensitivity in the RNA detection 50 times higher concentrations had to be used.

Archaeal L7Ae exhibits strong binding towards the archaeal kink-loop from sR26 (sR26-kl), with an estimated  $K_D \sim 0.4 \mu\text{M}$ . We determined a slightly lower affinity for L7Ae binding the wild-type internal kink-turn of eukaryotic guide RNA snR51 (snR51-kl1), which contains only the 1n-1b G•A base-pair directly before the kink-turn in stem II, but lacks the 2n-2b G•A base pair (**Figure 3.1.1 A** and **Figure 3.1.2**). The re-introduction of the second G•A base pair in snR51-kl1 resulting in snR51-kl2 (**Figure 3.1.1 B**) does not increase the affinity of L7Ae towards this RNA (**Figure 3.1.2** and **Table 3.1.1**). By replacing stem I in snR51-kl1 with the loop structure from sR26 (snR51-kl3) the affinity of L7Ae visibly drops but can be partly rescued by re-introducing the second G•A into snR51-kl3, as in snR51-kl4 (**Figure 3.1.1 B**, **Figure 3.1.2** and **Table 3.1.1**).

**A**



**B**



**Figure 3.1.2.** Electrophoretic mobility shift assays (EMSA) monitoring binding of archaeal L7Ae and eukaryotic Snu13 to a selected set of kink-loops and internal kink-turns. **(A)** EMSA titrations of increasing amounts of archaeal L7Ae onto a constant amount of different RNA constructs. **(B)** EMSA titrations of increasing amounts of eukaryotic Snu13 onto a constant amount of different RNA constructs. In **(A)** and **(B)** the RNA concentration was kept constant at  $2 \mu\text{M}$ . The first lane for each assay contains RNA only and serves as a control. Protein concentrations are indicated in  $\mu\text{M}$  on top of the respective lane in each assay.

These results showed that binding is affected negatively by substitution of the helix following the kink by a loop in the absence of the G•A base pair in position 2n-2b.

The natively occurring internal kink-turn of eukaryotic guide RNA snR41 (snR41-kl1) contains only the 2n-2b G•A base pair but lacks the 1n-1b G•A base pair in stem II. Archaeal L7Ae cannot bind this internal kink-turn structure (**Figure 3.1.1 A** and **Figure 3.1.2 A**).



By re-introducing the 1n-1b G•A base-pair the binding of L7Ae to snR41-kl2 can be rescued, with an affinity in the  $\mu\text{M}$  range (**Figure 3.1.2 A** and **Table 3.1.1**). Also, for this internal kink-turn structure, the replacement of stem I with the loop structure from sR26-kl does not improve the affinity of L7Ae to this RNA as compared to the wild type. The results for snR41-kl1 and its mutants in combination with the ones for snR51-kl1 and its mutants demonstrated that the 1n-1b G•A in stem II is critical for binding of L7Ae to internal kink-turn structural motifs; the presence of both G•A base-pairs is critical when the helix after the kink-turn is replaced by a loop.

Contrary to our expectations, we could not observe any binding of L7Ae toward the native internal kink-turn structure from eukaryotic guide RNA snR54 (snR54-kl1), even though the 1n-1b G•A base pair is present in this motif. Also, after re-introducing the 2n-2b G•A base-pair (snR54-kl2) no binding could be observed. As expected, replacement of stem I with the sR26-kl loop structure did not elicit any binding (**Figure 3.1.2 A**). We concluded that this box C'/D' motif does not form a kink-turn structure.

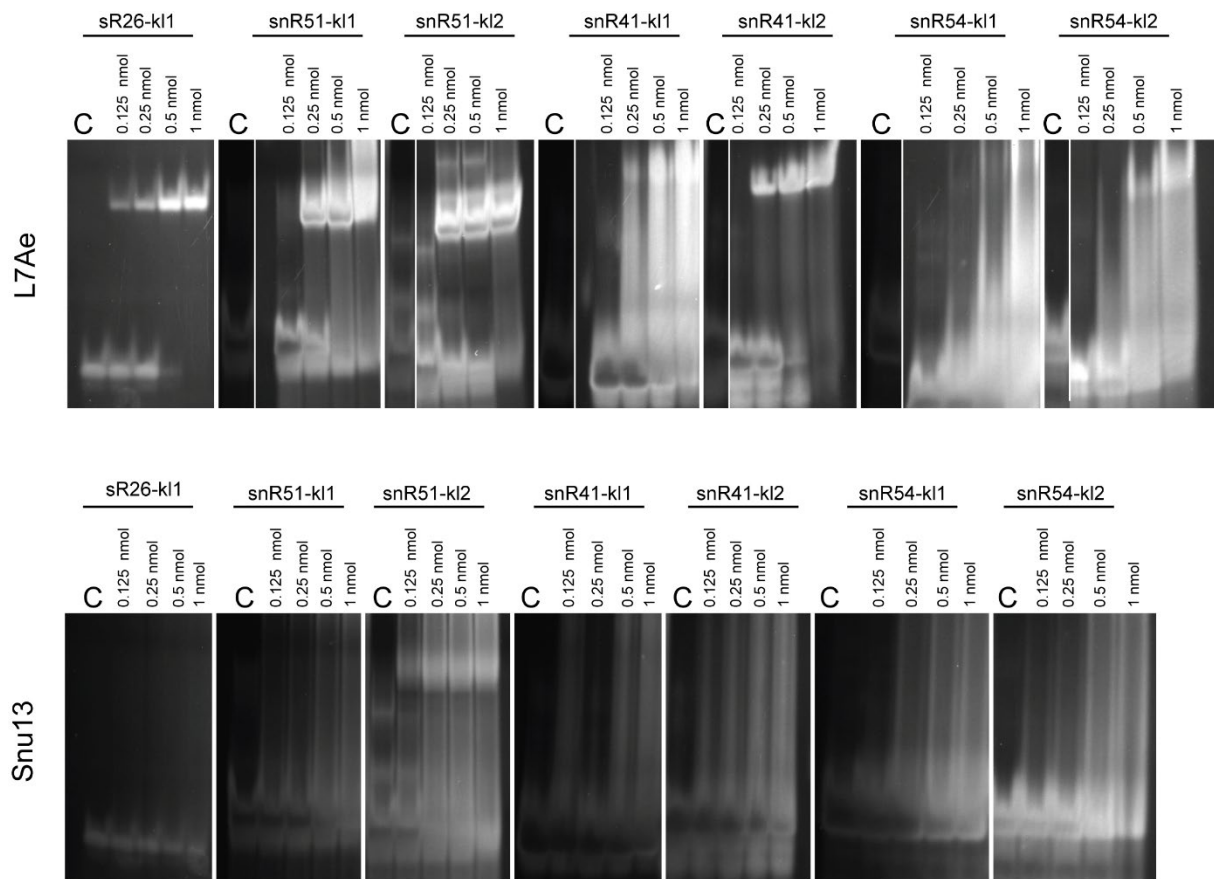
**Table 3.1.1.** Calculated  $K_D$  values and fitting parameters for L7Ae binding to sR26-kl, snR51-kl1, snR51-kl2, snR51-kl3, snR51-kl4 and snR41-kl2. Errors are given as standard deviations. For detailed methods description see chapter 5.5.

|   | sR26-kl                           | snR51-kl1                                   | snR51-kl2                         | snR51-kl3                         | snR51-kl4                         | snR41-kl2                         |
|---|-----------------------------------|---|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| A1  | $0.001 \pm 0.02$                  | $0.001 \pm 5.59\text{E-}4$                  | $0.002 \pm 0.05$                  | $0.005 \pm 0.005$                 | $0.009 \pm 0.17$                  | $0.003 \pm 0.02$                  |
| A2  | $0.91 \pm 0.02$                   | $0.82 \pm 5.15\text{E-}4$                   | $0.84 \pm 0.054$                  | $0.46 \pm 0.01$                   | $0.79 \pm 0.39$                   | $1.58 \pm 1.14$                   |
| p   | $2.59 \pm 0.63$                   | $3.98 \pm 0.016$                            | $3.50 \pm 0.37$                   | $12.84 \pm 0.77$                  | $1.91 \pm 1.39$                   | $1.39 \pm 0.41$                   |
| R <sup>2</sup>                                  | 0.99                              | 1   | 0.97                              | 0.99                              | 0.65                              | 0.99                              |
| <b>K<sub>D</sub> [<math>\mu\text{M}</math>]</b> | <b><math>0.41 \pm 0.02</math></b> | <b><math>0.56 \pm 5.44\text{E-}4</math></b> | <b><math>0.55 \pm 0.05</math></b> | <b><math>1.17 \pm 0.01</math></b> | <b><math>0.81 \pm 0.59</math></b> | <b><math>3.55 \pm 3.35</math></b> |

We also tested the same set of RNAs with the eukaryotic homolog Snu13 under the same conditions. We found that Snu13 binds only snR51-kl2 in the tested affinity range, which contains both G•A base pairs and stable helical structure after the kink. A very low-affinity binding is also observed for snR41-kl2, which again contains both G•A base pairs. However, Snu13 does not bind snR51-kl4, which has both G•A base pairs but a loop in the position of stem I. All other wild type and mutant RNA sequences derived from sR26, snR51 and snR41 do not bind Snu13 (**Figure 3.1.2 B**). Also, Snu13 does not bind any kinked structure derived from snR54, which again hinted at the fact that this sequence does indeed not form a kink-turn structure. These results indicated that both G•A base pairs are required for binding of Snu13 as well as a helical structure after the kink-turn.

These findings showed that Snu13 has a generally lower binding affinity towards internal kink-turns and kink-loops, which also indicates that auxiliary factors might be needed to help Snu13 bind these structures and keep it bound in the cellular context.

We showed that archaeal L7Ae can bind both internal kink-turn as well as kink-loops, requires both G•A base-pairs to bind kink-loops, while only the 1n-1b G•A base pair is necessary for L7Ae to bind an internal kink-turn structure. On the other hand, eukaryotic Snu13, requires both G•A bases-pairs and a helix in stem I.

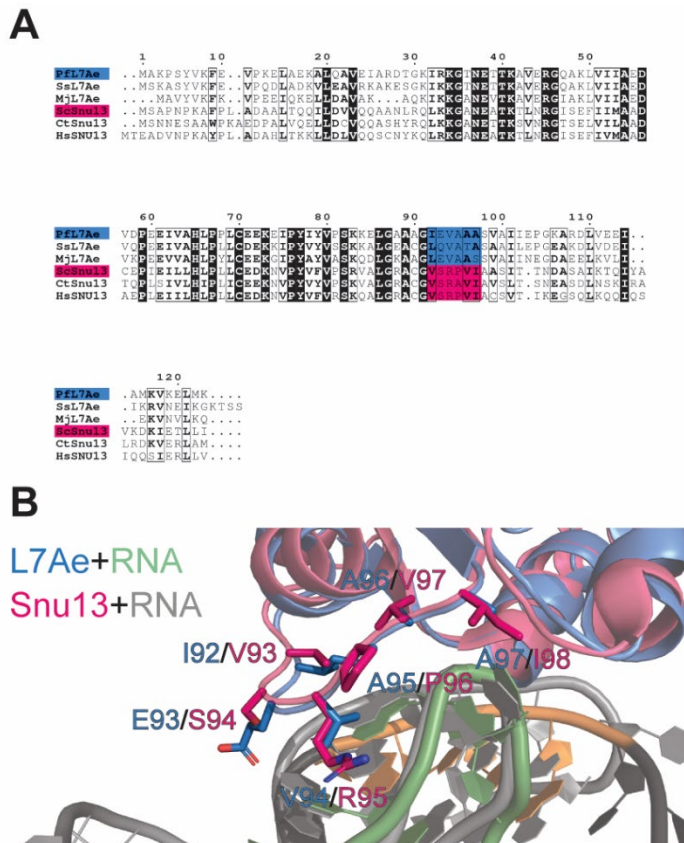


**Figure 3.1.3.** Electrophoretic mobility shift assays of unlabeled native C'/D' elements and mutants with archaeal L7Ae and eukaryotic Snu13. The first lane contains 0.5 nmol of RNA as a control, denoted C. The respective protein, denoted on the left side of the panel, was titrated at 0.125, 0.25, 0.5 and 1 nmol. The RNA used is indicated on top of each assay. Note that for assays stained with ethidium bromide 50 times higher concentrations were used as compared to the fluorescence-based assays, this can cause a slightly different binding behavior.

### 3.1.2 Six amino acids fine-tune binding of L7Ae and Snu13 to internal kink-turn structures

Results from the previous section indicated that eukaryotic Snu13 cannot bind kink-loop structures, while archaeal L7Ae can bind these structures in the presence of both G•A (1n-1b and 2n-2b) base pairs. In a previous study (Gagnon et al., 2010) proposed that six residues located in loop 9 of *M. jannaschii* L7Ae and *M. musculus* Snu13 influence the ability of the archaeal or eukaryotic protein to bind kink-loops.

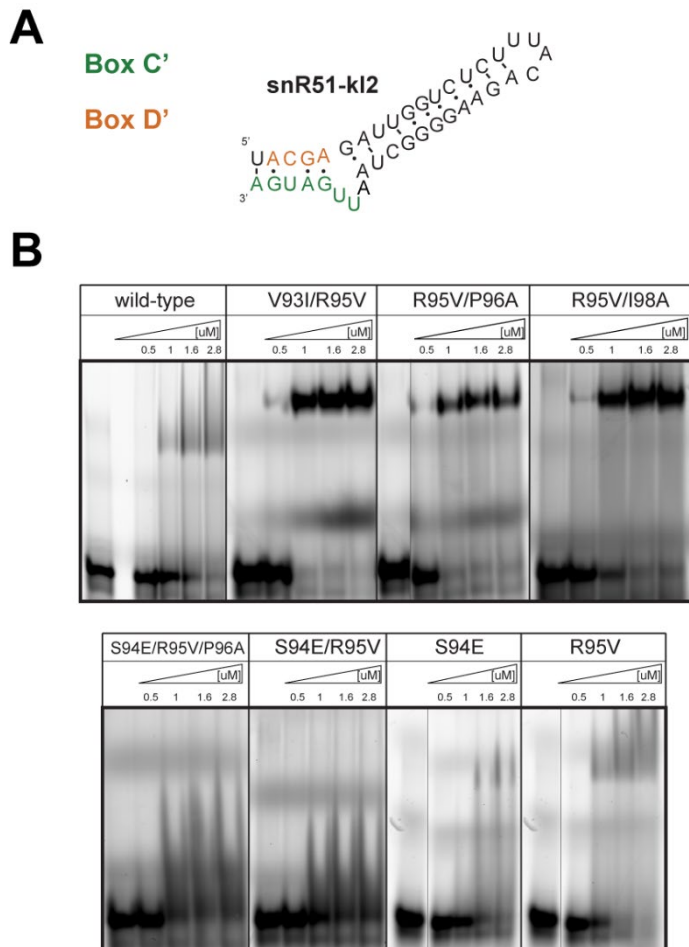
The amino acid composition of loop 9 differs significantly in the two proteins, which otherwise share 58.7% similarity and 34.1% identity (calculated for *P. furiosus* L7Ae and *S. cerevisiae* Snu13) and a high level of conservation in the RNA binding surface (Robert and Gouet, 2014; Madeira et al., 2019). Nevertheless, within each kingdom of life, the residues of loop 9 are conserved (**Figure 3.1.4 A-B**).



**Figure 3.1.4.** Comparison of residues in loop 9 of archaeal L7Ae and eukaryotic Snu13. **(A)** Multiple sequence alignment of archaeal L7Ae (blue) and eukaryotic Snu13 (pink). Residues of loop 9 are lighted in blue or pink, respectively. Pf – *P. furiosus*, Ss – *S. solfataricus*, Mj – *M. jannaschii*, Sc – *S. cerevisiae*, Ct – *C. thermophilum*, Hs – *H. sapiens*. The alignment was generated with the Clustal Omega Web service and displayed with ESPript 3.0. **(B)** Structural overlay of archaeal L7Ae (blue) bound to RNA (green) (structure presented in section 3.1.3) and human Snu13 (pink) bound to RNA (gray) (PDB-ID: 2OZB, Liu et al., 2007) highlighting residues in loop 9 of both proteins.

Based on the previous findings from (Gagnon et al., 2010), we tested whether we could manipulate the binding affinity of *S. cerevisiae* Snu13 toward our selected set of native and mutated RNAs by replacing individual residues in loop 9 with the corresponding amino acid from *P. furiosus* L7Ae.

To do so, we generated seven Snu13 mutants: S94E, R95V, V93I/R95V, S94E/R95V, R95V/P96A, R95V/I98A, and S95E/R95V/P96A. We tested the binding affinity of each mutant towards the complete set of RNAs (**Figure 3.1.1 A-B**, **Figure 3.1.5** and **Figure 3.1.6**) under the same experimental condition as in section 3.1.1.



**Figure 3.1.5.** Electrophoretic mobility shift assay of Snu13 wild type and mutants with snR51-kl2. **(A)** Predicted secondary structure of RNA snR51-kl2. **(B)** Electrophoretic mobility shift assays of Snu13 and Snu13 mutants with sR51-kl2 k-turn RNA. The Snu13 variant (wild type or mutant) is indicated above each assay. The first lane contains RNA alone as control. The concentration of protein in each lane is indicated above. Note that for mutants S94A and R95V and V93IR95V and P95VP96A the same control lane was used, as indicated with the separation line between control and titration lanes.

We found that none of the mutants changed the binding affinity of Snu13 as much as to induce detectable binding of mutant Snu13 to RNAs to which the wild type Snu13 did not bind (**Figure 3.1.6**).

Nevertheless, the mutations were able to modulate the affinity of Snu13 for mutant RNA snR51-kl2 (**Figure 3.1.5** and **Table 3.1.2**), to which wild type Snu13 binds with an approximate  $K_D$  value of  $\sim 1.1 \mu\text{M}$ .

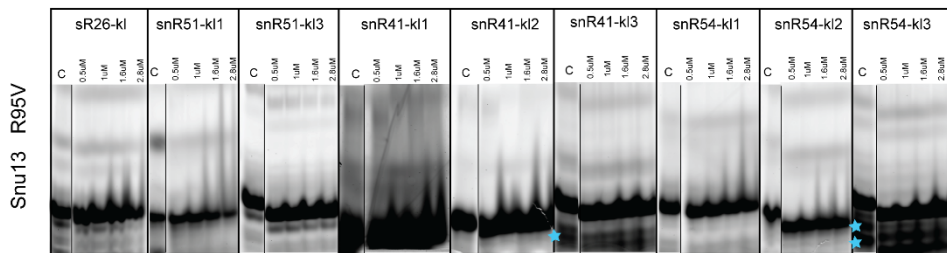
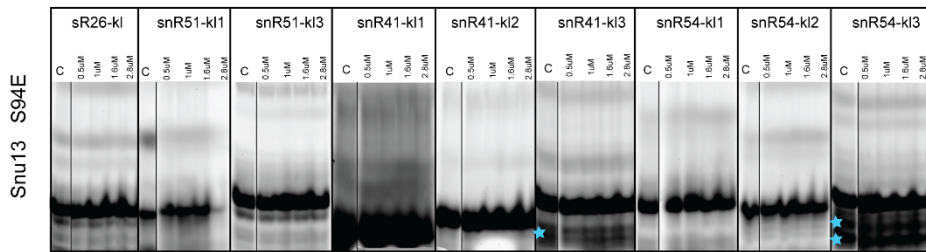
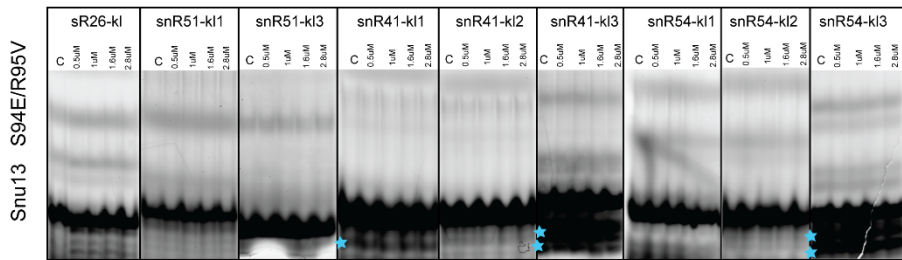
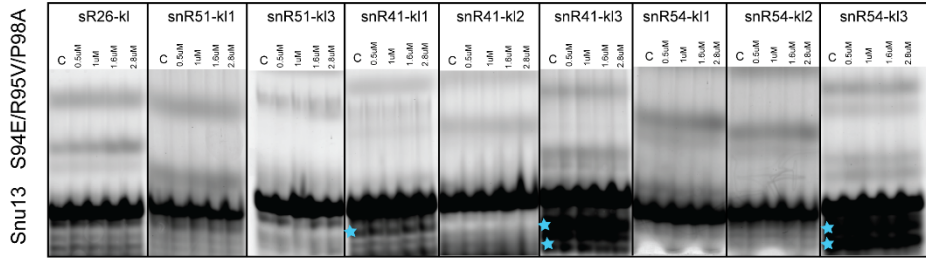
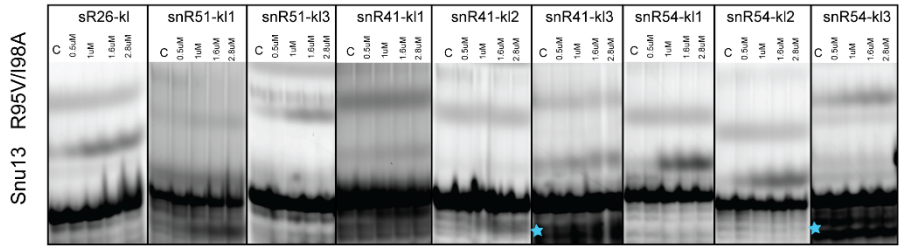
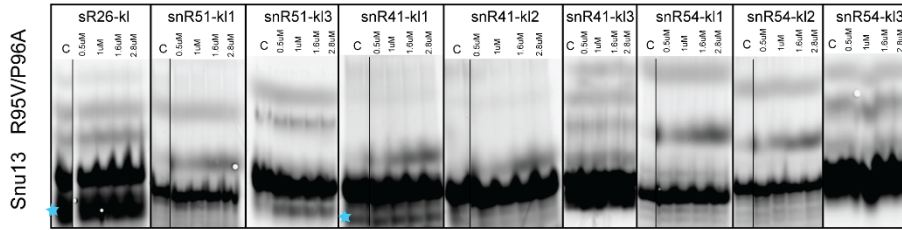
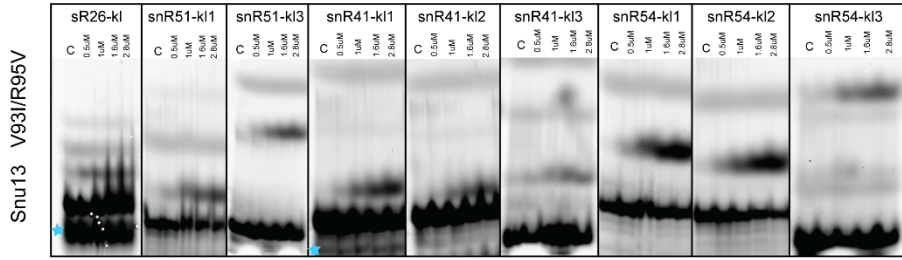
For Snu13 S94E, we observed a drop in binding affinity, indicating an adverse effect of the additional negative charge of the glutamic acid on the binding (**Figure 3.1.5 B** and **Table 3.1.2**). This effect is even more pronounced in Snu13 S94E/R95V and S94E/R97V/P96A, where the high-affinity binding is almost completely abolished.

Surprisingly, the removal of the positive charge of R95 in Snu13 R95V leads to an improved affinity (**Figure 3.1.5 B** and **Table 3.1.2**). This effect is magnified in Snu13 V93I/R95V and R95V/P96A and partly in Snu13 R95V/I97V (**Figure 3.1.5 B** and **Table 3.1.2**).

**Table 3.1.2.** Calculated  $K_D$  values and fitting parameters for Snu13 and Snu13 mutants. For detailed explanation see chapter 5.5. Errors are given as standard deviations and N/A indicates unsuccessful fitting.

|                           | <b>WT</b>          | <b>S94E</b>         | <b>R95V</b>        | <b>R95V/P96A</b> | <b>R95V/I97A</b> | <b>V93I/R95V</b>   |
|---------------------------|--------------------|---------------------|--------------------|------------------|------------------|--------------------|
| A1                        | 0.00 ± 0.00        | 0.01 ± 0.01         | 0.00 ± 0.05        | 0.00 ± 0.08      | 0.00 ± 0.06      | 0.00 ± 0.03        |
| A2                        | 0.39 ± 0.00        | 0.28 ± 0.02         | 1.11 ± 0.28        | 0.68 ± 0.06      | 0.77 ± 0.06      | 0.82 ± 0.02        |
| p                         | 5.08 ± 0.20        | 7.49 ± 2.38         | 1.09 ± 0.62        | N/A              | N/A              | 5.4 ± 0.86         |
| R <sup>2</sup>            | 0.99               | 0.98                | 0.98               | N/A              | N/A              | 0.99               |
| <b>K<sub>D</sub> [μM]</b> | <b>1.08 ± 0.01</b> | <b>1.44 ± 0.068</b> | <b>0.59 ± 0.28</b> | <b>N/A</b>       | <b>N/A</b>       | <b>0.64 ± 0.03</b> |

In conclusion, these results indicate that the residues in loop 9 of L7Ae or Snu13 are not the sole responsible factor for the ability of the protein to bind kink-loop or non-canonical kink-turn structures. However, loop 9 residues fine-tune the binding affinity of the protein to kink-turns. Increasing the proportion of flexible hydrophobic amino acids has a positive effect on binding, whereas introducing a negative charge in Snu13 loop 9 has a robust negative effect on the binding.

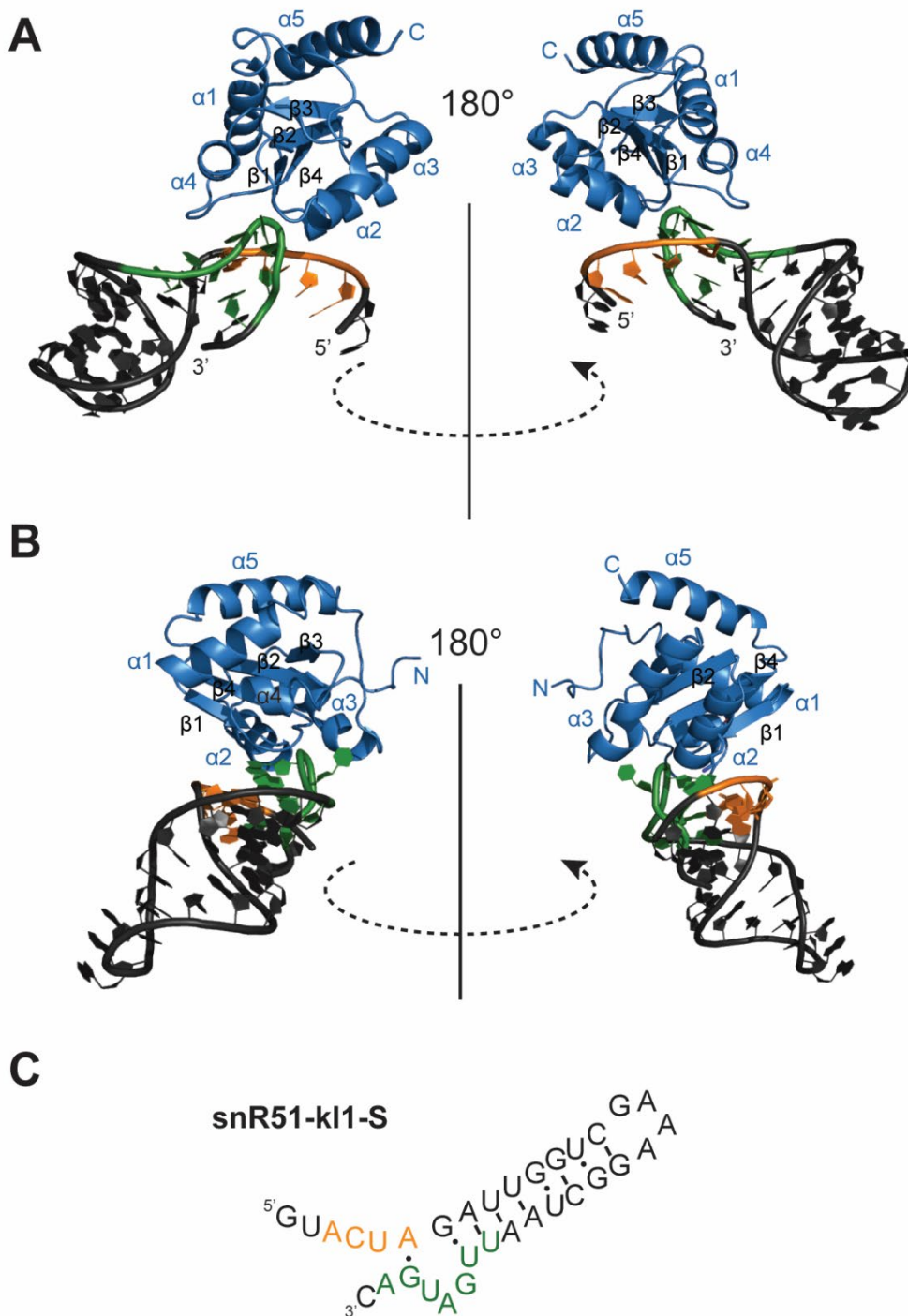


**Figure 3.1.6.** Electrophoretic mobility shift assays of Snu13 mutants with native Box C'/D' elements and mutants that did not show binding events. Electrophoretic mobility shift assays were performed with Snu13 mutants V93I/R95V, R95V/P98A, R95V/I98A, S94E-R95V/P96A, S94E/R95V, S94E and R95V and 5'-Cy5 labeled RNA constructs sR26-kl, snR51-kl1, snR51-kl3, snR41-kl1, snR41-kl2, snR41-kl3, snR54-kl1, snR54-kl2 and snR54-kl3. The first lane of each assay contains only RNA at 2 μM and serves as a control. All other lanes contain 2 μM RNA and 0.5 μM, 1 μM, 1.6 μM or 2.8 μM of the respective protein, as indicated on the left side of each panel. In case the control lanes were not run next to the titration lanes, the control lane of the same gel is pasted next to the titration lanes and separated by a thin line. The blue asterisks mark lanes corresponding to RNA degradation products.

### 3.1.3 Crystal structure of archaeal L7Ae bound to eukaryotic non-standard internal kink-turn structure

After determining essential elements for protein binding in both kink-turn or kink-loop forming RNA elements and the protein sequence, we wanted to understand the mechanism by which L7Ae and Snu13 bind non-canonical kink-turns and kink-loops.

To do so, we tried to crystallize L7Ae in complex with internal kink-turn snR51-kl1 and kink-loop sR26-kl as well as Snu13 in complex with the modified internal kink-turn snR51-kl2. We could not obtain any crystals of L7Ae in complex with sR26-kl (see section 5.7.1), while from the solution of the Snu13–snR51-kl2 complex, Snu13 crystallized alone, most likely due to the low affinity of the complex (see section 5.7.1). Using a shortened version of snR51-kl1, snR51-kl1-S (**Figure 3.1.7 C**), we could crystallize the L7Ae–snR51-kl1-S complex and solve its structure at 1.9 Å resolution (**Figure 3.1.7 A-B**).

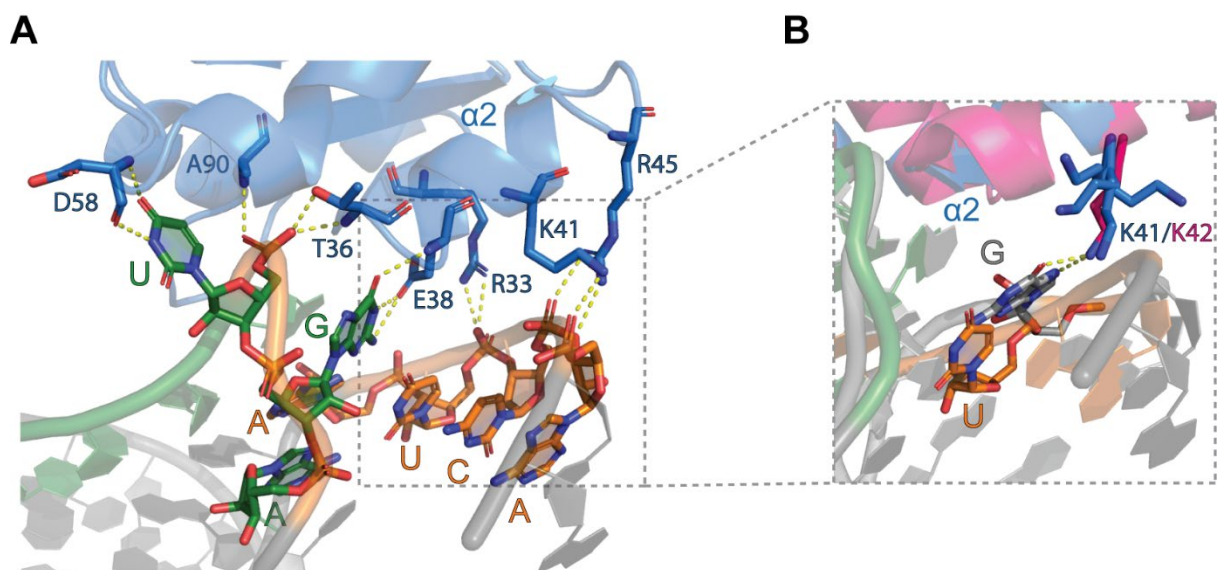


**Figure 3.1.7.** 1.9 Å crystallographic structure of *P. furiosus* L7Ae bound to the snR51-kl1-S non-canonical kink-turn. Both **(A)** and **(B)** show an overview of the structure (L7Ae in blue; snR51-kl1-S in black with box C' in green and box D' in orange). **(C)** shows the sequences and secondary structure of the snR51-kl1-S non-canonical kink-turn RNA. Box C' and D' are highlighted in green and orange, respectively.



*P. furiosus* L7Ae exhibits the same fold as already determined in previous studies (Charron et al., 2004; Hamma and Ferré-D'Amaré, 2004; Huang and Lilley, 2013; Suryadi et al., 2005), which consists of two pairs of parallel central  $\beta$ -sheets surrounded by five  $\alpha$ -helices. In contrast to all other published structures, we find that the 1n-1b G•A base-pair is instead a 1n-4b G•A base-pair in snR51-kl1-S. This results in a shift of all the remaining bases to a different position within the kink (Huang and Lilley, 2013, 2014; Ahmed et al., 2020).

The interactions between the protein and the RNA on the G•A base-pair side of the kink are dominated by polar interactions and hydrogen bonds between arginine (R) or lysine (K) residues with the phosphate backbone of the RNA (**Figure 3.1.8 A**). The guanine of the 1n-4b G•A base-pair forms a crucial hydrogen bonds engaging the N2 amino group and N1 imino group with the carboxy group of L7Ae E38. This interaction has been consistently observed in all so far characterized L7Ae–RNA and Snu13–RNA complexes and thus appears to be a requirement for the formation of a stable complex. This fact indicated that the formation of a G•A base-pair directly next to the RNA kink is essential (Hamma and Ferré-D'Amaré, 2004; Moore et al., 2004; Xue et al., 2010; Huang and Lilley, 2013; Huang et al., 2019) and explained why Box C'/D' element of the eukaryotic guide RNA snR41 (**Figure 3.1.2 A-B**), where the guanine is replaced by a cytosine, does not bind either L7Ae or Snu13.



**Figure 3.1.8.** Detailed view of archaeal L7Ae interactions with non-canonical internal kink-turn snR51-kl1-S. **(A)** Polar interactions (yellow dashed lines) between archaeal L7Ae (blue) and snR51-kl1-S on the G•A base-pair side of the kink. Interacting amino acids and nucleotides are shown as sticks. **(B)** Detailed view of human Snu13 (pink) interacting with U4 snRNA (light grey) (PDB-ID: 1E7K, Vidovic et al., 2000) overlaid with archaeal L7Ae (blue) interacting with snR51-kl1-S (Box C': green, Box D': orange, remaining RNA gray). Snu13-K42 forms hydrogen bonds with the 2n guanine, whereas L7Ae-K41 is flexible in all four chains present in the asymmetric unit and does not form defined hydrogen bonds. Oxygen atoms are indicated in red; nitrogen atoms are indicated in blue.

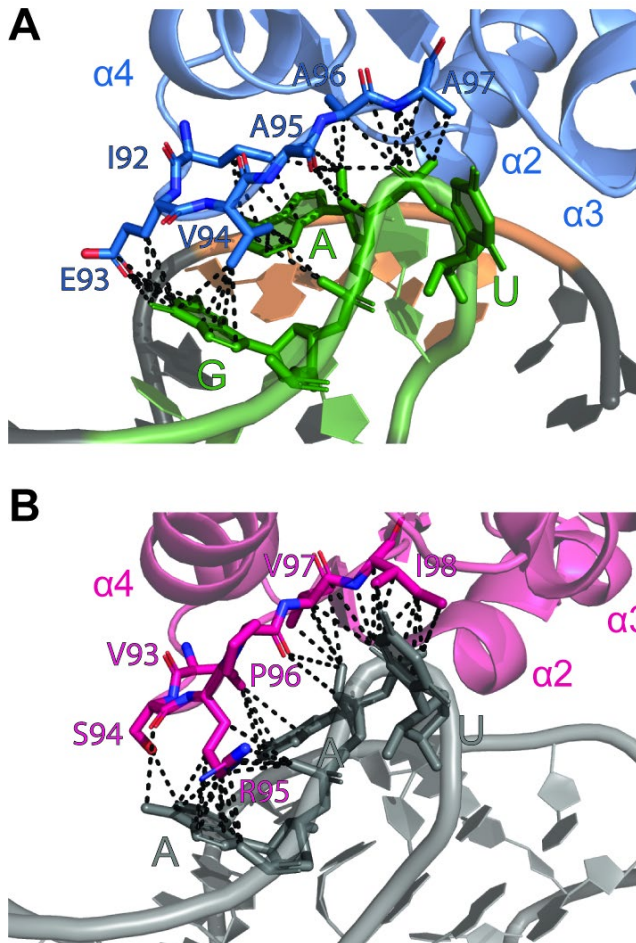
Our structure in combination with available structures of Snu13 bound to kink-turn motifs (PDB-ID: 2OZB, 1E7K, 3SIU, 3SIV, 5WLC, 6QW6, 6AHD, 5GAP, 6QX9, 5ZWO) (Vidovic et al., 2000; Liu et al., 2007a, 2011; Barandun et al., 2017; Nguyen et al., 2016; Charenton et al., 2019a; Bai et al., 2018; Zhan et al., 2018) suggests a reason why Snu13 requires the formation of the 2n-2b G•A base-pair for RNA recognition, while L7Ae does not. Snu13-K42 makes an hydrogen bond with the O6 of the guanine from the 2n-2b G•A base-pair (Vidovic et al., 2000; Liu et al., 2011, 2007a). This interaction is only randomly conserved in the available L7Ae–RNA structures and seems to be dispensable for high affinity (Moore et al., 2004; Li and Ye, 2006a; Huang and Lilley, 2013, 2014; Ahmed et al., 2020; Huang and Lilley, 2016; Xue et al., 2010; Ye et al., 2009; Oshima et al., 2016, 2018; Huang et al., 2019). In our structure, where the 2n-2b G•A base-pair cannot form, the corresponding amino acid K41 is not involved in any specific hydrogen bonds but contributes diffuse electrostatic interactions with the RNA phosphate backbone. Thus, we conclude that the involvement of the lysine (K) in a hydrogen bond is necessary for high affinity binding of Snu13 but is dispensable for L7Ae, probably due to compensating energy contributions. This fact would rationalize why L7Ae can bind to both the native snR51-kl1 internal kink-turn structure as well as the mutant with two G•A base pairs, whereas Snu13 can only bind the mutant snR51-kl2 (see section 3.1.1).

Next, we analyzed the interactions between archaeal L7Ae loop 9 residues and snR51-kl1-S, to better understand the results of section 3.1.2. I92 and V94 form extensive hydrophobic contacts with the adenine and the guanine in position L2 and L3 (**Figure 3.1.9 A**). E93 has a polar contact with the guanine in position L3; this contact is present in only ~50% of all published structures of L7Ae bound to RNA (Hamma and Ferré-D'Amaré, 2004; Moore et al., 2004; Xue et al., 2010; Huang and Lilley, 2013, 2014; Huang et al., 2019) and thus seems to be dispensable for binding. It is noteworthy that V94 connects the bases in position L2 and L3 by forming hydrophobic interactions with both bases. I92, on the other hand, forms hydrophobic contacts to the base at position L2 and the corresponding sugar. By means of these contacts I92 and V94 create a network of connections between nucleotides in position L2 and L3 and stabilize the RNA on the stem I side of the kink.

In Snu13 I92 is replaced by V93, which lacks one carbon in its side chain and therefore can only form hydrophobic contacts with the base in position L2 (**Figure 3.1.9 B**). In place of L7Ae V94, R95 forms hydrogen bonds with the phosphate backbone and some hydrophobic contacts with the base in position L3 but does not connect the L2 and L3 bases.

From these structural observations and the binding assay data from section 3.1.2, we can conclude that replacing R95 in eukaryotic Snu13 with valine (V), like in L7Ae, introduces the bridging interactions between the base in position L2 and L3, which have a beneficial effect on the binding affinity (see section 3.1.2). Similarly, it is beneficial to replace V93 with isoleucine

(I). Both substitutions improve Snu13 binding affinity towards snR51-kI2 compared to the wild type. On the other hand, replacing S95 with a glutamic acid (E) shows a robust negative effect on the binding affinity, especially if combined with the R95V mutation, demonstrating that the presence of a negative charge in the context of Snu13 is not tolerated when binding RNA.



**Figure 3.1.9.** Detailed view of interactions between residues in loop 9 of archaeal L7Ae and eukaryotic Snu13 with kink-turn structured RNA. **(A)** Interactions between residues in loop 9 of *P. furiosus* L7Ae and bound snR51-kI1-S RNA. Interactions (polar and apolar) are displayed as black dashed lines. Important hydrophobic interactions are established between I92 and the L2 adenosine base and sugar and between V94 and L3 guanine and L2 adenine. **(B)** Detailed view of interactions between residues in loop 9 of human Snu13 and the kink-turn structure of snoRNA U14 (PDB-ID:2OZB, Liu et al., 2007). Interactions (polar and apolar) are displayed as black dashed lines. R95 and S94 have polar bonds with the phosphate backbone of the RNA and L3 adenine, respectively.

Finally, we conclude that the interaction between residues in loop 9 of L7Ae and Snu13 and the RNA is dominated by non-sequence specific hydrophobic interactions and that the presence of polar or charged residues in this area is essential for fine-tuning the affinity of the protein towards different kink-turn and kink-loop structures.

### 3.1.4 Binding stoichiometry in fully assembled chimeric Box C/D s/snoRNP

In the fully assembled and catalytically active archaeal Box C/D enzyme, the complex between the kink-turn and kink-loop elements and L7Ae forms a composite protein–RNA surface that interacts with the C-terminal domain of the scaffolding protein Nop5. This interaction has a stabilizing effect on the binding of L7Ae to the respective kink-turn or kink-loop structures in

the guide RNA (Ghalei et al., 2010; Lapinaite et al., 2013). A similar binding mode of the C-terminal domains of Nop56 and Nop58 has been found in the eukaryotic Box C/D-like complex assembled on the non-methylating snoRNA U3 during ribosome biogenesis (Barandun et al., 2017; Cheng et al., 2017).

Up to this point, we have analyzed the interaction of archaeal L7Ae and eukaryotic Snu13 to a diverse set of kink-turns and kink-loops in the absence of the other proteins of the Box C/D enzyme. As a next step, we studied the binding of eukaryotic Snu13 to the internal kink-turns and kink-loops in the context of the fully assembled enzyme.

The C-terminal domain of Nop5 and Nop56 and Nop58 is highly conserved between archaea and eukaryotes, with the domains of Nop5 and Nop56 sharing 44.4% identity and 64.1% similarity and the domains of Nop5 and Nop58 sharing 41.9% identity and 60.5% similarity in the primary sequence. Similarity and identities were calculated, excluding the C-terminal extended tails of Nop5, Nop56, and Nop58 \*.

As we encountered difficulties in obtaining full-length Nop56 and Nop58 from recombinant and endogenous sources in a soluble form, we decided to use archaeal Nop5, and Fibrillarin, from *P. furiosus* for the following experiments, based on the high sequence similarity of the C-terminal domains.

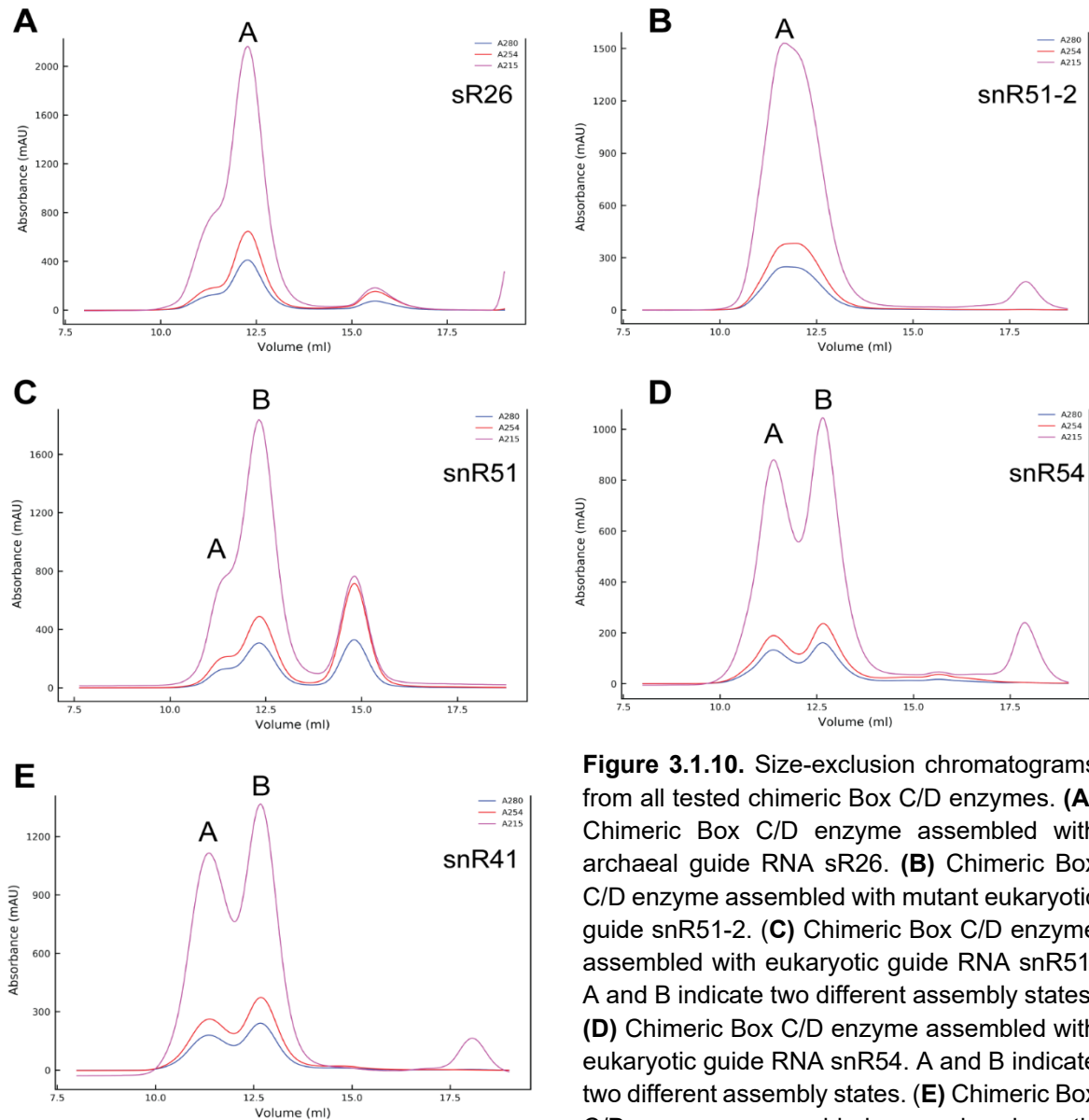
To determine whether wild type Snu13 can bind to the snR26-kl, snR51-kl1, snR51-kl2, snR54-kl1, and snR41-kl1 in the context of a fully assembled Box C/D enzyme, we assembled chimeric complexes *in vitro* using the corresponding full-length guide RNAs, Snu13 or L7Ae and the archaeal Nop5 and Fibrillarin proteins, as described in section 5.4. The resulting size-exclusion profiles for each chimeric complex are displayed in **Figure 3.1.10 A-E** and **Figure 3.1.11 A**. For fully assembled complexes showing two elution peaks, indicating different assembly states, both states (A and B) were analyzed separately.

