Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Andrea Graziadei, MBiochem

Born in:

Torino (Italy), 21 July 1989

Oral examination: 8 February 2019

The mechanism and regulation of rRNA methylation by the Box C/D sRNP enzyme in solution

Referees:
Dr. Martin Beck
Prof. Dr. Frauke Melchior

SUMMARY

The biogenesis of the ribosome requires a series of essential modifications of ribosomal RNAs (rRNAs) and their precursor pre-rRNAs. The most abundant of such modifications is the methylation of the ribose 2´-OH, which occurs at over 100 rRNA sites in humans. rRNA methylation is known to increase the stability of the ribosome and to be required for accurate and efficient protein translation. While 2'-O methylation sites are known to cluster around the functional centres of the ribosome, the abundance of methylation at each site is known to vary, which may provide a mechanism to fine tune ribosomal function, creating specialized ribosome populations.

In eukaryotes and archaea, rRNA 2'-O methylation is mediated by Box C/D ribonucleoprotein particles (RNPs). These particles, referred to as small nucleolar RNPs (snoRNPs) in eukaryotes and small RNPs (sRNPs) in archaea, use a guide RNA in order to direct the methylation of a specific nucleotide on the substrate rRNA. In archaea, each small guide RNA (sRNA) is responsible for the methylation of two rRNA sites using two different separate guide regions.

Despite several structures of archaeal Box C/D sRNPs being available, the molecular basis for the regulation of the enzyme and the consequent generation of varying methylation abundances across different rRNA sites remains elusive.

In order to understand the mechanism and regulation of the enzyme, I investigated the biochemical properties of archaeal Box C/D sRNPs reconstituted *in vitro*. Through a combination of biochemical and nuclear magnetic resonance (NMR)-based assays, I could show that archaeal RNPs catalyse the methylation of different substrate rRNA sites with varying degrees of efficiency and cooperativity. Furthermore, using low-resolution small angle scattering (SAS) techniques, I could show that addition of substrate RNAs onto some sRNPs is correlated with the complex undergoing a transition between different oligomeric and/or conformational states, thereby contextualising the multiple sRNP structures observed in previous studies.

In the second part of my work, I used a combination of distance restraints derived from NMR and low-resolution information from SAS to obtain the structures of an archaeal sRNP bound to either of its two substrate RNAs by an integrative structural biology approach. As this particle contains flexible regions, the work required the development of a novel algorithm capable of dealing with NMR/SAS signals arising from ensembles, rather than single conformers. Using this tool, I could derive the populations of conformers within ensembles of RNPs bound to different substrate RNAs, which provide a structural basis for the varying methylation efficiency of the enzyme.

Ultimately, the work presented here provides a model for understanding one of the mechanism through which specialised ribosome populations are generated *in vivo* and contributes to the development of novel techniques for integrative structure modelling of flexible systems.

ZUSAMMENFASSUNG

Ribosomale Ribonukleinsäuren (rRNAs) und deren Vorläufer prä-rRNAs durchlaufen im Zuge der Ribosomen-Biogenese eine Serie an essentiellen Modifikationen. Die häufigste dieser Modifikationen ist die Methylierung der Ribose am 2'-OH. Solche rRNA Methylierungen stabilisieren das Ribosom und ermöglichen somit effizientere und genauere Proteinsynthese. 2'-O-Methylierungen treten verstärkt an Positionen an und um funktionelle Zentren des Ribosoms auf, allerdings variiert die Häufigkeit der Methylierungen an einzelnen Stellen. Dies könnte eine Möglichkeit darstellen, die Funktion des Ribosoms gezielt zu steuern und somit spezialisierte Ribosomenpopulationen zu schaffen.
Die 2'-O-Methylierung der rRNA wird in Eukaryonten und Archaeen von Box C/D Ribonukleoproteinpartikeln (RNPs) übernommen. Diese Partikel, in Eukaryonten als snoRNPs und in Archaeen als sRNPs bezeichnet, methylieren spezifische Nukleotide der Substrat-RNA unter Verwendung einer Führer-RNA. In Archaeen ist jede Führer-RNA (sRNA) für die Methylierung der rRNA an zwei Positionen verantwortlich. Trotz der Verfügbarkeit mehrerer Strukturen von Box C/D sRNPs ist die molekulare Basis für die Regulierung des Enzyms und die dadurch folgende unterschiedliche Methylierungshäufigkeit an verschiedenen rRNA-Stellen immer noch ungeklärt.

In dieser Arbeit wurden die biochemischen Eigenschaften *in vitro* rekonstituierter archaealer Box C/D sRNPs untersucht, um Einblicke in die Regulation des Enzyms zu gewinnen. Durch Kombination biochemischer Techniken mit Kernspinresonanzspektroskopie (NMR) wurde gezeigt, dass archaeale sRNPs verschiedene Substrat-rRNA Positionen mit unterschiedlicher Effizienz und Kooperativität methylieren. Zusätzlich zeigten Kleinwinkelstreuungstechniken (SAS), dass einige sRNPs nach Zugabe der Substrat-RNA zwischen unterschiedlichen oligomeren und/oder Konformationszuständen wechslen. Diese Beobachtungen erklären auch die bisher beobachteten, unterschiedlichen sRNP Strukturen.
Im zweiten Teil dieser Arbeit wurden halb-geladene Strukturen eines archaealen sRNPs mit jeweils einer Substrat-RNA durch einen integrativen Strukturbiologieansatz gelöst, nämlich einer Kombination aus Entfernungseinschränkungen aus NMR und niedrigaufgelöster Strukturinformation aus SAS Experimenten. Da das Partikel flexible Regionen enthält, war es notwenig, einen neuen Algorithmus zu entwickeln, welcher NMR und SAS Signale aus mehreren statt wie bisher nur einzelnen Konformeren berücksichtigt. Mithilfe dieses Algorithmus wurden RNP-Populationen aus unterschiedlichen Konformeren gebunden an jeweils verschiedene Substrat-RNAs detektiert, welches die strukturelle Basis für die unterschiedliche Methylierungseffizienz des Enzyms darstellt. Zusammengefasst beschreibt diese Arbeit ein neues mechanistisches Modell wie spezialisierte Ribosomenpopulationen *in vivo* generiert werden können.

Abbreviations and symbols

CSA: chemical shift anisotropy
CSP: chemical shift perturbation
CTD: C-terminal domain

D_{max}: maximum dimension of the particle
EM: electron microscopy
Fib: Fibrillarin
HMQC: heteronuclear multiple-quantum coherence
HSQC: heteronuclear single-quantum coherence
IMP: integrative modelling platform
I_{para}/I_{dia}: ratio of peak intensity in the paramagnetic and diamagnetic states

K-loop: kink-loop
K-turn: kink-turn
LSU: large subunit of the ribosome
PRE: paramagnetic relaxation enhancement
MX: macromolecular crystallography
NMR: nuclear magnetic resonance
NTD: N-terminal domain

R_2^{dia}: transverse relaxation rate

R_2^{diaH}: ^{1}H single-quantum coherence transverse relaxation rate

R_2^{diaHC}:^{1}H-^{13}C multiple-quantum coherence transverse relaxation rate

RMSD: root-mean-square deviation
RNP: ribonucleoprotein
R_g: radius of gyration
SANS: small-angle neutron scattering
SAS: small-angle scattering
SAXS:small-angle X-ray scattering
SLD: scattering length density
SSU: small subunit of the ribosome
TROSY: transverse relaxation optimized spectroscopy
pre-rRNA: pre-ribosomal RNA
rRNA: ribosomal RNA
sRNA: small ribonucleoprotein particle
sRNP: small ribonucleoprotein
snoRNA: small nucleolar RNA
snoRNP: small nucleolar ribonucleoprotein particle
st-sR: stabilised sRNA construct

Γ_2: paramagnetic contribution to the transverse relaxation rate

τ_C: correlation time for the electron-nucleus interaction vector

1. Introduction

1.1 Ribosomal Biogenesis and rRNA modification in Eukaryotes and Archaea

Ribosomes are complex, MegaDalton (MDa)-size ribonucleoprotein (RNP) particles that read the genetic information present in mRNA and translate it into protein sequence by catalysing the formation of peptide bonds between specific amino acids. Ribosomes are present in all domains of life, and are always composed of a large and small subunit, which for historic reasons take the name of their Svedberg sedimentation coefficient. Thus, full eukaryotic ribosomes are referred to as 80S ribosomes, and the small and large ribosomal subunits as 40S and 60S subunits or SSU and LSU (see table 1).

Table 1. Nomenclature of ribosomes and their subunits

	Eukarya	Archaea	Bacteria
rRNAs	28S, 5.8S, 5S, 18S	23S, 5S, 16S	23S, 5S, 16S
ribosome	80S	70S	70S
small ribosomal subunit (SSU)	40S	30S	30S
large ribosomal subunit (LSU)	60S	50S	50S
RNA-modifying factors	snoRNPs	sRNPs	site-specific enzymes

Adapted from Yip et al, 2013

The correct assembly of ribosomes is crucial to cell survival and as such is very tightly controlled with ample redundancy. While the overall function of the ribosome is broadly conserved across the domains of life, ribosomal biogenesis has diverged and presents interesting similarities and differences between species, both in terms of the rRNA processing events and in terms of the proteins (r-proteins) involved in catalysis, structural support and regulation of the pathway.

In eukaryotes, the biogenesis of the two subunits and of the whole ribosome is a highly regulated process that involves a huge number of protein and RNA assembly factors, spread across multiple subcellular compartments, from the nucleolus to the cytoplasm.

Broadly speaking, a long pre-ribosomal RNA (pre-rRNA) precursor is transcribed from the genome in the nucleolus by RNA polymerase I, processed and folded together with other ribosomal components by assembly factors. The particle is then exported to the cytoplasm where mature ribosomes can be formed. All of the steps taking place in the nucleolus, be it folding or modification of the pre-rRNA, can occur both post- and co-transcriptionally. Furthermore, chemical modifications are placed at specific sites in the pre-rRNA sequence by *trans*-acting small ribonucleolar RNA protein particles (snoRNPs) assembled on small nucleolar RNAs (snoRNAs).

In the yeast *Saccharomyces cerevisiae*, pre-rRNA is transcribed from specific rDNA repeats located on chromosome 12, each of which encodes a 35S primary pre-rRNA transcript that is processed into the 18S, 5.8S and 25S pre-rRNAs via multiple exo- and endo-nucleolytic steps (see Figure 1). The 5' section of the transcript, encoding the precursor of the 18S rRNA, is separated from the 3' section via endonucleolytic cleavage. As the separation of the two fragments and processing of the 5' can occur at three different cleavage sites, the processing splits into multiple pathways, each of which removes successive spacer regions from the long precursor to eventually yield mature rRNAs. Eventually, the 20S pre-rRNA is packaged into 43S pre-SSU particles that are exported to the cytoplasm, where it undergoes removal of the final spacer region. On the other hand, the LSU RNA precursor $27SA_2$ can end up in 66S pre-LSU particles via two pathways: a major route involving protein cleavage at the A_3 site by the MRP RNAse, followed by digestion by endonucleases all the way to the $B1_S$ site; or a minor pathway proceeding via a cleavage step at the $B1_L$ site. These two pathways result in 5.8S rRNAs with differing 5' ends, both of which are found in mature ribosomes at a relative abundance of 85 and 15%.

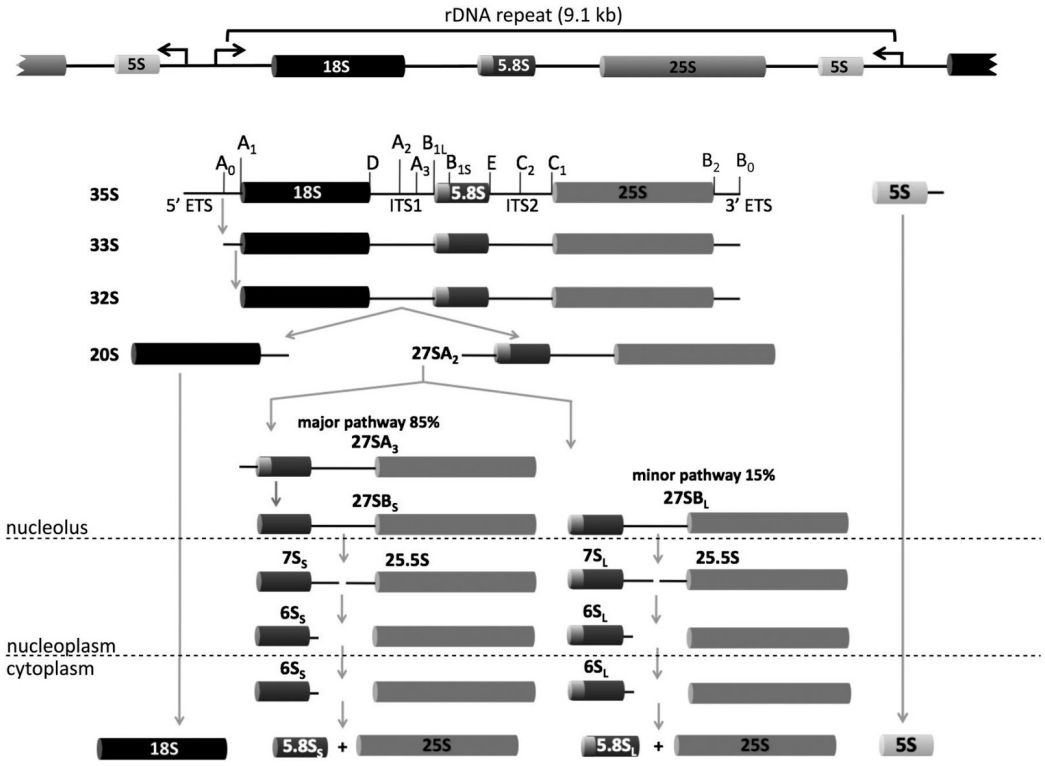


Figure 1.1.1. The steps of rRNA biogenesis in *Saccharomyces cerevisiae*. The figure shows the processing of the rRNA coming from the rDNA cluster on chromosome 12. Initially, the 5' end of the 32S precursor is generated by cleavage of the 5' ETS by successive cleavage steps, while the 3' of the precursor is generated by a single endonucleolytic cleavage by Rnt1. Cleavage of the 32S precursor yields the 20S rRNA precursor and the $27SA_2$ precursor, which encodes for th 25S and the 5.8S RNA. Rcl1p has been proposed as the endonuclease for the cleavage at the A_2 site that separates the LSU and SSU RNAs (Horn *et al*, 2011). While the 20S RNA is ready for cytoplasmic export, the $27SA_2$ precursor needs to be further processed: this can occur via a major pathway involving cleavage steps by MRP, Rat1p and Rrp17p (Oeffinger *et al*, 2009), or via direct cleavage at site $B1_L$. Reproduced with permission from (Woolford & Baserga, 2013).

Historically, all cleavage steps were thought to be post-transcriptional, since initial metabolic labelling experiments did not detect large population of early cleaved transcripts. However, novel time-resolved metabolic labelling and RNA extraction approaches revealed that, in *S. cerevisiae,* approximately 70% of nascent rRNAs undergo co-transcriptional cleavage, with the remainder of the transcripts following slower kinetics (Koš & Tollervey, 2010). Furthermore, although the cleavage steps are depicted as a sequence in Figure 1, rRNA processing may not necessarily be ordered (Torchet & Denmat, 2000), but rather depend on the rate of each step *in vivo*.

Although the nature of the rRNA cleavage pathway has been established, the protein composition of each pre-ribosomal particle has been tougher to define, mainly because r-proteins are so abundant that they contaminate pull-down experiments even when they are not tagged, which makes it challenging to disentangle what factors are associated with which step. In the case of bacterial ribosomes, *in vitro* reconstitution experiments show that ribosomal proteins are required in a hierarchical order, consistent with the existence of a specific pathway of loading ribosomal proteins onto pre-rRNAs. Such results have been replicated in yeast, where protein depletion by mutation leads to the accumulation of specific intermediates (Gamalinda *et al*, 2014). In some cases, it has been possible to prove that specific proteins are responsible for the correct folding of individual rRNA structures: for example, Rps4 is tasked with folding two helices in the 16S rRNA precursors in *E. Coli,* and manages to remodel its rRNA partner by affecting the conformational landscape the RNA can access (Kim *et al*, 2014).

Over recent years, both structural and proteomic efforts have led to considerable progress in defining the protein composition of each pre-ribosomal particle (Kater *et al*, 2017; Sun *et al*, 2017). These efforts have not only contributed towards a structural view of the ribosome biogenesis process, but also helped to identify the order and regulation of each step, further confirming that ribosomal proteins associate with rRNA precursors in an ordered manner.

For eukaryotes, it is clear that early ribosomal biogenesis, up to the formation of the 90S pre-ribosome, occurs mainly co-transcriptionally. Indeed, depletion of SSU proteins blocks biogenesis at the early cleavage states, impairing cleavage at the A_2 site (Ferreira-Cerca *et al*, 2005). Furthermore, SSU proteins are known to bind the nascent pre-rRNA transcript(de la Cruz *et al*, 2015), which is known to be cleaved predominantly co-transcriptionally (Koš & Tollervey, 2010). On the other hand, most of the assembly of the LSU occurs

post-transcriptionally, with r-proteins found in the LSU and their associated loading factors found in precursors downstream of A_2 site cleavage (Dez *et al*, 2004).

In several cases, proteins found in mature ribosomes have paralogs that are found in pre-ribosomes, which are removed in the process of ribosomal maturation. The best characterised example of such "placeholder" behaviour in *S. cerevisiae* is Mrt4, a 60S ribosomal factor that is then replaced by the N-terminus of protein P0 in the cytoplasm (Rodríguez-Mateos *et al*, 2009), in a step that is thought to act as quality control for the correct folding of the GTPase-activating center in the large ribosomal subunit, a universally conserved site involved in recruiting and regulating critical ribosomal GTPases.

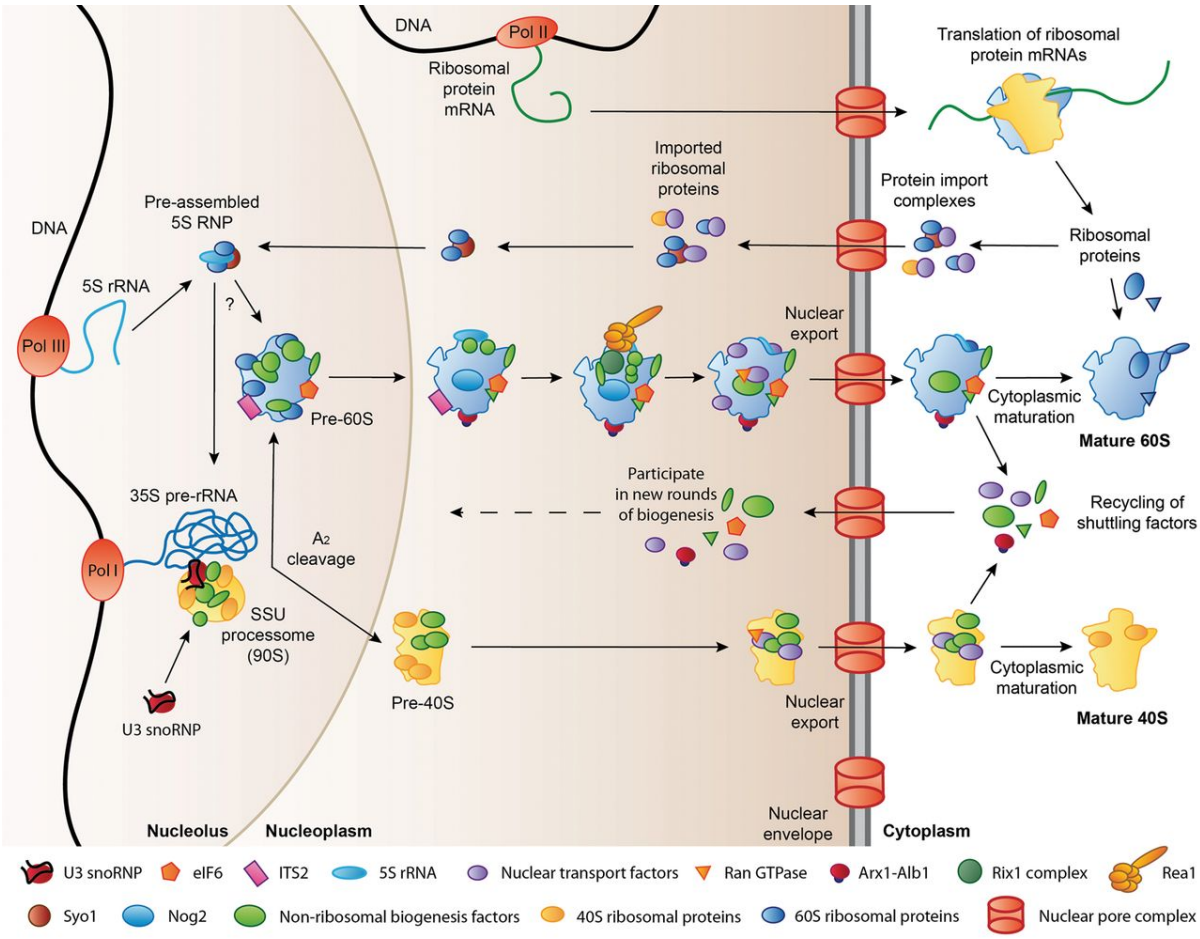


Figure 1.1.2. The ribosome biogenesis pathway in *Saccharomyces cerevisiae*. Binding of assembly factors and RNA modifying enzymes to rRNAs begins co-transcriptionally in the nucleolus, where the A2 cleavage separates the 90S pre-ribosome (also known as the SSU processome) into the nascent SSU and LSU. In the case of the 5S pre-rRNA, the timing of the binding is not clear. The LSU and SSU are exported to the cytosol independently of each other, where they can form full pre-ribosomes that undergo quality control steps where placeholder proteins are

removed, including a “test drive” (Strunk *et al*, 2012), before taking part in translation. Figure reproduced from Greber, 2016.

Chemical nucleotide modifications are also thought to play a key role in rRNA folding, stability and function. The two most abundant chemical modifications found in rRNAs are methylation of the 2’-hydroxyl group (2’-O methylation), and isomerisation of uridine into pseudouridine. Both modifications are highly abundant in rRNAs, conserved from the rRNA precursors all the way to the 80S ribosome. As in the case of rRNA cleavage, metabolic labelling studies initially led to the belief that rRNA modification occurs post-transcriptionally on newly released transcripts. These studies detected that ribose methylation immediately followed RNA cleavage, which at the time was thought to occur post-transcriptionally (Trapman *et al*, 1975).

However, in the case of the 18S RNA precursor, RNA methylation seem to occur almost entirely on the nascent transcript, while the modification of the precursor to the 25S rRNA occurs in two phases, one on the nascent transcript and one on the released chain (Koš & Tollervey, 2010; Birkedal *et al*, 2015). Recent high-resolution structures (Kornprobst *et al*, 2016; Sun *et al*, 2017), as well as proteomic data, show that U3 snoRNPs, which are essential for the A_0, A_1 and A_2 cleavage events, are bound to fully formed 90S pre-ribosomal particles prior to A_2 site cleavage (Kornprobst *et al*, 2016).

In mature ribosomes, modified nucleotides cluster around functional centers, including the peptidyl transfer center and the decoding region(Decatur & Fournier, 2002; Polikanov *et al*, 2015). 2’-O methyl sites have been shown to play a role in translation efficiency and accuracy (Baxter-Roshek *et al*, 2007; Baudin-Baillieu *et al*, 2009), as well as ribosome assembly(Basu *et al*, 2011). Thus, the regulation of occupancy of 2’-O methyl sites may be a way to modulate ribosome function(Sloan *et al*, 2017).

The mapping of such nucleotides has been tackled by both high-resolution structures of ribosomes and next-generation sequencing, on top of the more traditional primer-extension assays. Along these lines, recent sequencing-based approaches have shown that rRNA methyl sites are modified substoichiometrically, rather than constitutively (Birkedal *et al*, 2015). A similar phenomenon has also been observed for pseudouridylation in *S. cerevisiae* (Taoka *et al*, 2016). In general, a consensus seems to be emerging that in humans, there are tissue-specific ribosome populations displaying different levels of rRNA methylation,

contributing to the idea that higher organisms have “specialised ribosomes”. These specialised ribosome populations have been proposed to contribute to tissue-specific functions. In neurons, a subset of ribosomes localised in the dendrites are capable of translating local pools of mRNAs upon activation, in a process that contributes to long-term potentiation and memory formation (Xue & Barna, 2012).

2’-O RNA methylation is catalysed by box C/D snoRNPs, while box H/ACA snoRNPs are responsible for pseudouridylation (Kiss, 2001). Both classes of snoRNPs are made up by protein factors assembled onto snoRNAs which have short (up to 20 nucleotides) stretches that act as complementary guides for placing the chemical modification onto specific rRNA sites (Kiss-László *et al*, 1998; Nicoloso *et al*, 1996). Both snoRNP classes are named after the conserved elements found in their snoRNAs, which are recognised by the protein factors in snoRNP assembly. In *S. cerevisiae,* the conserved C and D boxes are recognised by the protein Snu13p (see section 1.2).

It is worth noting that, while the whole process of ribosome biogenesis is upregulated in several cancers, it has been shown that specific 2’-O methylation sites are highly modified in some aggressive breast cancer cell lines, leading to a decrease in translation accuracy (Belin *et al*, 2009). Moreover, changes in levels of the methyltransferase (FBL in humans) can affect translation start site selection. In cancers that have hypermethylated rRNAs due to FBL overexpression, IRES-independent translation initiation is stimulated, leading to an overall increase in gene expression (Marcel *et al*, 2013).

A small subset of snoRNAs is not used by either class of snoRNP for RNA modification but rather act as chaperones in the formation of specific structures: for example, the U3 snoRNP guides the formation of central pseudoknot of the 18S rRNA. Furthermore, certain snoRNAs are known to have dual functions, such as acting as guides for both tRNA and rRNA methylation, or being involved in pre-mRNA splicing.

Archaeal ribosome biogenesis. Although archaea do not have a nucleus, they nevertheless share many of the factors involved in eukaryotic ribosome biogenesis, including homologs for several snoRNAs and proteins involved in pre-rRNA processing (Hartman *et al*, 2006; Kuhn *et al*, 2002; Aittaleb *et al*, 2003), besides the 34 ribosomal proteins that are universally conserved.

Indeed, the overall components of archaeal ribosome biogenesis are more closely related to eukaryotes than bacteria, in terms of sequence homology and domain architecture, as well as overall number of components (Ebersberger *et al*, 2014). Thus, while rRNA processing, modification and folding in bacteria is carried out by site-specific protein enzymes, both archaeal and eukaryotic ribosome biogenesis rely on small-RNA guided enzymes to perform RNA modification steps.

However, little is known regarding the overall sequence of events in ribosome biogenesis in archaea, possibly due to the lack of advanced genetic tools to study these organisms. It is clear that there are differences in the overall organisation of the pathway. For example, *S. solfataricus* rRNA genes are found in two loci rather than in a single "operon-like" region, one housing the 16S/23S rRNAs, the other housing the 5S rRNA (Yip *et al*, 2013). Similarly, *M. jannaschii* contains two copies of each rRNA gene, one set arranged in a yeast-like operon, plus an operon containing 16S and 23S and a locus containing only the 5S gene.

As several archaea are thermophiles, the proteins and RNAs they use are more stable and amenable to *in vitro* studies than their eukaryotic counterparts. Thus, the archaeal homologs of snoRNPs have been successfully used as models for studying the biochemistry and structural biology of pre-rRNA processing *in vitro.*

1.2 rRNA methylation and the structural biology of Box C/D snoRNP system

2'-O methylation was identified as a chemical modification of rRNA over 50 years ago, by identifying nucleotides conferring protection from alkaline hydrolysis in ribonucleoprotein extracts of wheat germ (Singh & Lane, 1964) and HeLa cells (Wagner *et al*, 1967) .

In terms of chemistry, the presence of a 2'-O methyl group on nucleotides can prevent base-catalysed hydrolysis of RNA by blocking nucleophilic attack by the 2'OH ribose group, thereby increasing the stability of an RNA molecule. Furthermore, the presence of a methyl group at the 2' position seems to stabilise the C3'-endo conformation of the ribose ring (Lubini *et al*, 1994), as well as favour the formation of certain secondary structural elements over others, because of the absence of a hydrogen bonding hydroxyl group.

In *S. cerevisiae*, around 50 rRNA methylation sites have been identified, with sequences complementary to the guide regions of 76 different snoRNAs. The core *S. cerevisiae* Box

C/D snoRNP particle assembles around each snoRNA and is composed by the RNA-binding protein Snu13p, the scaffold proteins Nop56/Nop58, and the methyltransferase, Nop1p.

Each Box C/D snoRNA contains two conserved sequence elements, the C (RUGAUGA, where R is purine) and D (CUGA) box (Kiss-László *et al*, 1996). Box C/D elements come together to form a kink-turn (K-turn) structure, a well-characterised structural element that introduces a 50° bend into an RNA duplex (Huang & Lilley, 2016), (see Fig. 1.2.1). Several eukaryotic snoRNAs also have an additional set of C and D boxes, known as C'/D', which are however much less conserved (Tycowski *et al*, 1996; Kiss-László *et al*, 1998). These elements are proposed to fold into a kink-loop (K-loop) structure, though predicting the structures adopted by these elements is more challenging due to their lower sequence conservation.

Kink-turn motifs are recognised by Snu13p without the need for any additional factors *in vitro*. It has been shown by electrophoretic mobility shift assays that the human counterpart of Snu13p, 15.5K, is not able to bind K-loop structures *in vitro (Gagnon et al, 2009).* Thus, it is not clear whether Snu13p is bound to Box C'/D' elements *in vivo*, given the low conservation of these elements and the low affinity displayed by the protein. Indeed, in *Xenopus* oocytes, the RNP proteins have been found to assemble asymmetrically around the snoRNA (Cahill, 2002). However, additional factors may aid recognition of K-loops by Snu13p *in vivo*.

The guide region containing the methylation site is placed between these structural elements (see figure 1.2.1). In both eukaryotes and archaea, the nucleotide complementary to the methylated rRNA nucleotide is located 5 bases upstream of the CUGA D or D' motif.
Pull-down experiments and high resolution structures show that Snu13p is bound by the C-terminal domain (CTD) of Nop56 and Nop58. These two proteins form a coiled-coil heterodimer and their N-terminal domains (NTDs) can bind Nop1p, positioning it at the correct methylation location on the snoRNA(Gautier *et al*, 1997).

In thermophilic archaea such as *P. furiosus*, the set of 52 sRNAs all have highly similar sequences: both C/D and C'/D' boxes and two guide regions made up of 10-21 nucleotides complementary to the rRNA, with no or at most 3 nucleotides inserted between the boxes and the guide region (Fig. 1.2.1 a). Archaeal snoRNAs all have highly conserved box C/D and C'/D' elements, which fold into signature kink-turn (K-turn) and kink-loop (K-loop)

structures, respectively (Fig.1.2.1b). Both of these structures are bound by the protein L7Ae, the archaeal homolog of Snu13p, with a very high affinity. Unlike Snu13p, L7Ae is a ribosomal protein located in the LSU, highlighting a potential moonlighting role in ribosomal assembly. L7Ae is recognised by the CTD of the archaeal homolog of Nop56 and Nop58, known as Nop5. Unlike the yeast proteins, Nop5 forms its coiled-coil structure via homo-dimerization. The NTD of Nop5 can bind the archaeal methyltransferase, fibrillarin (Fib), homolog of Nop1p.

Structures of Snu13p, its human human counterpart 15.5K and L7Ae from different archaeal species show that all of these kink-turn binders adopt the same tertiary fold, though they display different specificities, since L7Ae can bind both kink-loop and kink-turn structures with very high affinity, while the eukaryotic proteins cannot (Oruganti *et al*, 2005; Soss & Flynn, 2007; Szewczak *et al*, 2005). This may be partially due to the kink-turn having multiple structures in equilibrium in solution, as shown by Förster resonance energy transfer (FRET) experiments (Goody *et al*, 2004). From the protein point of view, the difference in behaviour has been suggested to arise from a 5-amino acid motif in loop 9: introducing the snu13p sequence into L7Ae renders the protein unable to bind kink-loop structures (Gagnon *et al*, 2009; Špačková *et al*, 2010).

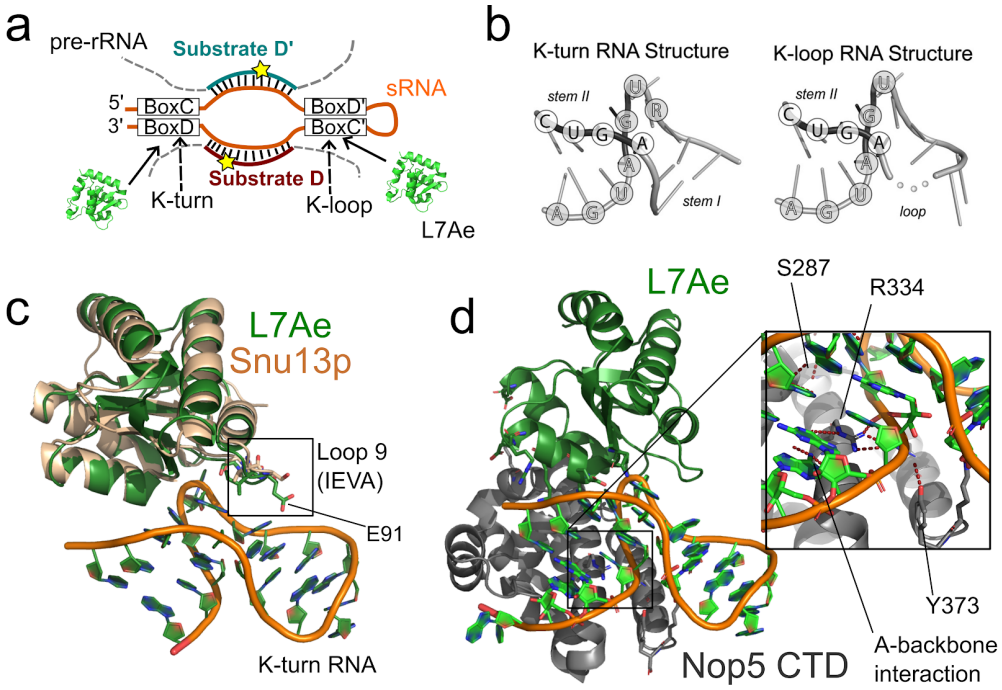


Figure 1.2.1. sRNAs, Snu13p and L7Ae. A) The conserved elements of archaeal Box C/D sRNAs: the kink-loop and kink-turn structures, which are recognised by L7Ae, and the guide regions, which bind the pre-rRNA substrates and direct methylation to the nucleotide complementary to the fifth nucleotide before Box D and Box D'. B) The K-turn and K-loop folds formed by Box C/D and Box C'/D', respectively. Reproduced with permission from Gagnon *et al*, 2010. C) Although yeast Snu13p (pdb accession code 2ALE) and *P. furiosus* L7Ae (pdb accession code 3NMU) are homologs sharing the same tertiary fold, only L7Ae recognises both K-loop and K-turn structures with a high affinity. This is at least partly due to differences in loop 9 of the two proteins, which in L7Ae is made up by the conserved IEVA sequence, while in Snu13p and other eukaryotic proteins this motif is VSRP. D) The recognition of L7Ae/sRNA by the C-terminal Nop domain of *P. furiosus* Nop5. The Nop5 CTD makes contact with the bases of Box C via R334 (see inset), and uses indirect readout to recognise the phosphate backbone of the conserved box elements. (pdb accession code 3NMU)

Nop56 and Nop58 were first identified as core components of ribosomal processing in genetic screening (Gautier *et al*, 1997), which also identified their interaction with Nop1p. Despite their 45% sequence identity, the two proteins differ in behaviour: while the NTD of Nop56 is known to form a tight, purifiable complex with Nop1p, pull-down seems to indicate that the affinity of Nop58 for the methyltransferase is lower, or at least that only a small fraction of Nop58 is bound to Nop1p at any given time (Gautier *et al*, 1997). However, both proteins are found bound to Nop1p in the structure of U3 snoRNP within the 90S pre-ribosome (Sun *et al*, 2017).

As stated above, Nop56 and Nop58 are known to heterodimerize via their coiled-coil domain, which also contains a 56-amino acid α-helical "insert", which does not play a role in dimerization. The proteins also share a C-terminal Nop domain that can interact with the Snu13p/snoRNA complex. The C-terminus of both proteins in eukaryotes also contains several KKxD repeats, though these motifs are not required for protein function *in vivo*. The archaeal homolog of Nop56 and Nop58, Nop5, shares their overall domain arrangement and an NTD capable of binding the Fib methyltransferase, a central coiled-coil domain containing a 56-amino acid insert and a C-terminal Nop domain that can recognise L7Ae/K-turn or L7Ae/K-loop RNA; however Nop5 lacks the C-terminal lysine-rich region. In both eukaryotic and archaeal proteins, the three main domains are connected by flexible linìkers that ensure that the enzyme can move the methyltransferase on the sRNA/pre-rRNA duplex, or away from it.

Sequence alignments of Nop5, Nop56 and Nop58 also identify the short α9 helix in the CTD and the ALFR motif as highly conserved (see Fig. 1.2.2). Indeed, high-resolution studies and

cross-linking show that this region of the protein forms an α-helical protrusion that interacts specifically with the RNA by inserting itself between the guide RNA regions and plying the end of kink-turn apart in both eukaryotes and archaea (Ghalei *et al*, 2010; Ye *et al*, 2009; Lapinaite *et al*, 2013) (see Fig 1.2.3).

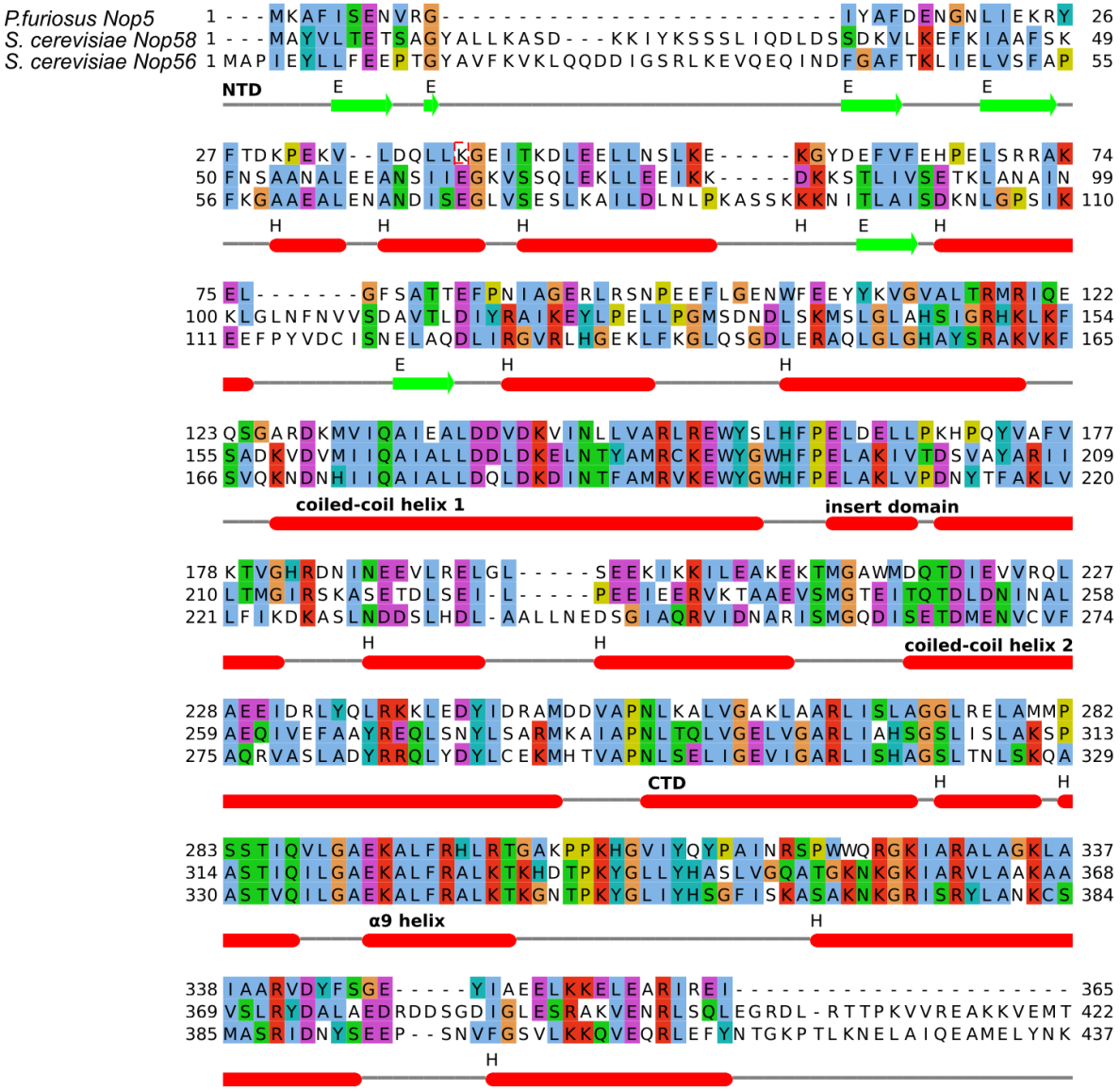


Fig. 1.2.2. Structure-based sequence alignment of Nop56/58 proteins Secondary structure annotation derived from the structure of *P. furiosus* Nop5. Alignment performed with PROMALS3D (Pei *et al*, 2008) by superimposing pdb accession codes 4BY9 (Nop5) and 5WYK (Nop56, Nop58).

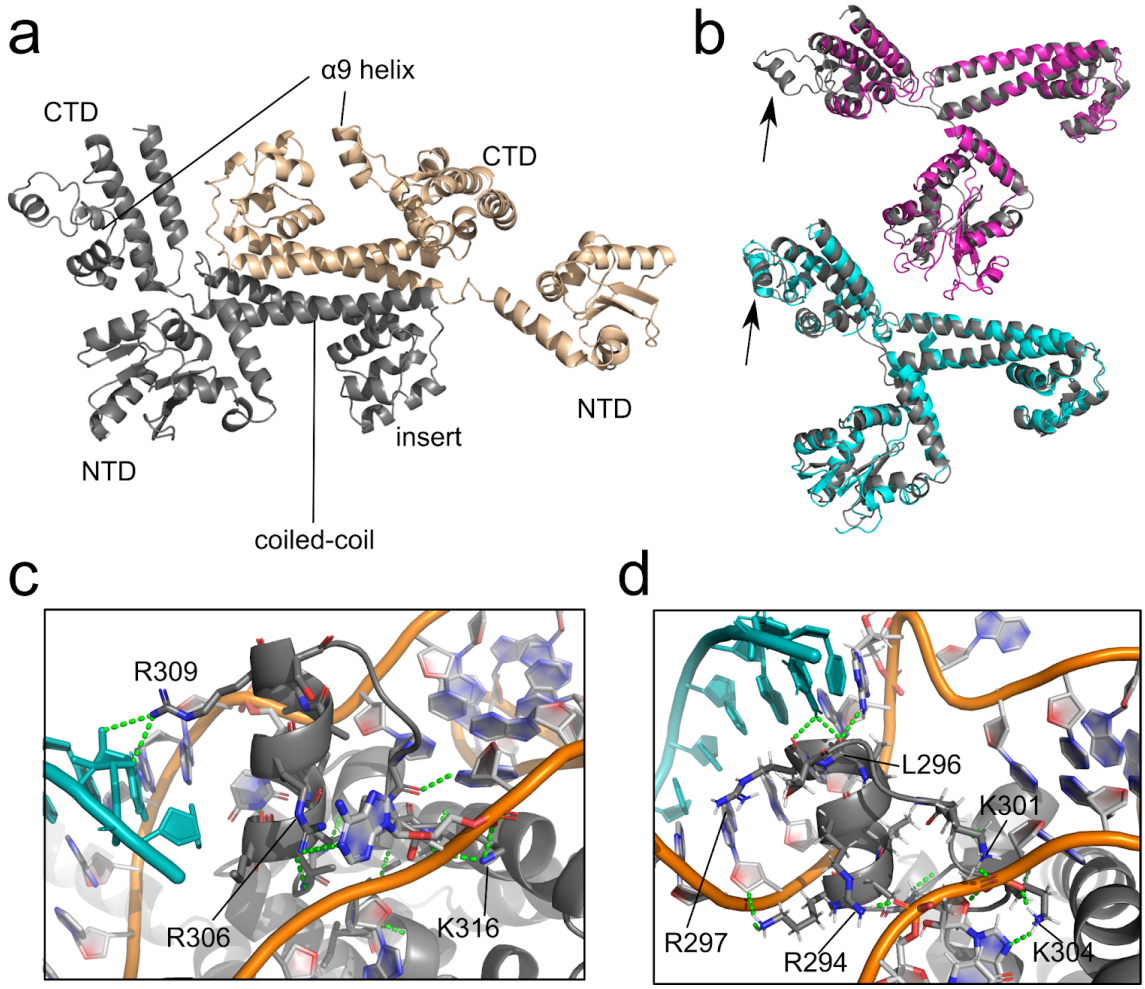


Figure 1.2.3. Nop56, Nop58 and archaeal Nop5 A) The structure of the *P. furiosus* Nop5 homodimer (PDB accession code 4BY9). The N-terminal domain is responsible for binding Fib; homo-oligomerization occurs via the central coiled-coil domain, and the C-terminal Nop domain recognizes L7Ae/RNA. B) Superposition of Nop5 (gray, PDB accession code 4BY9) onto *S. cerevisiae* Nop56 (magenta, PDB accession code 5WYK) and Nop58 (cyan, pdb accession code 5WYK) structures. Each domain is aligned individually. The position of the α9 helix is marked by arrows. The α9 helix of Nop56 is not resolved in the electron density. The structures further confirm the conservation of Nop56/58 folds between archaea and eukaryotes. C) The interaction between the Nop5 α9 helix and the sRNA K-turn and guide regions, as seen in the crystal structure of the mono-RNP reconstituted with a two-piece sRNA (PDB accession code 3PLA). The substrate RNA is displayed in cyan. In the holo mono-RNP model, the helix sits above the two sRNA guide regions. D) The interaction between the Nop5 α9 helix and the sRNA K-turn and guide regions, as seen in the *P. furiosus* holo di-RNP structure (PDB accession code 4BY9). The substrate RNA is displayed in cyan. In the holo di-RNP model, the helix sits in the plane of the sRNA, between the two guide regions. The R294, K301 and K304 interact with the sRNA downstream of the box C and box C' elements, while interaction with the substrate-guide duplex is made by the backbone of R297 and L296.

Archaeal Fib, yeast Nop1p and human fibrillarin methyltransferases share a highly conserved C-terminal domain that adopts the canonical nucleotide-binding Rossman fold, comprising of a seven-stranded β-sheet flanked by three α-helices on either side (see Fig.1.2.4), commonly found in methyltransferases (Martin & McMillan, 2002). The CTD of these methyltransferases contains the consensus motif for binding the S-adenosyl-methionine (SAM) cofactor, which provides the methyl group for the enzymatic reaction. After loss of the S-methyl group, SAM converts into S-adenosyl homocysteine (SAH). Despite the high conservation of the catalytic CTD, the NTD is highly variable: most eukaryotic box C/D RNA methyltransferases have a GAR domain, necessary for nuclear localisation, in their N-terminus, while archaeal NTDs are shorter and vary between species.

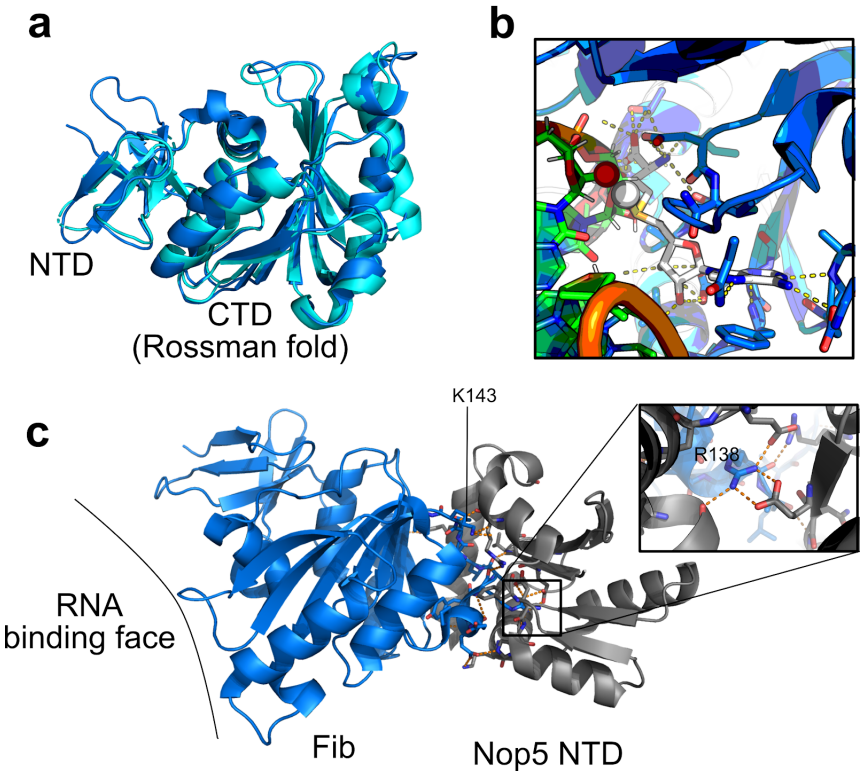


Figure 1.2.4. Fibrillarin. A) Human Fibrillarin (cyan, PDB accession code 2IPX) superimposed onto *P. furiosus* Fib. A (dark blue, PDB accession code 4BY9). The Cα-Cα RMSD is 1.1Å. The CTD adopts the typical methyltransferase Rossman fold B) The structure of Fib from the archaeon *Aeropyrum pernix* (PDB accession code 4DF3, shown in blue), which has been solved in the SAM-bound state, shows the arrangement of the SAM-binding pocket in the active site of the enzyme. Here, the cofactor is displayed in white and also shown relative to the position of rRNA containing the 2'-OH that accepts the methyl group from SAM. The 2'-OH group is shown as a red sphere, while the

SAM methyl group is shown as a white sphere. The RNA position is modelled by superposition onto the structure of Fib bound to RNA (PDB accession code 3PLA). C) The Nop5 N-terminal domain-Fib interaction is highly specific. The loop made up by residues 134-144 of *P. furiosus* Fib (blue) makes several polar contacts (orange dashes) with the Nop5 NTD (gray, PDB accession code 4BY9). In particular, arginine 138 of Fib forms hydrogen bonds to residues lying on three separate secondary structure elements within the Nop5 NTD (inset), effectively reading the tertiary fold of the domain.

While it has been possible to assemble biochemically active archaeal box C/D complex *in vitro* from proteins expressed in *E. coli* and *in vitro* transcribed sRNAs, the reconstitution of the yeast or human box C/D complex has so far been unsuccessful, due to the low solubility of Nop56 and Nop58, as well as their tendency to form homo-dimers rather than hetero-dimers, among other factors. Nevertheless, eukaryotic box C/D complexes have been purified by pull-down, or reconstituted with co-expression strategies in *E. coli*, although these complexes proved to be not active (Peng *et al*, 2014). Thus, it is possible that archaeal sRNPs may assemble *in vivo* without any additional factors, while eukaryotic snoRNPs are known to require a complex assembly pathway that includes a protein-only pre-complex, where the inhibitory factor Rsa1p binds Snu13p and Nop1p is kept in an inactive state. In later stages of snoRNP biogenesis, Rsa1p is released upon binding of the methyltransferase and snoRNA (Bizarro *et al*, 2014; Quinternet *et al*, 2016; Lechertier *et al*, 2009). The archaeal sRNP assembly may instead proceed via a methylation-competent Nop5-Fib sub-complex prior to the assembly of the full sRNP (Tomkuvienė *et al*, 2017; Zhang, 2006).

