

# 1 Introduction

## 1.1 The Lsm proteins: Ring architectures for RNA capture

### 1.1.1 Introduction: life cycle of mRNA

It is today recognised that the vast majority of the cellular pool of RNA (nearly 98 % in humans) comprises non-coding RNA (ncRNA) species (Mattick 2001), with only a small proportion serving as direct template for protein synthesis. The diverse ncRNA forms are themselves capable of function, involved in a plethora of tasks such as protein scaffolding, *cis* and *trans* regulatory roles and catalysis (Lilley 2005; Mattick and Makunin 2006). Many of these functions are carried out in tight partnership with specific ancillary proteins within large ribonucleo-protein complexes (RNPs) (Eddy 2001).

Various types of ncRNA, as well as RNPs containing tRNA, rRNA or snRNA, directly interact with mRNA at different stages of its life. Figure 1.1 presents an overview of the maturation of pre-mRNA and the fate of the mRNA generated. Pre-mRNA initially undergoes modification to enhance its stability: a 5' methyl guanosine (m<sub>7</sub>G) cap added during transcription (Wen and Shatkin 1999) and a poly(A)-tail placed in the 3' region by the polyadenylation machinery (Proudfoot et al. 2002; Balbo and Bohm 2007). Following initiation of spliceosomal assembly by recruitment of core particles in the cytoplasm, non-coding introns are spliced from the pre-mRNA sequence by the mature spliceosome in the nucleus (Crick 1979; Pozzoli et al. 2002). This multi-megadalton complex itself contains 170 protein

components and various types of snRNA, rivaling the ribosome in molecular complexity (Wahl et al. 2009).