To determine the ability of wild type Snu13 to bind to non-canonical kink-turns and kink-loops in the presence of the scaffolding protein Nop5, we determined the copy number of each protein in each complex using a quantitative mass spectrometry approach (see section 5.8).

From previous studies of the archaeal Box C/D enzymes (Ye et al., 2009; Xue et al., 2010; Lapinaite et al., 2013), the stoichiometry of Nop5 and Fibrillarin has been established to be two copies each per guide RNA molecule, corresponding to a 1:1 ratio. Depending on the oligomeric state of the complex (mono- or di-RNP), there are one or two guide RNA copies per particle. A Nop5 (or Fibrillarin) to Snu13 ratio of 1:1 indicates two copies of Snu13 per RNA

---

\* All similarity and identity values are derived from pair-wise primary sequence alignments calculated using the Needleman-Wunsch algorithm implemented in the EMBL alignment tool NEEDLE (Madeira et al., 2019).

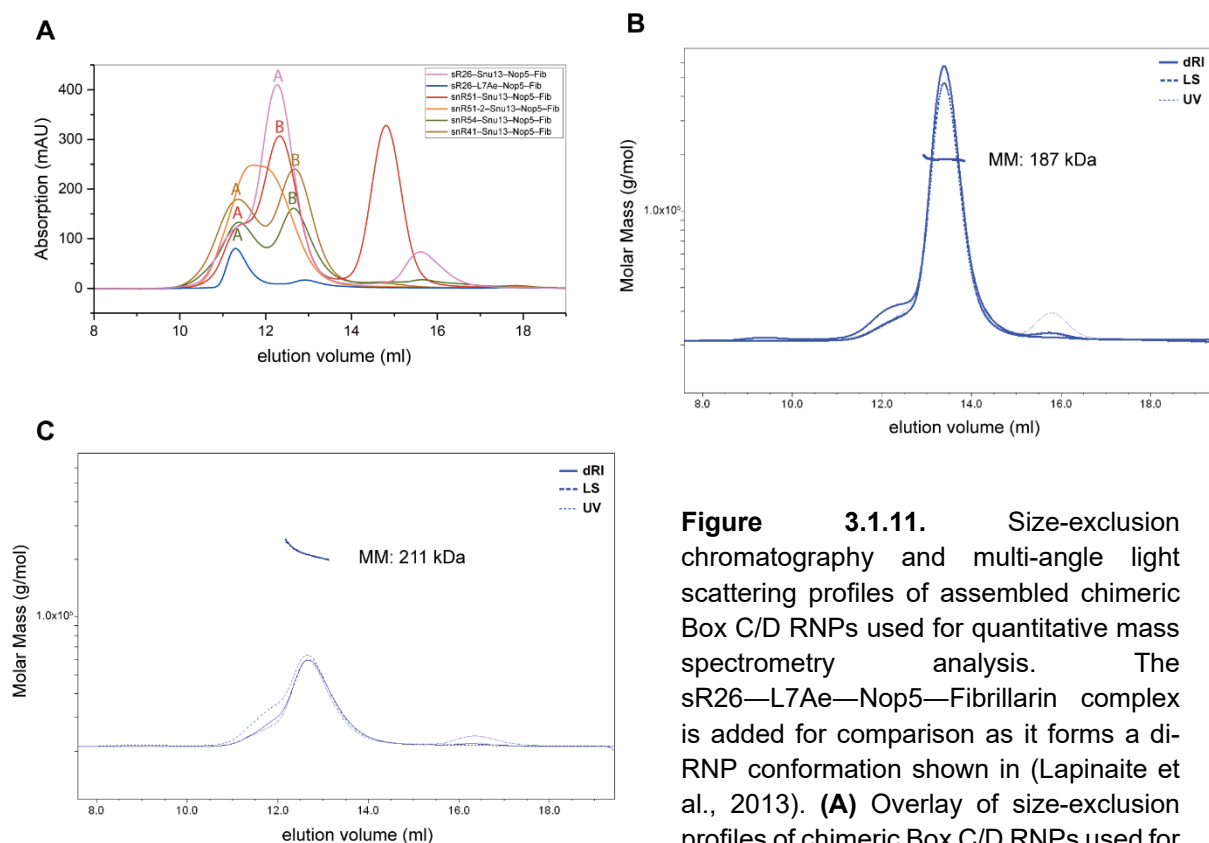


**Figure 3.1.10.** Size-exclusion chromatograms from all tested chimeric Box C/D enzymes. **(A)** Chimeric Box C/D enzyme assembled with archaeal guide RNA sR26. **(B)** Chimeric Box C/D enzyme assembled with mutant eukaryotic guide snR51-2. **(C)** Chimeric Box C/D enzyme assembled with eukaryotic guide RNA snR51. A and B indicate two different assembly states. **(D)** Chimeric Box C/D enzyme assembled with eukaryotic guide RNA snR54. A and B indicate two different assembly states. **(E)** Chimeric Box C/D enzyme assembled around eukaryotic guide RNA snR41. A and B indicate two different assembly states. The UV traces at 280 nm, 254 nm or 215 nm are shown in blue, red and pink, respectively.

copy, while a ratio of 2:1 indicates one copy of Snu13 per RNA copy. Thus, in the first case both Box C/D kink-turn and Box C'/D' kink-turn or kink-loop are bound by one copy of Snu13, while in the second case only the Box C/D kink-turn is bound by the protein. To confirm that our observation also correlates with the mass of the analyzed complex we determined the molecular mass of the snR51—Snu13—Nop5—Fibrillarin (**Figure 3.1.11 B**) and the sR26—Snu13—Nop5—Fibrillarin (**Figure 3.1.11 C**) complex by Multi-angle light scattering (MALS). We determined a molecular mass of ~187 kDa for the

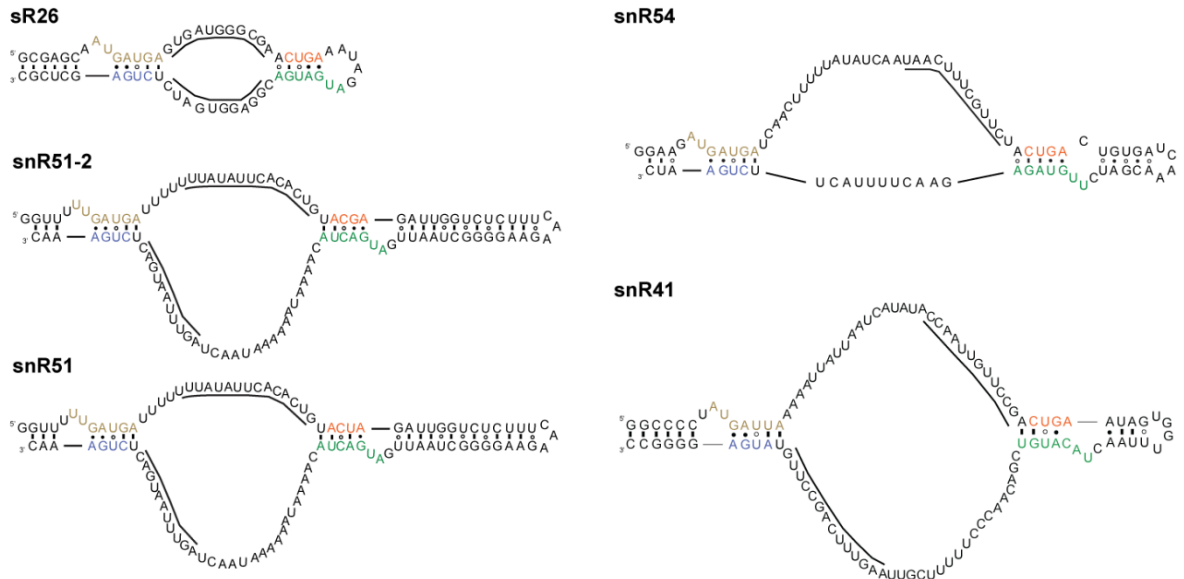
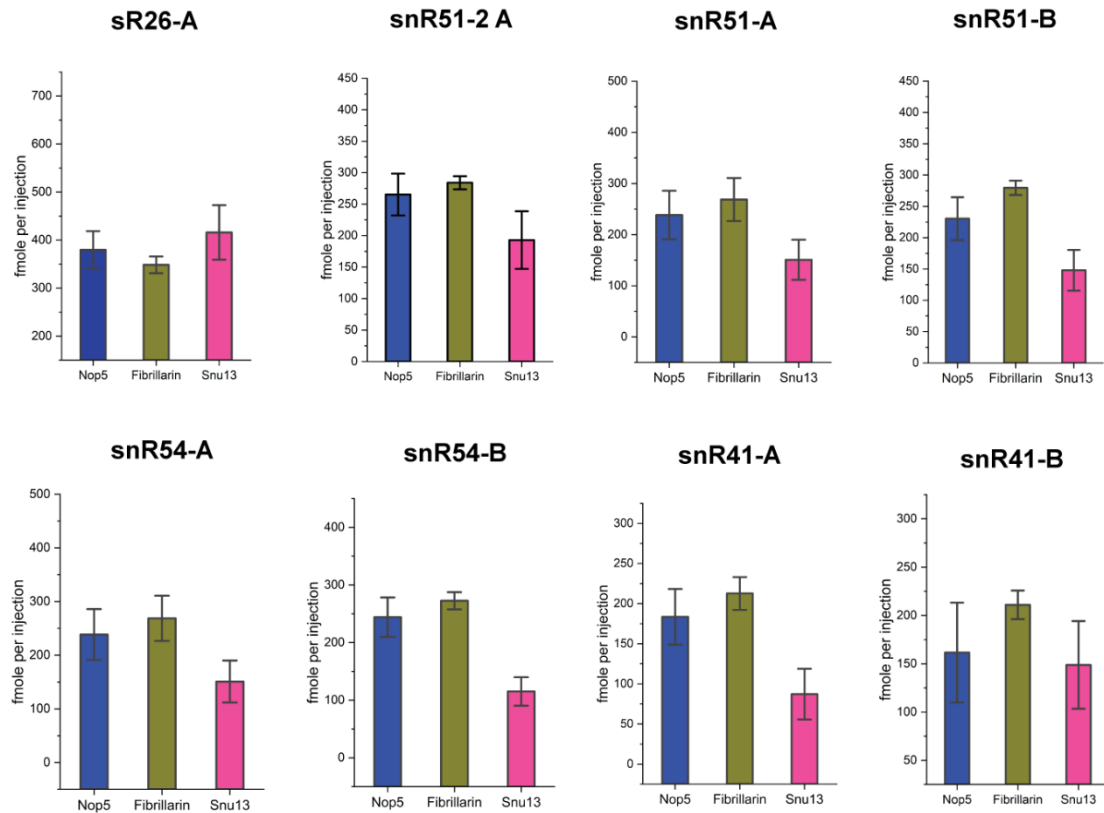
snR51—Snu13—Nop5—Fibrillarin complex and ~211 kDa for the sR26—Snu13—Nop5—Fibrillarin complex, which both represents a mono-RNP conformation.

Using quantitative mass spectrometry, we found that Snu13 stably binds the sR26-kl kink-loop in the context of the chimeric complex assembled with sR26 (**Figure 3.1.12**). This result contrasts with what we observed in gel shift assays for Snu13 and the sR26-kl k-loop RNA in isolation, where we could not detect any binding (section 3.1.1). Therefore, we conclude that the additional interaction with the C-terminal domain of Nop5 promotes the binding of Snu13 to snR26-kl.



**Figure 3.1.11.** Size-exclusion chromatography and multi-angle light scattering profiles of assembled chimeric Box C/D RNPs used for quantitative mass spectrometry analysis. The sR26—L7Ae—Nop5—Fibrillarin complex is added for comparison as it forms a di-RNP conformation shown in (Lapinaite et al., 2013). **(A)** Overlay of size-exclusion profiles of chimeric Box C/D RNPs used for quantitative analysis and the

sR26—L7Ae—Nop5—Fibrillarin complex for comparison. The profile for sR26—Snu13—Nop5—Fibrillarin is shown in pink, for sR26—L7Ae—Nop5—Fibrillarin in blue, for snR51—Snu13—Nop5—Fibrillarin in red, for snR51-2—Snu13—Nop5—Fibrillarin in orange, for snR54—Snu13—Nop5—Fibrillarin in green and for snR41—Snu13—Nop5—Fibrillarin in brown. **(B)** Multi-angle light scattering profile for snR51—Snu13—Nop5—Fibrillarin. Traces from the differential refractometer (dRI), the light scattering detector (LS) and the UV detector (UV) are indicated. The determined molecular mass (MM) is ~ 187 kDa. **(C)** Multi-angle light scattering profile for sR26—Snu13—Nop5—Fibrillarin. Traces from the differential refractometer (dRI), the light scattering detector (LS) and the UV detector (UV) are indicated. The determined molecular mass (MM) is ~ 211 kDa.

**A****B**

**Figure 3.1.12.** Protein stoichiometries in individual chimeric Box C/D enzymes as determined by quantitative mass spectrometry. **(A)** Schematic representation of the predicted secondary structure of used guide RNAs. Box C and D are indicated in brown and blue, respectively. Box C' and D' are indicated in green and orange, respectively. **(B)** Detected amount of peptides for each protein (Nop5, Fibrillarin, Snu13) in every analyzed RNP in fmol per injection. The label on top of each bar graph indicates the guide RNA used for the assembly of the complex. A and B after the guide RNA indicate which assembly state was analyzed. The height of each bar shows the mean fmol per injection detected for both peptides per protein over all measurements and the error bars represent the resulting standard error.

For the chimeric complex assembled with snR51-2, which contains the snR51-kl2 sequence instead of the wild-type snR51-kl1, we found a ratio between Nop5/Fibrillarin and Snu13 of 2:1,5 (**Figure 3.1.12**). This ratio indicates an approximately equimolar mixture of complexes with one and two copies of Snu13 per guide RNA copy. This result confirmed that Snu13 binds snR51-kl2 with moderate affinity and is therefore only bound to the motif in 50% of all formed complex particles. This data correlate well with the result of the gel shifts from section 3.1.1, where binding was observed in the low  $\mu\text{M}$  range.

For snR51 in both oligomeric states A and B, we find only one copy of Snu13 per guide RNA molecule. Thus, Snu13 is unable to bind snR51-kl1 also in the presence of Nop5 (**Figure 3.1.12**). We made the same observation for snR54-kl1 in snR54 A and B complexes as well as for snR41-kl1 in the snR41 A complex (**Figure 3.1.12**). For all three non-canonical kink-turns, these results correlate with the binding behavior observed in isolation (section 3.1.1). For the chimeric Box C/D RNP snR41-B we observed the same copy number for Snu13 as for Nop5 and Fibrillarin. The peak, however, has an elution volume that corresponds to something smaller than a mono-RNP, therefore it cannot be excluded that this peak contains only one copy of Nop5 and Fibrillarin (**Figure 3.1.11 A**).



## 3.2 Molecular basis of Nop1 and Nop56 interaction

### 3.2.1 Eukaryotic methyltransferase Nop1 cannot be complemented by the archaeal ortholog Fibrillarin

In the chimeric Box C/D complexes that we investigated in section 3.1.4 the catalytic unit was the archaeal SAM-dependent methyltransferase Fibrillarin. The catalytic unit of the Box C/D enzyme in the eukaryotic model system *S. cerevisiae* is the SAM-dependent methyltransferase Nop1 (Tollervey et al., 1991). In comparison to archaeal Fibrillarins, their eukaryotic orthologs have an additional ~ 80 amino acids long, potentially disordered arginine- and glycine-rich domain termed the RGG box/motif or GAR domain (Rodriguez-Corona et al., 2015). The RGG motif is found in more than 1000 human proteins, which are involved in transcription, DNA damage signaling, and apoptosis among other processes (Thandapani et al., 2013). It is known that the RGG motif is methylated at specific arginine residues by methyltransferase Hmt1p (Yagoub et al., 2015; Smith et al., 2020). There is some evidence that this motif plays a role in the nuclear and nucleolar localization of Fibrillarin and it has been recently proposed that a potential ribonuclease activity is associated with it; however, the precise function of this motif remains unclear (Guillen-Chable et al., 2020; Shubina et al., 2020). In the following study, we omitted the RGG-rich domain in the Nop1 construct, to facilitate expression and stability of the protein.

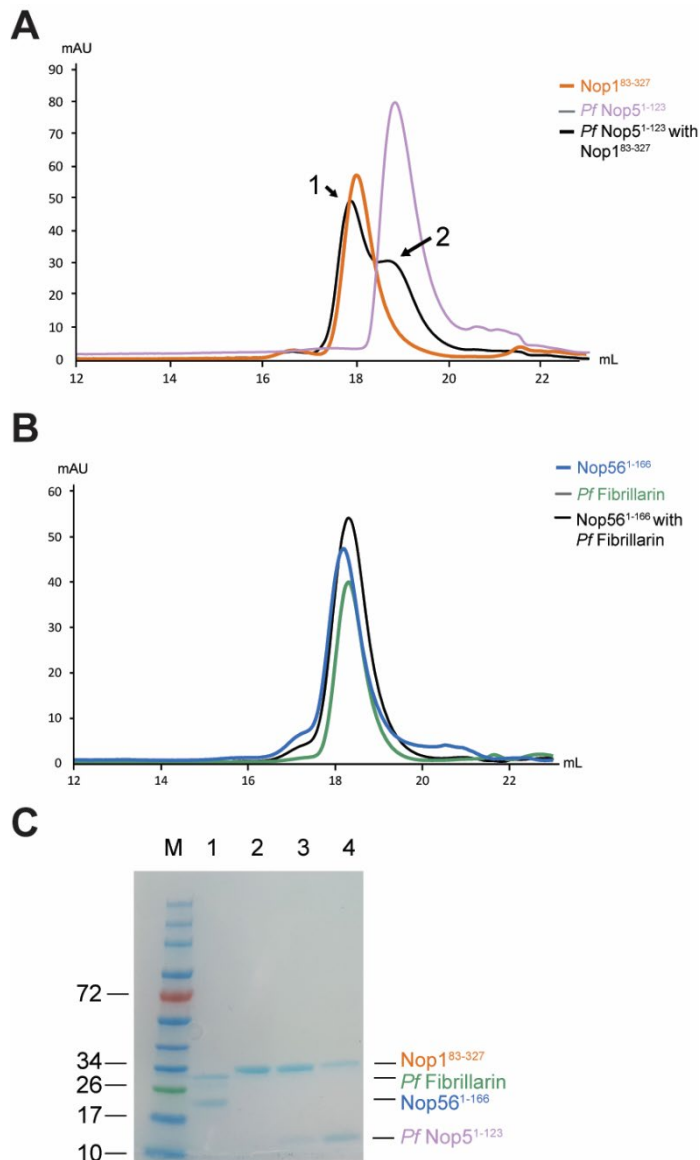
Similar to the archaeal enzyme, Nop1 is integrated into the RNP by interacting with the N-terminal domains of Nop56 and Nop58 (Cahill et al., 2002; Barandun et al., 2017). Nop1 shares a 60.9% (44.8%) primary sequence similarity (identity) with *P. furiosus* Fibrillarin whereas the N-terminal domains of Nop56 and Nop58 only share 31.3%(17.6%) and 28.6%(11.9%) similarity (identity) with the N-terminal domain of archaeal *P. furiosus* Nop5 \*. Prompted by the relatively high sequence similarity between Nop1 and Fibrillarin, we investigated whether the eukaryotic methyltransferase Nop1 can form a stable complex with the N-terminal domain of the archaeal scaffolding protein Nop5 and thus be integrated into the archaeal or chimeric Box C/D enzyme (see section 3.3).

To answer this question we used recombinantly expressed Nop1<sup>83-327</sup>, Nop56<sup>1-166</sup> (N-terminal domain of Nop56) (see section 5.2.2), archaeal Fibrillarin and Nop5<sup>1-123</sup> (Nop5 N-terminal domain) from *P. furiosus* (see section 5.2.1) and tested their ability to form stable cross-species protein-protein complexes by size-exclusion chromatography, as described in section 5.6.

---

\* All similarities and identities were derived from pair-wise primary sequence alignments calculated using the Needleman-Wunsch algorithm implemented in the EMBL alignment tool NEEDLE (Madeira et al., 2019).

It has been shown that Nop1 and Nop56 interact with each other *in vivo* (Gautier et al., 1997; Cahill et al., 2002; Lechertier et al., 2009). To test whether Nop1<sup>83-327</sup> forms a stable complex with its native binding partner Nop56<sup>1-166</sup> *in vitro*, we mixed the two proteins and subjected the mixture to size-exclusion chromatography. **Figure 5.3** in section 5.7 shows that the two eukaryotic proteins form a stable complex also *in vitro*. The stable interaction between archaeal Fibrillarlin and Nop5 *in vitro* had already been confirmed by (Aittaleb et al., 2003; Oruganti et al., 2007; Ye et al., 2009; Lapinaite et al., 2013) and by us in section 3.4.



**Figure 3.2.1** Size-exclusion chromatograms testing Nop1–Nop5 and Fibrillarlin–Nop56 interaction. **A)** Overlay of size-exclusion chromatograms. The orange trace shows the 280 nm absorption for the isolated Nop1<sup>83-327</sup> protein. The violet trace shows the 280 nm absorption of isolated *P. furiosus* Nop5<sup>1-123</sup>. The black trace shows the 280 nm absorption of Nop1<sup>83-327</sup> mixed with *P. furiosus* Nop5<sup>1-123</sup>. **(B)** Overlay of size-exclusion chromatograms. The blue trace shows the 280 nm absorption of isolated Nop56<sup>1-166</sup> protein. The green trace shows the 280 nm absorption of *P. furiosus* Fibrillarlin. The black trace shows the 280 nm absorption of Nop56<sup>1-166</sup> mixed with *P. furiosus* Fibrillarlin. **(C)** SDS gel confirming the content of elution peaks shown in A and B. Lane 1 shows the content of the black trace from B, containing Nop56<sup>1-166</sup> and archaeal Fibrillarlin. Lanes 2 to 4 show the content of the black trace from A, containing Nop1<sup>83-327</sup> and Nop5<sup>1-123</sup>.

Next, we tested the interaction between Nop1<sup>83-327</sup> and Nop5<sup>1-123</sup>. We found that, despite the high sequence similarity between the archaeal and eukaryotic methyltransferases, the two proteins did not form a stable complex (**Figure 3.2.1 A**). Similarly, archaeal Fibrillarlin and Nop56<sup>1-166</sup> were also unable to form a stable complex (**Figure 3.2.1 B**). The content of all elution peaks was confirmed by SDS gel-electrophoresis (**Figure 3.2.1 C**).

With this, we prove that eukaryotic and archaeal Fibrillarin cannot complement each other as they do not bind to the N-terminal domain of the scaffolding protein of the other species. As Nop1 and Nop56 share 73% and 58.5% primary sequence similarity with human Fibrillarin and Nop56 \*, it can be assumed that these findings also hold for the human Box C/D RNP enzyme.

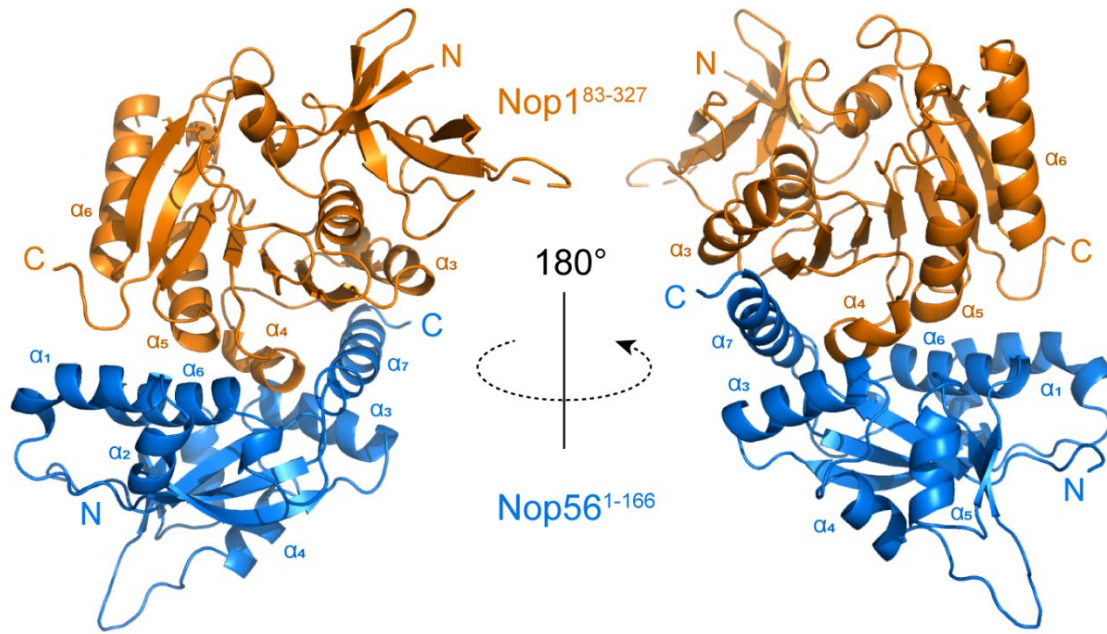
### 3.2.2 Structural differences between eukaryotic Nop1—Nop56 and archaeal Fibrillarin—Nop5 complex

#### *Crystal Structure of Nop1<sup>83-327</sup> bound to Nop56<sup>1-166</sup>*

To understand which interactions, govern the complex formation between Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>, we solved the structure of the complex by X-ray crystallography to a resolution of 1.7 Å (**Figure 3.2.2**). The structure is deposited in the Protein Data Bank (PDB) under the accession number 6ZDT (available upon publication of the corresponding manuscript). Nop1<sup>83-327</sup> forms a compact structure consisting of a small N-terminal lobe and a bigger C-terminal lobe. The N-terminal lobe is comprised of residues 83 to 146 and directly follows the N-terminal RGG motif in the full-length protein. The small lobe consists of five anti-parallel β-sheets (β). The C-terminal lobe consists of seven β-sheets, which are surrounded by six α-helices (α); together, they adopt a fold characteristic for SAM-dependent methyltransferases, termed Rossmann-fold (Chouhan et al., 2019; Wang et al., 2000). Nop1 from *S. cerevisiae* adopts a similar fold to the human ortholog (**Figure 3.2.3**) as well as to characterized archaeal Fibrillarins (**Figure 3.2.4**).

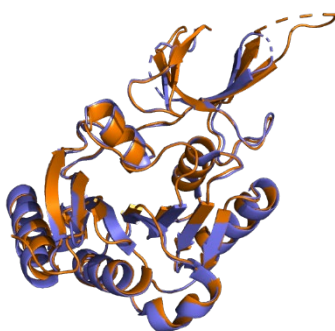
---

\* All similarities and identities were derived from pair-wise primary sequence alignments calculated using the Needleman-Wunsch algorithm implemented in the EMBL alignment tool NEEDLE (Madeira et al., 2019).

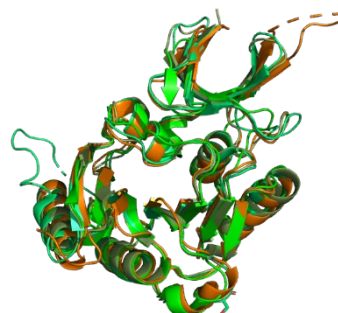


**Figure 3.2.2** Crystallographic structure of Nop1<sup>83-327</sup> in complex with Nop56<sup>1-166</sup> at 1.7 Å resolution. Nop1<sup>83-327</sup> (orange), lacking the N-terminal RGG motif, interacts with the N-terminal domain of Nop56 (blue) via  $\alpha$ -helix 3-5 and the C-terminal tail. Nop56<sup>1-166</sup> engages in interactions with Nop1<sup>83-327</sup> via  $\alpha$ -helix 1, 6 and 7 as well as  $\beta$ -sheet 1 and the loop region between  $\beta$ -sheet 4 and 5. N-termini and C-termini of both protein are labeled with N and C, respectively.

The N-terminal domain of Nop56 is comprised of residue 1 to 166 and is followed by the coiled-coil domain and the C-terminal domain in the full-length protein. The N-terminal domain itself consists of five  $\beta$ -sheets.  $\beta$ -sheet 1-3 are in anti-parallel orientation, while  $\beta$ -sheet 4-5 are in parallel orientation. These five  $\beta$ -sheets are surrounded by seven  $\alpha$ -helices, which also form the majority of the binding interface with Nop1.



**Figure 3.2.3.** Comparison of yeast Nop1 (orange) with human Fibrillarin (blue) (PDB-ID: 2IPX).

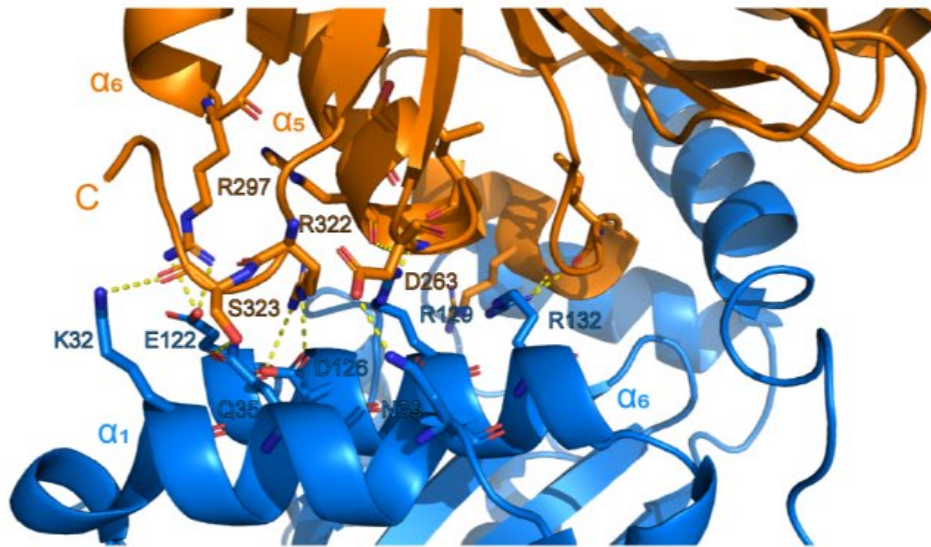
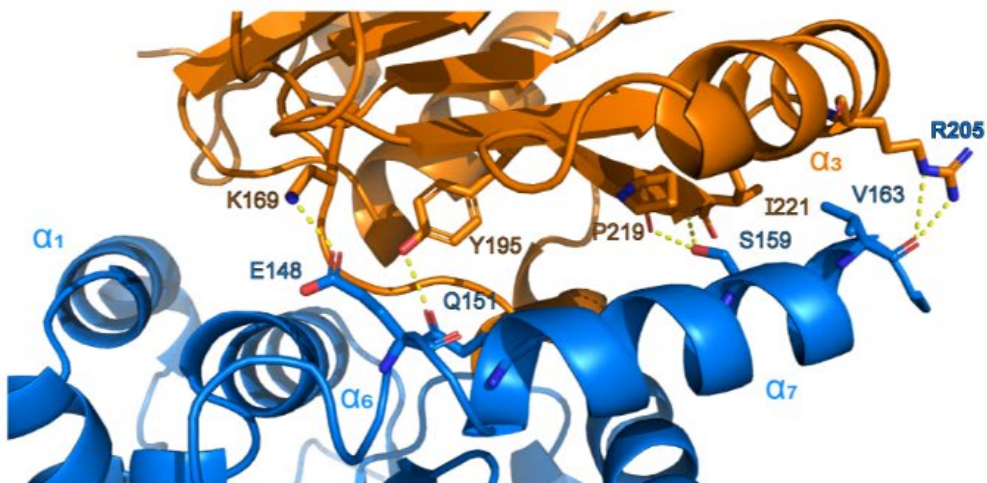
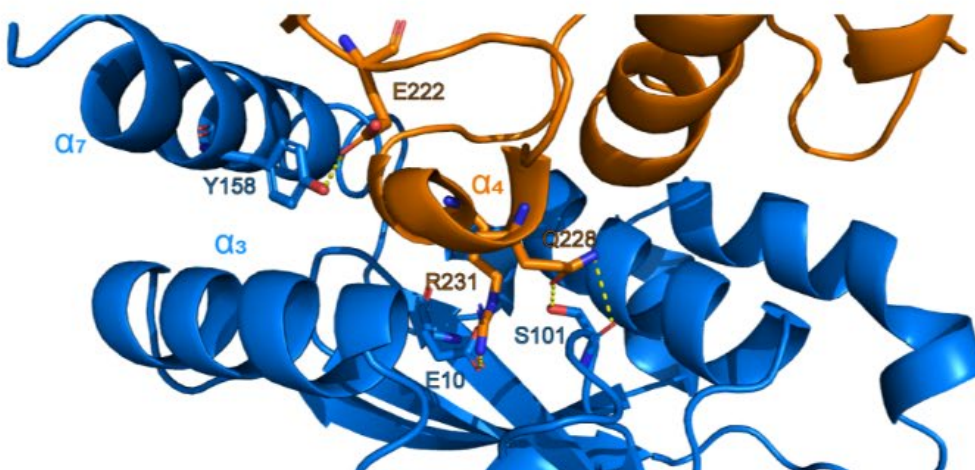


**Figure 3.2.4.** Comparison of yeast Nop1 (our structure) with archaeal Fibrillarin (green) (PDB-ID: 2NNW, 3ID5 and 1NT2).

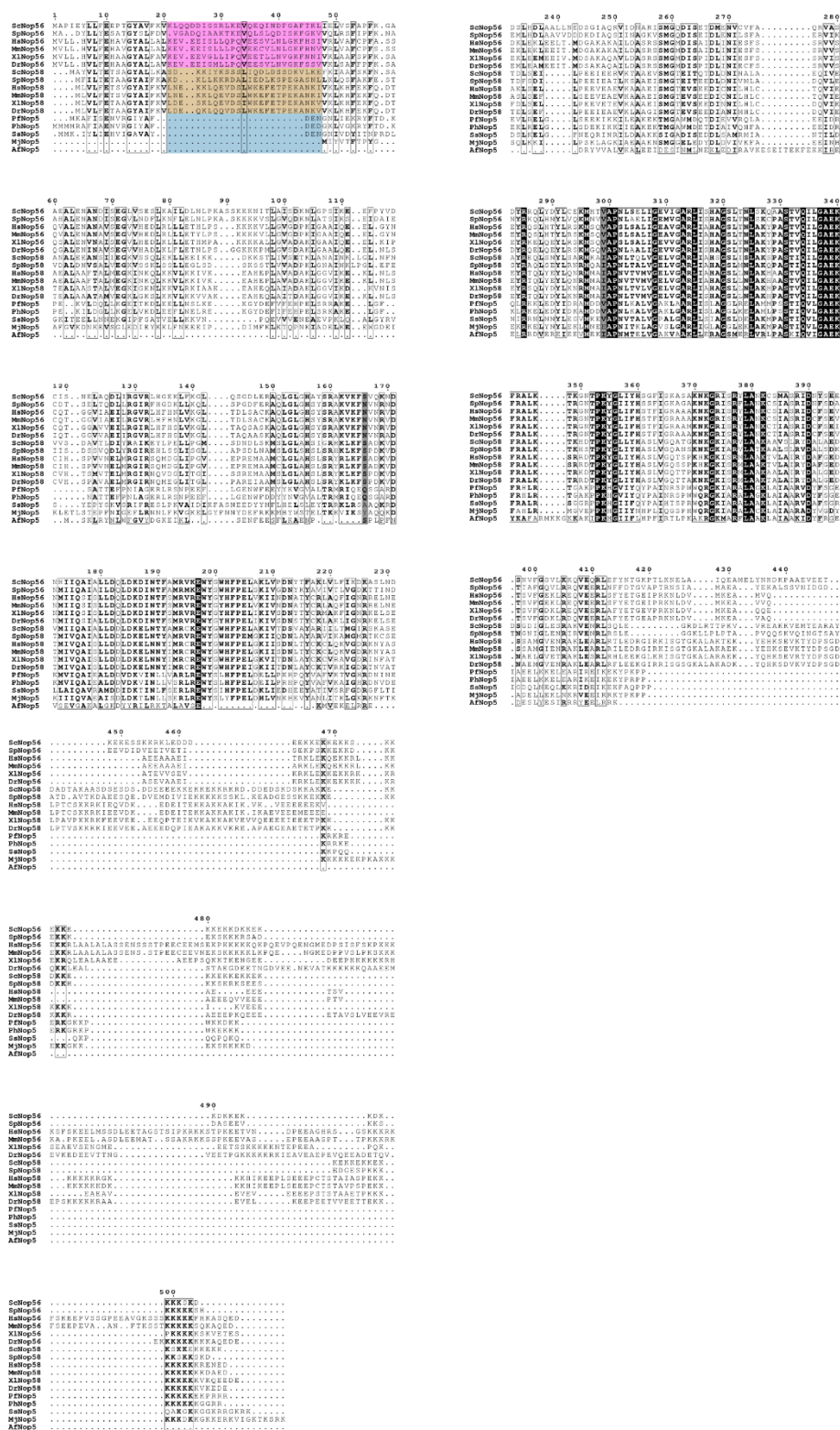
With respect to the N-terminal domain of archaeal Nop5, Nop56 acquired an additional 26 residues between  $\beta$ -sheet 1 and 2 (**Figure 3.2.5**). These additional residues form  $\alpha$ -helix 1 and 2 at the binding interface with Nop1. Interestingly, this insertion is present in all Nop56 orthologs as well as in Nop58. In the course of this project, we also intended to determine the

structure of Nop1<sup>83-327</sup> bound to the N-terminal domain of Nop58 (residue 1 to 155), but we were unable to do so due to the instability of Nop58 (see section 3.6 and 3.7).

From our structure, we could determine the areas and types of interactions that drive the formation of the complex between Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>. Nop1<sup>83-327</sup> engages  $\alpha$ -helix 3,4 and 5 and the C-terminal tail to interact with  $\alpha$ -helix 1, 6 and 7 as well as  $\beta$ -sheet 1 and the loop region between  $\beta$ -sheet 4 and 5 of Nop56<sup>1-166</sup>. These interactions can be divided into three main areas (**Figure 3.2.6**).

**A****B****C**

**Figure 3.2.6.** The three main areas of interactions between Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup> in the protein-protein complex. **(A)** Polar interactions and hydrogen bonds (yellow dashed lines) between Nop1<sup>83-327</sup> (orange)  $\alpha$ -helix 5, 6 and the C-terminal tail and Nop56<sup>1-166</sup> (blue)  $\alpha$ -helix 1 and 6. **(B)** Polar interactions and hydrogen bonds (yellow dashed lines) between Nop1<sup>83-327</sup> (orange)  $\alpha$ -helix 3 and  $\beta$ -sheet 7 and Nop56<sup>1-166</sup> (blue)  $\alpha$ -helix 7. **(C)** Polar interactions and hydrogen bonds (yellow dashed lines) between Nop1<sup>83-327</sup> (orange)  $\alpha$ -helix 4 and Nop56<sup>1-166</sup>  $\alpha$ -helix 3 and 7 and  $\beta$ -sheet 1.



**Figure 3.2.5.** Primary sequence alignment of archaeal Nop5 and eukaryotic Nop56 and Nop58. Sequence alignment was generated with Clustal Omega and displayed with ESPript 3.0. The  $\alpha 1$  and  $\alpha 2$  insertion is highlighted in pink for Nop56 and beige for Nop58. The lack of insertion in Nop5 is highlighted in blue. Sc – *S. cerevisiae*, Sp – *S. pombe*, Hs – *H. sapiens*, Mm – *M. musculus*, XI – *X. laevis*, Dr – *D. rerio*, Pf – *P. furiosus*, Ph – *P. horikoshi*, Ss – *S. solfataricus*, Af – *A. fulgidus*, Mj – *M. jannaschii*

In the first interaction area  $\alpha 1$  and  $\alpha 6$  from Nop56<sup>1-166</sup> contact the extended C-terminal tail of Nop1<sup>83-327</sup> (**Figure 3.2.6 A**); this area includes  $\alpha 1$  (Nop56<sup>1-166</sup>), which is part of the additional residues in eukaryotic Nop56 and Nop58. Nop56-K32 forms a hydrogen bond with the carbonyl-group of Nop1-S323, whereas Nop56-Q35 forms a hydrogen bond with the side chain of the same serine residue in Nop1. Furthermore,  $\alpha 1$  Nop56-N39 forms another hydrogen bond with Nop1-D263. In Nop56<sup>1-166</sup>  $\alpha 6$  Nop56-E122 forms a polar interaction with Nop1-R297 at the C-terminal end of Nop1  $\alpha 5$ . Nop1-R322, in the extended C-terminal tail, forms a hydrogen bond with Nop56-D126, whereas Nop56-R129 forms polar contacts with carbonyl groups of residues at the N-terminal end of Nop1  $\alpha 5$ , and Nop1-R132 forms polar interactions with carbonyl groups of residues in Nop1  $\alpha 4$ .

The second region is centered around Nop56  $\alpha 7$  and Nop1  $\alpha 3$  and  $\beta 7$  (**Figure 3.2.6 B**). Nop56-E148, at the beginning of  $\alpha 7$ , forms a hydrogen bond with Nop1-K169, and Nop56-Q151 forms a hydrogen bond with conserved Nop1-Y195. At the end of Nop56  $\alpha 7$ , S159 forms polar contacts with the backbone of Nop1-P219 and I221, and the backbone of Nop56-V163 forms hydrogen bonds with Nop1-R205.

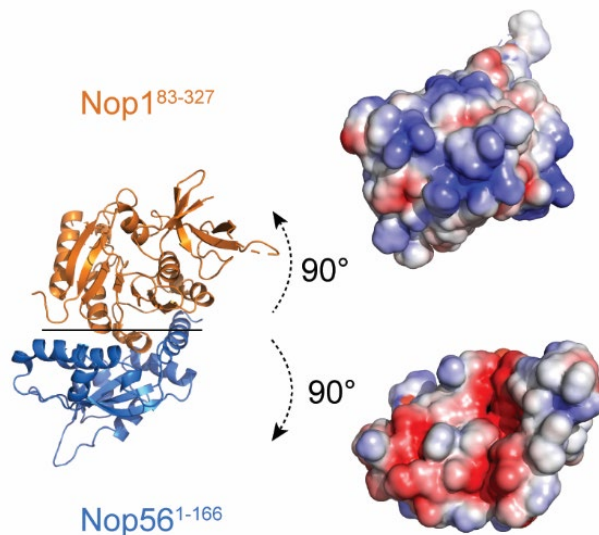
In the third region, Nop1  $\alpha 4$  inserts into the central cavity of the N-terminal domain of Nop56 (**Figure 3.2.6 C**). Nop1-R231 and Q228 form hydrogen bonds with Nop56-E10 and S101, respectively. Further down Nop1  $\alpha 4$ , E222 forms another hydrogen bond with Nop56-Y158 in  $\alpha 7$ . These interactions are driven by the charge complementarity between Nop1  $\alpha 4$ , which is positively charged, and the cavity in Nop56 N-terminal domain, which is mainly negatively charged.

Besides these three major areas of interaction, the complementarity of the charge distribution on the binding surface of both proteins, with Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>, displaying an overall positive and negative charge, respectively, likely drives complex formation (**Figure 3.2.7**).