For these reasons, there have been several high-resolution structures of archaeal sRNPs, while full eukaryotic snoRNPs have only been characterised as part of the 90S pre-ribosome, in the special case of U3 snoRNP. Further insights into eukaryotic snoRNPs may also be gained by analysis of the snRNPs forming the spliceosome, which have been characterised at a high resolution (Nguyen *et al*, 2015; Zhang *et al*, 2017). However, the available sRNP structures, combined with *in vitro* biochemistry and sequencing data, portray a highly complex and diverse set of architectures and mechanisms for the structure and function of the archaeal box C/D complex and the regulation of its methylation cycle.

Initial structural information on the archaeal box C/D core proteins led to the proposal of a 200 KiloDalton (KDa) “mono-RNP” architecture for the complex, where the Nop5 coiled-coil holds together two copies of Fib around a single sRNA and two copies of L7Ae, which bind the C/D kink-turn and the C’/D’ kink-loop (Aittaleb *et al*, 2003; Nolivos *et al*, 2005). However, a low-resolution negative stain electron microscopy map of the full *M. jannaschii* box C/D

particle assembled around a physiological sRNA showed that the complex adopts a "di-RNP" architecture *in vitro*, where four copies of Nop5, L7Ae and FIb surround a pair of sRNAs, with a molecular weight of approximately 400 KDa (Bleichert *et al*, 2009). This result was followed up by a high-resolution structure obtained by macromolecular crystallography (MX) consistent with an sRNA path going across a diRNP particle, rather than along the Nop5 coiled-coil domain, as predicted by the monoRNP model (Xue *et al*, 2010). Recently, a structure of the di-RNP complex in obtained by cryo-electron microscopy at 9 Å resolution was reported (Yip *et al*, 2016).

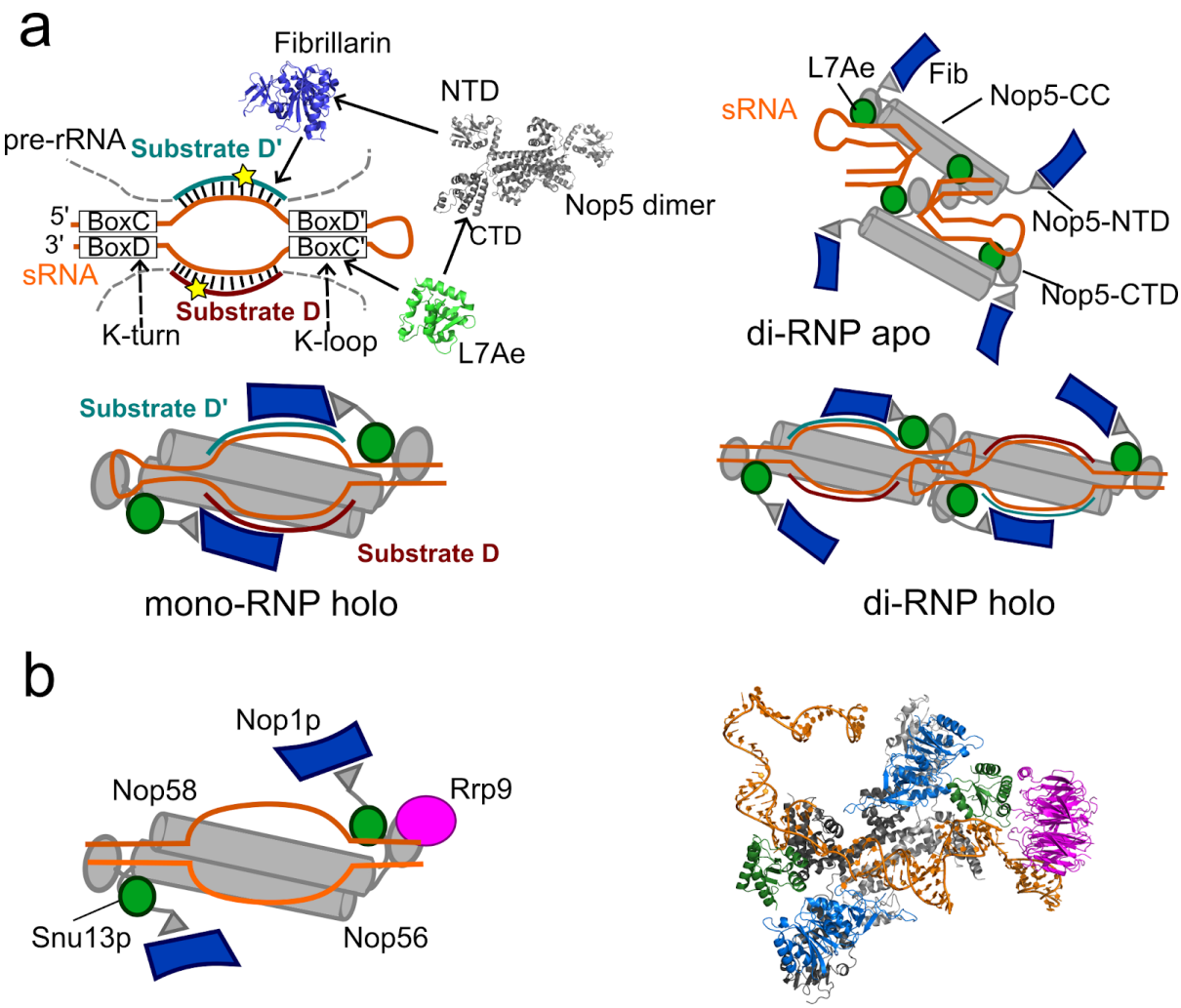


Figure 1.2.5. The mono-RNP and di-RNP model for the holo Box C/D sRNP particle. A) Interactions in the Box C/D sRNP in archaea and observed oligomeric states of the particle. In the mono-RNP model (Lin *et al*, 2011; Yang *et al*, 2016), two copies of Nop5, L7Ae and Fib are assembled around a single sRNA. In the di-RNP model (Bleichert *et al*, 2009), four copies of each protein bind two sRNAs, with the complex being held together by interactions between Nop5 domains. B) The *S. cerevisiae* snoRNP particle has been proposed to form as mono-RNP, as seen in the U3

snoRNP found in the *S. cerevisiae* pre-ribosome (PDB accession code 5WYK) However, as Snu13p does not bind the K-loop structure predicted to be formed by Box C'/D', additional loading factors may be required for the formation of the full snoRNP. **"Apo"** indicates the full sRNP without the substrate RNAs; **"holo"** indicates the full sRNP with both kinds (substrate D and substrate D') of substrate pre-rRNA bound.

High-resolution structures of archaeal particles assembled in a mono-RNP architecture have also been reported. In all of the mono-RNP structures, the proteins are assembled around a pair of complementary RNAs, rather than a single physiological sRNA sequence. Thus, the high-resolution structure of a Box C/D sRNP from *S. solfataricus* loaded with two complementary pre-rRNA sequences was obtained by reconstituting the proteins with a "two-piece" sRNA (Lin *et al*, 2011; Yang *et al*, 2016) (see Fig.1.2.7). Moreover, some of the structures from complexes reconstituted with incomplete or split sRNAs can be interpreted in the framework of a mono-RNP model (Ye *et al*, 2009), or as part of a di-RNP architecture.

Table 2. Structures of the archaeal Box C/D sRNP

PDB/EM DB code	Organism	Method	sRNA	state	Architecture	citation
3pla	*S. solfataricus*	MX	two-piece	holo	monoRNP	1
3nvi	*P.furiosus*	MX	K-turns only	apo	half-RNP	2
3nmu	*P. furiosus*	MX	K-turn +1 guide sequence	holo	half-RNP	2
3id5	*S. solfataricus*	MX	K-turn +1 guide sequence	holo	half-RNP	3
5gin, 5gio, 5gip	*S. solfataricus*	MX	two-piece	holo	monoRNP	4
4by9	*P. furiosus*	NMR/SAS	symmetric sRNA	holo	diRNP	5
1636	*M. jannaschii*	negative stain EM	sRNA	apo	diRNP	6
5419	*S. solfataricus*	negative stain EM	sRNA	apo	diRNP	7
8146	*M. jannaschii*	Cryo-EM	sRNA	apo	diRNP	8

1: Lin *et al*, 2011; **2**: Xue *et al*, 2010; **3**: Ye *et al*, 2009; **4**: Yang *et al*, 2016; **5**: Lapinaite *et al*, 2013; **6**: Bleichert *et al*, 2009; **7**: Bower-Phipps *et al*, 2012; **8**: Yip *et al*, 2016

In 2013, our group published the structure of the Box C/D sRNP from *P. furiosus* obtained from an integrative structural biology approach that combined nuclear magnetic resonance (NMR) and small-angle scattering (SAS) data (Lapinaite *et al*, 2013)(See figure 1.2.6). The complex was reconstituted with a near-physiological one-piece sRNA. This sRNA, referred to as ssR26, is a modified version of the physiological *P. furiosus* sR26 sRNA with symmetric features -two identical guide regions, and two kink-turn motifs, rather than a kink-turn and a kink-loop.

NMR investigation of Box C/D reconstituted with ssR26 showed that this complex has a di-RNP architecture in both **apo** and **pre-rRNA loaded (holo)** states, and is catalytically active. Moreover, the splitting of Fib resonances upon adding the substrate pre-rRNA indicates that two copies of Fib are bound to the substrate-guide duplex, while two copies remain in away from the active site.

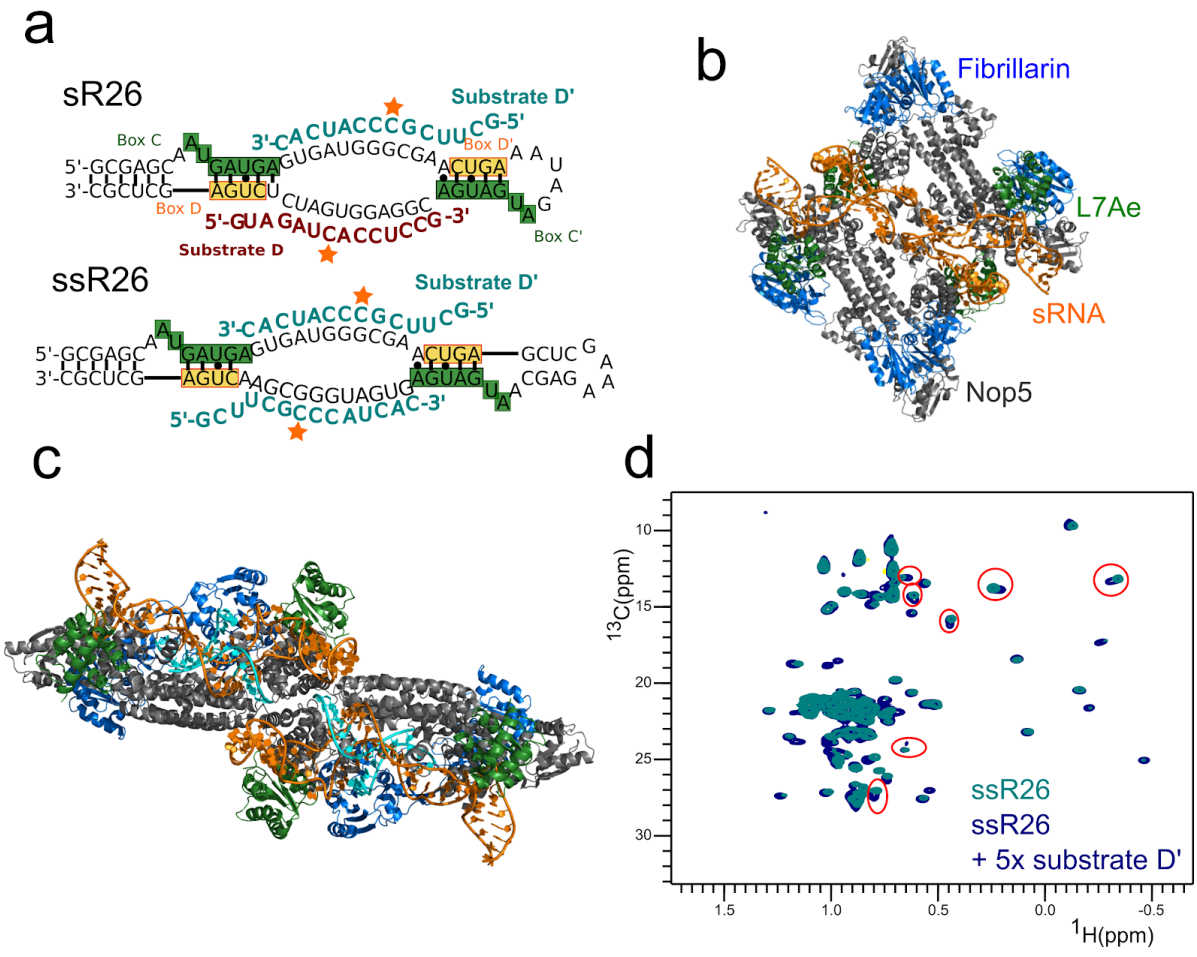


Figure 1.2.6. The structure of the archaeal Box C/D sRNP reconstituted with ssR26. A) The *P. furiosus* guide RNA, sR26, and its counterpart used in structure determination, ssR26 (Lin *et*

al, 2011; Lapinaite *et al*, 2013). B) The apo structure, showing a di-RNP in a flat arrangement with the four Fib molecules positioned away from the RNA. C) In the holo state, the complex adopts an extended conformation and positions two copies of Fib onto the substrate-guide duplex. Thus, the di-RNP model for the holo complex implies that only half the sites can be methylated by the enzyme at any given time. D) ^{1}H-^{13}C HMQC Spectra of ILV-labelled Fib within the full Box C/D sRNP reconstituted with ssR26 in the apo(light blue) and holo (dark blue) states. The splitting of the resonances highlighted by red circles indicates that Fib experiences two separate chemical environments, consistent with the fact that two out of the four copies of the methyltransferases are away from the RNA.

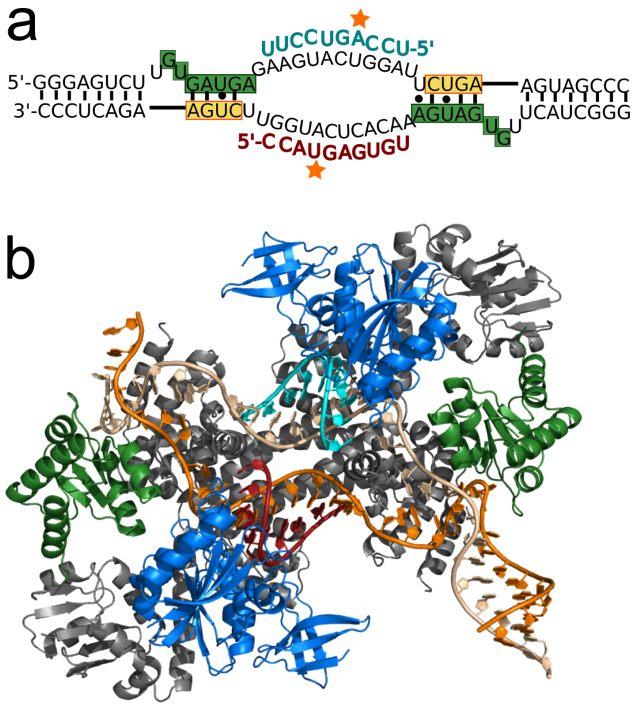


Figure 1.2.7. The structure of the holo archaeal Box C/D sRNP in the mono-RNP state. A) The two-piece RNA construct used for crystallisation of the holo mono-RNP (Lin *et al*, 2011) . B) The structure shows the *S. solfataricus* structure (PDB accession code 3PLA) with both Fib positioned onto the substrate-guide duplex and the Nop5 coiled-coil running at a 70° angle with respect to the sRNA. Color scheme as in Fig. 1.2.6.

The wide range of contradicting structural information is summarised in table 2 and figures 1.2.5, 1.2.6, 1.2.7. The fact that there is no clear agreement on the oligomeric state and topology of the complex can be attributed to the fact that the available models differ from one

another in RNA constructs and reconstitution methods used, as well as whether they are in the apo or in the holo state. However, it is also possible that this plasticity is reflective of the various oligomeric states the complex could adopt *in vivo*.

Several biochemical experiments have investigated the principles behind the architecture of the archaeal Box C/D *in vitro.* One of the key elements seems to be the presence of the apical loop, which is missing in “two-piece” sRNA structures. Studies on particles reconstituted with two-piece RNAs report the formation of both a mono and a diRNP particle during purification. Reconstitution followed by native gel electrophoresis of complexes assembled with and without apical loop sequences show that this region of the sRNA is necessary for the formation of a diRNP across a number of sRNAs and archaeal species, at least in the apo state (Bower-Phipps *et al*, 2012). However, this does not address whether the apical loop influences the architecture of the sRNP in the holo state. Another factor that influences sRNA flexibility and perhaps complex assembly is the interaction between the Nop5 α9 helix and the sRNA guide regions: while deleting this helix leads to increased sRNA flexibility, it may also ultimately affect the oligomeric state of the particle, since the interpretation of structures missing this helix is controversial (Xue *et al*, 2010). Finally, the recently solved cryo-EM structure in the apo state hypothesizes that the stem regions of the two sRNAs in a diRNP interact with each other, leading to increased stability of this architecture.

One of the main drawbacks of the monoRNP model is that, since the sRNA has to form a path dictated by the length of the Nop5 coiled-coil, it cannot really explain how an sRNA can recognise a target sequence that is longer than 10 nucleotides, despite such sequences being present in archaea. Indeed, recent monoRNP structures show that while the space between K-turn and K-loop can be shorter or longer than 10 nucleotides, optimal methylation activity is observed with a 10-nucleotide duplex between guide and substrate pre-rRNA, with longer sequences being unwound in order to achieve the correct duplex orientation (Yang *et al*, 2016). This is supported by sequencing data showing that archaeal substrate-guide pairing are mostly 10 or 11 nucleotide long, but is not consistent with the observation that a complementary RNA:RNA duplex covering the whole length of the complementary region is required for optimal methylation *in vitro (Appel & Maxwell, 2007)*. These apparently contradicting results may be reconciled by considering that the 2016 study revealed that, while complementarity may be required beyond the 10-nucleotide duplex, this complementary region does not need to form an A-form helix, though it still does not tolerate mutations. Indeed, adding nucleotides beyond the target:substrate duplex is found to

stimulate methylation activity, perhaps by interacting with other parts of the complex, a phenomenon first reported *in vivo* in *S. cerevisiae* (Appel & Maxwell, 2007), (Cavaillé & Bachellerie, 1998)

Thus, although the structures of the individual components of the Box C/D sRNP particle are well understood, it is not clear how the proteins and the sRNA come together to produce a regulated catalytic cycle, or whether the enzyme does assemble in a single oligomeric state *in vivo*. However, the high sequence conservation and structural similarity with the eukaryotic proteins, coupled with what seems a similarly complex set of sRNA sequences, lead us to believe that the archaeal system is a valid model system for eukaryotic rRNA methylation. Therefore, understanding the rules underlying Box C/D assembly and activity may provide critical insights into the regulation and activity of the system in higher organisms.

In order to understand the structure and dynamics of the system, we used a combination of structural approaches in solution and biochemical assays, briefly reviewed in the "methodological background" section below.

1.3 Methodological Background - solution nuclear magnetic resonance of high molecular weight macromolecules

Macromolecular NMR traditionally relies on a set of heteronuclear correlation experiments that lead to the observation of protons bound to nitrogen or carbon atoms. In solution, the line width of the observable NMR signal depends on the relaxation properties of the sample, quantified in the longitudinal and transverse relaxation rates R_1 and R_2. R_1, the rate governing the growth of the signal along the main magnetic field, is dependent largely on the strength of the external field and its fluctuations. On the other hand, R_2 governs the rate of decay of magnetisation in the transverse plane, and is dependent on fluctuations in the local magnetic field in any direction.

As molecular motion is one of the main sources of fluctuating magnetic fields in the sample, R_2 is dependent on the rotational correlation time of the molecule being observed. Thus, NMR line widths become broader with increasing protein size. As peaks broaden, the signal-to-noise ratio and effective resolution of NMR spectra decreases due to lower peak intensity at equal concentration and higher peak overlap. This effectively limits "traditional" NMR structure determination approaches to systems smaller than roughly 30 KDa.

Over the past 15 years, a combination of novel NMR pulse sequences, experiments and sample preparation techniques has extended the range of macromolecular NMR into the

hundreds of KDa by addressing the relaxation and spectral crowding problems presented by large macromolecules.

Two of the main factors influencing the transverse relaxation rate R_2 are dipole-dipole interactions and chemical shift anisotropy (CSA). Both these mechanisms lead to a broadening of the observed resonances, be it via a fluctuation of the local chemical environment (CSA), or via the fluctuations of local dipolar fields (dipole-dipole interaction). As both these relaxation mechanisms are coupled to the rotation of the macromolecular frame, their relative contribution depends on the size of the macromolecule. Thus, so-called "transverse relaxation optimisation spectroscopy" (TROSY) approaches have been developed in order to select the NMR line shapes that derive from pathways with minimal effective contribution of the CSA and dipole-dipole relaxation terms (Pervushin *et al*, 1997). Specifically, the TROSY approach ensures that these two terms are combined in such a way that they cancel each other out.

Although TROSY approaches were initially developed for ^{15}N-^{1}H correlations, it is their extension to ^{1}H-^{13}C experiments and their use in the NMR detection of methyl groups that unlocked the use of NMR for very large (above 100 KDa) systems.

In an approach known as methyl-TROSY, first developed in the Kay laboratory (Ollerenshaw *et al*, 2003; Tugarinov *et al*, 2003; Tugarinov & Kay, 2004), perdeuterated proteins containing ^{1}H- ^{13}C-labelled isoleucine, leucine and valine methyl groups are measured using ^{1H}- ^{13C} TROSY or HMQC experiments; nowadays it is possible to selectively label the methyl groups of methionine and alanine as well. The ability to observe methyl groups in large systems depends not only on the TROSY principle, but also on the fast rotation of methyl groups around the single bond connecting them to the rest of the amino acid side chains, which partially unclouples the methyl group from the slow tumbling of the large macromolecule. Furthermore, perdeuteration ensures minimal ^{1}H-^{1}H cross-relaxation via spin diffusion, as no protons are available to provide a "bath" conducive to magnetization loss.

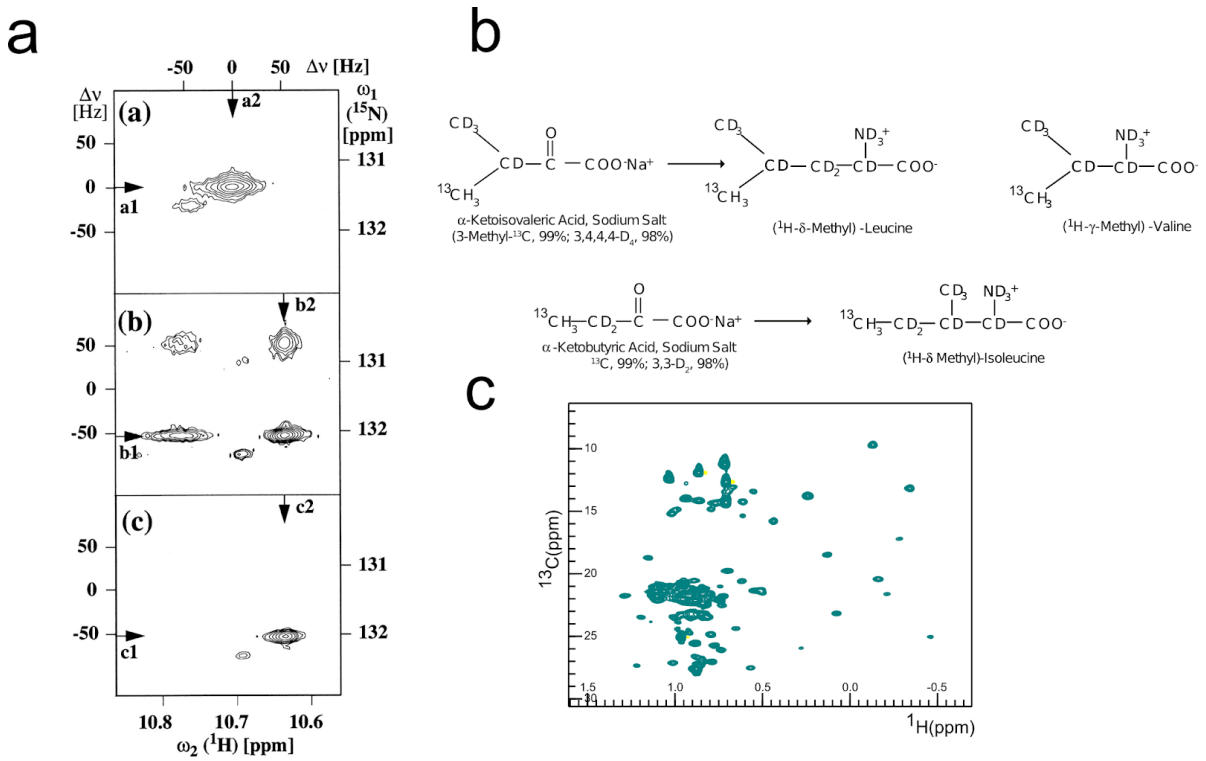


Figure 1.3.1. Solution NMR of large systems. A) Top: the peaks in a conventional decoupled ^{15}N-^{1}H HSQC experiment. Middle: the non-decoupled line shapes of ^{15}N-^{1}H HSQC experiment. Bottom: The line shape improvement achieved by a TROSY spectrum, which compensates for dipole-dipole interaction and CSA. Figure reproduced with permission from Pervishin *et. al*, 1997. B) Labelling of isoleucine, leucine and valine methyl groups is achieved by adding labelled biosynthetic precursors to minimal media prior to induction (Goto *et al*, 1999). The precursors shown here result in the labelling of the terminal methyl group; though precursors that generate full side-chain ^{13}C-labelled amino acids, required for NMR assignment experiments, are also available. C)The ^{1}H-^{13}C HMQC spectrum of ^{2}H,[^{1}H,^{13}C]-ILV methyl labelled Fibrillarin within the 400 KDa Box C/D sRNP complex. The methyl-TROSY approach relies on ^{1}H-^{13}C HMQC experiments acquired on perdeuterated, [^{1}H-^{13}C]-ILV methyl labelled proteins. The resulting spectra display narrow line shapes even when the sample molecular weight is in the hundreds of KDa.

Although the methyl-TROSY approach has extended the range of NMR experiments considerably, it still is limited to probing methyl groups, which means that normal triple-resonance experiments relying on transfer of magnetization through the amino acid side chains may not be used in systems as large as the Box C/D RNP. Therefore, assignment of resonances in such a complex relies on the transfer of assignments from the spectra of individual subunits, in a process known as the “divide and conquer” approach.

1.4 Methodological background - paramagnetic relaxation enhancement and distance restraints.

Since its description for biomolecules more than thirty years ago, the paramagnetic relaxation enhancement (PRE) effect has become a key tool used by NMR spectroscopists to obtain precise distance information in the 15-30 Å range. Historically, the use of PRE in biomolecular NMR was limited to metal-containing proteins with a naturally occurring paramagnetic centre. However, nowadays paramagnetic centres can be introduced in proteins by site-directed mutagenesis followed by spin labelling; the paramagnetic centre (spin label) is chemically conjugated to a sulphur atom of an exposed cysteine that is introduced at a selected position. A commonly used spin label is 3-(2-iodoacetoamido)-PROXYL, which contains a paramagnetic nitroxide group within a TEMPO moiety that links to the protein via rotatable bonds (See Fig. 1.4.1).

The PRE effect relies on the interaction between a free electron and the NMR-active nuclei being observed in NMR experiments. This interaction results in an increase in the nuclear relaxation rates R_1 (longitudinal) and R_2 (transverse), which, for paramagnetic systems with an isotropic electron g-factor such as nitroxides, is dominated by the dipolar relaxation mechanism.

As first described by Solomon and Bloembergen (Solomon & Bloembergen, 1956), such an enhancement in transverse relaxation rates of nuclei caused by an electron depends on the inverse sixth power of distance between the nucleus being observed and the paramagnetic centre (r^{-6}).

$$\Gamma_2 = \tfrac{1}{15} S(S+1)\gamma^2 g^2 \mu_B^2 \{4J_{SB}(0) + 3J_{SB}(\omega) \qquad (1)$$

Here, S is the spin quantum number of the electron, γ is the gyromagnetic ratio of the nucleus being observed and μ_B is the Bohr magneton, which expresses the magnetic moment of the electron **Γ_2 is the transverse PRE rate** and J_{SB} is the generalised spectral density function of the correlation being observed, and includes the r^{-6} dependency of the term:

$$J_{SB}(\omega_I) = r^{-6} \frac{\tau_c}{1+\omega^2 \tau_C{}^2} \quad (2)$$

τ_c is the correlation time of the electron-nucleus interaction, which includes both the rotational correlation time of the whole macromolecule and the effective electron relaxation time.

The PRE rate **Γ_2**, which is the **difference between the transverse relaxation rate in the diamagnetic state and in the paramagnetic state**, can be quantified by measuring the peak intensities of spectra of the two states (I_{para} and I_{dia}), provided the spectra display Lorentzian line shapes.

$$\frac{I_{para}}{I_{dia}} = \frac{e^{(-\Gamma_2 t)} R_2^{dia}}{R_2^{dia} + \Gamma_2} \qquad (3)$$

Where R_2^{dia} is the diamagnetic relaxation contribution, and *t* is the evolution time during which magnetisation is in the transverse plane.

In practice, this quantification is done by measuring of peak intensities in 2D heteronuclear correlation spectra (in our case, ^{1}H-^{13}C HMQC spectra, see section 1.3) of the spin labelled system in the paramagnetic and diamagnetic states; i.e. before and after addition of a reducing agent, ascorbic acid. Other approaches, where the effective relaxation rate is directly measured by time-point experiments, are recommended, though in practice they may not be feasible for high-molecular weight complexes that require methyl-TROSY approaches, because of either low sample concentrations or fast relaxation rates.

The ratio between peak intensities in the paramagnetic (I_{para}) and diamagnetic (I_{dia}) state may be translated to distances between the NMR-active nuclei and the paramagnetic electron.

First, the set of diamagnetic relaxation rates (R_2^{dia}) corresponding to the resonances being observed is quantified using experiments modified to introduce relaxation delays (see section 3.7 and Fig 3.7.1 for details); then, Γ_2 can be used to derive a distance once the correlation time of the electron-nuclear interaction vector, τ_c, is known, using equation (2). τ_C is made up by the correlation time of the whole molecule and the correlation time of motions of the vector between the electron and the nucleus being observed in the molecular frame.

Measuring τ_c directly is often not experimentally feasible; however, it can be derived empirically by quantifying PREs deriving from known distances and diamagnetic relaxation

rates, in an approach developed by the Clore laboratory (Iwahara *et al*, 2004) (See Fig.1.4.1). In this approach, the τ_c of the interaction between the nucleus being observed and an ensemble of conformations of a paramagnetic tag is minimised against experimentally-derived Γ_2 in a simulated annealing procedure.

The distances deriving from multiple spin label positions and labelling schemes can then be applied as restraints in traditional structure calculation approaches or in docking, or for scoring *ab initio*-derived models. However, as they only provide sparse information, they are often combined with approaches providing data on the position and orientation of whole subunits in a complex, such as small-angle X-Ray or neutron scattering experiments.

It is worth mentioning that, as probing high-molecular weight complexes relies on labelling of methyl groups, no information on nucleic acids may be obtained by PRE approaches on large macromolecules in solution, although solid-state approaches have shown promise in that regard. This technology is however still in development and has not been used in this work.

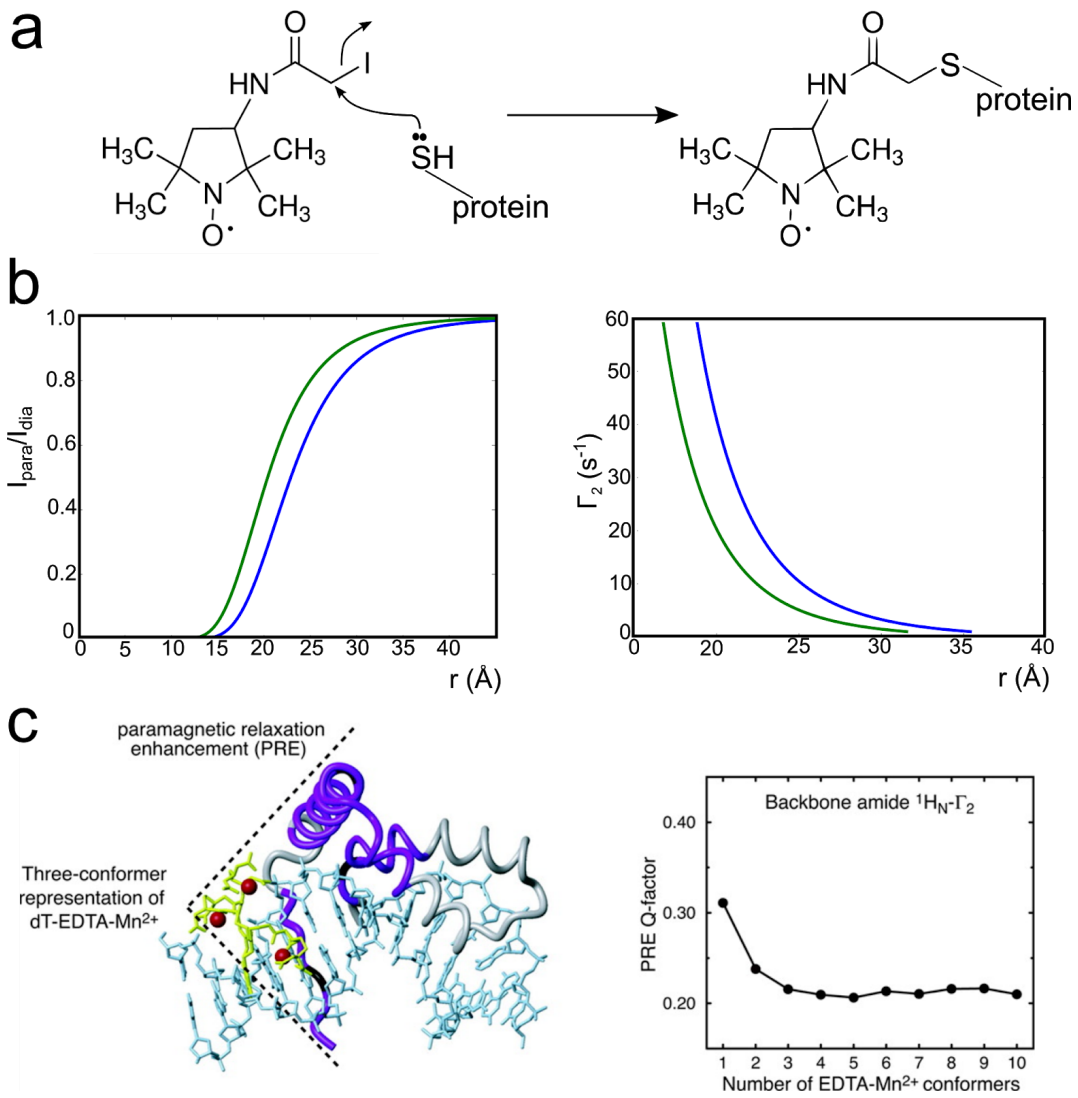


Figure 1.4.1. Paramagnetic relaxation enhancement. A) The reaction of 3-(2-iodoacetoamido)-PROXYL with cysteine side chains. B) The correlation between PRE rate Γ_2 and distance between the paramagnetic electron and a proton in a methyl group. Left: derivation of electron-nucleus distances (r) from the experimentally-measured ratio between paramagnetic and diamagnetic state intensities (I_{para}/I_{dia}). Right: dependence of Γ_2 on distance. Green line: τ_c= 20ns; blue line: τ_c= 40ns. C) The implementation of multiple spin label conformations to extract an apparent distance via r^{-6} averaging, as developed by the Clore group. Briefly, the paramagnetic group (in this case, a tag chelating a Mn^{2+} ion, shown as a red sphere) is represented as an ensemble of conformers, from which an r^{-6}-averaged distance is extracted. This distance is used to back-calculate a theoretical Γ_2 from computed structures, which is compared to experimental data using the Q-factor statistic. As the plot on the right shows, agreement with experimental data increases upon incorporation of multiple spin label conformers. This method can be applied to a set of PRE restraints within a known structured domain in order to extract the electron-nucleus correlation time τ_c. Reproduced with permission from Iwahara *et al.,* 2004.

1.5 Methodological background - Small angle X-ray and neutron scattering.

Small-angle scattering (SAS) can provide key low-resolution information about the overall shape and quaternary structure of biological assemblies. Together with NMR, SAS techniques have the advantage of probing the sample in near-physiological conditions. Furthermore, small-angle neutron scattering (SANS) does not perturb the sample, as neutrons do not cause radiation damage.

A typical SANS experiment uses neutrons produced by spallation sources and moderated to the Å wavelengths required by biological samples. On the other hand, small-angle X-ray scattering (SAXS) is often performed using synchrotron X-ray radiation with wavelengths in the order of 0.5-2Å. The sample, typically an isotropic diluted solution in the mg/ml concentration range, generates a 2D scattering pattern that is radially symmetric around the beam center and is a function of s, the momentum transfer along the detector, which has units of inverse nm or Å (see Fig.1.51a).

$$s = \frac{4\pi\, sin\theta}{\lambda} \qquad (4)$$

Where θ is the angle between the incident and the scattered beam and λ is the wavelength of the neutron or X-ray beam.

The dependency of the net X-ray or neutron scattering intensity I on the momentum transfer s can be expressed as a function of the scattering of a particle averaged over all possible orientations:

$$I(s) = < \int |(\rho(r) - \rho_s) e^{-i(s \cdot r)} dr|^2 > \qquad (5)$$

Where r is the coordinates of a scattering point in the molecular frame, and $(\rho(r) - \rho_s)$ is the difference between scattering density of the individual point and the scattering density of the solute. This quantity, known as scattering length density (SLD) depends on different parameters for X-ray and neutrons. In the case of X-rays, this quantity is the difference in electron density between the scattering point and the solvent. For neutrons, $\rho(r)$, and therefore SLD, is a physical property of the nuclei being considered, and varies randomly across the periodic table, with different isotopes yielding differend SLDs. Therefore, in

neutron scattering experiments the SLD of biomolecules may be manipulated by deuteration This forms the basis of the “contrast matching” technique (see below).

In an experiment, two measurements are made, one for the sample and one for the buffer. The signal from the buffer is then subtracted from the sample in order to obtain the scattering pattern of the solute.

From the scattering curve, a set of structural parameters can be derived. The Guinier approximation (2) (Feigin & Svergun, 1987; Guinier & Fournet, 1955) dictates that in the initial region of the curve intensity decays in a manner that is dependent on the radius of gyration R_g, defined as the root-mean-squared distance to the center of scattering mass (1). The y-intercept of the curve, defined as I(0), is proportional to the volume occupied by the solute and therefore, if the sample concentration is known, to the molecular mass.

$$R_g{}^2 = \frac{1}{N}\sum_{i=1}^{N}(r_i - r_{mean})^2 \qquad (6)$$

$$I(s) = I(0)exp(\frac{-s^2 R_g{}^2}{3}) \qquad (7)$$

In the definition of R_g in reported in equation 6, N is the number of atoms in the molecule, r is the position of a given atom and r_{mean} is the center of scattering mass. For globular systems, the Guinier approximation is valid up to sR_g <1.3, and can be applied from s values that are greater than $1/D_{max}$, the maximum dimension of the particle (Feigin & Svergun, 1987). Furthermore, plotting the natural logarithm of the intensity versus S^2 can indicate whether the system behaves ideally- that is if the solution is isotropic and the solute is not affected by aggregation or inter-particle effects. In case such effects are present, a dilution series may be used to quantify the second virial coefficient A_2 and treat the initial region of the curve such that it represents the scattering at infinite dilution (see Figure 1.5.1).

In addition to R_g and I(0), the scattering curve is the fourier transform of the pair-distance distribution function (PDDF), also known as P(r). This function reports the distribution of all pairwise interatomic distances in the scattering particle.

$$P(r) = \frac{r^2}{2\pi^2}\int_{s=0}^{s=\infty} s^2 I(s)\frac{sin(sr)}{sr}ds \qquad (8)$$

However, deriving the P(r) from the scattering curve would require the integration of a continuous scattering function that extends from s=0 to s=∞, something which in practice is not available, due to the discrete nature of digitised intensities on a detector, the presence of the beam stop at s=0 and signal-to-noise limitations at higher s ranges. In order to resolve this ill-posed problem, the curve is fit to a smooth curve made up by spline functions and then integrated. In the common program GNOM (Svergun, 1992), the choice of the smoothed curve depends on a set of perceptual criteria, such as curve smoothness and absence of systematic deviations, as well as fit to the experimental data, quantified by χ^2 (see equation 15).

Finally, in high s-ranges the scattering curve follows Porod's law - that is the intensity decays in a way that is proportional to s^{-4} and to the parameter K, dependent on the smoothness o f the surface of the particle. Thus, this region of the curve which may be used to derive the volume of the particle in a concentration-independent manner. By assuming a spherical particle, the volume may also be translated into an apparent molecular weight using the stokes-Einstein equation.

$$I(s) \sim Ks^{-4} \qquad (9)$$

Although SAXS has the advantage of having higher signal-to-noise and throughput, SANS is especially powerful for biomolecules, as careful manipulation of solvent and solute deuteration patterns can be used to "highlight" or "mask" components in a system of interest by matching SLDs (Gabel, 2015). The SLDs of ^{1}H and ^{2}H differ significantly, being -0.37×10^{-12} cm^{-2} and 0.67×10^{-12} cm^{-2}, respectively. Therefore, in SANS experiments individual subunits of protein, lipids or nucleic acids can be highlighted or masked within a multi-subunit complex by varying the deuteration of the sample and the solvent (see Fig.1.5.1). Furthermore, the absence of radiation damage in SANS allows for the design of titration experiments, as the same sample can be measured again and again with increasing ligand concentrations. Since the 1970s, these physical properties of SANS have been exploited to obtain low-resolution information on a wide range of biological complexes, such as the nucleosome and the ribosome (Capel *et al*, 1987). More recently, SANS has been successfully combined with other techniques such as NMR in order to obtain information on complexes and large assemblies at a high resolution (Lapinaite *et al*, 2013; Falb *et al*, 2010; Madl *et al*, 2013).

One of the fundamental properties of all SAS data however, is degeneracy: as the scattering curve reflects the distance distribution within the particle, particles of different overall shapes can have identical scattering curves. However, only a single theoretical curve corresponds to a given particle.

Thanks to improvements in both data collection and analysis, SAS data can be used to model multicomponent systems *ab initio* or to score models or ensembles generated *in silico* using data coming from other techniques. Because of the degeneracy problem, it has been shown that using scattering data in scoring, rather than sampling, provides more accurate and discriminative information. Programs such as CRYSOL (for SAXS data) (Svergun *et al*, 1995) and CRYSON (for SANS data) (Svergun *et al*, 1998) use spherical harmonics to model the solvent shell around a macromolecule and report the fit of the theoretical curve generated from models to the experimental data. As the treatment of the solvent is especially important in smaller systems, where the contribution of the solvent shell to the overall scattering pattern is more significant, efforts have been made in deriving more accurate solvent models for the prediction of theoretical curves.

As SAS provides low-resolution, global (typically nm-range) information, the full characterisation of a system of interest has to rely on high-resolution local information coming from other techniques such as NMR. Indeed, SAS has been used to improve traditional NMR structure calculations, which rely on short distance restraints to drive a simulated annealing that optimises local contacts, by scoring or based on the global shape of the system (Gabel *et al*, 2008, 2006; Mertens & Svergun, 2017).

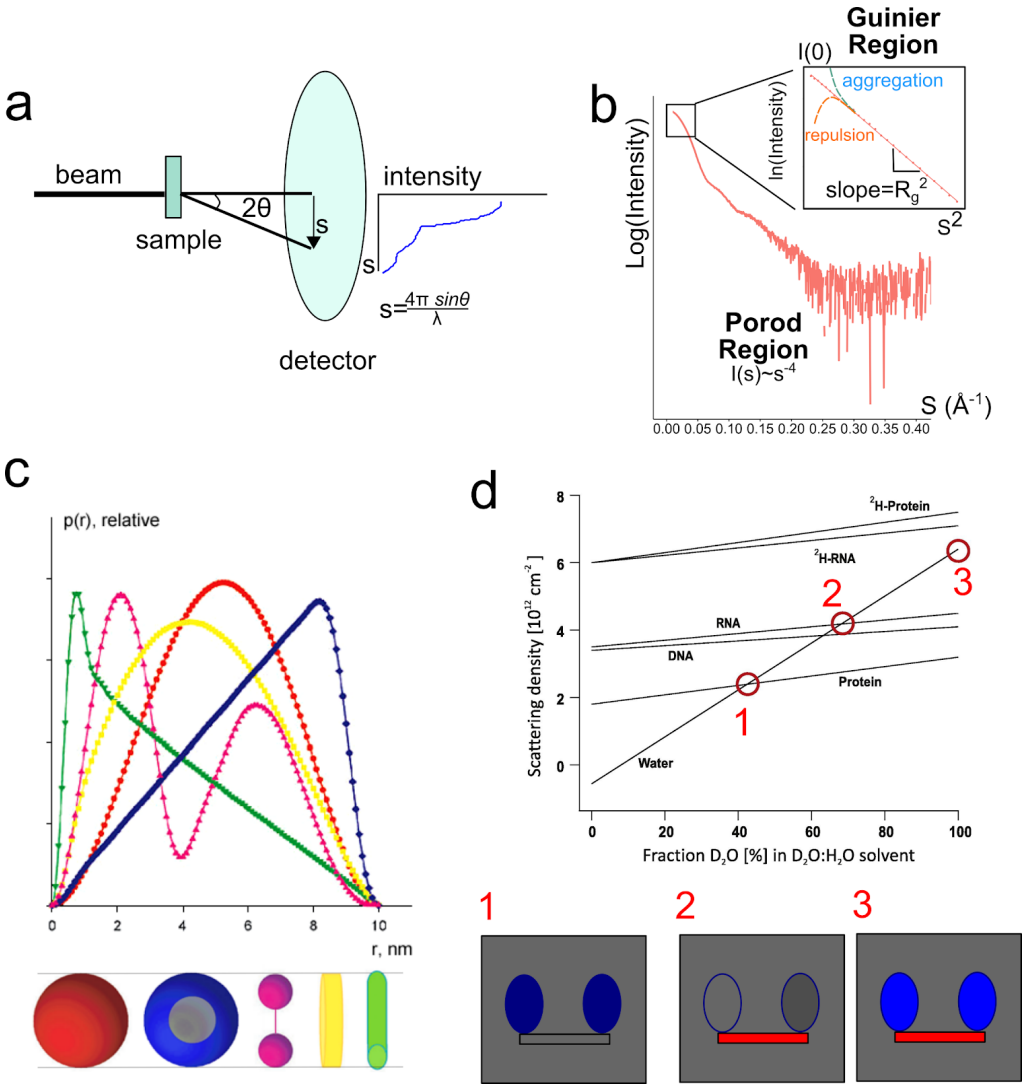


Figure 1.5.1. Small-angle scattering. A) The typical set-up of a SAS experiment. The outcome of the experiment is a radially-average set of intensities along the scattering vector s. B) The scattering curve can be divided into the Guinier region (up to sRg=1.3 for globular particles) and the Porod region. In the Guinier region, it is possible to ascertain the interparticle behaviour of the solute and extract the structural parameters R_g and I(0) from the Guinier plot (inset). The Porod region, at high s values, can show the excluded volume of the particle. C) The P(r) for particles of various shapes. Reproduced with permission from Svergun & Koch, 2003. D) In SANS, the SLD of the samples and solvent can be matched, thereby enabling the investigation of individual subunits within a complex. In this example, ^{1}H-proteins are depicted in blue and ^{1}H-RNA in red. Top reproduced with permission from Madl *et al.*, 2011.

1.6 Methodological background - Integrative structural biology.

Molecular machines and large biological assemblies mediate many of the key functions required by life. Although many of these large complexes have, over the years, been either purified or reconstituted biochemically from purified components, the structure, regulation and function of molecular machines remains one of the key questions in structural biology.

While in notable cases it has been possible to crystallise or solve by cryo-electron microscopy a whole, MDa-sized complex (Yusupov *et al*, 2014; Zhang *et al*, 2018), in other cases efforts to produce a high-resolution structural model have relied on integrating a wide range of structural and biochemical data into computational pipelines capable of sampling a vast conformational space and score models according to experimental data (von Appen *et al*, 2015; Dauden *et al*, 2017). This has been achieved by the fitting of high-resolution structures into lower resolution electron-microscopy maps, or in solution, by integrating information distance restraints from NMR or cross-linking mass spectrometry with low-resolution information derived by SAS.

The advantage of the solution approach, besides the fact that it probes the system in near-physiological conditions, is that it is capable of dealing with flexible or disordered systems, even when the nature of the motion is not known. Furthermore, NMR is capable of providing information on the kinetics of conformational exchange in a system, by monitoring the broadening or shifting of resonances.

Among the more successful platforms for integrative structural modelling are the integrative modelling platform (IMP) (Russel *et al*, 2012), the docking platform HADDOCK (Dominguez *et al*, 2003) and its extension for modelling of molecular machines, M3 (Karaca *et al*, 2017), and more specialised pipelines such as MDFF, dedicated to flexible fitting of structures inside electron microscopy maps (Trabuco *et al*, 2009), or Rosetta, which can integrate *ab initio* modelling with restraints coming from NMR and a number of other sources (Sgourakis *et al*, 2011).

Broadly speaking, all integrative structural biology approaches involve three steps: the generation of models, the scoring of models according to a combination of experimental data and physical potentials, and the selection of a final ensemble of models that fits all restraints given to the system.

In the docking-based approach of HADDOCK, the building blocks are kept in a full-atom representation and initially docked as rigid bodies using distance restraints in combinations with a physical and an empirical potential, before allowing for flexibility in a local search. With

M3, this approach is extended by then scoring the complexes using data that is orthogonal to the data driving the sampling, such as EM density or SAS data. Other implementations of this type of approach include the custom Aria-based protocol used here to solve the structure of the Box C/D complex, or XL-mod, used to have full flexibility in docking with cross-linking data (Ferber *et al*, 2016).

On the other hand, approaches like IMP have relied on placing the models into a diverse set of arrangements based on low-resolution information (usually EM maps), then scoring using various types of local information, stemming from NMR, proteomics or other sources. This approach has been particularly successful in deriving the arrangement of complexes where an electron microscopy map is available, or to model flexible regions interacting with a complex of known structure.

As both the data types and computational approaches used in integrative modelling can be so diverse, efforts have been made to establish a way to assess the precision and accuracy of the structures derived with such approaches (Schneidman-Duhovny *et al*, 2014). On a first level, the input data all have different levels of uncertainties, which need to be combined, and, in the case of EM maps, the resolution of the data is not uniform across the model. On top of this, judging model precision may be achieved by considering the convergence of the models that score best against the experimental data. However, gauging model accuracy is more challenging: while it is theoretically possible to set aside a small percentage of restraints to use in final scoring, in a way that is reminiscent to the crystallographic R_{free} parameter, the fact that restraints are sparse and do not contain equal amounts of information may mean that repeating calculations with different sets of restraints drives sampling to different solutions.