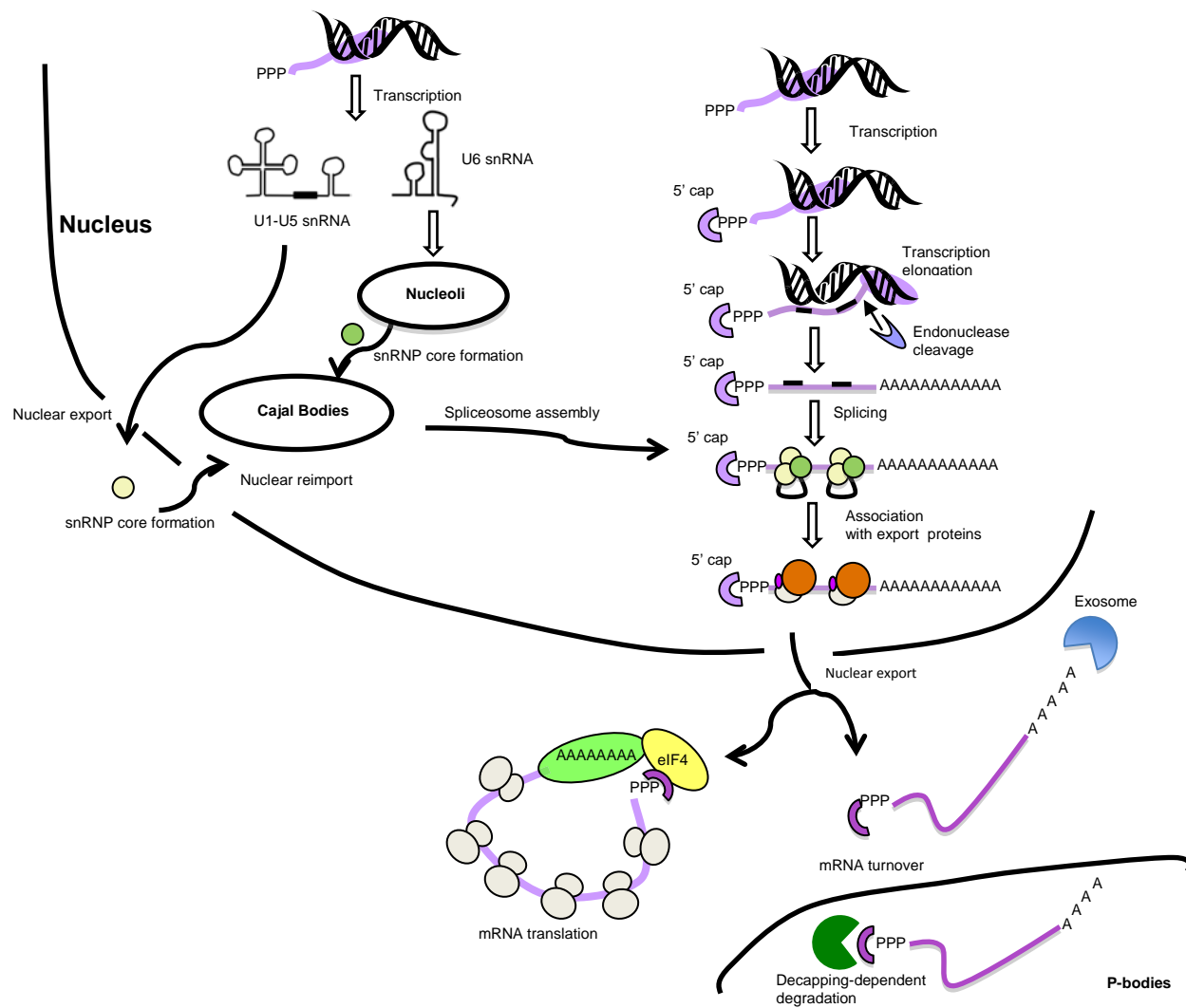
Within the spliceosome, several distinct small nuclear RNP (snRNP) core complexes each contain snRNA organised around specific ring-structured protein assemblies. For those known as U1-, U2-, U4- and U5-snRNPs, these ring scaffolds are provided by members of the Sm protein family (Luhrmann et al. 1990), recruited to their specific snRNA partners in the cytoplasm at a distinct Sm-site of bases (Urlaub et al. 2001; Peng and Gallwitz 2004). The core snRNPs are reimported into the nucleus for further processing and spliceosome assembly (Will and Luhrmann 2001; Patel and Bellini 2008). In contrast, U6 snRNA is first modified within the nucleoli and then engages with a related protein ring, in this case containing Lsm (“Sm-like”) proteins Lsm2-Lsm8. Together with the U1-U5 particles, the U6 snRNP is translocated to Cajal bodies for formation of the U4/U6\*U5 tri-snRNP (Patel and Bellini, 2008). The mature snRNPs eventually assemble on pre-mRNA for intron removal steps (Will and Luhrmann 2001; Patel and Bellini 2008). Following excision of introns, mRNA enters the cytoplasm via the nuclear pore complex to be either translated or degraded. In eukaryotes, two pathways are utilised for mRNA decay: i) 3'-to-5' degradation by the exosome or ii) 5'-decapping, followed by 5'-to-3' exonuclease degradation (Garneau et al. 2007). In either event, decay is initiated by shortening of the poly(A)-tail by deadenylases (Tucker et al. 2001; Garneau et al. 2007; Nissan et al. 2010). Protein machinery required for the 5'-decapping pathway is found enriched in cytoplasmic foci known as processing or P-bodies (Sheth and Parker 2003), which appear to control the sorting and storage of mRNA. Within P-bodies, a specific assembly of Lsm proteins (Lsm1-Lsm7) and ancilliary protein factors expedites

mRNA decapping and subsequent breakdown by ribonuclease (Nissan et al. 2010). While the extent to which mRNA decay is restricted to P-bodies is unclear, sequestered mRNA species are observed to leave P-bodies and may re-enter translation (Brenques et al. 2005).

### 1.1.2 Functional roles for Lsm proteins

Sm and Lsm proteins are known to interact with a diversity of RNA partner species. Specific RNA sequences recognised by various Lsm complexes include the Sm-site (A<sub>2</sub>U<sub>5</sub>GA) (Raker et al. 1999), U-rich stretches at the 3' end of oligoadenylated mRNA (Chowdhury et al. 2007) and RNA polymerase III transcripts, including snRNA (Achsel et al. 1999). Other binding partners include snoRNA (Kufel et al. 2003), P RNA (Kufel et al. 2002), tRNA (Kufel et al. 2002) and rRNA (Kufel et al. 2003). Depletion of Lsm proteins 2-5 and 8 in yeast results in defects in post-transcriptional processing of tRNA, P RNA, rRNA, snoRNA and snRNA precursors (Kufel et al. 2002; Kufel et al. 2003a; Kufel et al. 2003b). Yet only minor (or no) effects are observed on depletion of Lsm6 and Lsm7. A summary of some specific Lsm-ncRNA interactions is presented in Table 1.1.