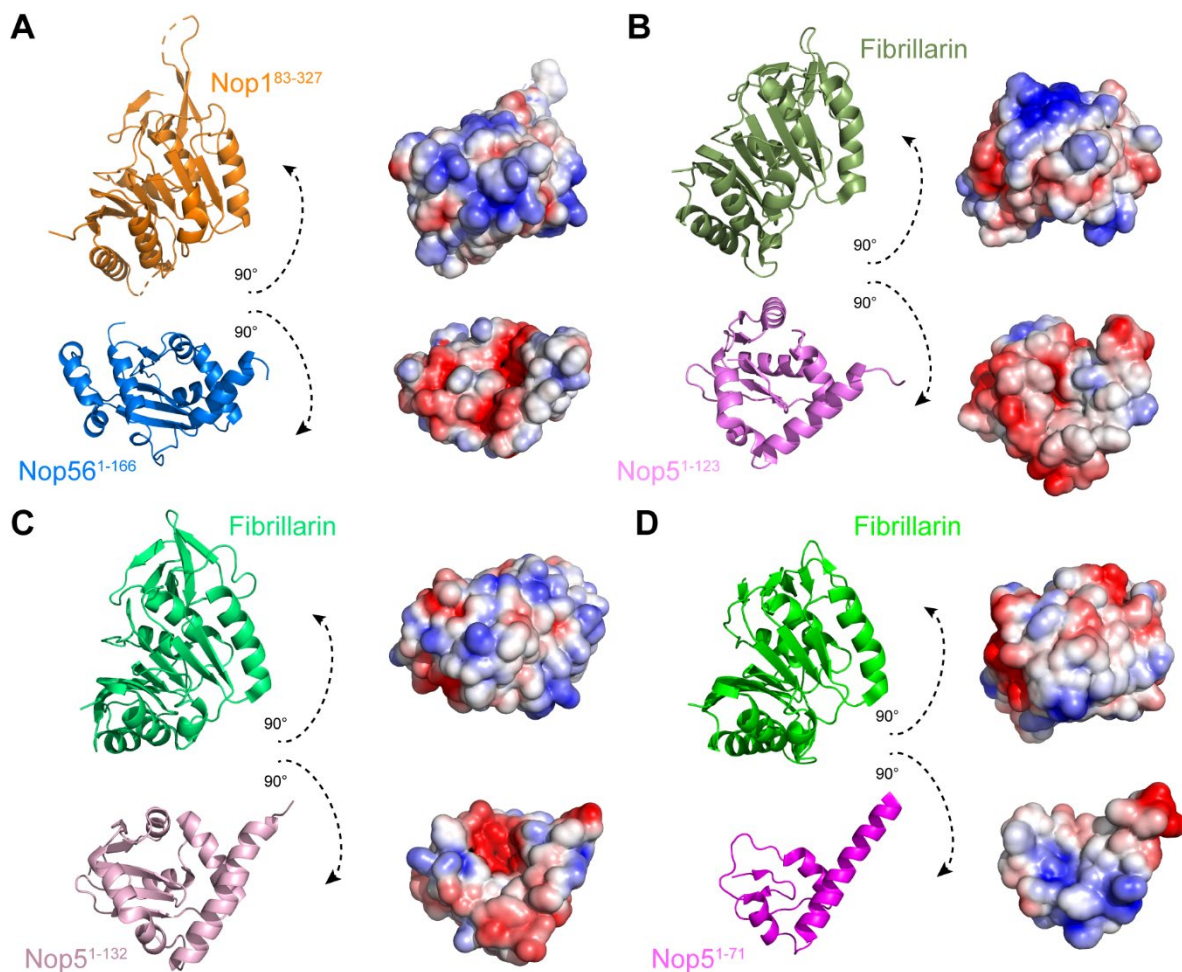
### Differences in charge distribution on the interaction surfaces

After we determined and analyzed the structure of the eukaryotic Nop1<sup>83-327</sup>–Nop56<sup>1-166</sup> protein complex, we wanted to investigate which differences between the eukaryotic and archaeal protein-protein complex are responsible for the incompatibility of the archaeal and eukaryotic proteins (see section 3.2.1).



**Figure 3.2.7.** Charge complementarity on the binding interface of Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>. The charge distribution on the binding interface between Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>. Nop1<sup>83-327</sup> has a mostly positively charged (blue) solvent accessible binding interface, whereas Nop56<sup>1-166</sup> has a mainly negatively charged (red) binding interface.

First, we compared the charge distributions on the solvent-accessible interaction surfaces in all orthologous complexes with available structure. To calculate the charge distribution we used the Adaptive Poisson-Boltzmann Solver (APBS) (Jurrus et al., 2018). From archaea, there is structural data available for three different organisms: *P. furiosus* (PDB-ID: 2NNW) (Oruganti et al., 2007), *S. solfataricus* (PDB-ID: 3ID5) (Ye et al., 2009), and *A. fulgidus* (PDB-ID: 1NT2) (Aittaleb et al., 2003). In **Figure 3.2.8 A-D**, we compare the charge distribution on the interaction surface for the eukaryotic and archaeal complexes.

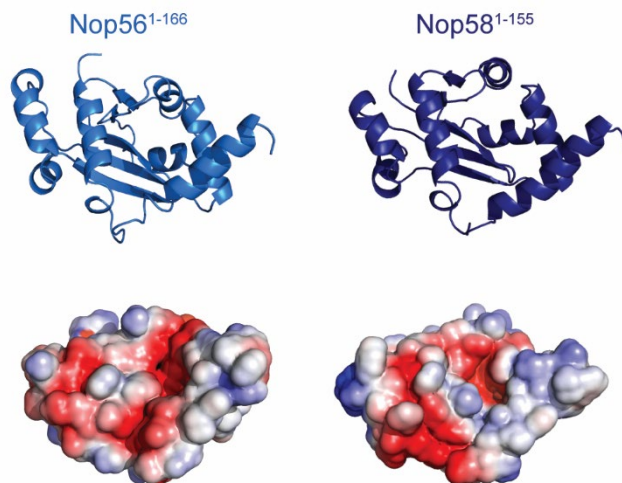


**Figure 3.2.8.** Comparison of charge distribution at the binding interface between Nop1/Fibrillarlin and Nop56/Nop5 N-terminal domain. **(A)** Nop1<sup>83-327</sup> (orange) and Nop56<sup>1-166</sup> (blue) **(B)**, Fibrillarlin (smudge) and Nop5<sup>1-123</sup> (rosa) from *P. furiosus*. **(C)** Fibrillarlin (first green) and Nop5<sup>1-132</sup> (light pink) from *S. solfataricus*. **(D)** Fibrillarlin (bright green) and Nop5<sup>1-71</sup> (magenta) from *A. fulgidus*. In each panel the proteins are twisted 90° away from each other to expose the binding surface. Electrostatic charge distribution is displayed for each protein surface. Blue, white and red represent neutral and negative charges, respectively. Ranges from +/- 3 were chosen in APBS.

In comparison to the eukaryotic complex (**Figure 3.2.8 A**), we saw a weaker charge complementary for the Fibrillarlin-Nop5 complex from the hyperthermophile archaeum *P. furiosus* (**Figure 3.2.8 B**). Despite the fact that the interfaces of Nop5 and Fibrillarlin are predominantly negatively and positively charged, the charges are weaker as compared to the eukaryotic proteins. The same applies for Fibrillarlin and Nop5 from *S. solfataricus* (**Figure 3.2.8 C**). All three complexes share the presence of a negatively charged central cavity of Nop56/Nop5, which fits at least a small positively charged area on Nop1/Fibrillarlin  $\alpha 4$ . For *A. fulgidus* Fibrillarlin and Nop5, the charge distribution diverges entirely from the other three examples (**Figure 3.2.8 D**). First, the Nop5 interface has no dominating charge and lacks the negatively charged central cavity characteristic of the other three systems. On the other side of the complex, Fibrillarlin also does not show a clear pattern of charge distribution, with the

interaction interface being mostly neutral. These factors indicate that the complex formation in *A. fulgidus* is mostly driven by hydrophobic rather than polar interactions. In addition, (Aittaleb et al., 2003) showed that the N-terminal domain of Nop5 from *A. fulgidus* can fold only in the presence of Fibrillarin, which supports the existence of hydrophobic interactions typically required for folding of protein cores.

The N-terminal domain of Nop58 was unable to fold in the absence of Nop1 during recombinant expression in *E. coli*. Therefore, we generated a model of Nop58<sup>1-155</sup> using MODELLER 9.23 (Sali and Blundell, 1993; Martí-Renom et al., 2000; Fiser et al., 2000; Webb and Sali, 2016) and the crystal structure of Nop56<sup>1-166</sup> as a template. From this model, we generated the solvent-accessible surface charge distribution, as described above. The surface charge distribution of Nop58<sup>1-155</sup> was found to be nearly identical to that of Nop56<sup>1-166</sup>, suggesting a similar interaction mode with Nop1.



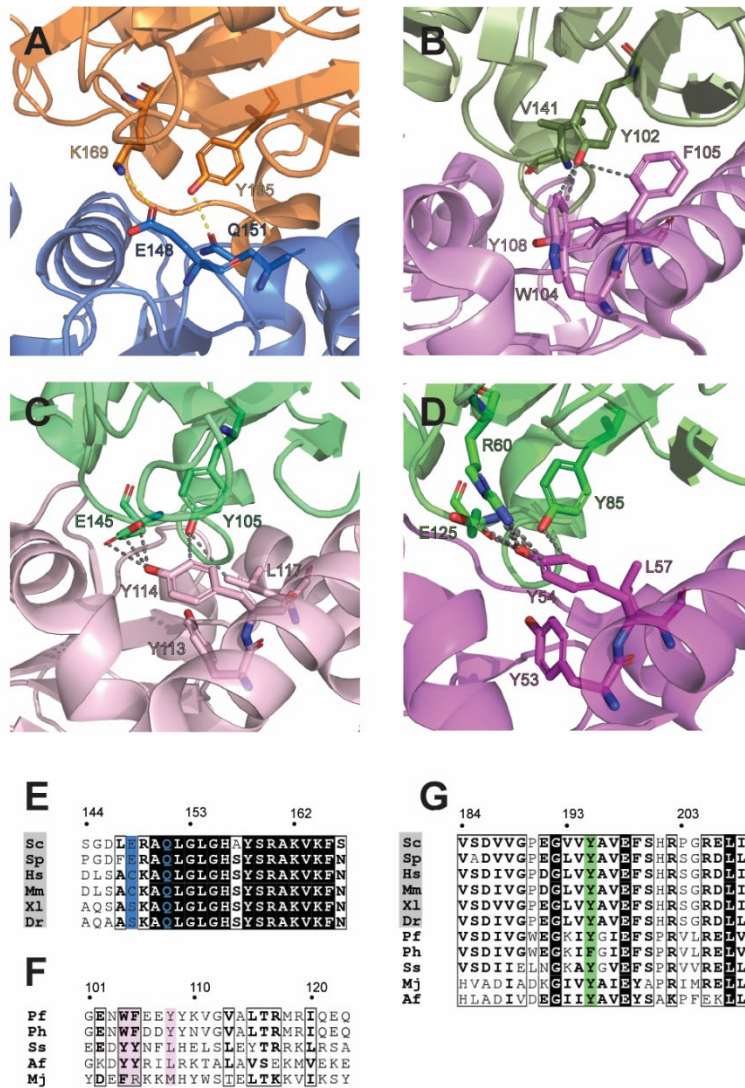
**Figure 3.2.9.** Comparison of solvent-accessible surface charge distribution between the N-terminal domains of Nop56 and Nop58 (model). The structure of Nop58<sup>1-155</sup> was modelled with MODELLER using our crystal structure of Nop56<sup>1-166</sup> as a template. Red indicates mainly negatively charged areas, white a neutral charge and blue positively charged areas. Ranges were chosen from +/- 3 in APBS.

In conclusion, complex formation is driven by charge complementarity for Nop1–Nop56, Nop1–Nop58, and the archaeal Fibrillarin–Nop5 from *P. furiosus* and *S. solfataricus*. For *A. fulgidus*, the driving force for complex formation appears to be of hydrophobic nature.

### *Eukaryotic Nop1—Nop56 complex lacks aromatic interactions found in archaeal orthologs*

After analyzing the charge distribution on the respective interaction surfaces in eukaryotes and archaea complexes, we analyzed the protein–protein contacts in atomic detail. Nop1-Y195 is highly conserved across eukaryotes and archaea (**Figure 3.2.10 G**) but sharply differs in the type of interactions it engages in in the two kingdoms of life. In eukaryotes, it interacts mainly with Q151 from Nop56  $\alpha$ 7 (**Figure 3.2.10 A**), which is also highly conserved in eukaryotes

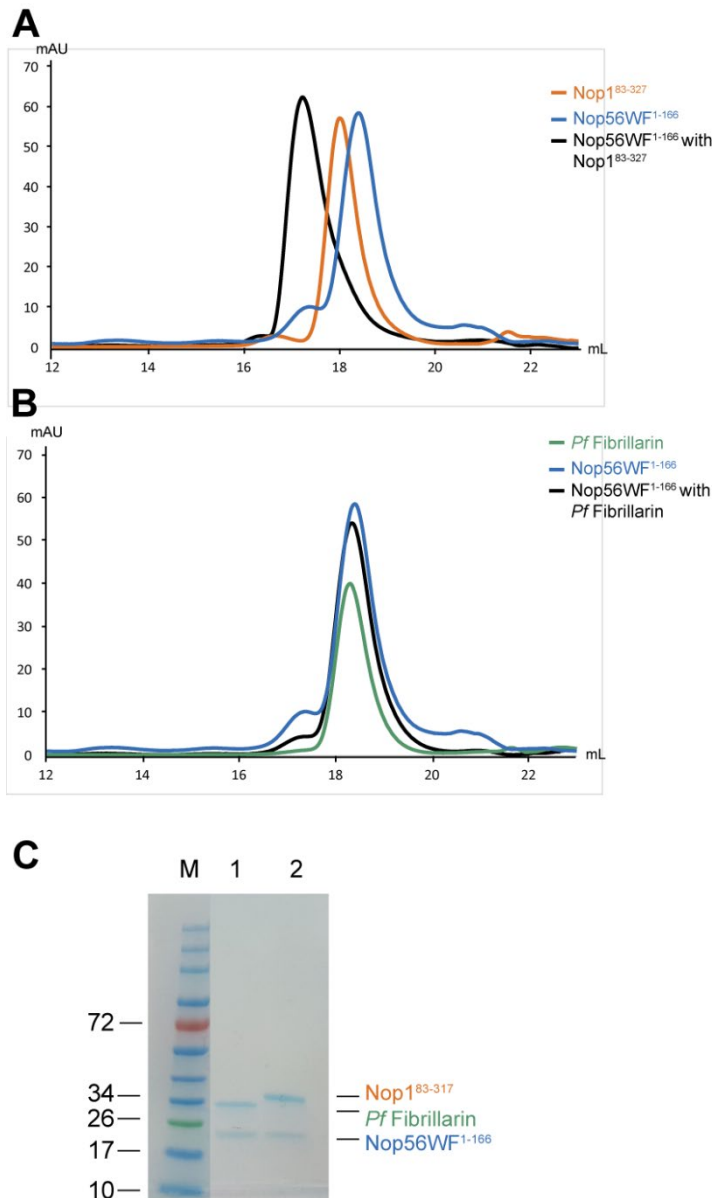
(Figure 3.2.10 E). Next to the tyrosine, the other interacting residue Nop56-E148, which forms a polar contact with Nop1-K169, is only present in fungi and is replaced by small polar residues like cysteine and serine in higher eukaryotes.



**Figure 3.2.10.** Interactions involving the conserved tyrosine in archaeal and eukaryotic Fibrillar/Nop1–Nop5/Nop56 complexes. **(A)** Eukaryotic Nop1 (orange)–Nop56 (blue) lack hydrophobic interactions around the conserved tyrosine residue (Y195). **(B)** Aromatic cluster around the conserved tyrosine (Y102) of *P. furiosus* Fibrillar (smudge) consisting of W104, F105 and Y108 of *P. furiosus* Nop5 (rosa) (PDB-ID: 2NNW, Oruganti et al., 2007). **(C)** Hydrophobic interactions between the conserved tyrosine (Y105) of *S. solfataricus* Fibrillar (forest green) and Y113, Y114 and L117 of *S. solfataricus* Nop5 (light pink) (PDB-ID: 3ID5, Ye et al., 2009). Fib-E145 makes additional electrostatic contacts with Nop5-Y114. **(D)** Interactions between the conserved tyrosine (Y85), R60 and E125 of *A. fulgidus* Fibrillar (green) with Y54 of *A. fulgidus* Nop5 (magenta) (PDB-ID: 1NT2, Aittaleb et al., 2003). Polar interactions are shown by yellow dashed lines, hydrophobic interactions are shown in dark gray. **(E)** Partial sequence alignment of Nop56  $\alpha 7$  in different eukaryotic species. **(F)** Partial sequence alignment of Nop5  $\alpha 6$  in different archaeal species. Important residues are highlighted in blue and light pink in E and F, respectively. **(G)** sequence alignment highlighting the well

conserved tyrosine residue (green) in eukaryotic and archaeal species. Numbering in E, F and G is according to the residue numbers in *S. cerevisiae*. Alignment is displayed with ESPript 3.0.

Next, we asked whether these aromatic contacts are crucial in determining the selectivity of Nop56 N-terminal domain for Nop1 versus archaeal Fibrillarin. We generated the Nop56<sup>1-166</sup>-L147W, E148F mutant (Nop56<sup>1-166</sup>WF) and tested its interaction with Nop1<sup>83-327</sup> and *P. furiosus* Fibrillarin. In this mutant we introduced aromatic residues in the Nop56 region proximal to Nop1-Y195 or Fibrillarin-Y102 (**Figure 3.2.10**) and tested the effect on the binding affinity for eukaryotic and archaeal methyltransferases.



**Figure 3.2.11.** Comparison of size-exclusion chromatograms for Nop56<sup>1-166</sup>WF mixed with Nop1<sup>83-327</sup> and archaeal Fibrillarin. All traces report the absorption at 280 nm wavelength. **(A)** Size-exclusion chromatogram of Nop56<sup>1-166</sup>WF mixed with Nop1<sup>83-327</sup>. The orange and blue traces show the elution profiles of isolated Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>WF, respectively. The black trace shows the elution profile of Nop1<sup>83-327</sup> incubated with Nop56<sup>1-166</sup>WF. **(B)** Size-exclusion chromatogram of Nop56<sup>1-166</sup>WF mixed with *P. furiosus* Fibrillarin. The green and blue traces show the elution profiles of isolated Fibrillarin and Nop56<sup>1-166</sup>WF, respectively. The black trace shows the elution profile of Nop56<sup>1-166</sup>WF incubated with archaeal Fibrillarin. **(C)** SDS gel of black elution peaks from **(A)** and **(B)**. Lane 1 shows the content of the black trace from **(B)**. Lane 2 shows the contents of the black trace from **(A)**. M is the protein size standard.

The presence of these aromatic residues did not disturb the binding of Nop56<sup>1-166</sup> to Nop1<sup>83-327</sup> (**Figure 3.2.11 A**) and had no effect on the inability of Nop56 to bind archaeal Fibrillarin (**Figure 3.2.11 B**). These results suggest that the binding specificity is not achieved by interactions in defined areas but is rather dependent on multiple interactions and thus cannot be changed by mutation of individual interaction points.

### Nop56 $\alpha$ -helix 1 is vital for the interaction with Nop1 and is lacking in archaeal Nop5

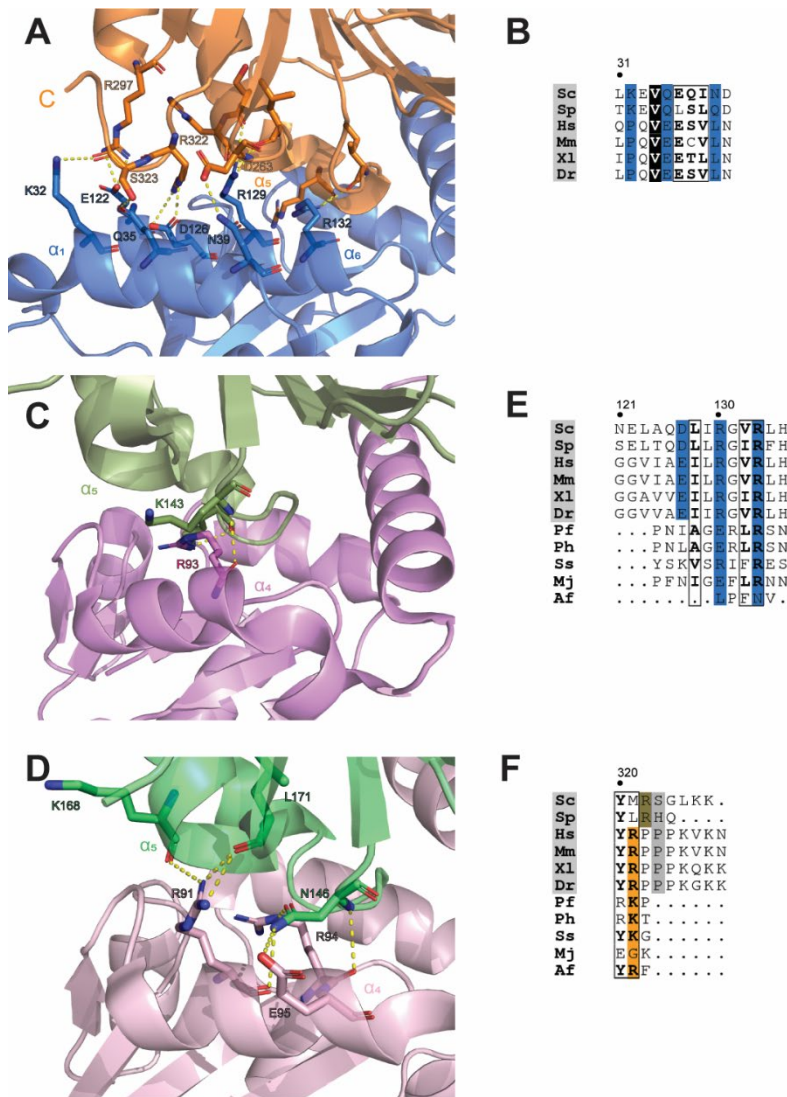
Eukaryotic Box C/D scaffolding proteins Nop56 and Nop58 acquired a 23-26 amino acid long insertion between  $\beta$ 1 and  $\beta$ 2. This 23-26 amino acids fold into two  $\alpha$ -helices,  $\alpha$ 1 and  $\alpha$ 2. From our structure (**Figure 3.2.12 A**), we find that  $\alpha$ 1 is positioned at the interface to Nop1 and therefore plays a role in the formation of the Nop1–Nop56 complex and possibly also in the binding specificity.

Archaeal Nop5 proteins miss these 23-26 amino acids and therefore lack this additional  $\alpha$ -helix at the interface with Fibrillarin (Oruganti et al., 2007; Ye et al., 2009; Lapinaite et al., 2013). Nop5 can engage only  $\alpha$ 4 (equivalent to eukaryotic Nop56  $\alpha$ 6) in interactions with Fibrillarin in this area of the binding surface (**Figure 3.2.12 C-D**). The interaction between Nop5  $\alpha$ 4 and Fibrillarin is mostly electrostatic and involves polar or charged residues of Nop5  $\alpha$ 4. Interestingly, the amino acid composition of Nop5  $\alpha$ 4 and Nop56  $\alpha$ 6 is conserved (**Figure 3.2.12 E**). *A. fulgidus* Nop5 even lacks helix  $\alpha$ 4 and has almost no interaction at this site. The additional helix  $\alpha$ 1 of eukaryotic Nop56 contacts the extended C-terminal tail of eukaryotic Nop1; fittingly, archaeal Fibrillarin has a shorter C-terminal tail (**Figure 3.2.12 F**).

The acquisition of an additional 26 amino acids in Nop56 and the extension of the C-terminal tail in Nop1 reshaped the interaction landscape between the eukaryotic proteins and thereby contributed to the mutual incompatibility between eukaryotic and archaeal orthologs.

The primary sequence of Nop56  $\alpha$ 1 is reasonably well conserved across various eukaryotic species, though it contains slightly more hydrophobic residues in higher eukaryotes as compared to fungal species (**Figure 3.2.12 B**). The amino acid composition of the C-terminal tail of eukaryotic Nop1/Fibrillarin is far less conserved and varies considerably between fungi and higher eukaryotes (**Figure 3.2.12 F**). In high eukaryotes, the C-terminal tail contains a hydrophobic poly-proline stretch, while in fungi, these positions are replaced by polar residues. The increased hydrophobicity in Nop56  $\alpha$ 1 and the Fibrillarin C-terminal tail in higher eukaryotes demonstrates that these two proteins have co-evolved.

Nevertheless, all eukaryotic proteins have a predominantly negatively charged surface on Nop56  $\alpha$ 1, which matches the mainly positively charged Nop1/Fibrillarin C-terminal tail and therefore contributes to the overall charge complementarity between these two proteins.

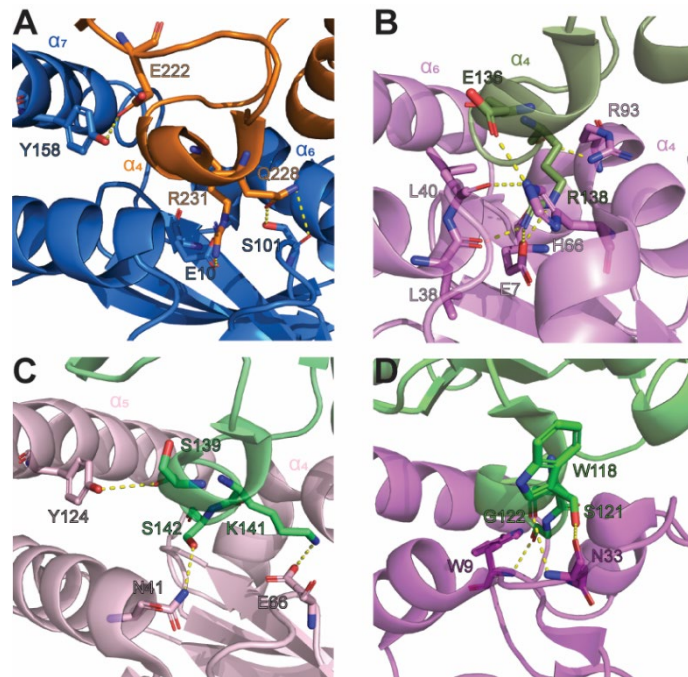


Sc – *S. cerevisiae*, Sp – *S. pombe*, Hs – *H. sapiens*, Mm – *M. musculus*, Xl – *X. laevis*, Dr – *D. rerio*, Pf – *P. furiosus*, Ph – *P. horikoshi*, Ss – *S. solfataricus*, Af – *A. fulgidus*, Mj – *M. jannaschii*

**Figure 3.2.12.** Detailed comparison of interactions between Nop56  $\alpha 1$  and Nop1 and the corresponding region in archaeal complexes. **(A)** Nop56 (blue) engages  $\alpha 1$  and  $\alpha 6$  to interact with Nop1 (orange) C-terminal tail and  $\alpha 5$ , respectively. **(B)** The primary sequence alignment of Nop56- $\alpha 1$  across various eukaryotic species. Interacting residues are highlighted in blue. **(C)** *P. furiosus* Nop5 (rosa) lacks Nop56  $\alpha 1$  and can engage only  $\alpha 4$  in the interaction with Fibrillarin (smudge). Fibrillarin lack the extended C-terminal tail (PDB-ID: 2NNW, Oruganti et al., 2007). **(D)** In *S. solfataricus* Nop5 (light pink) engages  $\alpha 4$  in the interaction with Fibrillarin (forest green)  $\alpha 5$ . Also, here Fibrillarin lacks the extended C-terminal tail (PDB-ID: 3ID5, Ye et al., 2009). Polar interactions are shown as yellow dashed lines. **(E)** Primary sequence alignment of eukaryotic  $\alpha 6$  and archaeal  $\alpha 4$  from Nop56<sup>1-166</sup> and Nop5, respectively. Interacting residues are highlighted in blue. **(F)** Primary sequence alignment of Nop1/Fibrillarin C-terminal tails. Interacting residues are highlighted in orange and grey. Residue numbering is according to *S. cerevisiae* Nop1.

Interaction around Nop1  $\alpha 4$  are highly conserved across eukaryotes and archaea

At the center of the Nop1-Nop56 interaction, Nop1  $\alpha 4$  inserts into the central cavity of Nop56 (**Figure 3.2.13 A**). This area of interaction is also the “hot-spot” of surface charge complementarity, as we discussed previously. Across both kingdoms of life, the methyltransferase  $\alpha 4$  inserts between Nop56  $\alpha 6$  and  $\alpha 7$ , or archaeal Nop5  $\alpha 4$  and  $\alpha 5$ . In eukaryotes, Nop1-R231 forms hydrogen bonds with E10 in Nop56  $\beta 1$ . This interaction is also conserved in *P. furiosus*, where Fibrillarin-R138 forms hydrogen bonds with E7 in Nop5  $\beta 1$  (**Figure 3.2.13 B**). In *S. solfataricus* the arginine is replaced by a serine (S142), which forms a hydrogen bond with Nop5-N41 located in  $\alpha 1$  (**Figure 3.2.13 C**).



**Figure 3.2.13.** Comparison of eukaryotic Nop1 and archaeal Fibrillarin engaging  $\alpha 4$  to interact with Nop56 and Nop5 respectively. **(A)** Nop1 (orange)  $\alpha 4$  interacting with Nop56 (blue). **(B)** *P. furiosus* Fibrillarin (smudge) interacting with Nop5 (rosa) (PDB-ID: 2NNW). **(C)** *S. solfataricus* Fibrillarin (forest green) binding to Nop5 (light pink) (PDB-ID: 3ID5). **(D)** Fibrillarin from *A. fulgidus* interacting with Nop5 (pink) (PDB-ID: 1NT2). Polar interactions are shown as yellow dashed lines.

The hydrogen bond formed by Nop1-Q228 in  $\alpha 4$  and Nop56-S101, in the loop connecting  $\beta 4$  and  $\alpha 5$ , is less conserved. In *P. furiosus*, this interaction is replaced by a hydrogen bond between Fibrillarin-E136 in  $\alpha 4$  and Nop5-H66 in  $\alpha 3$ ; in *S. solfataricus* Fibrillarin-K141 in  $\alpha 4$  forms a polar interaction with Nop5-E66 in the loop between  $\beta 4$  and  $\alpha 3$ . In *A. fulgidus* the interactions in this area occur mostly between sidechains and backbone (**Figure 3.2.13 D**). Overall, the interactions differ more between *A. fulgidus* and the other two archaeal species than they do between eukaryotes and archaea.

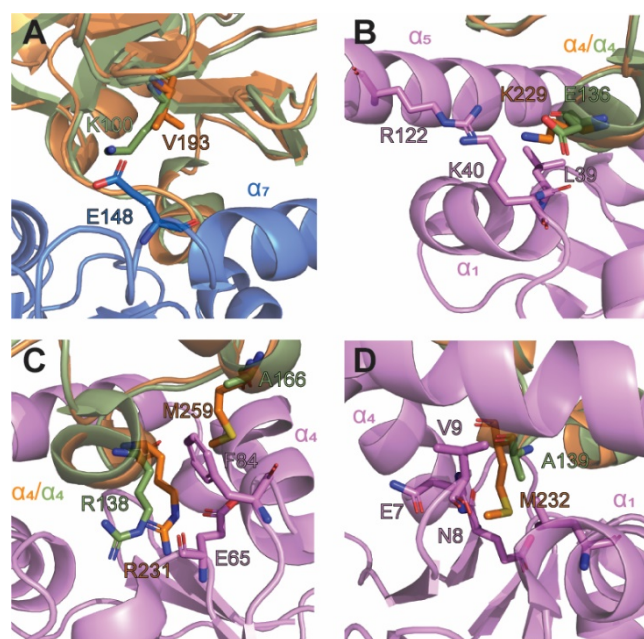


Steric hindrances are critical to the incompatibility between eukaryotic and archaeal protein

After careful examination of the similarities and differences of the attractive forces between the methyltransferase and the scaffolding protein of the Box C/D enzyme in eukaryotes and archaea, we asked whether repulsive forces and sterical hindrances could actively prevent the formation of the complex between Nop1 and archaeal Nop5 or archaeal Fibrillarin and Nop56. To do so, we compared our Nop1<sup>83-327</sup>–Nop56<sup>1-166</sup> structure with that of the Fibrillarin–Nop5 complexes from *P. furiosus* (PDB-ID: 2NNW) (Oruganti et al., 2007).

We found several places where the nature of the respective sidechain of the orthologous binding partner would cause steric clashes or charge repulsion and thereby prevent complex formation. When substituting Nop1 with archaeal Fibrillarin, Fibrillarin-K100 creates a steric clash with Nop56-E148, which is typically engaged in a hydrogen bonding with Nop1-K169. In Nop1 K100 is replaced by V193, which is short enough not to cause a steric clash (Figure 3.2.14 A). At the interface of Nop1  $\alpha_4$  and Nop5  $\alpha_1$  and  $\alpha_5$ , a charge repulsion is caused by Nop1-K229 coming close to Nop5-R122 and K40 (Figure 3.2.14 B).

When Nop56 is replaced by Nop5 in the Nop1–Nop56 complex Nop5-F84 potentially clashes with Nop1-M259 (Figure 3.2.14 C). This residue is replaced by the much smaller A166 in archaeal Fibrillarin. Lastly, Fibrillarin-A139 is replaced by M232 in Nop1, which disturbs the sidechain packing of Nop5 in this central region of interaction (Figure 3.2.14 D).



**Figure 3.2.14.** Potential steric clashes between Nop1<sup>87-327</sup> and Nop5 (*P. furiosus*) and archaeal Fibrillarin and Nop56<sup>1-166</sup>. (A) Fibrillarin  $\alpha_3$  K100 would clash with Nop56<sup>1-166</sup>  $\alpha_7$  E148 in a putative Fibrillarin–Nop56<sup>1-166</sup> complex. (B) Electrostatic repulsion between Nop5<sup>1-123</sup>-R122 and K40 with Nop1<sup>87-327</sup>-K229 in a putative Nop1<sup>87-327</sup>–Nop5<sup>1-123</sup> complex. (C) Steric clash between Nop5<sup>1-123</sup>  $\alpha_4$  F84 and Nop1<sup>87-317</sup>-M259 and another potential steric clash between Nop1<sup>87-327</sup>  $\alpha_4$  R231 and Nop5<sup>1-123</sup>-E65 in a putative complex. (D) Potential steric clash between Nop1<sup>87-327</sup>-M232 and Nop5<sup>1-123</sup>-E7, N8 and V9. The structure of archaeal Nop5<sup>1-123</sup> and Fibrillarin from *P. furiosus* is from (Oruganti et al., 2007) (PDB-ID: 2NNW). Nop1<sup>83-327</sup> (orange), Nop56<sup>1-166</sup> (blue), archaeal Fibrillarin (smudge) and Nop5<sup>1-123</sup> (rosa).

### 3.3 Activity of fully assembled chimeric Box C/D s/snoRNPs

After studying the interaction of individual components in the eukaryotic Box C/D complex and characterizing the differences to the archaeal complexes, we tested whether we were able to reconstitute a chimeric eukaryotic-archaeal complex with methylation activity. In archaeal complexes, the correct assembly of each component in the correct stoichiometry is required for catalytic activity (Cahill et al., 2002; Galardi et al., 2002; Aittaleb et al., 2003; Lin et al., 2011; Lapinaite et al., 2013). To test whether the inability of Snu13 to bind non-canonical kink-turns and kink-loops affects the correct assembly of chimeric complexes and thereby the catalytic activity, we tested the methylation efficiency of each chimeric complex assembled with both archaeal L7Ae and eukaryotic Snu13.

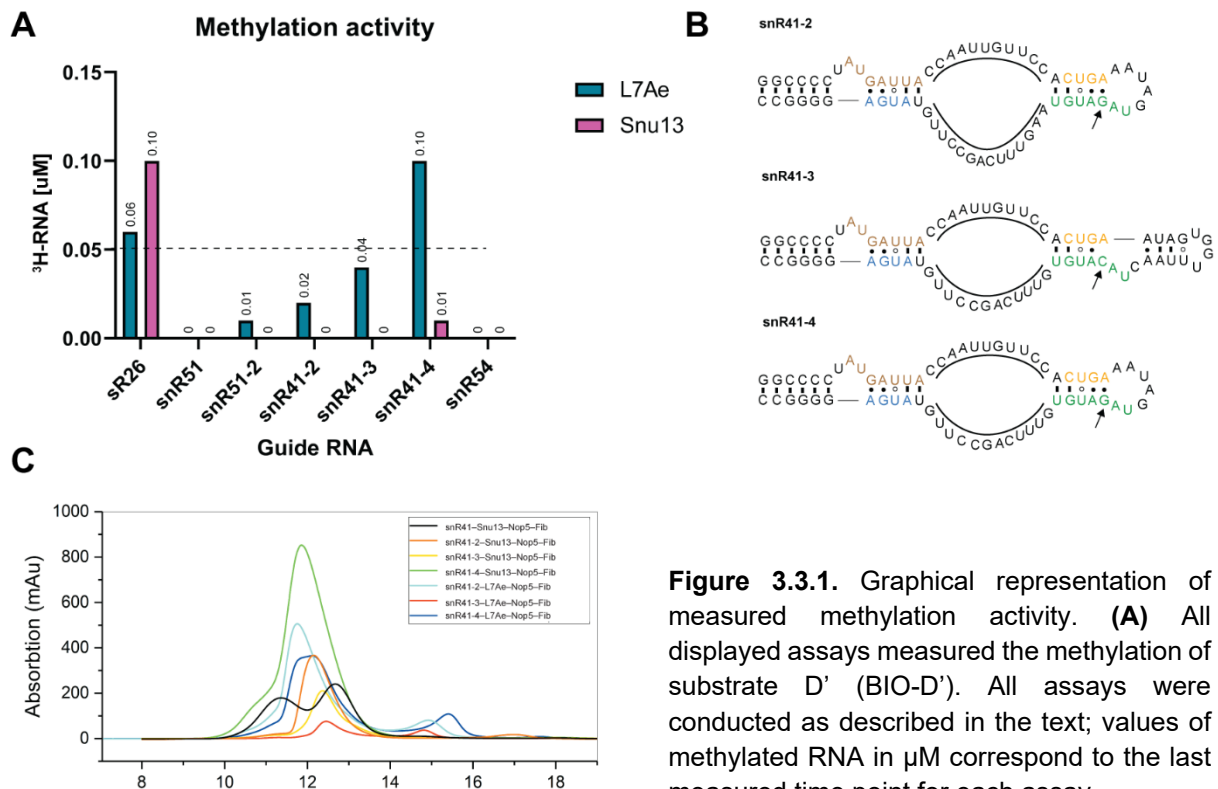
**Table 3.3.1.** Summary of results from methylation assays. The column “guide RNA” notes which sRNA or snoRNA was used for complex assembly. The column “RNA-binding protein” notes whether archaeal L7Ae or eukaryotic Snu13 was used for assembly. All complexes contained archaeal Nop5 and archaeal Fibrillarin. The column “RNA targets” reports which substrate RNAs were used for the assays: the label BIO indicates that this target RNA was 5'-biotinylated and therefore tested for the presence of the methyl; mark (a list of all substrate RNAs can be found in section 5.8). The column “reaction temperature” states at which temperature in °C the assays was conducted. The last column indicates whether methylation activity was detected. A complex was considered active if a value  $\geq 0.05 \mu\text{M}$  of methylated RNA was detected at the last point of measurement.

| guide RNA | RNA-binding protein | RNA targets      | reaction temperature (°C) | Methylation activity |
|-----------|---------------------|------------------|---------------------------|----------------------|
| sR26      | Snu13p              | BIO-D' and D     | 30                        | Yes                  |
| sR26      | L7Ae                | BIO-D' and D     | 30                        | Yes                  |
| snR51     | Snu13p              | BIO-D'           | 37                        | No                   |
|           | Snu13p              | BIO-D            | 37                        | No                   |
| snR51     | L7Ae                | BIO-D' and D     | 50                        | No                   |
|           | L7Ae                | BIO-D and D'     | 50                        | No                   |
| snR51-2   | Snu13p              | BIO-D' and D     | 30                        | No                   |
|           | Snu13p              | BIO-D and D'     | 30                        | No                   |
| snR51-2   | L7Ae                | BIO-D' and D     | 50                        | No                   |
|           | L7Ae                | BIO-D and D'     | 50                        | No                   |
| snR41 -2  | Snu13p              | BIO-D' and BIO-D | 30                        | No                   |
| snR41 -3  | Snu13p              | BIO-D' and BIO-D | 30                        | No                   |
| snR41 -4  | Snu13p              | BIO-D' and BIO-D | 30                        | No                   |
| snR41 -2  | L7Ae                | BIO-D' and BIO-D | 50                        | No                   |
| snR41 -3  | L7Ae                | BIO-D' and BIO-D | 50                        | No                   |
| snR41 -4  | L7Ae                | BIO-D' and BIO-D | 50                        | Yes                  |
| snR54     | Snu13p              | BIO-D'           | 30                        | No                   |
| snR54     | L7Ae                | BIO-D'           | 50                        | No                   |

Assays containing Snu13 were conducted at 30 or 37 °C, while assays containing only archaeal proteins were conducted at 50 °C. This was dictated by the fact that Snu13 is not stable at temperatures higher than 37°C for extended time. The optimal temperature for archaeal Fibrillar activity is  $\geq 70^\circ\text{C}$ ; therefore, the enzyme is potentially less active at 30 or 37°C.

For both wild-type and mutant snR51 guide RNAs, we did not detect any activity with either L7Ae or Snu13. In this case, the fact that L7Ae binds to both the canonical and non-canonical k-turns (and thus the complex should contain two copies of L7Ae) does not seem to support activity. The same holds true for snR51-2 (**Table and Figure 3.3.1**).

Mutant guide RNA snR41-2 contains the snR41-k13 sequence. Both complexes assembled with this guide RNA and either L7Ae or Snu13 showed no activity (**Table and Figure 3.3.1**). In snR41-3, the guide sequence upstream of Box D was shorted to 12 nt from the 5' side to



**Figure 3.3.1.** Graphical representation of measured methylation activity. **(A)** All displayed assays measured the methylation of substrate D' (BIO-D'). All assays were conducted as described in the text; values of methylated RNA in  $\mu\text{M}$  correspond to the last measured time point for each assay.

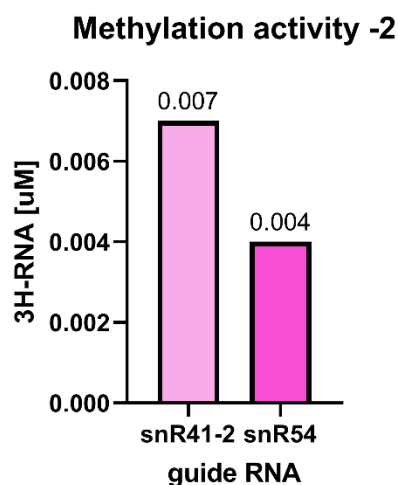
A complex is only considered active if a threshold of 0.05  $\mu\text{M}$  methylated RNA is surpassed (dashed line). Complexes containing archaeal L7Ae or eukaryotic Snu13 are indicated in blue and pink, respectively. The guide RNA used in each assay is noted on the x-axis. **(B)** Schematic representation of sequence predicted secondary structure of mutant Box C/D guide RNAs snR41-2, snR41-3 and snR41.4 from *S. cerevisiae*. In snR41-2 the G•A base in the 1n-1b position has been reintroduced and stem I has been replaced by a loop structure. In snR42-3 the guide sequence upstream of box D has been shorted to 12 nt. In snR41-4 the mutations from snR41-2 and snR41-3 are combined. In all secondary structures box C is in brown, box D in blue, box C' in green and box D' in orange. **(C)** Size-exclusion chromatography profile of chimeric Box C/D RNPs assembled around guide RNA snR41 or guide RNAs derived from snR41. Absorption trace at 280 nm is shown.

comply better with previous observations on archaeal complexes, which indicated an optimal guide sequence length of 10-12 nucleotides for methylation (Tran et al., 2005; Yang et al., 2016b). The guide sequence upstream of Box D' did not need to be shorted as it is only 12 nt long in the wild-type guide RNA. Also, in this case, we could not detect any activity in both chimeric complexes assembled with L7Ae or Snu13 (**Table and Figure 3.3.1**).

Guide RNA snR41-4 combines the mutations from snR41-2 and snR41-3. In this case, the optimized k-loop sequence in combination with short guide sequences led to an active enzyme with archaeal L7Ae but and an inactive enzyme in combination with Snu13 (**Table and Figure 3.3.1**).

Lastly, complexes assembled around snR54 do not show any activity with either L7Ae or Snu13 (**Table and Figure 3.3.1**).

To rule out that the lack of methylation activity observed for all eukaryotic guide RNAs with wild-type guide sequences is due to differences between the activity of archaeal and eukaryotic methyltransferases; we tested snR41-2 and snR54 with Box C/D complexes containing yeast Nop1<sup>87-317</sup>. To incorporate Nop1<sup>87-327</sup> into the complex, we engineered an archaeal Nop5, which contains the N-terminal domain of eukaryotic Nop56 instead of the natural Nop5 N-terminal domain. By engineering the Nop56 N-terminal domain into Nop5, Nop1<sup>87-327</sup> could be integrated



**Figure 3.3.2.** Methylation activity of chimeric complexes containing eukaryotic methyltransferase Nop1<sup>87-327</sup>. For both assays the amount of methylated RNAs at the last point of measurement is plotted.

into the chimeric Box C/D complex. The assays for these complexes were conducted at 30°C. However, regardless of the presence of the native eukaryotic methyltransferase, no activity could be detected for both of the tested complexes (**Figure 3.3.2**).

These results suggest that the lack of activity of eukaryotic complexes may be due to the incorrect folding of

the more complex eukaryotic guide RNA; RNA folding and functional Box C/D snoRNP assembly likely requires the assistance of additional factors, some of which have been characterized *in vivo*.

## 4. Discussion

Deregulation of 2'-O-RNA methylation and the associated enzymatic machinery is involved in several diseases, including various types of cancer, neurological pathologies, and premature aging, as discussed in chapter 1.3. Different techniques showed site-specific, cell-type-dependent differential methylation levels of rRNA that hint at cell-type-specific “specialized” ribosomes mediated via rRNA modifications (Erales et al., 2017; Sharma et al., 2017a). Therefore, a detailed understanding of the molecular details governing the interactions within the Box C/D multi-subunit enzyme, the assembly process and its activity mechanism would allow to target the enzymes in disease treatments as well as advance our understanding on the biogenesis and function of “specialized” ribosomes.

A comprehensive structural and functional understanding of the eukaryotic 2'-O-RNA methylation machinery, the Box C/D snoRNP, has been so far hindered by the lack of protocols to reconstitute the active enzyme homogeneously *in vitro* (Peng et al., 2014; Yang et al., 2020). In contrast, the archaeal Box C/D sRNP enzyme has been studied extensively and used as a proxy to understand the structure and the function of eukaryotic Box C/D snoRNPs (Moore et al., 2004; Aittaleb et al., 2003; Ye et al., 2009; Lin et al., 2011; Lapinaite et al., 2013).

In this work, we aimed to understand the molecular mechanisms that underly the eukaryotic Box C/D snoRNPs and challenged the applicability of the archaeal Box C/D sRNP as a proxy. We investigated the molecular basis of the interaction between the archaeal and the eukaryotic primary RNA-binding proteins L7Ae and Snu13, respectively, with non-standard Box C'/D' motifs and integrated these results into the context of full RNPs. Furthermore, we elucidated and presented the detailed molecular basis for the interaction between the eukaryotic methyltransferase Nop1 and the N-terminal domain of Nop56 and show that the interactions between these two proteins differ substantially from their archaeal counterparts Fibrillarin and Nop5.

### 4.1 L7Ae and Snu13 recognize non-standard Box C'/D' motifs differently

As described in section 1.2.2, the Box C/D guide RNA is the central component of the Box C/D enzyme. In archaea, it has been demonstrated that the symmetric arrangement of all three core protein components (L7Ae, Nop5, and Fibrillarin) is required for efficient methylation (Omer et al., 2006; Rashid et al., 2003; Tran et al., 2003). This arrangement of the core proteins requires the stable interaction of the primary RNA-binding protein L7Ae with the kinked

structures, k-turn or k-loop, formed by both the terminal Box C/D and the internal Box C'/D' motifs (Charron et al., 2004; Moore et al., 2004; Rozhdestvensky et al., 2003; Suryadi et al., 2005). Binding of two copies of L7Ae subsequently provides a platform for integration of Nop5<sub>2</sub>, via the C-terminal domains, and two copies of Fibrillarin that interact with the N-terminal domains of Nop5<sub>2</sub>. (see section 1.2.2). These data are supported by structural information on the complete and functional archaeal Box C/D enzyme from different species and is present irrespectively of the oligomeric state of the RNP (mono- or di-RNP) (**Figure 1.2.9**) (Lapinaite et al., 2013; Lin et al., 2011; Xue et al., 2010; Ye et al., 2009; Yu et al., 2018).