Desp

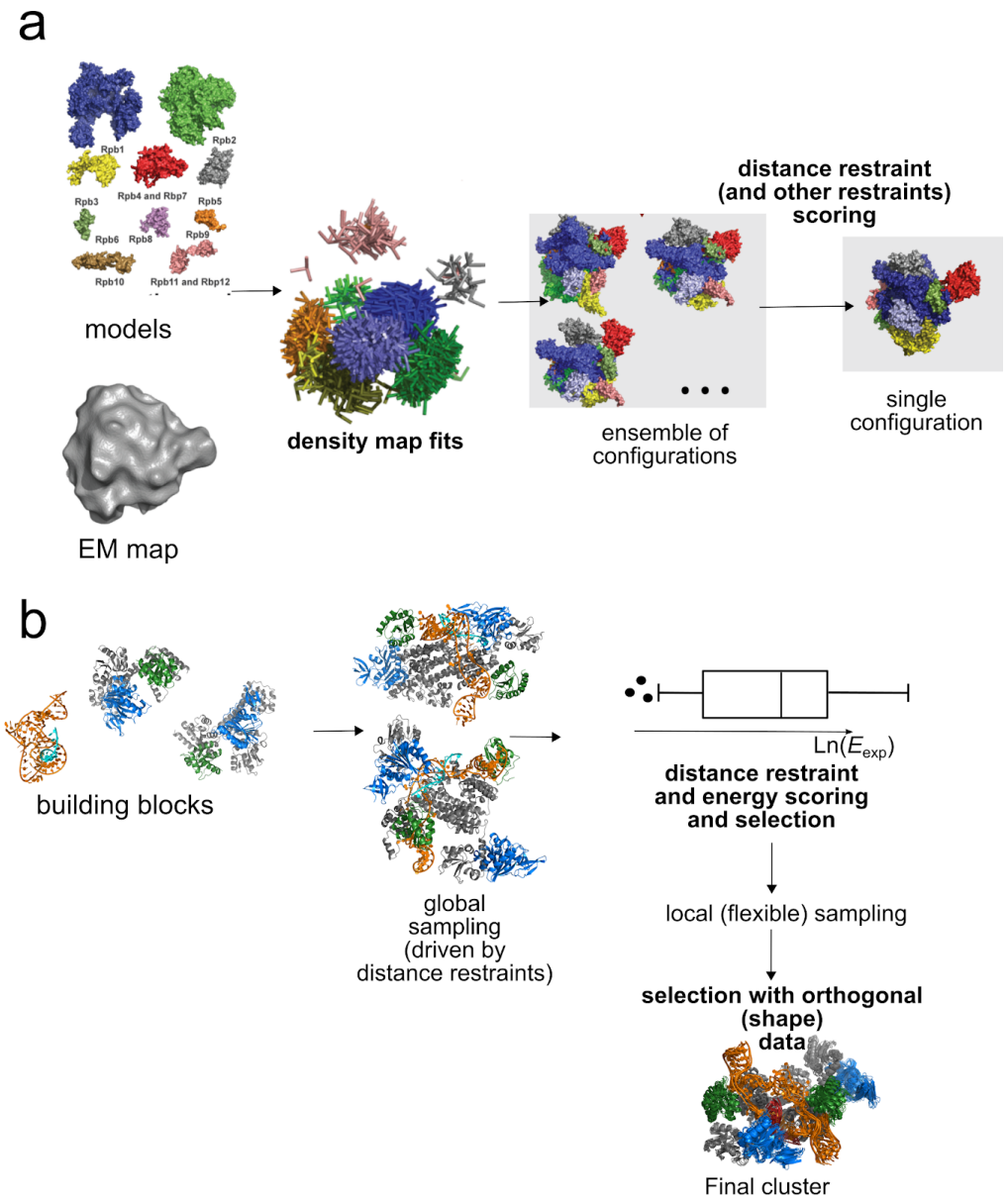


Figure 1.6.1. Integrative structural biology approaches. A) In IMP and other approaches used in combination with electron microscopy data, the system is assembled from components *ab initio* using various type of simulation approaches and, if possible, placed within an electron density map. The various arrangements are then scored against the data using experiment-specific potentials, which may include distance restraints, a physical potential, or several other types of information. IMP may in principle be modified by experts to adopt other workflows. Panel adapted from Russel *et al.* 2012. B) In HADDOCK-M3 and other approaches derived from docking and structure calculation, the components of the system are docked as rigid bodies or with various degrees of flexibility. The sampling in the docking phase is often driven by distance restraints, which may for example come from NMR experiments. The models are then scored with an orthogonal set of data, usually containing global information, either in the form of an electron density map or using SAS curves.

2. Goals of the thesis work.

Previous studies have shown that the archaeal Box C/D enzyme has a highly regulated activity, which may be correlated with the different assembly modes discovered for this kind of RNPs. For the *P. furiosus* Box C/D enzyme reconstituted with sR26 sRNA, for example, our laboratory has found that substrate D' can be ethylated and released independently of substrate D, while release of substrate D requires binding of substrate D' (REF). The extent of methylation and turnover at the two sites has been found to be dependent on the exact sequence of the sRNA at the guide sites. However, it is not clear how the complex achieves such a regulated catalytic activity and whether this depends on the transition pathways between the apo and holo forms.

In order to achieve an understanding of the structure and dynamics involved in the methylation cycle, and how these may influence methylation efficiency, we set out to

- Reconstitute the archaeal Box C/D complex with RNAs that yields stable Box C/D particles capable of binding physiological substrate D and D' sequences, and determine its oligomeric state.
- Understand the basis for the regulation of methylation by solving the structure of the complex bound to either substrate D or substrate D' using structural biology in solution based on NMR and SAS, complemented by performing biochemical experiments.
- Understand the dynamic behaviour of the enzyme in solution, and in particular of the Fib methyltransferase, at different stages of the catalytic cycle.
- Understand the determinants of the regulation of the catalytic activity of the enzyme by performing biochemical experiments on Box C/D sRNPs reconstituted using different sRNAs and protein mutants.

In order to achieve these goals I had to develop some methodology, especially designed for our questions. This included:

- Developing integrative modelling methods capable of using global information derived from SAXS/SANS to address conformational flexibility and structural heterogeneity.
- Deriving a more rigorous assignment of NMR resonances and relaxation rates in the full Box C/D sRNP, which improves on the derivation of distance restraints.

3. Methods

3.1 Protein purification.

The text of the following section was adapted from a joint manuscript and is originally written by myself:

The constructs used in this study, corresponding to *P. furiosus* L7Ae (UniProtKb Q8U160), fibrillarin (Fib, Q8U4M2) and Nop5 (Q8U4M1) in pETM-11 vector (EMBL collection) were available in the lab at the time of my study. The cloning procedure has been reported in (Lapinaite *et al*, 2013). The Nop5 construct also contains two characterised mutations, L148K and V223E, which are beneficial in preventing aggregation of the purified protein.

Cells harbouring the expression vector were grown at 37 °C until OD_{600} ~0.7, when they were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein expression was carried out for 18-22 hours at 22 °C.

After spin-down, cells were resuspended in lysis buffer (50 mM Tris, 10 mM imidazole, 1M NaCl, 5 mM β-mercaptoethanol, pH 7.5). EDTA-free protease inhibitors (Roche), DNAseI, and 0.25 mg/ml lysozyme were added, and lysis was allowed to proceed for 30 min at room temperature, prior to sonication. The lysate was then spun down for 30 min at 30000g at 16 °C. For Nop5 and Fib, *E. coli* proteins were removed from the mixture by incubation for 15 min at 80 °C, followed by 30 min centrifugation at 30000g. In the case of L7Ae, the supernatant was not incubated at high temperature.

Proteins were then purified by nickel affinity with a HisTrap FF column (GE healthcare). The protein bound to the column was washed extensively with alternating gradients of lysis buffer and wash buffer (50mM Tris, 10 mM imidazole, 1M NaCl, 2M LiCl, 5 mM β-mercaptoethanol pH 7.5). Proteins were then eluted from the column using a gradient of elution buffer (50 mM Tris, 300 mM imidazole, 1M NaCl, 5 mM β-mercaptoethanol pH 7.5). For Nop5 and Fib, the proteins were exchanged into lysis buffer with a HiPrep 26/10 Desalting column (GE Healthcare). As L7Ae is an RNA-binding protein, a further purification by ion exchange is required in order to remove any bound bacterial RNA. Indeed, at this stage of the purification eluted L7Ae displays high absorbance at 260 nm (A_{260}/A_{280} ratio of ~2), indicative of the presence of bound nucleotides, which strongly absorb at this wavelength.

Thus, L7Ae eluted from the affinity column was exchanged into a low salt buffer (25 mM Tris, 150 mM NaCl, 2 mM DTT, pH 7.5) with a HiPrep 26/10 Desalting column, and then loaded to a HiTrap Q FF strong anion exchange column (GE Healthcare), where the RNA-free L7Ae protein flows through and the nucleic acid is eluted from the column using lysis buffer.

At this step, Nop5, Fib and L7Ae preparations are each tested for the presence of RNAses using an RNAse Alert Kit (Ambion). RNAse-contaminated samples were re-purified by affinity purification. In order to remove the His-tag, 1mg TEV protease/50mg protein was added to RNAse-free samples, and the reaction was allowed to proceed overnight at room temperature. The protease and the His-tag were then separated from the samples by affinity purification on a HisTrap column using lysis buffer and elution buffer. Protein concentrations were then measured by absorbance at 280 nm (Molar extinction coefficient ε for L7Ae: 4470 M^{-1} cm^{-1}; ε Nop5: 53860 $M^{-1}cm^{-1}$;ε Fibrillarin = 31400 M^{-1} cm^{-1}). Nucleic acid contamination was checked by ensuring that the purified proteins maintained an A_{260}/A_{280} ratio of ~0.55. Proteins were stored at 4 °C.

3.2 Reconstitution of Box C/D sRNPs.

Reconstitution of the archaeal Box C/D sRNP proceeds via reconstitution of the L7Ae/sRNA and Nop5-Fib subcomplexes, which are then mixed together to produce a full sRNP.

The sRNA/L7Ae subcomplex was prepared by mixing L7Ae (previously stored in 50 mM Tris, 150 mM NaCl, 2 mM DTT pH 7.5) with the desired sRNA in a molar ratio of 2.2:1 L7Ae:sRNA. The mixture was then heated to 80 °C, allowed to cool to room temperature and, after concentration with a 10 KDa-cutoff centrifugal filter (Merck), purified by size exclusion chromatography in sample buffer (50 mM NaPi, 500 mM NaCl pH 6.6).
The Nop5-Fib subcomplex was obtained by mixing the two proteins, stored in lysis buffer at a molar ratio of 1:1.1 Nop5-Fib, heated to 80 °C, allowed to cool to room temperature, concentrated and purified by size exclusion chromatography in sample buffer. As the elution volume of this subcomplex is dependent on concentration, final elution volume was checked to be consistent with a Nop5-Fib dimer of dimers.

The full Box C/D complex was reconstituted by mixing the two previously assembled subcomplexes at a L7Ae/sRNA:Nop5-Fib ratio of 1:2.2. After incubation for 15 min at 80 °C, the sample was allowed to cool down to room temperature and concentrated. The sample

was then purified by size exclusion chromatography in sample buffer. For RNPs where higher order aggregates were observed, only the main peak, usually corresponding to a mass of around 400 KDa, was pooled.

In the case of samples for NMR or SANS, which require buffer containing D_2O, the final reconstitution step was carried out directly in the D_2O-containing buffer. Complexes were then concentrated to the desired volume using a 30-KDa cutoff centrifugal filter.
The sRNA concentration (and hence the concentration of the Box C/D sRNP) was estimated using equation (Porterfield *et al*, 2010).

$$[RNA] = \frac{A_{260} - (\frac{\varepsilon_{protein260}}{\varepsilon_{protein280}}) A_{280}}{\varepsilon_{RNA260} - (\frac{\varepsilon_{protein260}}{\varepsilon_{protein280}}) \varepsilon_{RNA280} L} \qquad (10)$$

Where A_{260} is the absorbance at 260 nm (nucleic acids maximum); A_{280} is the absorbance at 280 nm (tryptophan maximum); $\varepsilon_{protein260}/\varepsilon_{protein280}$ is the ratio of the molar extinction coefficients for the protein at 260 and 280 nm, i.e. 0.6; ε_{RNA260} is the molar extinction coefficient of the RNA at 260 nm (8000 $M^{-1}cm^{-1}$ per nucleotide) and ε_{RNA280} is the molar extinction coefficient of RNA at 280 nm (4000 $M^{-1}cm^{-1}$ per nucleotide); L is the cuvette path length.

For complexes in 20mM HEPES 150mM NaCl pH 7.5, the gel filtration peak was re-purified by a further gel filtration column in the buffer of choice.

Substrate RNA-bound complexes were generated by addition of the specified amount of substrate D or substrate D'. ε substrate D: 181300 $M^{-1}cm^{-2}$; ε substrate D':177800 $M^{-1}cm^{-2}$ [1].

[1] Molar extinction coefficients for the substrate RNAs are obtained with the Scripps oligo extinction coefficient calculator available at
https://www.scripps.edu/researchservices/old/corefac/biopolymercalc2.html

as

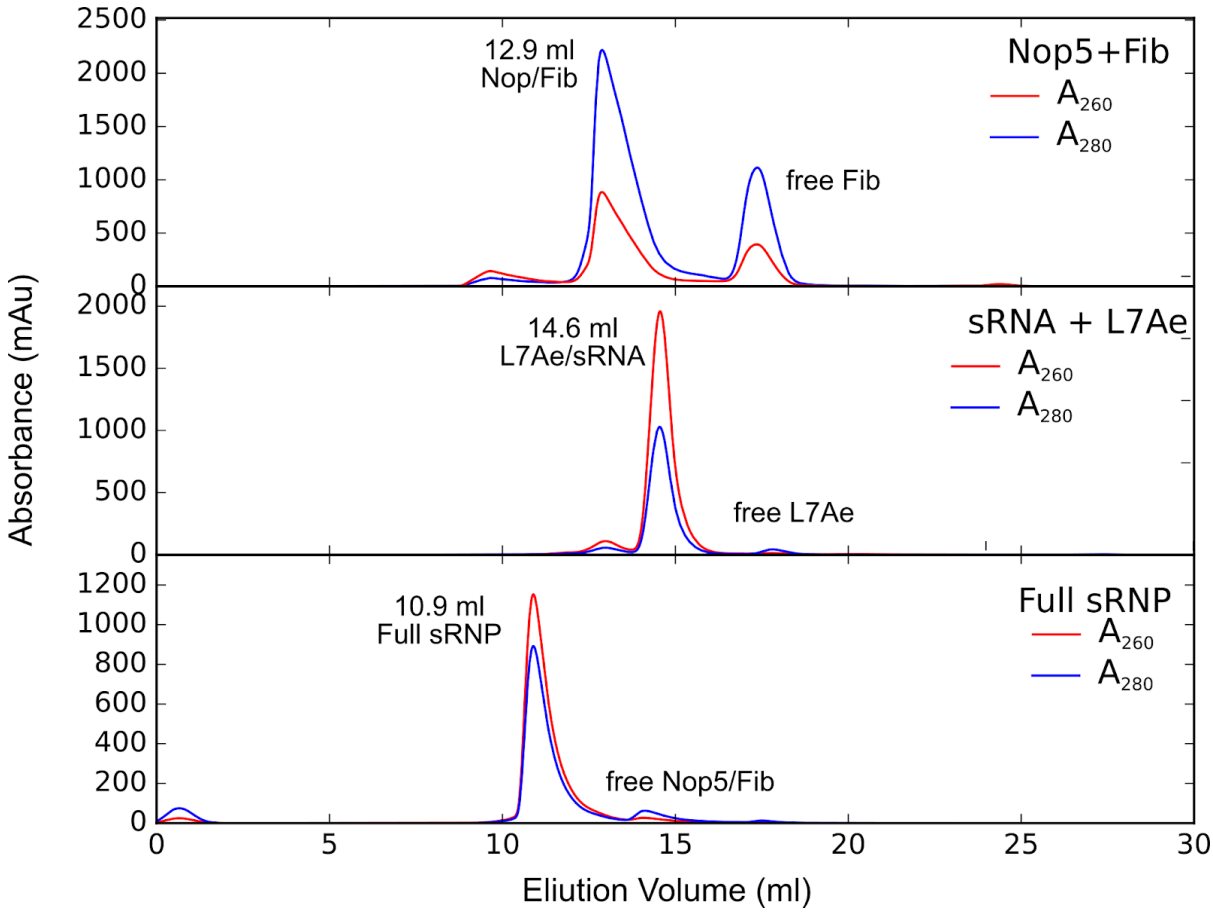


Figure 3.2.1. Reconstitution of the Box C/D sRNP. Top: Size exclusion chromatography of the Nop5-Fib subcomplex reconstitution step. Middle: Size exclusion chromatography of the L7Ae/sRNA step. In this case, the chromatogram displays the reconstitution of L7Ae and st-sR26 RNA. Bottom: After mixing the L7Ae/RNA and Nop5-Fib subcomplexes in a 1:2.2 ratio, the full sRNP is purified by size exclusion chromatography. A_{260}: absorbance at 260 nm wavelength (maximum for nucleic acids). A_{280}: absorbance at 280 nm wavelength (maximum for proteins containing tryptophan).

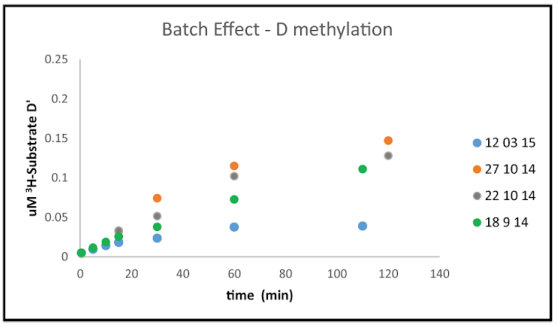
3.3 Activity assays

Activity of the Box C/D sRNP was measured by quantification of transfer of a ^{3}H-labelled methyl group from s-adenosyl methionine (SAM) to RNAs labelled with a 5' biotin group, which is then purified by affinity purification with streptavidin beads.

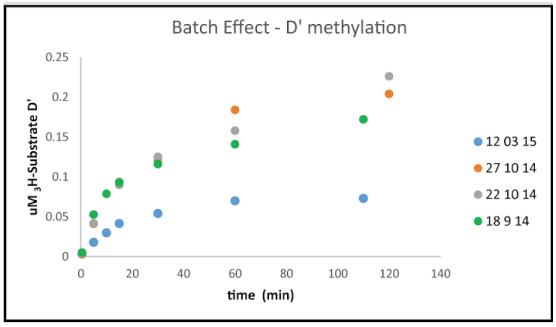
It is worth pointing out that the SAM cofactor has been reported to be extremely unstable in aqueous solutions at neutral pH, with increasing destabilisation as temperature is increased (Matos & Wong, 1987). Unhelpfully, one of the degradation products of SAM is s-adenosyl homocysteine (SAH), which is also the byproduct of the methylation reaction catalysed by Fib. Thus, results of methylation assays display strong batch effects, as Fib. 3.2.1 shows.

Therefore, results of such assays should be interpreted in terms of relative methylation efficiency of particles measured with the same ^{3}H-SAM batch, rather than absolute methylation efficiency.

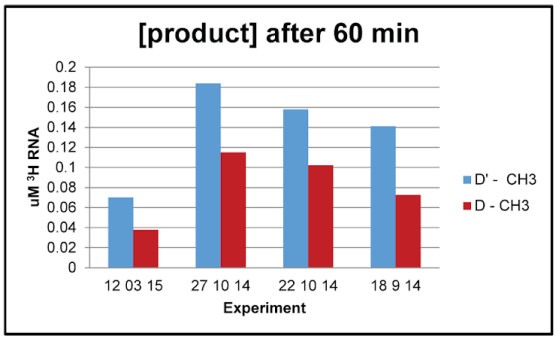
A



B



C



D

	Turnover Number (min-1)	
Batch	Substrate D	Substrate D'
12 03 15	0.012	0.035
27 10 14	0.032	0.081*
22 10 14	0.027	0.076*
18 9 14	0.019	0.103*

Figure 3.2.1: Effect of batch of [^{3}H]S-adenosyl methionine (SAM) and experimental reproducibility. A. Methylation of substrate D of Box C/D reconstituted using sR26 guide RNA. All experiments performed using different batches of ^{3}H-SAM, except for 22/10/14 and 27/10/14, which share the same batch, for the purpose of internal comparison. B. Same as in (A), but for substrate D'. C. Concentration of methylated product after 60 minutes of reaction for the experiments depicted in (A) and (B), indicating the reproducibility of the ~2-fold higher substrate D' methylation. D. Apparent turnover number estimated from the linear region of the methylation curve (0-15 minutes) for the experiments in (A) and (B). Values marked with an asterisk indicate a lower-boundary estimate extracted from only two points. Data produced jointly with Dr. Pawel Masiewicz.

Reconstituted Box C/D RNPs (in 50 mM NaPi pH 6.6, 500 mM NaCl) were mixed with 5'-end biotinylated substrate D or D' (Integrated DNA Technologies) in 1:3-5 molar ratio guide:substrate RNA (0.15:0.5-75 µM, depending on experimental design). For samples where addition of the second substrate was required, equal amounts of unmodified substrate D or D' was added to the mixture. Reactions were initiated by addition of methyl-^{3}H-SAM (6-15-fold excess over guide RNA, 15 Ci/mmol; Hartmann Analytic) onto samples

pre-incubated at 50 °C and allowed to proceed at the same temperature. 10 µL aliquots were taken at specific time points and mixed with 15 µL of stop buffer (0.1 mg/mL proteinase K [Thermo Scientific], 20 mM Tris pH 7.8, 20 mM NaCl, 0.1% SDS, 20 mM EDTA). Each aliquot was incubated for an additional hour at 50°C. Afterwards 5 µL were taken from each aliquot and added to NeutrAvidin Agarose beads (25 µL bead volume; Thermo Scientific) previously equilibrated in low salt buffer (20 mM HEPES pH 7.5, 150 mM NaCl). After addition to the beads the samples were incubated at 25°C for 1h under constant shaking. Next, the beads were washed 3-4 times with low salt buffer, transferred into scintillation vials and mixed with scintillation liquid (Aquasafe 500 plus; Zinsser Analytic). Incorporation of ^{3}H-SAM into the biotinylated substrate RNA was determined by scintillation counting (Wallac 1410 Liquid Scintillation Counter, Pharmacia or Tri-Carb Liquid Scintillation Counter; Perkin Elmer, depending on experimental set-up). Two separate aliquots were taken at a given time point and experiments were carried out in duplicates.

3.4 Small-angle X-Ray scattering

All SAXS data was acquired at BM29 in ESRF in Grenoble, France (Pernot *et al*, 2013). Box C/D sRNPs were reconstituted with the relevant buffer. Prior to acquisition, 2mM dithiothreitol (DTT) was added to the sample and to the buffer blank, as a protection against radiation damage. The measurements were carried out at 40 °C, unless specified otherwise. The samples were exposed for 10 frames of 1 second each.

The curves were compared, merged and buffer subtracted by the beamline software BsxCube. Scattering curves were displayed and analyzed using the ATSAS package (Petoukhov et al. 2012). For curves displaying severe radiation damage, the frames were compared using the CORMAP algorithm (Franke *et al*, 2015). R_g values were extracted using the Guinier approximation for globular particles, as described in section 3.3. P(r) functions were extracted using the GNOM program (Svergun, 1992) in its automatic version (datgnom) whenever possible. If the resulting fit between the smoothed curve and the experimental data was poor, the P(r) was searched by testing a range of D_{max} values (50-200 Å) and checking the quality of the results to pick the best-fitting P(r) curve.

For *ab initio* modelling of low-resolution envelopes, the program DAMMIN (Svergun, 1999) and DAMMIF (Franke & Svergun, 2009) were used in the “slow setting”. In each case, 20 models were generated, superimposed and averaged with DAMAVER, and used as a search volume for a new round of DAMMIN modelling, which yielded final results.

3.5 Small-angle neutron scattering

^{2}H-L7Ae, ^{2}H-Nop5, ^{2}H-fibrillarin, ^{2}H-RNA, ^{2}H-fibrillarin/^{2}H-RNA and ^{2}H(70%)-Nop5/^{2}H-RNA samples were measured in 50mM NaPi pH 6.6, 500mM NaCl, 42%/58% D_2O/H_2O solutions, in order to mask the contribution of the ^{1}H-proteins.

^{2}H-L7Ae, ^{2}H-Nop5, ^{2}H-RNA and ^{2}H(70%)-Nop5/^{2}H-RNA were acquired at D22 at the Institut Laue Langevin in Grenoble, France , with a neutron wavelength of 6 Å . The ^{2}H-fibrillarin and ^{2}H-fibrillarin/$^{2-}$H-RNA curves were acquired at KWS-1 at JCNS in Munich, Germany (Feoktystov *et al*, 2015) with a neutron wavelength of 5 Å. Both instruments were set up with sample-detector distances of 4 m and collimation lengths of 4 m.

Data reduction and radial integration were done with standard procedures using beamline-specific software. Buffer subtraction was done in PRIMUS (Svergun *et al*, 2013). Pair-wise distance-distribution functions P(r) were calculated from experimental data using GNOM (Svergun, 1992), testing a range of maximum dimension parameters in order to obtain a satisfactory curve. All SANS curves were acquired at 55 °C. P(r) functions were extracted using the GNOM program (Svergun, 1992) in its automatic version (datgnom) whenever possible. If the resulting fit between the smoothed curve and the experimental data was poor, the P(r) was searched by testing a range of D_{max} values (50-200 Å) and checking the quality of the results to pick the best-fitting P(r) curve.

Ab initio modelling of SANS envelopes was done with DAMMIN (Svergun, 1999) and DAMMIF (Franke & Svergun, 2009). For curves corresponding to ^{2}H-Fib, ^{2}H-Fib/^{2}H-RNA and ^{2}H-L7Ae, modelling was performed both with and without the assumption of a continuous solute density, as the deuterated subunits may not be in contact in the full complex. In each case, 20 models were generated, superimposed and averaged using DAMAVER (Volkov & Svergun, 2003), and used search volume for a final round of DAMMIN.

3.6 NMR assignment of methyl groups

Stereospecific assignment of Fib methyl groups in the full Box C/D RNP was achieved by a strategy combining novel isotope labelling approaches. $^{1}H,^{13}C$ labelling of methyl groups of isoleucine, leucine and valines in a perdeuterated protein was achieved following methods developed in the Kay laboratory (Tugarinov & Kay, 2003). Briefly, BL21(DE3) cells harboring the Fib-pETM-11 plasmid were adapted to growth in D_2O minimal media with ^{2}H-glycerol as the carbon source by successive rounds of growths in increasing D_2O/H_2O ratios, up to 99%. Selective labelling of the required amino acid type was achieved by addition of corresponding precursors to the final growth media. Protein expression and purification then proceeded as for unlabelled protein.

For ILV-methyl labelled samples of Fib or L7Ae ([U-99%-^{2}H, 99%-$^{1}H\delta,^{13}C\delta$-IL; 99%-$^{1}H\gamma,^{13}C\gamma$-V,]-Fib or L7Ae), 120 mg/L culture of (3-methyl-^{13}C, 99%; 3,4,4,4-D4, 98%) α-ketoisovaleric acid (Cambridge Isotope Labs) and 60 mg/L culture of (methyl-^{13}C, 99%; 3,3-D2, 98%) α-ketobutyric (Cambridge Isotope Labs) acid were added 40 minutes prior to induction at OD_{600} 0.8. For the production of Leucine-methyl labelled samples, 120 mg/L culture of α-ketoisocaproate precursor was used, obtained from Dr. Lichtenecker at the University of Vienna (Lichtenecker *et al*, 2013) . Stereospecific labelling of LV methyl proS group was achieved by using the TLAM-ILV^{proS} labelling kit from NMRBio. In the full sRNP samples, or in measurements of the Nop5-Fib subcomplex, Nop5 was perdeuterated ([U-99%-^{2}H]-Nop5).

Stereospecific assignments in the free state were obtained by a combination of ^{1}H-^{13}C-HMQC, ^{1}H-^{13}C NOESY-HMQC experiments on Fib samples with ILV-methyl, LV^{proS}-methyl or L-methyl labelling. This allowed us to improve on previous assignments obtained with ^{1}H-^{13}C-TOCSY experiments on ILV-methyl labelled Fib. Resonances belonging to pairs of diastereotopic methyl groups of each L/V residue were connected by the NOESY spectra. Residue typing was done by analysing HMQC spectra of L-methyl labelled Fib. The stereospecificity was assigned using the HMQC spectra of the Fib wherein only the proS groups were labeled.

The assignments were then transferred to the Nop5-NTD−Fib subcomplex by means of ^{1}H-^{13}C NOESY-HMQC and ^{1}H-^{13}C HMQC.

Assignments were transferred from the Nop5-NTD−Fib subcomplex to the Nop5-Fib subcomplex by acquiring ^{1}H-^{13}C HMQC spectra on ILV-methyl, LV^{proS}-methyl and L-methyl labelled Fib samples and comparing resonances.

Finally, assignments were transferred to the full Box C/D sRNP by acquiring ^{1}H-^{13}C HMQC spectra on ILV-methyl, LV^{proS}-methyl and L-methyl labelled Fib samples and comparing resonances. Residue typing in the Box C/D sRNP was done by analysing HMQC spectra of L-methyl labelled Fib. The stereospecificity was assigned using the HMQC spectra of the Fib wherein only the proS groups were labeled.

The resulting assignments in free Fib, Nop5-NTD−Fib, Nop5-Fib and the full Box C/D sRNP are displayed in section 4.5.

Chemical shift perturbation of methyl resonances between the different spectra is extracted by:

$$CSP = \sqrt{\delta_H{}^2 + (0.3\ \delta_C)^2} \quad (11)$$

Where δ_H is the distance between the ^{1}H chemical shifts in the two spectra, δ_C is the distance between the ^{13}C chemical shifts in the two spectra and 0.3 is the scaling factor that has been determined empirically for comparing carbon and proton chemical shift perturbation (Williamson, 2013).

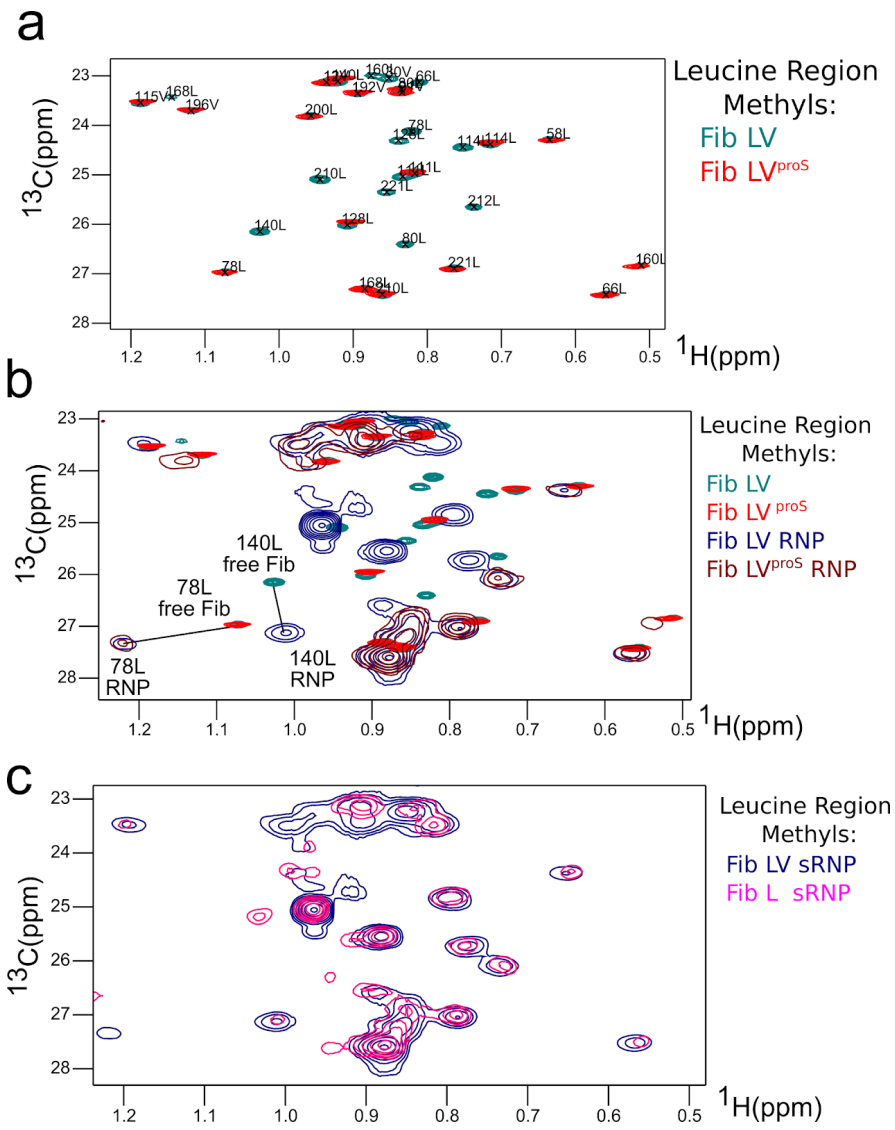


Figure 3.6.1: Strategy for assignment transfer from free Fib to the full sRNP. A. Overlay of ^{1}H-^{13}C HMQC spectra of ILV-methyl (light blue) and ILV^{proS}-methyl (red) labelled Fib. ILV^{proS}-methyl labelling of free Fib enables stereospecific assignments of methyl groups and reduces spectral crowding in 2D experiments. B) Overlay of ^{1}H-^{13}C HMQC spectra of ILV-methyl (light blue) and ILV^{proS}-methyl (red) labelled free Fib onto ILV-methyl (dark blue) and ILV^{proS}-methyl (brown) labelled Fib in the full st-sR26 Box C/D RNP. As highlighted by the example of leucines 78 and 140, LV^{proS} allows us to more easily transfer assignments between the free state and the full sRNP. In this case, we can see the proS resonance of leucine 78 shifting downfield to 1.2 ^{1}H ppm, which is a peak in the full RNP with the corresponding proS resonance. As no proS resonance is observed for the neighbouring peak at 1 ^{1}H ppm, we can deduce this peak comes from leucine 140. This assignment has been verified by acquisition of ^{1}H-^{13}C HMQC spectra on a Fib L140A mutant. C) Overlay of ^{1}H-^{13}C HMQC spectra of ILV methyl-labelled Fib and L-methyl labelled Fib, with both proteins being in the full st-sR26 Box CD RNP. In this case, we can identify which resonances arise from leucines, and which from valines in a group of overlapped peaks.

3.7 Paramagnetic relaxation enhancement NMR experiments and derivation of distance restraints.

We acquired PRE datasets probing the distances in 7 different paramagnetic tag/protein pairs (L7Ae-Q45C Fib; L7Ae-E58C Fib; L7Ae-C68 Fib; Nop5-E65C Fib Nop5-E196C Fib; Nop5-D247C Fib;Nop5-S343C Fib; see section 4.6) in the st-sR26 RNP bound to either substrate D or substrate D'.

Cysteine residues were introduced by site-directed mutagenesis following the QuikChange XL protocol (Agilent Technologies) and purified in the presence of 5 mM β-mercaptoethanol in order to prevent disulfide bond formation. For L7Ae, the native C68 was mutated to serine prior to introduction of cysteine residues at other sites. The purified protein was then buffer exchanged into 50 mM NaPi, 500 mM NaCl, pH 6.6 using a HiPrep 26/10 desalting column (GE Healthcare) and eluted directly into tubes containing a 10-fold molar excess of the 3-(2-iodoacetoamido)-PROXYL radical (Sigma-Aldrich) in the dark. The spin-labelling reaction (Fig.1.4.1) was allowed to proceed overnight at room temperature.
Spin-labelled proteins were then used for complex reconstitution and the free spin-label was removed during the gel-filtration step. The final reconstitution step was carried out in 99% D_2O buffer (50 mM NaPi, 500 mM NaCl, pH 6.6), prior to concentration with a 30 kDa-cutoff Amicon centrifugal concentrator (Merck Millipore).

^{1}H-^{13}C HMQC spectra were acquired for both the paramagnetic and diamagnetic state of the spin-labelled protein on Bruker Avance 800 and 850 MHz spectrometers, equipped with TCI cryoprobes, at 328 K. The diamagnetic spectrum was recorded after reduction of the spin-label by addition of ascorbic acid to a final concentration of 5 mM.

The peak heights in the paramagnetic and diamagnetic states were used to calculate the distance between the nitroxide group of the paramagnetic tag and the respective methyl group.

All spectra were processed using apodization with an exponential function in order to preserve Lorentzian line shapes.

In order to derive the peak intensity in the oxidised, paramagnetic state (I_{para}) and the reduced, diamagnetic state (I_{dia}), peaks of the diamagnetic and paramagnetic spectra were fitted with the program FUDA (http://www.ucl.ac.uk/hansen-lab/fuda/) assuming Lorentzian line-shapes. Where necessary, overlapped peaks were fitted as groups. The fitted volumes and line-widths were then converted into peak-heights using:

$$I = \frac{V}{Lw_H Lw_C} \quad (12)$$

Where I is the peak height, V is the peak volume and Lw_H and Lw_C represent the fitted line widths in the ^{1}H and ^{13}C dimensions, respectively. For each PRE dataset, this generated a list of I_{para}/I_{dia} ratios. Experimental errors on I_{para}/I_{dia} were derived by error propagation of the standard deviation of the noise in the two spectra.

As discussed in section 1.4, quantification of the PRE rate Γ_2 requires quantification of the overall diamagnetic relaxation rate, R_2^{dia}. For the ^{1}H-^{13}C HMQC experiment, this requires quantification of the diamagnetic R_2 rates corresponding to the transverse relaxation rates of ^{1}H single-quantum coherence (R_2^{diaH}) and ^{1}H-^{13}C multiple-quantum coherence (R_2^{diaHC}) for each resonance in the spectrum.

R_2^{diaH} values were quantified using the pulse scheme reported in (Tugarinov & Kay, 2006). R_2^{diaHC} values were quantified using the pulse schemes reported in (Tugarinov & Kay, 2013), modified to remove the fast-relaxing-component purging-element. These experiments use delays to monitor the decrease in observed signal intensity during the J-coupling evolution time of the HMQC. Relaxation delays were chosen as 0, 2, 3, 4, 6, 7, 10 & 16 ms for fibrillarin, and 0, 2, 3, 4, 6, 7 & 10 ms for L7Ae. The peak-heights were fitted to a mono-exponential decay-function to extract R_2^{diaH} and R_2^{diaHC} (see Fig.3.7.1).

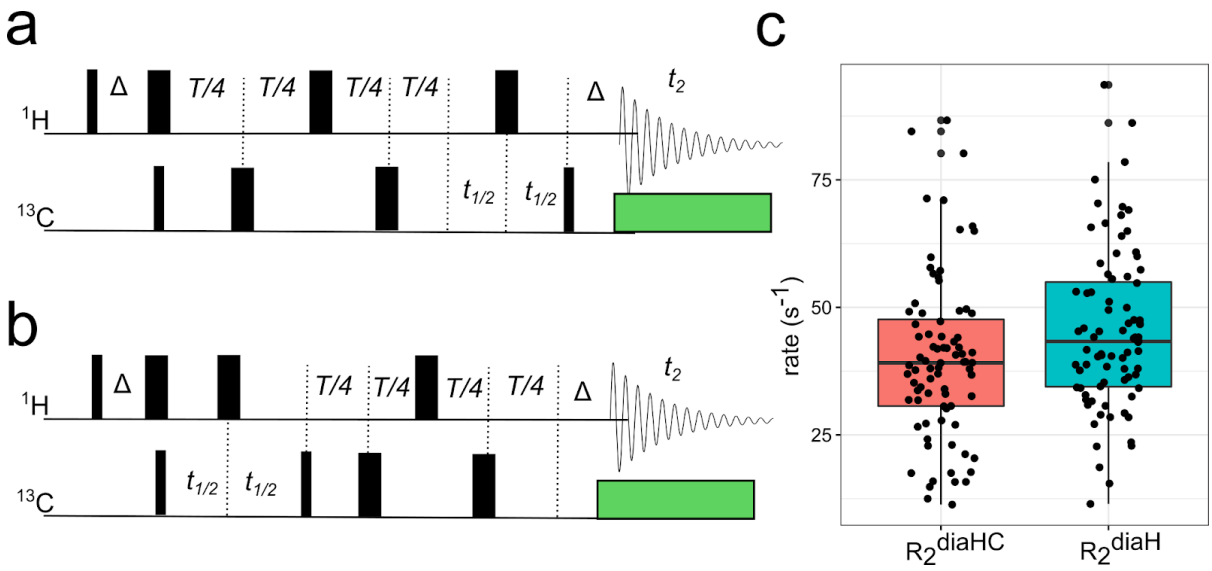


Figure 3.7.1: Extraction of diamagnetic relaxation rates. A) Simplified representation of the pulse sequence used to extract R_2^{diaHC}. Δ: ½J_{CH} ~3.8ms; *T*: variable relaxation delay (in our case, 0-16 ms); $t_{1/2}$: ½ of the ^{13}C chemical shift evolution time; t_2: ^{1}H chemical shift evolution time; thin bar: 90° hard pulse; thick bar:180° hard pulse. B) simplified representation of the pulse sequence used to extract R_2^{diaH}. C) Extracted R_2^{diaHC} and R_2^{diaH} for Fib ILV methyl resonances. R_2 rates were extracted by fitting a monoexponential decay function to the decrease in peak intensity as a function of *T.*

After estimation of resonance specific R_2^{diaH} and R_2^{diaHC} parameters, the ratios of peak-heights were converted into PREs, Γ_2 , using equation (13) and the R_2^{diaHC} and R_2^{diaH} rates measured for the respective peaks.

$$\frac{I_{para}}{I_{dia}} = \frac{exp(-\Gamma_2 t_{hmqc}) R_2^{diaHC} R_2^{diaH}}{(R_2^{diaHC}+\Gamma_2)(R_2^{diaH}+\Gamma_2)} \qquad (13)$$

This equation is an adaptation of equation (3) for the specific case of a ^{1}H-^{13}C HMQC experiment. Here, t_{HMQC} represents the magnetization transfer time in the HMQC sequence (7.6 ms). As this equation is non-invertible, Γ_2 was derived by plotting the simulated bleaching ratio, I_{para}/I_{dia} as a function of Γ_2 for a given set of diamagnetic rates, with the experimental errors on I_{para}/I_{dia}, R_2^{diaH} and R_2^{diaHC} used to determine the upper and lower bounds of the derived PRE.

Ultimately, translation of Γ_2 into distances requires the estimation of τ_C, the correlation time for the electron-nucleus interaction vector (see section 1.4). This value can be estimated using known distances and experimentally determined Γ_2 rates.

In order to derive τ_C, we quantified Γ_2 rates corresponding to known distances within fibrillarin in complexes reconstituted with the Fib-R109C mutant, using the procedure described

above. For L7Ae, we used known distances between Nop5-CTD and L7Ae in complexes reconstituted with the Nop5-S343C mutant.

These two calibration datasets were then used as restraints in the protocol developed by Clore *et al.* (Iwahara & Clore, 2004), which optimizes an ensemble of multiple spin-label conformations in combination with τ_C. For L7Ae, we used isoleucine resonances only. The minimization was run using the recommended "obsig" setting for the weighting of the different PREs. After minimization of 20 structures, τ_C was 51.8 ± 5.7 ns for fibrillarin and 50.4 ± 9.4 ns for L7Ae.

Once the τ_C value was obtained, we could translate Γ_2 rates for each PRE dataset into distances.

For a given value of τ_C, distances *r* between the unpaired electron and the methyl protons, were extracted from the equation:

$$r = \left(\frac{K}{\Gamma_2}\left(4\tau_c + \frac{3\tau_c}{1+\omega^2\tau_c{}^2}\right)\right)^{1/6} \quad (14)$$

where K is a constant (1.233×10^{-23} cm^6 s^{-2}) comprising the physical constants in equation (1) and ω is the proton Larmor frequency in rad/s. The errors on the distances were estimated by again using the errors in τ_C, experimental I_{para}/I_{dia} ratios and R_2 rates to yield upper and lower bounds on a calibration curve. Errors lower than 2 Å were set to 2 Å, in order to account for tag flexibility.

In the structure calculations, distances were imposed from the nitrogen atom of the nitroxide group of iodoacetoamido-PROXYL to the carbon atom of fib methyl groups. For L7Ae, where stereospecific assignment of LV methyl groups is not available, the distance restraint was imposed to both methyl group carbons.

3.8 Structure calculation

In this study, we tackled the structure of the st-sR26 RNP in its two half-loaded states, i.e. bound to either substrate D or substrate D'.

As discussed in section 4.4, it became apparent throughout the study that one of the Nop5-NTD−Fib modules in the substrate D'- or substrate D-bound st-sR26 RNP exchanges

between at least two families of conformations in solution. In one set of conformers, one Fib copy is assumed to be bound to the substrate-guide duplex ([on,off]-state). In the other, both Fib copies are positioned away from the RNA ([off,off]-state). The two conformer classes are depicted schematically in Fig. 4.4.1. Therefore, four separate states were calculated: substrate D'-bound st-sR26 RNP [on,off]-state; substrate D'-bound st-sR26 RNP [off,off]-state; substrate D-bound st-sR26 RNP [on,off]-state; substrate D-bound st-sR26 RNP [off,off]-state. Each of these states was calculated individually, after classification of distance restraints (see Fig. 3.8.3).

Structures were calculated using protocols that rely on the Aria1.2 framework that is built upon the crystallography and NMR system (CNS) (Brünger *et al*, 1998; Linge *et al*, 2003; Gabel *et al*, 2008). The protocol is a modified version of the one employed in Lapinaite *et al.* (*2013*), adapted to perform calculations of different mono-RNP particles around the st-sR26 RNA bound to either substrate D or substrate D'.

The protein building modules of the Box C/D sRNP, namely L7Ae, Fib, and Nop5, were derived from the crystal structure of the *P. furiosus* split di-RNP (PDB accession code 3NMU). The interfaces that are not perturbed in the complex, Nop5-NTD—Fib and Nop5-CTD—7Ae, were also derived from the same crystal structure and kept fixed during calculations. This is supported by chemical shift perturbation experiments, in which we verify that the interfaces formed in the Nop5-Fib subcomplex and the Nop5-CTD—L7Ae—K-turn subcomplex are retained in the full sRNP. The chemical shifts perturbation experiments performed on ILV methyl-labelled Fib are analysed in section 4.5, while checking of chemical perturbation in L7Ae upon sRNP assembly was performed in a previous study (Lapinaite *et al*, 2013).

The loops connecting these fixed regions were kept fully flexible during structure calculations. The helical region of the Nop5 coiled-coil domain was also kept rigid.

The two kink-turn regions of the st-sR26 sRNA were kept rigid during the structure calculation. Moreover, the substrate-guide duplex was fixed to adopt A-form helical geometry, consistent with previous biochemical and structural data (Appel & Maxwell, 2007). The guide sequence not bound to substrate RNA was kept flexible.

At the beginning of the calculation process, the two Nop5-Fib-L7Ae subcomplexes are separated in a randomised orientation around the coordinate center, with the substrate and sRNA kept away from the proteins in a randomised orientation that does not produce steric clashes (see figure 3.8.1).

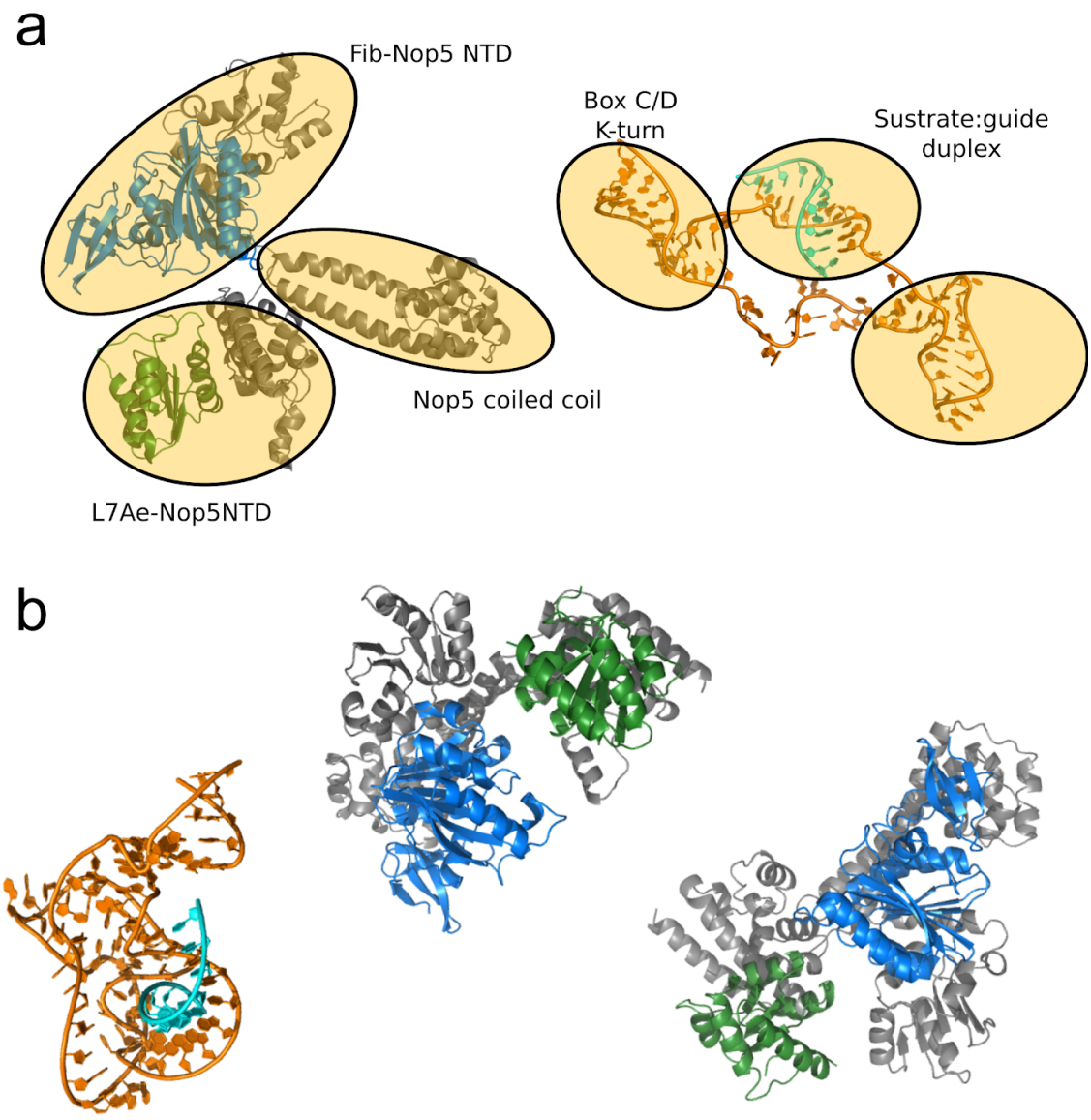


Figure 3.8.1. The structure building modules. A) The building modules of the Box C/D sRNP in the st-sR26 structure calculation, with the blocks that are kept rigid highlighted by circles. The substrate:guide duplex, the helical region of Nop5, the Nop5 CTD/L7Ae and Nop5-NTD—Fib modules are kept fixed. Loops connecting these sections remain fully flexible. B) The randomised starting arrangement of the Box C/D sRNP components with respect to each other, with the modules kept rigid. Here we show the substrate D'-bound state.