The Lsm2-Lsm8 complex plays a key role in U6 snRNA maturation, so impacting on the formation of spliceosomal snRNPs (Karaduman et al. 2006). U6 snRNA is the most conserved of all snRNA species and key to the catalytic activity of the



**Figure 1.1|** Lifecycle of mRNA from transcription to decay.

**Table 1.1|** Lsm binding interactions with ncRNA

<b>RNA species</b>	<b>Lsm function</b>	<b>Selected experimental evidence</b>	<b>References</b>
snRNA	assembly, processing	Lsm2-8 binds 3' end of U6 snRNA Lsm2-8 initiates structural rearrangements of U6 snRNA Depletion of Lsm2-Lsm8 results in splicing defects Splicing activity recovered through recombinant Lsm proteins	Achsel et al. 1999 Karaduman et al. 2006; 2008 Mayes et al. 1999 Verdone et al. 2004
	and nuclear localisation	Lsm2-8 localises U6 snRNA to the nucleus	Spiller et al. 2007
tRNA	splicing, 3' and 5' end-processing	Accumulation of unprocessed pre-tRNA and reduced La/Lhp1 binding upon Lsm2-Lsm5 and Lsm8 depletion	Kufel et al. 2002
		Direct interaction of Lsm3 with tRNA and its splicing factors	Fromont-Racine et al. 1997
P RNA	chaperone	Depletion of Lsm2-Lsm5 and Lsm8 reduces pre-PRNA levels	Mayes et al. 1999
		Reduced La/Lhp1 binding upon Lsm2-Lsm5 and Lsm8 depletion	Kufel et al. 2002
		Lsm2-Lsm7 proteins coprecipitate with pre-PRNA	Salgado-Garrido et al. 1999
rRNA	3' and 5' end-processing	Depletion of Lsm2-Lsm5 and Lsm8 delays pre-rRNA processing and increases rRNA decay rate	Kufel et al. 2003b
		Pre-rRNA coprecipitates with Lsm3 but not Lsm1	Kufel et al. 2003b
		Deletion of Lsm6 and Lsm7 genes impairs 20S pre-rRNA processing	Li et al. 2009
snoRNA	3' end-processing	Lsm2-Lsm5 and Lsm8 depletion results in U3-snoRNA degradation and loss of its 3' extended precursor	Kufel et al. 2003a
		Reduced La/Lhp1 binding upon Lsm3 or Lsm5 depletion	Kufel et al. 2003a
		Lsm2-Lsm7 but not Lsm1 or Lsm8 coprecipitate with snR5 snoRNA	Fernandez et al. 2004
		Lsm2-4 and 6-8 but not Lsm5 coprecipitate with U8 snoRNA	Tomasevic and Peculis, 2002

spliceosome (Brow 2002). Newly transcribed U6 pre-snRNA is targeted to the nucleoli following binding of the La protein (Lhp1 in yeast) at its U-rich 3' region (Wolin and Cedervall 2002). Following cyclic phosphorylation, the La (or Lhp1) protein is displaced from the U6 snRNA by the Lsm2-Lsm8 assembly (Achsel et al. 1999; Licht et al. 2008), which induces conformational changes that stimulate binding of a recycling factor (p110 or Prp24) (Rader and Guthrie 2002; Ryan et al. 2002; Karaduman et al. 2006). These conformational changes have been suggested to assist in the formation and recycling of the U4/U6 di-snRNP by exposing single stranded nucleotides for base pairing (Beggs 2005; Karaduman et al. 2006; Karaduman et al. 2008). The Lsm2-Lsm8 complex is also implicated in decapping steps of mRNA in the nucleus. This was suggested by the finding that Lsm6 and Lsm8 were required for nuclear mRNA decay (Kufel et al. 2004).

A specific role for Lsm1-Lsm7 concerns activation of mRNA decay in P-bodies; depletion of individual yeast Lsm proteins results in the accumulation of capped, oligoadenylated mRNA transcripts (Boeck et al. 1998; Bonnerot et al. 2000; Bouveret et al. 2000; Tharun et al. 2000). This specific Lsm complex is recruited alongside other decay factors to U-rich tracts by the protein Pat1, after its displacement of cap-binding translation factors (Parker and Sheth 2007). It is likely that Pat1 and Lsm1-Lsm7 are then involved in subsequent activation of the Dcp1-Dcp2 enzyme (Nissan et al. 2010). A variety of studies have demonstrated the interaction of Lsm1-Lsm7 with decapping factors and exoribonuclease Xrn1 (Bonnerot et al. 2000; Bouveret et al. 2000; Tharun et al. 2000; Coller et al. 2001).