The formation of the symmetric archaeal Box C/D sRNP enzyme, therefore, strongly depends on the ability of L7Ae to bind to both structures formed by the Box C/D and the Box C'/D' motifs. Several *in vitro* studies have demonstrated a broad affinity of L7Ae from different archaeal species to k-turn and k-loops from various sources, including Box C/D sRNA (Charron et al., 2004; Moore et al., 2004; Suryadi et al., 2005) and other archaeal RNAs (Huang and Lilley, 2013, 2014). This broad affinity also includes the ability to bind k-loop structures (Appel and Maxwell, 2007; Gagnon et al., 2010). Additionally to L7Ae having a broad specificity, archaeal Box C/D sRNAs also show a high degree of conservation in the terminal Box C and D and the internal Box C' and D' motifs, which helps to stabilize the interactions at both sites (Omer et al., 2000, 2006).

Whether the arrangement of the eukaryotic Box C/D snoRNPs is symmetric is still under debate with different data sets supporting either a symmetric or an asymmetric geometry. The cryo-EM structure of the U3 Box C/D snoRNP within the 90S pre-ribosomes showed two copies of Snu13 bound to two sites of the U3 snoRNA, yielding a symmetric arrangement of the four protein components (Snu13, Nop56, Nop58, and Nop1) (Barandun et al., 2017; Cheng et al., 2019). *In vivo* pull-down assays with *S. cerevisiae* U24 Box C/D snoRNPs also showed Snu13 bound to both Box C/D and Box C'/D' motifs, suggesting a symmetric arrangement of the proteins around the two sites (Qu et al., 2011). On the other hand, *in vivo* cross-linking data from *Xenopus leavis* U25 Box C/D snoRNP revealed that only one copy of Snu13 binds to the Box C/D motif of this RNA (Cahill et al., 2002). The hypothesis of an asymmetric assembly of the proteins at the Box C/D and Box C'/D' sites of snoRNA is supported by the fact that Box C' and D' motifs are poorly conserved in eukaryotes (van Nues et al., 2011) and that Snu13 shows a decreased affinity towards kinked structures as compared to L7Ae (Charron et al., 2004).

As available data do not provide a conclusive answer, we first asked whether the absence of a conserved Box C'/D' in eukaryotic Box C/D snoRNAs is the main factor preventing the recruitment of a second copy of Snu13 to the complex and second whether the lack of a second Snu13 binding site on the guide RNA impacts the ability to assemble a functional eukaryotic enzyme *in vitro*. As an initial step, we aimed at determining the sequence

requirements of Box C'/D' motifs necessary to form a kinked structure competent to bind eukaryotic Snu13 and, in comparison, archaeal L7Ae *in vitro*. We found that Snu13 cannot be recruited to k-loops with conserved canonical Box C'/D' motifs and to non-standard Box C'/D' k-turns, but is recruited to canonical Box C'/D' k-turns. The lack of binding of Snu13 to k-loops is consistent with a previous study from (Gagnon et al., 2010). Our results showed that Snu13 can be recruited to non-canonical k-turn with low affinity, provided the presence of both sheared G•A base pairs at the 1n-1b and 2n-2b positions in stem II as well as of a stable stem I. This is the first time that the binding preferences of Snu13 were analyzed systematically.

In comparison, L7Ae is recruited with high affinity to both canonical k-loops and k-turns. We also showed that L7Ae can bind non-standard kink-loops, though with lower affinity, as well as non-standard k-turns with the same or only slightly decreased affinity. Our biochemical data demonstrated that the presence of a G•A base pair in position 1n-1b in stem II is the major requirement for L7Ae recruitment to k-turns *in vitro*. We could confirm this observation in a high-resolution structure of L7Ae bound to the non-standard Box C'/D' k-turn motif from yeast guide snoRNA snR51, which shows a distorted stem II with just one G•A base pair (in this case formed between 1n and 4b) bound to L7Ae.

Notably, we observed a generally lower affinity of Snu13 to standard, canonical Box C'/D' k-turns as compared to L7Ae. This observation was recapitulated by attempts to crystallize Snu13 with canonical k-turns, which yielded crystals of unbound protein, and finds support in previous literature reports (Charron et al., 2004; Gagnon et al., 2010; Marmier-Gourrier et al., 2003; Turner and Lilley, 2008).

Based on available structures of Snu13 and L7Ae bound to canonical k-turns and our structure of L7Ae bound to a non-standard k-turn, we proposed that an essential contact between Snu13 and the RNA is responsible for the requirement of both stem II G•A base pairs. The formation of a critical hydrogen bond between Snu13  $\alpha$ 2-K42 and the guanosine of the 2n-2b G•A is observed in all available Snu13–RNA structures (Bai et al., 2018; Barandun et al., 2017; Charenton et al., 2019b; Cheng et al., 2019; Liu et al., 2007b, 2011; Nguyen et al., 2016; Vidovic et al., 2000). Conversely, the hydrogen bond between the corresponding amino acid L7Ae  $\alpha$ 2-K41 and the guanosine of the 2n-2b base pair is dispensable for the recruitment of L7Ae to k-turns or k-loops, and it is present in only a small subset of the over 100 available L7Ae–RNA structures (Ahmed et al., 2020; Huang and Lilley, 2013, 2014, 2016; Huang et al., 2019; Li and Ye, 2006b; Moore et al., 2004; Oshima et al., 2016, 2018; Xue et al., 2010; Ye et al., 2009). In our L7Ae–RNA structure, the guanine in the 2n-2b base pair is replaced by uracil, which makes it impossible for the L7Ae  $\alpha$ 2-K41 to form any hydrogens bonds with the base. Therefore, the  $\alpha$ 2-K41 adopts four different alternative conformations in the four chains present in the asymmetric unit of the crystal, but the ability of L7Ae to bind this RNA is not compromised.

Snu13-K42 and L7Ae-K41 are located in helix  $\alpha 2$ , which is highly conserved between archaea and eukaryotes; thus, the differences in the contacts made by Snu13-K42 and L7Ae-K41 may only be explained invoking differences in the packing of the secondary structure elements as well as protein dynamics. However, the dependence of the formation of a Snu13–RNA complex on one interaction highlights that the ensemble of contacts between Snu13 and k-turns is weaker as compared to L7Ae.

As mentioned above, our data, together with data from (Gagnon et al., 2010), showed that Snu13 cannot be recruited to canonical k-loops, in contrast to archaeal L7Ae that binds canonical k-loops with high affinity. As both proteins share a nearly identical tertiary structure, with the residues at the binding interface being highly conserved, this difference in binding specificity cannot be due to differences in the protein structures. (Gagnon et al., 2010) proposed that the residues in loop 9 of both proteins are responsible for the difference in binding specificity and the ability of L7Ae to bind k-loops. The amino acid composition of loop 9 in the two proteins differ in that L7Ae loop 9 contains branched hydrophobic residues while Snu13 loop 9 contains polar, charged, and short hydrophobic residues. Our systematic mutational analysis showed that partial conversion of the sequence of Snu13 loop 9 to that of L7Ae loop 9 does not confer Snu13 the ability to be recruited to canonical k-loop elements. However, increasing the hydrophobicity of Snu13 loop 9 improved the stability of the complex between Snu13 and the canonical k-turn from eukaryotic guide RNA snR51-2. In particular, the mutation R95V together with the mutation of the hydrophobic residue in position 93, 96 or 98 to the corresponding amino acid in L7Ae have beneficial effect on the affinity of the Snu13–RNA complex. Hydrophobic residues are ideal to interact with base rings and ribose sugars in a sequence- and structure-independent manner. Nevertheless, the fact that none of the Snu13 mutants could be recruited to canonical k-loops indicates that the mode by which the Snu13 mutants interact with k-loops differs from L7Ae.

In conclusion, we demonstrated that Snu13 can only be recruited the canonical k-turn elements with a stable stem I and II with high affinity *in vitro*. Our observations together with the sequences of all characterized Box C'/D' motifs from *S. cerevisiae*, which frequently lack the 2n-2b G•A base pair, suggest that the binding of Snu13 to the eukaryotic Box C'/D' sites is not required for enzymatic activity of the snoRNP (van Nues et al., 2011). Our data support an asymmetric eukaryotic Box C/D snoRNP model in which the binding of Snu13 to the Box C'/D' motif is not essential for the assembly of a functional enzyme, but can nevertheless occur in some Box C/D snoRNAs.



## **4.2 The assembly of active eukaryotic Box C/D snoRNPs is not dependent on the binding of Snu13 to the Box C'/D' motif**

In the fully assembled Box C/D enzyme, Snu13 does not interact with the guide snoRNA in isolation but is integrated into the RNP complex. In the archaeal complex, the L7Ae–RNA sub-complex interacts with the C-terminal domain of the scaffolding protein Nop5 (described in section 1.2.2), which leads to the integration of the Nop5 homodimer and the methyltransferase Fibrillarin into the Box C/D sRNP (Lapinaite et al., 2013; Lin et al., 2011; Xue et al., 2010; Ye et al., 2009; Yu et al., 2018). A similar interaction between the Snu13–RNA sub-complex and the C-terminal domains of Nop56 and Nop58 is present in the U3 Box C/D snoRNP structure in the 90S pre-ribosome (Barandun et al., 2017; Cheng et al., 2019).

We showed that L7Ae is recruited to canonical k-turns and k-loops as well as to non-standard k-turns and k-loops with high affinity, which is in agreement with literature data (Appel and Maxwell, 2007; Charron et al., 2004; Gagnon et al., 2010). Thus, the interaction of the C-terminal domain of Nop5 is not necessary to stabilize the L7Ae–RNA sub-complex. Snu13, on the other hand, displays a generally reduced affinity to canonical k-turns, and we could not observe any interaction with canonical k-loops. However, in the context of the full snoRNP, the interaction between the composite Snu13–RNA surface and the C-terminal domain of Nop56 might promote recruitment of Snu13 to the Box C'/D' site and enable the assembly of a symmetric eukaryotic Box C/D snoRNP.

The C-terminal domains of archaeal Nop5 and eukaryotic Nop56 and Nop58 share 50-55% sequence similarity; this number excludes the potentially disordered KDE-rich tails. Thus, we used the archaeal Nop5<sub>2</sub>-Fibrillarin<sub>2</sub> sub-complex as a suitable proxy for the Nop56–Nop58–Nop1<sub>2</sub> complex. This approach allowed us to reconstitute a chimeric Box C/D enzyme, with Nop5, archaeal Fibrillarin, Snu13, and eukaryotic (or archaeal) Box C/D guide RNAs and assess the stoichiometry of each of the three proteins in dependence of the snoRNA sequence.

Strikingly, we found that the interaction between Snu13 and the C-terminal domain of Nop5 could stabilize the interaction between Snu13 and the k-loop formed by the Box C'/D' motif of archaeal guide sRNA sR26 (sR26-kl), as we could detect two copies of Snu13 bound to this RNA. Motivated by this observation, we hypothesized that a similar stabilization could also occur for Snu13 in complex with non-standard k-turns from Box C'/D' motifs of eukaryotic snoRNAs. We tested this hypothesis by determining the protein stoichiometry of chimeric Box C/D snoRNPs assembled with yeast guide RNAs snR41, snR51, and snR54. We found that Snu13 is not recruited to the Box C'/D' motif, even in the presence of Nop5. This result leads to the conclusion that the absence of a stem I, as in a k-loop, can be compensated by the presence of Nop5, if a canonical Box C'/D' element is present, but cannot be compensated in

the absence of a canonical Box C'/D' motif. In other words, the presence of Nop5 enables Snu13 to bind canonical Box C'/D' k-loops, but not non-canonical Box C'/D' elements.

Our finding explains contradictory literature on the ability of Snu13 to bind snoRNA Box C'/D' motifs: Snu13 could not be recruited to the non-canonical Box C'/D' k-loop of *Xenopus leavis* U25 snoRNA (Cahill et al., 2002), whereas it was found to bind the canonical Box C'/D' k-loop of human U24 snoRNA (Qu et al., 2011).

We clearly showed that Snu13 can only bind to canonical Box C'/D' motifs. In the case of k-turns, Snu13 is recruited to the Box C'/D' element with high affinity also in isolation, whereas in the case of k-loops, the presence of Nop56 and Nop58 is essential for stabilizing the Snu13–RNA interaction. Given these results, we can postulate that the recruitment of a second copy of Snu13 to the guide RNA requires a canonical C'/D' motif. In eukaryotic Box C/D snoRNAs only Box C and D are highly conserved, and the internal Box C' and D' motifs are often degenerate (van Nues et al., 2011). Thus, we concluded that the second Snu13 binding site is absent in most snoRNAs, which again suggests recruitment of Snu13 to the Box C'/D' element is not essential for assembly and functionality of the eukaryotic Box C/D snoRNP enzyme.

If the second Snu13–RNA platform is not needed for the RNP assembly, the question arises on how Nop56 is recruited to the RNP and whether it interacts with the poorly conserved Box C'/D' motif of the guide RNA. Multiple studies demonstrated that chaperones and assembly factors are involved in the assembly of the eukaryotic Box C/D snoRNP *in vivo*. Almost all characterized auxiliary factors, including Rvb1/2, Rsa1, Naf1, and Bcd1, are involved at different stages of the RNP assembly in mediating interaction between the core protein components Snu13, Nop56, Nop58, and Nop1 or Fibrillarin (Khoshnevis et al., 2019; Massenet et al., 2017; Rothé et al., 2014a, 2014b, 2017). On the other hand, little is known on the RNA processing factors and chaperones involved in the processing and remodeling of snoRNAs in parallel to protein recruitment.

In contrast to the archaeal Box C/D sRNP, the Box C'/D' motif is dispensable for methylation upstream of Box D in eukaryotes, whereas the absence of Box C/D leads to an inactive enzyme by impairing methylation upstream of both Box D and D' *in vivo* (Qu et al., 2011). This observation, together with our data, suggests that the Nop56–Nop58 dimer is recruited to the RNP via the Snu13–RNA sub-complex at the Box C/D site, whereas the Box C'/D' motif is not capable of recruiting Nop56 by itself. This assembly mechanism and the resulting eukaryotic RNP differ substantially from those of the archaeal counterpart, where the anchoring of the Nop5 dimer to the two L7Ae–RNA sub-complexes is essential for correct positioning of the methyltransferase with respect to the guide RNA, and consequently for efficient methylation.

According to the data discussed above, we hypothesized that the failure to recruit Snu13 to the Box C'/D' site *in vitro* is not the reason why many laboratories failed to reconstitute a functional snoRNP *in vitro*. To test this, we measured the activity of chimeric RNP assembled with Nop5<sub>2</sub>-Fibrillarin<sub>2</sub> and either Snu13 or L7Ae: we found that only RNPs containing archaeal sRNAs possessed methylation activity, in the presence of either Snu13 or L7Ae. None of the RNPs containing wild type snoRNAs from yeast (snR41, snR51, and snR54) showed any activity with either L7Ae or Snu13, despite the fact that L7Ae could bind the Box C'/D' element of snR51. Strikingly, also, the RNP containing snR51-2, whose Box C'/D' motif is recognized by both L7Ae and Snu13, did not show any activity. Only after eliminating all eukaryotic specific features from snR41 (i.e. reducing the length of the guide sequence to 12 nt, introducing a canonical Box C'/D' motif, removing the AU-rich spacer elements), we could detect activity in combination with L7Ae. This result indicates that the eukaryotic specific features, especially the AU-rich spacers, require RNA chaperones for the snoRNA to adopt a methylation-competent conformation.

In conclusion, our data suggest that the binding of Snu13 to the Box C'/D' element of the guide snoRNA is not required for the assembly of a functional eukaryotic Box C/D snoRNP enzyme and thus is not crucial for the successful reconstitution of an active enzyme *in vitro*. Instead, the data indicate that the critical step during the assembly processes is the remodeling of the guide snoRNA into a methylation-competent conformation in parallel to the recruitment of the core proteins Snu13, Nop56, Nop58, and Nop1 or Fibrillarin. These results might assist in the pursuit to identify RNA chaperones involved in the assembly of the Box C/D enzyme and therefore help the successful *in vitro* reconstitution of the enzyme for detailed structural and functional analysis. Furthermore, our data highlights crucial differences between the archaeal and the eukaryotic Box C/D enzymes and suggests that the lesson learned from the archaeal enzyme cannot be entirely transferred to eukaryotic enzymes.

### **4.3 Interaction between the Box C/D scaffolding proteins and the methyltransferase differ between archaea and eukaryotes**

Archaeal and eukaryotic Box C/D enzymes both contain a SAM-dependent methyltransferase, which is the catalytic unit responsible for the transfer of a methyl group from the SAM cofactor to the 2'-OH position of the ribose sugar of a defined nucleotide in the target RNA. In comparison to other components of the Box C/D RNP, the methyltransferase Fibrillarin or Nop1 in human or yeast, respectively, is well conserved between archaea and eukaryotes and shares, both a highly conserved tertiary structure (see section 1.2.2, **Figure 1.2.11**), and a ~

70% sequence similarity (Rodriguez-Corona et al., 2015). The methyltransferase is specifically well conserved in its catalytic site and the RNA binding motif, with an average 85.7% and 79.5% sequence similarity between Nop1 and archaeal Fibrillarins in residues in a range of 4 Å from either the cofactor or RNA substrate, respectively. The most striking difference between eukaryotic and archaeal Fibrillarins is the presence of an ~ 80 amino acids long N-terminal RGG motif or GAR domain in eukaryotic Fibrillarins (Rodriguez-Corona et al., 2015). The RGG motif is present in more than 1000 characterized human proteins and is involved in transcription, DNA damage signaling, and apoptosis, among other things (Thandapani et al., 2013). The exact role of the RGG motif in eukaryotic Fibrillarins, including yeast Nop1, remains unclear. The RGG motif of several nucleolar proteins, including Nop1, is subject to methylation by the Htm1p methyltransferase, although the function of this modification is also still unknown (Yagoub et al., 2015).

The integration of the catalytic subunit Fibrillarin or Nop1 into the Box C/D enzyme is essential for the assembly of a functional RNP both in archaea and in eukaryotes. In archaea, Fibrillarin interacts with the N-terminal domain of Nop5 and thereby is recruited to the sRNP. The interaction interface has been characterized for Box C/D sRNPs from several different archaeal species and is very well understood (Aittaleb et al., 2003; Lin et al., 2011; Xue et al., 2010; Ye et al., 2009). From low-resolution cryo-EM structures of the U3 snRNP, we know that Nop1 is also integrated into the Box C/D snoRNP through interaction with the N-terminal domains of Nop56 and Nop58 (Barandun et al., 2017; Cheng et al., 2019). Strikingly though, the N-terminal domains of archaeal Nop5 and eukaryotic Nop56 and Nop58 share on average only 29% sequence similarity. As a result, we found that the Nop1 and the N-terminal domain of Nop56 from *S. cerevisiae* cannot bind their respective orthologous binding partner from archaeon *P. furiosus*, namely the Nop5 N-terminal domain or Fibrillarin.

Here we presented the first 1.7 Å high-resolution structure of the Nop56–Nop1 complex from *S. cerevisiae* and show that the driving force of the interaction is of electrostatic nature, similar to the archaeal Nop5–Fibrillarin complex; however, the specific interactions differ substantially between eukaryotes and archaea. Furthermore, we demonstrated that two additional  $\alpha$ -helices of Nop56 expand the Nop56–Nop1 interface by interacting with the extended C-terminal tail of Nop1; these structural elements are absent in archaea. With the high-resolution structure of the N-terminal domain of Nop56 in hand we also generated a homology model of the N-terminal domain of Nop58, which revealed an almost identical electrostatic surface charge distribution to Nop56, suggesting that Nop58 interacts in a very similar way with Nop1.

The differences in the interaction surface between archaeal and the eukaryotic protein-protein complex raises the question, which factors might have influenced the subtle differences between the archaeal and eukaryotic proteins that lead to their incompatibility with each other.

The most noticeable difference consists in the cellular environments under which the proteins operate. Archaea often inhabit hostile and extreme ecological niches and live under high-salt, high-temperature, or extreme pH conditions. To survive and multiply under such extreme conditions, archaeal proteins have acquired specific adaptations that protect them from denaturation; these include the preferred formation of higher-order oligomers, tighter protein packing due to favorable hydrophobic interactions, increased number of salt-bridges and surface charges, among others (Reed et al., 2013). Furthermore, archaea lack a dedicated nucleus. In eukaryotes, on the other hand, the Box C/D RNP core proteins are located in the nucleolus, a sub-department of the nucleus, dedicated to the production of rRNA (Pederson, 2011). Operation in these two very different environments posed different requirements on the proteins involved in the Box C/D RNP enzymes and forces them to appropriate adaptations.

Additionally, eukaryotic Fibrillarin or Nop1 also acquired additional roles as compared to the archaeal Fibrillarin. In human and *S. cerevisiae*, Fibrillarin or Nop1 also methylates Q104 or Q105 of histone H2A, respectively (Iyer-Bierhoff et al., 2018; Tessarz et al., 2014). The factors involved in this process are so far unknown, but most likely engage the same interface as Nop56 and Nop58. Furthermore, human Fibrillarin also has a proven interaction with nucleophosmin (NPM1) in an RNA-independent manner (Nachmani et al., 2019); also for this interaction, no molecular or structural details are available so far.

On the other side, also Nop56 and Nop58 have gained more interaction partners in eukaryotes, which might have influenced the altered sequence and structure as compared to archaeal Nop5. While the AAA+ ATPase R2TP complex is recruited to the C-terminal unstructured KDE-rich tail of Nop58 (Yu et al., 2019), other interactions partners contact yet unknown surfaces of Nop56 or Nop58. This includes the *Drosophila* protein hoip, which interacts with both Nop56 and Nop58 (Murata et al., 2008) and the R2TP complex component Nop17, which associates with Nop58 during Box C/D maturation (Prieto et al., 2015).

These findings, together with our high-resolution structure, suggest that eukaryotic Fibrillarin and the N-terminal domains of Nop56 and Nop58 coevolved to enable the expanded interaction with additional partners while ensuring stable complex formation between Nop1/Fibrillarin and Nop56 or Nop58.

In conclusion, our high-resolution structure of the Nop56–Nop1 complex again challenges the applicability of the archaeal Box C/D sRNP model to the eukaryotic system. We showed that the archaeal and eukaryotic methyltransferase cannot complement each other as they lack the ability to interact with the respective orthologous Box C/D scaffolding protein Nop5 or Nop56/Nop58, respectively. Furthermore, we characterized the differences in the binding surface at a single amino acid level, which confirmed the substantial incompatibility between the archaeal and eukaryotic proteins.

As the level of similarity between the archaeal and eukaryotic Box C/D enzyme is still unclear, the data presented in this work represent a considerable step towards understanding the similarity and differences between the enzymes from these two domains of life and how they are evolutionarily related. In view of the many human diseases associated with the Box C/D machinery, it is vital to understand how the eukaryotic enzyme differs from the archaeal counterpart to guide the development of targeted treatments.

## 5. Methods and Materials

*Section 5.1 – 5.8 have been partially taken from two manuscripts submitted for publication and have been entirely written by me.*

All used materials mentioned in section 5.1 to 5.8 are listed in section 5.9.

### 5.1 Cloning and Mutagenesis

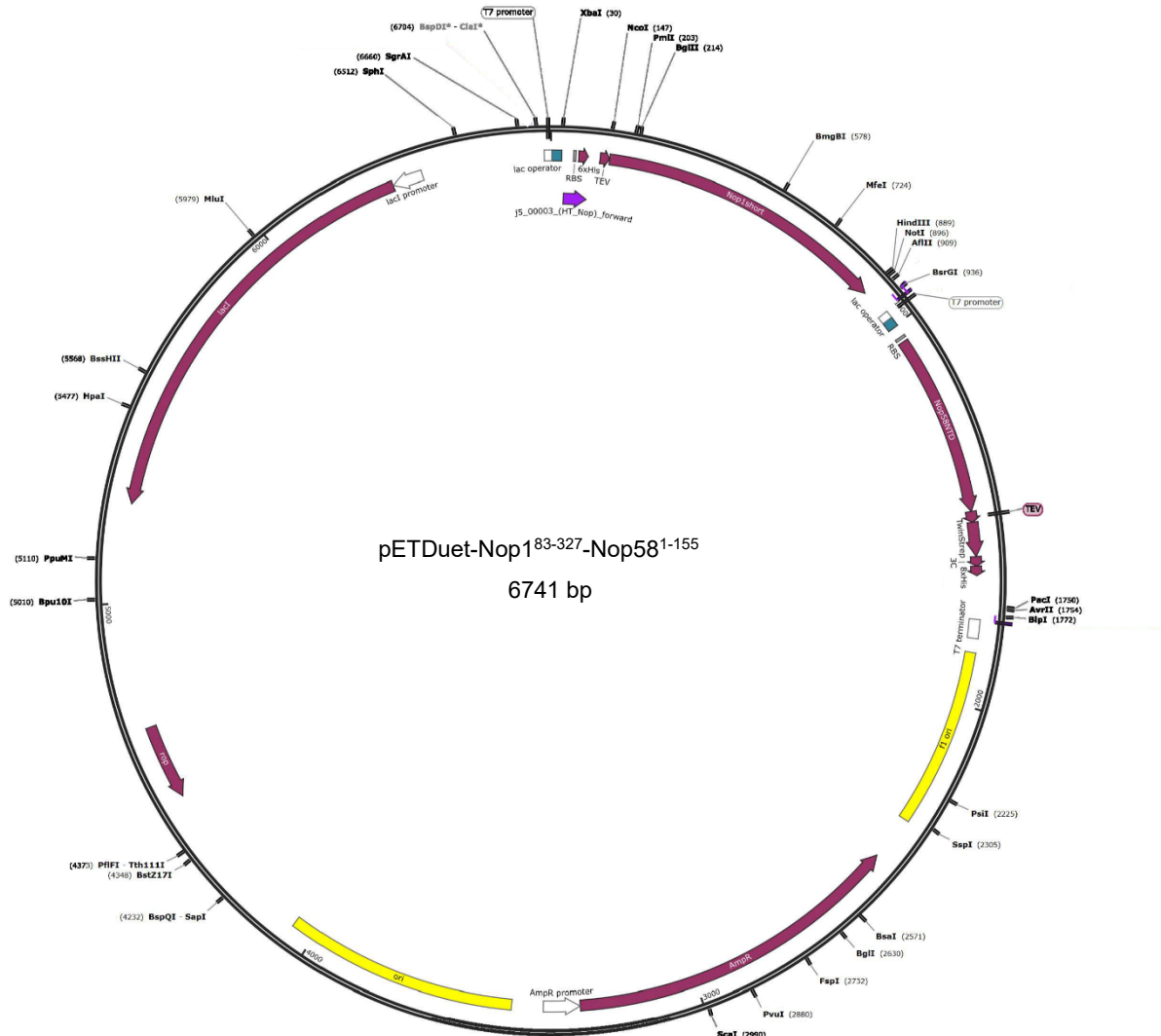
#### 5.1.1 Cloning

##### Proteins

Genes for full-length *Pyrococcus furiosus* (*P. furiosus*) 50S ribosomal protein L7Ae (UniProtKB accession number Q8U160, NCBI Gene-ID: WP\_014835393.1), Nop5 (Q8U4M1, WP\_011011172.1), and Fibrillarin (Q8U4M2, WP\_011011171.1) were obtained from genomic *P. furiosus* DNA by polymerase chain reaction (PCR) and cloned into the bacterial expression vector pET-M11 (EMBL collection). Nop5<sup>1-123</sup>, containing the N-terminal domain of Nop5, was amplified from the full-length gene by PCR and cloned as described. All archaeal gene constructs were cloned in such a way to integrate a TEV (tobacco etch virus) protease-cleavable N-terminal His<sub>6</sub>-Tag in the expression product. These plasmids had already been prepared at the start of this project and as in described in (Lapinaite et al., 2013).

Genes for *Saccharomyces cerevisiae* (*S. cerevisiae*) Snu13 (P39990, 856687), Nop1 (P15646, 851548), Nop56 (Q12460, 850894), Nop58 (Q12499, 854487), and Nop5 with the N-terminal domain of Nop56 (Nop5-56) were ordered as synthetic genes from Invitrogen (Thermo Fisher Scientific) and were codon-optimized for expression in *Escherichia coli*. The genes were amplified via PCR using DNA oligos (primers) containing the cleavage site for restriction enzymes NcoI-HF (5' CCATGG) and XhoI (5' CTCGAG) at the 5'- and the 3'-end, respectively. The amplified fragments were purified and cleaved with the above-mentioned restriction enzymes (both New England Biolabs). Cleaved inserts were ligated into the cleaved pET-M11 bacterial expression vector with T4 DNA ligase (New England Biolabs) retaining an N-terminal TEV protease-cleavable His<sub>6</sub>-Tag. Insertion and insert sequence were confirmed by DNA sequencing (Eurofins). Shorter constructs of Nop1 and Nop56 were also cloned into expression vector pET-M11 using the above-described protocol. These constructs correspond to Nop1<sup>83-327</sup> (from residue 83 to 327, excluding the N-terminal RGG-rich domain) and Nop56<sup>1-166</sup> (from residue 1 to 166, complete N-terminal domain). The N-terminal domain of Nop58,

ranging from residue 1 to 155, was cloned with a C-terminal TEV-cleavable TwinStrep-Tag and 3C protease cleavable His<sub>6</sub>-Tag into a pETDuet co-expression vector together with Nop1<sup>83-327</sup> with N-terminal TEV-cleavable His<sub>6</sub>-Tag (**Figure 5.1**). All primers used for cloning are listed in **Appendix 3**.



**Figure 5.1.** Plasmid map for Nop1<sup>83-327</sup> and Nop58<sup>1-155</sup> pETDuet co-expression vector

## RNA

DNA templates for full-length *S. cerevisiae* guide RNAs snR51 (NCBI Gene-ID: 9164983), snR41 (NCBI Gene-ID: 9164986), and snR54 (NCBI Gene-ID: 9164960) were ordered as synthetic genes from GENEWIZ (Sigma-Aldrich) in *E. coli* cloning vector pUC57. Each construct was designed and ordered with a 5' EcoRI cleavage site (5' GAATTC), a 3' HindIII cleavage site (5' AAGCTT), both for subcloning, and a PstI cleavage site (5' CTGCAG) directly upstream of the HindIII site for plasmid linearization. The pUC19-sR26 template plasmid has



already been prepared as described in (Graziadei et al., 2016). All other DNA templates were ordered as complementary single-stranded DNA oligos, also containing an EcoRI, HindIII, and PstI cleavage sites as described above. For subcloning, DNA oligos were annealed and ligated into EcoRI-HF and HindIII-HF cleaved pUC19 cloning vector using T4 DNA ligase (New England Biolabs). Insertion and insert sequence were verified by DNA sequencing (Eurofins). For each RNA construct, one guanosine was added at the 5' end, to optimize transcription yield by T7 RNA polymerase. At the 3' end, one additional cytosine was added as a result of the PstI cleavage site. These sites are indicated in bold in **Table 5.2**. All used DNA oligos are listed in **Appendix 3**.

### 5.1.2 Mutagenesis

#### Proteins

Point-mutations in the full-length genes of *S. cerevisiae* Snu13 and Nop56<sup>1-166</sup> were introduced using Pfu Plus polymerase (Roboklon) according to the manufacturer protocol. The primers used to generate each mutation are listed in **Table 5.1**. The following mutants were generated: S94E, R95V, S94ER95V, S94ER95VP96A, V93IR95V, R95VP96A, and R95VI98A for Snu13; L147VE148F for Nop56<sup>1-166</sup>. PCR products were freed from the starting DNA by specific digestion of methylated DNA with restriction enzyme DpnI (New England Biolabs). DpnI was inactivated by heat treatment (80 °C for 20 min) and the remaining DNA transformed into chemically competent *E. coli* OmniMax cells (produced in-house). Positive clones were verified by sequencing (Eurofins).

**Table 5.1.** Mutagenesis primers used for Snu13 and Nop56<sup>1-166</sup> point-mutations

| Primer ID                   | Sequence (5' → 3')                             |
|-----------------------------|--|
| S94E_Foward                 | 5' - GGGTCGTGCATGTGGTGTGAACGT                  |
| S94E_Reverse                | 5' - GCTGCAATAACCGGACGTTCAAC                   |
| R95V_Foward                 | 5' - GGGTCGTGCATGTGGTGTAGCGTGCCG               |
| R95V_Reverse                | 5' - GGTAATGCTTGCTGCAATAACCGGCACGCT            |
| S94ER95V_Foward             | 5' - ACTGGGTCGTGCATGTGGTGTGAAGTGCCG            |
| S96ER95V_Reverse            | 5' - GGTAATGCTTGCTGCAATAACCGGCACTTCAAC         |
| S94ER95VP96A_Foward         | 5' - ACTGGGTCGTGCATGTGGTGTGAAGTGCGGTT          |
| S94ER95VP96A_Reverse        | 5' - GTGGTAATGCTTGCTGCAATAACCGCCACTTCAAC       |
| V93IR95V_Foward             | 5' - CGTGCATGTGGTATTAGCGTGCCGTTATTGCAGC        |
| V93IR95V_Reverse            | 5' - GCTGCAATAACCGGCACGCTAATACCACATGCACGACC    |
| R95VP96A_Foward             | 5' - TGTGGTGTAGCGTGGCGGTTATTGCAGCAAGCATTACCACC |
| R95VP96A_Reverse            | 5' - GCTGCAATAACCGCCACGCTAACACCACATGCACGACC    |
| R95VI98A_Foward             | 5' - GTGGTGTAGCGTGCCGTTGCGGCAGCAAGCATTAC       |
| R95VI98A_Reverse            | 5' - CTTGCTGCCGCAACCGGCACGCTAACACCACATG        |
| Nop56NTD_L147WE148F_Foward  | 5' - CAGAGCGGTGATTGGTTTCGTGCACAGCTGGGTCT       |
| Nop56NTD_L147WE148F_Reverse | 5' - CCAGCTGTGCACGAAACCAATCACCGCTCTGCAGAC      |

## 5.2 Protein expression and purification

### 5.2.1 *P. furiosus* proteins

N-terminal His<sub>6</sub>-tagged L7Ae, Nop5, Nop5<sup>1-123</sup>, and Fibrillarin were expressed in *E. coli* expression strain BL21(DE3). Plasmids (described in section 5.1.1) were transformed into chemically competent *E. coli* BL21(DE3) cells and transformants selected with the kanamycin resistance encoded in the pET-M11 expression vector. Cells were grown in LB (lysogeny broth) culture medium (Carl Roth) with kanamycin at 37°C and 200rpm shaking for oxygenation until an OD 600 (optical density at 600 nm wavelength) of 0.6 -0.8; at this point the temperature was lowered to 20 °C. Expression was induced at 20°C with a final IPTG (isopropyl β-D-1-thiogalactopyranoside) concentration of 1 mM. The cells were harvested 18-20 hours after induction by centrifugation at 4500 rpm (revolutions per minute) for 45 min at 4 °C.

All cells were resuspended in buffer A (50 mM Tris-HCl, 1 M NaCl, 10% glycerol, 10 mM imidazole, 10 mM β-mercaptoethanol (BME), pH 7.5). For each pellet of 1-2 liters of LB medium one tablet of cComplete EDTA-free protease inhibitor cocktail (Roche) and 1mg of lysozyme (Carl Roth) was added and the cells were incubated for 30 min on ice. After incubation, cells were lysed by sonication on ice (50% amplitude, 5 seconds pulse, 10 seconds break). All lysates were cleared by centrifugation at 18500 rpm at 16 °C for 1 hour.

For His<sub>6</sub>-tagged L7Ae the supernatant after centrifugation was mixed 1:3 with buffer A containing 8 M guanidinium chloride, to reach a final concentration of 6 M guanidinium chloride and unfold all soluble proteins. This solution was loaded on a 5 ml HisTrap FF column (Cytiva) using an Äkta pure system with an additional sample pump. After loading the bound target protein was slowly refolded by reducing the guanidinium chloride concentration gradually to 0 M over the course of 40 column volumes (CV). The refolded target protein was eluted in a 0-50% gradient of buffer B (50 mM Tris-HCl, 1 M NaCl, 10% Glycerol, 1 M imidazole, 10 mM BME, pH 7.5). This protocol was adapted from (Ahmed et al., 2020).

For His<sub>6</sub>-tagged Nop5, Nop5<sup>1-123</sup> and Fibrillarin the supernatant was boiled for 15 min at 80 °C in a water bath after the first centrifugation and then centrifuged again for 1 hour at 16 °C and 18500 rpm. The final supernatant was loaded onto a 5 mL HisTrap FF column (Cytiva) using an Äkta pure system. The bound protein was washed six times with 3 CV of buffer C (50 mM Tris-HCl, 1 M NaCl, 10% glycerol, 10 mM imidazole, 2 M LiCl, 10 mM BME, pH 7.5) and eluted with a 0-50% gradient of buffer B (described above). This protocol was adapted from (Lapinaite et al., 2013).

After affinity purification all three proteins were buffer exchanged into buffer A using a HiPrep 26/10 desalting column (Cytiva). The N-terminal His<sub>6</sub>-Tag was removed by overnight TEV protease (produced in-house) cleavage under constant gentle shaking at room temperature.

Target protein and TEV protease were mixed in a 1:50 ratio for optimal results. Cleaved proteins were purified from TEV protease and cleaved N-terminal affinity tag by reverse affinity chromatography using a HisTrap FF column (Cytiva) with buffer A and buffer B (described above). All purification steps were performed at room temperature. All proteins were tested for remaining RNA or RNase contamination using the RNaseAlert™ Lab Kit from Thermo Fisher Scientific. Proteins were used for further experiments only when no contamination was detected after incubation of the protein sample with the test kit overnight. All RNaseAlert™ include a positive control (RNase A) and negative control (sterile water). The purity of all proteins was tested by SDS gel electrophoresis.

Final protein concentration was determined using UV absorption at 280 nm with molar extinction coefficients calculated with the online tool ProtParam on the ExPASy Server (Gasteiger et al., 2005). The following molar extinction coefficients ( $M^{-1}cm^{-1}$ ) were used; L7Ae: 4470, Nop5: 53860, Nop5<sup>1-123</sup>: 12950, Fibrillarin: 31400.

### 5.2.2 *S. cerevisiae* proteins

N-terminal His<sub>6</sub>-tagged Snu13 (including all mutants), Nop1<sup>83-327</sup>, Nop56<sup>1-166</sup>, Nop56<sup>1-166</sup>-WF, Nop5-56, and Nop1<sup>83-327</sup>-Nop58<sup>1-155</sup> (co-expression) were also expressed in *E. coli* BL21(DE3). Chemically competent *E. coli* cells were transformed with the expression plasmids (section 5.1.1 and 5.1.2) for each protein and grown in LB medium (with kanamycin or ampicillin) at 37 °C until an OD 600 of 0.6-0.8 was reached. Then, the temperature was reduced to 16 °C and expression was induced with a final IPTG concentration of 1 mM for Snu13 and Snu13 mutants and 0.5 mM for Nop1<sup>83-327</sup>, Nop56<sup>1-166</sup>, Nop5-56, and Nop1<sup>83-327</sup>-Nop58<sup>1-155</sup> expression. Cells were harvested in all cases 18-20 hours after induction by centrifugation at 4500 rpm and 4 °C for 45 min.

Cell pellets for Snu13 and Snu13 mutants were resuspended in buffer D (50 mM Tris-HCl, 1 M NaCl, 5% glycerol, 10 mM imidazole, 10 mM BME, pH 7.5); one tablet of cOmplete, EDTA-free protease inhibitor cocktail (Roche) and 1mg of the lysosome (Carl Roth®) was added and the mixture incubated for 30 min on ice. The cells were lysed by sonication on ice (50% amplitude, 5 seconds pulse, 10 seconds break) and the lysate was cleared by centrifugation at 18500 rpm and 4 °C for 1 hour. The supernatant was loaded and purified using a 5 ml HisTrap FF affinity column (Cytiva). The loaded protein was washed six times with 3 CV of buffer E (50 mM Tris-HCl, 1 M NaCl, 5% glycerol, 10 mM imidazole, 2 M LiCl, 10 mM BME, pH 7.5) and eluted using a 0-50% gradient of buffer F (50 mM Tris-HCl, 1 M NaCl, 5% glycerol, 1 M imidazole, 10 mM BME, pH 7.5). Eluted protein was buffer exchanged into buffer G (50 mM Tris-HCl, 150 mM NaCl, 5% glycerol, 10 mM BME, pH 8.0) using a HiLoad desalting 26/10 column (Cytiva). To remove bound RNA and RNase the protein was further cleaned using a 5 ml QTrap HP anion exchange column (Cytiva) where the free target-protein was collected in

the flow-through and RNA and RNA-bound protein was eluted with buffer H (50 mM Tris-HCl, 2 M NaCl, 5% glycerol, 10 mM BME, pH 8.0). The purified protein was cleaved with TEV protease (produced in-house, added in a 1: 50 ratio) to remove the N-terminal His<sub>6</sub>-tag overnight at 4 °C. TEV protease and the cleaved tag were removed by reverse affinity chromatography using a 5 ml HisTrap FF column (Cytiva) and buffer D and F. All purification steps were performed at 4 °C. Snu13 and Snu13 mutants were also tested for RNA and RNase contamination using the RNaseAlert™lab kit from Thermo Fisher Scientific as described in section 5.2.1.

Cells that expressed Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>, Nop56<sup>1-166</sup>-WF mutant, Nop5-56, and Nop1<sup>83-327</sup>-Nop58<sup>1-155</sup> were resuspended in buffer A (50 mM Tris-HCl, 1 M NaCl, 10% glycerol, 10 mM imidazole, 10 mM BME, pH 7.5). One tablet of cOmplete EDTA-free protease inhibitor cocktail (Roche) and 1 mg of lysozyme (Carl Roth) was added per cell pellet, and the mixture was incubated for 30 min on ice. Afterwards, cells were lysed by sonication (50% amplitude, 5 seconds pulse, 10 seconds break) for 30 min on ice. The lysates were cleared by centrifugation at 18500 rpm for 1 hour at 4 °C. The supernatant was purified by affinity chromatography using a 5 ml HisTrap FF column (Cytiva). Once loaded, the bound proteins were washed three times with 3 CV of buffer C (50 mM Tris-HCl, 1 M NaCl, 10% glycerol, 10 mM imidazole, 2 M LiCl, 10 mM BME, pH 7.5) and eluted with a 0-50% gradient of buffer B (50 mM Tris-HCl, 1 M NaCl, 10% glycerol, 1 M imidazole, 10 mM BME, pH 7.5). The purified protein was buffer-exchanged into buffer A using a HiPrep 16/10 desalting column (Cytiva) and incubated overnight at 4 °C with TEV protease (1:50) to remove the N-terminal His<sub>6</sub>-tag. The TEV protease and cleaved His<sub>6</sub>-tag were purified from the proteins by reverse affinity purification (5 ml HisTrap FF column). The purified proteins were collected in the flow-through. The purity of all proteins was tested by SDS gel-electrophoresis.

The final protein concentration was determined using UV absorption at 280 nm with molar extinction coefficients calculated with the online tool ProtParam on the ExPASy Server (Gasteiger et al., 2005). The following molar extinction coefficients (M<sup>-1</sup>cm<sup>-1</sup>) were used; Snu13 and Snu13 mutant: 3105, Nop1<sup>83-327</sup>: 17420, Nop56<sup>1-166</sup>: 5960, Nop56<sup>1-166</sup>-WF: 11460, Nop5-56: 46870 and Nop1<sup>83-327</sup>-Nop58<sup>1-155</sup>: 40465.

### **5.3 RNA *in vitro* transcription and purification**

All RNAs used for crystallization and mass spectrometric analysis and the guide RNAs used in activity assays were produced in-house using *in vitro* transcription. Linearized plasmid DNA was used as a template for all transcription reactions. DNA template containing plasmids (see

section 3.1.1) were transformed into *E. coli* Top10 and transformed cells were grown in LB medium overnight at 37 °C with 200 rpm shaking and harvested by centrifugation at 4500 rpm and 4 °C for 45 min. Plasmids were extracted using the Qiagen Plasmid Mega Kit (Qiagen). Extracted Plasmid DNA was linearized by cleavage with PstI (New England Biolabs). Linearized plasmid DNA was purified by means of phenol/chloroform/isoamyl alcohol (Carl Roth) and chloroform/isoamyl alcohol (Carl Roth) extraction and concentrated by precipitating the DNA with pure ethanol and NaCl. Purified and linearized template DNA was stored at 1 mg/ml in pure water at -20 °C.

RNAs were synthesized by *in vitro* transcription using T7 RNA polymerase (produced in-house) and linearized DNA templates. All transcription reactions were performed using 10X transcription buffer (400 mM Tris-HCl, 10 mM spermidine, 50 mM DTT, 0.1% (v/v) Triton X-100, pH 8.0). Transcription conditions were optimized in 20 µl reactions for each construct and every new batch of NTPs (nucleotide triphosphate mixture). Optimized variables were MgCl<sub>2</sub> concentration (20–50mM), NTP concentration (10–40mM), DNA template concentration (0.025 – 0.2 mg/ml) and T7 polymerase concentration (0.025–0.1 mg/ml). The 100 mM NTP stock solution was prepared from ATP, CTP, GTP and UTP powder (Carl Roth) solubilized in water at pH 7.0 and mixed 1:1:1:1. Optimization reactions were incubated for three hours at 37 °C and stopped by the addition of 2x denaturing RNA loading-dye (95% formamide, 2.5% (w/v) bromophenol blue, 2.5% (w/v) xylene cyanol, 0.25 mM EDTA, pH 8.0). Optimization results were checked on a small analytical denaturing polyacrylamide gel (12-20% acrylamide, 8 M urea, 1x TBE buffer pH 8.3) using a 1x TBE buffer as running buffer at 15 W power. The RNA was stained in a 1x TBE buffer solution with ethidium bromide for 10 min at room temperature. For visualization, a Gel Doc XR+ gel documentation system (BioRad) was used. The condition with the best yield was subsequently used for large-scale transcription reactions.