In the next step, the RNAs are moved to the center of the coordinate frame, and the calculation enters a simulated annealing step in torsion angle space driven by experimentally-derived distance restraints, together with distance restraints for the formation of the Nop5 coiled-coil and the ideal L7Ae/K-turn interaction, dihedral restraints on the unbound guide region of the sRNA and a center-of-mass restraint enforcing the experimentally-derived L7Ae/L7Ae center-of-mass distance.

For structures where one Fib copy is assumed to be bound to the substrate-guide duplex ([on,off]-state), distances between the A-form helix and the RNA-binding face of Fib, modelled on the mono-RNP structure of *S. solfataricus* (PDB accession code 3PLA), were also included.
An analysis of the interactions between the Nop5 α9 helix and the sRNA led to assuming the presence of specific interactions between the helix and free purine residues in the guide region to which no RNA is bound, which are available in the substrate D-bound case, but not in the substrate D'-bound case (see Fig. 3.8.2).

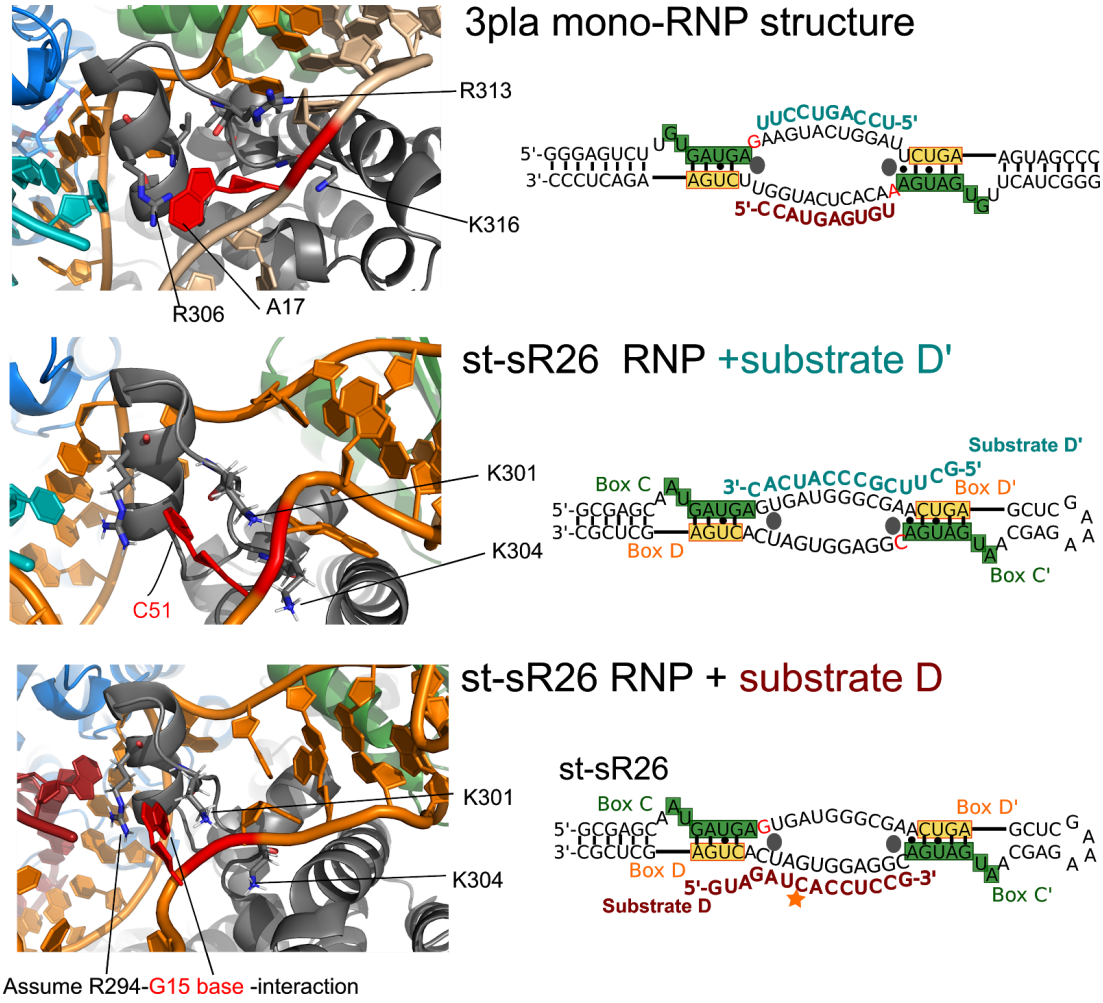


Figure 3.8.2. Assumptions on the Nop5 α9 helix-free guide RNA interaction. Top: the position of the Nop5 α9 helix in the holo mono-RNP structure from *S. solfataricus* (pdb accession code 3pla) relative to the free nucleotide after the box C/D or C'/D' element shows two basic residues on either side of the RNA phosphate backbone, R313 and K316, and a stacking interaction between R306 and the purine base. On the right hand side, the sRNA sequence is shown, with the α9 helix shown as gray cylinders and the interacting base highlighted in red. Middle: the Nop5 α9 helix interacting with the free guide region in the substrate D'-bound st-sR26 sRNP from *P. furiosus*. As the first free nucleotide is a pyrimidine, only the interaction between the basic residues and the phosphate was assumed throughout structure calculation. Interactions mapped onto the average structure of the [on,off]-state ensemble (see below and section 4.4). Bottom: the α9 helix interaction with the free guide region in the substrate D-bound st-sR26 RNP. In this case, both the interactions of R301 and R304 with the phosphate backbone and the interaction between R294 and G15 were assumed, by analogy with the holo mono-RNP structure.

For complexes where Fib is positioned away from the RNA ([off, off]-state), the two copies of Fib were considered equivalent. Thus, PRE-derived distances were imposed on both Fib or L7Ae subunits as such:

```
assign (resid 247 and name NAI and (segid NOP1 or segid NOP2))
       (resid  71 and name CD1 and segid FIB1) 24.4 2.0 2.0

assign (resid 247 and name NAI and (segid NOP1 or segid NOP2))
       (resid  71 and name CD1 and segid FIB2) 24.4 2.0 2.0
```

This represents a distance restraint between the nitrogen on the paramagnetic tag (NAI) on either Nop5 copy to the Cδ1 of isoleucine 71 on either Fib copy with an error of ±2Å. On the other hand, in complexes where one Fib copy is on the substrate:guide duplex ([on,off]-state), the two copies of Fib are non-equivalent. Thus, distance restraints were imposed with an OR statement indicating that the paramagnetic relaxation enhancement may have been induced by proximity to either Fib or L7Ae copy:

```
assign (resid 247 and name NAI and (segid NOP1 or segid NOP2))
       (resid  71 and name CD1 and segid FIB1 or segid FIB2) 24.4 2.0 2.0
```

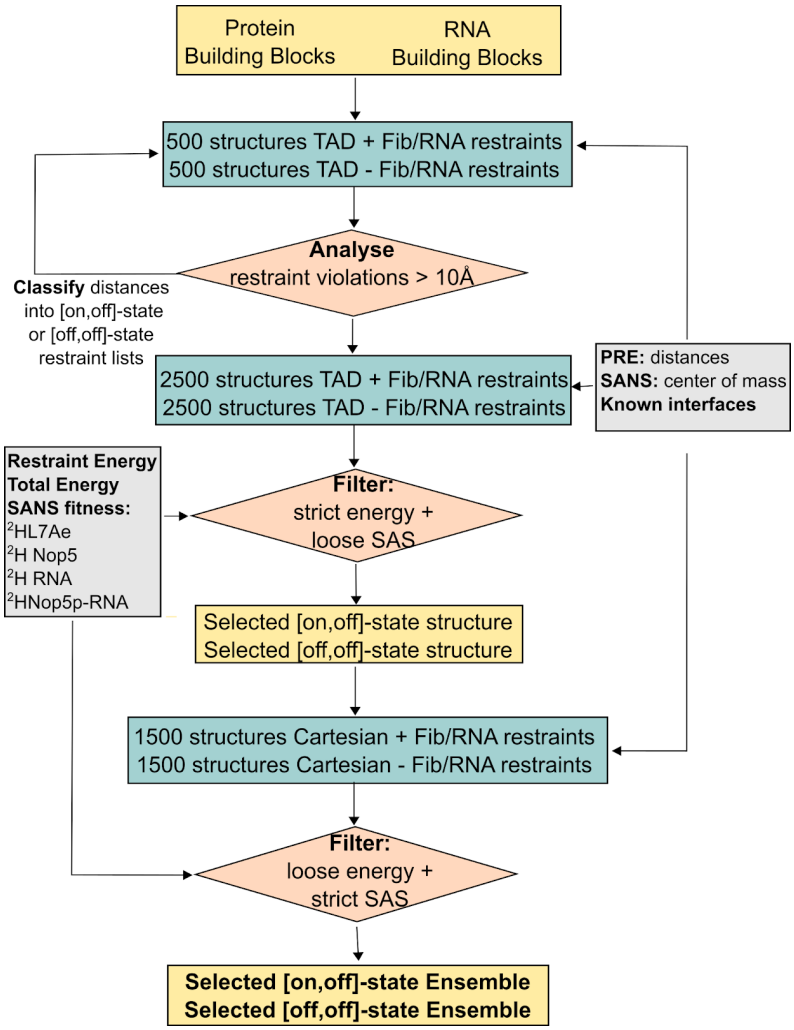


Figure 3.8.3. The st-sR26 structure calculation workflow. Calculations are run in parallel assuming that one Fib copy is either in the methylation-competent position ([on,off]-state) or both Fib copies are away from the RNA ([off,off]-state) and restraint violations larger in either state than 10Å are classified, so that two separate peak lists representing the on and off state emerge. After 4 torsion angle dynamics (TAD) iterations, no large violations are left and the restraint lists are used in a global TAD conformational search. Initial simulated annealing in torsion angle space brings the two halves of the complex together around the RNA, which lies at the center of the coordinate frame. Subsequent cooling steps bring the temperature of the system to 50 °C. A first selection step is applied based on a stringent violation energy filter, combined with cutoffs based on fitness to the ^{2}H-Nop5, ^{2}H-L7Ae, ^{2}H-RNA and ^{2}H-Nop5RNA 42%D_2O SANS curves. The selected structure is then used as the basis for a refinement in cartesian space. The final ensemble is chosen by a strict filter based on the fitness to the ^{2}H-Nop5, ^{2}H-L7Ae, ^{2}H-RNA and ^{2}H-Nop5/^{2}H-RNA 42%D_2O SANS curves, combined with a loose violation energy filter. The SAXS, ^{2}H-Fib and ^{2}H-Fib/^{2}H-RNA SANS curves were reserved for ensemble scoring as discussed in section 3.9.

For each sRNP the proteins and RNA were submitted to two sets of preliminary torsion-angle simulated-annealing procedures in parallel. One included a set of restraints positioning one of the two fibrillarin copies onto the methylation site of the substrate–guide duplex ([on,off]-state). In another set, no restraints were imposed between fibrillarin and the RNA ([off,off]-state). 500 structures were calculated per iteration.

The conformational sampling was driven by PRE-derived distance restraints, distance restraints positioning the two L7Ae–Nop5-CTD modules onto the RNA K-turns and a loose distance restraint between the centres of mass of the two L7Ae modules (90±15 Å), which was derived from the P(r) curve of ^{2}H-L7Ae in 42%/58% D_2O/H_2O.

For both substrate D-bound and substrate D'-bound cases, restraints positioning the Nop5-α9 helix between the two guide regions were also used (see Fig. 3.8.2). At the end of each iteration, restraint violations were evaluated: restraints violated by more than 10 Å in either set of calculations were classified, eliminated from that particular set, but kept in the other. This led to two restraint-lists per sRNP, corresponding to the [on,off]- and [off,off]-states of the sRNP. With these four sets of restraints (two for the substrate D-loaded and two for the substrate D'-loaded sRNP), four separate runs of torsion-angle simulated-annealing calculations were performed; we generated 2500 structures per run, using the settings described in Lapinaite et al. (2013) . Fitting of SAS curves was performed fitness was calculated with the programs CRYSOL and CRYSON, from the ATSAS suite, version 2.7.5.

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{I_{exp}(s_i)-cI_{calc}(s_i)}{\sigma(s_i)}\right]^2 \quad (15)$$

where I_{calc} represent the back-calculated value (I_{para}/I_{dia} or SAS intensities), I_{exp} is the corresponding experimental value, N is the number of experimental points, σ represents the experimental error and c is the scaling factor, required to scale theoretical SAS curves to experimental data:

$$c = \frac{\sum_{i=1}^{N}\frac{I_{exp}(s_i)I_{calc}(s_i)}{\sigma(s_i)^2}}{\sum_{i=1}^{N}\frac{I_{calc}(s_i)}{\sigma(s_i)^2}} \quad (16)$$

The structures ranking within the top 2% in both total energy and restraint energy were selected. To further narrow down the selection on the basis of the SAS data, we evaluated the χ^2 distribution of ^{2}H-Nop5, ^{2}H-L7Ae and ^{2}H-RNA SANS curves and picked loose cutoffs to eliminate the conformers that did not fit the scattering data. The SAS cutoffs were placed so as to prevent the worst-fitting structures from entering Cartesian refinement, which is a local search step.

Since we expected the position of the Nop5-NTD–Fib module to be variable, the SAS curves including the contribution of Fib were not used at this point, and reserved for ensemble scoring (see section 3.9). Therefore, after the global torsion angle dynamics step **structures in the top 2% by restraint energy and total energy that passed the loose SAS cutoffs were selected.** The cutoffs applied are shown in Fig. 3.8.4 and table 3.8.1.

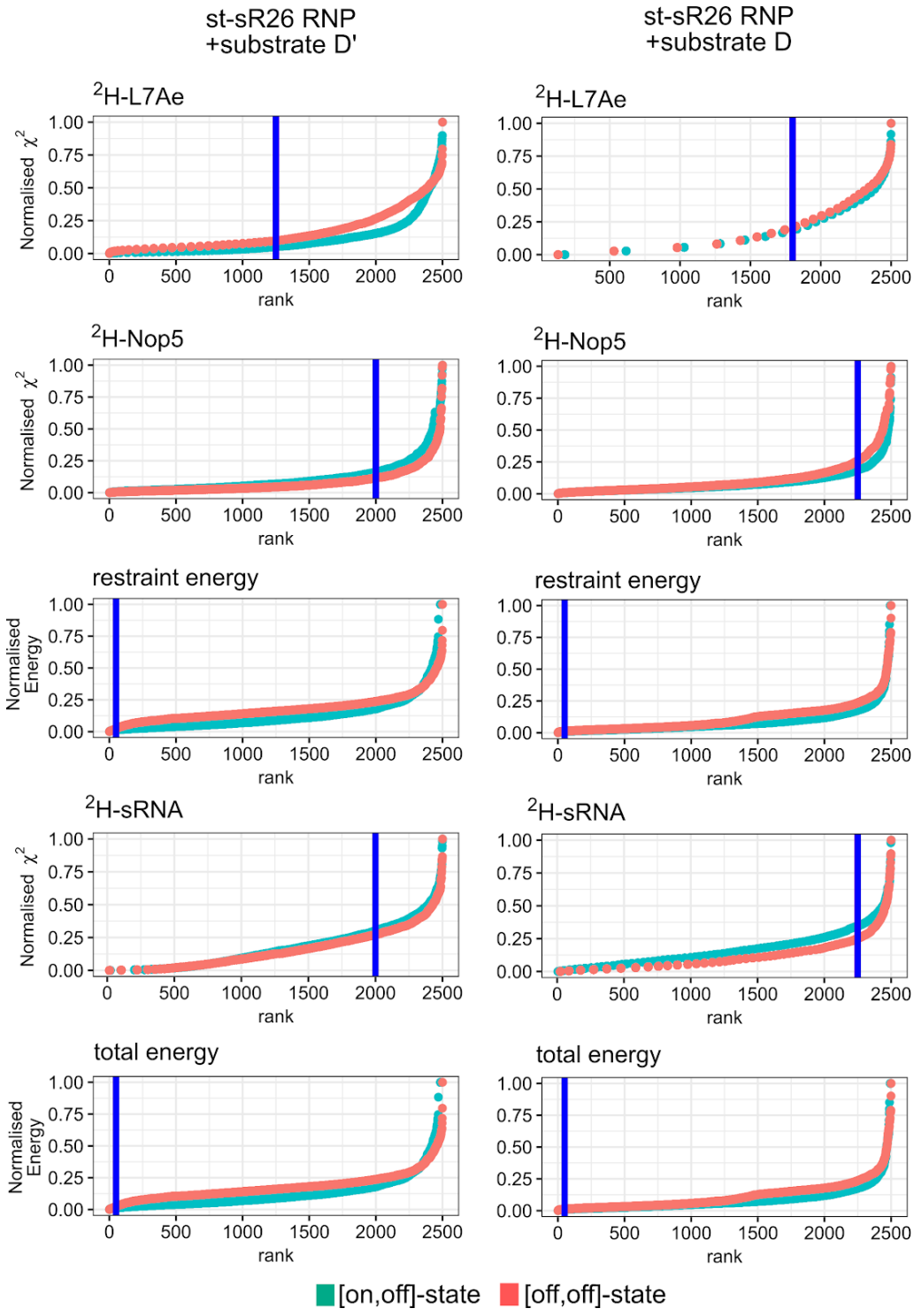


Figure 3.8.4. Selection after torsion angle dynamics. Consensus NMR-SAS selection after generation of 2500 conformers by torsion-angle simulated annealing for each state (st-sR26 RNP+substrate D' [on,off]-state; st-sR26 RNP+substrate D' [off,off]-state;st-sR26 RNP+substrate D [on,off]-state;st-sR26 RNP+substrate D' [off,off]-state. The vertical lines represent the cutoff point. At this stage, a string energy filter, which selects for the structures best fitting the experimental distance restraints, was used. All values shown are normalized between 0 and 1 for comparison purposes using equation (17).

SAS curve (42% D_2O)	st-sR26 RNP + substrate	Cutoff [off,off]-state	χ^2_{min} [off,off]-state	Cutoff [on,off]-state	χ^2_{min} [on,off]-state
^{2}H-sRNA	D'	80% (χ^2 <7.3χ^2_{min})	0.210	80%(χ^2<8.7χ^2_{min})	0.210
$^{0.2}$H-L7Ae	D'	50% (χ^2 <1.4χ^2_{min})	0.650	50% (χ^2<1.8χ^2_{min})	0.640
^{2}H-Nop5	D'	80% (χ^2 <6.7χ^2_{min})	0.510	80% (χ^2<8.9χ^2_{min})	0.490
^{2}H-sRNA	D	90% (χ^2 <3χ^2_{min})	0.260	90% (χ^2<2.9χ^2_{min})	0.270
^{2}H-L7Ae	D	75% (χ^2 <2.5χ^2_{min})	0.060	75% (χ^2<2.2χ^2_{min})	0.060
^{2}H-Nop5	D	90% (χ^2 <8.3χ^2_{min})	0.850	90% (χ^2<9.3χ^2_{min})	0.870

Table 3.8.1. SAS Selection after torsion angle dynamics. The cutoffs applied to the SAS curves for selection after global sampling by torsion-angle simulated annealing. The "cutoff" column reports the percentage of structures allowed by the cutoff and where the cutoff is placed in relation to χ^2_{min}, the minimum χ^2 of the run against that particular curve. The χ^2_{min} column reports the actual $\chi^2_{min.}$. As χ^2 factors in experimental error, the raw χ^2 values are not comparable between rows (each curve has different experimental error), only across columns.

Among the selected structures of each of the four runs ([on,off]- and [off,off]-states of both substrate D- and substrate D'-loaded sRNPs), the one in each run with the lowest restraint-violation energy that maintained the correct RNA topology was chosen as the starting point for a refinement in Cartesian space.

The four refinement runs were comprised of 1500 structures each spanning up to 9 Å root-mean-square deviation (RMSD) of Cα and P atoms to the starting structure (number calculated for substrate D'-loaded [on,off]-state). **At the end of the refinement**, we applied **stringent selection criteria with respect to the SAS curves** and **loose criteria with respect to the energy** (**top 33% total energy, restraint energy and van der Waals energy** in most cases, see Fig.3.8.5). The cut-offs for the SAS data were set upon visual inspection of the χ^2 distributions for each run and curve, whereby we selected different numbers of structures depending on the shape of the distribution (fewer structures were selected for steeper distributions) and ensuring that common solutions to all curves were found.

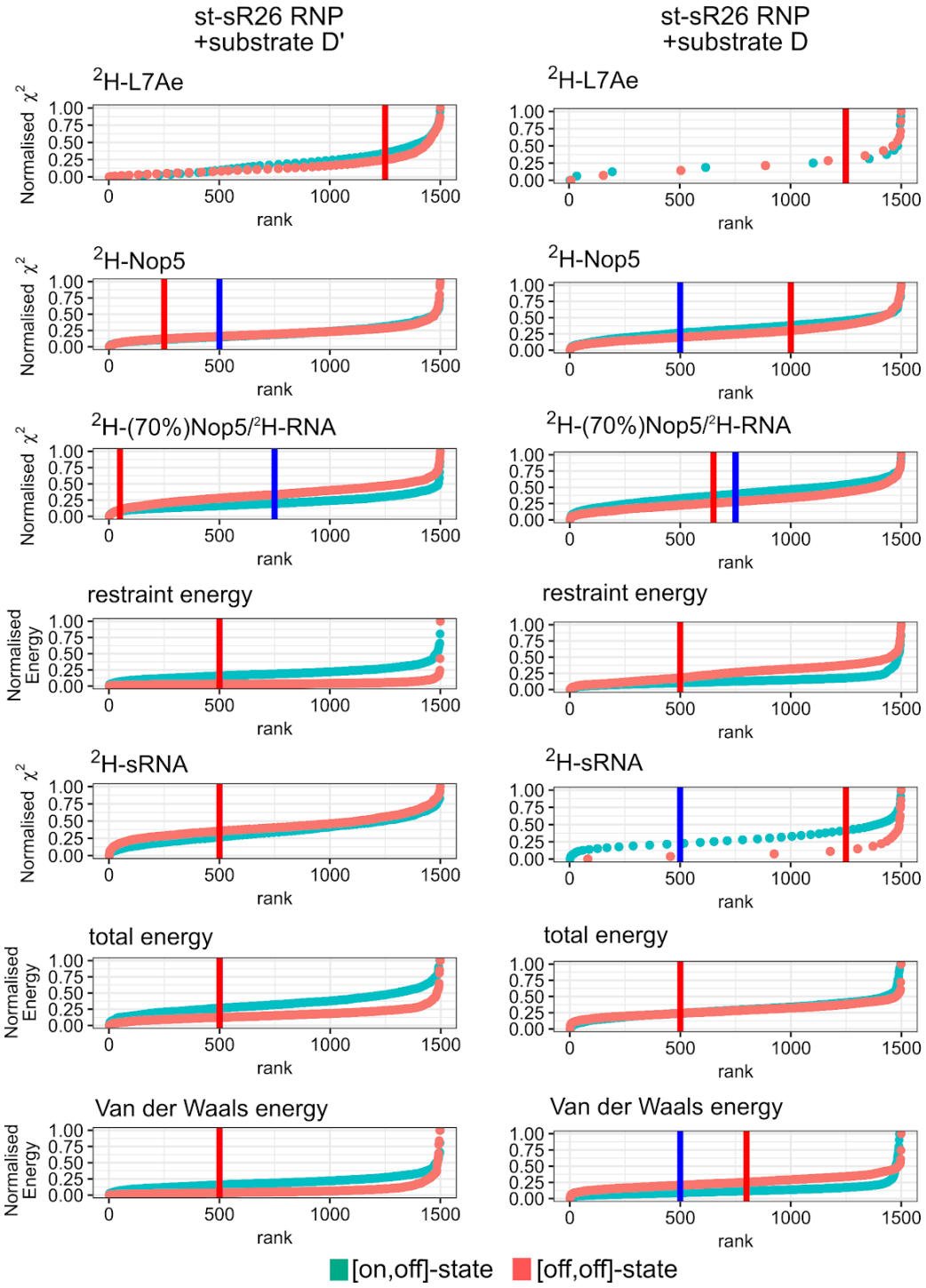


Figure 3.8.5. Selection after Cartesian refinement. Consensus NMR-SAS selection after generation of 1500 conformers by simulated annealing in Cartesian space for each state (st-sR26 RNP+substrate D' [on,off]-state; st-sR26 RNP+substrate D' [off,off]-state;st-sR26 RNP+substrate D [on,off]-state;st-sR26 RNP+substrate D' [off,off]-state. When two vertical lines are present, the blue one is the cutoff point for the [on,off]-state and the red one is the cutoff point for the [off,off]-state. At this stage, a loose energy filter was applied, combined with more stringent SAS filters. All values shown are normalized between 0 and 1 for comparison purposes using equation (17).

SAS curve (42% D_2O)	st-sR26 RNP + substrate	Cutoff [off,off]-state	χ^2_{min} [off,off]-state	Cutoff [on,off]-state	χ^2_{min} [on,off]-state
^{2}H-sRNA	D'	33% (χ^2<3.1χ^2_{min})	0.470	33% (χ^2<1.6χ^2_{min})	0.790
$^{0.2}$H-L7Ae	D'	83%(χ^2 <1.6χ^2_{min})	0.680	83%(χ^2 <1.6χ^2_{min})	0.640
^{2}H-Nop5	D'	16% (χ^2 < 2.71χ^2_{min})	0.600	33% (χ^2<2.1χ^2_{min})	0.710
^{2}H(70%)-Nop5 /^{2}H-RNA	D'	3% (χ^2 <6.7χ^2_{min})	1.670	50% (χ^2<3.1χ^2_{min})	5.150
^{2}H-sRNA	D	88% (χ^2 <1.1χ^2_{min})	0.260	33% (χ^2 <1.1χ^2_{min})	0.310
^{2}H-L7Ae	D	83% (χ^2 <1.6χ^2_{min})	0.060	83% (χ^2 <3.7χ^2_{min})	0.060
^{2}H-Nop5	D	66% (χ^2 <2.0χ^2_{min})	0.940	33% (χ^2 <3.7χ^2_{min})	0.310
^{2}H(70%)-Nop5/ ^{2}H-RNA	D	43% (χ^2 <3.5 χ^2_{min})	2.200	17% (χ^2 <4.3χ^2_{min})	4.100

Table 3.8.2. SAS Selection after Cartesian refinement. The cutoffs applied to the SAS curves for selection after global sampling by torsion-angle simulated annealing. The "cutoff" column reports the percentage of structures allowed by the cutoff and where the cutoff is placed in relation to χ^2_{min}, the minimum χ^2 of the run against that particular curve. The χ^2_{min} column reports the actual $\chi^2_{min.}$. As χ^2 factors in experimental error, the raw χ^2 values are not comparable between rows (each curve has different experimental error), only across columns.

Applying these criteria we selected a single structure for the **substrate D'-loaded** [off,off]-state and 11 structures for the [on,off]-state. The [on,off]-state structures displayed an average RMSD of **1.97 Å**, calculated on all C α and P atoms (Supplementary Fig. S4) excluding the fully flexible regions, namely the free guide region of the RNA (nucleotides 51–62), the loops connecting the Nop5-NTD to the coiled-coil domain (residues 116–122), and the loops connecting the coiled-coil domain to the Nop5-CTD (residues 249–251).

The final ensembles for the **substrate D-loaded** [off,off]- and [on,off]-states consist of 6 and 5 structures, respectively, with a Cα/P RMSDs of **3.05** and **2.51 Å,** respectively. All structures in the final ensembles were minimized in explicit water using Amber14 and the corresponding Amber99SB force field.

Thus, the structure calculation process generated four sets of pdb files: substrate D'-bound [on,off]-state; substrate D'-bound [off,off]-state; substrate D-bound [on,off]-state; substrate D-bound [off,off]-state.

The structures are described in section 4.7.

3.9 Consensus NMR-SAS scoring of conformational ensembles.

The following text is adapted from a manuscript in preparation and was originally written by myself:

The peak-shapes in the ^{1}H-^{13}C methyl spectra, the PRE data and the scattering data indicated the presence of a conformational equilibrium between states with one Fib copy onto the substrate-guide duplex ([on,off]) and states with both Fib copies away from the RNA ([off,off]), as discussed in section 4.4. The ^{2}H-Fib, ^{2}H-fFib/^{2}H-RNA SANS and SAXS curves were therefore fitted to a mixture of structures in the [on,off]-and [off,off]-states.

For this procedure, we chose as representative structures for the [on,off]- and [off,off]-states of each sRNP (substrate D- and substrate D'-loaded) the structure of the corresponding ensemble that is closest to the mean structure of the ensemble. When each SAS curve reporting on different subunits of the RNP is fit as an equilibrium of the [on,off]- and [off,off]-state conformers selected by Cartesian refinement, no single population ratio can fit the all the data. This indicates that, in addition to coexistence with the major [on,off]- and [off,off]-state conformers, the flexible Nop5-NTD−Fib module adopts different orientations, which significantly affect the fitting of curves reporting on Nop5, Fib and the overall shape of the complex, but not the rest of the SAS data.

In order to address the flexibility of the Nop5-NTD−Fib modules not in contact with the RNA, we sought to generate ensembles containing different orientations of these modules that

would improve the fit to the SAS curves. This conformational diversity is in addition to the equilibrium between the [on,off]- and [off,off]-states, resulting in a pool of structures containing both [on,off]- and [off,off]- states and multiple conformations of Nop5-NTD−Fib modules in each of these states.

To generate these ensembles we proceeded as follows. We chose the representative structure of each ensemble coming out of Cartesian refinement (Fig. 4.7.1, 4.8.1) as a template for a further simulated-annealing step. Here, the loops connecting the Nop5-NTD−Fib modules placed away from the substrate-guide duplex to the rest of the Box C/D particle were allowed to adopt random orientations, while the rest of the particle was kept rigid. At this stage, we generated 2000 structures with randomized Nop5-NTD−Fib positions, from which we removed structures containing steric clashes. The structures also contained all spin-labels, which were left flexible, in order to allow calculation of the fit to the PRE data (see below).
In a separate run comprising of 300 structures, the template structures were kept entirely rigid while the spin-label side chains were allowed to rotate, in order to generate different orientations, as multiple conformations of the spin-label have been demonstrated to fit the PRE data more accurately than a single one (Iwahara *et al., 2004*).

Ensemble scoring was carried out for substrate D'- and substrate D-loaded sRNPs via the pseudo-genetic algorithm shown in Fig. 3.9.1, which we developed for this application. First, we grouped the structures into four pools, containing 1700, 1700, 300 and 300 structures: [on,off]-state with randomized Nop5-NTD−Fib positions, [off,off]-state with randomised Nop5-NTD−Fib positions, [on,off]-state with randomized spin-label orientations and [off,off]-state with randomized spin-label orientations. Each iteration, The algorithm sampled four "parent" ensembles, each comprising of 4–10 conformers randomly chosen from the pools. These ensembles were merged into a single parent pool, which was sub-sampled, yielding 20 "children" sub-samples ranging from 3–10 conformers in size. Each sub-sampling event had a 30% probability of duplicating a conformer or replacing one with another from the main pool. The process of parent selection, sub-sampling and scoring was repeated 250 times.

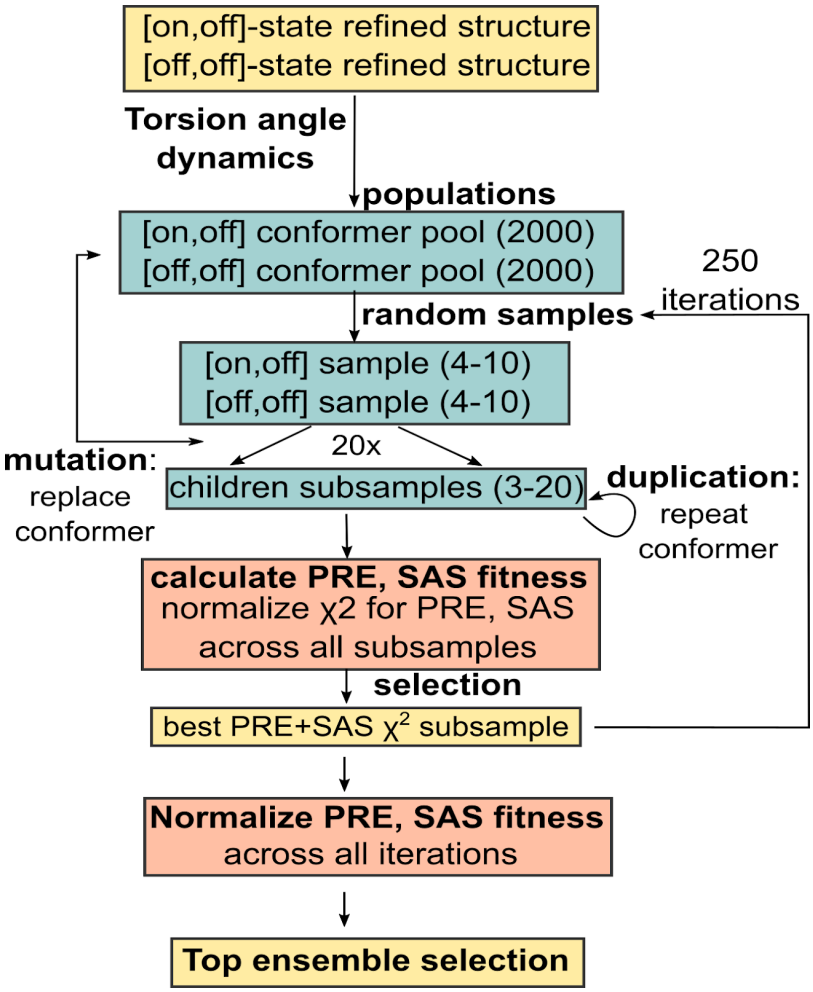


Figure 3.9.1. Pseudo-genetic algorithm developed for selection of structural ensembles with consensus fit to all PRE and SAS data. The number of conformers generated or selected at each step is given in parentheses. Representative mean structures of the four ensembles resulting from Cartesian refinement of [on,off]- and [off,off]-state structure calculations for both substrate D- and substrate D'-loaded st-sR26 RNPs (section 4.7) are used as templates for a further round of torsion-angle simulated-annealing. For both [on,off]- and [off,off]-states, 300 structures are generated where the conformations of the spin-labels are randomized, and a further 1700 structures with randomized positions of the Nop5-NTD−Fib modules not in contact with the RNA. This generates 4 conformer pools of [on,off]- and [off,off]-state structures, which are then sampled in 250 successive iterations, each containing 20 sub-sampling events. For each sub-sample, theoretical ensemble PRE effects and SAS scattering curves are computed, and fitness to the experimental data is calculated via χ^2. Normalized consensus SAS fitness and PRE score are equally weighted in selecting the best sub-sample from each iteration.

The theoretical scattering curve of a child ensemble was computed as the linear combination of the scattering curves of each individual conformer (scaling the populations to represent

molar fractions rather than volume fractions, which is the standard ATSAS output). The χ^2 to the experimental data was calculated by OLIGOMER (Konarev *et al*, 2003). The normalization of χ^2 of all sub-sampled ensembles and across iterations was done according to equation (16):

$$\chi_{norm}{}^2 = \frac{\chi_{ensemble}{}^2 - \chi_{min}}{\chi_{max} - \chi_{min}} \qquad (17)$$

Where $\chi^2_{ensemble}$ is the fitness of an individual ensemble, and χ^2_{min} and χ^2_{max} are the respective minimum and maximum values across the iterations or sub-samples being considered. Five SAS curves were used for scoring: ^{2}H-Nop5, ^{2}H-Fib, ^{2}H-Fib/^{2}H- RNA, ^{2}H(70%)-Nop5/^{2}H-RNA and SAXS. The normalized χ^2 values for each curve were then summed and renormalized into a single value, obtained with the same equation (17), which then represented the overall SAS-fitness.

In order to calculate theoretical I_{para}/I_{dia} ratios from mixed [on,off]- and [off,off]-state ensembles, we considered the Nop5-NTD−Fib module to be in fast-exchange between the two states on the relevant NMR time-scales, consistent with the spectra in Fig. 4.4.1. Back-calculated PRE effects could therefore be calculated using $\langle r^{-6} \rangle$ ensemble-averaged distances over all [on,off]- and [off,off]-states. Each methyl group of each fibrillarin or L7Ae copy is influenced by two PRE tags (SL1 and SL2). The resulting Γ_2 values for the methyl groups of the two copies are given by:

$$\Gamma_2^{Methyl1} = \Gamma_{2,SL1}^{Methyl1} + \Gamma_{2,SL2}^{Methyl1}$$

$$\Gamma_2^{Methyl2} = \Gamma_{2,SL1}^{Methyl2} + \Gamma_{2,SL2}^{Methyl2} \qquad (18)$$

where Methyl1 and Methyl2 refer to the two copies of L7Ae or fibrillarin. Because Methyl1 and Methyl2 have almost indistinguishable chemical shifts, the resulting I_{para}/I_{dia} ratios for Methyl1 and Methyl2, back-calculated using equations (12,13), were averaged before comparison to the experimental data.

The PRE fitness was quantified using χ^2 to all experimental PRE values using equation (14). Distances were computed from the PDB files using the Biopython Bio.PDB module (Hamelryck & Manderick, 2003). The fitness of PRE data was normalized using equation (11) and summed with the SAS- fitness score, to yield a consensus PRE-SAS score for each ensemble within the 20 sub-sampling events, and across the 250 iterations.

Three independent runs of the scoring algorithm were performed for substrate D'- and substrate D-loaded sRNPs, with the top scoring ensemble, judged by the consensus

PRE-SAS score, displayed in Fig. 4.9.1. The theoretical I_{para}/I_{dia} ratios of the final selected ensemble were also assessed used the PRE Q-factor parameter, which is a quality-of-fit score commonly used for PRE assessment (Clore & Iwahara, 2009).

$$Q = \left(\frac{\sum_i \{\frac{I_{para}}{I_{dia}} exp(i) - \frac{I_{para}}{I_{dia}} theor(i)\}^2}{\sum_i \frac{I_{para}}{I_{dia}} exp(i)^2}\right)^{1/2} \qquad (18)$$

Where $\frac{I_{para}}{I_{dia}} exp$ is the experimentally-determined I_{para}/I_{dia} ratio for an individual resonance, and $\frac{I_{para}}{I_{dia}} theor$ its corresponding value back-calculated from the ensemble of conformers.

4. Results

4.1 Box C/D oligomeric state depends on the sRNA used in reconstitution.

In order to investigate the role of the physiological guide regions in the assembly and activity of Box C/D, we used the ssR26 scaffold to obtain a version of the sR26 guide RNA that yields a stable complex capable of methylating two different substrates which directs methylation of C1364 and C1490 in 16S rRNA in *P. furiosus* with its guide D' and guide D regions.

This new RNA construct, while retaining the guide regions from sR26 (see Figure 4.1.1), includes the features present in ssR26 that induce increased stability of the reconstituted RNP, hence the name st-sR26. st-sR26 therefore contains a GAAA tetraloop as the apical loop, and symmetrical box C/D box C'/D' elements expected to fold into two K-turn structures, instead of one K-turn and one K-loop structure. st-sR26 was used for biochemical and structural studies on the apo, substrate D'-bound, substrate D-bound and holo complexes, after establishing that it correctly reproduces the methylation behaviour of RNPs reconstituted with the sR26 RNA.

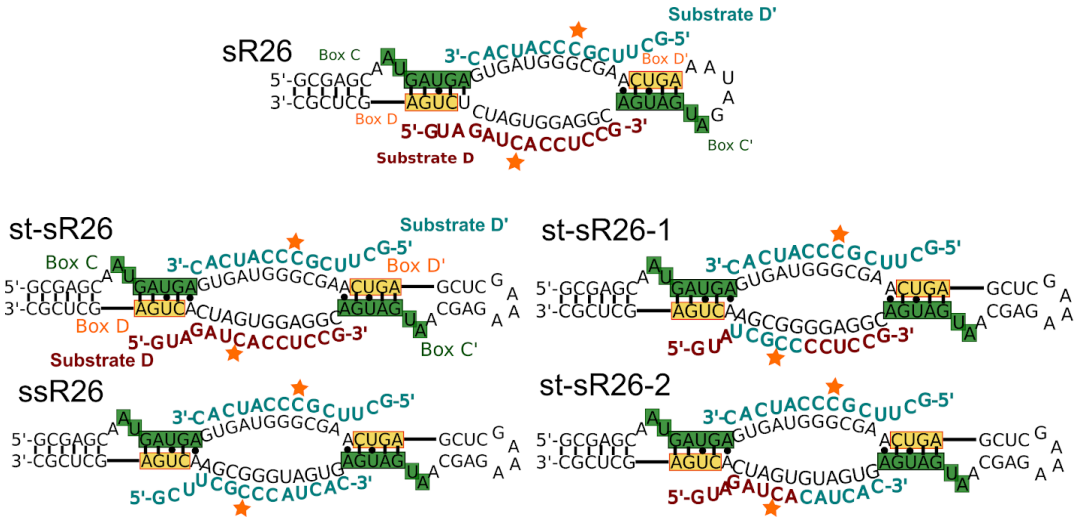


Figure 4.1.1. RNA sequences used to probe mono- and di-RNP formation. All the sequences used in structural studies are stabilised (st-) with respect to their physiological counterpart by making box C/D and box C'/D' elements identical to each other, generating two kink-loop elements. Moreover, the apical loop is converted to a stable GAAA tetraloop. st-sR26 is the guide RNA used for structural studies in this work. ssR26 is the guide RNA used for the di-RNP structure previously

published by our group (Lapinate *et al.,* 2013). st-sR26-1 and st-sR26-2 are chimeric constructs using half the guide sequence of st-sR26 and half the guide sequence of ssR26.

When reconstituting the sRNP with st-sR26, the resulting apo particle is slightly smaller than the one previously observed for complexes reconstituted with ssR26: its radius of gyration, R_g, derived from small-angle X-Ray scattering, equals 54.3 Å versus the 60.7 Å measured for the complex reconstituted with ssR26 (Fig.4.1.5). The apparent difference in size may either be due to a different conformation in the apo state, or to the presence of higher order aggregates in the complex containing ssR26, as suggested by the presence of a front shoulder in the final gel filtration reconstitution step (see Fig 4.1.2). Nevertheless, the R_g of the apo st-sR26 RNP is consistent with a di-RNP particle, as the analysis in Fig. 4.1.6 shows.

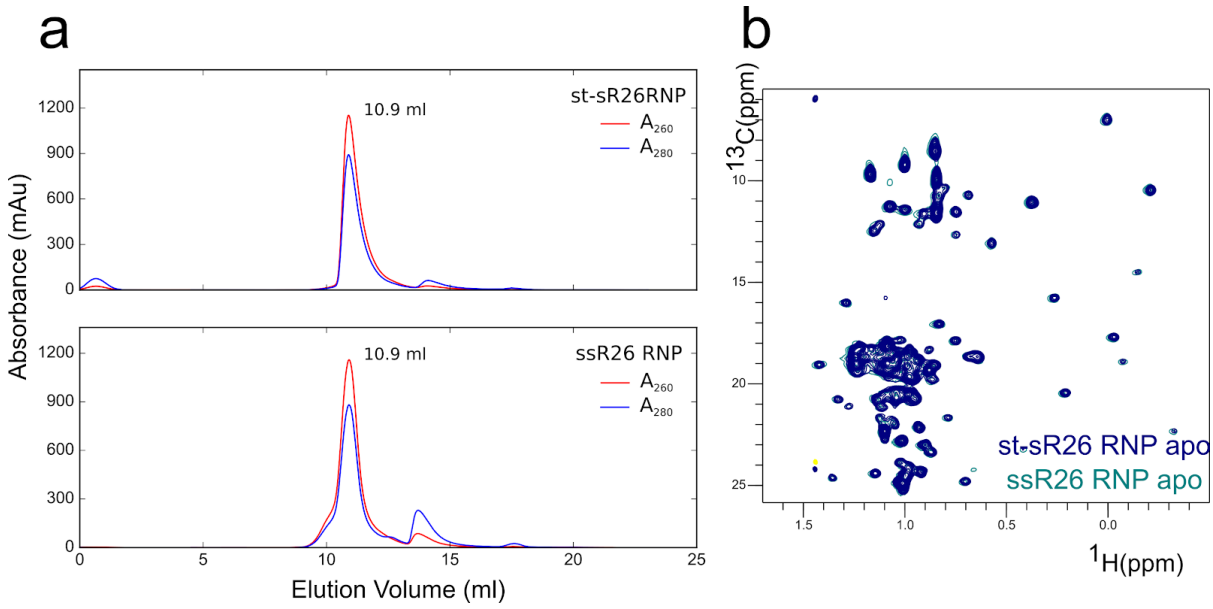


Figure 4.1.2. Reconstitution of complexes assembled using st-sR26 and ssR26 sRNAs. A) Size exclusion chromatography of ssR26 and st-sR26 Box C/D RNP reconstitution. The RNAs induce the formation of similar apo RNPs (eluting at 10.9 ml, ~400 KDa M_r), albeit with different degrees of monodispersity. The second peak at at ~14 ml contains excess proteins from the reconstitution. The samples are run on a superdex S200 10/300 increase (GE healthcare) at room temperature on the same set-up[2]. B) Overlay of ^{1}H-^{13}C HMQC NMR spectra of ILV-methyl Fib in the apo st-sR26 RNP and ssR26 RNPs. Each methyl-group yields a single peak at

[2] During my studies, the gel filtration set-up in the laboratory (loop lengths, specific column and purifier system) changed over time. M_r values reported are obtained from calibrations with molecular standards on the same set up as the sample measured. However, elution volumes are not directly comparable between figures (chromatograms are shifted by a constant).

the same position in both sRNPs, indicating that Fib is in the same conformation and chemical environment.

Another factor in determining the oligomeric state of the apo form of the sRNP is the salt concentration. Both gel filtration and SAXS measurements indicate that lowering the salt concentration to 150 mM results in the formation of higher order assemblies (see Fig. 4.1.3), presumably mediated by higher order oligomerisation of the Nop5 coiled-coil domain. -

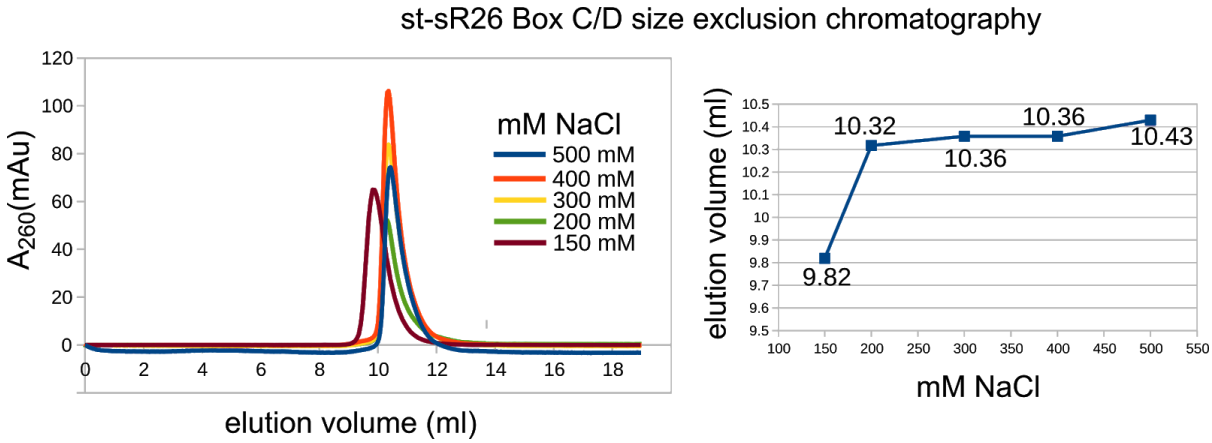


Figure 4.1.3. Dependence of Box C/D reconstitution on salt concentration. Box C/D RNPs reconstituted with st-sR26 RNA in 50 mM NaCl, 500 mM NaPi pH 6.6 run on size exclusion chromatography with buffers of varying salt concentration. Using molecular weight calibration on the same set-up, we could judge that the complex shifts from an apparent M_r of 480 KDa (10.43 ml) at 500 mM NaCl to an apparent M_r of 700 KDa at 150 mM NaCl. All runs are performed at room temperature.

Upon addition of either substrate D or substrate D', the st-sR26 RNP transitions from a di-RNP state to a mono-RNP state, which is preserved in the holo particle. This is in contrast to the ssR26 RNP, which is a di-RNP in both apo and holo states (Fig. 4.1.5 and Lapinaite *et al.,* 2013). We have confirmed this finding by a series of biochemical and biophysical methods, including NMR, SAXS, SANS. In NMR experiments probing ILV-methyl labelled Fib within the context of a deuterated st-sR26 RNP, the Fib methyl group intensities are found to increase by a factor of 1.3, consistent with a reduction in the correlation time of the particle.

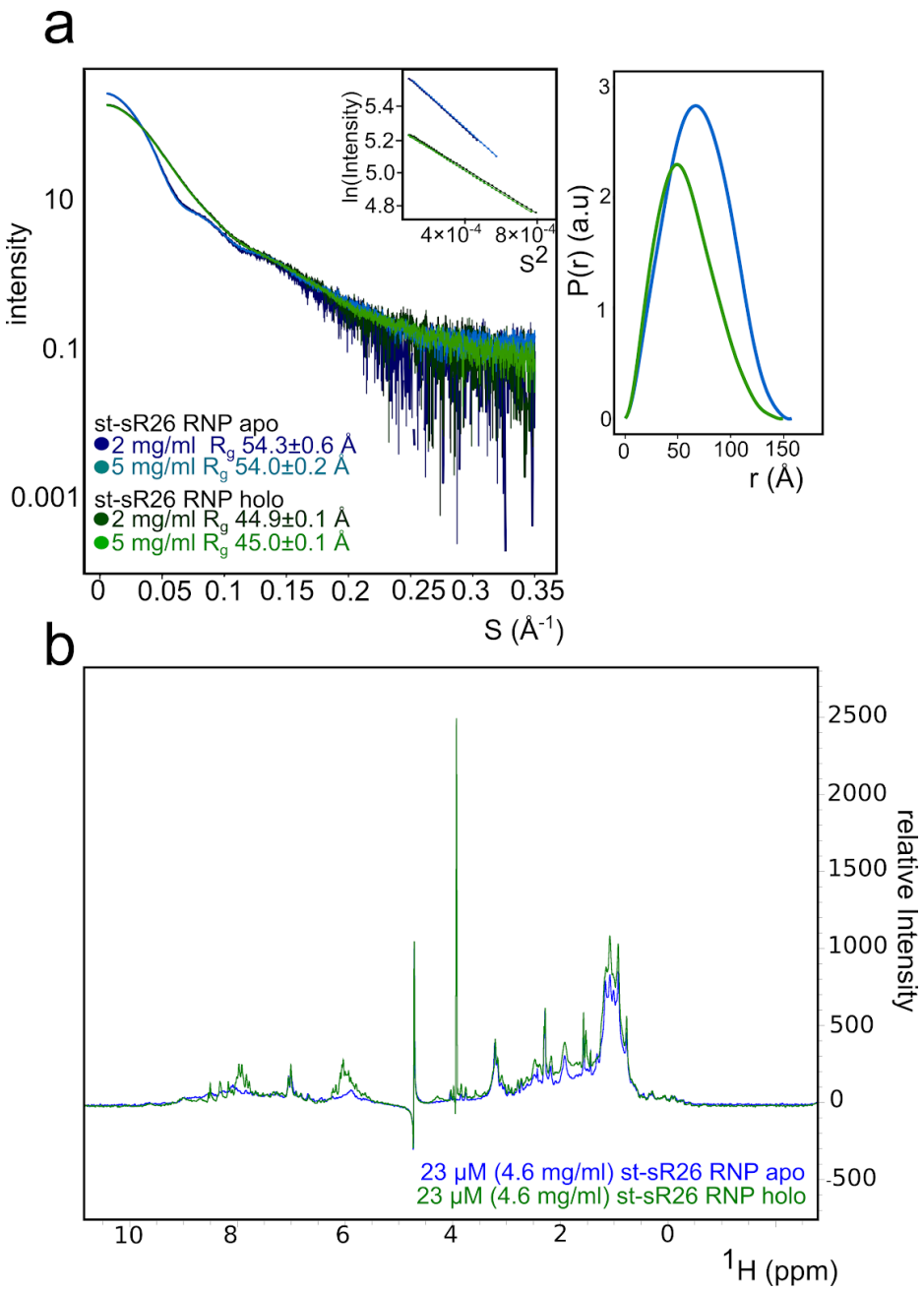


Figure 4.1.4. SAXS of st-sR26 Box C/D shows di-RNP to mono-RNP transition independent of concentration. A) Left: SAXS curves of complexes reconstituted with st-sR26 RNA show a transition from di-RNP to mono-RNP architecture, with Guinier plot and fit displayed in the inset. The reduction in R_g, as well as the extension of S_{max} of the Guinier region from 42 to 54 points are consistent with the transition to a mono-RNP particle. The fact that the apparent R_g is independent of concentration and that the curve is linear in the Guinier region (inset) indicates that the sample does not display interparticle attraction or repulsion. Right: The Pair distance distribution function P(r) extracted from the experimental scattering curves at 5 mg/ml with the program GNOM (Svergun, 1992); the plot represents the distribution of all atom-atom distances in the system as extracted from SAXS data. All samples measured at 40 °C. The holo sample is measured in the

presence of 1.25 molar equivalents of substrate D and substrate D'. B) Increase in NMR ^{1}H intensities upon addition of substrate RNA onto Box C/D RNPs reconstituted with st-sR26 RNA, ^{2}H-Nop, and ILV-methyl labelled Fib. The methyl region (2-0 ppm) displays an increase in intensity of a factor of 1.3 upon addition of substrate RNAs. The appearance of sharp peaks in the region between 5 and 7 ppm is indicative of the presence of free excess substrate RNA. The measurement is carried out at 55 °C, and the holo complex is generated by addition of 3x molar equivalents of substrate D and substrate D'.

In an initial set of SAXS experiments, we investigated the concentration dependence of the apo and holo st-sR26 complexes, and found that the transition from a di-RNP to a mono-RNP is independent of concentration, down to 2 mg/ml, at 40 °C (see Fig. 4.1.4). In addition, we found that the complex does not display concentration-dependent scattering behaviour or interparticle effects, enabling us to rely on the R_g parameter extracted from the Guinier region of the small-angle scattering curves. Moreover, we measured the ssR26 complex at the same concentration (Fig.4.1.5). Upon addition of substrate RNA in SAXS, the ssR26 complex displays an elongated structure, consistent with the holo di-RNP features previously described for this RNP. This particle does not show a reduction in apparent molecular weight upon addition of the substrate RNA, judging from R_g, I(0) and particle volume extracted from the small-angle scattering curve, unlike RNPs reconstituted with the st-sR26 sRNA.

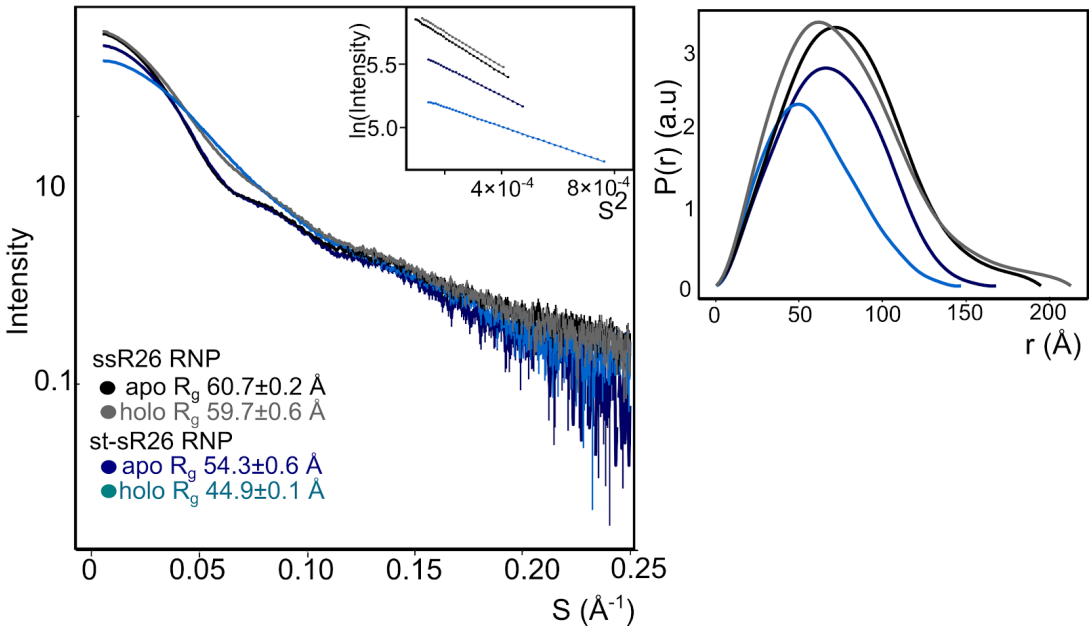
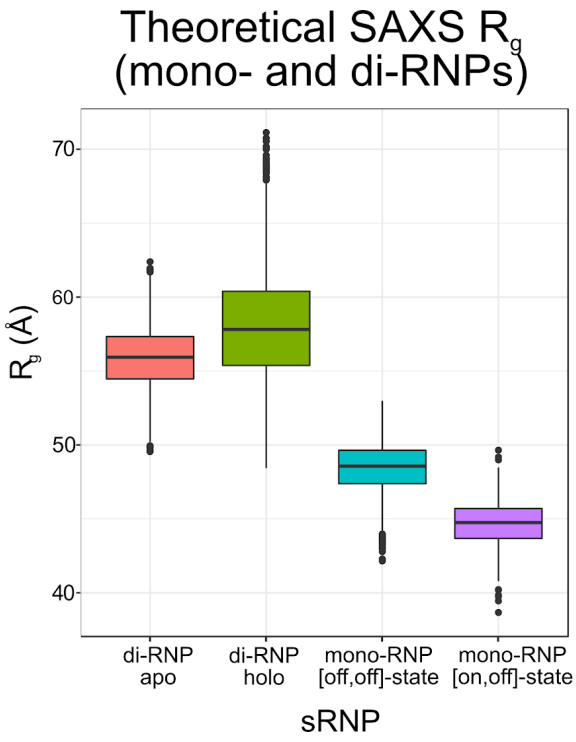


Figure 4.1.5. SAXS analysis of Box C/D reconstituted with st-sR26 and ssR26. ssR26-reconstituted Box C/D retains a di-RNP architecture upon addition of 2.25x substrate D':sRNA (4.5x substrates:complex, due to the fact that ssR26 binds two identical substrates). Upon substrate

binding, the RNP displays a slight particle elongation, as observed in the increase of the maximum distance of the P(r) function, but no apparent size decrease, as judged by R_g and forward scattering intensity I(0). This reproduces the features of the structure of the holo di-RNP complex characterised in Lapinaite *et al.* (2013) for this sRNP. The st-sR26 particle transitions from a di-RNP architecture, albeit one with a lower R_g, to a mono-RNP structure upon addition of 1.25 molar equivalents of substrate D and substrate D'. All particles measured at 2 mg/ml concentration at 40 °C.

We used the models generated by the exhaustive global search of the conformational space of mono- and di-RNPs (given the topological constraints of the complex such as protein connectivity, conserved structural elements) to interpret the observed SAXS R_g values, in order to establish which R_g values are compatible with the mono and di-RNP arrangements,. We computed the theoretical SAXS R_g for each particle after generation of models in the global torsion angle dynamics step of our structure calculation protocol (see section 3.7).



sRNP	Experimental SAXS Rg (Å)
st-sR26 apo	54.3
st-sR26 + substrate D	50.0
st-sR26 + substrate D'	47.3
st-sR26 holo	44.9
ssR26 apo	60.7
ssR26 holo	59.7

Figure 4.1.6. Theoretical SAXS R_g for various sRNP arrangements. The theoretical SAXS R_g values for Box C/D complexes from the global torsion angle dynamics search steps were computed using CRYSOL. For the di-RNP apo and holo state, 5000 conformers generated for calculation of the ssR26 di-RNP structure in Lapinaite *et al.* were used. In the case of the holo di-RNP, two out of four copies of Fib were assumed to be in proximity of the substrate-guide duplex. For the mono-RNP

particles, two states were computed, each with 2500 conformers where the RNP is bound to substrate D'. In the [off,off]-state both Fib copies are placed away from the substrate-guide duplex. In the [on,off]-state one Fib copy is positioned onto the substrate-guide duplex. The theoretical R_g of the particle where both Fib copies are in the on-position ([on,on]-state) is taken to be represented by the crystal structure of the holo mono-RNP (PDB accession code 3PLA), which has a theoretical SAXS R_g of 40.0 Å. The plots are consistent with the st-sR26 complex transitioning to a mono-RNP arrangement upon binding of substrate RNA.

Thus, we were able to correlate the observed SAXS R_g values of apo st-sR26 RNP, substrate D'-bound st-sR26 RNP, substrate D-bound st-sR26 RNP and holo st-sR26 RNP to the range of theoretical values expected for mono- and di-RNP particles, extracted from known structures and from modelling efforts (see figure 4.1.6).

Additionally, in SANS experiments probing the low-resolution features of the Fib subunits within the complex, the P(r) function obtained for D-bound, D'-bound and holo st-sR26 Box C/D are consistent with the presence of only two Fib copies, something which becomes apparent once the SANS data is used in *ab initio* modelling of the Fib scattering envelope (Fig. 4.1.7).

When the ^{2}H-RNA SANS curve from substrate D or D'-bound st-sR26 Box C/D is used in *ab initio* modelling with the programs DAMMIN (Svergun, 1999) or DAMMIF (Franke & Svergun, 2009), the resulting volume of the derived RNA shape cannot accommodate two RNA molecules.

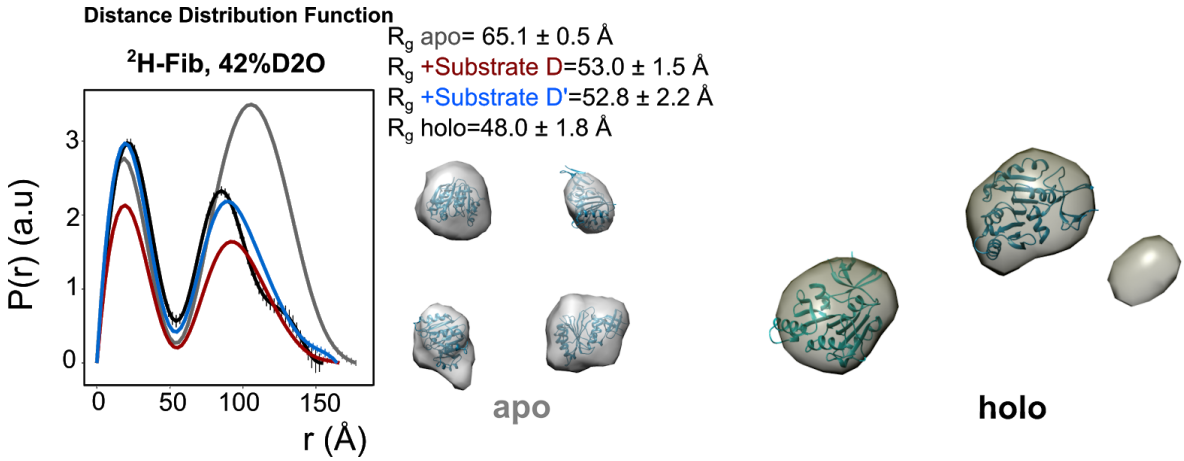


Figure 4.1.7. SANS analysis of st-sR26 Box C/D Fib subunits. P(r) functions and R_g values extracted from a SANS titration of substrate RNAs onto Box C/D reconstituted with st-sR26 RNA, and ^{2}H-Fib in 42% D_2O buffer. At this D_2O concentration, the contribution of all other proteins in the complex is masked at low s-values (see Fig. 1.5.1). Substrate RNAs are added at a 1.25 molar ratio. *Ab initio* low-resolution envelopes were computed using DAMMIN from the plotted P(r) curves. Fib is fitted into the densities using CHIMERA. As such, the Fib position is not meant to represent the orientation of the methyltransferase in the complex. The extra density in the holo complex is too small to fit a Fib subunit.

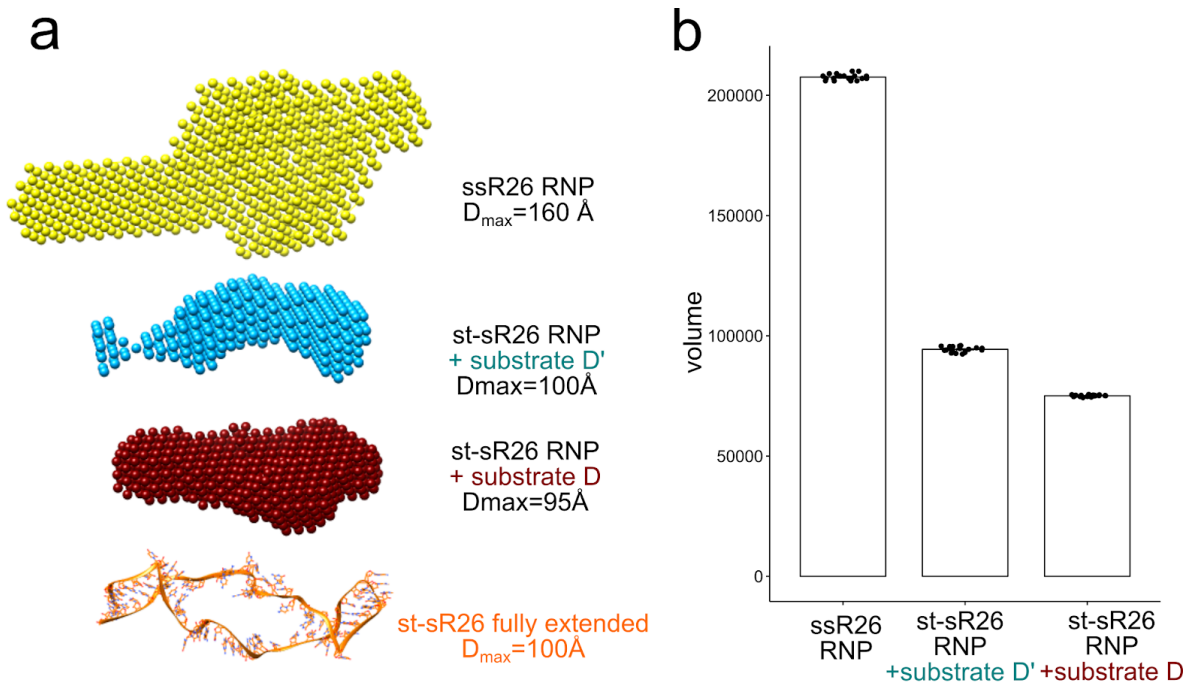


Figure 4.1.8. *Ab initio* RNA models from SANS of RNPs reconstituted with ^{2}H-sRNA in 42% D_2O. A)*Ab initio* dummy atom envelopes generated from the SANS data for ^{2}H-sRNA samples in 42% D_2O. At this D_2O concentration, all the ^{1}H-proteins are masked and the model is made up only by the contribution of the ^{2}H-sRNA. The density of the ssR26 RNP is large enough to accomodate two sRNA molecules, while the models of the sRNA within st-sR26 RNPs bound to substrate D or substrate D' can only accomodate a single molecule. B) The dummy atom volumes extracted from the *ab initio* models reflect the transition from 2 sRNA copies to one upon addition of substrate RNA for the st-sR26 RNP. The points represent the individual volumes of 20 independently generated envelopes.

In order to understand the principles underlying the different behaviours induced by the ssR26 and st-sR26 guide RNAs upon sRNP reconstitution, we designed two RNA chimeras,

st-sR26-1 and st-sR26-2 (Fig. 4.1.1), each of which shares half of the guide D sequence with the st-sR26. Reconstituting the RNP with these sequences leads to different behaviours both in the apo and substrate-loaded states. The st-sR26-1 RNP remains a di-RNP throughout the titration, while the st-sR26-2 RNP adopts a mono-RNP form upon substrate binding (see Fig.4.1.9). Thus, we conclude that the oligomeric state of the Box C/D sRNP is heavily influenced by the guide sequence of the RNA used in reconstitution.

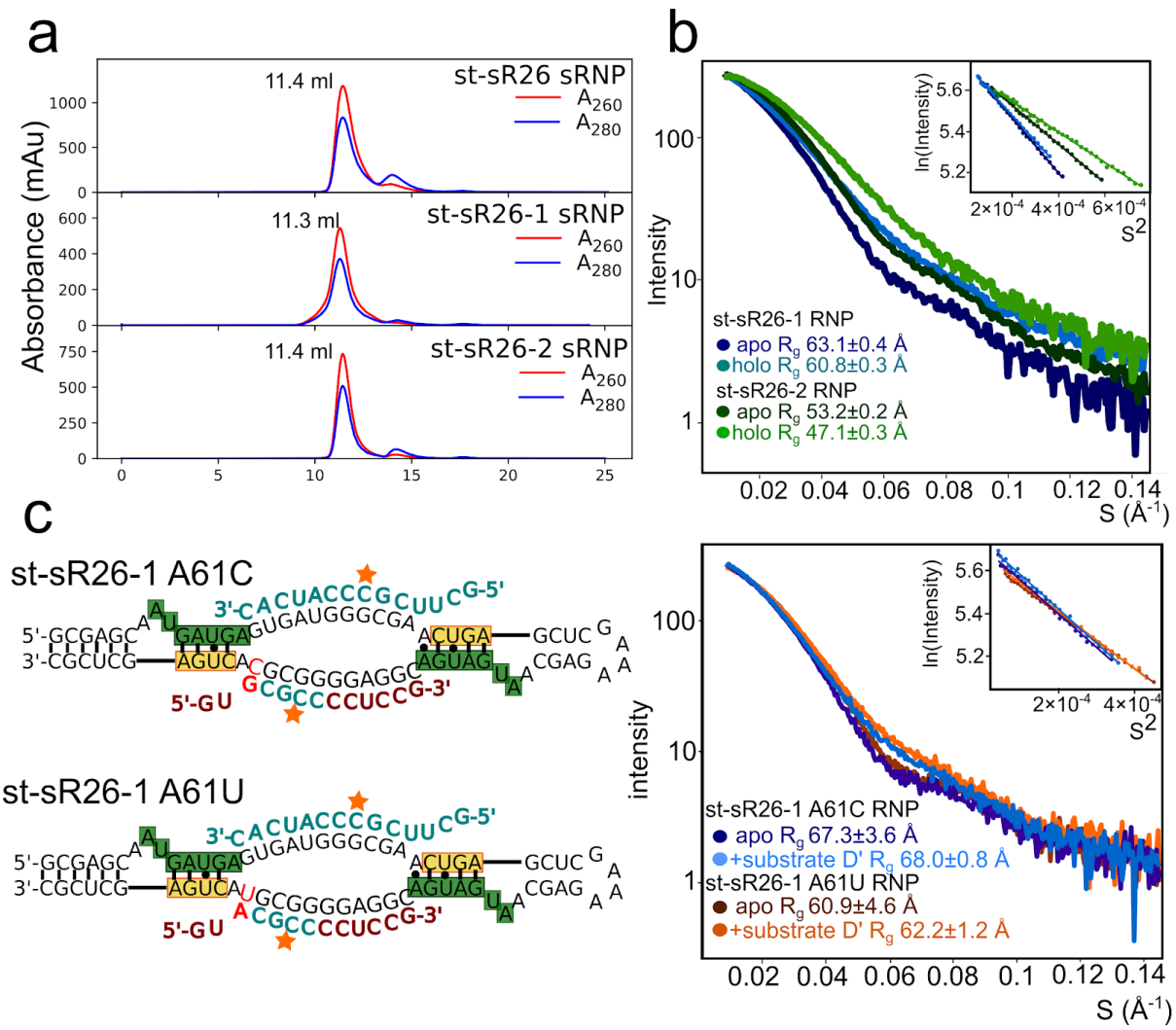


Figure 4.1.9. Reconstitution of complex assemblies using st-sR26-1 and st-R26-2 chimeric sRNAs shows difference in the oligomeric state of the holo complex. A) Size exclusion chromatography of complexes reconstituted with the st-sR26-1 and st-sR26-2 variants. Reconstitution carried out on the same set-up. B) SAXS profiles of Box C/D RNPs reconstituted with st-sR26-1 and st-sR26-2. The st-sR26-2 particle behaves similarly to the st-sR26 RNP, in that it shifts from a di-RNP particle to a mono-RNP. On the other hand, the st-sR26-1RNP retains a di-RNP architecture, similar to the ssR26 construct. The measurements are carried out at 50 °C on particles

at 0.5 mg/ml concentration. The holo complex is generated by addition of 1.25 molar equivalents of substrate D and substrate D' RNA. Curves scaled to match I(0) for comparison purposes. C) SAXS profiles of complexes reconstituted with variants of the st-sR26-1 RNA (mutations highlighted in red in the sequence). The complexes retain a di-RNP architecture upon addition of 1.1x molar equivalents of substrate D' RNA. Measurements carried out at 40 °C on complexes at 0.9 mg/ml concentration.

We then hypothesized that this difference in behaviour is due to the different nature of the base pairs at the ends of the substrate-guide duplex, as in other experiments we could show that the terminal base pair affects sRNP methylation activity (see section 4.3). In st-sR26-1 and ssR26, guide D ends with an adenosine (A61) base paired to a uracil on substrate D (see Fig. 4.1.9). We generated two additional constructs, st-sR26-1A61C and st-sR26-1A61U, where this nucleotide is mutated to cytosine or uracil. However, upon addition of substrate D', these complexes retain a di-RNP architecture, as was the case for st-sR26-1 (see Fig. 4.1.9)

In summary, the behaviour in solution and oligomeric state of the archaeal Box C/D complex strongly depends on the sRNA sequence it assembles on. In the apo state, different substrate RNAs lead to different mixtures of Box C/D particles, with the main population retaining a di-RNP architecture. Upon binding of substrate RNA, complexes reconstituted with different sRNAs adopt different oligomeric states, depending on the nature of the RNA guide regions. Thus, the sRNA does not act as a simple platform for the proteins to perform their function; rather, it plays an active role in determining the overall structure and assembly state of the RNP.

4.2 The Box C/D enzyme displays cooperative behaviour.

In order to characterise the methylation activity of the Box C/D sRNP in solution, we performed a series of methylation assays on complexes reconstituted with the wild-type sR26 guide RNA (See Fig.4.2.1-2). The assays monitor the transfer of a ^{3}H-labelled methyl group from the s-adenosyl methionine (SAM) cofactor to a biotinylated substrate RNA, which is then purified by affinity pull-down with streptavidin beads. In an initial set of experiments, we monitored the methylation of substrate D or substrate D' by the Box C/D complex in isolation. Under these conditions, we observed that substrate D' is reproducibly methylated more efficiently than substrate D.

In a second set of experiments, we observed that substrate D or D' methylation is stimulated by addition of the other substrate, indicating that the enzyme displays cooperative behaviour between methylation sites. Interestingly, the addition of 1 mM MgCl did not affect the preference for substrate D' methylation. This is in conflict with earlier reports, conducted with gel-based assays, which indicated that the presence of Mg^{2+} can affect the specificity of methylation of a sRNP from *M. Jannaschii* (Appel & Maxwell, 2007).

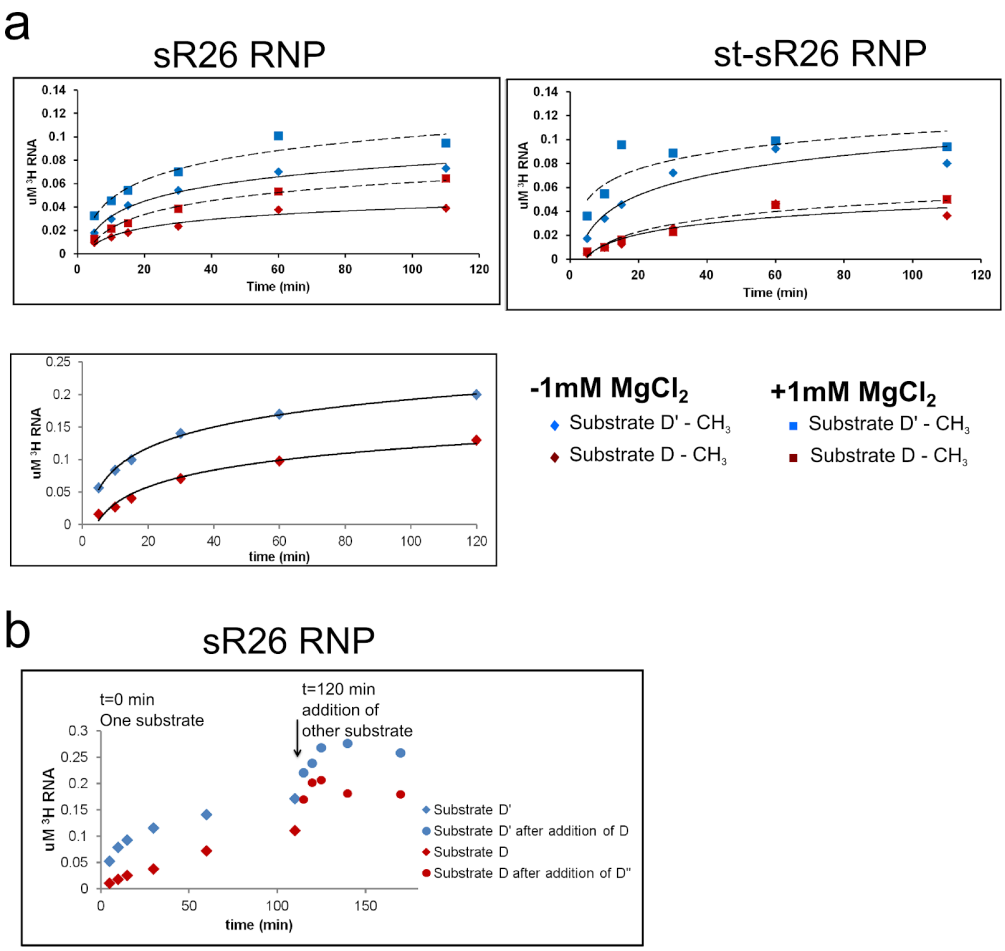


Figure 4.2.1. RNPs reconstituted with sR26 RNA and st-sR26 preferentially methylate substrate D' and show cooperative behaviour. A) Methylation of each substrate in isolation by sR26 and st-sR26 RNPs. The assays on the top left and top right, comparing sR26 and st-sR26 RNPs, were performed using the same batch of ^{3}H-SAM cofactor. The assays on the left were performed with the same experimental set-up, but two different batches of SAM cofactor. B) Stimulation of substrate methylation upon addition of the other substrate. After the methylation of a substrate RNA has reached a plateau, the other substrate is added. This results in a stimulation of methylation of the original substrate. Assays performed with 0.15 µM Box C/D, 1µ ^{3}-H SAM and 0.5 µM substrate RNAs at 50 °C. Continuous line: log fitting of -Mg^{2+} points; dashed line: log fitting of + Mg^{2+} points. Figure adapted from Graziadei *et al.,* 2016. Data produced jointly with Dr. Pawel Masiewicz.

In addition to the *in vitro* methylation assays, we performed an NMR-based activity assay, capable of monitoring the release of methylated product by the Box C/D sRNP. The assay was performed by adding SAM cofactor labelled with the NMR-active nucleus ^{13}C to unlabelled sR26 Box C/D complex in the presence of an excess of substrate RNA. Unlike the methyl groups of isoleucine, leucine and valines, we find that the 2'-O CH_3 resonances are too broad in this molecular weight range to be observed when bound to the Box C/D complex. The RNA methyl peaks observed in these spectra therefore arise from the methyl group bound to the released enzymatic product.

Here, we observed the fact that substrate D can be methylated and released by the Box C/D complex only after addition of substrate D'. While no resonances (i.e. released methylated RNA) are observed upon addition of substrate D alone, adding substrate D' leads to the appearance of both sets of resonances. Biochemical assays, which measure the total amount of product, showed that the enzyme is capable of methylating substrate D in isolation. However, we do not find a significant signal for released methylated substrate D in the NMR-based assay, indicating that the majority of methylated substrate D RNA is still bound to the enzyme.

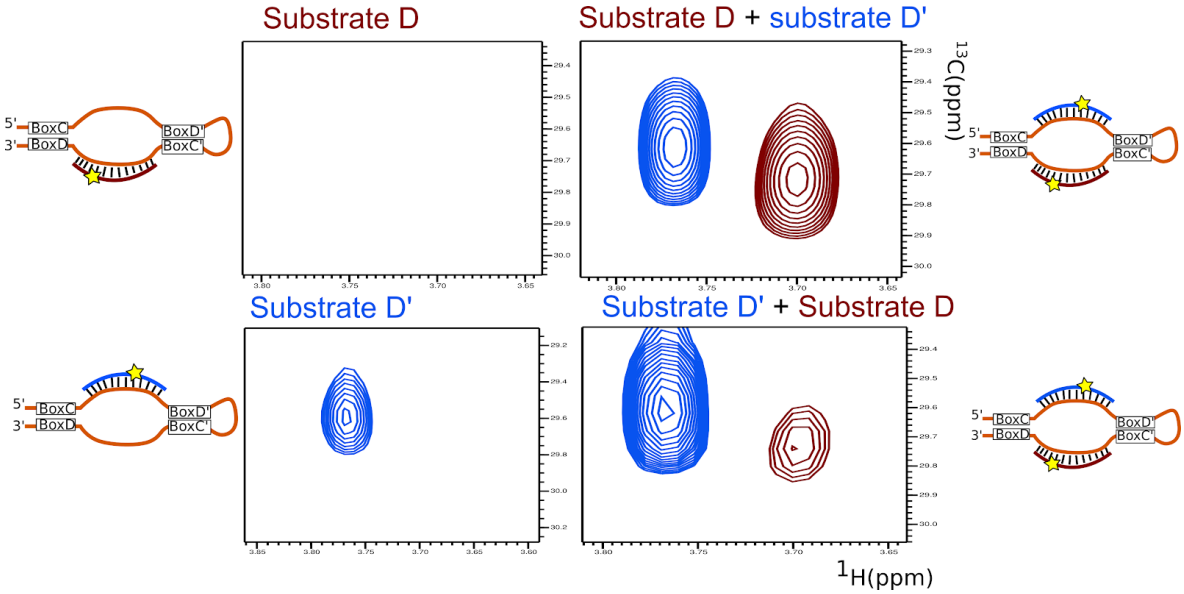


Figure 4.2.2. Substrate RNA 2'-O methylation monitored by NMR spectroscopy. ^{1}H-^{13}C HMQC spectra of sR26-reconstituted Box C/D RNPs in the presence of ^{13}C-labelled SAM and substrate RNAs. Top panels: addition of substrate D followed by addition of substrate D'. Addition of substrate D' stimulates methylation and release of substrate D. Bottom panels: addition of substrate D' followed by substrate D, showing substrate D' is efficiently methylated and released by the enzyme in isolation, although addition of substrate D stimulates methylation. Upper panels: 12.5 μM Box C/D

enzyme, 75 µM substrate RNAs, 150 µM ^{13}C-SAM. Lower panels: 14.5 µM Box C/D enzyme, 87 µM substrate RNAs, 174 µM ^{13}C-SAM. Figure adapted from Graziadei *et al.*, 2016.

We then monitored the influence of the oligomeric state on the cooperative behaviour of the enzyme. We used Box C/D complexes reconstituted with st-sR26-1 and st-sR26-2 RNAs in biochemical assays where the methylation of substrate D or D' is measured in isolation or in the presence of the other substrate (Fig.4.2.3). The two substrates stimulate the methylation of one another in the case of RNPs reconstituted with st-sR26 and st-sR26-2.

As shown in section 4.1, these RNPs display a transition from a di-RNP architecture to mono-RNP architecture upon addition of substrate RNA. However, in the case of Box C/D RNPs reconstituted with st-sR26-1 RNA, addition of substrate D' reduces the methylation of substrate D and vice versa.
Thus, while the half-loaded RNPs generated in each case retain the more efficient methylation of substrate D' as compared to D, the cooperative behaviour of the enzyme is different in mono- and di-RNP states, in a manner that may be correlated to the dynamics of the Fib methyltransferase within the complex (see discussion).

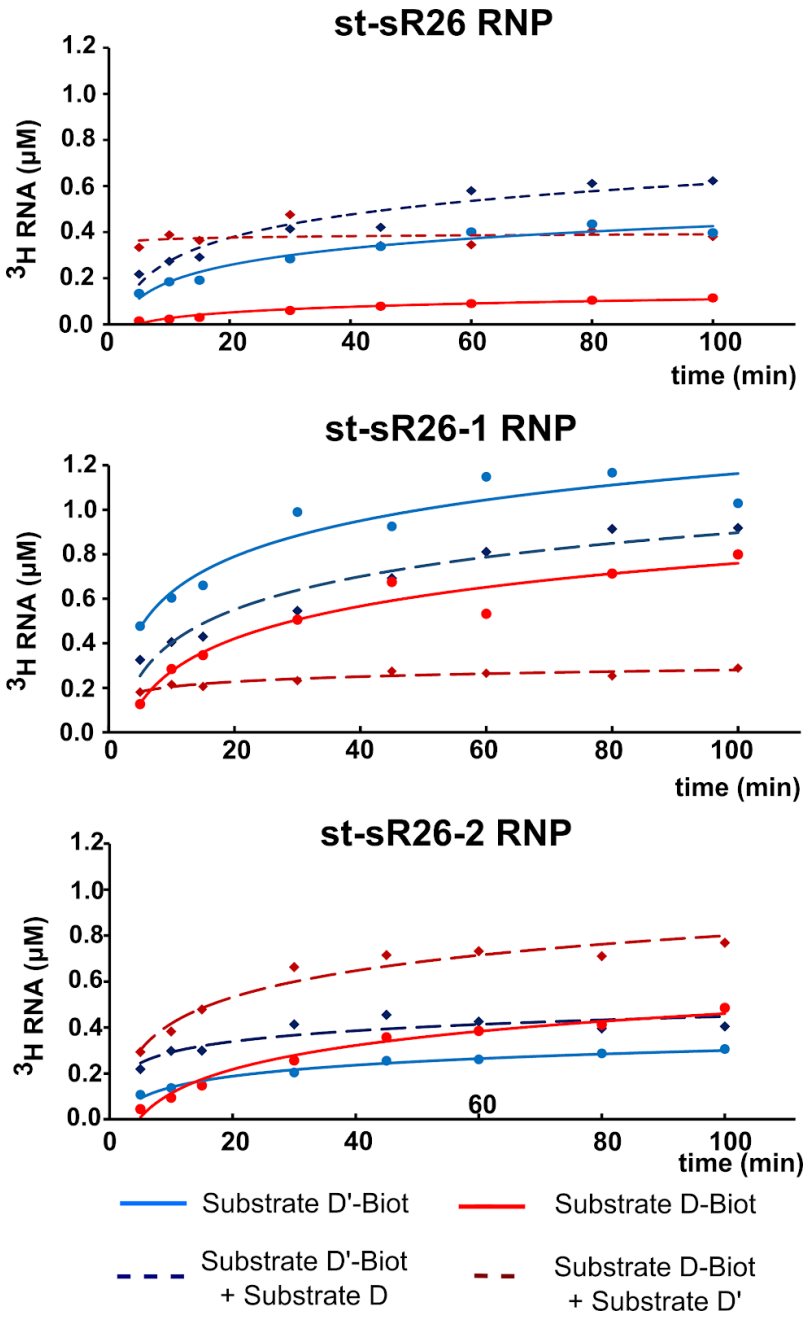


Figure 4.2.3. Activity profiles of st-sR26 variants. *In vitro* methylation assays performed with Box C/D sRNPs assembled with st-sR26 (top), st-sR26-1 (middle) and st-sR26-2 (bottom) guide-RNAs. 0.75 µM of each substrate was added to 0.15 µM enzyme in the presence of 2.25 µM [^{3}H]-SAM at 50 °C. For both enzymes transitioning to a mono-RNP(st-sR26 and st-sR26-2), the methylation at each site is stimulated by addition of the other substrate (continuous vs dashed lines). For the di-RNP enzyme (st-sR26-1), the methylation at each site is diminished upon addition of the second substrate. Data produced jointly with Simone Hoefler.

4.3 Box C/D activity depends on the nature of the sRNA guide regions.

Further biochemical assays highlight one of the principles behind the more efficient methylation of sR26 substrate D'. Despite the fact that the two guide sequences have identical G:C content, they differ in sequence. In a set of experiments designed to probe the role of the guide sequence in determining methylation efficiency, we swapped the terminal A:U base pair in substrate D' with the preceding G:C base pair (see fig. 4.3.1).

The resulting complex displays a highly reduced methylation efficiency of substrate D'. This is consistent with a "zipper" product release mechanism that involves melting the substrate-guide duplex from the ends, in a consistent with the hypothesis that methylated product release relies on melting the substrate-guide duplexes from the ends. In this case, a G:C base pair would stabilise the bound form of the product and inhibit enzymatic turnover.

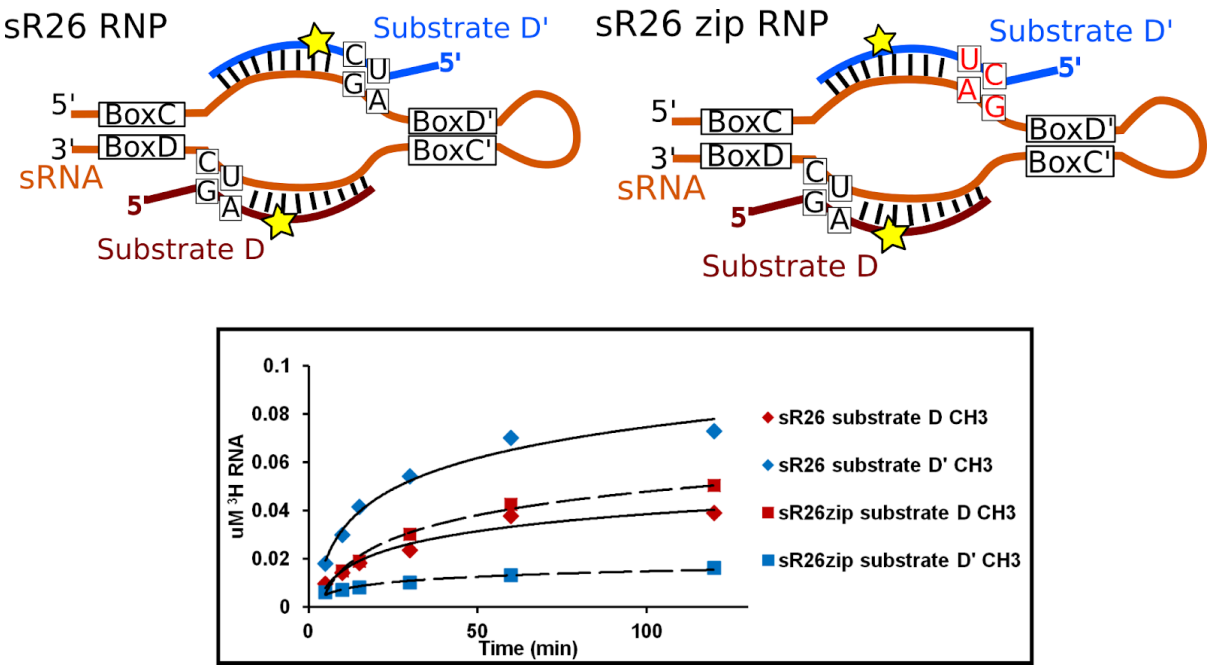


Figure 4.3.1. Box C/D activity depends on substrate-guide duplex ends. Swapping the order of the nucleotides at the 5' end of the substrate D' base pairs knocks down methylation of substrate D'. Assays performed with 0.15 µM Box C/D, 1µ ^{3}-H SAM and 0.5 µM substrate RNAs at 50 °C. Data produced jointly with Dr. Pawel Masiewicz. Figure adapted from Graziadei *et al.*, 2016.

In a second set of experiments, we inverted the relative position of guide D and D' within the sR26 guide RNA. Complexes reconstituted using this inverted sRNA methylation of substrate D' was still carried out at a higher efficiency (see Fig. 4.6.2). In fact, inversion of

the guide sequences resulted in an even more marked knock-down of methylation of substrate D.

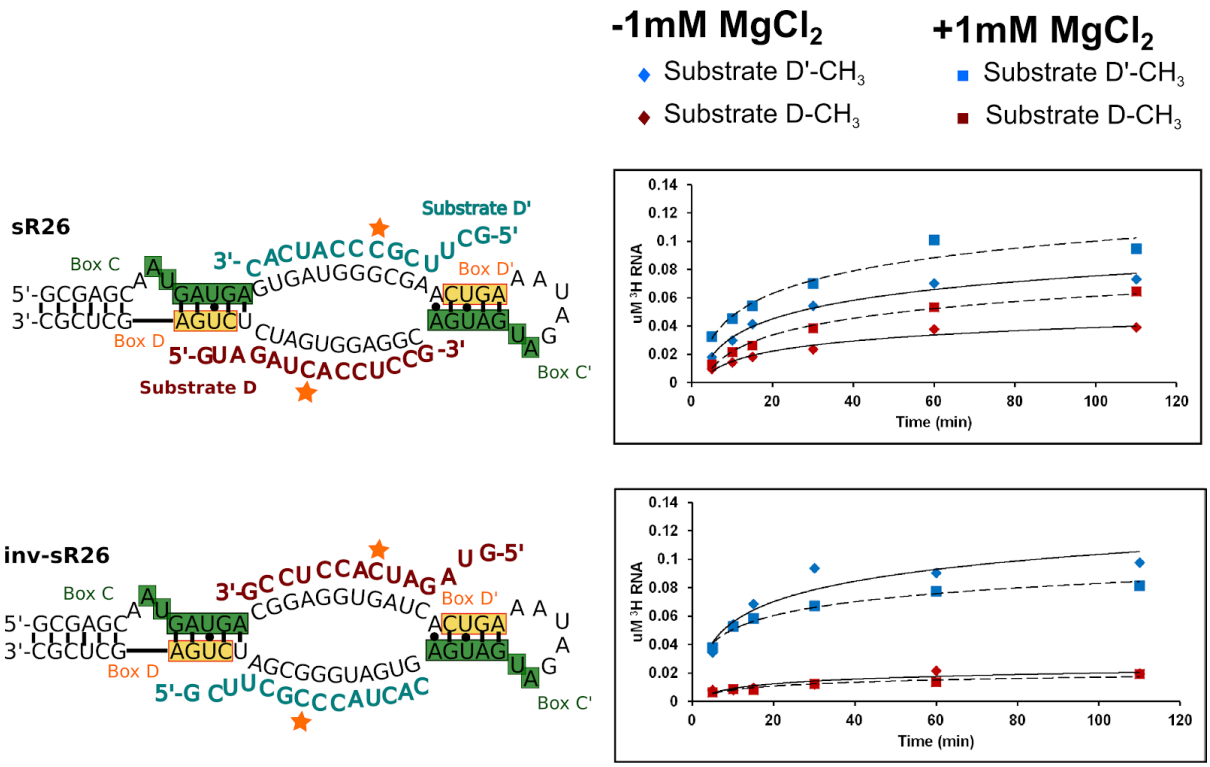


Figure 4.3.2. Substrate preference is independent of relative position. Methylation assays performed with RNPs reconstituted with sR26 and inv-sR26 show that inverting the guide position does not result in a swap of methylation efficiency, as substrate D' methylation remains more efficient. However, we could observe a further inhibition of substrate D methylation. Assays performed with 0.15 µM Box C/D, 1µ ^{3}-H SAM and 0.5 µM substrate RNAs at 50 °C. Data produced jointly with Dr. Pawel Masiewicz. Figure adapted from Graziadei *et al.*, 2015.

An additional set of biochemical experiments was performed to monitor the effect of reconstitution in 150 mM NaCl on the methylation preference of the sRNP (see Fig. 4.3.3). As the experiments show, reconstitution in 150 mM NaCl bears no effect on the preference of methylation, despite it resulting in the assembly of higher-order oligomeric states of the RNP particle (see Fig. 4.1.3).

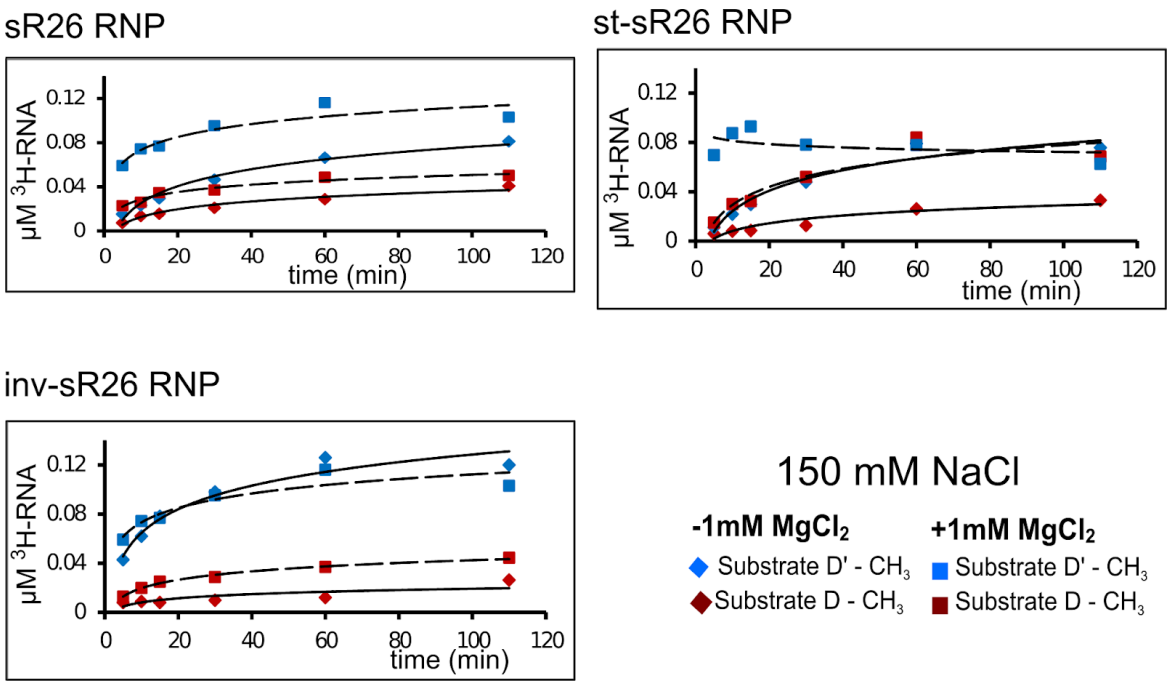


Figure 4.3.3. Substrate D' preference is retained in low-salt conditions. Assays performed with 0.15 μM Box C/D, 1μ ^{3}-H SAM and 0.5 μM substrate RNAs at 50 °C in 20 mM HEPES, 150 mM NaCl, pH 7.5. Data produced jointly with Dr. Pawel Masiewicz. Figure adapted from Graziadei *et al.*, 2016.

Thus, we conclude that the higher methylation efficiency of substrate D' by Box C/D sRNPs reconstituted with variants of the sR26 sRNA depends on the specific sequence of the substrate-guide duplex, rather than its relative position with respect to the conserved structural elements.

4.4 Fib exists in a conformational equilibrium between "off" and "on" states.

Binding of substrate RNA to st-sR26 Box C/D not only inducthe transition from a di-RNP to a mono-RNP, but also a conformational change that places one Fib copy onto the substrate-guide duplex, where it may methylate the substrate RNA. The two resulting conformational state, which we term [on,off]- and [off,off]-states, are depicted in fig. 4.4.1.

We observed this transition by NMR experiments, were we see a few Fib methyl resonances adopting an elongated shape upon substrate addition(Fig. 4.4.2a). indicative of an exchange between conformations of a fast NMR time scale upon binding of substrate RNA. In the holo di-RNP structure, where two Fib copies are stably bound to the RNA, we see that

these same resonances split into two distinct set of peaks, corresponding to the two Fib copies bound to the RNA and the two Fib copies away from the substrate-guide duplex (see Fig.4.4.2). In the holo di-RNP, where two of the four copies of Fib are stably bound to the substrate-guide duplex, the Fib methyl resonances split into two distinct peaks (see Fig.4.4.2, 1.2.6). Thus, we can use the frequency distance between the peaks corresponding to the two Fib states found in the holo di-RNP to quantify the lower limit of the exchange rate between the two states in the mono-RNP, where these two peaks are not resolved. In the holo di-RNP, the frequency distance of the two states varies between 20 and 60 Hz depending on the peak. As the same set of resonances in the substrate-bound st-sR26 RNP exchange faster than the frequency separating the two states in the di-RNP spectrum, we estimate the rate of exchange to be faster than 20 Hz.

As Fig. 4.4.3 shows, the Fib methyl resonances displaying chemical shift perturbation in the presence of RNA are located on the RNA binding interface of Fib.

Next, we measured SAS curves of two RNP forms, with either substrate D or substrate D' bound to the corresponding guide sequence. The R_g measured in SAXS upon substrate RNA binding (+substrate D'= 47.3 Å, +substrate D=50.0 Å) is too large for a mono-RNP structure where both Fib proteins are on the substrate-guide duplex (reported as 40.0 Å), suggesting that at most one Fib copy is in contact with the RNA. This is consistent with the fact that only one of the two guide sequences is base paired with the substrate in each complex.

SANS curves highlighting the Fib methyltransferase show that the experimental P(r) distribution upon binding of either substrate D or substrate D' lies in between that of a mono-RNP with two Fib copies away from the RNA ([off,off]-state) and one with one of the Fib copies placed on the substrate-guide duplex([on,off-state]) (see Fig. 4.4.2).

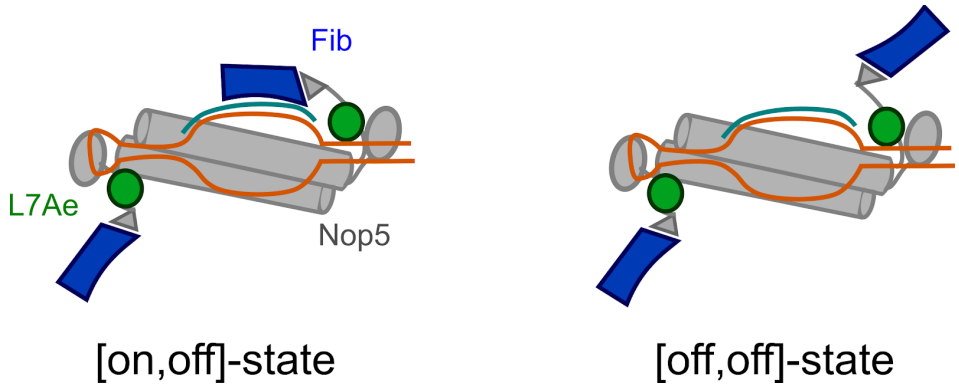


Figure 4.4.1 half-bound mono-RNP arrangements. Cartoon representation of the [on,off]- and [off,off]-states of the RNP bound to either substrate D or substrate D'. The guide sRNA is shown in orange and substrate in cyan.