Large-scale transcription reactions (2–10ml) were set up for five hours at 37 °C and stopped by the addition of 0.1 reaction volumes of 0.5 M EDTA solution. Before further purification steps, the reaction mixture was concentrated using a vacuum concentrator at 45 °C. After concentration, the RNA and remaining T7 RNA polymerase were denatured through addition of 2x de-naturing loading-dye (95% formamide, 2.5% (w/v) bromophenol blue, 2.5% (w/v) xylene cyanol, 0.25 mM EDTA, pH 8.0) was added. RNAs were then further purified using preparative denaturing polyacrylamide gels (12–20% acrylamide, 8 M urea, 1x TBE buffer pH 8.3) and 1x TBE running buffer at 15–20 W power at room temperature. The acrylamide content was adjusted to the size of the RNA to be purified (20%: 10–30 nt; 15% 30–100 nt; 12%: 100–150nt). In the preparative gel the RNA was visualized with UV shadowing and the gel was cut out in slices. The gel slices were crushed using sterile plastic syringes and the pieces soaked in gel extraction buffer (40 mM MES, 10 mM EDTA, pH 6.0) and frozen at -80°C

for 20 min (“crush-and-soak” method) (Petrov et al., 2013). After freezing the gel pieces were soaked in the extraction buffer overnight at 4 °C under constant shaking. Gel pieces were removed by centrifugation at 4000 rpm and 4 °C for 45 min and the supernatant decanted carefully. The supernatant, containing the purified RNA, was filtered (0.22 µm syringe filter) and the nucleic acid precipitated by the addition of 2.3 volumes of pure ethanol and 0.1 volumes of 5 M NaCl and incubation at –20 °C for at least 2 hours. After precipitation, the pure RNA was pelleted by centrifugation at 8500 rpm for 1.5 hours, the supernatant was discarded, and the pellet was once washed with 70% pure ethanol, dried at 37 °C and resuspended in pure water. All purified RNAs were stored at –20°C. Final purity was checked using analytical denaturing polyacrylamide gels. This protocol was loosely adapted from (Nilsen, 2013). Concentrations were determined using a spectrophotometer (Nanodrop), applying an extinction coefficient of 0.025 (µg/ml)<sup>-1</sup>cm<sup>-1</sup> for single-stranded RNA at 260 nm wavelength (Green et al., 2012). All transcribed RNAs are listed in **Table 5.2**.

**Table 5.2.** RNA constructs used in this study. Nucleotides at the 5' and 3' end (indicated in bold) do not belong to the native sequence and are the result of cloning and restriction enzyme cleavage sites.

| <b>guide RNA</b> | <b>Sequence (5' → 3')</b>  |
|------------------|--|
| sR26             | 5'- <b>GCGAGCAAUGAUGAGUGAUGGGCGAACUGAAAUGAUGAUGA</b><br><b>CGGAGGUGAUCUCUGAGCUCGC</b>  |
| snR51            | 5' - <b>GGUUUAUGAUGAUUUUUUUUAUUUCACACUGUACUAGAUUGG</b><br>UCUCUUUAACGAAGGGGCUAUUUGAUGACUACAAAUAUUUUUUUAACUGAUUUAAUGACUCUGAAA<br><b>C</b>     |
| snR51-2          | 5' - <b>GGUUUAUGAUGAUUUUUUUUAUUUCACACUGUACGAGAUUGG</b><br>UCUCUUUAACGAAGGGGCUAUUUGAUGACUACAAAUAUUUUUUUAACUGAUUUAAUGACUCUGAAA<br><b>C</b>     |
| snR41            | 5' - <b>GGCCCCUAUGAUUAAAAUUUUUAUUAUCAUUAACCAUUGUU</b><br>CCGACUGAAUAGUGGUUUUAACUACAUGUCGACAACCCUUUUUCGUUAAGUUUCAGCCUUGUAUGA<br><b>GGGGCC</b> |
| snR41 -2         | 5' - <b>GGCCCCUAUGAUUACCAUUGUUCGACUGAAAUGAUGAUG</b><br>UAAGUUUCAGCCUUGUAUGAGGGGGCC   |
| snR41 -3         | 5' - <b>GGCCCCUAUGAUUACCAUUGUUCGACUGAAUAGUGGUUU</b><br>AAUACAUGUGUUUCAGCCUUGUAUGAGGGGGCC   |
| snR41 -4         | 5' - <b>GGCCCCUAUGAUUACCAUUGUUCGACUGAAAUGAUGAUG</b><br>UGUUUCAGCCUUGUAUGAGGGGGCC   |
| snR54            | 5' - <b>GGAAGAUGAUGAUCAACUUUUUAUCAUAACUUUCGUUCU</b><br>ACUGACUGUGAUCAAACGAUCUUGUAGAGAACUUUUACUCUGAAUC  |
| snR51-kl1-S      | 5' - <b>GUACUAGAUUGGUCGAAAGGCUAUUUGAUGAC</b>   |
| snR51-kl2-S      | 5' - <b>GUACGAGAUUGGUCGAAAGGCUAUUUGAUGAC</b>   |
| sR26-kl          | 5' - <b>GACUGAAAUGAUGAUGAC</b>   |

## 5.4 Reconstitution of chimeric Box C/D s/snoRNP complexes

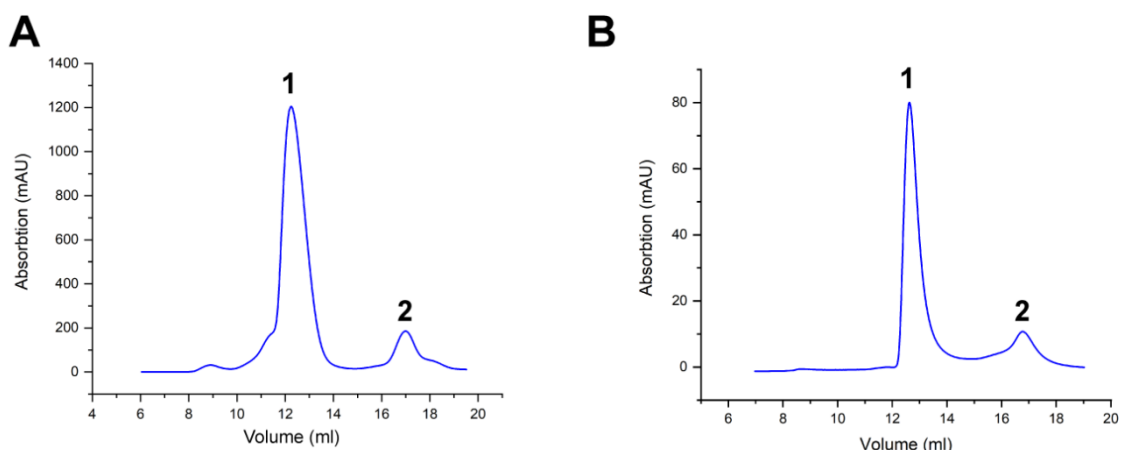
In the following paragraphs, “chimeric” indicates RNP complexes assembled *in vitro* from both eukaryotic and archaeal components.

All s/snoRNA complexes were assembled in complex buffer (20 mM sodium phosphate, 500 mM NaCl, pH 6.6) using a Superdex S200 Increase 10/300 GL (Cytiva) and an Äkta pure system. Purified Nop5 and Fibrillarin were mixed in a 1:1.1 ratio, incubated for 15 min at 80 °C and slowly cooled down to room temperature. The Nop5–Fibrillarin complex was purified via gel filtration (**Figure 5.2 A**). Purified Nop1<sup>83-327</sup> was mixed with Nop5-56 in a 1:1.2 ratio and incubated at RT for 15 min before being purified via gel-filtration (**Figure 5.2 B**).

The reconstitution protocol of complexes containing only archaeal proteins L7Ae, Nop5, and Fibrillarin was adapted from (Graziadei et al., 2016, 2020; Lapinaite et al., 2013).

For RNP complexes consisting of eukaryotic Snu13 and archaeal Nop5 or Nop5-56 and Fibrillarin, the guide RNA was annealed in annealing buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.5) by incubation at 90 °C for 2 min and snap-cooling on ice. The Nop5–Fibrillarin or Nop5-56–Nop1<sup>83-327</sup> complex, Snu13, and the guide RNA were mixed in a 2:2:1 ratio and incubated for 15 min before purification using size-exclusion chromatography.

All purification steps were carried out at room temperature. For all size-exclusion runs, UV traces at 280, 260, and 215 nm were recorded to monitor the presence of RNA and proteins in the elution peak. The presence of all protein components was verified by SDS gel electrophoresis.



**Figure 5.2.** Size-exclusion chromatography elution profile at 280 nm detection (blue line) of Nop5–Fibrillarin complex (**A**) and Nop1<sup>83-327</sup>–Nop5-56 complex (**B**) on a Superdex S200 Increase 10/300 in complex buffer. The peak in both chromatograms (**A**) and (**B**) labeled with **1** represents the respective protein complex, whereas the peak labeled **2** represents the excess of free Fibrillarin or Nop1<sup>83-327</sup>, respectively.

## 5.5 Electrophoretic mobility shift assays (EMSA)

All RNAs used for the electrophoretic mobility shift assays (EMSA) were purchased from Integrated DNA Technologies (IDT) purified by HPLC purified. Purchased RNAs contain a 5'-Cy5<sup>TM</sup> label used for detection. A list of all purchased and used 5' Cy5<sup>TM</sup>-labeled RNAs can be found in **Table 5.3**.

For each assay, 10 pmol of RNA were mixed with pure and sterile LC-MS grade water (Merck) and annealing buffer (final concentrations: 10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.5) in a total volume of 5  $\mu$ l and annealed by heating-up the sample to 80 °C for 1 min and cooling it down in 1.5 °C steps to 4 °C using a T100 Thermo Cycler (BioRad). After annealing, 0 pmol, 2.5 pmol, 5 pmol, 10 pmol, and 20 pmol of the protein (Snu13, Snu13 mutants, and L7Ae) were added and incubated for up to 30 min at 4 °C. After incubation 5x native loading dye (50 mM Tris-HCl, 0.25% xylene cyanol, 0.25% bromophenol blue, 30% glycerol, pH 7.5) was added to each sample. This protocol was loosely adapted from (Ruscher et al., 2000).

All samples were analyzed on 10% native polyacrylamide gel at 4 °C to prevent diffusion and degradation of the protein and the RNA. Each gel was prepared with a 40% acrylamide solution (Carl Roth), TBE buffer and sterile pure water and pre-run for 30 min to 1 hour before sample loading at 4 °C and 5 mA current. The complete sample volume was loaded to ensure the same amounts of labeled RNA in each sample. After sample loading gels were run overnight at 4°C with 10mA electrical current.

To ensure that the additional nucleotides added to the 5' end of each RNA as a linker did not affect the folding of the RNA, a subset of assays (L7Ae or Snu13 with sR26-kl, snR51-kl1, snR51-kl2, snR41-kl1, snR41-kl2, snR54-kl1, and snR54-kl2) was also conducted with non-labeled RNAs lacking the three additional nucleotides. For each of these assays, 0.5 nmol of non-labeled RNA were treated as described above and then mixed with 0 nmol, 0.125 nmol, 0.25 nmol, 0.5 nmol, and 1 nmol of protein (L7Ae or Snu13), and the mixture was incubated for 30 min at 4 °C. All samples were further treated as described above. Gels were stained with ethidium bromide, and the RNA was visualized using a Gel Doc XR+ gel documentation system (BioRad).

Gels with Cy5-labeled RNAs were analyzed using a Typhoon Trio system (GE Healthcare) applying a 670 nm-bandpass (670 BP 30) emission filter. Intensities were extracted using Fiji (Schindelin et al., 2012). From extracted intensities, the bound RNA fraction  $f$  was calculated according to **equation 5.1**, where  $[RP]$  is the amount of RNA bound to the protein and  $[R]$  is the amount of free RNA. The bound RNA fraction  $f$  was then plotted against the total protein concentration  $[P]_T$  and the curve was fitted to **equation 5.2**,



with  $A_1$  being the initial  $f$  value,  $A_2$  the final  $f$  value,  $f_{1/2}$  the  $[P]_T$  at which  $f = (A_2+A_1)/2$  (corresponding to the  $K_D$ ) and  $p$  the Hill coefficient. The data was fitted using OriginPro.

$$f = \frac{[RP]}{[R]+[RP]} \quad \text{Eq. (5.1)}$$

$$f = A_2 + \frac{(A_1-A_2)}{1+\left(\frac{[P]_T}{f_{1/2}}\right)^p} \quad \text{Eq. (5.2)}$$

**Table 5.3.** RNA sequences for Cy5-labeled RNAs ordered from Integrated DNA Technologies (IDT). Underlined nucleotides were added to the desired sequences as a linker to the fluorescent Cy5 label.

| RNA       | Sequence (5' → 3')                                      |
|-----------|---|
| sR26-kl   | Cy5 - <u>AUA</u> ACUGAAAUAGAUGAUGA                      |
| snR51-kl1 | Cy5 - <u>AUA</u> ACUAGAUUGGUCUCUUUCAAGAAGGGGCUAAUUGAUGA |
| snR51-kl2 | Cy5 - <u>AUA</u> ACGAGAUUGGUCUCUUUCAAGAAGGGGCUAAUUGAUGA |
| snR51-kl3 | Cy5 - <u>AUA</u> ACUAAAUAGUUGAUGA                       |
| snR51-kl4 | Cy5 - <u>AUA</u> ACGAAAUAGUUGAUGA                       |
| snR41-kl1 | Cy5 - <u>UAU</u> ACUGAAUAGUGGUUUAACUACAUGU              |
| snR41-kl2 | Cy5 - <u>UAU</u> ACUGAAUAGUGGUUUAACUAGAUGU              |
| snR41-kl3 | Cy5 - <u>UAU</u> ACUGAAAUAGUACAUGU                      |
| snR54-kl1 | Cy5 - <u>AUA</u> ACUGACUGUGAUCAAACGAUCUUGUAGA           |
| snR54-kl2 | Cy5 - <u>AUA</u> ACUGACUGUGAUCAAACGAUCUUGAAGA           |
| snR54-kl3 | Cy5 - <u>AUA</u> ACUGAAAUAGUUGUAGA                      |

## 5.6 Size-exclusion Chromatography and Multi-angle Light Scattering

Protein interactions between Nop1<sup>83-327</sup>, Nop56<sup>1-166</sup>, Nop56<sup>1-166</sup>-L147WE148F and archaeal Fibrillarin and Nop5<sup>1-123</sup> were detected using size-exclusion chromatography. Proteins were mixed in a 1:1 ratio in the following combination: Nop1<sup>83-327</sup>-Nop56<sup>1-166</sup>, Nop1<sup>83-327</sup> – Nop56<sup>1-166</sup>-L147WE148F, Nop1<sup>83-327</sup>-Nop5<sup>1-123</sup>, Fibrillarin-Nop56<sup>1-166</sup>, Fibrillarin-Nop56<sup>1-166</sup>-L147WE148F. The control experiment for the interaction of archaeal Fibrillarin and Nop5 can be found in Figure 5.2 in section 5.4. Each protein was analyzed in isolation as a control. Each protein mixture was incubated for 15 min at room temperature and the formation of a protein-protein complex checked by size-exclusion chromatography. For each analysis, an Äkta pure system (GE Healthcare) and a Superdex 200 Increase 10/300 GL size-exclusion chromatography column (Cytiva) were used. The same running buffer was used for all systems for comparability (50 mM sodium phosphate, 100 mM NaCl, 10 mM BME, pH 7.0). The results were monitored by UV light absorption at 280 nm.

To visualize the proteins, present in the elution peaks, each peak was analyzed by SDS gel-electrophoresis using 4-20% Mini-PROTEAN® TGX™ Precast Protein gels (BioRad) with a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BioRad). All gels were stained with Pierce™ PageBlue™ protein stain (Thermo Fisher Scientific) according to the manufacturer's protocol.

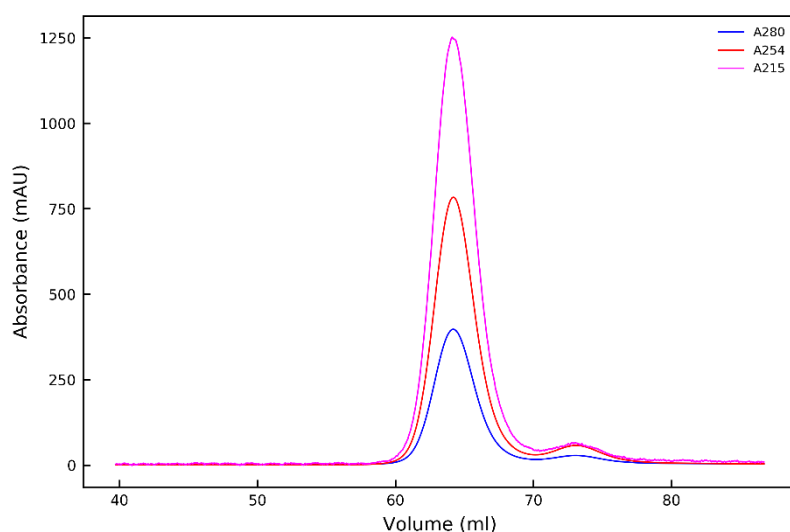
Multi-angle-light scattering (MALS) data were collected using an on-line SEC-MALS set-up, consisting of a Superdex S200 Increase 10/300 GL (Cytiva) column on an Äkta pure system coupled to a MALS miniDawn TREOS system (Wyatt Technologies) and Optilab T-rex refractive index detector (Wyatt Technologies). Data were analyzed using the ASTRA 7.0 software package (Wyatt Technologies).

## 5.7 X-Ray Crystallography

### 5.7.1 Crystallization

#### L7Ae and snR51-kl1-S

Purified archaeal L7Ae and snR51-kl1-S RNA were mixed in a 1:1 ratio, incubated for 15 min at 80 °C and afterward slowly cooled down to room temperature. The protein-RNA complex was purified in crystallization buffer (50 mM MOPS, 50 mM NaCl, 1 mM DTT, pH 6.6) using a HiLoad 16/600 Superdex 75 pg column (Cytiva) on an Äkta pure system at room temperature. The elution profile is shown in **Figure 5.3**. The purified complex was concentrated using Amicon® Ultra-15 3K centrifugal filters (Merck) with 4500 rpm at 4 °C.

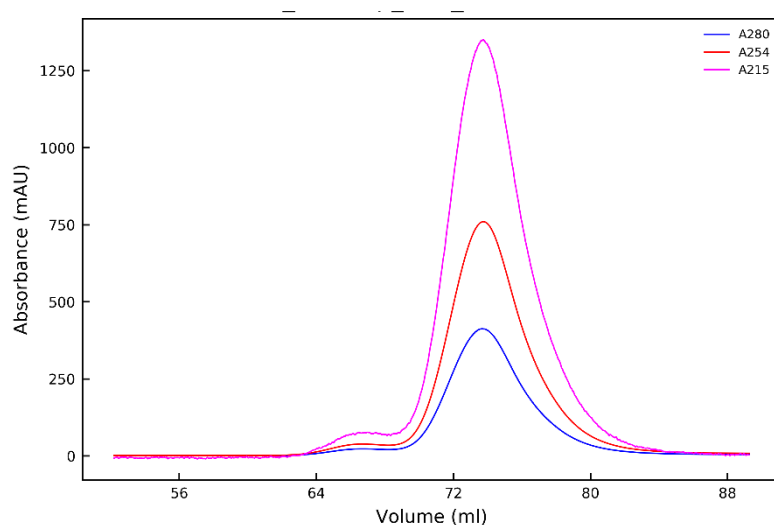


**Figure 5.3.** Size-exclusion profile of the complex between archaeal L7Ae and eukaryotic snR51-kl1-S RNA obtained with a HiPrep 16/600 Superdex 75 pg column (Cytiva). A280 (blue), A254 (red) and A215 (pink) are the absorptions at 280 nm, 254 nm and 215 nm, respectively.

A concentrated solution of ~ 10 mg/ml of L7Ae–snR51-kl1-S in the crystallization buffer was used for crystallization via sitting drop vapor diffusion. Initial crystallization screens were set up with a Crystal Phoenix crystallization robot (Art Robbins Instruments) using NeXtal DWBlock Suites (Qiagen); JCSG Core I Suite, JCSG Core II Suite, JCSG Core II Suite, JCSG Core IV Suite, Nucleix Suite, PEG Suite, and PEG II Suite. The drop was mixed from 0.25 µl of the sample solution and 0.25 µl of reservoir solution equilibrated against 50 µl of reservoir solution at 18 °C. Crystals appeared after one week in several conditions across all initial screens. The best condition corresponded to 0.02 M CaCl<sub>2</sub>, 0.1 M sodium acetate, 30% 2-Methyl-2,4-pentandiol (G12) from Qiagen JCSG Core I Suite. Cryo-protection was achieved by the addition of 10% (2*R*, 3*R*)-2,3-butanediol before flash-freezing.

### L7Ae and sR26-kl

Purified archaeal L7Ae and sR26-kl RNA were mixed in a 1:1 ratio, incubated for 15 min at 80 °C and afterward slowly cooled down to room temperature. The protein–RNA complex was purified in crystallization buffer (50 mM MOPS, 50 mM NaCl, 1 mM DTT, pH 6.6) using a HiLoad 16/600 Superdex 75 pg column (Cytiva) on an Äkta pure system at room temperature (**Figure 5.4**). The purified complex was concentrated using Amicon® Ultra-15 3K centrifugal filters (Merck) at 4500 rpm and 4 °C.

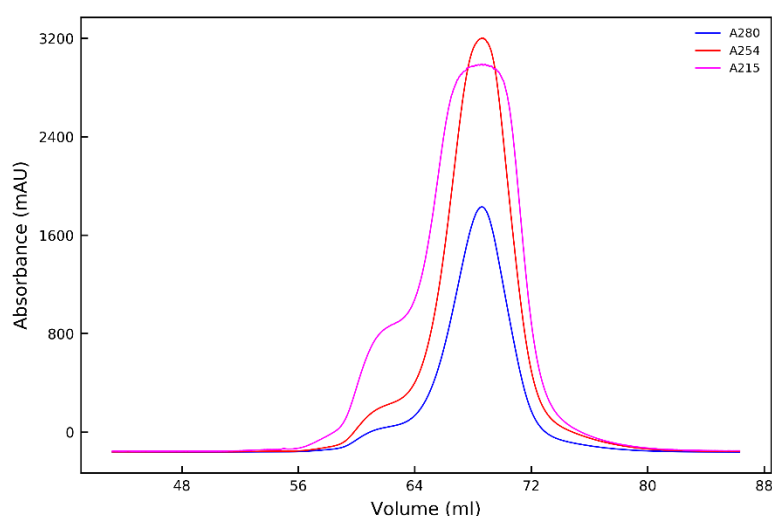


**Figure 5.4.** Size-exclusion profile of the complex between archaeal L7Ae and archaeal sR26-kl RNA obtained with a HiPrep 16/600 Superdex 75 pg column (Cytiva). A280 (blue), A254 (red) and A215 (pink) are the absorptions at 280 nm, 254 nm and 215 nm, respectively.

A concentrated solution of ~ 7 mg/ml of L7Ae–sR26-kl in the crystallization buffer was used for crystallization via sitting drop vapor diffusion. Initial crystallization screens were set up with a Crystal Phoenix crystallization robot (Art Robbins Instruments) using NeXtal DWBlock Suites (Qiagen); JCSG Core I Suite, JCSG Core II Suite, JCSG Core II Suite, JCSG Core IV Suite, Nucleix Suite, PEG Suite, and PEG II Suite. The drop was mixed from 0.25 µl of the sample solution and 0.25 µl of reservoir solution equilibrated against 50 µl of reservoir solution at 18 °C. After one month of incubation and step-up of screens with new samples preps no crystal formation could be observed.

### Snu13 and snR51-kl2-S

Eukaryotic RNA snR51-kl2-S was first annealed by heating the RNA sample to 80 °C followed by slow cooling to 4 °C. Afterwards Snu13 was added to the RNA in a 1:1 ratio and incubated for 30 min at 4 °C. The protein–RNA complex was purified in crystallization buffer (50 mM MOPS, 50 mM NaCl, 1 mM DTT, pH 6.6) using a HiLoad 16/600 Superdex 75 pg column (Cytiva) on an Äkta pure system at 4 °C (**Figure 5.5**). The purified complex was concentrated using Amicon® Ultra-15 3K centrifugal filters (Merck) at 4500rpm and 4 °C.

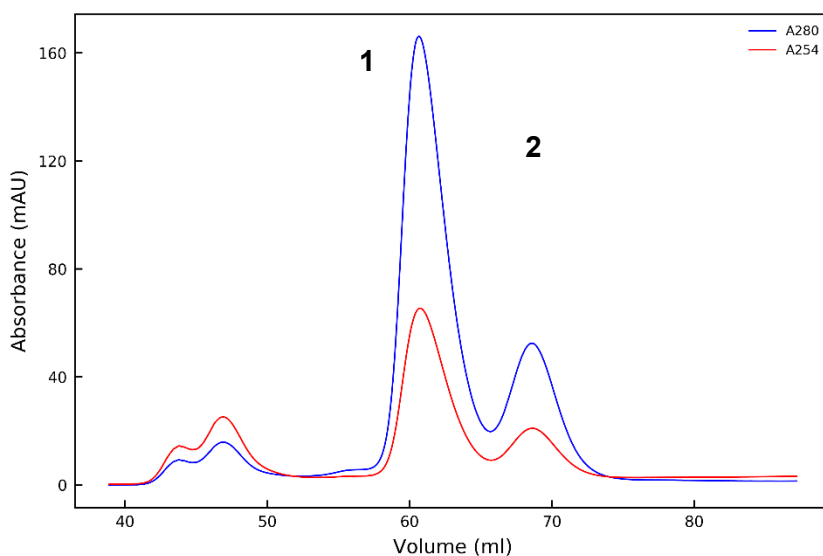


**Figure 5.5.** Size-exclusion profile of the complex between eukaryotic Snu13 and snR51-kl2-S RNA obtained with a HiPrep 16/600 Superdex 75 pg column (Cytiva). A280 (blue), A254 (red) and A215 (pink) are the absorptions at 280 nm, 254 nm and 215 nm, respectively.

A concentrated solution of ~ 7 mg/ml of Snu13–sn51-kl2-S in the crystallization buffer was used for crystallization via sitting drop vapor diffusion. Initial crystallization screens were set up with a Crystal Phoenix crystallization robot (Art Robbins Instruments) using NeXtal DWBlock Suites (Qiagen); JCSG Core I Suite, JCSG Core II Suite, JCSG Core II Suite, JCSG Core IV Suite, Nucleix Suite, PEG Suite, and PEG II Suite. The drop was mixed from 0.25 µl of the sample solution and 0.25 µl of reservoir solution equilibrated against 50 µl of reservoir solution at 12 °C. After two weeks of incubation and step-up of screens, crystal formation was observed in JCSG Suite II condition D10 and E10, JCSG Suite III condition C5 and Nucleix condition D9 and H6. Unfortunately, the crystals contained the protein only.

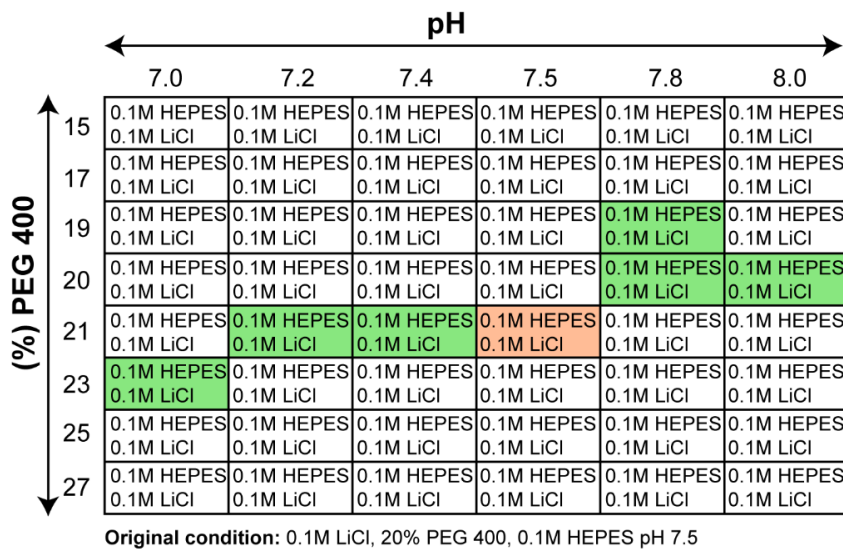
### Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>

Purified Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup> were mixed in a 1:1 ratio and incubated for 15 min at 4 °C. The protein-protein complex was purified a HiLoad 16/600 Supdedex S75 pg column (Cytiva) in protein crystallization buffer (50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 7.5). The elution profile is shown in **Figure 5.6**. The purified complex was concentrated using Amicon® Ultra-15 10K centrifugal filters (Merck) at 4500 rpm and 4 °C.

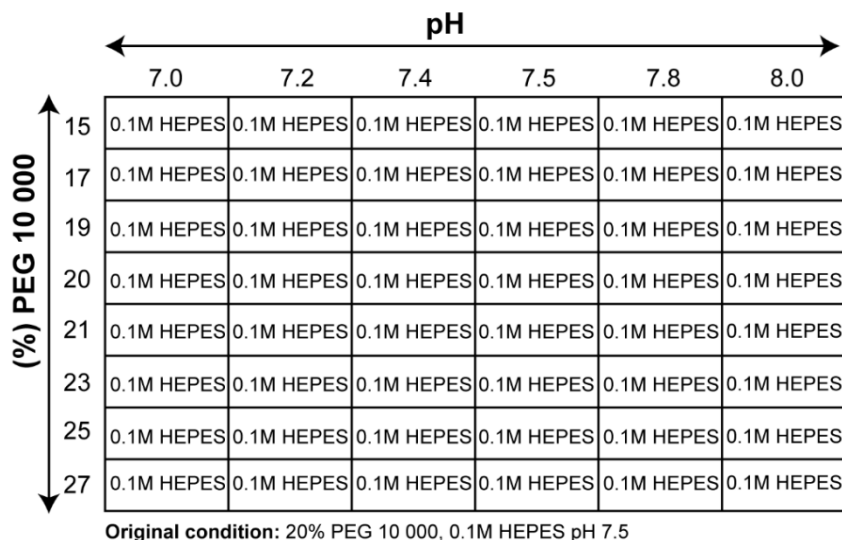


**Figure 5.6.** Size-exclusion profile of the complex between Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup> (peak 1) obtained with a HiPrep 16/600 Superdex 75 pg column (Cytiva). Peak 2 shows excess Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup> that did not form the complex. A280 (blue) and A254 (red) are the absorptions at 280 nm and 254 nm, respectively.

A concentrated solution of ~ 10 mg/ml of Nop56<sup>1-166</sup>-Nop1<sup>83-327</sup> in protein crystallization buffer was used for crystallization using sitting drop vapor diffusion. Initial crystallization screens were set up with a Crystal Phoenix crystallization robot (Art Robbins Instruments) using NeXtal DWBlock Suites (Qiagen); JCSG Core I Suite, JCSG Core II Suite, JCSG Core II Suite, JCSG Core IV Suite, Protein Complex Suite, Nucleix Suite, PEG Suite, and PEG II Suite. The drop was mixed from 0.25 µl of protein solution and 0.25 µl of reservoir solution equilibrated against 50 µl of reservoir solution at 12 °C. Crystals appeared in the Protein Complex Suite screen as well as in the PEG Suite screen after one week. Grid screen were set up based on the two best hit conditions and incubated at 12 °C (**Figure 5.7 and 5.8**). After one week crystals were obtained in multiple conditions. The best condition was 0.1 M HEPES (pH 7.5), 0.1 M LiCl, and 40% PEG400 based on a condition of the NeXtal DWBlock Protein Complex Suite. Cryo-protection was achieved by the addition of 10% (2*R*, 3*R*)-2,3-butanediol before flash-freezing.



**Figure 5.7.** Crystallization fine screen design based on the original condition A3 from the NeXtal DWBlock Protein Complex Suite Screen. Green conditions produced crystals and the orange condition produced the crystal for structure determination.

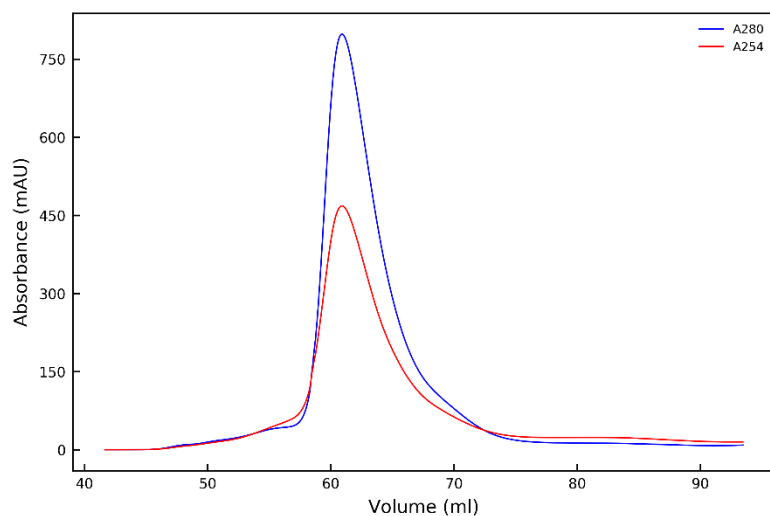


**Figure 5.8.** Crystallization fine screen design based on the original condition D5 from the NeXtal DWBlock PEG Suite Screen. None of the conditions produced crystals.

## Nop1<sup>83-327</sup> and Nop58<sup>1-155</sup>

Nop1<sup>83-327</sup> and Nop58<sup>1-155</sup> were co-expressed and co-purified as described in section 5.2. The co-purified proteins were concentrated using an Amicon® Ultra-15 10K centrifugal filters (Merck) at 4500 rpm and 4 °C; the protein-protein complex was isolated using a HiLoad 16/600 Superdex S75 pg size-exclusion chromatography column (Cytiva) in protein crystallization buffer (50 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 7.5). The isolated protein complex (**Figure 5.9**) was again concentrated using the above-mentioned concentrating system.

A concentrated solution of ~ 8 mg/ml of Nop58<sup>1-155</sup>-Nop1<sup>83-327</sup> in protein crystallization buffer was used for crystallization using sitting drop vapor diffusion. Initial crystallization screens were set up with a Crystal Phoenix crystallization robot (Art Robbins Instruments) using NeXtal DWBlock Suites (Qiagen); JCSG Core I Suite, JCSG Core II Suite, JCSG Core II Suite, JCSG Core IV Suite, Protein Complex Suite, Nucleix Suite, PEG Suite, and PEG II Suite. The drop was mixed from 0.25 µl of protein solution and 0.25 µl of reservoir solution equilibrated against 50 µl of reservoir solution at 12 °C. After incubation up to 3 month no crystals appeared in any of the tested conditions.



**Figure 5.9.** Size-exclusion profile of the complex between Nop1<sup>83-327</sup> and Nop58<sup>1-155</sup>, obtained with a HiPrep 16/600 Superdex 75 pg column (Cytiva). A280 (blue) and A254 (red) are the absorptions at 280 nm and 254 nm, respectively.

### 5.7.2 Data collection and processing

The datasets were collected at PETRA III, beamline P11 of DESY (Deutsches Elektronen-Synchrotron, Hamburg, Germany). The datasets were recorded at 100K and a wavelength of 1.0 Å using a Pilatus 6M detector (Dectris). The data were processed using the AutoPROC (Vonrhein et al., 2011) toolbox (Global Phasing) executing XDS (Kabsch, 2010), Pointless (Evans, 2006), Aimless (Evans and Murshudov, 2013) from the CCP4 software suit (Winn et al., 2011).

### 5.7.3 Structure determination

#### L7Ae and snR51-k11 short

The crystal structure of L7Ae–snR51-k11-S was solved by molecular replacement. To determine the number of molecules in the asymmetric unit Xtriage (Zwart et al.) from the Phenix software package (Adams et al., 2010) was used. For archaeal L7Ae, L7Ae from *P. furiosus* (PDB-ID: 4WB0, sequence identity: 100%) was found as a working search model executing Balbes (Long et al., 2008) from the CCP4 suite (Winn et al., 2011). The initial structure of the protein was first built and refined using AutoBuild (Terwilliger et al., 2008) from the Phenix software package (Adams et al., 2010). Afterwards, the missing RNA was built using AutoBuild (Terwilliger et al., 2008) from the Phenix software package (Adams et al., 2010) around the fixed model of the protein. The models were analyzed and completed in iterative cycles with Coot (Emsley et al., 2010); the refinement was done using Phenix-refine (Afonine et al., 2012). Data collection and refinement statistics are shown in **Table 5.4**.



**Table 5.4.** Crystallographic statistics for archaeal L7Ae bound to eukaryotic snR51-kl1-S.

| <b>Structure</b>  | L7Ae—snR51-kl1-S       |
|---|------------------------|
| <b>PDB-ID:</b>  | <b>6ZDS</b>            |
| <b>Data collection</b>  |                        |
| Beamline  | P11, PETRA III, DESY   |
| Wavelength (Å)  | 1.03                   |
| Space Group   | C2                     |
| Cell dimensions   |                        |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                                | 109.39, 61.60, 138.61  |
| $\alpha$ , $\beta$ , $\gamma$ (°)                                 | 90.00, 108.36, 90.00   |
| Resolution (Å) <sup>a</sup>                                       | 1.91-54.50 (1.91-1.94) |
| $R_{\text{merge}}$ (%) <sup>a,1</sup>                             | 8.1 (77.9)             |
| $R_{\text{pim}}$ (%) <sup>a,2</sup>                               | 3.4 (32.0)             |
| $I/\sigma I^a$  | 12.4 (2.1)             |
| Completeness (%) <sup>a</sup>                                     | 100 (100)              |
| Redundancy <sup>a</sup>   | 6.8 (6.8)              |
| CC <sub>1/2</sub> (%) <sup>a</sup> (Karplus and Diederichs, 2012) | 99.8 (84.9)            |
| <b>Refinement</b>   |                        |
| Resolution (Å)  | 1.91-54.30             |
| No. reflections   | 68473                  |
| $R_{\text{work}}/ R_{\text{free}}$ (%)                            | 18.79/22.84            |
| No. atoms   | 7092                   |
| Protein/RNA   | 6517                   |
| Ligand/ion  | 10                     |
| Water   | 565                    |
| B-factors (Å <sup>2</sup> )                                       | 41.52                  |
| Protein   | 38.59                  |
| RNA   | 45.19                  |
| Ligand/ion  | 58.65                  |
| Water   | 43.36                  |
| R.m.s deviations  |                        |
| Bond lengths (Å)  | 0.004                  |
| Bond angles (°)   | 0.73                   |
| Ramachandran statistics (%)                                       |                        |
| Favored   | 99.59                  |
| Allowed   | 0.41                   |
| Outliers  | 0                      |

<sup>a</sup> Values for the highest resolution shell are shown in parentheses.

<sup>1</sup>  $R_{\text{merge}} = \sum_h \sum_i |<I_h> - I_{h,i}| / \sum_h \sum_i I_{h,i}$ , where *h* enumerates the unique reflections and *i* are their symmetry-equivalent contributions

<sup>2</sup>  $R_{\text{pim}} = \sum_h [1/(n_h - 1)]^{1/2} \sum_i |<I_h> - I_{h,i}| / \sum_h \sum_i I_{h,i}$ , where *h* enumerates the unique reflections and *i* are their symmetry-equivalent contributions

### *Nop1<sup>83-327</sup> and Nop56<sup>1-166</sup>*

The crystal structure of Nop56<sup>1-166</sup>-Nop1<sup>83-327</sup> was solved by molecular replacement executing Balbes (Long et al., 2008) from the CCP4 suite (Winn et al., 2011). For Nop1<sup>83-327</sup> human Fibrillarin (PDB-ID: 2IPX, sequence identity: 74%) was found to as a working search model. For Nop56<sup>1-166</sup> a model was built using the Phyre2 web server (Kelley et al., 2015) and energy minimized using Maestro from the Schrödinger 2018 Suite (Schrödinger, LLC). This model was used as a search model using Phaser (McCoy et al., 2007) from the Phenix suite (Adams et al., 2010), including the fixed partial solution containing only the model for Nop1<sup>83-327</sup>, which had been obtained from Balbes. Initial structural models were built with AutoBuild (Terwilliger et al., 2008) from the Phenix software package (Adams et al., 2010). The models were analyzed and completed by iterative model-building and refinement cycles using Coot (Emsley et al., 2010); and Phenix.refine (Afonine et al., 2012), including TLS-refinement and the addition of hydrogens in riding positions. Data collection and refinement statistics are summarized in **Table 5.5**.

**Table 5.5.** Crystallographic statistics for Nop1p<sup>83-327</sup>-Nop56p<sup>1-166</sup> complex

| <b>Structure</b>  | Nop56 <sup>1-166</sup> -Nop1 <sup>83-327</sup> |
|---|--|
| PDB-ID:   | <b>6ZDT</b>                                    |
| <b>Data collection</b>  |  |
| Beamline  | P11, PETRA III, DESY                           |
| Wavelength (Å)  | 1.03   |
| Space group   | P2 <sub>1</sub> 2 <sub>1</sub> 2               |
| Cell dimensions   |  |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                                | 69.04, 118.29, 48.83                           |
| $\alpha$ , $\beta$ , $\gamma$ (°)                                 | 90.00, 90.00, 90.00                            |
| Resolution (Å) <sup>a</sup>                                       | 1.71-48.83 (1.71-1.74)                         |
| <i>R</i> <sub>merge</sub> (%) <sup>a,1</sup>                      | 7.7 (121.4)                                    |
| <i>R</i> <sub>pim</sub> (%) <sup>a,2</sup>                        | 2.1 (35.5)                                     |
| <i>I</i> / $\sigma$ <sup>a</sup>                                  | 22.2 (2.1)                                     |
| Completeness (%) <sup>a</sup>                                     | 100 (100)                                      |
| Redundancy <sup>a</sup>   | 13.2 (12.6)                                    |
| CC <sub>1/2</sub> (%) <sup>a</sup> (Karplus and Diederichs, 2012) | 100 (85.6)                                     |
| <b>Refinement</b>   |  |
| Resolution (Å)  | 1.71-48.83                                     |
| No. reflections   | 44211  |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)           | 16.28/20.49                                    |
| No. atoms   | 3870   |
| Protein   | 3503   |
| Ligand/ion  | -  |
| Water   | 367  |
| B-factors (Å <sup>2</sup> )                                       | 33.20  |
| Protein   | 32.56  |
| Ligand/ion  | -  |
| Water   | 39.36  |
| R.m.s deviations  |  |
| Bond lengths (Å)  | 0.005  |
| Bond angles (°)   | 0.686  |
| Ramachandran statistics (%)                                       |  |
| Favored   | 98.18  |
| Allowed   | 1.82   |
| Outliers  | 0.00   |

<sup>a</sup> Values for the highest resolution shell are shown in parentheses.

<sup>1</sup>  $R_{merge} = \sum_h \sum_i |<I_h> - I_{h,i}| / \sum_h \sum_i I_{h,i}$ , where *h* enumerates the unique reflections and *i* are their symmetry-equivalent contributions

<sup>2</sup>  $R_{pim} = \sum_h [1/(n_h - 1)]^{1/2} \sum_i |<I_h> - I_{h,i}| / \sum_h \sum_i I_{h,i}$ , where *h* enumerates the unique reflections and *i* are their symmetry-equivalent contributions

## 5.8 Quantitative Mass spectrometry

To determine suitable peptides for quantification, each protein was first prepared and measured in isolation as described below. Two suitable peptides per proteins were chosen and purchased as synthetic peptides in exactly 1 mg quantity with the C-terminal residue of each peptide  $^{13}\text{C}/^{15}\text{N}$ -labeled (New England Peptides). The peptide sequences are given in **Table 5.6**. This labeling scheme resulted in an additional mass of 10 Da for arginines or 8 Da for lysines as C-terminals residues as compared to the unlabeled peptide.

RNP complexes for analysis by quantitative mass spectrometry were assembled as described in section 5.4. All complexes used in this section contain eukaryotic protein Snu13 and archaeal proteins Nop5 and Fibrillarin. The amount of sample ( $\mu\text{g}$ ) refers to the amount of the protein without RNA.

For each sample 20  $\mu\text{g}$  of protein were mixed with 5x gel loading dye (250 mM Tris-HCl, 50% glycerol, 10% SDS, 500 mM DTT, 0.5% bromophenol blue, pH 6.8), topped up with water to a final volume of 60  $\mu\text{l}$  and boiled for 5 min at 95 °C. Afterwards, 1  $\mu\text{l}$  of 40% acrylamide solution was added and the sample was incubated for 30 min at room temperature. The three proteins in each sample were separated via SDS gel-electrophoresis using 4-20% Mini-PROTEAN® TGX™ Precast Protein gels (BioRad) with a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BioRad), stained with Pierce™ PageBlue™ protein stain (Thermo Fisher Scientific) according to the manufacturer protocol. Afterwards, the lane corresponding to each of the three proteins (Nop5, Fibrillarin, and Snu13) was cut out and de-stained in a solution of 50% acetonitrile and 20 mM ammonium bicarbonate and incubated for 30 min at 37 °C under constant shaking; this step was repeated twice. Then, 100% acetonitrile was added, and the samples were incubated for 10 min at room temperature. The sample was dried in a vacuum concentrator. All samples were trypsin-digested by the addition of a 10n g/ $\mu\text{l}$  trypsin solution in 10% acetonitrile and 20 mM ammonium bicarbonate and incubated for 1 hour on ice and overnight at 37 °C. The trypsin-digested proteins were extracted by adding a 50% acetonitrile and 5% trifluoroacetic acid solution and incubating for 30 min at room temperature under constant shaking. Solid parts were removed by centrifugation; a 50% acetonitrile, 0.5% trifluoroacetic acid solution was added to the supernatant, incubated for 30 min at room temperature, and again centrifuged. As a final step, 100% acetonitrile was added and incubated for 20 min at room temperature. The final sample was dried completely using a vacuum concentrator. Final samples were resuspended in 2% acetonitrile and 0.1% formic acid solution.