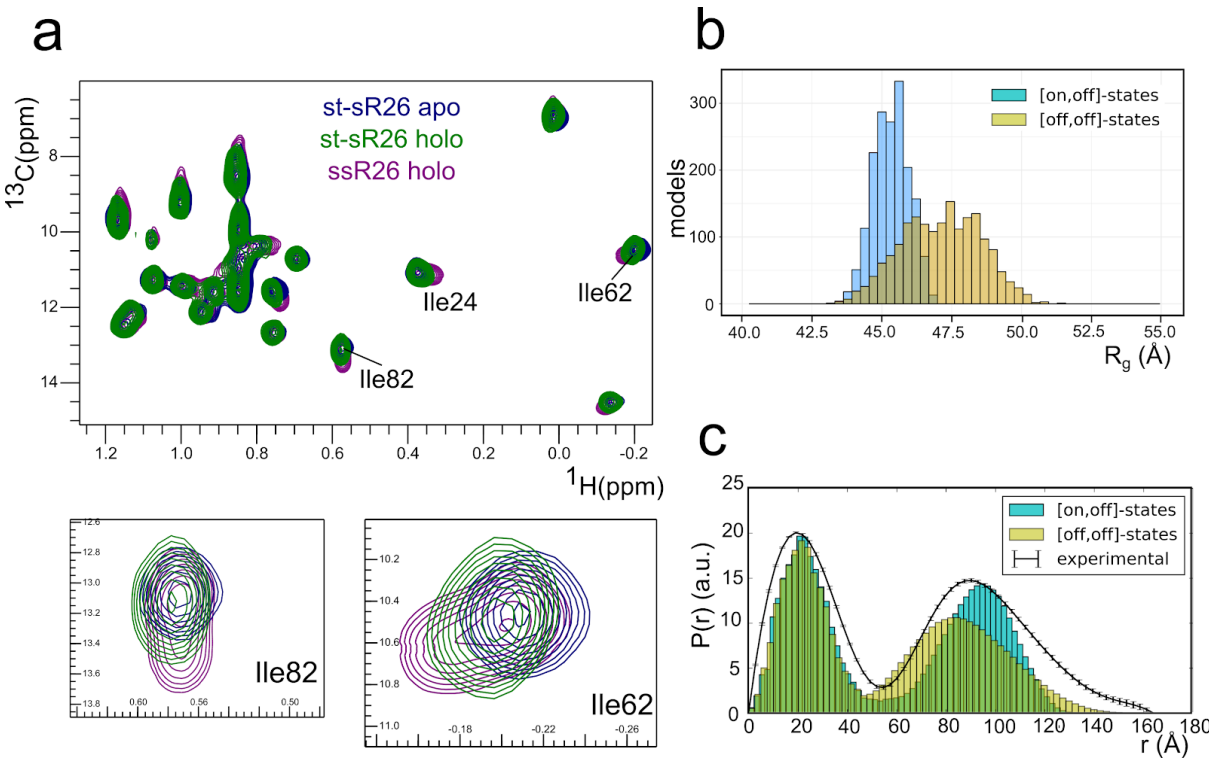


Figure 4.4.2. Fib shuttles between on- and off-states in the presence of substrate RNA. A) Addition of substrate RNAs onto RNPs reconstituted with st-sR26 induces elongation in Fib methyl resonances, indicating the presence of a exchange process that is fast on the NMR time scale. On the other hand, RNPs reconstituted with ssR26 display splitting of Fib methyl resonances into two states. B) Comparison of the theoretical SAXS R_g of models with one Fib copy on the substrate-guide duplex, and models with both Fib copies away from the substrate-guide duplex. The crystal structure of a mono-RNP with both Fib copies bound to RNA (PDB accession code 3pla) has a theoretical R_g of

40.4 Å. The experimental SAXS R_g of substrate D'-bound st-sR26 Box C/D is 47.3 Å, while that of substrate D-bound st-sR26 Box C/D is 50.0 Å. C) Theoretical P(r) function for ^{2}H-Fib 42% D_2O SANS profiles, which highlight that the observed average distance between Fib subunits can be generated by an equilibrium between conformers with both Fib copies away from the RNA and conformers with one Fib copy placed on the substrate-guide duplex. The distribution is calculated on the pool of conformers generated by randomising Nop5-NTD—Fib position after Cartesian refinement (see section 3.7, 3.8). This corresponds to a sampling around the sRNP structure best fitting the experimental restraints for the substrate D'-bound RNP, and is therefore narrower than the R_g distributions in Fig. 4.1.6.

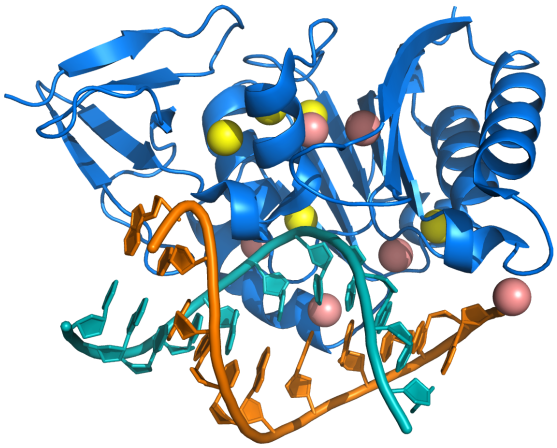


Figure 4.4.3. Fib methyl peaks affected by the presence of RNA mapped onto the model of Fib bound to the RNA duplex. Peaks displaying elongated or shifted behaviour upon addition of 3x molar equivalents of substrate D and substrate D' onto 23 μM st-sR26 Box C/D sRNP. Isoleucine Cδ1 atoms are shown as yellow spheres, while leucine Cδ1 and valine Cγ1 are displayed in pink. The guide sRNA is shown in orange. Substrate D' is shown in light blue. The residues highlighted are Isoleucines 24, 62, 82, 71, 159; Leucines 58, 114, 221; Valines 110, 151, 185.

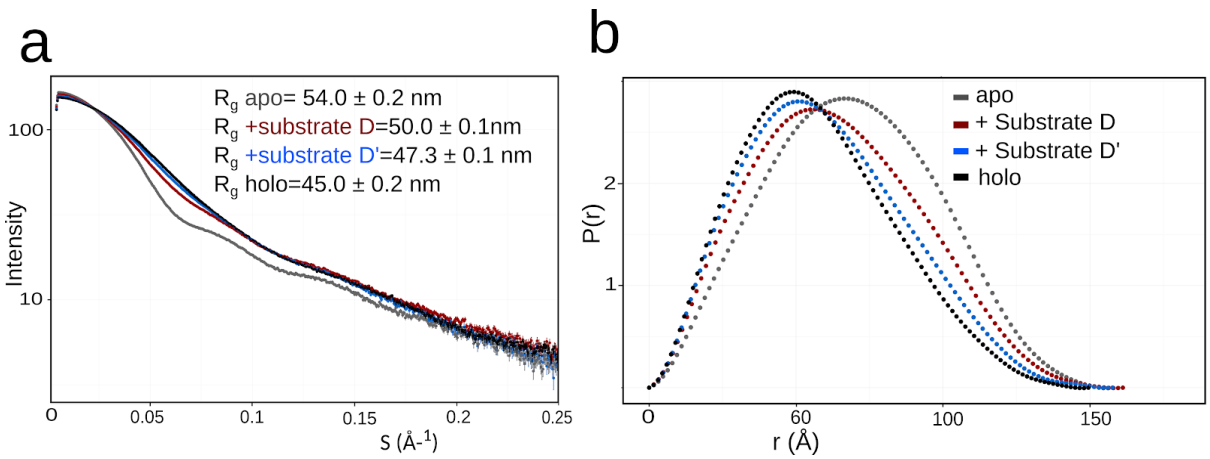


Figure 4.4.4. Addition of different substrate RNAs induces different RNP states. Titration of substrate RNAs on top of Box C/D RNPs reconstituted with st-sR26 RNAs at 5 mg/ml concentration. B. The pairwise-distance distribution function P(r) extracted from the same curves. The P(r) functions show a marked difference between the enzymes bound to substrate D or D', with the substrate D'-bound enzyme resembling the holo complex. Curves scaled to match I(0) for comparison purposes. Substrate RNAs added to 1.25x molar ratio. Experiments carried out at 40 °C. Figure adapted from (Graziadei *et al*, 2016).

In NMR experiments, a set of PREs is observed that is uniquely compatible with at least one of the two Fib copies being placed on the substrate-guide duplex upon binding of substrate RNA. These PREs, arising from a spin label placed in the Nop5 NTD (Nop5 E65 to methyl groups located on L7Ae, indicate the presence of Nop5NTD-L7Ae contacts, which can only be satisfied by placing Fib onto the substrate RNA. Interestingly, the PRE effects observed are more intense in substrate D'-bound samples than in substrate D-bound sample, which led us to hypothesize that a higher proportion of Fib is near the substrate-guide duplex upon binding of substrate D' than upon binding of substrate D.

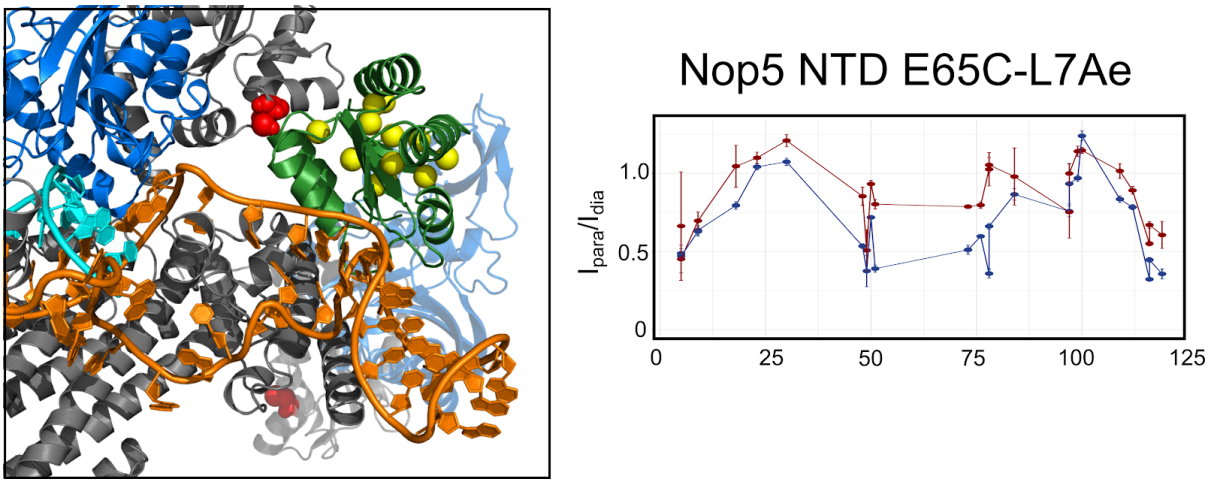


Figure 4.4.5. Distances corresponding to the on-state in substrate-bound RNPs. Right: The relative position of the Nop5-NTD—Fib module in the on- (solid blue & grey ribbons) and off-state (transparent ribbons). Upon substrate D' binding, the Nop5-E65C spin-label (red) induces PRE intensity-ratios <0.75 for the L7Ae-ILV residues shown as yellow spheres. Left: PRE intensity ratios (I_{para}/I_{dia}) induced by the binding of substrate D' (blue) or substrate D (red). The difference in these effects, uniquely compatible with the Nop5-NTD—Fib being near the substrate-guide duplex, is consistent with the presence of a higher proportion of Fib in the on-state upon binding of substrate D' than of binding of substrate D.

We conclude that addition of substrate RNA to the st-sR26 containing BoxC/D RNP induces a dynamic behaviour of the Fib methyltransferase, which shuttles between off and on position with an apparent frequency higher than ~20 Hz. Furthermore, this data is consistent with a higher proportion of Fib in the on-state in the substrate D'-bound RNP, compared to the substrate D-bound RNP.

4.5 Stereospecific assignment of Fib methyl groups in the free protein and the full RNP.

Fib methyl groups were assigned in the free state by combining the information from 3D ^{1}H-^{13}C NOESY-HMQC experiments acquired on ILV and ILV^{proS}-methyl labelled samples with ^{1}H-^{13}C TOCSY-HMQC experiments on ILV-labelled samples, ^{1}H-^{13}C HMQC spectra on ILV, ILV^{proS} and L-methyl labelled spectra and NOEs expected from the crystal structure of free *P. furiosus* Fib (see section 3.5).

The assignment of Fib methyl resonances in the free state, displayed in Fig. 4.5.1, was transferred to the Nop5-NTD—Fib subcomplex (Fig. 4.5.2), the Nop5-Fib subcomplex and finally the full Box C/D RNP (Fig.4.5.3) by combining ^{1}H-^{13}C HMQC spectra of ILV, ILV^{proS}, L-methyl labelled samples (see section 3.6).
The labelling schemes used in this study enabled us to obtain stereospecific assignment of leucine and valine methyl groups.

We followed the formation of the Nop5-NTD—Fib subcomplex by chemical shift perturbation (see Fig.4.5.4 a-b), which probes the changes in chemical environment around the methyl group, and verified that the interface formed in solution is consistent with previous crystal structures for this interaction. Furthermore, we confirmed that the interfaces formed in the Nop5-NTD—Fib subcomplex are retained in the Nop5-Fib subcomplex and in the full sRNP (see Fig. 4.5.4c-d). This justifies our use of the Nop5/Fib interfaces present in previous crystal structures in structure calculations of the full Box C/D sRNP.

We were able to leverage the simplified leucine-valine region present in LV^{proS}- or Leu-methyl labelled sample to aid interpretation of the 3D NOESY-^{13}C HMQC spectrum of the free protein and to transfer assignments to Nop5-Fib (~110 KDa) and the full Box C/D sRNP (in the apo state, ~400 KDa) (see Fig. 4.5.3). Assignments are also reported as a table in Appendix 2.

The strategies developed here enabled us to verify and correct the assignments used in our previous study, as we could follow methyl group resonances from the free protein to the full complex with a higher confidence than what was previously achieved, thereby enabling a more accurate analysis of paramagnetic relaxation enhancement effects in the full Box C/D sRNP.

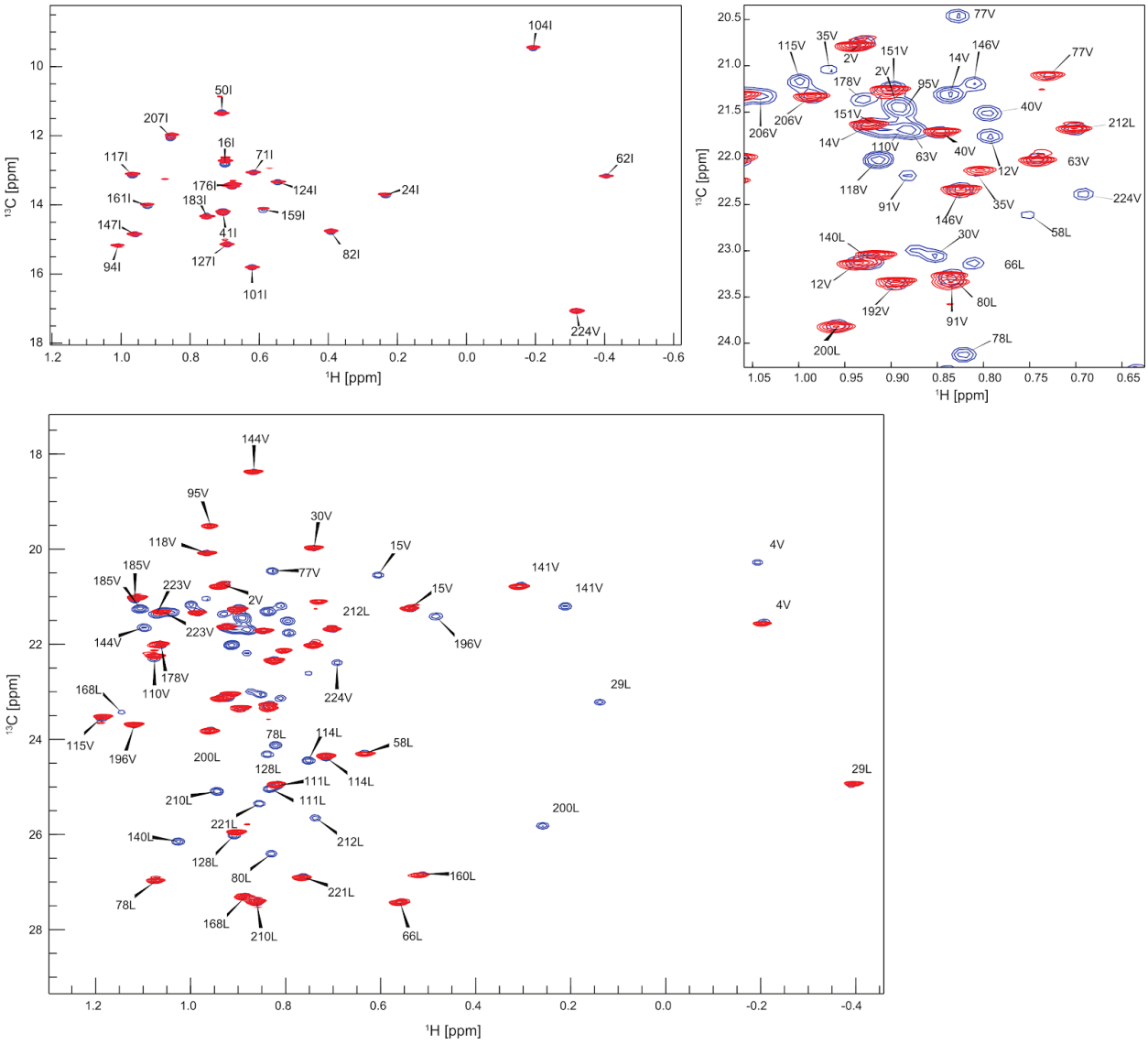


Figure 4.5.1 Stereospecific assignment of ILV-methyl resonances in free Fib. Overlay of ^{1}H-^{13}C HMQC spectra acquired on ILV-methyl (blue) and ILV^{proS}-methyl labelled Fib in isolation. LV resonances present in the red spectrum correspond to the proS methyl group of that specific amino acid. Assignment derived as described in methods. Top left panel panel: Isoleucine region; Top right panel: leucine/valine region, zoom on crowded region; Bottom panel: leucine/valine region, full view.

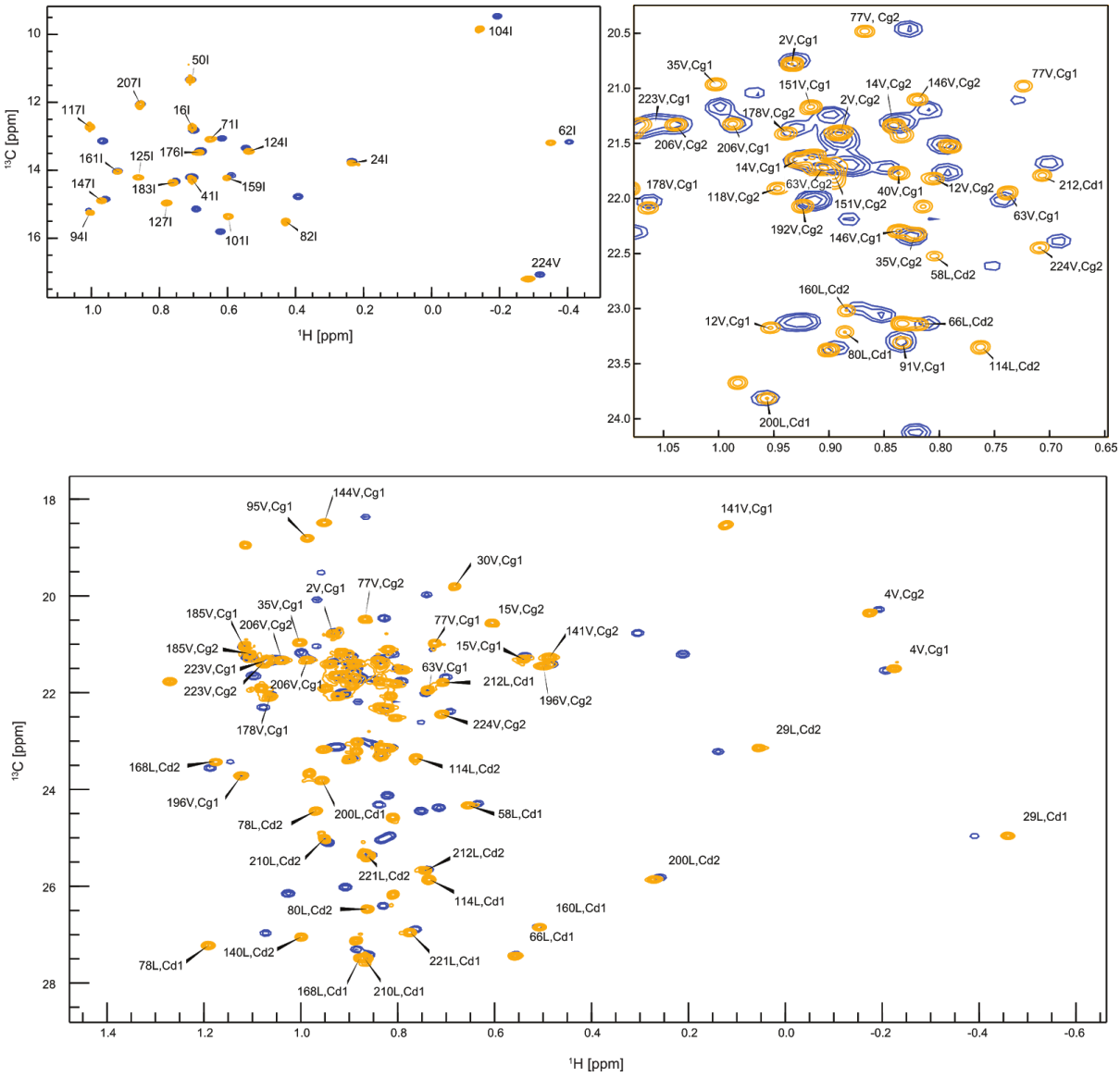
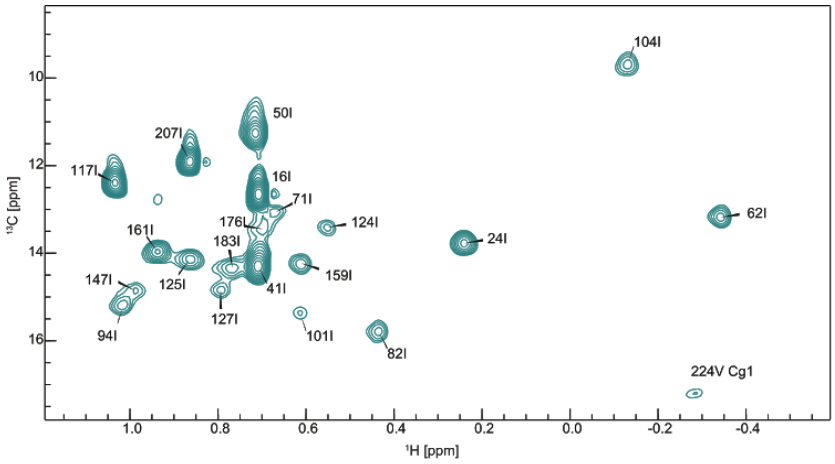
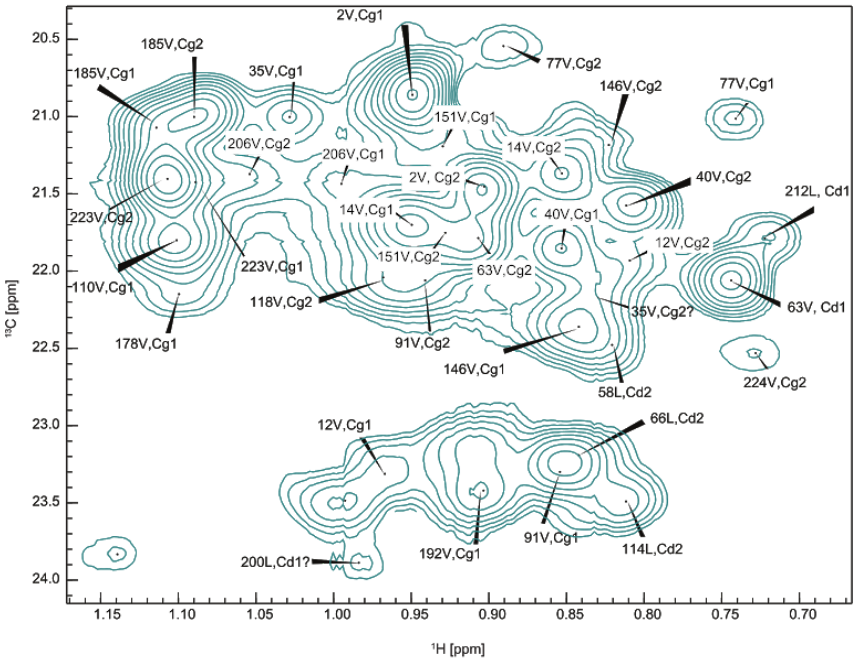


Figure 4.5.2 Transfer of methyl group assignments from free Fib to Nop5-NTD−Fib subcomplex. Overlay of ^{1}H-^{13}C HMQC spectra acquired on ILV-methyl labelled Fib in isolation (blue) and in complex with the N-terminal domain of Nop5 (orange). Top left panel panel: Isoleucine region; Top right panel: leucine/valine region, zoom on crowded region; Bottom panel: leucine/valine region, full view.


104I
50I
207I
117L
16I
71I
176L
124I
62I
161I
183I
24I
147I
125I
159I
41I
127I
94I
101I
82I
224V Cg1
^{13}C [ppm]
^{1}H [ppm]


2V,Cg1
185V,Cg2
185V,Cg1
35V,Cg1
77V,Cg2
146V,Cg2
77V,Cg1
151V,Cg1
206V,Cg2
206V,Cg1
14V,Cg2
2V, Cg2
40V,Cg2
212L, Cd1
223V,Cg2
14V,Cg1
40V,Cg1
12V,Cg2
151V,Cg2
223V,Cg1
63V,Cg2
110V,Cg1
118V,Cg2
35V,Cg2?
63V, Cd1
178V,Cg1
91V,Cg2
146V,Cg1
58L,Cd2
224V,Cg2
12V,Cg1
66L,Cd2
91V,Cg1
200L,Cd1?
192V,Cg1
114L,Cd2
^{13}C [ppm]
^{1}H [ppm]

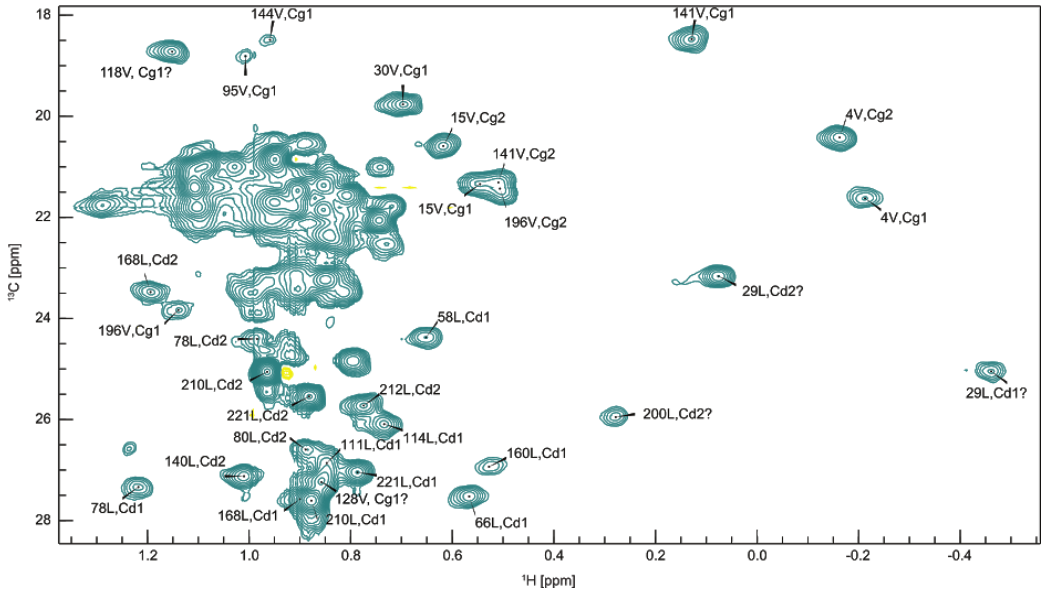

144V,Cg1
141V,Cg1
118V, Cg1?
95V,Cg1
30V,Cg1
15V,Cg2
4V,Cg2
141V,Cg2
15V,Cg1
196V,Cg2
4V,Cg1
168L,Cd2
29L,Cd2?
196V,Cg1
58L,Cd1
78L,Cd2
210L,Cd2
212L,Cd2
29L,Cd1?
221L,Cd2
200L,Cd2?
80L,Cd2
111L,Cd1
114L,Cd1
160L,Cd1
140L,Cd2
221L,Cd1
78L,Cd1
128V, Cg1?
168L,Cd1
210L,Cd1
66L,Cd1
^{13}C [ppm]
^{1}H [ppm]

Figure 4.5.3 Assignment of the apo sRNP. ^{1}H-^{13}C HMQC spectra acquired on ILV-methyl labelled Fib, 2H-Nop5 in the ssR26 apo sRNP. Stereospecific assignment of ILV methyl group resonances of Fib in the apo sRNP. As the spectra of ILV-methyl Fib inthe ssR26 RNP and the st-sR26 RNP match, the assignment is used for samples reconstituted with both RNAs. Peaks marked with the question marks are tentative assignments and were not used for extraction of distances in PRE experiments.

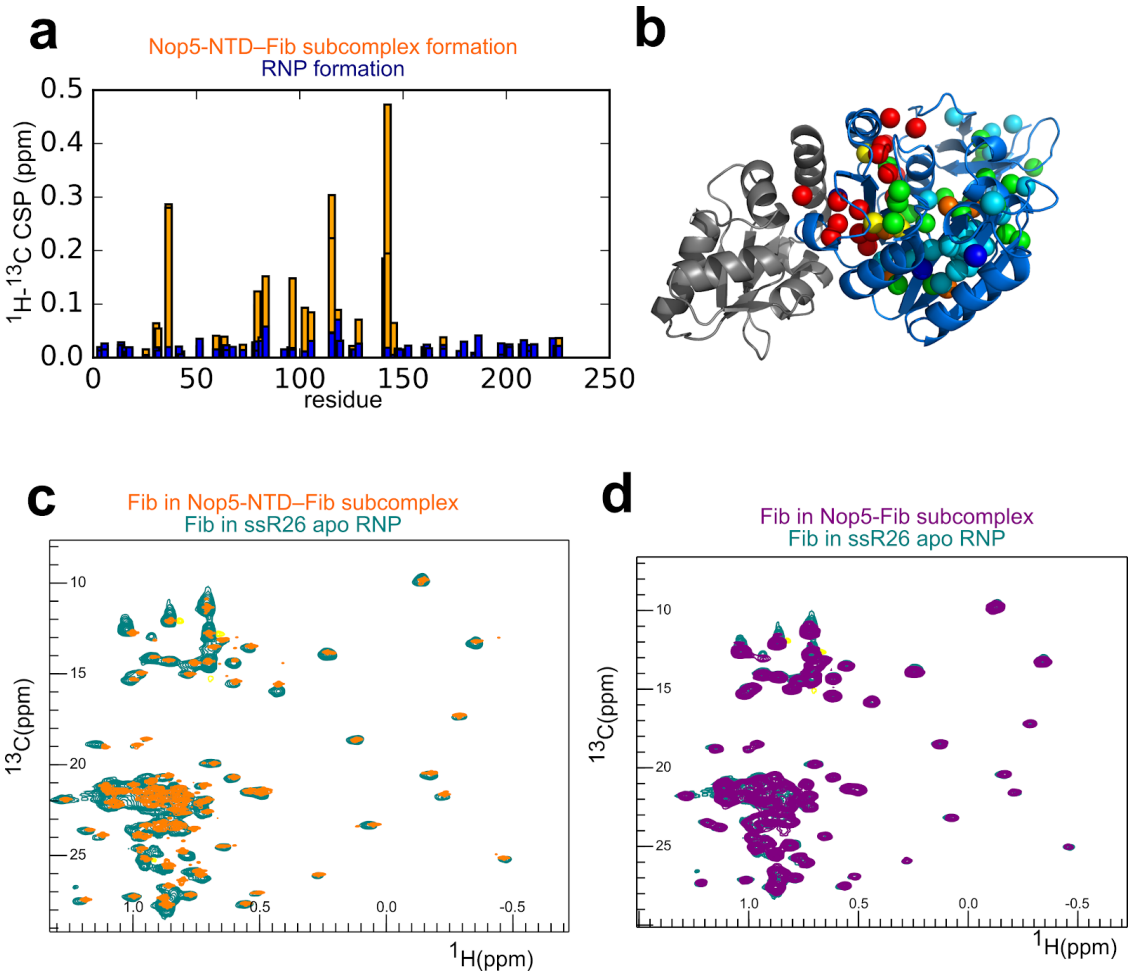


Figure 4.5.4. The Nop5-NTD−Fib module. A) Analysis of the chemical shift perturbation in the ILV-methyl resonances of Fib upon binding of Nop5-NTD (orange) and going from the Nop5-NTD−Fib subcomplex to the full sRNP (blue). While severe chemical shift perturbation is observed upon binding of the Nop5 NTD, the Fib resonances are largely unaffected in the transition from the Nop5-NTD−Fib subcomplex to the full RNP. Chemical shift perturbation computed using equation (11) in section 3.5. B) The chemical shift perturbation (CSP) values for Nop5-NTD−Fib formation mapped onto the corresponding methyl groups on the structure of the Nop5-NTD−Fib subcomplex (PDB accession code 3NMU, Nop5 NTD in gray). Color scale from blue (CSP=0 ppm) to red (CSP > 0.05ppm) and Fib in deep blue. C) Overlay ^{1}H-^{13}C HMQC spectra acquired on ILV-methyl labelled Fib in the Nop5-NTD−Fib subcomplex (orange) and in the full RNP (teal). D) Overlay of ^{1}H-^{13}C HMQC spectra acquired ILV-methyl labelled Fib in the Nop5-Fib subcomplex (purple) and in the full RNP (teal)

showing that Fib retains the same conformation and chemical environment between the Nop5-Fib subcomplex and the full sRNP.

4.6 Distance restraints and small-angle scattering characterisation of substrate D and substrate D'-bound sRNPs.

In order to solve the structures of the substrate D and substrate D'-bound RNPs, we set out to acquire a combination of distance restraints and low-resolution structural information to define the structure and dynamics of the RNP in solution. PRE experiments were acquired and quantified on substrate D and substrate D'-bound RNPs reconstituted with st-sR26 RNA as described in sections 3.2 and 3.6.

Distance restraints were defined from several PRE datasets probing the relative orientations of the Nop5-NTD—Fib, L7Ae/K-turn, Nop5 coiled-coil and substrate-guide duplex modules in the context of the [on,off]- and [off,off]-state in both the substrate D'- and the substrate D bound RNPs. Each PRE experiment was repeated for the substrate D-bound and the substrate D'-bound RNP. The positions of the spin labels within Nop5, L7Ae and Fib are shown in Fig. 4.6.1.

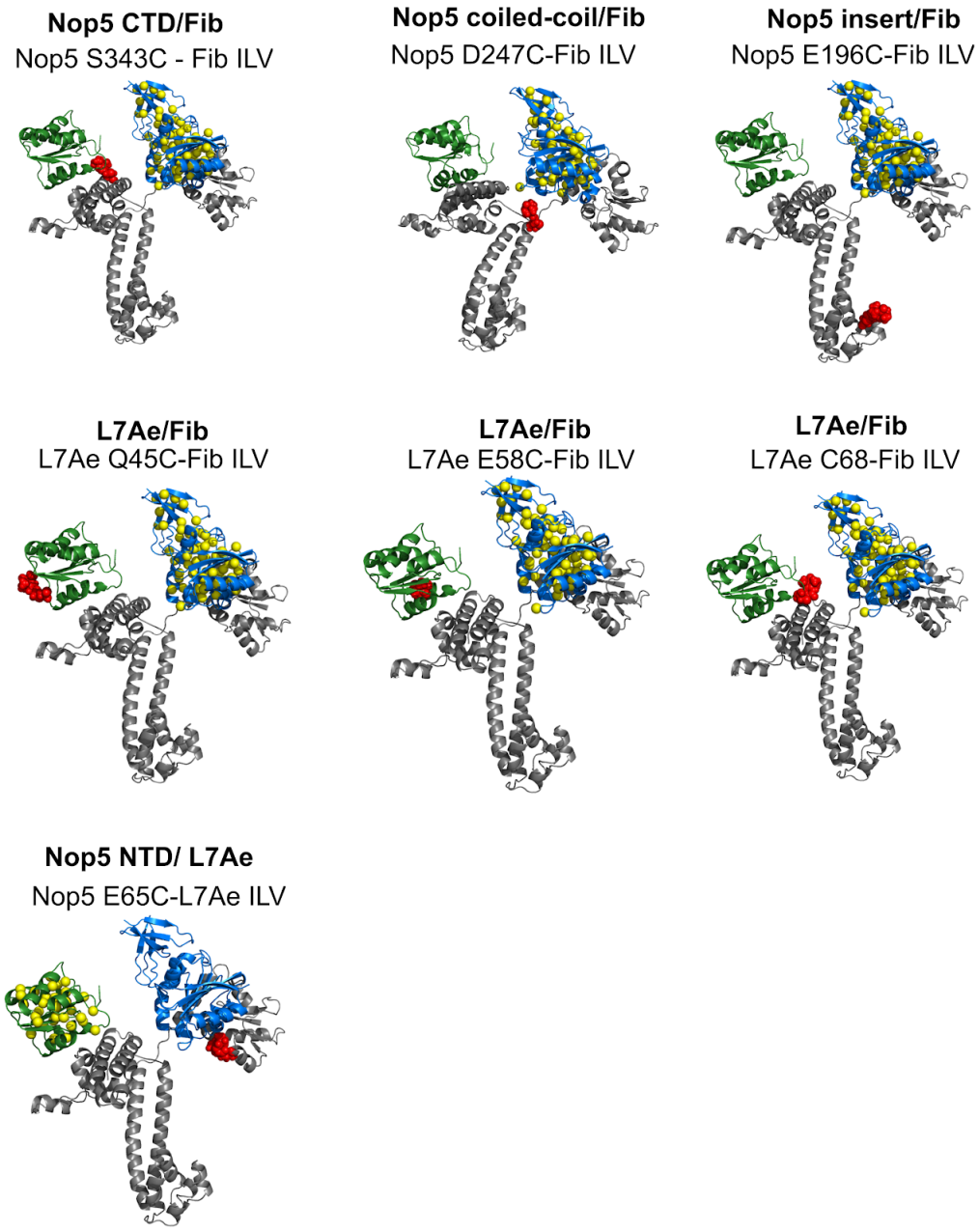


Figure 4.6.1 Positions of the paramagnetic tags used in the study. PRE tag positions shown on a single Nop5-Fib/L7Ae unit. Cysteine residues were introduced by site-directed mutagenesis and labelled with the iodoacetoamido-PROXYL tag (red spheres). Distance restraints between the paramagnetic tags and the ILV-labelled methyl groups (yellow spheres) were acquired in PRE experiments.

For each PRE spectrum, the ratios between peak intensities in the paramagnetic and diamagnetic states (I_{para}/I_{dia}) were quantified and translated into distances as described in section 3.7. This resulted in the acquisition of 424 distance restraints between tags and methyl groups for the substrate D'-bound RNP, and 398 distance restraints for the substrate

D-bound RNP from 7 PRE experiments on each RNP (see tables 4.6.1, 4.6.2). This difference is due to the fact that one of the experiments on the substrate D-bound RNP was carried out on an ILV^{proS}-methyl labelled sample, in which only one of the methyl groups of leucine and valine produces a signal in the ^{1}H-^{13}C HMQC spectrum, rather than on a full ILV-methyl labelled sample.

While the majority of the distances were compatible with both [on,off]- and [off,off]-states, a small set had to be classified as belonging uniquely to the [on,off]- or the [off,off]-state (see tables 4.6.1, 4.6.2). This included mostly distances arising from the Nop5 E65C label to the ILV methyl groups of Fib, which could only be satisfied by having a Fib copy in the proximity of the substrate-guide duplex (on-state).

Since Nop5, Fib and L7Ae are present in two copies in the mono-RNP, each methyl-tag distance arises from the averaging of the four distances between the two methyl groups (one for each L7Ae, Fib copy) and the two paramagnetic tags (one for each nop5, L7Ae copy). Unfortunately, these contributions cannot be estimated *a priori* when deriving distances from experimental data, without making strong assumptions on the structure of the system being studied. For these reasons, a single distance between paramagnetic tags and methyl groups was extracted, and imposed as appropriate in the [on,off]- or [off,off]-states as discussed in section 3.7. The positions of the paramagnetic tag pairs in the context of the full mono-RNP is shown in Fig. 4.6.2.

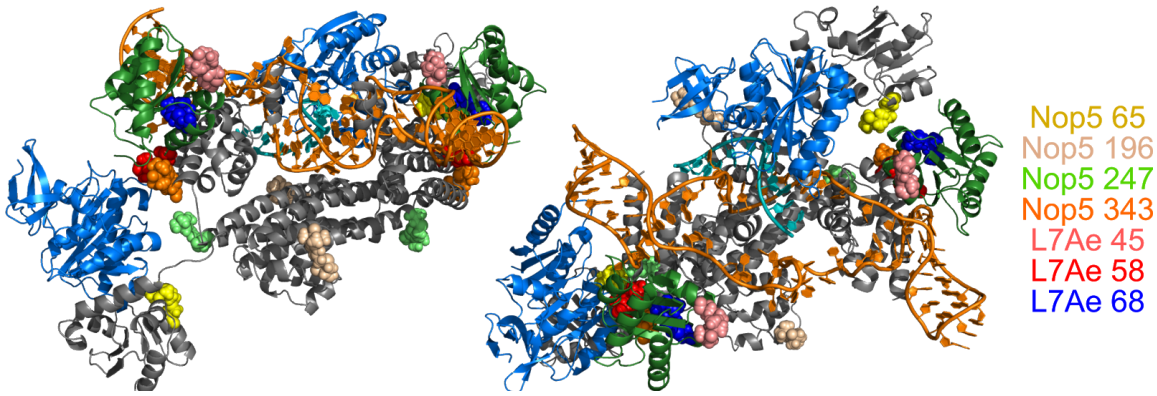


Figure 4.6.2 Positions of the paramagnetic tags in the full RNP. PRE tags (spheres) visualized on the mean structure of the selected ensemble of the substrate D'-bound [on,off]-state RNP. Each tag is present in two copies in the complex, each of which can give rise to PREs in both copies of Fib or L7Ae . Structure presented in top and side views.

The PRE measurements are summarised in **table 4.6.1 and table 4.6.2.** The observed I_{para}/I_{dia} ratios are shown together with to the theoretical values calculated from the selected ensembles of [on,off]- and [off,off]-state structures in Fig. 4.9.2.

dataset	restraints	I_{para}/I_{dia} < 0.8	[on,off]-state	[off,off]-state
Nop5 65-L7Ae	25	17	19	0
Nop5 196-Fib	66	0	2	0
Nop5 247-Fib	65	37	0	1
Nop5 343-Fib	67	7	0	0
L7Ae 45-Fib	63	3	4	0
L7Ae 58-Fib	66	37	0	2
L7Ae 68-Fib	71	0	0	3

Table 4.6.1 Substrate D'-bound PRE restraints. Summary of PRE restraints extracted from ^{1}H-^{13}C HMQC spectra of spin labelled, substrate D'-bound Box C/D particles reconstituted with st-sR26 RNA. PREs were extracted from the ratio between peak intensities in the paramagnetic and diamagnetic states (I_{para}/I_{dia}).

dataset	restraints	I_{para}/I_{dia} < 0.8	[on,off]-state	[off,off]-state
Nop5 65-L7Ae	25	9	6	0
Nop5 196-Fib	63	3	9	7
Nop5 247-Fib	67	18	0	2
Nop5 343-Fib	67	9	2	0
L7Ae 45-Fib	68	1	1	0
L7Ae 58-Fib[1]	45	10	10	0
L7Ae 68-Fib	64	4	0	0

Table 4.6.2 Substrate D-bound PRE restraints. Summary of PRE restraints extracted from ^{1}H-^{13}C HMQC spectra of spin labelled, substrate D-bound Box C/D particles reconstituted with st-sR26 RNA. PREs were extracted from the ratio between peak intensities in the paramagnetic and diamagnetic states (I_{para}/I_{dia}).

Restraints: the total number of distance extracted from each dataset; **I_{para}/I_{dia} < 0.8**: the number of resonances affected by the presence of the tag, i.e. the number of distances shorter than ~30 Å; **[on,off]-state:** the number of restraints classified as belonging exclusively to [on,off]-state conformers after preliminary structure calculation rounds (>10Å violation in [off,off]); **[off,off]-state:** the number of restraints classified as belonging exclusively to [off,off]-state conformers after preliminary structure calculation rounds (>10Å violation in [on,off]).

1: This dataset was acquired with ILV^{proS} labelling on Fib; hence the lower number of distances extracted (only one peak per LV residue appears in the spectrum).

To complement the local information derived from NMR, we acquired multiple small angle scattering curves for the substrate D and substrate D'-bound RNPs. The curves are displayed in Fig. 4.6.3, with structural parameters derived from the experiments summarised in table 4.6.3.

Scattering curve	+ substrate	Concentration (mg/ml)	R_g (Å)	D_{max} (Å)
SAXS	D'	5	47.3	164
^{2}H-Fib 42%D_2O	D'	8.9	50.2	160
^{2}H-Nop5 42%D_2O	D'	5.4	40.0	135
^{2}H-L7Ae 42%D_2O	D'	12	40.2	120
^{2}H-sRNA/^{2}H-Fib 42%D_2O	D'	4.4	47.7	160
^{2}H-sRNA 42%D_2O	D'	11.9	28.1	90
^{2}H(70%)-Nop5/^{2}H-sRNA 42%D_2O	D'	7.4	40.0	140
SAXS	D	5	50.0	170
^{2}H-Fib 42%D_2O	D	7.7	49.4	164
^{2}H-Nop5 42%D_2O	D	5.4	41.2	135
^{2}H-L7Ae 42%D_2O	D	12	39.2	120
^{2}H-sRNA/^{2}H-Fib 42%D_2O	D	4.3	51.2	140
^{2}H-sRNA 42%D_2O	D	11.9	27.9	95
^{2}H(70%)-Nop5/^{2}H-sRNA 42%D_2O	D	7.3	40.2	135

Table 4.6.3 SAS curve summary. Summary of small-angle scattering curves used in this study. D_{max} is the maximum dimension of the particle extracted from the P(r) curve after fitting with the program GNOM (svergun, 1992) (see section 1.5).

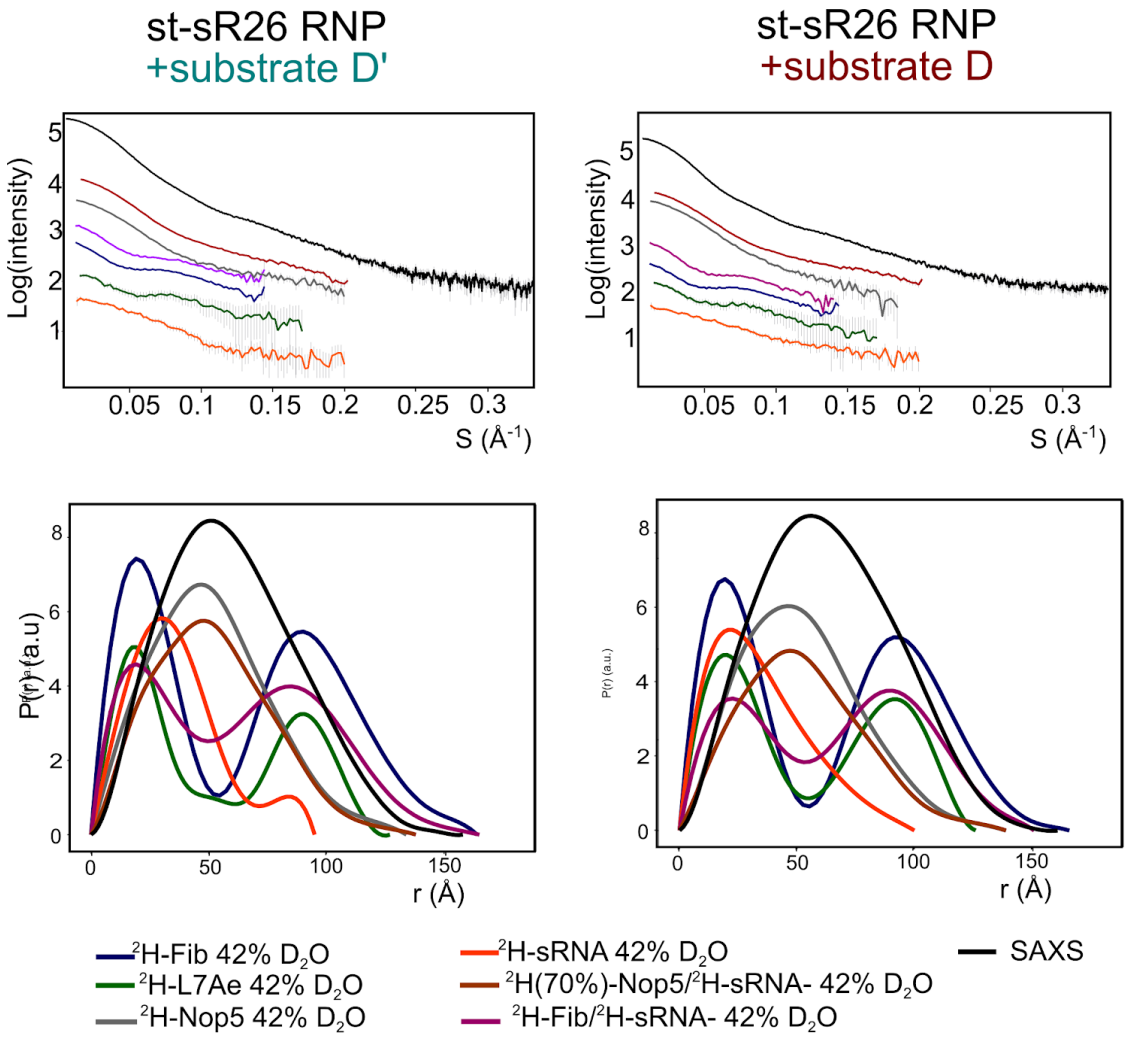


Figure 4.6.3 SAS curves used in structure calculations. Summary of small-angle scattering curves used in this study. SAXS (black) reports on the overall low-resolution features of the complex, while SANS curves at 42% D_2O (colors) report on the deuterated subunits. P(r) functions allow to derive a center of mass-center of mass distance for the two copies of L7Ae (green).

From the P(r) function of ^{2}H-L7Ae, we could extract the center of mass-center of mass distance between the two L7Ae copies by quantifying the position of the maximum of the second peak in the distribution. This corresponds to 90 Å for both the substrate D-bound and the substrate D'-bound RNP, and this distance was implemented as a restraint in structure calculations (see section 3.7).

Since the Fib subunit undergoes a conformational exchange (section 4.7), the apparent Fib subunit-subunit distance from the P(r) was not used as a restraint in structure calculations, as this distance is likely to arise from the averaging of multiple conformers.

4.7 The structures of Box C/D reconstituted with st-sR26 RNA bound to substrate D' or substrate D.

Upon binding of substrate RNA, the box C/D sRNP assumes a mono-RNP configuration. The observed PREs may be fit by two families of conformers: one with a copy of Fib in contact with the methylation site on the substrate-guide duplex, and one with both copies away from the RNA. Both conformers retain the same overall topology observed in a previous mono-RNP crystal structure from *S. solfataricus* (PDB accession code 3pla). However, this structure was obtained in presence of both substrate RNAs on an artificial two-piece guide RNA (see section 1.2).

The distance restraints described in section 4.6 were used to drive the generation of models for the st-sR26 RNPs, which were then scored using a subset of the available small-angle scattering curves.

Models were generated by successive torsion angle dynamics and cartesian refinement stages. At each stage, the models were evaluated by a consensus of total energy, restraint violation energy, and scattering curves monitoring the rigid parts of the complex (see section 3.8). Since the Nop5-NTD—Fib module was found to be flexible, these structural ensembles were selected using scattering data reporting mainly on the core of the RNP, namely the SANS curves for ^{2}H-L7Ae, ^{2}H-Nop5, ^{2}H-sRNA and ^{2}H(70%)-Nop5/^{2}H-RNA in 42%D_2O.

The remaining scattering curves, reporting on the low-resolution features of the two Fib copies and the RNP overall, were kept for scoring the conformational ensemble made up by [on,off]- and [off,off]-state conformers (see section 4.9).

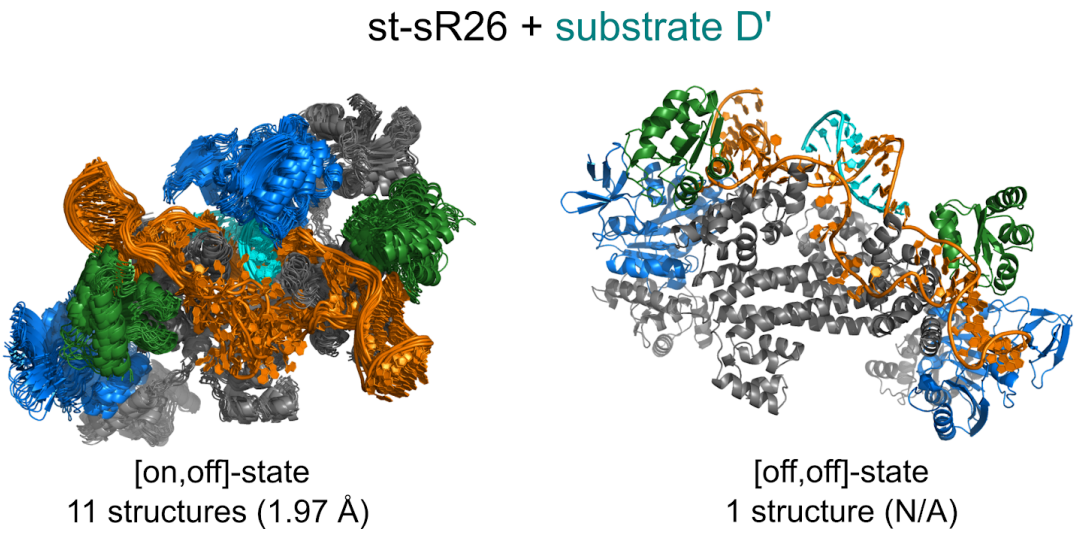


Figure 4.7.1. Structures of substrate D'-bound st-sR26 Box C/D RNP. The structures represent the ensembles selected after Cartesian refinement with distance restraint lists binned into [on,off] and [off,off]-state restraints. The mean Cα/P RMSD to the average structure excluding fully flexible regions is reported in parenthesis.

st-SR26 RNP + substrate D'. In the substrate D'-bound [on,off]-state, the conformer ensemble selected by consensus NMR and SAS score after Cartesian refinement comprises 11 structures, with a Cα-P RMSD of less than 2 Å. On the other hand, consensus scoring only selects a single structure after Cartesian refinement of the substrate D'-bound off-state (see Fig. 4.7.1).

In the [on,off]-state, the selected conformer ensemble retains the 70° angle between the Nop5 coiled-coil and the guide RNA as seen in the structure of the holo mono-RNP (Lin *et al*, 2011), along with contacts between the L7Ae C-terminus and the Nop5 NTD that positions the Fib methyltransferase. Moreover, the Nop5 α9 helix, placed between the two guide regions, is poised to make several interactions with the phosphate first base pair of the substrate-guide duplex, possibly contributing to the formation of a stable A-helix structure, which has been shown to be required for methylation (see Fig. 4.7.2).

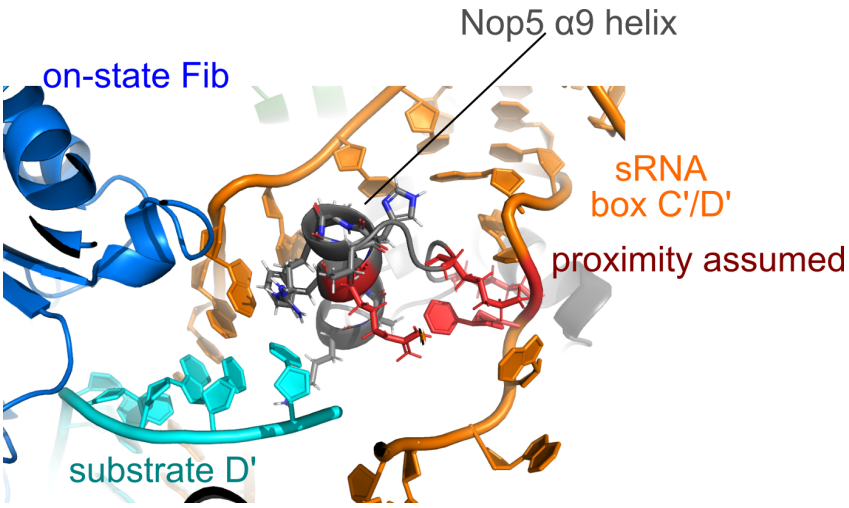


Figure 4.7.2. The Nop5 α9 helix in the substrate D'-bound [on,off]-state RNP. The placement of the Nop5 α9 helix in the average structure of the selected ensemble of [on,off]-state structures of the substrate D'-bound st-sR26 RNP. The helix is poised to make interactions with the 5' end of the substrate RNA and of the substrate-guide duplex. The residues colored in red, belonging to Nop5 and to the free guide RNA, are assumed to be in close proximity throughout the structure calculation process, by analogy with other mono-RNP structures. However, no assumptions are made in regards to the relative position of the substrate-guide duplex and the box C'/D' elements.

In the [off,off]-state, the substrate-guide duplex is placed above the plane of the sRNA and away from Nop5, in a position that presumably ensures flexibility. The two Fib copies are also facing solution, in proximity to the opposite face of L7Ae from the one proximal to the methylation position. From this arrangement, the lack of protein-RNA or protein-protein interactions makes the Nop5-NTD—Fib module unrestrained, leaving the linker formed by residues 116-124 in Nop5 free to adopt alternative conformations. This linker is responsible for positioning the Nop5-NTD—Fib module with respect to the rest of the RNP.

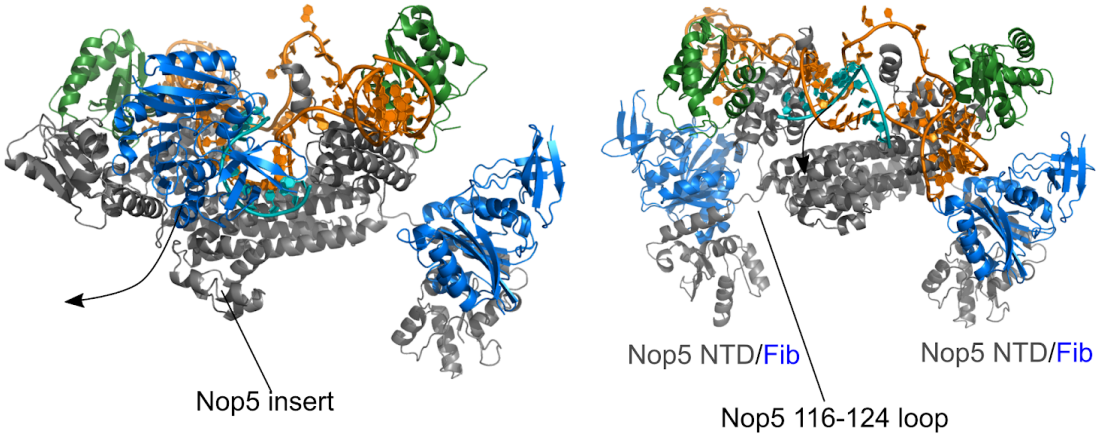


Figure 4.7.3. The motion of the Nop5-NTD—Fib module. The [on,off]-state average structure (left) and [off,off]-state (right) structure differ in the placement and dynamics of the Nop5-NTD—Fib module. In the [on,off]-state one of the two copies of this module is locked onto the substrate-guide duplex by protein-RNA interactions, interactions between the Nop5 NTD and L7Ae, and the barrier provided by the Nop5 insert region. In the [off,off]-state the 116-124 loop in Nop5 is unrestrained. This allows both Nop5-NTD—Fib modules to explore multiple positions.

Additionally, the angle between the Nop5 coiled-coil and the sRNA is retained between the [on,off]- and the [off,off]-state RNPs, enabling us to hypothesize a path for the translation of the Nop5-NTD—Fib NTD module from the substrate-guide duplex to the off position. Interestingly, the most direct version of this path would involve clashing with the Nop5 insert domain, which may therefore be providing a barrier that helps keep Fib onto the methylation site.

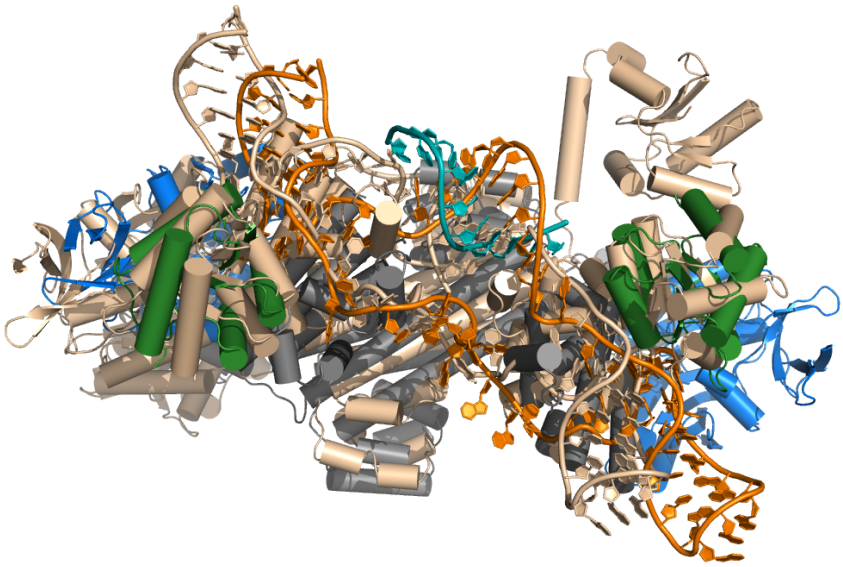


Figure 4.7.4. The relative orientation of the Nop5 coiled-coil and sRNA is retained. The structure of the substrate D'-bound [off, off]-state RNP (colors) is superimposed onto the average structure of the [on,off]-state (beige). Superposition carried out on the L7Ae subunits. The structures reveal that the two structures retain the relative orientation of each domain, despite differing in the placement of the substrate-guide duplex and the Nop5-NTD—Fib module.

st-SR26 RNP + substrate D. In the substrate D-bound st-sR 26 RNP, 5 and 6 structures are selected after Cartesian refinement of the [on,off]-state and the [off,off]-state, respectively (Fig. 4.7.5).

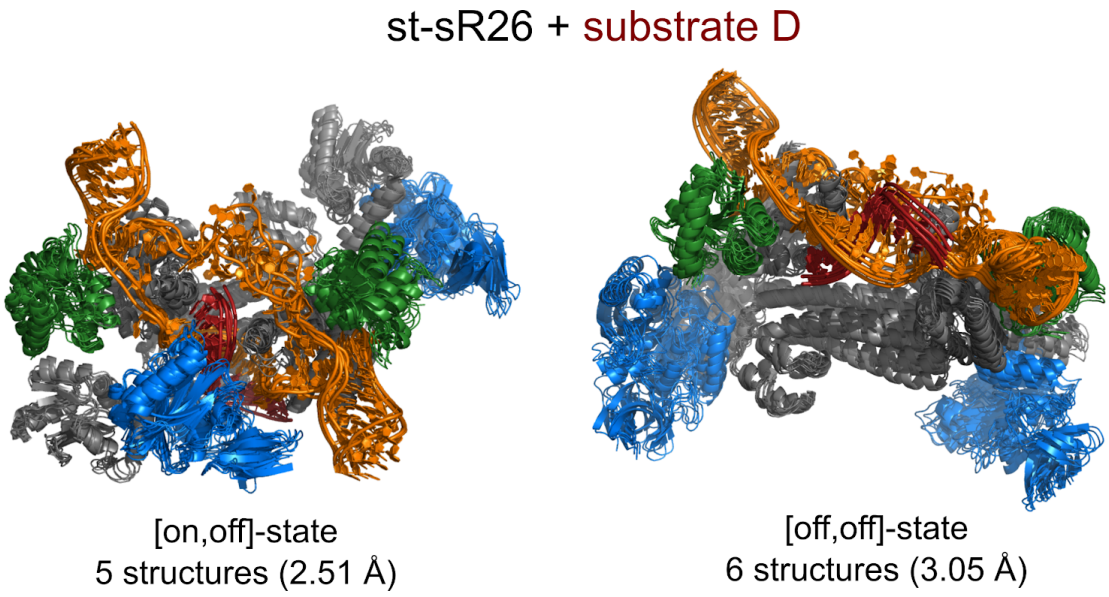


Figure 4.7.5. Structures of substrate D-bound st-sR26 Box C/D RNP. The structures represent the ensembles selected after Cartesian refinement with classified restraint lists. The mean Cα-P RMSD to the average structure excluding fully flexible regions is reported in parenthesis.

Overall, the substrate D-bound RNPs retain very similar (nearly mirror-image) conformations to the substrate D'-bound ensembles, which is consistent with the very similar PRE effects observed upon binding of either substrate (see Fig. 4.9.2).

In the substrate D-bound RNPs, the α9 helix is placed in the same position as in the substrate D'-bound RNP, where it can interact with the final base pair of the substrate-guide duplex (see Fig. 4.7.8). In this case, F293 is in close proximity to a purine on the substrate RNA, G1, and may potentially form a favourable stacking interaction. In the substrate D'-bound structure, the purine lies on the sRNA.

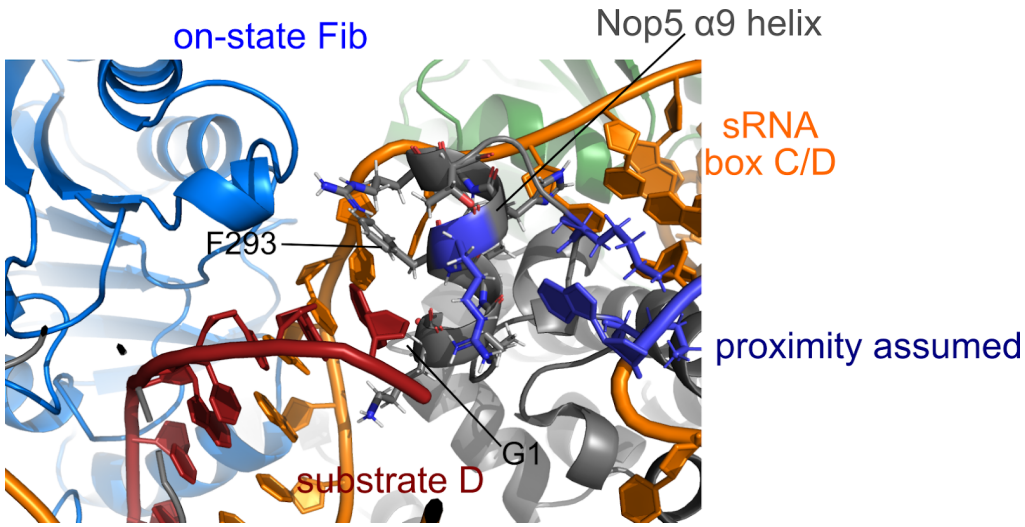


Figure 4.7.6. The Nop5 α9 helix in the substrate D-bound [on,off]-state RNP. The placement of the Nop5 α9 helix in the average structure of the selected ensemble of [on,off]-state structures of the substrate D-bound st-sR26 RNP. The helix is poised to make interactions with the 5' end of the substrate RNA and of the substrate-guide duplex.

The main difference between the substrate D'-bound and the substrate D-bound RNP lies in the different orientation of the Nop5 coiled-coil domain with respect to the sRNA in the [off,off]-state. In the average selected substrate D-bound [off,off]-state structure, the Nop5 coiled-coil runs parallel to the sRNA, rather than at a 70° angle (Fig.4.7.7).

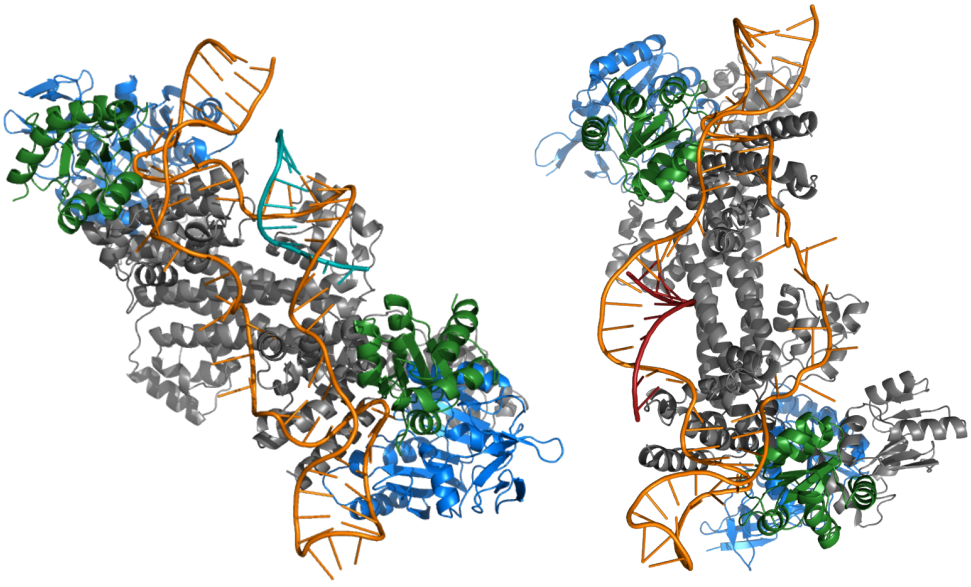


Figure 4.7.7. sRNA-Nop5 coiled-coil orientations in the [off,off]-state structures. Left, top view of the selected substrate D'-bound [off,off]-state structure. Right, top view of the average selected substrate D-bound [off,off]-state structure. The Nop5 coiled coil runs parallel to the path of the sRNA in the substrate D-bound RNP, but retains the 70° angle observed in the [on,off]-state in the case of the substrate D'-bound RNP.

4.8 Development of an algorithm capable of choosing a conformational ensemble that best fits a consensus of multiple SAS curves and PRE data.

In order to improve the fit to the experimental data and model the solution behaviour of the sRNP, we developed a scoring and selection algorithm capable of accounting for multiple coexisting conformation.

The main advantages of this algorithm over existing protocols for scoring flexible ensembles is the fact that we have developed a tool capable of selecting ensembles from a conformer pool using a consensus of NMR data and multiple small angle scattering data sets coming from both SAXS and SANS. The algorithm scores ensembles by back-calculating scattering curves and PRE effects from full-atom representations of the structures, and averaging the contribution of each conformer in the sampled ensemble, to produce theoretical SAS and PRE profiles to compare against the experimental data (see section 3.8).

We set out to obtain the equilibrium between [on,off]-state and [off,off]-state conformers of sRNPs bound to substrate D or substrate D', and to address the flexibility of the loops connecting the Nop5-NTD—Fib module to the Nop5 coiled coil module in solution. Thus, we generated two large conformer pools (2000 models) representing [on,off]- and [off,off]-state sRNPs by randomising the orientation of residues 116-124 in Nop5 copies where the Nop5-NTD—Fib module is positioned away from the RNA (see Fig. 4.9.1) while keeping the rest of the residues and RNA fixed. The orientation of the paramagnetic tags were randomised as well. These conformer pools had a bimodal distribution in terms of SAXS R_g, which can be observed in Fig.4.8.1b, and were used as the starting pools for consensus SAS/NMR ensemble selection.

The conformer pool also contained 300 models each of the [on,off]- and [off.off]-state starting structures with all protein positions kept fixed, but with fully randomised spin label positions. This is to account for PREs due to different orientations of the paramagnetic tag in the template structures.

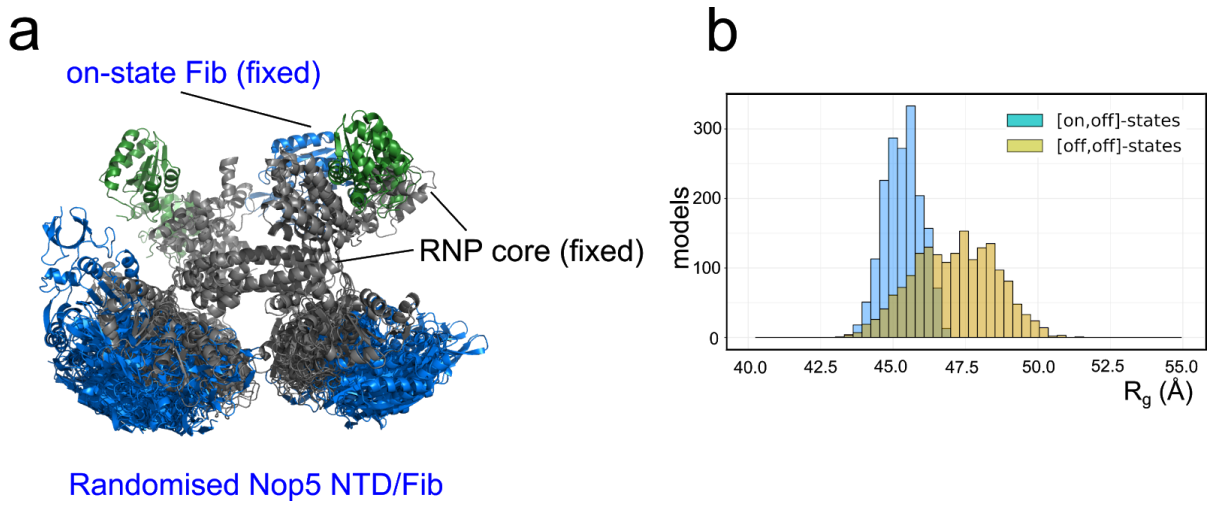


Figure 4.8.1. Scrambling of the Nop5-NTD—Fib module A)The average structure in the ensembles shown in figures 4.7.1 (substrate D'-bound) and 4.7.5 (substrate D-bound) is used as a starting point for the randomisation of the conformation of the 116-124 loop in Nop5. The resulting pool of structures forms the starting point for the selection of a mixed ensemble. 8 [on,off]-state structures are overlaid onto 8 [off,off]-state structures. The Fib copy near the substrate-guide duplex is kept fixed. RNAs not shown. B) The distribution of SAXS R_g values in the conformer pool for the selection of the substrate D'-bound RNP ensemble.

The pseudo-genetic algorithm relies on multiple sampling events to select small (3-20 conformers) ensembles, which are assigned a consensus SAS/NMR fitness score, based on normalized χ^2 to the experimental data (see section 3.8). Implementation of consensus NMR scoring relies on the assumption that Fib exchanges between on- and off-state on the time scale of the NMR experiment (7.6 ms), which is consistent with the exchange regime detailed in Fig. 4.4.2. Standard small-angle scattering programs were used to back-calculate theoretical scattering curves corresponding to each conformer, and their output was adapted to account for the number of each conformer in the ensemble (see methods).

The SAS/NMR fitness scoring was developed keeping in mind the dependence of χ^2 on experimental error, which varies between data points, experiments and techniques. As such, the assessment of SAS/PRE fitness relies on normalising across the ensembles sampled within the scoring run, placing equal weight to the total SAS fitness and the fitness to the NMR data. This normalisation method was also implemented during my studies in a separate, generalised integrative modelling approach, the HADDOCK-M3 method, to score restraint energy violation or experimental fitness and select best fitting models out of a docking run.

The consensus scoring was performed separately for substrate D-bound and substrate D'-bound sRNPs, and repeated three independent times for each, in order to characterise the variance between selected ensembles.

4.9 Binding of substrate D and substrate D' leads to different fractions of Fib bound to RNA.

The combination of small-angle scattering and NMR data showed that binding of different substrate RNAs to st-sR26 RNA induces different conformer ensembles in solutions. Indeed, no single mono-RNP conformer can adequately account for the experimental scattering curves observed upon substrate RNA binding. Specifically, all average [on,off]-state and [off,off]-state conformers selected after Cartesian refinement display very high χ^2 values for the SAXS, ^{2}H-Fib and ^{2}H-Fib/^{2}H-RNA curves.This indicates that the flexible regions of the complex, namely the Nop5-NTD−Fib modules, which make up a large fraction of the macromolecular volume, should be fit by multiple conformations, comprising both [on,off]-state and [off,off]-state structures.

The different relative populations of [on,off]-state and [off,off]-state conformers in the presence of the two substrate RNAs was derived using the ensemble scoring approach described in sections 4.8 and 3.9. As expected from the SAXS profiles and the difference in PRE effects observed for the Nop5 E65C PRE tag, the fraction of [on,off]-state conformers in the ensemble is higher upon binding of substrate D' (46.7%±8.5%) than of substrate D (21.6±9.9%).

The ensembles containing a mixed populations of [on,off]- and [off,off]-state conformers, fitted to the scattering data used by the selection algorithm, are shown in Fig. 4.9.1. For both the substrate D'-bound and the substrate D-bound ensemble, theoretical I_{para}/I_{dia} values were calculated and compared to the values extracted from PRE experiments (Fig.4.9.2).

The selected conformer ensembles provide a good fit of the experimentally observed PRE effects, and improve the fitting of the scattering curves (see table 4.9.1).

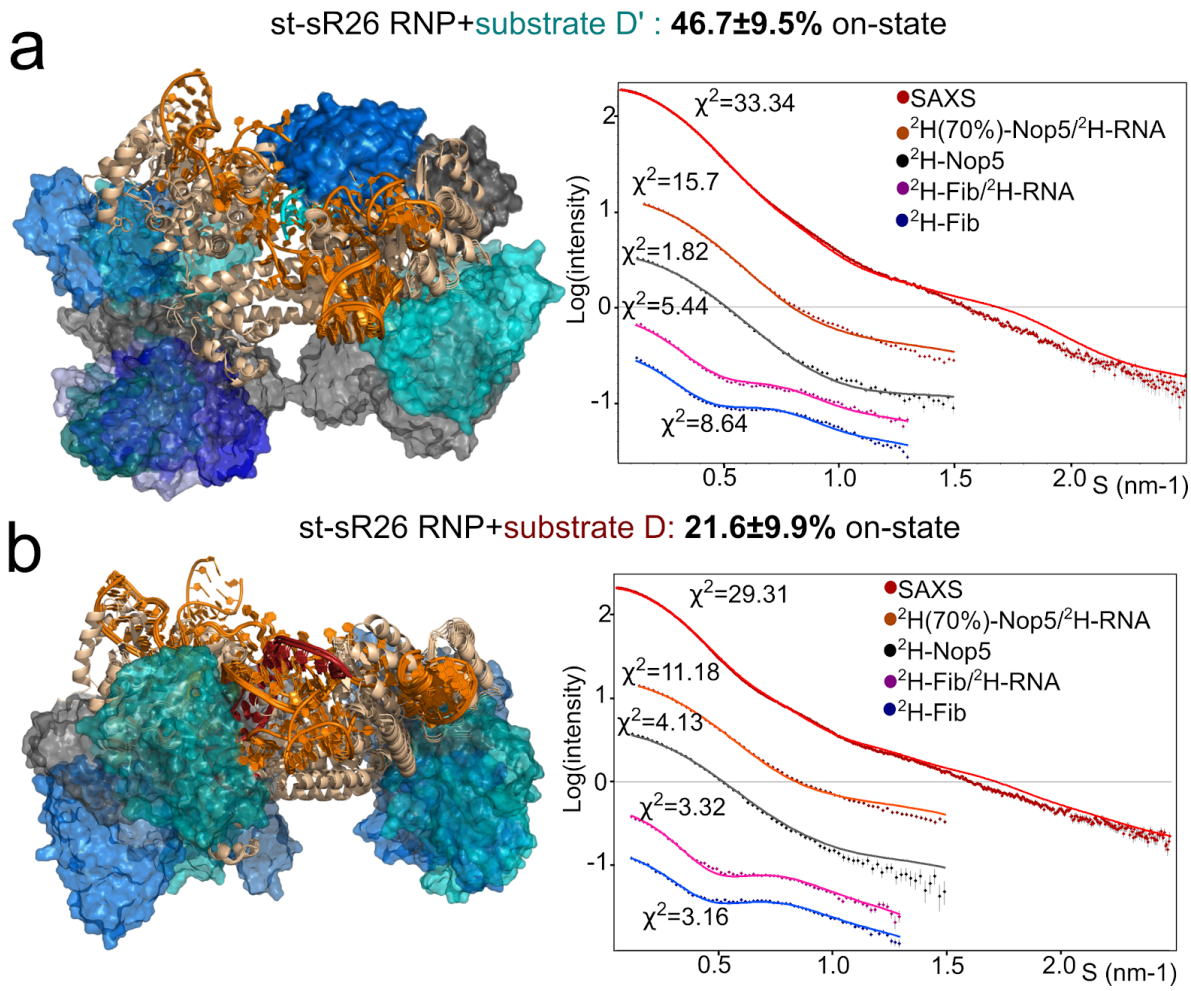


Figure 4.9.1. Selection of ensembles for the substrate D-bound and substrate D'-bound st-sR26 RNPs A) The structural ensemble selected by the pseudo-genetic scoring algorithm (see section 3.8) for the substrate D'-bound sRNP, containing five [on,off]- and five [off,off]-state conformers, with the position of the flexible Nop5-NTD—Fib module represented as surfaces. The Fib copies of each conformer in the ensemble are shown in shades of blue. The fits to the experimental SAS curves are shown on the right. B) Structural ensemble selected by the pseudo-genetic scoring algorithm for the substrate D-bound sRNP, containing one [on,off]- and six [off,off]-state conformers. In both A and B, the mean and standard deviation of the percentage of on-state structures in the three top-scoring ensembles across three independent scoring runs is shown in the title.

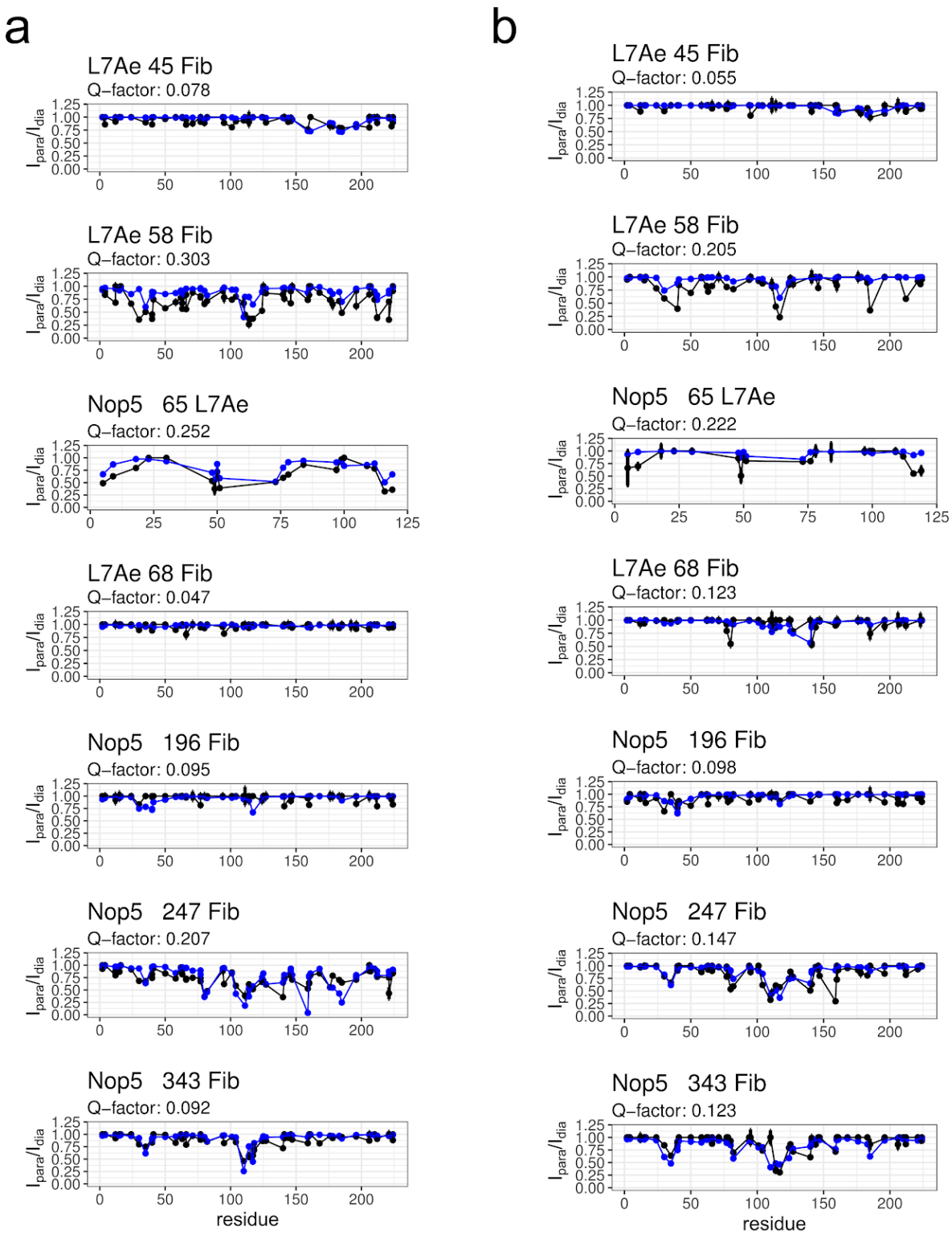

a
L7Ae 45 Fib
Q−factor: 0.078
L7Ae 58 Fib
Q−factor: 0.303
Nop5 65 L7Ae
Q−factor: 0.252
L7Ae 68 Fib
Q−factor: 0.047
Nop5 196 Fib
Q−factor: 0.095
Nop5 247 Fib
Q−factor: 0.207
Nop5 343 Fib
Q−factor: 0.092
b
L7Ae 45 Fib
Q−factor: 0.055
L7Ae 58 Fib
Q−factor: 0.205
Nop5 65 L7Ae
Q−factor: 0.222
L7Ae 68 Fib
Q−factor: 0.123
Nop5 196 Fib
Q−factor: 0.098
Nop5 247 Fib
Q−factor: 0.147
Nop5 343 Fib
Q−factor: 0.123
Ipara/Idia
residue

Figure 4.9.2 Back-calculated and experimental PRE intensity-ratios for substrate D'-bound and substrate D-bound st-sR26 RNP ensembles. a. Comparison of I_{para}/I_{dia} ratios back-calculated from the final selected conformer ensemble shown in Figure 4.9.1a (blue) with the experimental ratios (black) for the substrate D'-bound st-sR26 RNP. Each plot represents the intensity ratio observed for experiments monitoring the distances between a paramagnetic tag placed at a specific residue and the methyl groups of a an ILV-methyl labelled protein. Thus, Nop5 343 Fib indicates distances between a paramagnetic tag on residue 343 of Nop5 to ILV methyl groups of Fib. B) Comparison of I_{para}/I_{dia} ratios back-calculated from the final selected conformer ensemble shown in Figure 4.9.1b (blue) with the experimental ratio (black) for the substrate D-bound st-sR26 RNP. The PRE Q factor, reporting on the quality of fit, is calculated as described in sections 3.8.

dataset	+substrate	[on,off]-state χ^2	[off,off]-state χ^2	Ensemble χ^2
SAXS	D'	201.87	59.67	33.34
^{2}H-Fib 42%D_2O	D'	23.88	9.63	8.64
^{2}H-Nop5 42%D_2O	D'	1.30	1.01	1.82
^{2}H(70%)-Nop5/^{2}H-RNA	D'	13.83	11.70	15.7
^{2}H-Fib/^{2}H-RNA 42%D_2O	D'	12.66	10.21	5.44
SAXS	D	617.14	55.73	29.31
^{2}H-Fib 42%D_2O	D	8.75	3.99	3.16
^{2}H-Nop5 42%D_2O	D	5.76	5.87	4.13
^{2}H(70%)-Nop5/^{2}H-RNA	D	17.80	28.34	11.8
^{2}H-Fib/^{2}H-RNA 42%D_2O	D	7.17	6.87	3.32

Table 4.9.1 Fitting of conformational ensembles to experimental SAS data. The table reports the χ^2 of the fits of the individual [on,off]- and [off,off]-state structures used as starting point for the randomisation of the Nop5-NTD—Fib modules to the experimental data, and the χ^2 of the fit of the final selected ensembles.

4.10 other sRNAs.

In addition to complexes reconstituted using variants of st-sR26, we investigated complexes reconstituted using the guide sequences from other sRNAs in *P. furiosus.* In these constructs, we used the same st-sR framework used for our structural studies (identical box C/C', D/D', GAAA tetraloop) but swapped the guide regions of sR26 with the guide regions of other sRNAs. We characterised the oligomeric state of apo and substrate-bound RNPs reconstituted with these constructs by SAS. The results are reported in table 4.10.1.

We generated four constructs, containing the guide D and D' sequences of Pf sR2, sR5, sR11 and sR40. The guide regions of these sRNAs can be found in appendix 1. The sR11- and sR40-like RNPs behaved as the st-sR26 RNP, converting from a di-RNP to a mono-RNP upon substrate binding, while the sRNP containing the sR2-like RNA maintained its di-RNP form in the holo-state. The sR5-like RNP displayed a mixed behaviour.

Interestingly, the constructs also display a range of R_g values in the apo state, despite all being consistent with a di-RNP (SAXS R_g 53.7-57.3 Å), further highlighting the role of the sRNA guide regions in directing RNP assembly. Moreover, the different sRNPs behaved differently in size exclusion chromatography (see appendix 3).

sRNA	Method	concentration (mg/ ml)	labelling	Apo R_g, Å	Holo R_g, Å
st-sR2	SANS	1.36	^{2}H-Nop5 42% D_2O	46.1 ± 2.5	42.8 ± 0.9
st-sR2	SAXS	2.0	-	54.0 ± 0.4	53.6 ± 0.1
st-sR5	SAXS	2.0	-	57.3 ± 0.2	49.9 ± 0.7
st-sR11	SAXS	2.0	-	53.7 ± 0.2	48.7 ± 0.5
st-sR40	SANS	2.1	^{2}H-Nop5 42% D_2O	44.8 ± 0.8	39.7 ± 0.6

Table 4.10.1 SAS characterisation of RNPs reconstituted with different sRNAs. The table reports the R_g values of the apo and holo states of RNPs reconstituted with st-sR RNA constructs containing the guide sequences of different *P. furiosus* sRNAs[3].

[3] Experimental ^{2}H-Nop5 R_g di-RNP apo, measured on the ssR26 sRNP: 50.1 Å (Lapinaite *et al*, 2013) Experimental ^{2}H-Nop5 R_g di-RNP holo, measured on the ssR26 sRNP: 48.2 Å (Lapinaite *et al*, 2013). Experimental ^{2}H-Nop5 R_g di-RNP apo, measured on the st-sR26 sRNP: 45.2 Å (Lapinaite *et al*, 2013). Experimental ^{2}H-Nop5 R_g mono-RNP holo, measured on the st-sR26 sRNP: 38.8 Å (Lin *et al*, 2011)(concentration: 5.3 mg/ml)

5. Discussion

5.1 The structures of the substrate-bound RNPs help to explain the Box C/D functional cycle.

The structures of the substrate D and substrate D'-bound sRNPs presented here help fill in important details in the overall picture of the methylation mechanism of the Box C/D sRNP, by showing the conformation of the RNP half-way through its catalytic cycle, when it is bound to only substrate D or substrate D'. To our initial surprise, the structures of the half-bound RNPs reconstituted with st-sR26 adopt a mono-RNP configuration, despite the sRNA being very similar to the ssR26 RNA used for the reconstitution of the holo di-RNP structure. Nevertheless, activity assays indicate that these structures represent physiologically relevant states of the RNP, since st-sR26 recapitulates the behaviour of the native sR26 RNA in biochemical activity assays, including retaining a preference for methylation of substrate D' and displaying stimulation upon addition of another substrate type.

The structures of the half-loaded sRNPs presented here are determined at a precision better than 4 Å, as defined by the Cα/P RMSD over the rigid parts of the structure (see section 3.8). They show the features of the [on,off]- and [off,off]-state structures adopted by the complex upon binding of substrate D or substrate D'. Interestingly, while the [on,off]-state structures of the substrate D-bound and substrate D'-bound particles are very similar, the two [off,off]-state structures differ in the position of the sRNA relative to the Nop5 coiled-coil domain.

Our structures of the half-loaded sRNPs in the [on,off]-state retain the main features of the previously known holo mono-RNP structure (Fig. 5.1.1). This structure of the sRNP from *S. solfataricus* has both substrate RNAs bound and both Fib copies positioned onto the substrate-guide duplex ([on,on]-state). In the [on,off]-state structures presented here, the section of the RNP containing the Fib retains the contacts between Nop5 NTD and L7Ae that help position the Fib onto the ribose lying 5 nucleotides away from Box D or Box D'. Other features that are retained are the path of the sRNA and its position relative to the Nop5 coiled-coil.

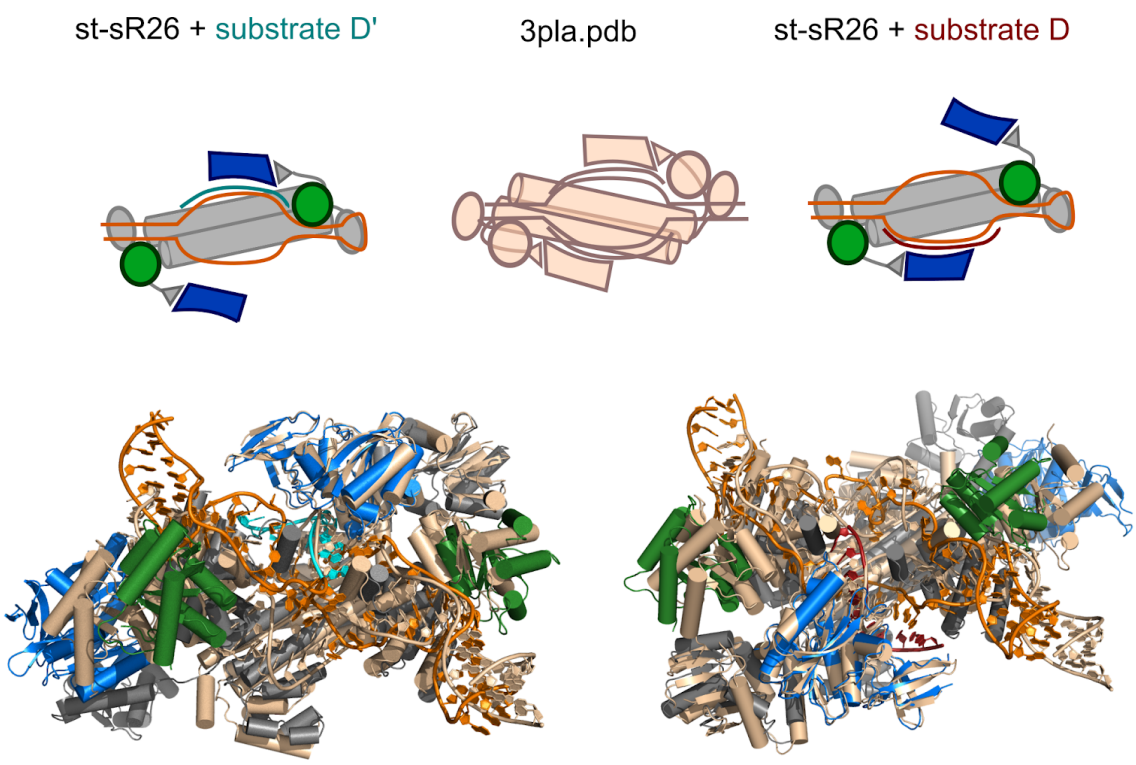


Figure 5.1.1. Comparison of the [on,off]-state structures with the *S. solfataricus* holo mono-RNP structure. Overlay of [on,off]-state st-sR26 structures (colors) onto the crystal structure of the holo mono-RNP from *S. solfataricus* (beige, PDB accession code 3PLA). Left:substrate D'-bound; Right: Substrate D-bound. Structures are overlaid on the Fib copy that is in contact with the RNA.The structures displayed are the ones closest to the mean of the ensemble selected after Cartesian refinement. The [on,off]-state structures retain the contacts between Nop5-NTD (grey) and L7Ae(green) that places Fib (blue) onto the substrate-guide duplex, as well as the angle between the Nop5 coiled-coil and the sRNA.

Additionally, while no mono-RNP structure has been solved in the apo state or in the [off,off]-state, the structures of the [off,off]-state RNPs presented here may be compared to the structures of the split RNPs obtained with truncated protein or truncated RNA constructs (see section 1.2). Interestingly, these structures also adopt a conformation in which the Nop5 coiled-coil runs at an angle relative to the putative path of the sRNA. This feature is retained in the substrate D'-bound structure, in which the sRNA path is close to the one observed in the [on,off]-state, perhaps enabling an easier transition between [on,off]- and [off,off]-state. However, the sRNA plane in the st-sR26 substrate D'-bound structure lies closer to the Nop5 coiled-coil than in the split RNP structures.

On the other hand, the Nop5 coiled-coil in substrate D-bound [off,off]-state structure runs parallel to the path of the sRNA guide regions. From this structure, a larger motion would be needed to transition into the [on,off]-state, as the Nop5 coiled-coil would need an additional rotation relative to the sRNA on top of the necessary repositioning of the Nop5-NTD—Fib module.

The structures presented here also highlight a potential role for the 56 amino acid insert in the coiled-coil domain of Nop5. Looking at the conformational changes required for the Nop5/NTD module to move between the on- and the off-states structures, we can see that the flexible loop comprising residues 116-124 of Nop5 would be clashing with the insert if it were to move along the shortest available path between the two conformations. Thus, the insert may be aiding methylation efficiency by providing a physical barrier to lock Fib on the substrate-guide duplex for a long enough time for the methylation reaction to occur.

5.2 The dynamic equilibrium of the substrate-bound sRNP particle.

As distance restraints and small angle scattering data highlights, the Nop5-NTD—Fib module is in exchange between multiple positions over the time frame of the NMR experiment. Indeed, the observed PREs and scattering curves could not be fit by any single conformer in isolation.
This dynamic behaviour of Nop5-NTD—Fib is consistent with the functional cycle of the enzyme, as it provides a mechanism for the release of the methylated substrate RNA and for enzymatic turnover.
The approach presented here to generate a conformational ensemble that fits the data consists in treating the core of the RNP as a rigid body and randomising the conformation of the linker connecting the Nop5-NTD—Fib module to the RNP core. This generates a conformer pool from which structures may be selected in order to generate an ensemble with improved fitting to the experimental data.

Ultimately, our ensemble fitting approaches also identified the fact that the experimental data is best fit by a higher ratio of [on,off]-state conformers in the substrate D'-bound case. This is

to be expected considering the stronger PRE effects observed for the Nop5 E65C spin label to L7Ae, as well as from the SAXS data, both of which highlight that a higher fraction of Fib is closer to L7Ae and the center of mass of the sRNP in the substrate D'-bound case.
It is worth mentioning that the specific conformers selected by our ensemble fitting approach only represent the "fundamental components" required to generate the best fitting ensemble to the experimental data; they do not represent the specific structures adopted by the RNP in solution. This is because the particles in solution experience dynamics on multiple time scales and shuffle between an unknown number of conformations, many of which may not be included in the conformer pool used for the search.

The higher proportion of Fib in the on-state when bound to substrate D' correlates with biochemical assays highlighting the methylation preference for substrate D'. Taken together, the data presented here suggest a role for the sequence of the substrate-guide duplex in determining substrate RNA methylation efficiency by fine tuning the equilibrium between on-state and off-state Fib populations.

However, as solution NMR methods relying on methyl-TROSY approaches are insensitive to conformational changes in RNAs, we cannot detect whether this fine tuning occurs through structural or dynamic changes in the local RNA structure around the methylation site, or via mechanisms involving protein-nucleic acid interactions.
Nevertheless, it is possible to rule out the fact that this increased methylation efficiency and higher fraction of Fib bound is due to increased substrate D'-sRNA helix stability, at least as derived from simple melting temperature models, as substrate D and substrate D' binding are predicted to generate duplexes with identical stability.
It is clear that the nucleotides at the ends of the substrate-guide duplex play an important role in sRNP function; modification of these nucleotides greatly affects the behaviour of the complex in biochemical assays. This may be due to potential stacking interactions between the duplex ends and the free nucleotide connecting them to the K-turn elements, which may help position the substrate-guide duplex relative to the plane of the RNA in a way that makes it more likely to contact the flexible Nop5-NTD—Fib module.

One of the differences between the two guide regions in sR26 (and st-sR26) is in the sequence around the methylation site. Both guide D and guide D' in sR26 direct the methylation of a cytosine in 16S pre-rRNA. In substrate D', the nucleotide to the 3' of the methylation site is a pyrimidine (guanine), while in substrate D this is a purine (adenosine).

Interestingly, an analysis of *P. furiosus* sRNAs reveals that the frequency of nucleotides at the methylation position is not equal, despite the fact that the methylation occurs on the ribose 2'-OH and not on the base (see Fig. 5.2.1). The fact that certain nucleotide combinations seem to be more frequent may also indicate that local structure around the methylation site could be stabilised by base stacking interactions.

Furthermore, certain combinations of nucleotide pairs around the methylation site seem to be very rare, consistent with the fact that rRNA sequences around 2'-OCH_3 sites may be optimised not only in terms of rRNA folding, but also in order to impact rRNA methylation kinetics and regulation (see Fig. 5.2.2).