**Table 5.8.** Heavy isotope-labeled peptides for quantification. Two peptides per protein with the best behavior during mass spectrometric measurements were selected. Each peptide was synthesized with  $^{13}\text{C}/^{15}\text{N}$ -labeled (\*) C-terminal residue. The labeled peptides were used as an external standard for quantification. For each peptide, both  $^{13}\text{C}/^{15}\text{N}$  labeled and unlabeled (light and heavy), the three most abundant fragment ions were detected (indicated as detected fragments).

| Protein                        | selected peptide                               | peptide mass (Da), heavy | detected fragments |
|--------------------------------|--|--------------------------|--------------------|
| Nop5<br><i>P. furiosus</i>     | H <sub>2</sub> N-AFISENVR* <u>R</u> -OH        | 944.49                   | y6<br>y5<br>y4     |
|                                | H <sub>2</sub> N-GIYAFDENGNLIEK* <u>K</u> -OH  | 1590.73                  | y11<br>y10<br>y9   |
| Fibrillarin <i>P. furiosus</i> | H <sub>2</sub> N-FPGVYVWIDDDGSEK* <u>K</u> -OH | 1647.79                  | y11<br>y10<br>y9   |
|                                | H <sub>2</sub> N-IYGIEFSPR* <u>R</u> -OH       | 1091.49                  | y7<br>y6<br>y5     |
| Snu13p<br><i>S. cerevisiae</i> | H <sub>2</sub> N-NVPYVFVPSR* <u>R</u> -OH      | 1187.54                  | y8<br>y6<br>y5     |
|                                | H <sub>2</sub> N-TQIYAVK* <u>K</u> -OH         | 830.41                   | y6<br>y5<br>y4     |

For each sample exactly 100 fmol of each of the six isotope labeled peptides (heavy) were spiked for absolute quantification (**Table 5.6**). Three biological replicates (three separately assembled, identical RNPs) were measured on a QTRAP4000 triple Quadrupol-MS (SCIEX) system, combined with an Ultimate 3000 liquid chromatography system (Dionex/Thermo Fisher Scientific) with Nano Flow (250 nl/min). An Acclaim™ PepMap™ 100 C18 LC column (Thermo Fisher Scientific) and Trap column from the same product line (Thermo Fisher Scientific) were used as well as a selected reaction monitoring (SRM) detection set-up (Lange et al., 2008; Chambers et al., 2015). For each light and heavy peptide, the three most abundant fragment ions were detected (**Table 5.6**). Data processing and peak integration were done using the Skyline software (Henderson et al., 2018; Pino et al., 2020; Schilling et al., 2012). Ratios between light and heavy fragments belonging to the same peptide (L/H) were calculated for each fragment separately and then averaged over all fragments to obtain the average ratio for this peptide. This ratio was multiplied by 100 (corresponding to 100 fmol of heavy peptide) and the value was then averaged over all biological replicates and both peptides for each protein. Ratios between proteins were determined using these final values. Raw values for all measurements can be found in **Appendix 1**. \*

\*All mass spectrometric measurements and all sample preparation were performed at the Central Research Facility Proteomics at the Hannover Medical School.

---

## 5.9 Activity assays

All RNPs used for activity assays were assembled as described in section 5.4. Activity assays were performed with RNPs containing the following guide RNAs (described in **Table 5.2**): sR26, snR51, snR51-1, snR41-2, snR41-3, snR41-4, and snR54.

All substrate RNAs (RNAs to be methylated shown in **Table 5.9**) with and without the 5'-biotin labels were ordered from Integrated DNA Technologies (IDT) with HPLC purification. <sup>3</sup>H-labeled SAM was purchased from Hartman Analytics, Inc and not used for more than one month after shipment. All reactions contained 0.75 μM of each target RNA (one or both carrying a 5'-biotin label), 0.15 μM of guide RNA (assembled in the RNP complex), and 2.25 μM of <sup>3</sup>H-labeled SAM. The assays were performed in methylation buffer (20 mM HEPES, 150 mM NaCl, pH 7.5). The RNP complex was first mixed with the substrate RNAs and incubated for 10 min at room temperature; afterwards, <sup>3</sup>H-labeled SAM was added and the reaction incubated at the respective reaction temperature. Twice 10 μL were taken from each reaction at time points 0 (right after the addition of SAM co-factor), 5, 15, 30, 60, 80, 100 and 120 min and the methylation reaction was stopped by the addition of 15 μL of stop buffer (20 mM Tris-HCl, 0.1% SDS, 20 mM EDTA, 20 mM NaCl) and proteinase K (New England Biolabs). Stopped reactions were incubated for 1 hour at 50 °C to degrade the proteins and then 5 μL were added to Pierce™ NeutrAvidin™ Agarose beads (Thermo Fisher Scientific) pre-equilibrated in methylation buffer. Each reaction was incubated with the beads for 1 hour at room temperature under constant shaking to allow binding of 5'-biotin-labeled RNAs to the beads. After incubation, the beads were washed three times with 500 μL of methylation buffer, then transferred into scintillation vials and mixed with a scintillation cocktail (Zinsser Analytic, Aquasafe 500 plus). Counts were measured for 1 min on either a Tri-Carb liquid scintillation counter (PerkinElmer), a Pharmacia Wallac 1410 liquid scintillation counter (Perkin Elmer) or a Hidex 300 SL liquid scintillation counter. Data, including CPM, DPM, and counting efficiency, was extracted with the MikroWin 300 SL software from the Hidex 330SL system or via direct read-out from the counter. CPM/fmol was calculated with **Equation 5.3** (Kahl et al., 2004), with 1 Ci = 2.22x10<sup>12</sup> DPM as a conversion factor. The concentration of methylated target RNA in the final samples was calculated from CPM/fmol applying **Equation 5.3**.

$$\frac{CPM}{fmol} = [Specific\ activity\ \left(\frac{Ci}{mmol}\right) * \left[2.22 * 10^{12} \frac{DPM}{Ci}\right] * \left[\frac{mmol}{10^{12} fmol}\right] * Efficiency\left(\frac{CPM}{DPM}\right)] \quad Eq. (5.3)$$

$$nM = \frac{CPM / \text{Specific Activity} \left( \frac{CPM}{fmol} \right)}{\text{Volume} (\mu l)} \quad \text{Eq. (5.4)}$$

**Table 5.9.** Substrate RNAs used in the activity assays. All RNAs were purchased either with or without 5'-biotin label and base-pair r with the guide sequence upstream of either Box D' or Box D.

| guide RNA               | type of guide      | sequence (5' → 3')    |
|-------------------------|--------------------|-----------------------|
| sR26                    | upstream of Box D' | (BIO)-GCUUCGCCCAUCAC  |
|                         | upstream of Box D  | (BIO)-GUAGAUCACCUCGG  |
| snR51/snR51-2           | upstream of Box D' | (BIO)-GCCAGUGUGAAUA   |
|                         | upstream of Box D  | (BIO)-CCGUCAUUAUAUC   |
| snR41-2/snR41-3/snR41-4 | upstream of Box D' | (BIO)-CCUGGAACAAUUGG  |
|                         | upstream of Box D  | (BIO)-GCUCAAGGCUGAAAC |
| snR54                   | upstream of Box D' | (BIO)-GCAAGAACGAAAGUU |

The complete raw data for all activity assays can be found in **Appendix 2**.

## 5.10 Materials

All used materials, including chemicals, kits, and equipment, are listed in alphabetical order.

| Material  | Manufacturer                 |
|---|------------------------------|
| <sup>3</sup> H-SAM                              | Hartman Analytics            |
| 40% Acrylamide solution (37,5:1)                | Carl Roth                    |
| Acetonitrile                                    | Sigma-Aldrich (Merck)        |
| Ammonium bicarbonate                            | Sigma-Aldrich (Merck)        |
| Aquasafe 500 plus scintillation cocktail        | Zinsser Analytic             |
| ATP   | Carl Roth                    |
| BME   | Carl Roth                    |
| Boric Acid                                      | Carl Roth                    |
| Bromophenol blue                                | VWR                          |
| Chloroform/Isoamyl alcohol                      | Carl Roth                    |
| cComplete EDTA-free protease inhibitor cocktail | Roche (purchased from Merck) |
| CTP   | Carl Roth                    |
| Dpnl  | New England Biolabs          |
| DTT   | Carl Roth                    |
| <i>E. coli</i> BL21(DE3)                        | made in-house                |
| <i>E. coli</i> OmniMax                          | made in-house                |
| <i>E. coli</i> Top10                            | made in-house                |
| EcoRI-HF  | New England Biolabs          |
| EDTA  | Sigma-Aldrich (Merck)        |
| Formamide                                       | Carl Roth                    |
| Formic Acid                                     | Sigma-Aldrich (Merck)        |
| Glycerol  | Carl Roth                    |
| GTP   | Carl Roth                    |
| Guanidinium Chloride                            | Carl Roth                    |

|  |  |
|--|--|
| HEPES (BioUltra)                       | Sigma-Aldrich (Merck)                  |
| HiLoad 16/600 Superdex 75 pg column    | Cytiva                                 |
| HindIII-HF                             | New England Biolabs                    |
| HiPrep 26/10 Desalting column          | Cytiva                                 |
| HisTrap FF column (5ml)                | Cytiva                                 |
| Imidazole                              | Carl Roth                              |
| IPTG                                   | Carl Roth                              |
| LB Medium                              | Carl Roth                              |
| LiCl                                   | Carl Roth                              |
| LiCl (BioUltra)                        | Sigma-Aldrich (Merck)                  |
| Lysozyme                               | Carl Roth                              |
| MES                                    | Carl Roth                              |
| MgCl <sub>2</sub>                      | Carl Roth                              |
| MOPS                                   | Carl Roth                              |
| Na <sub>2</sub> HPO <sub>4</sub>       | Sigma-Aldrich (Merck)                  |
| NaCl                                   | Carl Roth                              |
| NaH <sub>2</sub> PO <sub>4</sub>       | Sigma-Aldrich (Merck)                  |
| NcoI-HF                                | New England Biolabs                    |
| NeXtal JCSG Core I Suite               | NeXtal (purchased via Qiagen)          |
| NeXtal JCSG Core II Suite              | NeXtal (purchased via Qiagen)          |
| NeXtal JCSG Core III Suite             | NeXtal (purchased via Qiagen)          |
| NeXtal JCSG Core IV Suite              | NeXtal (purchased via Qiagen)          |
| NeXtal Nucleix Suite                   | NeXtal (purchased via Qiagen)          |
| NeXtal PEG Suite                       | NeXtal (purchased via Qiagen)          |
| NeXtal PEG II Suite                    | NeXtal (purchased via Qiagen)          |
| NeXtal Protein Complex Suite           | NeXtal (purchased via Qiagen)          |
| PEG 10 000 (BioUltra)                  | Sigma-Aldrich (Merck)                  |
| PEG 400 (BioUltra)                     | Sigma-Aldrich (Merck)                  |
| Pfu Plus Polymerase                    | Roboklon                               |
| Phenol/Chloroform/Isoamyl alcohol      | Carl Roth                              |
| Pierce™ NeutrAvidin™ Agarose beads     | Thermo Fisher Scientific               |
| Pierce™ PageBlue™ protein stain        | Thermo Fisher Scientific               |
| Proteinase K                           | New England Biolabs                    |
| PstI-HF                                | New England Biolabs                    |
| Qiagen Plasmid Mega Kit                | Qiagen                                 |
| RNAaseAlert™ Kit                       | Invitrogene (Thermo Fisher Scientific) |
| SDS                                    | Carl Roth                              |
| Spermidine                             | Carl Roth                              |
| Superdex 200 Increase 10/300 GL column | Cytiva                                 |
| T4 DNA Ligase                          | New England Biolabs                    |
| T7 RNA polymerase                      | made in-house                          |
| TEV protease                           | made in-house                          |
| Trifluoroacetic acid                   | Sigma-Aldrich (Merck)                  |
| Tris-Base                              | Sigma-Aldrich (Merck)                  |
| Tris-HCl                               | Carl Roth                              |
| Triton X-100                           | Sigma-Aldrich (Merck)                  |
| Trypsin                                | Sigma-Aldrich (Merck)                  |



|               |                     |
|---------------|---------------------|
| UTP           | Carl Roth           |
| Xhol          | New England Biolabs |
| Xylene cyanol | VWR                 |

# Appendix 1 – Mass spectrometry raw data

Roman numbers after the sample name indicate biological replicates and arabic numbers indicate technical replicates.

## sR26

Peptide

Nop5-1

AFISENVR

| Light     | y6 - 717.3890+ | y5 - 604.3049+ | y4 - 517.2729+ | Heavy y   | y6 - 727.3972+ | y5 - 614.3132+ | y4 - 527.2812+ |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 26A-I-1   | 52132812       | 59702972       | 13337320       | 26-I-1    | 11985508       | 15295504       | 2060798.375    |
| 26A-I-2   | 42395388       | 49979292       | 10009695       | 26-I-2    | 9833323        | 12468476       | 1730206.125    |
| 26A-I-3   | 37199140       | 44558692       | 8843790        | 26-I-3    | 9147783        | 11524022       | 1627652.125    |
| 26A-II-1  | 45982468       | 52923124       | 10884664       | 26-II-1   | 10145129       | 12257687       | 1725589.5      |
| 26A-II-2  | 40411492       | 45977772       | 9131419        | 26-II-2   | 9161475        | 11534259       | 1558279.125    |
| 26A-II-3  | 37642476       | 44335896       | 8489119        | 26-II-3   | 9409422        | 11815644       | 1613639.5      |
| 26A-III-1 | 45604216       | 53768300       | 10540922       | 26-III-1  | 9924762        | 12193688       | 1615981.25     |
| 26A-III-2 | 37351288       | 43696444       | 8741140        | 26-III-2  | 8506428        | 10470189       | 1509600.75     |
| 26A-III-3 | 35170444       | 40828692       | 7854402        | 26-III-3  | 8541784        | 10594898       | 1468838.875    |

Peptide

Nop5-2

GIYAFDENG

LIEK

| Light     | y11 - 1249.6059+ | y10 - 1178.5688+ | y9 - 1031.5004+ | Heavy y   | y11 - 1257.6201+ | y10 - 1186.5830+ | y9 - 1039.5146+ |
|-----------|------------------|------------------|-----------------|-----------|------------------|------------------|-----------------|
| Replicate | Peak Area        | Peak Area        | Peak Area       | Replicate | Peak Area        | Peak Area        | Peak Area       |
| 26A-I-1   | 2646962.25       | 5564616.5        | 4369947.5       | 26-I-1    | 1175449.75       | 2108084.25       | 1960001         |
| 26A-I-2   | 1218096.125      | 2208257.5        | 1785592.25      | 26-I-2    | 565238.3125      | 1098971          | 1058040.5       |
| 26A-I-3   | 2030749.625      | 3515082.25       | 2769322.5       | 26-I-3    | 590593.3125      | 1210409.75       | 1106068         |
| 26A-II-1  | 2875988.75       | 5296381.5        | 4383471         | 26-II-1   | 1215723.125      | 2390626.75       | 2251443.25      |
| 26A-II-2  | 1975510.125      | 3473489.75       | 3011420.25      | 26-II-2   | 822730.25        | 1642289.5        | 1492010.75      |
| 26A-II-3  | 1955568.875      | 3710925          | 3065364.75      | 26-II-3   | 769991.5         | 1354558          | 1373819.125     |
| 26A-III-1 | 1658740.875      | 3119851.25       | 2481079         | 26-III-1  | 735692.25        | 1435963.875      | 1247944.625     |
| 26A-III-2 | 1289049.25       | 2471375          | 2058626.5       | 26-III-2  | 551408.1875      | 1093434.375      | 1007214.688     |
| 26A-III-3 | 1756433.125      | 3542772.75       | 2779139.5       | 26-III-3  | 644980.25        | 1283457.25       | 1235449.25      |

Peptide

Fib-1

FPGVYVIDD

DGSEK

| Light     | y11 - 1239.5739+ | y10 - 1076.5106+ | y9 - 977.4422+ | Heavy y   | y11 - 1247.5881+ | y10 - 1084.5248+ | y9 - 985.4564+ |
|-----------|------------------|------------------|----------------|-----------|------------------|------------------|----------------|
| Replicate | Peak Area        | Peak Area        | Peak Area      | Replicate | Peak Area        | Peak Area        | Peak Area      |
| 26A-I-1   | 4887037.5        | 2775825.75       | 2996190.25     | 26-I-1    | 1353418.875      | 907325.8125      | 1044908.625    |

|           |             |                 |                 |          |                 |                 |                 |
|-----------|-------------|-----------------|-----------------|----------|-----------------|-----------------|-----------------|
| 26A-I-2   | 1023114.25  | 622682.375      | 709330.5        | 26-I-2   | 225189.875      | 153528.203<br>1 | 184411.85<br>94 |
| 26A-I-3   | 1724278.625 | 958957.312<br>5 | 1096386.5       | 26-I-3   | 385448.218<br>8 | 256806.578<br>1 | 323686.96<br>88 |
| 26A-II-1  | 2955412.75  | 1733728.12<br>5 | 2134015         | 26-II-1  | 895853.687<br>5 | 623675.937<br>5 | 731872.25       |
| 26A-II-2  | 978810.1875 | 510284.625      | 594361.93<br>75 | 26-II-2  | 213124.343<br>8 | 131781.234<br>4 | 151521.48<br>44 |
| 26A-II-3  | 1926461.125 | 1126839.37<br>5 | 1334972.8<br>75 | 26-II-3  | 407159.187<br>5 | 279132.781<br>3 | 338912.93<br>75 |
| 26A-III-1 | 1088753.625 | 644876.687<br>5 | 759822.43<br>75 | 26-III-1 | 304471.593<br>8 | 170782.453<br>1 | 253460.54<br>69 |
| 26A-III-2 | 354910.3438 | 215050.734<br>4 | 245976.45<br>31 | 26-III-2 | 74652.5546<br>9 | 49940.9882<br>8 | 61872.140<br>63 |
| 26A-III-3 | 1145750.25  | 718193.437<br>5 | 831068.12<br>5  | 26-III-3 | 260854.359<br>4 | 182463.421<br>9 | 229104.46<br>88 |

Peptide

Fib-2 IYGIEFSPR

| Light     | y7 - 805.4203+ | y6 - 748.3988+ | y5 - 635.3148+ |          | y7 - 815.4285+ | y6 - 758.4071+  | y5 - 645.3230+ |
|-----------|----------------|----------------|----------------|----------|----------------|-----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Heavy    | Peak Area      | Peak Area       | Peak Area      |
| 26A-I-1   | 37601888       | 3589210        | 21674960       | 26-I-1   | 13722064       | 1187715.25      | 7466721        |
| 26A-I-2   | 32488424       | 3217056.25     | 18961180       | 26-I-2   | 10289465       | 916304.75       | 5584004        |
| 26A-I-3   | 34410052       | 3423790.5      | 19876490       | 26-I-3   | 11259034       | 980490.875      | 6134218        |
| 26A-II-1  | 38037796       | 3874061.75     | 21731932       | 26-II-1  | 13810867       | 1248193.25      | 7351309        |
| 26A-II-2  | 34031832       | 3248284        | 19485872       | 26-II-2  | 10534891       | 996631.375      | 5536988.5      |
| 26A-II-3  | 35922400       | 3627277.5      | 20689116       | 26-II-3  | 11993087       | 1092777.5       | 6577317        |
| 26A-III-1 | 33284292       | 3432570        | 19608046       | 26-III-1 | 12082931       | 1119447.25      | 6486502.5      |
| 26A-III-2 | 32273522       | 3171905.25     | 18728450       | 26-III-2 | 10432757       | 911407.5        | 5505750.5      |
| 26-III-3  | 35538496       | 3452242        | 19696384       | 26-III-3 | 12081388       | 1048760.87<br>5 | 6068038        |

Peptide

Snu13-1 NVPYVFPSPR

| Light     | y8 - 964.5251+ | y6 - 704.4090+ | y5 - 605.3406+ | Heavy     | y8 - 974.5333+ | y6 - 714.4173+ | y5 - 615.3488+ |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 26A-I-1   | 31615566       | 12937429       | 13352658       | 26-I-1    | 12724408       | 4481618.5      | 5099076        |
| 26A-I-2   | 25522038       | 10313937       | 11081436       | 26-I-2    | 8979422        | 3276583.5      | 3534866.2<br>5 |
| 26A-I-3   | 20301850       | 8396010        | 9081989        | 26-I-3    | 9113267        | 3470556.75     | 3822380.7<br>5 |
| 26A-II-1  | 28788436       | 11522831       | 12005962       | 26-II-1   | 11503723       | 4466472.5      | 4728192        |
| 26A-II-2  | 24224092       | 10427848       | 10748580       | 26-II-2   | 8600724        | 3248815.5      | 3501482.5      |
| 26A-II-3  | 21679824       | 8982751        | 9562554        | 26-II-3   | 9756761        | 3702330        | 3888438.5      |
| 26A-III-1 | 24154452       | 10075175       | 10664752       | 26-III-1  | 9613837        | 3667669.25     | 3971789.5      |
| 26A-III-2 | 24204952       | 9696899        | 10601720       | 26-III-2  | 8067489        | 3209832        | 3312531.2<br>5 |
| 26A-III-3 | 19735628       | 8052954.5      | 8396007        | 26-III-3  | 8692197        | 3281243.5      | 3478799.5      |

Peptide

Snu13-2 TQIYAVK

| Light | y6 - 721.4243+ | y5 - 593.3657+ | y4 - 480.2817+ | Heavy | y6 - 729.4385+ | y5 - 601.3799+ | y4 - 488.2959+ |
|-------|----------------|----------------|----------------|-------|----------------|----------------|----------------|
|-------|----------------|----------------|----------------|-------|----------------|----------------|----------------|

| Replicate | Peak Area | Peak Area | Peak Area | Replicate | Peak Area   | Peak Area | Peak Area   |
|-----------|-----------|-----------|-----------|-----------|-------------|-----------|-------------|
| 26A-I-1   | 7180615   | 82621536  | 20787130  | 26-I-1    | 1048481.5   | 15698282  | 2986606.25  |
| 26A-I-2   | 5036899   | 55163076  | 14764939  | 26-I-2    | 879248      | 12427088  | 2517952     |
| 26A-I-3   | 4451841.5 | 50144540  | 13181380  | 26-I-3    | 872980.75   | 12227402  | 2435358.5   |
| 26A-II-1  | 4303058   | 52663140  | 12440516  | 26-II-1   | 615468.25   | 8765042   | 1765005.125 |
| 26A-II-2  | 4874678.5 | 53584324  | 13699020  | 26-II-2   | 794029.625  | 11302909  | 2322002.25  |
| 26A-II-3  | 4368497   | 49641204  | 12355558  | 26-II-3   | 819570.5625 | 11419607  | 2299894.25  |
| 26A-III-1 | 5078529.5 | 59180932  | 14777741  | 26-III-1  | 762981.8125 | 10637897  | 2129874.25  |
| 26A-III-2 | 4490464   | 50118128  | 12847949  | 26-III-2  | 766337.1875 | 10833256  | 2152761.5   |
| 26A-III-3 | 4314840   | 48330528  | 12395221  | 26-III-3  | 817190      | 11091269  | 2264599     |

**snR51**

Peptide

Nop5-1

AFISENVR

| Light     | y6 - 717.3890+ | y5 - 604.3049+ | y4 - 517.2729+ | Heavy     | y6 - 727.3972+ | y5 - 614.3132+ | y4 - 527.2812+ |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 51A-I-1   | 26508642       | 31278624       | 5231460.5      | 51A-I-1   | 13949050       | 18330410       | 2889011.25     |
| 51A-I-2   | 15588548       | 20261396       | 3070594        | 51A-I-2   | 8900909        | 11393422       | 1779481.625    |
| 51A-II-1  | 29760882       | 35574072       | 6100036        | 51A-II-1  | 9980998        | 13117816       | 1935624.25     |
| 51A-II-2  | 25289616       | 30759478       | 5264752.5      | 51A-II-2  | 8571746        | 10594626       | 1728429        |
| 51A-III-1 | 45940700       | 55901200       | 10914474       | 51A-III-1 | 12329265       | 15900694       | 2422752.5      |
| 51A-III-2 | 29032812       | 41572916       | 6929737.5      | 51A-III-2 | 7840335.5      | 10158366       | 1606734.375    |
| 51B-I-1   | 36768588       | 46297496       | 8033269.5      | 51B-I-1   | 12277868       | 16515324       | 2576425.5      |
| 51B-I-2   | 27101736       | 31861294       | 5081904.5      | 51B-I-2   | 7745182        | 10225950       | 1862544        |
| 51B-II-1  | 43206804       | 53615956       | 9524736        | 51B-II-1  | 13438380       | 17998952       | 2930226        |
| 51B-II-2  | 27041666       | 30931122       | 5356196.5      | 51B-II-2  | 7632064        | 9351020        | 1505355.125    |
| 51B-III-1 | 36689124       | 44413816       | 8003462.5      | 51B-III-1 | 13638033       | 17829408       | 2820723.25     |
| 51B-III-2 | 17828728       | 23270072       | 4154056.5      | 51B-III-2 | 7056032.5      | 8739439        | 1374067.625    |

Peptide

GIYAFDENG

Nop5-2

LIEK

| Light     | y11 - 1249.6059+ | y10 - 1178.5688+ | y9 - 1031.5004+ | Heavy     | y11 - 1257.6201+ | y10 - 1186.5830+ | y9 - 1039.5146+ |
|-----------|------------------|------------------|-----------------|-----------|------------------|------------------|-----------------|
| Replicate | Peak Area        | Peak Area        | Peak Area       | Replicate | Peak Area        | Peak Area        | Peak Area       |
| 51A-I-1   | 394752.625       | 699118.75        | 538344.25       | 51A-I-1   | 379225.875       | 731103.9375      | 595462.0625     |
| 51A-I-2   | 192114.7344      | 354426.4063      | 304678.7188     | 51A-I-2   | 214793.2188      | 404174.0938      | 346266.2813     |
| 51A-II-1  | 553852.125       | 906352.375       | 742445.9375     | 51A-II-1  | 310125.6875      | 620789.875       | 506651.8438     |

|           |             |                 |                 |           |                 |                 |                 |
|-----------|-------------|-----------------|-----------------|-----------|-----------------|-----------------|-----------------|
| 51A-II-2  | 355348.3438 | 668925.062<br>5 | 535972.18<br>75 | 51A-II-2  | 227777.312<br>5 | 439346.906<br>3 | 351829.40<br>63 |
| 51A-III-1 | 963245.375  | 1720106.12<br>5 | 1313778.1<br>25 | 51A-III-1 | 277402.406<br>3 | 538857.375      | 446279.43<br>75 |
| 51A-III-2 | 619505.8125 | 1121911.87<br>5 | 915340.93<br>75 | 51A-III-2 | 189298.031<br>3 | 387066.843<br>8 | 315220.59<br>38 |
| 51B-I-1   | 620242.75   | 1026547.31<br>3 | 813561.43<br>75 | 51B-I-1   | 301136.562<br>5 | 601589.937<br>5 | 490356.15<br>63 |
| 51B-I-2   | 345029.875  | 622781.312<br>5 | 521345.78<br>13 | 51B-I-2   | 194979.062<br>5 | 398119.343<br>8 | 313234.18<br>75 |
| 51B-II-1  | 591032.875  | 1038742.37<br>5 | 872152.75       | 51B-II-1  | 299915.062<br>5 | 589418.312<br>5 | 503680.96<br>88 |
| 51B-II-2  | 198216.2969 | 348888.875      | 299749.28<br>13 | 51B-II-2  | 110652.109<br>4 | 221863.796<br>9 | 189532.40<br>63 |
| 51B-III-1 | 472150.5    | 841912.625      | 676388.62<br>5  | 51B-III-1 | 317314.781<br>3 | 613356.5        | 506038.28<br>13 |
| 51B-III-2 | 59730.17188 | 105284.273<br>4 | 89678.156<br>25 | 51B-III-2 | 45326.1210<br>9 | 93926.8359<br>4 | 77657.515<br>63 |

Peptide FPGVYVVIDD  
Fib-1 DGSEK

| Light     | y11 -<br>1239.5739+ | y10 -<br>1076.5106+ | y9 -<br>977.4422+ | Heavy<br>y    | y11 -<br>1247.5881+ | y10 -<br>1084.5248+ | y9 -<br>985.4564+ |
|-----------|---------------------|---------------------|-------------------|---------------|---------------------|---------------------|-------------------|
| Replicate | Peak Area           | Peak Area           | Peak Area         | Replic<br>ate | Peak Area           | Peak Area           | Peak Area         |
| 51A-I-1   | 1011501.875         | 652995.375          | 700257.75         | 51A-I-1       | 691310.625          | 381702.718<br>8     | 445575.28<br>13   |
| 51A-I-2   | 180040.25           | 107489.804<br>7     | 132342.93<br>75   | 51A-I-2       | 134206.921<br>9     | 77500.3593<br>8     | 84479.945<br>31   |
| 51A-II-1  | 1139019.125         | 730309.375          | 816425.43<br>75   | 51A-II-1      | 439921.187<br>5     | 237140.328<br>1     | 267067.62<br>5    |
| 51A-II-2  | 140088.8281         | 99081.1640<br>6     | 105211.60<br>16   | 51A-II-2      | 57535.0039<br>1     | 31584.1914<br>1     | 39247.242<br>19   |
| 51A-III-1 | 2112019.25          | 1285362.62<br>5     | 1396000.5         | 51A-III-1     | 553189.375          | 311326.937<br>5     | 359376.31<br>25   |
| 51A-III-2 | 481902.2813         | 313025.218<br>8     | 343594.59<br>38   | 51A-III-2     | 150817.109<br>4     | 84873.4531<br>3     | 94110.531<br>25   |
| 51B-I-1   | 1487128.75          | 919918.062<br>5     | 1013034.4<br>38   | 51B-I-1       | 559071.25           | 324954.718<br>8     | 355524.09<br>38   |
| 51B-I-2   | 234007.8594         | 151665.25           | 157854.39<br>06   | 51B-I-2       | 94576.0234<br>4     | 55702.3554<br>7     | 62740.187<br>5    |
| 51B-II-1  | 1608453.375         | 1005764.93<br>8     | 1152709.1<br>25   | 51B-II-1      | 541487.687<br>5     | 311161.25           | 331437.25         |
| 51B-II-2  | 306269.8125         | 209768.75           | 247317.64<br>06   | 51B-II-2      | 99109.2109<br>4     | 63299.0742<br>2     | 67231.976<br>56   |
| 51B-III-1 | 1126385.5           | 703695.812<br>5     | 802155.62<br>5    | 51B-III-1     | 460673.562<br>5     | 256670.75           | 291920.78<br>13   |
| 51B-III-2 | 145207              | 96583.875           | 108299.78<br>91   | 51B-III-2     | 58807.2812<br>5     | 34092.0781<br>3     | 38873.664<br>06   |

Peptide IYGIEFSPR  
Fib-2

| Light     | y7 - 805.4203+ | y6 -<br>748.3988+ | y5 -<br>635.3148+ | Heavy<br>y    | y7 -<br>815.4285+ | y6 -<br>758.4071+ | y5 -<br>645.3230+ |
|-----------|----------------|-------------------|-------------------|---------------|-------------------|-------------------|-------------------|
| Replicate | Peak Area      | Peak Area         | Peak Area         | Replic<br>ate | Peak Area         | Peak Area         | Peak Area         |
| 51A-I-1   | 27032566       | 2637702.5         | 15499834          | 51A-I-1       | 18218426          | 1663139.12<br>5   | 9842732           |
| 51A-I-2   | 22429882       | 2052386           | 11608348          | 51A-I-2       | 15433670          | 1328707.75        | 7689420.5         |
| 51A-II-1  | 31236060       | 3148621.5         | 18025620          | 51A-II-1      | 12565719          | 1151640           | 6705426.5         |
| 51A-II-2  | 35073404       | 3275912.25        | 19246412          | 51A-II-2      | 13657294          | 1141794.37<br>5   | 6790021           |
| 51A-III-1 | 45395160       | 5172517           | 27602782          | 51A-III-1     | 13706628          | 1228428.25        | 7346483           |
| 51A-III-2 | 36202492       | 3770409.75        | 20330726          | 51A-III-2     | 10126483          | 887835.187<br>5   | 5245378           |

|           |          |            |          |           |          |                 |           |
|-----------|----------|------------|----------|-----------|----------|-----------------|-----------|
| 51B-I-1   | 38652008 | 3945242    | 22619132 | 51B-I-1   | 15053712 | 1337332.62<br>5 | 7731036   |
| 51B-I-2   | 35251896 | 3653190.5  | 20407284 | 51B-I-2   | 13340603 | 1149968.37<br>5 | 6936391   |
| 51B-II-1  | 39674268 | 4079812    | 23708566 | 51B-II-1  | 15376853 | 1435798.12<br>5 | 8429090   |
| 51B-II-2  | 39319068 | 3797089.25 | 22273432 | 51B-II-2  | 15369541 | 1294675         | 8086380   |
| 51B-III-1 | 37756532 | 3962500.75 | 22802710 | 51B-III-1 | 16610551 | 1517801.62<br>5 | 8777071   |
| 51B-III-2 | 36490900 | 3527588.25 | 20795546 | 51B-III-2 | 15529925 | 1329197.75      | 7861318.5 |

Peptide

Snu13-1

NVPYVFPVPSR

| Light     | y8 - 964.5251+ | y6 - 704.4090+  | y5 - 605.3406+  | Heavy     | y8 - 974.5333+ | y6 - 714.4173+ | y5 - 615.3488+ |
|-----------|----------------|-----------------|-----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area       | Peak Area       | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 51A-I-1   | 3331874.25     | 1553606.25      | 1618289.7<br>5  | 51A-I-1   | 7220615        | 3187274        | 3431645.5      |
| 51A-I-2   | 3899207        | 1740240.5       | 1789988.3<br>75 | 51A-I-2   | 8130190        | 3640841.5      | 3652476.2<br>5 |
| 51A-II-1  | 2955466.5      | 1343041.75      | 1420304.2<br>5  | 51A-II-1  | 4661083        | 2071194.25     | 2164924.2<br>5 |
| 51A-II-2  | 4055890        | 1856837.37<br>5 | 1883158.3<br>75 | 51A-II-2  | 6367025        | 2831048.5      | 2890053.2<br>5 |
| 51A-III-1 | 4023400.5      | 1965358         | 1973266.3<br>75 | 51A-III-1 | 4247317.5      | 1953102.25     | 1998171.2<br>5 |
| 51A-III-2 | 4723804        | 2231705.5       | 2289248.5       | 51A-III-2 | 4930134.5      | 2159994.25     | 2318972.7<br>5 |
| 51B-I-1   | 4857022.5      | 2220928.75      | 2369861.5       | 51B-I-1   | 5483041.5      | 2487927        | 2614253        |
| 51B-I-2   | 5975087        | 2751138.5       | 2848790.5       | 51B-I-2   | 6543354        | 2898227.5      | 3008306.2<br>5 |
| 51B-II-1  | 3859484        | 1770966.5       | 1919961.5       | 51B-II-1  | 5535626        | 2427182        | 2512091.5      |
| 51B-II-2  | 5411201.5      | 2514202.75      | 2534778         | 51B-II-2  | 7516899.5      | 3298610.5      | 3429058.7<br>5 |
| 51B-III-1 | 4428380        | 2057235.12<br>5 | 2181180.2<br>5  | 51B-III-1 | 6606745        | 3007577        | 3203400.7<br>5 |
| 51B-III-2 | 5586774.5      | 2494890.25      | 2562701.7<br>5  | 51B-III-2 | 8101784.5      | 3650895.25     | 3733175.7<br>5 |

Peptide

Snu13-2

TQIYAVK

| Light     | y6 - 721.4243+ | y5 - 593.3657+ | y4 - 480.2817+  | Heavy     | y6 - 729.4385+  | y5 - 601.3799+ | y4 - 488.2959+  |
|-----------|----------------|----------------|-----------------|-----------|-----------------|----------------|-----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area       | Replicate | Peak Area       | Peak Area      | Peak Area       |
| 51A-I-1   | 498009.2813    | 6469649.5      | 1570140         | 51A-I-1   | 405213.437<br>5 | 6025293.5      | 1304420.5       |
| 51A-I-2   | 332185.0938    | 4650363.5      | 914265          | 51A-I-2   | 310022.843<br>8 | 3678973.5      | 798689.93<br>75 |
| 51A-II-1  | 582234.4375    | 8345768        | 1936608.8<br>75 | 51A-II-1  | 304723.343<br>8 | 4271495        | 926708.81<br>25 |
| 51A-II-2  | 521371.9063    | 7263043        | 1417368.6<br>25 | 51A-II-2  | 229762.390<br>6 | 2949939.5      | 792289.81<br>25 |
| 51A-III-1 | 1135452.375    | 15669824       | 3799664.5       | 51A-III-1 | 398443.968<br>8 | 5434471.5      | 1090050.8<br>75 |
| 51A-III-2 | 639681         | 8375111        | 1982795.7<br>5  | 51A-III-2 | 216278.625      | 2838071.75     | 583488.31<br>25 |
| 51B-I-1   | 922570.625     | 13869482       | 3090162.2<br>5  | 51B-I-1   | 385995.843<br>8 | 5824796.5      | 1288806.8<br>75 |
| 51B-I-2   | 507971         | 6879740        | 1380604.1<br>25 | 51B-I-2   | 199253.046<br>9 | 2640910.75     | 567601.93<br>75 |
| 51B-II-1  | 1002394.063    | 13628034       | 3264435.5       | 51B-II-1  | 510429          | 6670849        | 1483636.7<br>5  |

|           |             |           |                 |           |                 |            |                 |
|-----------|-------------|-----------|-----------------|-----------|-----------------|------------|-----------------|
| 51B-II-2  | 350753.0313 | 5004166.5 | 1203800.1<br>25 | 51B-II-2  | 173979.312<br>5 | 2531673.25 | 534626.68<br>75 |
| 51B-III-1 | 805104.1875 | 10722087  | 2380306.2<br>5  | 51B-III-1 | 396213.687<br>5 | 6026392    | 1215817.6<br>25 |
| 51B-III-2 | 398154      | 6027405   | 1386113.6<br>25 | 51B-III-2 | 207082.328<br>1 | 2683246.75 | 649540.12<br>5  |

**snR51-2**

Peptide

Nop5-1

AFISENVR

| Light     | y6 - 717.3890+ | y5 - 604.3049+ | y4 - 517.2729+ | Heavy y   | y6 - 727.3972+ | y5 - 614.3132+ | y4 - 527.2812+  |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|-----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area       |
| 512-I-1   | 36846832       | 44638204       | 8411222        | 512-I-1   | 12190516       | 16150123       | 2509400.5       |
| 512-I-2   | 21020722       | 25251000       | 4610748.5      | 512-I-2   | 6739133.5      | 9069333        | 1412329.5       |
| 512-II-1  | 41585184       | 51198976       | 9475199        | 512-II-1  | 13446139       | 16981656       | 2581913.7<br>5  |
| 512-II-2  | 27708116       | 35935664       | 5901749.5      | 512-II-2  | 7992871.5      | 9915261        | 1620631.5       |
| 512-III-1 | 42557388       | 53514024       | 10226464       | 512-III-1 | 13088626       | 16737140       | 2907700.2<br>5  |
| 512-III-2 | 31514252       | 40001596       | 6676421.5      | 512-III-2 | 9464363        | 10544336       | 1562744.3<br>75 |

Peptide

Nop5-2

GIYAFDENG

LIEK

| Light     | y11 - 1249.6059+ | y10 - 1178.5688+ | y9 - 1031.5004+ | Heavy y   | y11 - 1257.6201+ | y10 - 1186.5830+ | y9 - 1039.5146+ |
|-----------|------------------|------------------|-----------------|-----------|------------------|------------------|-----------------|
| Replicate | Peak Area        | Peak Area        | Peak Area       | Replicate | Peak Area        | Peak Area        | Peak Area       |
| 512-I-1   | 288455.6563      | 536792.75        | 419980.65<br>63 | 512-I-1   | 139710.343<br>8  | 270883.062<br>5  | 224257.70<br>31 |
| 512-I-2   | 49084.94141      | 89077.7734<br>4  | 73006.078<br>13 | 512-I-2   | 28063.3300<br>8  | 56561.5898<br>4  | 48297.925<br>78 |
| 512-II-1  | 850992           | 1507529          | 1182574.2<br>5  | 512-II-1  | 330857.343<br>8  | 649971.562<br>5  | 535440.06<br>25 |
| 512-II-2  | 527155.6875      | 935327           | 782066          | 512-II-2  | 239107.859<br>4  | 472913.718<br>8  | 394709.18<br>75 |
| 512-III-1 | 427242.9688      | 777492.687<br>5  | 580980.81<br>25 | 512-III-1 | 201648.906<br>3  | 392940.906<br>3  | 309112.12<br>5  |
| 512-III-2 | 142101.0156      | 267060.468<br>8  | 212426.93<br>75 | 512-III-2 | 76722.9531<br>3  | 154957.187<br>5  | 120301.42<br>97 |

Peptide

Fib-1

FPGVYVIDD

DGSEK

| Light     | y11 - 1239.5739+ | y10 - 1076.5106+ | y9 - 977.4422+  | Heavy y   | y11 - 1247.5881+ | y10 - 1084.5248+ | y9 - 985.4564+  |
|-----------|------------------|------------------|-----------------|-----------|------------------|------------------|-----------------|
| Replicate | Peak Area        | Peak Area        | Peak Area       | Replicate | Peak Area        | Peak Area        | Peak Area       |
| 512-I-1   | 772139.9375      | 486338.718<br>8  | 541758.06<br>25 | 512-I-1   | 306008.312<br>5  | 166029.984<br>4  | 191086.81<br>25 |
| 512-I-2   | 93636.67969      | 58528.1054<br>7  | 69263.789<br>06 | 512-I-2   | 38859.8437<br>5  | 19792.6132<br>8  | 26341.792<br>97 |
| 512-II-1  | 1714431.375      | 1080759.25       | 1181952.8<br>75 | 512-II-1  | 564393.937<br>5  | 316753.187<br>5  | 340736.46<br>88 |
| 512-II-2  | 314580.5625      | 201690.328<br>1  | 234728.84<br>38 | 512-II-2  | 102678.335<br>9  | 62079.2773<br>4  | 73927.570<br>31 |
| 512-III-1 | 1112136.375      | 700237.062<br>5  | 767609.75       | 512-III-1 | 409144.718<br>8  | 228483.140<br>6  | 245507.54<br>69 |
| 512-III-2 | 166113.5         | 101820.843<br>8  | 108589.58<br>59 | 512-III-2 | 61601.7148<br>4  | 36306.9257<br>8  | 39453.25        |

Peptide  
Fib-2

IYGIEFSPR

| Light     | y7 - 805.4203+ | y6 - 748.3988+ | y5 - 635.3148+ | Heavy<br>y | y7 - 815.4285+ | y6 - 758.4071+ | y5 - 645.3230+ |
|-----------|----------------|----------------|----------------|------------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate  | Peak Area      | Peak Area      | Peak Area      |
| 512-I-1   | 35368660       | 3725464        | 21456084       | 512-I-1    | 15398816       | 1391050.625    | 8224544        |
| 512-I-2   | 29856154       | 2928787.25     | 16704483       | 512-I-2    | 13049221       | 1148704.875    | 6599191        |
| 512-II-1  | 39125184       | 4268537        | 23166916       | 512-II-1   | 15075005       | 1343045        | 7814665.5      |
| 512-II-2  | 30729306       | 3046301        | 16909588       | 512-II-2   | 11578057       | 983226.625     | 5862800        |
| 512-III-1 | 35765460       | 3695436.5      | 21121392       | 512-III-1  | 13288388       | 1214292.375    | 7323288        |
| 512-III-2 | 31735460       | 3207035.75     | 17857600       | 512-III-2  | 12135266       | 1055989.875    | 6358103        |

Peptide  
Snu13-1

NVPYVFVPSR

| Light     | y8 - 964.5251+ | y6 - 704.4090+ | y5 - 605.3406+ | Heavy<br>y | y8 - 974.5333+ | y6 - 714.4173+ | y5 - 615.3488+ |
|-----------|----------------|----------------|----------------|------------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate  | Peak Area      | Peak Area      | Peak Area      |
| 512-I-1   | 5387082.5      | 2511388.5      | 2644663.5      | 512-I-1    | 5562380        | 2511719.75     | 2602413        |
| 512-I-2   | 6759619        | 3051558        | 3206103.75     | 512-I-2    | 7032751.5      | 2900202.25     | 3048015.75     |
| 512-II-1  | 3986537.5      | 1978375.5      | 2034657.5      | 512-II-1   | 4570015        | 2190959.5      | 2273384.75     |
| 512-II-2  | 4670131.5      | 2096196.125    | 2226074.5      | 512-II-2   | 5512363        | 2363781.5      | 2483946.5      |
| 512-III-1 | 3795736.25     | 1768631.875    | 1842135        | 512-III-1  | 4399874.5      | 1958613.375    | 2083086.125    |
| 512-III-2 | 5118710        | 2331824.5      | 2396691.25     | 512-III-2  | 5833134        | 2611733        | 2718933        |

Peptide  
Snu13-2

TQIYAVK

| Light     | y6 - 721.4243+ | y5 - 593.3657+ | y4 - 480.2817+ | Heavy<br>y | y6 - 729.4385+ | y5 - 601.3799+ | y4 - 488.2959+ |
|-----------|----------------|----------------|----------------|------------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate  | Peak Area      | Peak Area      | Peak Area      |
| 512-I-1   | 1153893.625    | 16048096       | 4076456.25     | 512-I-1    | 379960.0313    | 5856634        | 1212080        |
| 512-I-2   | 699829.75      | 8778402        | 2009721.75     | 512-I-2    | 173134.75      | 3277733.5      | 684109.0625    |
| 512-II-1  | 1129505.375    | 16091167       | 3501365.25     | 512-II-1   | 344246.4063    | 5657475        | 1301607.875    |
| 512-II-2  | 477482.4063    | 7805873        | 1627868.5      | 512-II-2   | 203789.0625    | 2648077.25     | 559669.125     |
| 512-III-1 | 1243212        | 16312946       | 3771552.5      | 512-III-1  | 399438.2813    | 6459693.5      | 1415183.5      |
| 512-III-2 | 743356.375     | 9148045        | 2337801.5      | 512-III-2  | 222259.1406    | 3810967.5      | 792178.8125    |

snR41

Peptide  
Nop5-1

AFISENVR

| Light     | y6 - 717,3890+ | y5 - 604,3049+ | y4 - 517,2729+ | Heavy<br>y | y6 - 727,3972+ | y5 - 614,3132+ | y4 - 527,2812+ |
|-----------|----------------|----------------|----------------|------------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate  | Peak Area      | Peak Area      | Peak Area      |
| 41A-I     | 75309864       | 333823840      | 65797268       | 41A-I      | 51343640       | 177563312      | 28971648       |
| 41A-II    | 101627128      | 400342080      | 81051000       | 41A-II     | 51493772       | 171778240      | 28113496       |



|         |           |           |               |         |          |           |          |
|---------|-----------|-----------|---------------|---------|----------|-----------|----------|
| 41A-III | 102018696 | 467224864 | 10239968<br>8 | 41A-III | 55100232 | 186393680 | 33510178 |
| 41B-I   | 106683440 | 454314432 | 79772240      | 41B-I   | 59858936 | 199769024 | 33658264 |
| 41B-II  | 108373704 | 522730432 | 11183892<br>8 | 41B-II  | 57553936 | 179144656 | 30191122 |
| 41B-III | 68077296  | 353060960 | 71123560      | 41B-III | 49327404 | 171605824 | 27544770 |