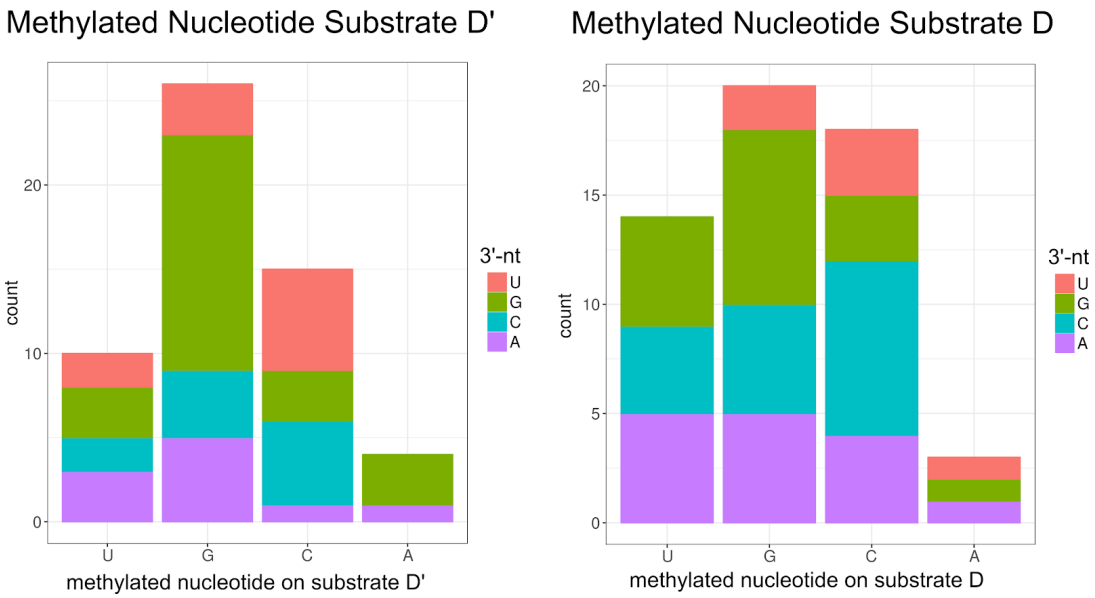


Figure 5.2.1. The methylation site features differ between substrate D and substrate D'. An analysis of the substrate RNA sequences in *P. furiosus* . Comparison of the methylated nucleotide between substrate D and substrate D' highlights that the methylation of G occurs more frequently in substrate D' than in substrate D, while methylation of A is discouraged overall. The colors highlight the nucleotide found at the 3' of the methylation site in the substrate RNA. This shows that GG base stacking is frequent in substrate D', while the distribution of stacking interactions in substrate D sequences is nearly random.

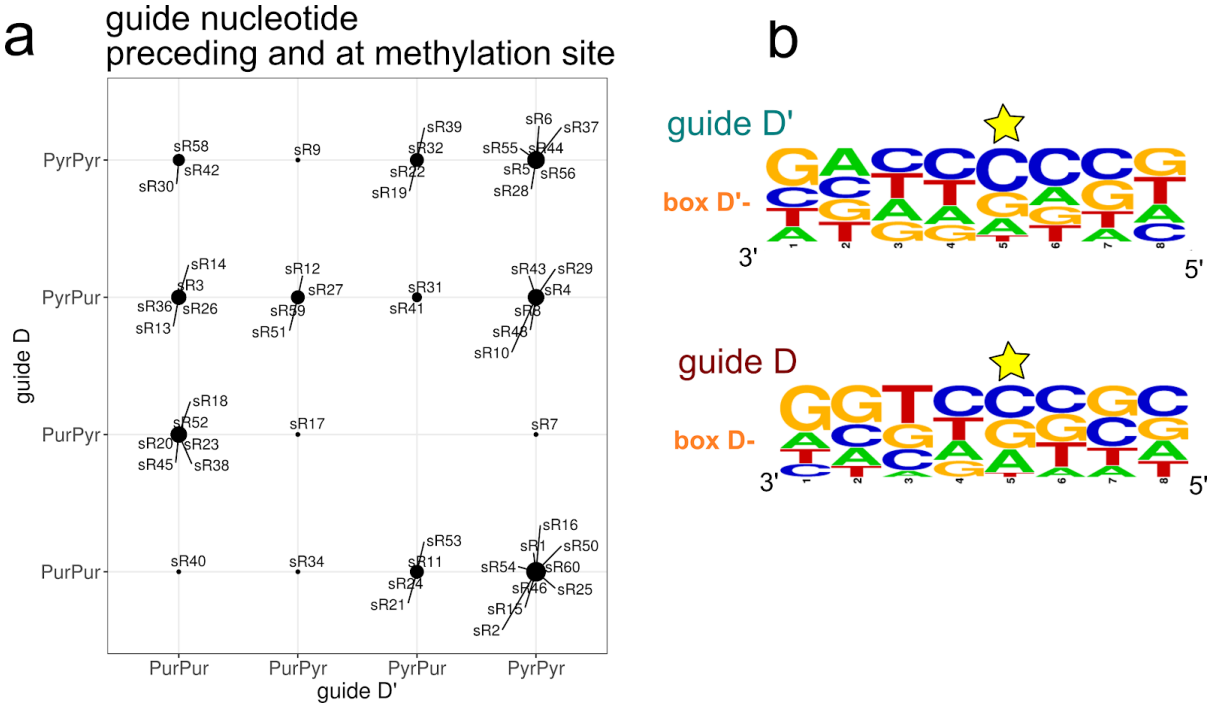


Figure 5.2.2. The guide sequence around the methylation site. An analysis of the stacking of guide nucleotides around the methylation site for the *P. furiosus* sRNAs. A) The stacking of guide nucleotides in guide D and guide D'. Pur: purine, Pyr:pyrimidine. The second nucleotide base pairs with the methylated nucleotide in substrate RNA. The first nucleotide is the one preceding it in the guide sequence. The sRNA used in this study, sR26, has the nucleotide pairs UG in guide D, and GG in guide D', and therefore is shown as as PyrPur/PurPur. B) the nucleotide frequency of guide D' and guide D from the conserved structural element to and beyond the base directing rRNA methylation, which lies 5 nucleotides upstream of the conserved box D or box D' structural element (marked by a star).

Indeed, the biochemical assays of section 4.3 show that it is mainly the nature of this guide sequence, and not its position relative to box D or box D' in the sRNA, that determines the efficiency of methylation of each substrate in isolation, though position does play a small role, as evidenced by the decrease in methylation efficiency of substrate D' upon inversion of the guide sequences. It is possible that this influence of the guide sequence on the methylation efficiency of the sRNP is mediated by interactions between the first base pair of the substrate-guide duplex and the Nop5 α9 helix, which can interact with the substrate-guide duplex differently depending on whether substrate D or substrate D' are bound.

5.3 The regulation of the activity and oligomeric state of the archaeal Box C/D sRNP.

Both the small-angle scattering and the biochemical data presented in this study highlight the important role of the sRNA guide regions in determining sRNP assembly and behaviour.

RNPs reconstituted with st-sR-26 and ssR26 display similar properties in the apo state, both assembling in a di-RNP configuration. However, while Box C/D reconstituted using the ssR26 RNA retain a di-RNP configuration upon substrate RNA binding, the st-sR26 particle transitions to a mono-RNP state. The marked difference in behaviour in these two RNPs which differ only in an 11-nucleotide stretch in the guide region for substrate D shows that the particle must be able to recognise different guide RNAs and adopt different oligomeric states based upon the nature of the sRNA upon which it is assembled.

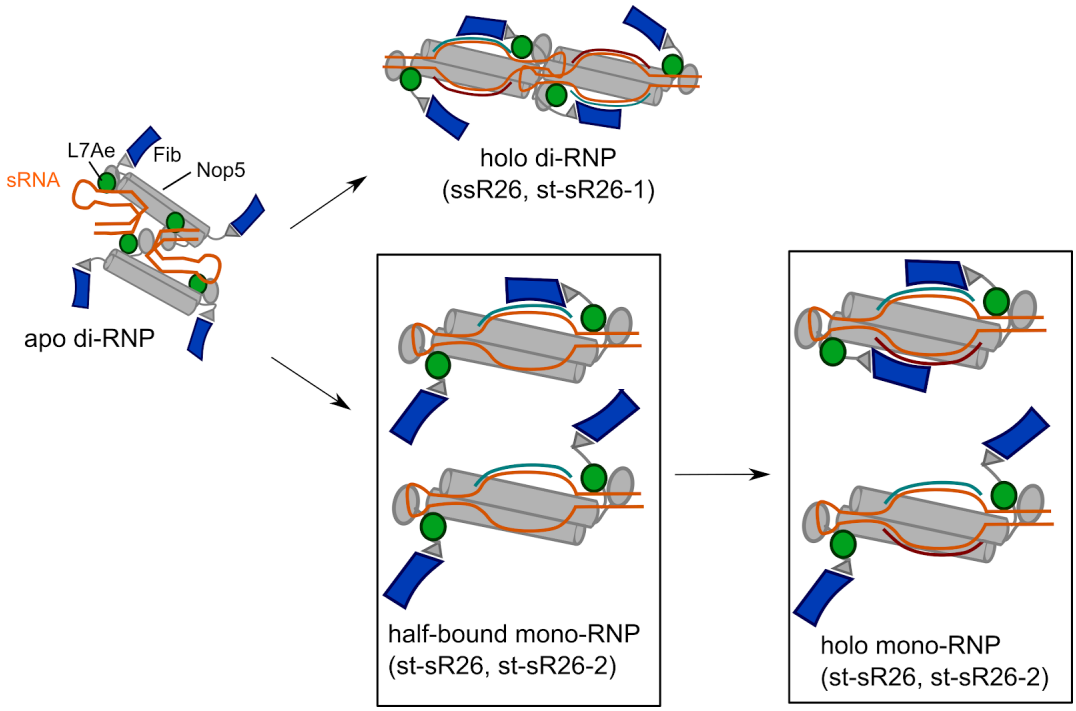


Figure 5.3.1. Dependence of sRNP architecture on sRNA sequence. While all sRNAs used in this study lead to the formation of a di-RNP apo particle, addition of substrate RNAs leads to two types of behaviour. The ssR26 and st-sR26-1 RNPs retain a holo di-RNP architecture, while st-sR26 and st-sR26-2 RNPs transition to a mono-RNP state upon binding of either substrate D or substrate D' (substrate D' shown here). The mono-RNP is in equilibrium between [on,off]- and [off,off]-state conformers.

These distinct oligomerization states are consistent with the existing literature on the archaeal Box C/D sRNP. In previous structures, the apo complex has been characterised as a di-RNP, while the holo complex has been shown to adopt both mono- and di-RNP architectures. Our experiments further confirm this heterogeneous behaviour, and identify the guide regions the elements responsible for the di- to mono-RNP transition.

The importance of the sRNA sequence has already been identified for other structural elements, since RNPs reconstituted with sRNAs lacking the apical loop assemble only partially into di-RNPs. This finding has led other groups to propose that di-RNPs particles are held together by interactions between the apical loops of two sRNA copies. Although mutational analysis shows that these interactions may play a role, this proposed dimerisation mechanism is in conflict with the apo and holo *P. furiosus* di-RNP structures, which show that protein-protein and protein-RNA interactions are necessary for di-RNP formation.
The studies presented here further confirm that the presence of the apical loops alone cannot be the determinant of the oligomeric state of the complex, since we observe that st-sR26-1 and st-sR26-2 display different low-resolution features in the apo state and adopt two different oligomeric states upon substrate RNA binding, despite both having GAAA tetraloops as their apical loop elements.

While the studies presented here do not identify a precise mechanism by which the guide regions determine the oligomeric state of the substrate-bound RNP, we can hypothesize that the nature of RNA-protein interactions involving the Nop5 α9 helix plays an important role. As shown in section 1.2, this helix interacts with the nucleotides at the ends of the substrate-guide duplex in different ways in mono- and di-RNP structures. As the α9 helix sits between the two guide regions and the conserved K-loop/K-turn elements, it may help determine the relative orientation of the sRNA and Nop5 coiled-coil domain, thereby enabling the cross-complex protein-RNA interactions required for di-RNP formation. Thus, the differences in guide D sequence between st-sR26-1 and st-sR26-2 may drive the formation of mono- or di-RNP particles by presenting different nucleotides to the Nop5 α9 helix, which would then influence the overall path of the sRNA.

The difference in oligomeric states may then be correlated to the behaviour of the complex in biochemical assays, where we observe that di-RNP particles do not exhibit the positive cooperative behaviour that is found in mono-RNPs. This is consistent with the holo di-RNP

structure, which features two of the four Fib copies stably bound to the substrate-guide duplex, and with the corresponding ^{1}H-^{13}C HMQC spectra, which shows that Fib shuttles on a fast NMR time scale between on- and off-states in mono-RNPs, but not in di-RNPs.

In the context of archaeal pre-rRNA processing, such a heterogeneous behaviour would be consistent with the enzyme coupling 2'-O methylation at multiple rRNA sites through both kinetics and spatial proximity.
Specifically, we can hypothesize that while mono-RNPs would promote the coupling between the methylation at two rRNA sites, di-RNPs would promote the alternative methylation of the site in substrate D or substrate D', consistent with the behaviour observed in biochemical assays.

This would be advantageous, as it would enable the archaeon to use 2'-O methylation to coordinate the folding of structural elements in rRNA that lie far away in sequence, but are recognised by the same sRNA. Furthermore, the methylation of multiple rRNA sites may be coupled in different ways, depending on whether an sRNA directs the reconstitution of a mono- or a di-RNP. This would then enable the folding of specific structural elements on the different time scales required by the various steps of rRNA processing.

Finally, the studies presented here are consistent with the hypothesis that Box C/D may undergo some sort of irreversible activation mechanism, as it does not seem possible to reconstitute a di-RNP from two mono-RNPs. This may indicate that the complex is produced by the cell in its di-RNP form and, upon binding of the correct substrate RNA, it becomes irreversibly activated into the corresponding holo mono- or di-RNPs, capable of processing specific rRNA sites with the required kinetics.

In the case of the di-RNPs, these particles may be specialized in the methylation of either substrate D or substrate D'. On the other hand mono-RNPs would turn over one substrate type after the other, since turnover of a mono-RNP does not require transition via an apo-like state. Another hypothesis is that sRNPs are not designed to be processive enzymes, but rather particles that assemble at specific points of the ribosomal biogenesis pathway, catalyse individual methylation events and dissociate.

5.4 Implications for the eukaryotic Box C/D snoRNP.

Recent developments in RNA sequencing have highlighted that rRNAs in higher eukaryotes display different methylation patterns across cell types, and that not all sites are methylated to the same extent, even within a specific cell type (Erales *et al*, 2017).
The idea that the static picture of rRNA methylation observed in current ribosome structures is not fully compatible with rRNA processing *in vivo* is consistent with both structural and biochemical data presented in this study.

In this study, we show that the archaeal Box C/D sRNP methylates different substrate RNAs with different efficiency. As the components of the archaeal sRNP are homologous to the eukaryotic snoRNP components, our findings provide a model for understanding the basis for different extent of methylations found in eukaryotes *in vivo* - in terms of enzymatic efficiency and shifts in conformational equilibria of sRNPs and, by analogy, snoRNPs.

However, the sequence elements of eukaryotic snoRNAs are less conserved and snoRNAs often only contain one guide region. Thus, it is probable that the overall dynamics of snoRNPs in rRNA folding differ from those of sRNPs, as eukaryotic proteins would not be able to couple methylation of multiple rRNA sites. In the future, understanding of the dynamics of the Box C/D snoRNP reconstituted with different snoRNAs could help explain how the current set of eukaryotic snoRNAs directs concerted methylation of pre-rRNA.

5.5 Accounting for flexibility in integrative structural biology.

Over the past decade, integrative structural biology approaches have matured into reliable methods to model and solve structures of large biological complexes . These approaches, at least at an initial stage, often rely on the relative placement of subunits or modules of a macromolecular complex as rigid bodies (Ward *et al*, 2013).
While these approaches have provided success, it has been recognised that biological macromolecules often carry out their function through conformational changes or interactions requiring flexible regions. This has provided a considerable challenge in terms of adapting the conformation and relative position of individual domains or full subcomplexes within a larger assembly, where their structures come from a different conformational or oligomeric state. One of the most basic approaches, which has worked well in optimising

local structure, has been to use normal mode analysis to simulate protein motions (Tama *et al*, 2004; Komolov *et al*, 2017). Alternatively, flexible fitting has relied on molecular dynamics (Trabuco *et al*, 2008) or custom energy functions (Zheng, 2011).

The data and general docking approach presented here has formed the basis for the implementation of the HADDOCK-M3 protocol (Karaca *et al*, 2017), which adapts the widely used HADDOCK docking protocol (Dominguez *et al*, 2003) to function with multiple different data types and on docking problems comprising a large number of molecules.

The ensemble scoring methods presented here, which are inspired by the ensemble optimisation method EOM (Tria *et al*, 2015) , ENSEMBLE (Krzeminski *et al*, 2013) and other approaches used for the characterisation of disordered proteins by SAXS or by NMR, deal with the conformational heterogeneity of proteins in solution by creating multiple pools of conformers in which the structure of the flexible element is randomised, while the rest of the particle is kept in the particular arrangement found in a given conformation.

These pools then provide a starting point for a pseudo-genetic algorithm which relies on successive sampling events scored according to fitness to NMR, SAXS and SANS data. While these methods were developed with the dynamics between [on,off]- and [off,off]-state RNPs in mind, they may be easily ported to several systems where flexible regions need to be characterised by a consensus of NMR and small angle scattering.

In particular, this is the first ensemble fitting method capable of properly normalising and scaling the contribution of NMR and multiple scattering datasets, as previous algorithms are only capable of dealing with individual scattering curves, or the combination of multiple NMR data sets with a single scattering curve. This enabled us to take full advantage of the SANS contrast matching method to extract the conformer ensemble best fitting scattering curves reporting on different subunits of the complex individually and as a whole[4].

[4] In its present version the ensemble scoring algorithm cannot deal with a mixture of different oligomeric states, due to the fact that the theoretical contribution of each individual conformer to the scattering of the ensemble is linearly scaled. This could be made possible by scaling the theoretical scattering curve of each component according to its volume fraction. Additionally, ensembles of systems interacting with each other may also not be scored using this method, as one would need to introduce an additional structure factor contribution to the calculation of the theoretical ensemble scattering curve.

Thus, the methods developed in this study may be adapted to find conformational equilibria in single or multi-component systems co-existing in several conformational states in solution, such as proteins undergoing a closed- to open-state transition, or intrinsically disordered proteins affected by binding events.

Appendix 1. RNA sequences.

sRNAs (guide D' in light blue; guide D in red)

construct	sequence
sR26	5'GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAAAUAGUGAUGA**CGGAGGUG AUC**UCUGAGCUCGC-3'
ssR26	5'-GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAGCUCGAAAGAGCAAUGAUG A**GUGAUGGGCGA**ACUGAGCUCGC-3'
inv-sR26	5'-GCGAGCAAUGAUGACGGAGGUGAUCACUGAAAUAGUGAUGA**GUGAUGGG CGA**UCUGAGCUCGC
st-sR26	5'-GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAGCUCGAAAGAGCAAUGAUG A**CGGAGGUGAUC**ACUGAGCUCGC-3'
st-sR26-1	5'-GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAGCUCGAAAGAGCAAUGAUG A**CGGAGGGGCGA**ACUGAGCUGCG-3'
st-sR26-2	5'-GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAGCUCGAAAGAGCAAUGAUG A**GUGAUGUGAUC**ACUGAGCUGCG-3'
st-sR26-1 A61U	st-sR26_1: 5'-GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAGCUCGAAAGAGCAAUGAUG A**CGGAGGGGCGU**ACUGAGCUGCG-3'
st-sR26-1 A61C	st-sR26_1: 5'-GCGAGCAAUGAUGA**GUGAUGGGCGA**ACUGAGCUCGAAAGAGCAAUGAUG A**CGGAGGGGCGC**ACUGAGCUGCG-3'
st-sR2	5'-GCGAGCAAUGAUGA**GUUUUUCCCUCAC**ACUGAGCUCGAAAGAGCAAUGA U GA**GGAGCCGAUGC**ACUGAGCUGCG-3'
st-sR5	5'-GCGAGCAAUGAUG**AGCGCCAUCCGAU**ACUGAGCUCGAAAGAGCAAUGAU G**ACCGGAUUCCUG**ACUGAGCUGCG-3'
st-sR11	5'-GCGAGCAAUGAUGA**GUUUGCCGAGU**ACUGAGCUCGAAAGAGCAAUGAUG A**AUAAACAGUCG**ACUGAGCUGCG-3'
st-sR40	5'-GCGAGCAAUGAUGA**UAGGCGGGUUU**ACUGAGCUCGAAAGAGCAAUGAUG A**ACCUUUGGAGC**ACUGAGCUGCG-3'

Substrate RNAs

Substrate RNA	Corresponding sRNA	sequence
Substrate D’	st-sR26, sR26	5′-GCUUCGCCCAUCAC-3’
Substrate D	st-sR26, sR26	5′-GUAGAUCACCUCCG-3’
Substrate D	st-sR26-1	5’-GUAUCGCCCCUCCG-3’
Substrate D	st-sR26-2	5’-GUAGAUCACAUCAC-3’
Substrate D’	st-sR2	5’-ACCGUGAGGGAAAAG-3’
Substrate D	st-sR2	5’-GAGGCAUCGGCUCC-3’
Substrate D’	st-sR5	5’-GCUAUCGGAUGGCGC-3’
Substrate D	st-sR5	5’-GUACAGGAAUCCGG-3’
Substrate D’	st-sR11	5’-GCUACUCGGCAAAC-3’
Substrate D	st-sR11	5’-GUACGACTGUUUAU-3’
Substrate D’	st-sR40	5’-GCUAAACCCGCCUA-3’
Substrate D	st-sR40	5’-GUAGCUCCAAAGGU-3’

Appendix 2. Assignment of Fib methyl groups in the apo sRNP.

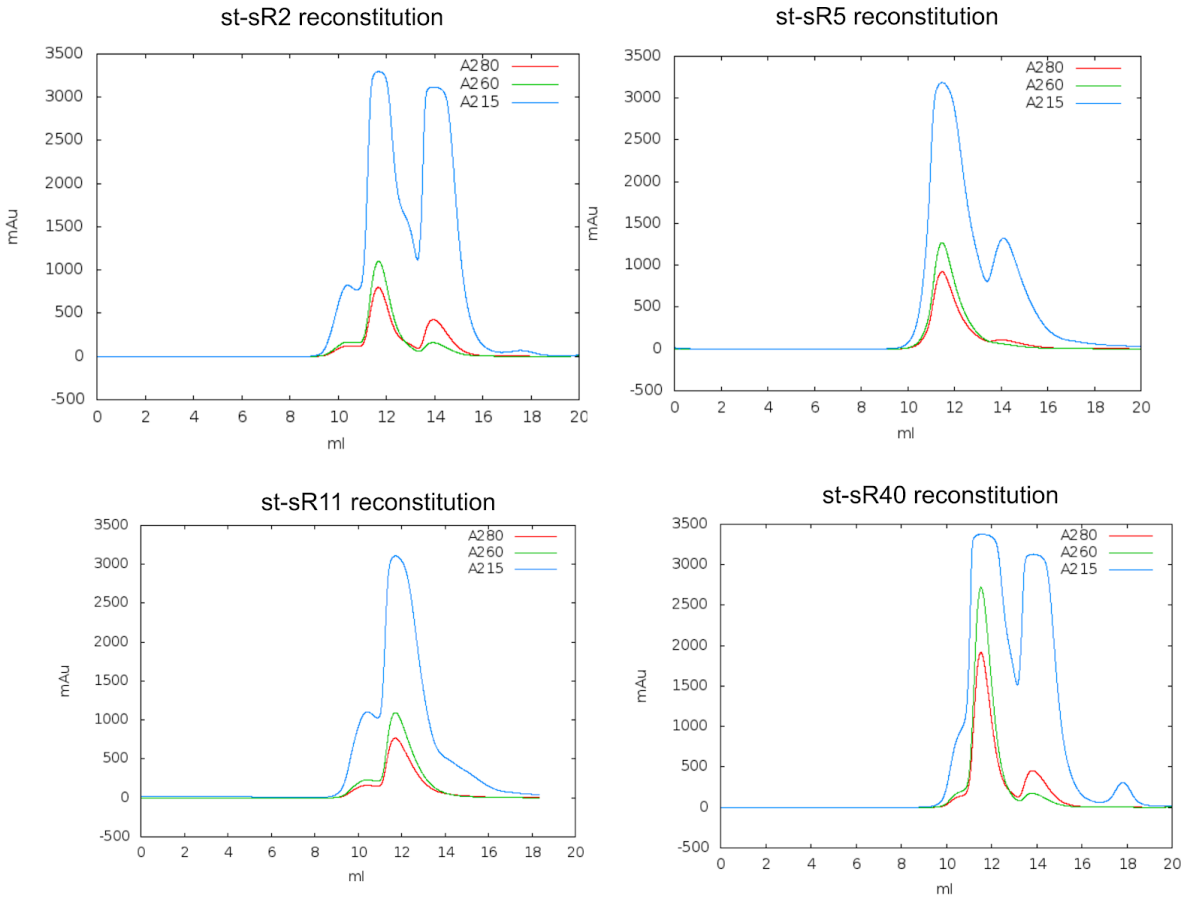
1H (ppm)	13C (ppm)	Assign F2	Details
0.94026	21.02744	2ValCg1	
0.894	21.62151	2ValCg2	
-0.22171	21.794	4ValCg1	
-0.17178	20.59013	4ValCg2	
0.95806	23.48054	12ValCg1	
0.80026	22.09954	12ValCg2	
0.94068	21.86902	14ValCg1	
0.84394	21.53585	14ValCg2	
0.53634	21.51207	15ValCg1	
0.60858	20.75737	15ValCg2	
0.69875	12.82336	16IleCd1	
0.23167	13.94084	24IleCd1	
-0.47005	25.21657	29LeuCd1	conflicting PREs, may be swapped with 200 from free state
0.06676	23.33358	29LeuCd2	conflicting PREs, may be swapped with 200 from free state
0.68817	19.93522	30ValCg1	
1.01933	21.17079	35ValCg1	
0.84427	22.02121	40ValCg1	
0.80254	21.74413	40ValCg2	
0.69978	14.46695	41IleCd1	
0.7058	11.42924	50IleCd1	
0.64266	24.54418	58LeuCd1	
0.8117	22.64643	58LeuCd2	
-0.35102	13.34904	62IleCd1	
0.73493	22.22873	63ValCg1	
0.8979	21.95461	63ValCg2	
0.5577	27.69067	66LeuCd1	

0.83312	23.36067	66LeuCd2
0.66031	13.23103	71IleCd1
0.732	21.18209	77ValCg1
0.88165	20.71229	77ValCg2
1.20902	27.50492	78LeuCd1
0.97504	24.57498	78LeuCd2
0.87952	26.76352	80LeuCd2
0.42596	15.95571	82IleCd1
0.84523	23.46789	91ValCg1
0.93204	22.22846	91ValCg2
1.00556	15.35244	94IleCd1
0.99812	18.98061	95ValCg1
0.6046	15.52616	101IleCd1
-0.14007	9.86156	104IleCd1
1.09209	21.96873	110ValCg1
0.83825	27.03071	111LeuCd1
0.72517	26.26017	114LeuCd1
0.8027	23.65939	114LeuCd2
1.0247	12.56152	117IleCd1
0.95907	22.21014	118ValCg2
0.54266	13.57757	124IleCd1
0.85509	14.30438	125IleCd1
0.7848	15.01303	127IleCd1
1.00101	27.29119	140LeuCd2
0.11953	18.64441	141ValCg1
0.50044	21.4781	141ValCg2
0.94984	18.65887	144ValCg1
0.83311	22.52956	146ValCg1
0.81377	21.35257	146ValCg2
0.97895	15.02514	147IleCd1
0.92073	21.36111	151ValCg1
0.91912	21.92011	151ValCg2

0.60244	14.39608	159IleCd1	
0.51864	27.10209	160LeuCd1	
0.92807	14.13574	161IleCd1	
0.89095	27.74325	168LeuCd1	
1.18507	23.6477	168LeuCd2	
0.68899	13.61532	176IleCd1	
1.09059	22.31755	178ValCg1	
0.75818	14.51198	183IleCd1	
1.10521	21.23994	185ValCg1	
1.08093	21.17039	185ValCg2	
0.8945	23.58878	192ValCg1	
1.1303	24.00147	196ValCg1	
0.49725	21.60089	196ValCg2	
0.97485	24.05653	200LeuCd1	conflicting PREs, may be swapped with 29 from free state
0.26938	26.12128	200LeuCd2	conflicting PREs, may be swapped with 29 from free state
0.98606	21.60375	206ValCg1	
1.04506	21.5404	206ValCg2	
0.85491	12.07407	207IleCd1	
0.86845	27.77246	210LeuCd1	
0.95542	25.22403	210LeuCd2	
0.71332	21.95643	212LeuCd1	
0.76536	25.89377	212LeuCd2	
0.77757	27.21417	221LeuCd1	
0.87287	25.7149	221LeuCd2	
1.07994	21.593	223ValCg1	
1.09808	21.57152	223ValCg2	
-0.29572	17.36346	224ValCg1	
0.71919	22.69922	224ValCg2	
1.14239	18.89209	{264}[692]	Tentative 118
0.84885	27.408	{271}[706]	Tentative 128
0.98361	23.65461	{272}[708]	

0.82055	22.21642	{275}[715]	Tentative 35
1.27954	21.93924	{277}[719]	
0.78621	25.0075	{282}[730]	
0.90175	23.32719	{283}[732]	LVpairs shows a Val, L spectrum a leu
0.9086	24.8935	{284}[733]	Tentative 111

Appendix 3. Reconstitution of RNPs with other sRNAs



Size exclusion chromatography of RNPs reconstitution with st-sR2,5,11 and 40 RNAs highlights difference in formation of higher order oligomers during reconstitution, as evidenced by the shoulder eluting in front of the main peak. Elution volumes: st-sR2 RNP: 11.7 ml; st-sR5 RNP: 11.5 ml; st-sR11 RNP: 11.7 ml; st.sR40 RNP: 11.4 ml .

Appendix 4. Publication list

The work presented in this thesis has been partially published in these works:

Graziadei A, Masiewicz PM, Lapinaite A, Carlomagno T. “Archaea box C/D enzymes methylate two distinct substrate rRNA sequences with different efficiency.” *RNA 2016; 22(5): 764-72*

Karaca E, Rodrigues JPGLM, **Graziadei A**, Bonvin AMJJ, Carlomagno T. “M3: an integrative framework for structure determination of molecular machines.”*Nat. Methods 2017; 14(9):897-902*

Additionally, I performed SAS analysis for this work:

Glatt S, Zabel R, Kolaj-Robin O, Onuma OF, Baudin F, **Graziadei A**, Taverniti V, Lin TY, Baymann F, Seraphin B, Breunig KD, Müller CW. “Structural basis for tRNA modification by Elp3 from Dehalococcoides mccartyi.” *Nat. Str. Mol. Biology 2016; 23(9):794802*

Acknowledgments

To my supervisor, Dr. Teresa Carlomagno, thank you for your support and guidance throughout my graduate studies and throughout the thesis writing process. Your expertise, patience and drive has been crucial throughout all the ups and downs of the PhD. I have learned a lot from your breath of knowledge and from the possibility you gave me to try many different approaches on this challenging work.

I would also like to thank my thesis advisory committee members, Prof. Frauke Melchior, Dr. Martin Beck and Dr. Ramesh Pillai for taking an interest in my research, and providing guidance, especially in the early stages of the work.

I also would like to thank Dr. Bernd Simon and Dr. Ezgi Karaca for all the help and teaching introducing me to experimental NMR and the world of structural bioinformatics and integrative modelling. Similarly, I would like to thank Dr. John Kirkpatrick and Dr. Teresa Carlomagno for all the input on the technical aspects of translating NMR signals into distance restraints, and Dr. Frank Gabel for assistance with recording and advice on interpretation of SAS data.

Thanks to Dr. Pawel Masiewicz and Simone Hoefler for helping me obtain biochemical activity data, and for the insightful discussions on this challenging complex.

I would like to thank all current and past members of the Carlomagno Group. Throughout these years, they have provided personal support and a great environment for scientific and less serious discussions. My immense gratitude goes to Dr. Megha Karanth and Leo Nesme, who have not only helped and provided valuable advice and personal friendship, but also proofread this manuscript, and to Nataliya Danilenko, whose support, love and enthusiasm got me through the tough and the good times. A special mention goes to Dr. Luca Codutti for the scientific guidance and dry humor he dashed out throughout my studies.

On a personal note, I am forever thankful to my parents, Ines and Michele and to my brother Nanni for their support during these 6 years.

Bibliography

Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ & Li H (2003) Structure and function of archaeal box C/D sRNP core proteins. *Nat. Struct. Biol.* **10:** 256–263

Appel CD & Maxwell ES (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

von Appen A, Kosinski J, Sparks L, Ori A, DiGuilio AL, Vollmer B, Mackmull M-T, Banterle N, Parca L, Kastritis P, Buczak K, Mosalaganti S, Hagen W, Andres-Pons A, Lemke EA, Bork P, Antonin W, Glavy JS, Bui KH & Beck M (2015) In situ structural analysis of the human nuclear pore complex. *Nature* **526:** 140–143

Basu A, Das P, Chaudhuri S, Bevilacqua E, Andrews J, Barik S, Hatzoglou M, Komar AA & Mazumder B (2011) Requirement of rRNA Methylation for 80S Ribosome Assembly on a Cohort of Cellular Internal Ribosome Entry Sites. *Mol. Cell. Biol.* **31:** 4482–4499

Baudin-Baillieu A, Fabret C, Liang X-H, Piekna-Przybylska D, Fournier MJ & 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

Baxter-Roshek JL, Petrov AN & Dinman JD (2007) Optimization of ribosome structure and function by rRNA base modification. *PLoS One* **2:** e174

Belin S, Beghin A, Solano-Gonzàlez E, Bezin L, Brunet-Manquat S, Textoris J, Prats A-C, Mertani HC, Dumontet C & Diaz J-J (2009) Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PLoS One* **4:** e7147

Birkedal U, Christensen-dalsgaard M, Krogh N, Sabarinathan R, Gorodkin J & Nielsen H (2015) Profiling of ribose methylations in RNA by high-throughput sequencing. *Angewandte Chemie - International Edition* **54:** 451–455

Bizarro J, Charron C, Boulon S, Westman B, Pradet-Balade B, Vandermoere F, Chagot M-E, Hallais M, Ahmad Y, Leonhardt H, Lamond A, Manival X, Branlant C, Charpentier B, Verheggen C & Bertrand E (2014) Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *J. Cell Biol.* **207:** 463–480

Bleichert F, Gagnon KT, Brown BA, Maxwell ES, Leschziner AE, Unger VM, Baserga SJ, Bleichen F, Gagnon KT, Brown BA, Stuart Maxwell E, Leschziner AE, Unger VM & Baserga SJ (2009) A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science* **325:** 1384–1387

Bower-Phipps KR, Taylor DW, Wang H-W & Baserga SJ (2012) The box C/D sRNP dimeric architecture is conserved across domain Archaea. *RNA* **18:** 1527–1540

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T & Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure

determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921

Cahill NM (2002) Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J.* **21:** 3816–3828

Capel MS, Engelman DM, Freeborn BR, Kjeldgaard M, Langer JA, Ramakrishnan V, Schindler DG, Schneider DK, Schoenborn BP & Sillers IY (1987) A complete mapping of the proteins in the small ribosomal subunit of Escherichia coli. *Science* **238:** 1403–1406

Cavaillé J & Bachellerie JP (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res.* **26:** 1576–1587

Clore GM & Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* **109:** 4108–4139

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

Dauden MI, Kosinski J, Kolaj-Robin O, Desfosses A, Ori A, Faux C, Hoffmann NA, Onuma OF, Breunig KD, Beck M, Sachse C, Séraphin B, Glatt S & Müller CW (2017) Architecture of the yeast Elongator complex. *EMBO Rep.* **18:** 264–279

Decatur WA & Fournier MJ (2002) rRNA modifications and ribosome function. *Trends Biochem. Sci.* **27:** 344–351

Dez C, Froment C, Noaillac-Depeyre J, Monsarrat B, Caizergues-Ferrer M & Henry Y (2004) Npa1p, a component of very early pre-60S ribosomal particles, associates with a subset of small nucleolar RNPs required for peptidyl transferase center modification. *Mol. Cell. Biol.* **24:** 6324–6337

Dominguez C, Boelens R & Alexandre M J (2003) HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **125:** 1731–1737

Ebersberger I, Simm S, Leisegang MS, Schmitzberger P, Mirus O, von Haeseler A, Bohnsack MT & Schleiff E (2014) The evolution of the ribosome biogenesis pathway from a yeast perspective. *Nucleic Acids Res.* **42:** 1509–1523

Erales J, Marchand V, Panthu B, Gillot S, Belin S, Ghayad SE, Garcia M, Laforêts F, Marcel V, Baudin-Baillieu A, Bertin P, Couté Y, Adrait A, Meyer M, Therizols G, Yusupov M, Namy O, Ohlmann T, Motorin Y, Catez F, et al (2017) Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **114:** 12934–12939

Falb M, Amata I, Gabel F, Simon B & Carlomagno T (2010) Structure of the K-turn U4 RNA: a combined NMR and SANS study. *Nucleic Acids Res.* **38:** 6274–6285

Feigin LA & Svergun DI (1987) Structure Analysis by Small-Angle X-Ray and Neutron Scattering

Feoktystov AV, Frielinghaus H, Di Z, Jaksch S, Pipich V, Appavou M-S, Babcock E, Hanslik R, Engels R, Kemmerling G, Kleines H, Ioffe A, Richter D & Brückel T (2015) KWS-1

high-resolution small-angle neutron scattering instrument at JCNS: current state. *J. Appl. Crystallogr.* **48:** 61–70

Ferber M, Kosinski J, Ori A, Rashid UJ, Moreno-Morcillo M, Simon B, Bouvier G, Batista PR, Müller CW, Beck M & Nilges M (2016) Automated structure modeling of large protein assemblies using crosslinks as distance restraints. *Nat. Methods* **13:** 515–520

Ferreira-Cerca S, Pöll G, Gleizes P-E, Tschochner H & Milkereit P (2005) Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol. Cell* **20:** 263–275

Franke D, Jeffries CM & Svergun DI (2015) Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12:** 419–422

Franke D & Svergun DI (2009) DAMMIF, a program for rapidab-initioshape determination in small-angle scattering. *J. Appl. Crystallogr.* **42:** 342–346

Gabel F (2015) Small-Angle Neutron Scattering for Structural Biology of Protein-RNA Complexes. *Methods Enzymol.* **558:** 391–415

Gabel F, Simon B, Nilges M, Petoukhov M, Svergun D & Sattler M (2008) A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. *J. Biomol. NMR* **41:** 199–208

Gabel F, Simon B & Sattler M (2006) A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. *Eur. Biophys. J.* **35:** 313–327

Gagnon KT, Zhang X, Qu G, Biswas S, Suryadi J, Brown BA & Maxwell ES (2009) 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

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

Gautier T, Bergès T, Tollervey D & Hurt E (1997) Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol. Cell. Biol.* **17:** 7088–7098

Ghalei H, Hsiao H-H, Urlaub H, Wahl MC & Watkins NJ (2010) A novel Nop5 – sRNA interaction that is required for efficient archaeal box C / D sRNP formation. **:** 2341–2348

Goody TA, Melcher SE, Norman DG & Lilley DMJ (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* **10:** 254–264

Goto NK, Gardner KH, Mueller GA, Willis RC & Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J. Biomol. NMR* **13:** 369–374

Graziadei A, Masiewicz P, Lapinaite A & Carlomagno T (2016) Archaea box C/D enzymes methylate two distinct substrate rRNA sequences with different efficiency. *RNA* **22:** 764–772

Guinier A & Fournet G (1955) Small-angle Scattering of X-rays [by] André Guinier [and]

Gérard Fournet. Translation by Christopher B. Walker

Hamelryck T & Manderick B (2003) PDB file parser and structure class implemented in Python. *Bioinformatics* **19:** 2308–2310

Hartman H, Favaretto P & Smith TF (2006) The archaeal origins of the eukaryotic translational system. *Archaea* **2:** 1–9

Huang L & Lilley DMJ (2016) The Kink Turn, a Key Architectural Element in RNA Structure. *J. Mol. Biol.* **428:** 790–801

Iwahara J, Schwieters CD & Clore GM (2004) Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126:** 5879–5896

Karaca E, Rodrigues JPGLM, Graziadei A, Bonvin AMJJ & Carlomagno T (2017) M3: an integrative framework for structure determination of molecular machines. *Nat. Methods* **14:** 897–902

Kater L, Thoms M, Barrio-Garcia C, Cheng J, Ismail S, Ahmed YL, Bange G, Kressler D, Berninghausen O, Sinning I, Hurt E & Beckmann R (2017) Visualizing the Assembly Pathway of Nucleolar Pre-60S Ribosomes. *Cell* **171:** 1599–1610.e14

Kim H, Abeysirigunawarden SC, Chen K, Mayerle M, Ragunathan K, Luthey-Schulten Z, Ha T & Woodson SA (2014) Protein-guided RNA dynamics during early ribosome assembly. *Nature* **506:** 334–338

Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M & Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85:** 1077–1088

Kiss-László Z, Henry Y & Kiss T (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* **17:** 797–807

Kiss T (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20:** 3617–3622

Komolov KE, Du Y, Duc NM, Betz RM, Rodrigues JPGLM, Leib RD, Patra D, Skiniotis G, Adams CM, Dror RO, Chung KY, Kobilka BK & Benovic JL (2017) Structural and Functional Analysis of a β-Adrenergic Receptor Complex with GRK5. *Cell* **169:** 407–421.e16

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ & Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36:** 1277–1282

Kornprobst M, Turk M, Kellner N, Cheng J, Flemming D, Koš-Braun I, Koš M, Thoms M, Berninghausen O, Beckmann R & Hurt E (2016) Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell* **166:** 380–393

Koš M & Tollervey D (2010) Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. *Mol. Cell* **37:** 809–820

Krzeminski M, Marsh JA, Neale C, Choy W-Y & Forman-Kay JD (2013) Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* **29:** 398–399

Kuhn JF, Tran EJ & Maxwell ES (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.* **30:** 931–941

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

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

Lichtenecker RJ, Coudevylle N, Konrat R & Schmid W (2013) Selective isotope labelling of leucine residues by using α-ketoacid precursor compounds. *Chembiochem* **14:** 818–821

Linge JP, Habeck M, Rieping W & Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19:** 315–316

Lin J, Lai S, Jia R, Xu A, Zhang L, Lu J & Ye K (2011) Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469:** 559–563

Lubini P, Zürcher W & Egli M (1994) Stabilizing effects of the RNA 2’-substituent: crystal structure of an oligodeoxynucleotide duplex containing 2'-O-methylated adenosines. *Chem. Biol.* **1:** 39–45

Madl T, Sattler M, Zhang Y, Bagdiul I, Kern T, Kang H, Zou P, Maeusbacher N, Sieber SA & Kraemer A (2013) Structure, phosphorylation and U2AF65 binding of the Nterminal Domain of splicing factor 1 during 3 splice site Recognition. Available at: http://dx.doi.org/10.2210/pdb2m0g/pdb

Marcel V, Ghayad SE, Belin S, Therizols G, Morel A-P, Solano-Gonzàlez E, Vendrell JA, Hacot S, Mertani HC, Albaret MA, Bourdon J-C, Jordan L, Thompson A, Tafer Y, Cong R, Bouvet P, Saurin J-C, Catez F, Prats A-C, Puisieux 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

Martin JL & McMillan FM (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12:** 783–793

Matos JR & Wong C-H (1987) S-adenosylmethionine: Stability and stabilization. *Bioorg. Chem.* **15:** 71–80

Mertens HDT & Svergun DI (2017) Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch. Biochem. Biophys.* **628:** 33–41

Nguyen THD, Galej WP, Bai X-C, Savva CG, Newman AJ, Scheres SHW & Nagai K (2015) The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523:** 47–52

Nicoloso M, Qu LH, Michot B & Bachellerie JP (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2’-O-ribose methylation of rRNAs. *J. Mol. Biol.* **260:** 178–195

Nolivos S, Carpousis AJ & Clouet-d'Orval B (2005) The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res.* **33:** 6507–6514

Oeffinger M, Zenklusen D, Ferguson A, Wei KE, El Hage A, Tollervey D, Chait BT, Singer RH & Rout MP (2009) Rrp17p is a eukaryotic exonuclease required for 5' end processing of Pre-60S ribosomal RNA. *Mol. Cell* **36:** 768–781

Ollerenshaw JE, Tugarinov V & Kay LE (2003) Methyl TROSY: explanation and experimental verification. *Magn. Reson. Chem.* **41:** 843–852

Oruganti SV, Zhang Y & Li H (2005) Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem. Biophys. Res. Commun.* **333:** 550–554

Pei J, Kim B-H & Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36:** 2295–2300

Peng Y, Yu G, Tian S & Li H (2014) Co-Expression and Co-Purification of Archaeal and Eukaryal Box C/D RNPs. *PLoS One* **9:** e103096

Pernot P, Round A, Barrett R, De Maria Antolinos A, Gobbo A, Gordon E, Huet J, Kieffer J, Lentini M, Mattenet M, Morawe C, Mueller-Dieckmann C, Ohlsson S, Schmid W, Surr J, Theveneau P, Zerrad L & McSweeney S (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20:** 660–664

Pervushin K, Riek R, Wider G & Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.* **94:** 12366–12371

Polikanov YS, Melnikov SV, Söll D & Steitz TA (2015) Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.* **22:** 342–344

Porterfield JZ, Zachary Porterfield J & Zlotnick A (2010) A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology* **407:** 281–288

Quinternet M, Chagot M-E, Rothé B, Tiotiu D, Charpentier B & Manival X (2016) Structural Features of the Box C/D snoRNP Pre-assembly Process Are Conserved through Species. *Structure* **24:** 1693–1706

Rodríguez-Mateos M, García-Gómez JJ, Francisco-Velilla R, Remacha M, de la Cruz J & Ballesta JPG (2009) Role and dynamics of the ribosomal protein P0 and its related trans-acting factor Mrt4 during ribosome assembly in Saccharomyces cerevisiae. *Nucleic Acids Res.* **37:** 7519–7532

Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B & Sali A (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10:** e1001244

Schneidman-Duhovny D, Pellarin R & Sali A (2014) Uncertainty in integrative structural

modeling. *Curr. Opin. Struct. Biol.* **28:** 96–104

Sgourakis NG, Lange OF, DiMaio F, André I, Fitzkee NC, Rossi P, Montelione GT, Bax A & Baker D (2011) Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings. *J. Am. Chem. Soc.* **133:** 6288–6298

Singh H & Lane BG (1964) THE ALKALI-STABLE DINUCLEOTIDE SEQUENCES IN 18S+28S RIBONUCLEATES FROM WHEAT GERM. *Can. J. Biochem.* **42:** 1011–1021

Sloan KE, Warda AS, Sharma S, Entian K-D, Lafontaine DLJ & Bohnsack MT (2017) Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* **14:** 1138–1152

Solomon I & Bloembergen N (1956) Nuclear Magnetic Interactions in the HF Molecule. *J. Chem. Phys.* **25:** 261–266

Soss SE & Flynn PF (2007) Functional implications for a prototypical K-turn binding protein from structural and dynamical studies of 15.5K. *Biochemistry* **46:** 14979–14986

Špačková N, Réblová K & Šponer J (2010) Structural Dynamics of the Box C/D RNA Kink-Turn and Its Complex with Proteins: The Role of the A-Minor 0 Interaction, Long-Residency Water Bridges, and Structural Ion-Binding Sites Revealed by Molecular Simulations. *J. Phys. Chem. B* **114:** 10581–10593

Strunk BS, Novak MN, Young CL & Karbstein K (2012) A Translation-Like Cycle Is a Quality Control Checkpoint for Maturing 40S Ribosome Subunits. *Cell* **150:** 111–121

Sun Q, Zhu X, Qi J, An W, Lan P, Tan D, Chen R, Wang B, Zheng S, Zhang C, Chen X, Zhang W, Chen J, Dong M-Q & Ye K (2017) Molecular architecture of the 90S small subunit pre-ribosome. *Elife* **6:** Available at: http://dx.doi.org/10.7554/eLife.22086

Svergun D, Barberato C & Koch MHJ (1995) CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28:** 768–773

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25:** 495–503

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76:** 2879–2886

Svergun DI, Koch MHJ, Timmins PA & May RP (2013) Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules OUP Oxford

Svergun DI, Richard S, Koch MH, Sayers Z, Kuprin S & Zaccai G (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.* **95:** 2267–2272

Szewczak LBW, Gabrielsen JS, Degregorio SJ, Strobel SA & Steitz JA (2005) Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. *RNA* **11:** 1407–1419

Tama F, Miyashita O & Brooks CL 3rd (2004) Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J.*

Mol. Biol. **337:** 985–999

Taoka M, Nobe Y, Yamaki Y, Yamauchi Y, Ishikawa H, Takahashi N, Nakayama H & Isobe T (2016) The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* **44:** 8951–8961

Tomkuvienė M, Ličytė J, Olendraitė I, Liutkevičiūtė Z, Clouet-d'Orval B & Klimašauskas S (2017) Archaeal fibrillarin-Nop5 heterodimer 2'--methylates RNA independently of the C/D guide RNP particle. *RNA* **23:** 1329–1337

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

Trabuco LG, Villa E, Mitra K, Frank J & Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16:** 673–683

Trabuco LG, Villa E, Schreiner E, Harrison CB & Schulten K (2009) Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49:** 174–180

Trapman J, Retèl J & Planta RJ (1975) Ribosomal precursor particles from yeast. *Exp. Cell Res.* **90:** 95–104

Tria G, Mertens HDT, Kachala M & Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2:** 207–217

Tugarinov V, Hwang PM, Ollerenshaw JE & Kay LE (2003) Cross-correlated relaxation enhanced 1H- 13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125:** 10420–10428

Tugarinov V & Kay LE (2003) Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **125:** 13868–13878

Tugarinov V & Kay LE (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR* **28:** 165–172

Tugarinov V & Kay LE (2006) A 2H NMR relaxation experiment for the measurement of the time scale of methyl side-chain dynamics in large proteins. *J. Am. Chem. Soc.* **128:** 12484–12489

Tugarinov V & Kay LE (2013) Estimating side-chain order in [U-2H;13CH3]-labeled high molecular weight proteins from analysis of HMQC/HSQC spectra. *J. Phys. Chem. B* **117:** 3571–3577

Tycowski KT, Shu MD & Steitz JA (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379:** 464–466

Volkov VV & Svergun DI (2003) Uniqueness ofab initioshape determination in small-angle scattering. *J. Appl. Crystallogr.* **36:** 860–864

Wagner EK, Penman S & Ingram VM (1967) Methylation patterns of HeLa cell ribosomal RNA and its nucleolar precursors. *J. Mol. Biol.* **29:** 371–387

Ward AB, Sali A & Wilson IA (2013) Integrative structural biology. *Science* **339:** 913–915

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73:** 1–16

Woolford JL Jr & Baserga SJ (2013) Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics* **195:** 643–681

Xue S & Barna M (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13:** 355–369

Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, Stuart Maxwell E & Li H (2010) Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol. Cell* **39:** 939

Yang Z, Lin J & Ye K (2016) Box C/D guide RNAs recognize a maximum of 10 nt of substrates. *Proceedings of the National Academy of Sciences* **113:** 10878–10883

Ye K, Jia R, Lin J, Ju M, Peng J, Xu A & Zhang L (2009) Structural organization of box C/D RNA-guided RNA methyltransferase. *Proceedings of the National Academy of Sciences* **106:** 13808–13813

Yip WSV, Shigematsu H, Taylor DW & Baserga SJ (2016) Box C/D sRNA stem ends act as stabilizing anchors for box C/D di-sRNPs. *Nucleic Acids Res.* **44:** 8976–8989

Yip WSV, Vincent NG & Baserga SJ (2013) Ribonucleoproteins in archaeal pre-rRNA processing and modification. *Archaea* **2013:** 614735

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD & Noller HF (2014) Crystal structure of the ribosome at 5.5 A resolution. Available at: http://dx.doi.org/10.2210/pdb4v42/pdbx

Zhang X (2006) The coiled-coil domain of the Nop56/58 core protein is dispensable for sRNP assembly but is critical for archaeal box C/D sRNP-guided nucleotide methylation. *RNA* **12:** 1092–1103

Zhang X, Yan C, Hang J, Finci LI, Lei J & Shi Y (2017) An Atomic Structure of the Human Spliceosome. *Cell* **169:** 918–929.e14

Zhang X, Yan C, Zhan X, Li L, Lei J & Shi Y (2018) Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28:** 307–322

Zheng W (2011) Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophys. J.* **100:** 478–488