Peptide  
Nop5-2      GIYAFDENG  
LIEK

| Light     | y11 -<br>1249,6059+ | y10 -<br>1178,5688+ | y9 -<br>1031,5004<br>+ | Heavy<br>y | y11 -<br>1257,6201+ | y10 -<br>1186,5830+ | y9 -<br>1039,5146<br>+ |
|-----------|---------------------|---------------------|------------------------|------------|---------------------|---------------------|------------------------|
| Replicate | Peak Area           | Peak Area           | Peak Area              | Replicate  | Peak Area           | Peak Area           | Peak Area              |
| 41A-I     | 173761.0625         | 1535814.25          | 3125863                | 41A-I      | 130135.0625         | 589312.125          | 3775074.75             |
| 41A-II    | 133761.3438         | 1317673.375         | 2773538.5              | 41A-II     | 54912.51172         | 668930              | 3239332.75             |
| 41A-III   | 130733.0313         | 384845.5938         | 2573602                | 41A-III    | 130880.0625         | 439106.3125         | 2821390.5              |
| 41B-I     | 15129.90723         | 573885              | 2432261.5              | 41B-I      | 20539.16016         | 386350.7813         | 2578881.5              |
| 41B-II    | 87688.79688         | 1597079.5           | 2500243.75             | 41B-II     | 0                   | 1125910.5           | 3889222                |
| 41B-III   | 46602.63281         | 325139.5625         | 2195300.5              | 41B-III    | 58766.67578         | 816036.625          | 3229494.75             |

Peptide  
Fib-1      FPGVYVIDD  
DGSEK

| Light     | y11 -<br>1239,5739+ | y10 -<br>1076,5106+ | y9 -<br>977,4422+ | Heavy<br>y | y11 -<br>1247,5881+ | y10 -<br>1084,5248+ | y9 -<br>985,4564+ |
|-----------|---------------------|---------------------|-------------------|------------|---------------------|---------------------|-------------------|
| Replicate | Peak Area           | Peak Area           | Peak Area         | Replicate  | Peak Area           | Peak Area           | Peak Area         |
| 41A-I     | 4397637             | 6431509             | 10462816          | 41A-I      | 1939029.75          | 2504835.75          | 10215372          |
| 41A-II    | 5000244             | 7063053             | 14928142          | 41A-II     | 2487345.5           | 3643649.75          | 10796472          |
| 41A-III   | 5617591             | 10410457            | 17102426          | 41A-III    | 2522267.75          | 4461657.5           | 12693405          |
| 41B-I     | 5328174.5           | 8181008.5           | 13085890          | 41B-I      | 1854427.75          | 3749846.75          | 13201334          |
| 41B-II    | 6160545.5           | 10662796            | 16432651          | 41B-II     | 3205612             | 4097920.25          | 13340327          |
| 41B-III   | 6172615.5           | 12194808            | 19748794          | 41B-III    | 3186597.75          | 4451154             | 13681541          |

Peptide  
Fib-2      IYGIEFSPR

| Light     | y7 - 805,4203+ | y6 -<br>748,3988+ | y5 -<br>635,3148+ | Heavy<br>y | y7 -<br>815,4285+ | y6 -<br>758,4071+ | y5 -<br>645,3230+ |
|-----------|----------------|-------------------|-------------------|------------|-------------------|-------------------|-------------------|
| Replicate | Peak Area      | Peak Area         | Peak Area         | Replicate  | Peak Area         | Peak Area         | Peak Area         |
| 41A-I     | 321440416      | 35143076          | 226315872         | 41A-I      | 109388400         | 24497616          | 140574224         |
| 41A-II    | 339209504      | 32451908          | 225181216         | 41A-II     | 99150432          | 19416836          | 105273944         |
| 41A-III   | 411055232      | 48970540          | 282820608         | 41A-III    | 114214104         | 25042094          | 117167704         |
| 41B-I     | 316879232      | 31587258          | 200813904         | 41B-I      | 114810272         | 22067370          | 135804032         |
| 41B-II    | 444803424      | 48426512          | 317459776         | 41B-II     | 119187704         | 28290644          | 149976064         |
| 41B-III   | 354352032      | 39779152          | 243599136         | 41B-III    | 106594760         | 23422250          | 135866432         |

Peptide  
Snu13-1      NVPYVFPSPR

| Light     | y8 - 964,5251+ | y6 -<br>704,4090+ | y5 -<br>605,3406+ | Heavy<br>y | y8 -<br>974,5333+ | y6 -<br>714,4173+ | y5 -<br>615,3488+ |
|-----------|----------------|-------------------|-------------------|------------|-------------------|-------------------|-------------------|
| Replicate | Peak Area      | Peak Area         | Peak Area         | Replicate  | Peak Area         | Peak Area         | Peak Area         |

|         |           |           |               |         |           |           |               |
|---------|-----------|-----------|---------------|---------|-----------|-----------|---------------|
| 41A-I   | 45891644  | 70517080  | 13398409<br>6 | 41A-I   | 121297784 | 417380544 | 30549193<br>6 |
| 41A-II  | 67523712  | 119775696 | 19602235<br>2 | 41A-II  | 118367352 | 452556160 | 32816064<br>0 |
| 41A-III | 131156312 | 199431792 | 38391782<br>4 | 41A-III | 107726576 | 418327104 | 31174457<br>6 |
| 41B-I   | 94748616  | 157788080 | 28049574<br>4 | 41B-I   | 159068320 | 477505184 | 34747145<br>6 |
| 41B-II  | 164796224 | 294170304 | 50041785<br>6 | 41B-II  | 135164016 | 515588064 | 36550716<br>8 |
| 41B-III | 143575728 | 210092896 | 39382640<br>0 | 41B-III | 118686776 | 419739232 | 30552236<br>8 |

Peptide  
Snu13-2 TQIYAVK

| Light     | y6 - 721,4243+ | y5 - 593,3657+ | y4 - 480,2817+ | Heavy<br>y | y6 - 729,4385+ | y5 - 601,3799+ | y4 - 488,2959+ |
|-----------|----------------|----------------|----------------|------------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate  | Peak Area      | Peak Area      | Peak Area      |
| 41A-I     | 0              | 2438195.75     | 23182434       | 41A-I      | 0              | 3565954.25     | 25118628       |
| 41A-II    | 0              | 2449489.75     | 31201368       | 41A-II     | 0              | 5483715        | 24989856       |
| 41A-III   | 0              | 4131985        | 68217208       | 41A-III    | 0              | 3359193.25     | 28766100       |
| 41B-I     | 0              | 5828788        | 67843608       | 41B-I      | 0              | 4870809        | 29937274       |
| 41B-II    | 0              | 7818964        | 98692072       | 41B-II     | 0              | 3972406.75     | 31112688       |
| 41B-III   | 0              | 6996492.5      | 57588048       | 41B-III    | 0              | 4883784        | 22510290       |

snR54  
Peptide  
Nop5-1 AFISENVR

| Light     | y6 - 717.3890+ | y5 - 604.3049+ | y4 - 517.2729+ | Heavy<br>y | y6 - 727.3972+ | y5 - 614.3132+ | y4 - 527.2812+  |
|-----------|----------------|----------------|----------------|------------|----------------|----------------|-----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate  | Peak Area      | Peak Area      | Peak Area       |
| 54A-I-1   | 42617456       | 52732456       | 10076662       | 54A-I-1    | 12732946       | 16966388       | 2686277.7<br>5  |
| 54A-I-2   | 22628238       | 27707762       | 4086274.7<br>5 | 54A-I-2    | 5851981.5      | 7304951.5      | 1097541.7<br>5  |
| 54A-II-1  | 32093328       | 38419388       | 6758003.5      | 54A-II-1   | 13477044       | 17515304       | 2626200         |
| 54A-II-2  | 19111552       | 22597790       | 3718376.7<br>5 | 54A-II-2   | 7228269        | 9640143        | 1395509.8<br>75 |
| 54A-III-1 | 37790184       | 45495492       | 8448499        | 54A-III-1  | 12893288       | 16734974       | 2638462.5       |
| 54A-III-2 | 24242510       | 31082102       | 4999881.5      | 54A-III-2  | 7992905.5      | 10436461       | 1523628.1<br>25 |
| 54B-I-1   | 41772436       | 51008804       | 9246477        | 54B-I-1    | 12415172       | 16659763       | 2485553.2<br>5  |
| 54B-I-2   | 30830936       | 36101936       | 6527919        | 54B-I-2    | 8790312        | 11082347       | 1795619.6<br>25 |
| 54B-II-1  | 40720184       | 48626872       | 9535464        | 54B-II-1   | 14201371       | 18329560       | 2900008.7<br>5  |
| 54B-II-2  | 25610416       | 32185054       | 5236663.5      | 54B-II-2   | 8263282        | 10414519       | 1515865.8<br>75 |
| 54B-III-1 | 35761856       | 43313464       | 7409361.5      | 54B-III-1  | 12787103       | 16921152       | 2785208         |
| 54B-III-2 | 25873916       | 28864512       | 4911621.5      | 54B-III-2  | 8450376        | 10253140       | 1562910         |

Peptide  
Nop5-2 GIYAFDENG  
LIEK

| Light | y11 - 1249.6059+ | y10 - 1178.5688+ | y9 - 1031.5004<br>+ | Heavy<br>y | y11 - 1257.6201+ | y10 - 1186.5830+ | y9 - 1039.5146<br>+ |
|-------|------------------|------------------|---------------------|------------|------------------|------------------|---------------------|
|-------|------------------|------------------|---------------------|------------|------------------|------------------|---------------------|

| Replicate | Peak Area   | Peak Area   | Peak Area   | Replicate | Peak Area   | Peak Area   | Peak Area   |
|-----------|-------------|-------------|-------------|-----------|-------------|-------------|-------------|
| 54A-I-1   | 479255.3438 | 871781.9375 | 667606      | 54A-I-1   | 166642.5781 | 346997.6875 | 275742.1875 |
| 54A-I-2   | 166527.3906 | 297780.4063 | 248640.8281 | 54A-I-2   | 67978.82031 | 137947.2656 | 117947.3984 |
| 54A-II-1  | 328832.2813 | 560294.125  | 463834.4063 | 54A-II-1  | 226059.25   | 429845.8438 | 351953.625  |
| 54A-II-2  | 93427.07031 | 181987.0781 | 150009.375  | 54A-II-2  | 72902.52344 | 132570.7656 | 132342.9375 |
| 54A-III-1 | 486997.1563 | 862361.9375 | 679399      | 54A-III-1 | 236367.6094 | 473687.375  | 386273.375  |
| 54A-III-2 | 248669.25   | 427553.6875 | 354931.0625 | 54A-III-2 | 127762.7578 | 256664.7656 | 206537.3594 |
| 54B-I-1   | 413141.9375 | 722273.5    | 572781.875  | 54B-I-1   | 159308.5938 | 302628.3438 | 269097.3125 |
| 54B-I-2   | 86240.375   | 159453.5625 | 124804.1563 | 54B-I-2   | 42954.5     | 82077.47656 | 64431.75    |
| 54B-II-1  | 408517.6875 | 728794.5625 | 580711.5625 | 54B-II-1  | 181142.2344 | 353253.25   | 302980.8438 |
| 54B-II-2  | 114925.0313 | 217899.8594 | 169684.7813 | 54B-II-2  | 54345.51953 | 114780.0547 | 92764.32031 |
| 54B-III-1 | 353054.625  | 639767.75   | 504804.9063 | 54B-III-1 | 176061      | 355666.0313 | 289158.6875 |
| 54B-III-2 | 109016.7656 | 191561.2813 | 146757.3125 | 54B-III-2 | 61747.85547 | 121445.0469 | 98908.42188 |

Peptide FPGVYVVIDD  
Fib-1 DGSEK

| Light     | y11 - 1239.5739+ | y10 - 1076.5106+ | y9 - 977.4422+ | Heavy y   | y11 - 1247.5881+ | y10 - 1084.5248+ | y9 - 985.4564+ |
|-----------|------------------|------------------|----------------|-----------|------------------|------------------|----------------|
| Replicate | Peak Area        | Peak Area        | Peak Area      | Replicate | Peak Area        | Peak Area        | Peak Area      |
| 54A-I-1   | 708347.625       | 449696.4688      | 501841.2188    | 54A-I-1   | 246754.8281      | 142216.6719      | 148434.25      |
| 54A-I-2   | 70693.82813      | 49346.82031      | 51978.92578    | 54A-I-2   | 23460.98242      | 16124.24512      | 17347.03516    |
| 54A-II-1  | 775527.5         | 444480.375       | 518005         | 54A-II-1  | 339533.9063      | 184018.5625      | 210099.9375    |
| 54A-II-2  | 112563.8828      | 71827.26563      | 86586.01563    | 54A-II-2  | 55102.07031      | 30262.66992      | 35520.34375    |
| 54A-III-1 | 1396719.5        | 885876.5625      | 994942.1875    | 54A-III-1 | 510139.1875      | 283852.25        | 311941.4688    |
| 54A-III-2 | 205846.0781      | 141207.2188      | 154752.1875    | 54A-III-2 | 79778.5625       | 49851.24609      | 57597.13672    |
| 54B-I-1   | 884841.5625      | 558017.125       | 623386.1875    | 54B-I-1   | 250319.0625      | 144482.5156      | 155453.25      |
| 54B-I-2   | 106551.8203      | 78712.46094      | 81363.82813    | 54B-I-2   | 32707.09961      | 20900.23047      | 21045.22656    |
| 54B-II-1  | 1136596          | 745884           | 833636.25      | 54B-II-1  | 347446.8125      | 199487.8281      | 222497.6875    |
| 54B-II-2  | 150361.6563      | 96939.33594      | 115921.0469    | 54B-II-2  | 50251.79297      | 30689.85938      | 33736.05078    |
| 54B-III-1 | 1237132.125      | 820391           | 862709.3125    | 54B-III-1 | 407627           | 235578.125       | 266876.6875    |
| 54B-III-2 | 143339.5313      | 100959.0938      | 114070.9297    | 54B-III-2 | 52136.64844      | 31733.55078      | 37036.28516    |

Peptide IYGIEFSPR  
Fib-2

| Light     | y7 - 805.4203+ | y6 - 748.3988+ | y5 - 635.3148+ | Heavy y   | y7 - 815.4285+ | y6 - 758.4071+ | y5 - 645.3230+ |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 54A-I-1   | 36303596       | 3877272.5      | 22145110       | 54A-I-1   | 13361524       | 1181039.125    | 6851464        |
| 54A-I-2   | 31533458       | 3102514        | 17873758       | 54A-I-2   | 11238607       | 1023054.875    | 5739319.5      |

|           |          |            |          |           |          |             |           |
|-----------|----------|------------|----------|-----------|----------|-------------|-----------|
| 54A-II-1  | 32055092 | 3246702.75 | 18926970 | 54A-II-1  | 15228915 | 1357500     | 8133690.5 |
| 54A-II-2  | 25934286 | 2475299.75 | 14566424 | 54A-II-2  | 12607065 | 1074673.75  | 6170643   |
| 54A-III-1 | 35862248 | 3649737.25 | 21154636 | 54A-III-1 | 13873814 | 1244551.25  | 7310749   |
| 54A-III-2 | 30817066 | 2931608.25 | 17044694 | 54A-III-2 | 11917407 | 1002373.313 | 6021340   |
| 54B-I-1   | 41774224 | 4560674.5  | 25659804 | 54B-I-1   | 13569314 | 1243070.875 | 7295886   |
| 54B-I-2   | 40307312 | 4165488.75 | 22565756 | 54B-I-2   | 12783251 | 1130072     | 6524030.5 |
| 54B-II-1  | 37426608 | 3991931    | 22482560 | 54B-II-1  | 13980136 | 1317796     | 7655738.5 |
| 54B-II-2  | 31959516 | 3049024    | 17727222 | 54B-II-2  | 11063642 | 1002534     | 5780797.5 |
| 54B-III-1 | 37346224 | 3984537.25 | 22350360 | 54B-III-1 | 14184388 | 1332355.75  | 7864277   |
| 54B-III-2 | 32488860 | 3293311.25 | 18103266 | 54B-III-2 | 12132060 | 1120443.25  | 6104241.5 |

Peptide  
Snu13-1

NVPYVFPVPSR

| Light     | y8 - 964.5251+ | y6 - 704.4090+ | y5 - 605.3406+ | Heavy     | y8 - 974.5333+ | y6 - 714.4173+ | y5 - 615.3488+ |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 54A-I-1   | 2369473        | 1122999        | 1182344        | 54A-I-1   | 4519089.5      | 2097190        | 2225305        |
| 54A-I-2   | 3430257        | 1585398.75     | 1602683.625    | 54A-I-2   | 6434527        | 2780955.5      | 2845058.75     |
| 54A-II-1  | 3396746.75     | 1606840.125    | 1716588.875    | 54A-II-1  | 6386375.5      | 2878307        | 3002545.5      |
| 54A-II-2  | 4252267.5      | 1975513.5      | 1998129.875    | 54A-II-2  | 8039533.5      | 3292087.5      | 3463719.25     |
| 54A-III-1 | 3865424        | 1847650.75     | 1935498        | 54A-III-1 | 5124678.5      | 2303106        | 2465440        |
| 54A-III-2 | 4435613        | 2060940.875    | 2135199.75     | 54A-III-2 | 5889639.5      | 2544276.75     | 2688797        |
| 54B-I-1   | 4938606.5      | 2319366.25     | 2361332.5      | 54B-I-1   | 5041036        | 2288523.5      | 2458542.25     |
| 54B-I-2   | 6642151        | 2913583.25     | 3038052.5      | 54B-I-2   | 6445183.5      | 2820847.5      | 3021668        |
| 54B-II-1  | 2771238        | 1325269.375    | 1318930.875    | 54B-II-1  | 4846595.5      | 2205212.5      | 2266007.5      |
| 54B-II-2  | 3361365.75     | 1497277.125    | 1634031.5      | 54B-II-2  | 5840611        | 2608274.25     | 2708536        |
| 54B-III-1 | 3231570.25     | 1531408.75     | 1552471.75     | 54B-III-1 | 4935770.5      | 2231314.5      | 2327661.75     |
| 54B-III-2 | 4696994        | 2016479.75     | 2080725.125    | 54B-III-2 | 6987500        | 2967733.5      | 3063066        |

Peptide  
Snu13-2

TQIYAVK

| Light     | y6 - 721.4243+ | y5 - 593.3657+ | y4 - 480.2817+ | Heavy     | y6 - 729.4385+ | y5 - 601.3799+ | y4 - 488.2959+ |
|-----------|----------------|----------------|----------------|-----------|----------------|----------------|----------------|
| Replicate | Peak Area      | Peak Area      | Peak Area      | Replicate | Peak Area      | Peak Area      | Peak Area      |
| 54A-I-1   | 495522.875     | 7070625        | 1601439.5      | 54A-I-1   | 403207.8125    | 6362945.5      | 1480541.375    |
| 54A-I-2   | 211818         | 2976981.5      | 729409         | 54A-I-2   | 180706.6406    | 2344736        | 533198.4375    |
| 54A-II-1  | 527171.375     | 7097862        | 1576646.75     | 54A-II-1  | 424355.1563    | 6001908        | 1454505.875    |
| 54A-II-2  | 230116.3594    | 3049636.75     | 665468.25      | 54A-II-2  | 186274.8125    | 2573258.5      | 546363.3125    |
| 54A-III-1 | 715373.4375    | 10146875       | 2042887.875    | 54A-III-1 | 389613.9688    | 5734013.5      | 1285836        |

|           |             |           |             |           |             |            |             |
|-----------|-------------|-----------|-------------|-----------|-------------|------------|-------------|
| 54A-III-2 | 439085.1563 | 5712210.5 | 1364094.5   | 54A-III-2 | 263107.0625 | 3707685    | 758979.375  |
| 54B-I-1   | 874911.4375 | 12897907  | 2804973.5   | 54B-I-1   | 368100.3438 | 5494206    | 1179880.5   |
| 54B-I-2   | 695210.875  | 9091490   | 1850899.25  | 54B-I-2   | 287301.375  | 3435085.25 | 788458      |
| 54B-II-1  | 603493.9375 | 8852266   | 2044613.875 | 54B-II-1  | 468057.125  | 7095271.5  | 1509697.375 |
| 54B-II-2  | 329409.625  | 4967366   | 1064388.25  | 54B-II-2  | 240449.9688 | 4018677.25 | 819402.3125 |
| 54B-III-1 | 652957.625  | 8870572   | 1838016.25  | 54B-III-1 | 366588.3438 | 6017691    | 1178617     |
| 54B-III-2 | 381036.4063 | 5000544   | 1156791.625 | 54B-III-2 | 218917.9219 | 3268451.25 | 799683.8125 |

## Appendix 2 – Enzymatic assay raw data

Please note that, if the counting efficiency E for a given assay shows the same value for all measurements within the assay, these assays were measured with a Pharmacia Wallac 1410 (Perkin Elmer) liquid scintillation counter without TDCR and the counting efficiency was determined using external calibration. Assays that show only DPM values were measured on a Tri-Carb liquid scintillation counter (Perkin Elmer), with TDCR and DPM as only read-out parameter. All other measurements were performed on a Hidex 300SL (Hidex) liquid scintillation counter with TDCR.

**sR26—L7Ae—Nop5—Fibrillar**in      Temperature: 30 °C  
BIO-D' and D

| time point (min) - replicate | CPM     | DPM     | efficiency |
|------------------------------|---------|---------|------------|
| 0-1                          | 75.20   | 313.33  | 0.24       |
| 0-2                          | 103.40  | 430.83  | 0.24       |
| 15-1                         | 423.60  | 1765.00 | 0.24       |
| 15-2                         | 328.50  | 1368.75 | 0.24       |
| 30-1                         | 566.20  | 2359.17 | 0.24       |
| 30-2                         | 589.30  | 2455.42 | 0.24       |
| 60-1                         | 906.20  | 3775.83 | 0.24       |
| 60-2                         | 964.20  | 4017.50 | 0.24       |
| 90-1                         | 1052.60 | 4385.83 | 0.24       |
| 90-2                         | 991.90  | 4132.92 | 0.24       |
| 120-1                        | 1236.60 | 5152.50 | 0.24       |
| 120-2                        | 1259.90 | 5249.58 | 0.24       |
| 150-1                        | 1582.20 | 6592.50 | 0.24       |
| 150-2                        | 1790.10 | 7458.75 | 0.24       |
| 180-1                        | 1324.10 | 5517.08 | 0.24       |
| 180-2                        | 1792.60 | 7469.17 | 0.24       |

**sR26—Snu13—Nop5—Fibrillar**in      Temperature: 30 °C  
BIO-D' and D

| time point (min) - replicate | CPM     | DPM      | efficiency |
|------------------------------|---------|----------|------------|
| 0-1                          | 57.30   | 286.50   | 0.20       |
| 0-2                          | 99.50   | 497.50   | 0.20       |
| 15-1                         | 657.80  | 3289.00  | 0.20       |
| 15-2                         | 630.60  | 3153.00  | 0.20       |
| 30-1                         | 1125.30 | 5626.50  | 0.20       |
| 30-2                         | 981.90  | 4909.50  | 0.20       |
| 60-1                         | 1477.60 | 7388.00  | 0.20       |
| 60-2                         | 1450.40 | 7252.00  | 0.20       |
| 80-1                         | 1867.40 | 9337.00  | 0.20       |
| 80-2                         | 1867.40 | 9337.00  | 0.20       |
| 100-1                        | 1952.40 | 9762.00  | 0.20       |
| 100-2                        | 2018.30 | 10091.50 | 0.20       |

|       |         |          |      |
|-------|---------|----------|------|
| 120-1 | 2126.50 | 10632.50 | 0.20 |
| 120-2 | 2140.80 | 10704.00 | 0.20 |

**snR51—L7Ae—Nop5—Fibrillarin**

Temperature: 50 °C

BIO-D' and D

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 264.99 | 904.35  | 0.29       |
| 0-2                          | 176.97 | 445.72  | 0.40       |
| 5-1                          | 271.96 | 687.07  | 0.40       |
| 5-2                          | 281.95 | 946.63  | 0.30       |
| 15-1                         | 261.94 | 908.73  | 0.29       |
| 15-2                         | 340.96 | 1072.71 | 0.32       |
| 30-1                         | 335.96 | 958.76  | 0.35       |
| 30-2                         | 293.94 | 1032.43 | 0.29       |
| 60-1                         | 276.95 | 878.41  | 0.32       |
| 60-2                         | 296.94 | 882.67  | 0.34       |
| 80-1                         | 310.94 | 1035.48 | 0.30       |
| 80-2                         | 256.94 | 788.96  | 0.33       |
| 100-1                        | 377.96 | 1238.34 | 0.31       |
| 100-2                        | 294.95 | 950.80  | 0.31       |
| 120-1                        | 328.95 | 958.12  | 0.34       |
| 120-2                        | 388.95 | 1137.51 | 0.34       |

BIO-D and D'

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 218.94 | 602.52  | 0.36       |
| 0-2                          | 238.94 | 596.41  | 0.40       |
| 5-1                          | 309.93 | 880.87  | 0.35       |
| 5-2                          | 293.95 | 852.69  | 0.35       |
| 15-1                         | 259.96 | 844.84  | 0.31       |
| 15-2                         | 245.94 | 764.42  | 0.32       |
| 30-1                         | 316.95 | 981.15  | 0.32       |
| 30-2                         | 280.94 | 744.14  | 0.38       |
| 60-1                         | 308.93 | 940.95  | 0.33       |
| 60-2                         | 283.95 | 719.59  | 0.40       |
| 80-1                         | 308.93 | 885.23  | 0.35       |
| 80-2                         | 305.95 | 1178.39 | 0.26       |
| 100-1                        | 369.93 | 1257.33 | 0.29       |
| 100-2                        | 359.93 | 936.54  | 0.38       |
| 120-1                        | 460.94 | 1337.10 | 0.35       |
| 120-2                        | 396.94 | 1388.46 | 0.29       |

**snR51—Snu13—Nop5—Fibrillarin**

Temperature: 37 °C

BIO-D'

| time point (min) - replicate | DPM    |
|------------------------------|--------|
| 0-1                          | 177.00 |

|       |        |
|-------|--------|
| 0-2   | 150.00 |
| 5-1   | 228.00 |
| 5-2   | 145.00 |
| 15-1  | 204.00 |
| 15-2  | 141.00 |
| 30-1  | 270.00 |
| 30-2  | 261.00 |
| 60-1  | 124.00 |
| 60-2  | 81.00  |
| 120-1 | 205.00 |
| 120-2 | 199.00 |
| 240-1 | 226.00 |
| 240-2 | 207.00 |

BIO-D

| time point (min) - replicate | DPM    |
|------------------------------|--------|
| 0-1                          | 284.00 |
| 0-2                          | 196.00 |
| 5-1                          | 230.00 |
| 5-2                          | 160.00 |
| 15-1                         | 265.00 |
| 15-2                         | 163.00 |
| 30-1                         | 242.00 |
| 30-2                         | 228.00 |
| 60-1                         | 262.00 |
| 60-2                         | 212.00 |
| 120-1                        | 166.00 |
| 120-2                        | 163.00 |
| 240-1                        | 139.00 |
| 240-2                        | 110.00 |

**snR51-2—L7Ae—Nop5—Fibrillarin** Temperature: 50 °C

BIO-D' and D

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 29.00  | 96.97   | 0.30       |
| 0-2                          | 29.00  | 78.87   | 0.37       |
| 5-1                          | 34.99  | 115.11  | 0.30       |
| 5-2                          | 32.99  | 78.75   | 0.42       |
| 15-1                         | 20.00  | 69.10   | 0.29       |
| 15-2                         | 42.99  | 170.66  | 0.25       |
| 30-1                         | 24.00  | 58.90   | 0.41       |
| 30-2                         | 26.99  | 50.70   | 0.53       |
| 60-1                         | 316.99 | 1499.75 | 0.21       |
| 60-2                         | 33.00  | 111.25  | 0.30       |
| 80-1                         | 520.97 | 1699.97 | 0.31       |
| 80-2                         | 237.95 | 932.38  | 0.26       |
| 100-1                        | 80.98  | 270.78  | 0.30       |



|       |       |        |      |
|-------|-------|--------|------|
| 100-2 | 31.99 | 72.99  | 0.44 |
| 120-1 | 31.99 | 67.33  | 0.48 |
| 120-2 | 40.99 | 148.83 | 0.28 |

BIO-D and D'

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 56.99  | 145.82  | 0.39       |
| 0-2                          | 43.99  | 127.09  | 0.35       |
| 5-1                          | 22.99  | 49.84   | 0.46       |
| 5-2                          | 33.99  | 82.33   | 0.41       |
| 15-1                         | 36.99  | 86.93   | 0.43       |
| 15-2                         | 71.99  | 186.04  | 0.39       |
| 30-1                         | 212.97 | 633.07  | 0.34       |
| 30-2                         | 44.99  | 113.66  | 0.40       |
| 60-1                         | 201.97 | 1106.78 | 0.18       |
| 60-2                         | 31.99  | 139.71  | 0.23       |
| 80-1                         | 54.98  | 162.77  | 0.34       |
| 80-2                         | 28.00  | 62.69   | 0.45       |
| 100-1                        | 49.99  | 130.53  | 0.38       |
| 100-2                        | 45.98  | 144.66  | 0.32       |
| 120-1                        | 51.99  | 170.34  | 0.31       |
| 120-2                        | 30.99  | 78.05   | 0.40       |

**snR51-2—Snu13—Nop5—Fibrillarin** Temperature: 30 °C

BIO-D' and D

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 19.99  | 66.57   | 0.30       |
| 0-2                          | 39.99  | 123.29  | 0.32       |
| 5-1                          | 20.99  | 47.07   | 0.45       |
| 5-2                          | 49.98  | 114.67  | 0.44       |
| 15-1                         | 22.99  | 75.63   | 0.30       |
| 15-2                         | 35.99  | 96.04   | 0.38       |
| 30-1                         | 44.99  | 151.67  | 0.30       |
| 30-2                         | 498.00 | 6926.45 | 0.07       |
| 60-1                         | 13.99  | 27.54   | 0.51       |
| 60-2                         | 140.96 | 1519.14 | 0.09       |
| 80-1                         | 50.98  | 76.73   | 0.66       |
| 80-2                         | 242.95 | 1051.01 | 0.23       |
| 100-1                        | 113.96 | 616.35  | 0.19       |
| 100-2                        | 37.99  | 146.32  | 0.26       |
| 120-1                        | 69.97  | 349.69  | 0.20       |
| 120-2                        | 24.99  | 73.38   | 0.34       |

BIO-D and D'

| time point (min) - replicate | CPM    | DPM    | efficiency |
|------------------------------|--------|--------|------------|
| 0-1                          | 115.96 | 351.77 | 0.33       |
| 0-2                          | 79.97  | 222.73 | 0.36       |
| 5-1                          | 55.98  | 155.91 | 0.36       |

|       |        |         |      |
|-------|--------|---------|------|
| 5-2   | 45.98  | 119.23  | 0.39 |
| 15-1  | 193.97 | 1249.62 | 0.16 |
| 15-2  | 19.99  | 55.68   | 0.36 |
| 30-1  | 20.99  | 65.77   | 0.32 |
| 30-2  | 30.98  | 79.01   | 0.39 |
| 60-1  | 37.98  | 83.00   | 0.46 |
| 60-2  | 7.99   | 22.25   | 0.36 |
| 80-1  | 93.96  | 258.57  | 0.36 |
| 80-2  | 61.98  | 252.64  | 0.25 |
| 100-1 | 51.98  | 157.68  | 0.33 |
| 100-2 | 33.98  | 130.88  | 0.26 |
| 120-1 | 74.97  | 248.65  | 0.30 |
| 120-2 | 34.98  | 119.38  | 0.29 |

**snR41-2—L7Ae—Nop5—Fibrillarin**      Temperature: 50 °C  
 BIO-D' and BIO-D

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 37.10  | 154.58  | 0.24       |
| 0-2                          | 20.90  | 87.08   | 0.24       |
| 5-1                          | 41.30  | 172.08  | 0.24       |
| 5-2                          | 152.10 | 633.75  | 0.24       |
| 15-1                         | 110.20 | 459.17  | 0.24       |
| 15-2                         | 71.10  | 296.25  | 0.24       |
| 30-1                         | 261.30 | 1088.75 | 0.24       |
| 30-2                         | 180.80 | 753.33  | 0.24       |
| 60-1                         | 243.50 | 1014.58 | 0.24       |
| 60-2                         | 248.70 | 1036.25 | 0.24       |
| 90-1                         | 522.20 | 2175.83 | 0.24       |
| 90-2                         | 411.00 | 1712.50 | 0.24       |
| 120-1                        | 479.40 | 1997.50 | 0.24       |
| 120-2                        | 510.30 | 2126.25 | 0.24       |

**snR41-2—Snu13—Nop5—Fibrillarin**      Temperature: 30 °C  
 BIO-D' and BIO-D

| time point (min) - replicate | CPM   | DPM    | efficiency |
|------------------------------|-------|--------|------------|
| 0-1                          | 32.40 | 135.00 | 0.24       |
| 0-2                          | 87.80 | 365.83 | 0.24       |
| 15-1                         | 62.70 | 261.25 | 0.24       |
| 15-2                         | 68.90 | 287.08 | 0.24       |
| 30-1                         | 98.70 | 411.25 | 0.24       |
| 30-2                         | 64.80 | 270.00 | 0.24       |
| 60-1                         | 81.00 | 337.50 | 0.24       |
| 60-2                         | 59.50 | 247.92 | 0.24       |
| 90-1                         | 69.50 | 289.58 | 0.24       |
| 90-2                         | 80.40 | 335.00 | 0.24       |

|       |       |        |      |
|-------|-------|--------|------|
| 120-1 | 82.00 | 341.67 | 0.24 |
| 120-2 | 92.90 | 387.08 | 0.24 |
| 150-1 | 78.40 | 326.67 | 0.24 |
| 150-2 | 58.50 | 243.75 | 0.24 |
| 180-1 | 98.20 | 409.17 | 0.24 |
| 180-2 | 66.90 | 278.75 | 0.24 |

**snR41-3—L7Ae—Nop5—Fibrillarin**      Temperature: 50 °C

BIO-D' and BIO-D

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 322.80 | 1345.00 | 0.24       |
| 0-2                          | 324.40 | 1351.67 | 0.24       |
| 5-1                          | 192.70 | 802.92  | 0.24       |
| 5-2                          | 128.00 | 533.33  | 0.24       |
| 15-1                         | 327.00 | 1362.50 | 0.24       |
| 15-2                         | 279.50 | 1164.58 | 0.24       |
| 30-1                         | 322.80 | 1345.00 | 0.24       |
| 30-2                         | 191.60 | 798.33  | 0.24       |
| 60-1                         | 432.40 | 1801.67 | 0.24       |
| 60-2                         | 432.40 | 1801.67 | 0.24       |
| 90-1                         | 414.10 | 1725.42 | 0.24       |
| 90-2                         | 512.80 | 2136.67 | 0.24       |

**snR41-3—Snu13—Nop5—Fibrillarin**      Temperature: 30 °C

BIO-D' and BIO-D

| time point (min) - replicate | CPM     | DPM     | efficiency |
|------------------------------|---------|---------|------------|
| 0-1                          | 99.70   | 415.42  | 0.24       |
| 0-2                          | 65.30   | 272.08  | 0.24       |
| 15-1                         | 51.20   | 213.33  | 0.24       |
| 15-2                         | 75.20   | 313.33  | 0.24       |
| 30-1                         | 35.50   | 147.92  | 0.24       |
| 30-2                         | 1398.90 | 5828.75 | 0.24       |
| 60-1                         | 70.00   | 291.67  | 0.24       |
| 60-2                         | 75.70   | 315.42  | 0.24       |
| 90-1                         | 61.60   | 256.67  | 0.24       |
| 90-2                         | 60.60   | 252.50  | 0.24       |
| 120-1                        | 53.30   | 222.08  | 0.24       |
| 120-2                        | 66.80   | 278.33  | 0.24       |
| 150-1                        | 108.10  | 450.42  | 0.24       |
| 150-2                        | 106.00  | 441.67  | 0.24       |
| 180-1                        | 42.80   | 178.33  | 0.24       |
| 180-2                        | 93.00   | 387.50  | 0.24       |

**snR41-4—L7Ae—Nop5—Fibrillarin**      Temperature: 50 °C

BIO-D' and BIO-D

| time point (min) - replicate | CPM     | DPM      | efficiency |
|------------------------------|---------|----------|------------|
| 0-1                          | 43.90   | 182.92   | 0.24       |
| 0-2                          | 160.90  | 670.42   | 0.24       |
| 5-1                          | 187.50  | 781.25   | 0.24       |
| 5-2                          | 287.20  | 1196.67  | 0.24       |
| 15-1                         | 593.90  | 2474.58  | 0.24       |
| 15-2                         | 767.70  | 3198.75  | 0.24       |
| 30-1                         | 1183.00 | 4929.17  | 0.24       |
| 30-2                         | 1273.90 | 5307.92  | 0.24       |
| 60-1                         | 2055.10 | 8562.92  | 0.24       |
| 60-2                         | 1980.80 | 8253.33  | 0.24       |
| 90-1                         | 2569.10 | 10704.58 | 0.24       |
| 90-2                         | 2372.60 | 9885.83  | 0.24       |
| 120-1                        | 2559.70 | 10665.42 | 0.24       |
| 120-2                        | 2666.80 | 11111.67 | 0.24       |

**snR41-4—Snu13—Nop5—Fibrillarin** Temperature: 30 °C

BIO-D' and BIO-D

| time point (min) - replicate | CPM    | DPM     | efficiency |
|------------------------------|--------|---------|------------|
| 0-1                          | 65.80  | 274.17  | 0.24       |
| 0-2                          | 78.30  | 326.25  | 0.24       |
| 15-1                         | 80.40  | 335.00  | 0.24       |
| 15-2                         | 54.30  | 226.25  | 0.24       |
| 30-1                         | 80.40  | 335.00  | 0.24       |
| 30-2                         | 46.50  | 193.75  | 0.24       |
| 60-1                         | 136.80 | 570.00  | 0.24       |
| 60-2                         | 110.20 | 459.17  | 0.24       |
| 90-1                         | 113.80 | 474.17  | 0.24       |
| 90-2                         | 111.70 | 465.42  | 0.24       |
| 120-1                        | 76.80  | 320.00  | 0.24       |
| 120-2                        | 65.80  | 274.17  | 0.24       |
| 150-1                        | 73.10  | 304.58  | 0.24       |
| 150-2                        | 64.80  | 270.00  | 0.24       |
| 180-1                        | 414.20 | 1725.83 | 0.24       |
| 180-2                        | 67.40  | 280.83  | 0.24       |

**snR54—L7Ae—Nop5—Fibrillarin** Temperature: 50 °C

BIO-D'

| time point (min) - replicate | CPM    | DPM    | efficiency |
|------------------------------|--------|--------|------------|
| 0-1                          | 45.90  | 191.25 | 0.24       |
| 0-2                          | 54.70  | 227.92 | 0.24       |
| 5-1                          | 65.10  | 271.25 | 0.24       |
| 5-2                          | 108.40 | 451.67 | 0.24       |
| 10-1                         | 54.20  | 225.83 | 0.24       |

|       |        |        |      |
|-------|--------|--------|------|
| 10-2  | 75.00  | 312.50 | 0.24 |
| 15-1  | 55.20  | 230.00 | 0.24 |
| 15-2  | 52.10  | 217.08 | 0.24 |
| 30-1  | 79.20  | 330.00 | 0.24 |
| 30-2  | 84.90  | 353.75 | 0.24 |
| 60-1  | 61.50  | 256.25 | 0.24 |
| 60-2  | 97.90  | 407.92 | 0.24 |
| 80-1  | 59.40  | 247.50 | 0.24 |
| 80-2  | 51.10  | 212.92 | 0.24 |
| 100-1 | 133.40 | 555.83 | 0.24 |
| 100-2 | 106.30 | 442.92 | 0.24 |

**snR54—Snu13—Nop5—Fibrillarin**      Temperature: 30 °C  
 BIO-D'

| time point (min) - replicate | DPM    |
|------------------------------|--------|
| 0-1                          | 183.00 |
| 0-2                          | 296.00 |
| 30-1                         | 292.00 |
| 30-2                         | 168.00 |
| 60-1                         | 154.00 |
| 60-2                         | 312.00 |
| 120-1                        | 270.00 |
| 120-2                        | 287.00 |
| 240-1                        | 375.00 |
| 240-2                        | 134.00 |

**snR41-2—Snu13—Nop5-56—Nop1<sup>83-327</sup>**      Temperature: 30 °C  
 BIO-D' and BIO-D

| time point (min) - replicate | CPM   | DPM     | efficiency |
|------------------------------|-------|---------|------------|
| 0-1                          | 113.9 | 474.58  | 0.24       |
| 0-2                          | 79.1  | 329.58  | 0.24       |
| 5-1                          | 232.5 | 968.75  | 0.24       |
| 5-2                          | 179.0 | 745.83  | 0.24       |
| 15-1                         | 242.4 | 1010.00 | 0.24       |
| 15-2                         | 252.3 | 1051.25 | 0.24       |
| 30-1                         | 322.0 | 1341.67 | 0.24       |
| 30-2                         | 387.1 | 1612.92 | 0.24       |
| 60-1                         | 153.5 | 639.58  | 0.24       |
| 60-2                         | 267.4 | 1114.17 | 0.24       |
| 90-1                         | 151.9 | 632.92  | 0.24       |
| 90-2                         | 252.3 | 1051.25 | 0.24       |
| 120-1                        | 99.9  | 416.25  | 0.24       |
| 120-2                        | 169.6 | 706.67  | 0.24       |

**snR54—Snu13—Nop5-56—Nop1<sup>83-327</sup>**      Temperature: 30 °C

BIO-D' and BIO-D

| time point (min) - replicate | CPM    | DPM    | efficiency |
|------------------------------|--------|--------|------------|
| 0-1                          | 204.80 | 853.33 | 0.24       |
| 0-2                          | 224.50 | 935.42 | 0.24       |
| 5-1                          | 194.40 | 810.00 | 0.24       |
| 5-2                          | 129.40 | 539.17 | 0.24       |
| 15-1                         | 80.00  | 333.33 | 0.24       |
| 15-2                         | 129.40 | 539.17 | 0.24       |
| 30-1                         | 107.60 | 448.33 | 0.24       |
| 30-2                         | 137.20 | 571.67 | 0.24       |
| 60-1                         | 86.30  | 359.58 | 0.24       |
| 60-2                         | 57.20  | 238.33 | 0.24       |
| 90-1                         | 95.60  | 398.33 | 0.24       |
| 90-2                         | 121.60 | 506.67 | 0.24       |
| 120-1                        | 69.60  | 290.00 | 0.24       |
| 120-2                        | 83.20  | 346.67 | 0.24       |



|                   |   |
|-------------------|---|
| snR51-kl2-Reverse | 5'<br>AGCTTCTGCAGGTCATCAATTAGCCTTTCGACCAATCTCGTACTATAGT<br>GACTCGTATTAG   |
| snR41-2-Forward   | 5'<br>AATTCTAATACGACTCACTATAGGCCCTATGATTACCAATTGTTCCGA<br>CTGAAATAGTAGATGTAAGTTTCAGCCTTGTATGAGGGGCCCTGCAGA            |
| snR41-2-Reverse   | 5'<br>AGCTTCTGCAGGGCCCTCATACAAGGCTGAAACTTACATCTACTATTT<br>CAGTCGGAACAATTGGTAATCATAGGGCCTATAGTGAGTCGTATTAG             |
| snR41-3-Forward   | 5'<br>AATTCTAATACGACTCACTATAGGCCCTATGATTACCAATTGTTCCGA<br>CTGAATAGTGGTTTAACTACATGTGGTTTCAGCCTTGTATGAGGGGCC<br>CTGCAGA |
| snR41-3-Reverse   | 5'<br>AGCTTCTGCAGGGCCCTCATACAAGGCTGAAACACATGTAGTTAAAC<br>CACTATTCAGTCGGAACAATTGGTAATCATAGGGCCTATAGTGAGTC<br>GTATTAG   |
| snR41-4-Forward   | 5'<br>AATTCTAATACGACTCACTATAGGCCCTATGATTACCAATTGTTCCCG<br>ACTGAAATAGTAGATGTGTTTCAGCCTTGTATGAGGGGCCCTGCAGA             |
| snR41-4-Reverse   | 5'<br>AGCTTCTGCAGGGCCCTCATACAAGGCTGAAACACATCTACTATTTCA<br>GTCGGAACAATTGGTAATCATAGGGCCTATAGTGAGTCGTATTAG               |



# Acknowledgments

I want to use this last section of my thesis to thank all the great people that supported and joined me during this journey and without whom I would have never reached this point.

First and foremost, I want to thank my supervisor Prof. Dr. Teresa Carlomagno for supporting me through these five years and never giving up on me or this project also in difficult times. Thank you for being my mentor and for always having an open ear to talk, I learned a lot from you. Furthermore, I want to thank Prof. Dr. Wulf Blankenfeldt and Prof. Dr. Claus-Peter Witte for being part of my thesis committee and taking an interest in my research.

I want to thank Junior-Prof. Dr. Alexander Marchanka for helping me get started in the RNA world and continuous advice and support. Petra Baruch and Dr. Roman Fedorov I want to thank for an amazing collaboration and teaching me so much about biomolecular crystallization, I am really glad I got to know you. On a similar note, I want to thank Dr. Peer Lukat also for great collaboration, it was very nice to work with you. Ulrike Schrameck and Prof. Dr. Andreas Pich, I also want to thank for a great collaboration, fancy mass spectrometry and inspiring conversations. Additionally, I want to thank Dr. Luca Codutti for providing a smoothly running lab environment. I want to thank all past and present members of the Carlomagno group for providing a challenging and inspiring work atmosphere.

On a more personal note I want to thank Marco Aurélio Wust, Dr. Juliane Buschmann, Dr. Neha Dhimole and Renate for their constant emotional support, for always taking time and for becoming very important people to me.

Finally, I want to thank my mother, my sister, my father and Christian and all members of my family, who believe in me and support me since day one.

## Bibliography

- Abeyrathne, P.D., Koh, C.S., Grant, T., Grigorieff, N., and Korostelev, A.A. (2016). Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome. *ELife* 5, e14874.
- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* 66, 213–221.
- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* 68, 352–367.
- Ahmed, M., Marchanka, A., and Carlomagno, T. (2020). Structure of a Protein-RNA Complex by Solid-State NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.* 59, 6866–6873.
- Aittaleb, M., Rashid, R., Chen, Q., Palmer, J.R., Daniels, C.J., and Li, H. (2003). Structure and function of archaeal box C/D sRNP core proteins. *Nature Structural & Molecular Biology* 10, 256–263.
- Amiri, K.A. (1994). Fibrillarin-like proteins occur in the domain Archaea. *J. Bacteriol.* 176, 2124–2127.
- Appel, C.D., and Maxwell, E.S. (2007). Structural features of the guide:target RNA duplex required for archaeal box C/D sRNA-guided nucleotide 2'-O-methylation. *RNA* 13, 899–911.
- Arai, T., Ishiguro, K., Kimura, S., Sakaguchi, Y., Suzuki, T., and Suzuki, T. (2015). Single methylation of 23S rRNA triggers late steps of 50S ribosomal subunit assembly. *Proc Natl Acad Sci USA* 112, E4707–E4716.
- Arndt, C., Koristka, S., Feldmann, A., and Bachmann, M. (2019). Native Polyacrylamide Gels. *Methods Mol. Biol.* 1855, 87–91.
- Ayadi, L., Galvanin, A., Pichot, F., Marchand, V., and Motorin, Y. (2019). RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1862, 253–269.
- Bae, S.-C., and Lee, Y.H. (2018). Meta-analysis of gene expression profiles of peripheral blood cells in systemic lupus erythematosus. *Cellular and Molecular Biology* 64, 40–49.
- Baer, R.J., and Dubin, D.T. (1981). Methylated regions of hamster mitochondrial ribosomal RNA: structural and functional correlates. *Nucleic Acids Res.* 9, 323–337.
- Bai, R., Wan, R., Yan, C., Lei, J., and Shi, Y. (2018). Structures of the fully assembled *Saccharomyces cerevisiae* spliceosome before activation. *Science* 360, 1423–1429.
- Barandun, J., Chaker-Margot, M., Hunziker, M., Molloy, K.R., Chait, B.T., and Klinge, S. (2017). The complete structure of the small-subunit processome. *Nat Struct Mol Biol* 24, 944–953.
- Baudin-Baillieu, A., Fabret, C., Liang, X., Piekna-Przybylska, D., Fournier, M.J., and Rousset, J.-P. (2009). Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res* 37, 7665–7677.

- Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Crystal Structure of the Eukaryotic Ribosome. *Science* *330*, 1203–1209.
- Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Engl.* *54*, 451–455.
- Bleichert, F., Granneman, S., Osheim, Y.N., Beyer, A.L., and Baserga, S.J. (2006). The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. *Proc. Natl. Acad. Sci. U.S.A.* *103*, 9464–9469.
- Bleichert, F., Gagnon, K.T., Brown, B.A., Maxwell, E.S., Leschziner, A.E., Unger, V.M., and Baserga, S.J. (2009). A Dimeric Structure for Archaeal Box C/D Small Ribonucleoproteins. *Science* *325*, 1384–1387.
- Blow, D. (2002). *Outline of Crystallography for Biologists* (Oxford, New York: Oxford University Press).
- Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. *Nature Reviews Molecular Cell Biology* *8*, 574–585.
- Bortolin-Cavaillé, M.-L., and Cavaillé, J. (2012). The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader–Willi locus generate canonical box C/D snoRNAs. *Nucleic Acids Res* *40*, 6800–6807.
- Bouffard, S., Dambrose, E., Brombin, A., Lempereur, S., Hatin, I., Simion, M., Corre, R., Bourrat, F., Joly, J.-S., and Jamen, F. (2018). Fibrillarin is essential for S-phase progression and neuronal differentiation in zebrafish dorsal midbrain and retina. *Dev. Biol.* *437*, 1–16.
- Bragg, W.H., and Bragg, W.L. (1913). The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* *88*, 428–438.
- Broda, R. (2003). A review of the triple-to-double coincidence ratio (TDCR) method for standardizing radionuclides. *Applied Radiation and Isotopes* *58*, 585–594.
- Broda, R., and Pochwalski, K. (1992). The enhanced triple to double coincidence ratio (ETDCR) method for standardization of radionuclides by liquid scintillation counting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* *312*, 85–89.
- Bs, S., T, K., E, G., Jr, K., H, S., and Te, D. (2017). Amino acid substrates impose polyamine, eIF5A, or hypusine requirement for peptide synthesis. *Nucleic Acids Res* *45*, 8392–8402.
- Buchwalter, A., and Hetzer, M.W. (2017). Nucleolar expansion and elevated protein translation in premature aging. *Nature Communications* *8*, 328.
- Cahill, N.M., Friend, K., Speckmann, W., Li, Z.-H., Terns, R.M., Terns, M.P., and Steitz, J.A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *The EMBO Journal* *21*, 3816–3828.
- Cassette, P., and Bouchard, J. (2003). The design of a liquid scintillation counter based on the triple to double coincidence ratio method. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* *505*, 72–75.

- Cavaillé, J. (2017). Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. *WIREs RNA* 8, e1417.
- Cavaillé, J., Nicoloso, M., and Bachellerie, J.-P. (1996). Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* 383, 732–735.
- Cavaillé, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C., and Bachellerie, J.-P. (2002). Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader–Willi/Angelman syndrome region. *Hum Mol Genet* 11, 1527–1538.
- Chambers, A.G., Percy, A.J., Yang, J., and Borchers, C.H. (2015). Multiple Reaction Monitoring Enables Precise Quantification of 97 Proteins in Dried Blood Spots. *Mol Cell Proteomics* 14, 3094–3104.
- Charenton, C., Wilkinson, M.E., and Nagai, K. (2019a). Mechanism of 5' splice site transfer for human spliceosome activation. 7.
- Charenton, C., Wilkinson, M.E., and Nagai, K. (2019b). Mechanism of 5' splice site transfer for human spliceosome activation. *Science* 364, 362–367.
- Charron, C., Manival, X., Cléry, A., Senty-Ségault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C., and Aubry, A. (2004). The Archaeal sRNA Binding Protein L7Ae has a 3D Structure Very Similar to that of its Eukaryal Counterpart While Having a Broader RNA-binding Specificity. *Journal of Molecular Biology* 342, 757–773.
- Chayen, N.E., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods* 5, 147–153.
- Chen, E.C.M. (1979). Selective Ion Monitoring. In *Practical Mass Spectrometry: A Contemporary Introduction*, B.S. Middleditch, ed. (Boston, MA: Springer US), pp. 127–149.
- Chen, W., Xie, Z., Yang, F., and Ye, K. (2017). Stepwise assembly of the earliest precursors of large ribosomal subunits in yeast. *Nucleic Acids Research* 45, 6837–6847.
- Cheng, J., Kellner, N., Berninghausen, O., Hurt, E., and Beckmann, R. (2017). 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nat Struct Mol Biol* 24, 954–964.
- Cheng, J., Baßler, J., Fischer, P., Lau, B., Kellner, N., Kunze, R., Griesel, S., Kallas, M., Berninghausen, O., Strauss, D., et al. (2019). Thermophile 90S Pre-ribosome Structures Reveal the Reverse Order of Co-transcriptional 18S rRNA Subdomain Integration. *Mol. Cell* 75, 1256-1269.e7.
- Choi, J., Indrisiunaite, G., DeMirici, H., leong, K.-W., Wang, J., Petrov, A., Prabhakar, A., Rechavi, G., Dominissini, D., He, C., et al. (2018). 2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat Struct Mol Biol* 25, 208–216.
- Chouhan, B.P.S., Maimaiti, S., Gade, M., and Laurino, P. (2019). Rossmann-Fold Methyltransferases: Taking a “β-Turn” around Their Cofactor, S-Adenosylmethionine. *Biochemistry* 58, 166–170.
- Cook Jr, E.H., and Scherer, S.W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature* 455, 919–923.

- Debye, P. (1913). Interferenz von Röntgenstrahlen und Wärmebewegung. *Annalen Der Physik* **348**, 49–92.
- Deng, L., Starostina, N.G., Liu, Z.-J., Rose, J.P., Terns, R.M., Terns, M.P., and Wang, B.-C. (2004). Structure determination of fibrillarin from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochem. Biophys. Res. Commun.* **315**, 726–732.
- Derenzini, M., Trerè, D., Pession, A., Montanaro, L., Sirri, V., and Ochs, R.L. (1998). Nucleolar function and size in cancer cells. *Am J Pathol* **152**, 1291–1297.
- Dever, T.E., Dinman, J.D., and Green, R. (2018). Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harb Perspect Biol* **10**.
- Dimitrova, D.G., Teysset, L., and Carré, C. (2019). RNA 2'-O-Methylation (Nm) Modification in Human Diseases. *Genes* **10**, 117.
- Dobbyn, H.C., McEwan, P.A., Krause, A., Novak-Frazer, L., Bella, J., and O'Keefe, R.T. (2007). Analysis of pre-mRNA and pre-rRNA processing factor Snu13p structure and mutants. *Biochem. Biophys. Res. Commun.* **360**, 857–862.
- Doe, C.M., Relkovic, D., Garfield, A.S., Dalley, J.W., Theobald, D.E.H., Humby, T., Wilkinson, L.S., and Isles, A.R. (2009). Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. *Hum Mol Genet* **18**, 2140–2148.
- Duval, M., Simonetti, A., Caldelari, I., and Marzi, S. (2015). Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics. *Biochimie* **114**, 18–29.
- El Hassouni, B., Sarkisjan, D., Vos, J.C., Giovannetti, E., and Peters, G.J. (2019). Targeting the Ribosome Biogenesis Key Molecule Fibrillarin to Avoid Chemoresistance. *Curr. Med. Chem.* **26**, 6020–6032.
- Elliott, B.A., Ho, H.-T., Ranganathan, S.V., Vangaveti, S., Ilkayeva, O., Abou Assi, H., Choi, A.K., Agris, P.F., and Holley, C.L. (2019). Modification of messenger RNA by 2'-O-methylation regulates gene expression in vivo. *Nat Commun* **10**, 3401.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
- Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S.E., Garcia, M., Laforêts, F., Marcel, V., Baudin-Baillieu, A., et al. (2017). Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc Natl Acad Sci USA* **114**, 12934–12939.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.
- Evans, P., and McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64**, 1–10.
- Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.
- Ewald, P.P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen Der Physik* **369**, 253–287.

- Fernandez, J.M., and Hoeffler, J.P. (1999). Gene expression systems: using nature for the art of expression (San Diego: Academic Press).
- Fiser, A., Do, R.K., and Sali, A. (2000). Modeling of loops in protein structures. *Protein Sci.* 9, 1753–1773.
- Fried, M.G. (1989). Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *ELECTROPHORESIS* 10, 366–376.
- Fried, M., and Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* 9, 6505–6525.
- Gagnon, K.T., Zhang, X., Qu, G., Biswas, S., Suryadi, J., Brown, B.A., and Maxwell, E.S. (2010). Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* 16, 79–90.
- Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C., and Bozzoni, I. (2002). Purified Box C/D snoRNPs Are Able To Reproduce Site-Specific 2J-O-Methylation of Target RNA In Vitro. *MOL. CELL. BIOL.* 22, 6.
- Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res* 9, 3047–3060.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, J.M. Walker, ed. (Totowa, NJ: Humana Press), pp. 571–607.
- Gautier, T., Bergès, T., Tollervey, D., and Hurt, E. (1997). Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Molecular and Cellular Biology* 17, 7088–7098.
- Ghalei, H., Hsiao, H.-H., Urlaub, H., Wahl, M.C., and Watkins, N.J. (2010). A novel Nop5-sRNA interaction that is required for efficient archaeal box C/D sRNP formation. *RNA* 16, 2341–2348.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* 27, 14–25.
- Golden, B.L., and Kundrot, C.E. (2003). RNA crystallization. *J. Struct. Biol.* 142, 98–107.
- Graziadei, A., Masiewicz, P., Lapinaite, A., and Carlomagno, T. (2016). Archaea box C/D enzymes methylate two distinct substrate rRNA sequences with different efficiency. *RNA* 22, 764–772.
- Graziadei, A., Gabel, F., Kirkpatrick, J., and Carlomagno, T. (2020). The guide sRNA sequence determines the activity level of box C/D RNPs. *ELife* 9, e50027.
- Green, R., and Noller, H.F. (1997). Ribosomes and Translation. *Annual Review of Biochemistry* 66, 679–716.
- Green, M.R., Sambrook, J., and Sambrook, J. (2012). *Molecular cloning: a laboratory manual* (Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press).

- Gruner, S.M., and Ealick, S.E. (1995). Charge coupled device X-ray detectors for macromolecular crystallography. *Structure* **3**, 13–15.
- Guillen-Chable, F., Corona, U.R., Pereira-Santana, A., Bayona, A., Rodríguez-Zapata, L.C., Aquino, C., Šebestová, L., Vitale, N., Hozak, P., and Castano, E. (2020). Fibrillar Ribonuclease Activity is Dependent on the GAR Domain and Modulated by Phospholipids. *Cells* **9**.
- Gutierrez, E., Shin, B.-S., Woolstenhulme, C.J., Kim, J.-R., Saini, P., Buskirk, A.R., and Dever, T.E. (2013). eIF5A promotes translation of polyproline motifs. *Mol. Cell* **51**, 35–45.
- Hamma, T., and Ferré-D'Amaré, A.R. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure* **12**, 893–903.
- Hellen, C.U.T. (2018). Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harb Perspect Biol* **10**, a032656.
- Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat Protoc* **2**, 1849–1861.
- Henderson, C.M., Shulman, N.J., MacLean, B., MacCoss, M.J., and Hoofnagle, A.N. (2018). Skyline Performs as Well as Vendor Software in the Quantitative Analysis of Serum 25-Hydroxy Vitamin D and Vitamin D Binding Globulin. *Clin. Chem.* **64**, 408–410.
- Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58.
- Higa-Nakamine, S., Suzuki, T., Uechi, T., Chakraborty, A., Nakajima, Y., Nakamura, M., Hirano, N., Suzuki, T., and Kenmochi, N. (2012). Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.* **40**, 391–398.
- Hinnebusch, A.G. (2014). The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* **83**, 779–812.
- Ho, C.S., Lam, C.W.K., Chan, M.H.M., Cheung, R.C.K., Law, L.K., Lit, L.C.W., Ng, K.F., Suen, M.W.M., and Tai, H.L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* **24**, 3–12.
- Hou, X. (2018). Liquid scintillation counting for determination of radionuclides in environmental and nuclear application. *J Radioanal Nucl Chem* **318**, 1597–1628.
- Huang, L., and Lilley, D.M.J. (2013). The molecular recognition of kink-turn structure by the L7Ae class of proteins. *RNA* **19**, 1703–1710.
- Huang, L., and Lilley, D.M.J. (2014). Structure of a rare non-standard sequence k-turn bound by L7Ae protein. *Nucleic Acids Res.* **42**, 4734–4740.
- Huang, L., and Lilley, D.M.J. (2016). A quasi-cyclic RNA nano-scale molecular object constructed using kink turns. *Nanoscale* **8**, 15189–15195.
- Huang, L., and Lilley, D.M.J. (2018). The kink-turn in the structural biology of RNA. *Q. Rev. Biophys.* **51**, e5.
- Huang, L., Ashraf, S., and Lilley, D.M.J. (2019). The role of RNA structure in translational regulation by L7Ae protein in archaea. *RNA* **25**, 60–69.

- Hunziker, M., Barandun, J., Petfalski, E., Tan, D., Delan-Forino, C., Molloy, K.R., Kim, K.H., Dunn-Davies, H., Shi, Y., Chaker-Margot, M., et al. (2016). UtpA and UtpB chaperone nascent pre-ribosomal RNA and U3 snoRNA to initiate eukaryotic ribosome assembly. *Nat Commun* 7, 12090.
- Incarnato, D., Anselmi, F., Morandi, E., Neri, F., Maldotti, M., Rapelli, S., Parlato, C., Basile, G., and Oliviero, S. (2017). High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. *Nucleic Acids Research* 45, 1433–1441.
- Iyer-Bierhoff, A., Krogh, N., Tessarz, P., Ruppert, T., Nielsen, H., and Grummt, I. (2018). SIRT7-Dependent Deacetylation of Fibrillar Controls Histone H2A Methylation and rRNA Synthesis during the Cell Cycle. *Cell Rep* 25, 2946-2954.e5.
- Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science* 27, 112–128.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.
- Kahl, S.D., Sittampalam, G.S., and Weidner, J. (2004). Calculations and Instrumentation used for Radioligand Binding Assays. In *Assay Guidance Manual*, G.S. Sittampalam, A. Grossman, K. Brimacombe, M. Arkin, D. Auld, C.P. Austin, J. Baell, B. Bejcek, J.M.M. Caaveiro, T.D.Y. Chung, et al., eds. (Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences), p.
- Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030–1033.
- Kasturi, K.N., Hatakeyama, A., Spiera, H., and Bona, C.A. (1995). Antifibrillar autoantibodies present in systemic sclerosis and other connective tissue diseases interact with similar epitopes. *J Exp Med* 181, 1027–1036.
- Ke, H. (1997). [25] Overview of isomorphous replacement phasing. In *Methods in Enzymology*, (Academic Press), pp. 448–461.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10, 845–858.
- Khoshnevis, S., Dreggors, R.E., Hoffmann, T.F.R., and Ghalei, H. (2019). A conserved Bcd1 interaction essential for box C/D snoRNP biogenesis. *J. Biol. Chem.* 294, 18360–18371.
- Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell* 85, 1077–1088.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal Structure of the Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6. *Science* 334, 941–948.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., and Ban, N. (2012a). Atomic structures of the eukaryotic ribosome. *Trends Biochem. Sci.* 37, 189–198.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., and Ban, N. (2012b). Atomic structures of the eukaryotic ribosome. *Trends in Biochemical Sciences* 37, 189–198.



Koh, C.M., Gurel, B., Sutcliffe, S., Aryee, M.J., Schultz, D., Iwata, T., Uemura, M., Zeller, K.I., Anele, U., Zheng, Q., et al. (2011). Alterations in Nucleolar Structure and Gene Expression Programs in Prostatic Neoplasia Are Driven by the MYC Oncogene. *The American Journal of Pathology* 178, 1824–1834.

Kondrat, R.W., McClusky, G.A., and Cooks, R.G. (1978). Multiple reaction monitoring in mass spectrometry/mass spectrometry for direct analysis of complex mixtures. *Anal. Chem.* 50, 2017–2021.

Königlich Preussische Akademie der Wissenschaften zu Berlin (1816). *Abhandlungen der physikalischen Klasse der Königlich-Preussischen Akademie der Wissenschaften* (Berlin : Realschul-Buchhandlung).

Konikkat, S., and Woolford, J.L. (2017). Principles of 60S ribosomal subunit assembly emerging from recent studies in yeast. *Biochem. J.* 474, 195–214.

Kossert, K. (2010). Activity standardization by means of a new TDCR-Čerenkov counting technique. *Applied Radiation and Isotopes* 68, 1116–1120.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292.

Krogh, N., Jansson, M.D., Häfner, S.J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., Lund, A.H., and Nielsen, H. (2016). Profiling of 2'- O -Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res* 44, 7884–7895.

Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology* 4, 222.

Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., and Carlomagno, T. (2013). The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* 502, 519–523.

Laue, M. (1913). Eine quantitative Prüfung der Theorie für die Interferenzerscheinungen bei Röntgenstrahlen. *Annalen Der Physik* 346, 989–1002.

Lechertier, T., Grob, A., Hernandez-Verdun, D., and Roussel, P. (2009). Fibrillarin and Nop56 interact before being co-assembled in box C/D snoRNPs. *Exp. Cell Res.* 315, 928–942.

Lee, K.-W., and Bogenhagen, D.F. (2014). Assignment of 2'-O-methyltransferases to modification sites on the mammalian mitochondrial large subunit 16 S ribosomal RNA (rRNA). *J. Biol. Chem.* 289, 24936–24942.

Lee, K.-W., Okot-Kotber, C., LaComb, J.F., and Bogenhagen, D.F. (2013). Mitochondrial ribosomal RNA (rRNA) methyltransferase family members are positioned to modify nascent rRNA in foci near the mitochondrial DNA nucleoid. *J. Biol. Chem.* 288, 31386–31399.

Li, L., and Ye, K. (2006a). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* 443, 302–307.

Li, L., and Ye, K. (2006b). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* 443, 302–307.

- Liang, X., Liu, Q., and Fournier, M.J. (2007). rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell* **28**, 965–977.
- 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.
- Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J., and Ye, K. (2011). Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469**, 559–563.
- Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T., and Wahl, M.C. (2007a). Binding of the Human Prp31 Nop Domain to a Composite RNA-Protein Platform in U4 snRNP. *Science* **316**, 115–120.
- Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T., and Wahl, M.C. (2007b). Binding of the Human Prp31 Nop Domain to a Composite RNA-Protein Platform in U4 snRNP. *Science* **316**, 115–120.
- Liu, S., Ghalei, H., Lührmann, R., and Wahl, M.C. (2011). Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. *RNA* **17**, 1655–1663.
- Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 125–132.
- Loveland, A.B., Demo, G., Grigorieff, N., and Korostelev, A.A. (2017). Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **546**, 113–117.
- Loza-Muller, L., Rodríguez-Corona, U., Sobol, M., Rodríguez-Zapata, L.C., Hozak, P., and Castano, E. (2015). Fibrillarin methylates H2A in RNA polymerase I trans-active promoters in *Brassica oleracea*. *Front. Plant Sci.* **6**.
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641.
- Marcel, V., Ghayad, S.E., Belin, S., Therizols, G., Morel, A.-P., Solano-González, E., Vendrell, J.A., Hacot, S., Mertani, H.C., Albaret, M.A., et al. (2013). p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer Cell* **24**, 318–330.
- Marchand, V., Blanloeil-Oillo, F., Helm, M., and Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res* **44**, e135–e135.
- Marmier-Gourrier, N., Cléry, A., Senty-Ségault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R., and Branlant, C. (2003). A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* **9**, 821–838.
- Martin, J.L., and McMillan, F.M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793.
- Martí-Renom, M.A., Stuart, A.C., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29**, 291–325.

- Massenet, S., Bertrand, E., and Verheggen, C. (2017). Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biology* *14*, 680–692.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* *40*, 658–674.
- McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E., and Watkins, N.J. (2007). A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol. Cell. Biol.* *27*, 6782–6793.
- Moazed, D., and Noller, H.F. (1989). Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* *57*, 585–597.
- Moore, T., Zhang, Y., Fenley, M.O., and Li, H. (2004). Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* *12*, 807–818.
- Murata, T., Suzuki, E., Ito, S., Sawatsubashi, S., Zhao, Y., Yamagata, K., Tanabe, M., Fujiyama, S., Kimura, S., Ueda, T., et al. (2008). RNA-binding protein hoip accelerates polyQ-induced neurodegeneration in *Drosophila*. *Biosci. Biotechnol. Biochem.* *72*, 2255–2261.
- Murray, J., Savva, C.G., Shin, B.-S., Dever, T.E., Ramakrishnan, V., and Fernández, I.S. (2016). Structural characterization of ribosome recruitment and translocation by type IV IRES. *ELife* *5*, e13567.
- Murray, K.K., Boyd, R.K., Eberlin, M.N., Langley, G.J., Li, L., and Naito, Y. (2013). Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry* *85*, 1515–1609.
- Nachmani, D., Bothmer, A.H., Grisendi, S., Mele, A., Bothmer, D., Lee, J.D., Monteleone, E., Cheng, K., Zhang, Y., Bester, A.C., et al. (2019). Germline NPM1 mutations lead to altered rRNA 2'-O-methylation and cause dyskeratosis congenita. *Nat. Genet.* *51*, 1518–1529.
- Narasimhan, K., Pillay, S., Huang, Y.-H., Jayabal, S., Udayasuryan, B., Veerapandian, V., Kolatkar, P., Cojocaru, V., Pervushin, K., and Jauch, R. (2015). DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. *Nucleic Acids Res* *43*, 1513–1528.
- Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I., and Klaholz, B.P. (2017). Visualization of chemical modifications in the human 80S ribosome structure. *Nature* *551*, 472–477.
- Nazar, R.N., Sitz, T.O., and Somers, K.D. (1980). Cytoplasmic methylation of mature 5.8 S ribosomal RNA. *J. Mol. Biol.* *142*, 117–121.
- Newton, K., Petfalski, E., Tollervy, D., and Cáceres, J.F. (2003). Fibrillarin Is Essential for Early Development and Required for Accumulation of an Intron-Encoded Small Nucleolar RNA in the Mouse. *Molecular and Cellular Biology* *23*, 8519–8527.
- Nguyen, T.H.D., Galej, W.P., Bai, X.-C., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* *530*, 298–302.
- Nilsen, T.W. (2013). Gel Purification of RNA. *Cold Spring Harb Protoc* *2013*, pdb.prot072942.

- van Nues, R.W., Granneman, S., Kudla, G., Sloan, K.E., Chicken, M., Tollervey, D., and Watkins, N.J. (2011). Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing: Extended snoRNA-rRNA base-pairing. *The EMBO Journal* *30*, 2420–2430.
- Ogle, J.M., Brodersen, D.E., Clemons, W.M., Tarry, M.J., Carter, A.P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* *292*, 897–902.
- Oliveira, C., and Domingues, L. (2018). Guidelines to reach high-quality purified recombinant proteins. *Appl Microbiol Biotechnol* *102*, 81–92.
- Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000). Homologs of Small Nucleolar RNAs in Archaea. *Science* *288*, 517–522.
- Omer, A.D., Zago, M., Chang, A., and Dennis, P.P. (2006). Probing the structure and function of an archaeal C/D-box methylation guide sRNA. *RNA* *12*, 1708–1720.
- Oruganti, S., Zhang, Y., Li, H., Robinson, H., Terns, M.P., Terns, R.M., Yang, W., and Li, H. (2007). Alternative Conformations of the Archaeal Nop56/58-Fibrillarin Complex Imply Flexibility in Box C/D RNPs. *Journal of Molecular Biology* *371*, 1141–1150.
- Oshima, K., Kakiuchi, Y., Tanaka, Y., Ueda, T., Nakashima, T., Kimura, M., and Yao, M. (2016). Structural basis for recognition of a kink-turn motif by an archaeal homologue of human RNase P protein Rpp38. *Biochem. Biophys. Res. Commun.* *474*, 541–546.
- Oshima, K., Gao, X., Hayashi, S., Ueda, T., Nakashima, T., and Kimura, M. (2018). Crystal structures of the archaeal RNase P protein Rpp38 in complex with RNA fragments containing a K-turn motif. *Acta Crystallogr F Struct Biol Commun* *74*, 57–64.
- Patel, B.A., Gospodarek, A., Larkin, M., Kenrick, S.A., Haverick, M.A., Tugcu, N., Brower, M.A., and Richardson, D.D. (2018). Multi-angle light scattering as a process analytical technology measuring real-time molecular weight for downstream process control. *MAbs* *10*, 945–950.
- Pauli, C., Liu, Y., Rohde, C., Cui, C., Fijalkowska, D., Gerloff, D., Walter, C., Krijgsveld, J., Dugas, M., Edemir, B., et al. (2020). Site-specific methylation of 18S ribosomal RNA by SNORD42A is required for acute myeloid leukemia cell proliferation. *Blood* *135*, 2059–2070.
- Pederson, T. (2011). The nucleolus. *Cold Spring Harb Perspect Biol* *3*.
- Peña, C., Hurt, E., and Panse, V.G. (2017). Eukaryotic ribosome assembly, transport and quality control. *Nat Struct Mol Biol* *24*, 689–699.
- Peng, Y., Yu, G., Tian, S., and Li, H. (2014). Co-Expression and Co-Purification of Archaeal and Eukaryal Box C/D RNPs. *PLoS ONE* *9*, e103096.
- Penzo, M., and Montanaro, L. (2018). Turning Uridines around: Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function. *Biomolecules* *8*.
- Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., and Coon, J.J. (2012). Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics* *11*, 1475–1488.

- Petrov, A., Wu, T., Puglisi, E.V., and Puglisi, J.D. (2013). Chapter Seventeen - RNA Purification by Preparative Polyacrylamide Gel Electrophoresis. In *Methods in Enzymology*, J. Lorsch, ed. (Academic Press), pp. 315–330.
- Piekna-Przybylska, D., Decatur, W.A., and Fournier, M.J. (2007). New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA* 13, 305–312.
- Pino, L.K., Searle, B.C., Bollinger, J.G., Nunn, B., MacLean, B., and MacCoss, M.J. (2020). The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* 39, 229–244.
- Pleiss, J.A., Derrick, M.L., and Uhlenbeck, O.C. (1998). T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* 4, 1313–1317.
- Pochwalski, K., Broda, R., Radoszewski, T., and Zelazny, P. (1981). Standardization of low-level beta-emitter solutions by using the liquid-scintillation triple-to-double coincidence ratio (TDCCR) method. *Methods of Low-Level Counting and Spectrometry*.
- Polikanov, Y.S., Melnikov, S.V., Söll, D., and Steitz, T.A. (2015). Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nature Structural & Molecular Biology* 22, 342–344.
- Prakash, A., Tomazela, D.M., Frewen, B., Maclean, B., Merrihew, G., Peterman, S., and Maccoss, M.J. (2009). Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J. Proteome Res.* 8, 2733–2739.
- Prieto, M.B., Georg, R.C., Gonzales-Zubiate, F.A., Luz, J.S., and Oliveira, C.C. (2015). Nop17 is a key R2TP factor for the assembly and maturation of box C/D snoRNP complex. *BMC Mol. Biol.* 16, 7.
- Prusiner, P., Yathindra, N., and Sundaralingam, M. (1974). Effect of ribose 2'-O-methylation on the conformation of nucleosides and nucleotides. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 366, 115–123.
- Qu, G., van Nues, R.W., Watkins, N.J., and Maxwell, E.S. (2011). The Spatial-Functional Coupling of Box C/D and C'/D' RNPs Is an Evolutionarily Conserved Feature of the Eukaryotic Box C/D snoRNP Nucleotide Modification Complex. *Molecular and Cellular Biology* 31, 365–374.
- Quinternet, M., Rothé, B., Barbier, M., Bobo, C., Saliou, J.-M., Jacquemin, C., Back, R., Chagot, M.-E., Cianférani, S., Meyer, P., et al. (2015). Structure/Function Analysis of Protein–Protein Interactions Developed by the Yeast Pih1 Platform Protein and Its Partners in Box C/D snoRNP Assembly. *Journal of Molecular Biology* 427, 2816–2839.
- Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell* 108, 557–572.
- Rashid, R., Aittaleb, M., Chen, Q., Spiegel, K., Demeler, B., and Li, H. (2003). Functional Requirement for Symmetric Assembly of Archaeal Box C/D Small Ribonucleoprotein Particles. *Journal of Molecular Biology* 333, 295–306.
- Rauniyar, N. (2015). Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int J Mol Sci* 16, 28566–28581.

- Reed, C.J., Lewis, H., Trejo, E., Winston, V., and Evilia, C. (2013). Protein Adaptations in Archaeal Extremophiles. *Archaea* 2013.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42, W320–W324.
- Rodriguez-Corona, U., Sobol, M., Rodriguez-Zapata, L.C., Hozak, P., and Castano, E. (2015). Fibrillarin from Archaea to human: Review on fibrillarin. *Biol. Cell* 107, 159–174.
- Rondeau, J.-M., and Schreuder, H. (2015). Protein Crystallography and Drug Discovery. In *The Practice of Medicinal Chemistry*, (Elsevier), pp. 511–537.
- Rothé, B., Back, R., Quinternet, M., Bizarro, J., Robert, M.-C., Blaud, M., Romier, C., Manival, X., Charpentier, B., Bertrand, E., et al. (2014a). Characterization of the interaction between protein Snu13p/15.5K and the Rsa1p/NUFIP factor and demonstration of its functional importance for snoRNP assembly. *Nucleic Acids Res.* 42, 2015–2036.
- Rothé, B., Saliou, J.-M., Quinternet, M., Back, R., Tiotiu, D., Jacquemin, C., Loegler, C., Schlotter, F., Peña, V., Eckert, K., et al. (2014b). Protein Hit1, a novel box C/D snoRNP assembly factor, controls cellular concentration of the scaffolding protein Rsa1 by direct interaction. *Nucleic Acids Res.* 42, 10731–10747.
- Rothé, B., Manival, X., Rolland, N., Charron, C., Senty-Ségault, V., Branlant, C., and Charpentier, B. (2017). Implication of the box C/D snoRNP assembly factor Rsa1p in U3 snoRNP assembly. *Nucleic Acids Res.* 45, 7455–7473.
- Rozhdestvensky, T.S., Tang, T.H., Tchirkova, I.V., Brosius, J., Bachellerie, J.-P., and Hüttenhofer, A. (2003). Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.* 31, 869–877.
- Ruscher, K., Reuter, M., Kupper, D., Trendelenburg, G., Dirnagl, U., and Meisel, A. (2000). A fluorescence based non-radioactive electrophoretic mobility shift assay. *Journal of Biotechnology* 78, 163–170.
- Ruse, C.I., Peacock, S., Ghiban, C., Rivera, K., Pappin, D.J., and Leopold, P. (2013). A tool to evaluate correspondence between extraction ion chromatographic peaks and peptide-spectrum matches in shotgun proteomics experiments. *Proteomics* 13, 2386–2397.
- Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W., and Beaudet, A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature Genetics* 40, 719–721.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Samarsky, D.A., Fournier, M.J., Singer, R.H., and Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *The EMBO Journal* 17, 3747–3757.
- Schilling, B., Rardin, M.J., MacLean, B.X., Zawadzka, A.M., Frewen, B.E., Cusack, M.P., Sorensen, D.J., Bereman, M.S., Jing, E., Wu, C.C., et al. (2012). Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol. Cell Proteomics* 11, 202–214.

- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9, 676–682.
- Shao, S., Murray, J., Brown, A., Taunton, J., Ramakrishnan, V., and Hegde, R.S. (2016). Decoding Mammalian Ribosome-mRNA States by Translational GTPase Complexes. *Cell* 167, 1229-1240.e15.
- Sharma, S., Marchand, V., Motorin, Y., and Lafontaine, D.L.J. (2017a). Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci Rep* 7, 11490.
- Sharma, S., Yang, J., Nues, R. van, Watzinger, P., Kötter, P., Lafontaine, D.L.J., Granneman, S., and Entian, K.-D. (2017b). Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLOS Genetics* 13, e1006804.
- Shubina, M.Y., Musinova, Y.R., and Sheval, E.V. (2018). Proliferation, cancer, and aging- novel functions of the nucleolar methyltransferase fibrillarin: Unexpected functions of fibrillarin. *Cell Biol Int* 42, 1463–1466.
- Shubina, M.Y., Arifulin, E.A., Sorokin, D.V., Sosina, M.A., Tikhomirova, M.A., Serebryakova, M.V., Smirnova, T., Sokolov, S.S., Musinova, Y.R., and Sheval, E.V. (2020). The GAR domain integrates functions that are necessary for the proper localization of fibrillarin (FBL) inside eukaryotic cells. *PeerJ* 8, e9029.
- Sirum-Connolly, K., Peltier, J.M., Crain, P.F., McCloskey, J.A., and Mason, T.L. (1995). Implications of a functional large ribosomal RNA with only three modified nucleotides. *Biochimie* 77, 30–39.
- Smith, J.L. (1998). Multiwavelength Anomalous Diffraction in Macromolecular Crystallography. In *Direct Methods for Solving Macromolecular Structures*, S. Fortier, ed. (Dordrecht: Springer Netherlands), pp. 211–225.
- Smith, D.-L., Erce, M.A., Lai, Y.-W., Tomasetig, F., Hart-Smith, G., Hamey, J.J., and Wilkins, M.R. (2020). Crosstalk of Phosphorylation and Arginine Methylation in Disordered SRGG Repeats of *Saccharomyces cerevisiae* Fibrillarin and Its Association with Nucleolar Localization. *J. Mol. Biol.* 432, 448–466.
- Smyth, M.S., and Martin, J.H.J. (2000). x Ray crystallography. *Mol Pathol* 53, 8–14.
- Spinner, D.S., Liu, S., Wang, S.-W., and Schmidt, J. (2002). Interaction of the myogenic determination factor myogenin with E12 and a DNA target: mechanism and kinetics 11 Edited by K. Yamamoto. *Journal of Molecular Biology* 317, 431–445.
- Su, H., Xu, T., Ganapathy, S., Shadfan, M., Long, M., Huang, T.H.-M., Thompson, I., and Yuan, Z.-M. (2014). Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene* 33, 1348–1358.
- Suryadi, J., Tran, E.J., Maxwell, E.S., and Brown, B.A. (2005). The crystal structure of the *Methanocaldococcus jannaschii* multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry* 44, 9657–9672.
- Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N., et al. (2018). Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* 46, 9289–9298.

- Terwilliger, T.C., and Eisenberg, D. (1987). Isomorphous replacement: effects of errors on the phase probability distribution. Erratum. *Acta Cryst A* **43**, 286–286.
- Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69.
- Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., and Kouzarides, T. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**, 564–568.
- Thandapani, P., O'Connor, T.R., Bailey, T.L., and Richard, S. (2013). Defining the RGG/RG motif. *Mol. Cell* **50**, 613–623.
- Thomson, E., Ferreira-Cerca, S., and Hurt, E. (2013). Eukaryotic ribosome biogenesis at a glance. *J. Cell. Sci.* **126**, 4815–4821.
- Tiku, V., Jain, C., Raz, Y., Nakamura, S., Heestand, B., Liu, W., Späth, M., Suchiman, H.E.D., Müller, R.-U., Slagboom, P.E., et al. (2017). Small nucleoli are a cellular hallmark of longevity. *Nature Communications* **8**, 1–9.
- Tollervey, D., Lehtonen, H., Carmo-Fonseca, M., and Hurt, E.C. (1991). The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *EMBO J.* **10**, 573–583.
- Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., and Hurt, E.C. (1993). Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* **72**, 443–457.
- Tran, E., Zhang, X., Lackey, L., and Maxwell, E.S. (2005). Conserved spacing between the box C/D and C'/D' RNPs of the archaeal box C/D sRNP complex is required for efficient 2'-O-methylation of target RNAs. *RNA* **11**, 285–293.
- Tran, E.J., Zhang, X., and Maxwell, E.S. (2003). Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs. *EMBO J.* **22**, 3930–3940.
- Triana-Alonso, F.J., Dabrowski, M., Wadzack, J., and Nierhaus, K.H. (1995). Self-coded 3'-Extension of Run-off Transcripts Produces Aberrant Products during in Vitro Transcription with T7 RNA Polymerase. *J. Biol. Chem.* **270**, 6298–6307.
- Turner, B., and Lilley, D.M.J. (2008). The Importance of G·A Hydrogen Bonding in the Metal Ion- and Protein-induced Folding of a Kink Turn RNA. *Journal of Molecular Biology* **381**, 431–442.
- Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* **6**, 1331–1342.
- Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302.
- Waller, I. (1923). Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen. *Z. Physik* **17**, 398–408.



- Wang, H., Boisvert, D., Kim, K.K., Kim, R., and Kim, S.-H. (2000). Crystal structure of a fibrillar homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *The EMBO Journal* *19*, 317–323.
- Watanabe, M., Ni, S., Lindenberger, A.L., Cho, J., Tinch, S.L., and Kennedy, M.A. (2013). Characterization of the Stoichiometry of HMGA1/DNA Complexes. *Open Biochem J* *7*, 73–81.
- Watanabe-Susaki, K., Takada, H., Enomoto, K., Miwata, K., Ishimine, H., Intoh, A., Ohtaka, M., Nakanishi, M., Sugino, H., Asashima, M., et al. (2014). Biosynthesis of Ribosomal RNA in Nucleoli Regulates Pluripotency and Differentiation Ability of Pluripotent Stem Cells. *STEM CELLS* *32*, 3099–3111.
- Watkins, N.J., Dickmanns, A., and Lührmann, R. (2002). Conserved Stem II of the Box C/D Motif Is Essential for Nucleolar Localization and Is Required, Along with the 15.5K Protein, for the Hierarchical Assembly of the Box C/D snoRNP. *Mol Cell Biol* *22*, 8342–8352.
- Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* *54*, 5.6.1–5.6.37.
- Wilson, D.N., and Doudna, J.H. (2012). The structure and function of the eukaryotic ribosome. *Cold Spring Harb Perspect Biol* *4*.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* *67*, 235–242.
- Woolford, J.L., and Baserga, S.J. (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* *195*, 643–681.
- Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* *272*, 1–40.
- Xue, S., Wang, R., Yang, F., Terns, R.M., Terns, M.P., Zhang, X., Maxwell, E.S., and Li, H. (2010). Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol. Cell* *39*, 939–949.
- Yagoub, D., Hart-Smith, G., Moecking, J., Erce, M.A., and Wilkins, M.R. (2015). Yeast proteins Gar1p, Nop1p, Npl3p, Nsr1p, and Rps2p are natively methylated and are substrates of the arginine methyltransferase Hmt1p. *Proteomics* *15*, 3209–3218.
- Yang, J., Sharma, S., Watzinger, P., Hartmann, J.D., Kötter, P., and Entian, K.-D. (2016a). Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS ONE* *11*, e0168873.
- Yang, Z., Lin, J., and Ye, K. (2016b). Box C/D guide RNAs recognize a maximum of 10 nt of substrates. *Proc Natl Acad Sci USA* *113*, 10878–10883.
- Yang, Z., Wang, J., Huang, L., Lilley, D.M.J., and Ye, K. (2020). Functional organization of box C/D RNA-guided RNA methyltransferase. *Nucleic Acids Res.*
- Ye, K., Jia, R., Lin, J., Ju, M., Peng, J., Xu, A., and Zhang, L. (2009). Structural organization of box C/D RNA-guided RNA methyltransferase. *PNAS* *106*, 13808–13813.
- Yin, S., Jiang, H., Chen, D., and Murchie, A.I.H. (2015). Substrate Recognition and Modification by the Nosiheptide Resistance Methyltransferase. *PLOS ONE* *10*, e0122972.

- Yip, W.S.V., Shigematsu, H., Taylor, D.W., and Baserga, S.J. (2016). Box C/D sRNA stem ends act as stabilizing anchors for box C/D di-sRNPs. *Nucleic Acids Res* *44*, 8976–8989.
- Yoshihama, M., Nakao, A., and Kenmochi, N. (2013). snOPY: a small nucleolar RNA orthological gene database. *BMC Res Notes* *6*, 426.
- Yu, G., Zhao, Y., and Li, H. (2018). The multistructural forms of box C/D ribonucleoprotein particles. *RNA* *24*, 1625–1633.
- Yu, G., Zhao, Y., Tian, S., Rai, J., He, H., Spear, J., Sousa, D., Fan, J., Yu, H.-G., Stagg, S.M., et al. (2019). Yeast R2TP Interacts with Extended Termini of Client Protein Nop58p. *Sci Rep* *9*, 20228.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science* *292*, 883–896.
- Zhan, X., Yan, C., Zhang, X., Lei, J., and Shi, Y. (2018). Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res.* *28*, 1129–1140.
- Zhao, H., Brown, P.H., and Schuck, P. (2011). On the Distribution of Protein Refractive Index Increments. *Biophys J* *100*, 2309–2317.
- Zhou, F., Liu, Y., Rohde, C., Pauli, C., Gerloff, D., Köhn, M., Misiak, D., Bäumer, N., Cui, C., Göllner, S., et al. (2017). AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nat Cell Biol* *19*, 844–855.
- Zhu, Y., Pirnie, S.P., and Carmichael, G.G. (2017). High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA* *23*, 1303–1314.
- Zimm, B.H. (1948a). The Scattering of Light and the Radial Distribution Function of High Polymer Solutions. *J. Chem. Phys.* *16*, 1093–1099.
- Zimm, B.H. (1948b). Apparatus and Methods for Measurement and Interpretation of the Angular Variation of Light Scattering; Preliminary Results on Polystyrene Solutions. *J. Chem. Phys.* *16*, 1099–1116.
- Zwart, P.H., Grosse-kunstleve, R.W., and Adams, P.D. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation.

# Curriculum Vitae

## Simone Höfler

|                               |  |
|-------------------------------|--|
| <b>Education</b>              | <b>Gottfried Wilhelm Leibniz University Hannover, Germany</b><br>Dr. rer. nat., Chemistry, <u>expected 2020</u><br><b>University of Vienna, Austria</b><br>MSc, Biological Chemistry, 2015<br>BSc, Molecular Biology, 2013   |
| <b>Research Experience</b>    | <b>Center of Biomolecular Drug Research,<br/>Gottfried Wilhelm Leibniz Universität Hannover, Germany</b><br><u>Doctoral Researcher</u><br><i>Project:</i> Structural and Functional Analysis of Eukaryotic snoRNP Complexes Catalyzing rRNA 2'-O-Ribose Methylation<br><i>Supervisor:</i> Prof. Dr. Teresa Carlomagno<br><i>Duration:</i> November 2015 to 2020<br><br><b>Institute of Organic Chemistry,<br/>University of Barcelona, Spain</b><br><u>ERASMUS+ Undergraduate Researcher</u><br><i>Project:</i> Insights into the modulation of the interactions between the Unique and SH3 domains in Src Family kinases<br><i>Supervisor:</i> Prof. Dr. Robert Konrat and Prof. Dr. Miquel Pons<br><i>Duration:</i> February to August 2015<br><br><b>Institute of Molecular Biotechnology (IMBA), Vienna, Austria</b><br><u>Undergraduate Student Researcher</u><br><i>Project:</i> Investigation of protein complex formation by the needle forming protein in the type III secretion system from Salmonella using transmission electron microscopy (TEM)<br><i>Supervisor:</i> Thomas Marlovits, PhD<br><i>Duration:</i> July to August 2014<br><br><b>Max Perutz Labs,<br/>University of Vienna, Austria</b><br><u>Undergraduate Student Researcher</u><br><i>Project:</i> NMR analysis of extracellular matrix protein osteopontin reveals binding of MAX homodimer and thermodynamic analysis shows thermodynamics of heparin and osteopontin<br><i>Supervisor:</i> Prof. Dr. Robert Konrat<br><i>Duration:</i> May to June 2013 |
| <b>Publications</b>           | <b>S. Höfler</b> , T. Carlomagno, Structural and functional roles of 2'-O-ribose methylations and their enzymatic machinery across multiple classes of RNAs. <i>Current Opinion in Structural Biology</i> . <b>65</b> , 42–50 (2020).<br><br>D. Kurzbach, T. C. Schwarz, G. Platzer, <b>S. Höfler</b> , D. Hinderberger, R. Konrat, Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex. <i>Angewandte Chemie International Edition</i> . <b>53</b> , 3840–3843 (2014).   |
| <b>Awards and Fellowships</b> | <b>Performance Scholarship awarded by the University of Vienna, Austria</b><br>2011, 2012, 2013, 2014  |

**Conference Talks  
and Posters**

**Keystone Symposium on RNA-Protein Interactions (X1)**

24<sup>th</sup> to 28<sup>th</sup> of February 2019, Whistler, Canada

Selected Talk and Poster presenter

*Titel:* Structural and Functional Analysis of Eukaryotic snoRNP Complexes  
Catalyzing rRNA 2'-O-Riobse Methylation

**5<sup>th</sup> Meeting of Regulating with RNA in Bacteria and Archaea**

19<sup>th</sup> to 22<sup>th</sup> of March 2018, Sevilla, Spain

Poster presenter

**NMR Winter Retreat of Protein-RNA Interactions 2017**

January 2017, Parpan, Switzerland

Poster presenter

**Biochemistry 2016**

6<sup>th</sup> to 8<sup>th</sup> of July 2016, Frankfurt am Main, Germany

Poster presenter

**Training and  
Workshops**

**EMBO Practical Course**

“Small Angle Neutron and X-ray Scattering from proteins in solution”

10<sup>th</sup> to 15<sup>th</sup> of September 2017, Grenoble, France