In contrast to its enhancement of mRNA decay, however, the Lsm1-Lsm7 complex can also protect mRNA against 3' end trimming (He and Parker 2001). This may involve steric hindrance of nuclease attack at mRNA locations on which Lsm1-Lsm7 and Pat1 proteins are bound.

### 1.1.3 Specific functions of bacterial Hfq

Bacterial Hfq is observed to interact with bacterial sRNA and so promote the formation of sRNA-mRNA complexes (Wassarman et al. 2001; Gottesman and Storz 2010). Bacterial sRNAs are small non-coding RNA species (50-500 nucleotides), which regulate gene expression via base pairing with mRNA transcripts in a similar mechanism to eukaryotic siRNA or miRNA (Storz et al. 2004; Majdalani et al. 2005; Livny and Waldor 2007; Gottesman and Storz 2010). Hfq controls gene expression either by rearranging the RNA secondary structure, or by increasing the concentration of RNA locally to promote RNA-RNA interactions (Moll et al. 2003; Lease and Woodson 2004; Afonyushkin et al. 2005). A similar mode of binding to sRNA was recently observed for the archaeal Lsm from *Haloferax volcanii* (Fischer et al. 2011).

As for the eukaryotic Lsm proteins, Hfq is required for deadenylation-dependent mRNA decay. An RNase E-Hfq-sRNA complex is thought to function in translational repression and subsequent mRNA destabilisation and degradation (Morita et al. 2005; Morita et al. 2006). Additional functions of Hfq include ATPase activity (Sukhodolets and Garges 2003), cellular stress response and modulation of

virulence in some bacterial strains (Tsui et al. 1994; Fantappie et al. 2009; Liu et al. 2010). Interestingly, the virulence of the multi-drug resistant human pathogen *S. aureus* was decreased in Hfq-deletion strains (Liu et al. 2010).

#### **1.1.4 Lsm proteins in human disease and viral replication**

Aberrations in functions of Lsm proteins have been associated with a number of human diseases. Sm proteins are known to be targeted by auto-antibodies in systemic lupus erythematosus (Lerner and Steitz 1979). In fact, the proteins were first identified in nuclear extracts of a patient suffering from this disease (Lerner and Steitz 1979). A mutation of the SMN gene resulting in diminished assembly of snRNPs is the cause of spinal muscular atrophy (Lefebvre et al. 1995; Wan et al. 2005). Three Lsm proteins (Lsm1, Lsm3 and Lsm7) have now been directly connected to different cancer types. Lsm1 (also named cancer associated Sm-like protein, CaSm) was upregulated in pancreatic, prostate and breast cancer, as well as in several cancer-derived cell lines (Schweinfest et al. 1997; Fraser et al. 2005; Streicher et al. 2007). Remarkably, overexpression of antisense Lsm1 has been demonstrated to promote tumor reduction (Kelley et al. 2000; Kelley et al. 2001; Yan et al. 2006). Elevated levels of Lsm7 have been identified in malignant thyroid tumors, and a reduction in Lsm7 expression was observed in breast cancers (Conte et al. 2002; Rosen et al. 2005). The copy number and expression for the Lsm3 gene was found to be elevated in cervical cancer (Lyng et al. 2006).



Some observed functions for Lsm proteins in viral replication underline their functional diversity, as well as functional conservation throughout the domains of life. Bacterial Hfq was initially described as a host factor required for phage Q $\beta$  replication (Franze de Fernandez et al. 1968). A role for Lsm1 as an effector of HIV replication has been reported (Chable-Bessia et al. 2009). It has also been suggested more recently that positive-strand RNA viruses may directly bind to the host Lsm1-7 protein complex via tRNA-like structures and A-rich stretches, so diverting normal mRNA regulation (Galao et al. 2010). The requirement of host Lsm proteins for the replication of this class of virus has additionally been demonstrated in plant brome mosaic virus (Diez et al. 2000; Noueirry et al. 2003; Mas et al. 2006) and human hepatitis C virus (Scheller et al. 2009).

### **1.1.5 Phylogeny of Lsm protein sequences**

The Lsm proteins recur as molecular chaperones for RNA during the many steps of its processing, sorting and regulation (Beggs 2005). While Sm proteins were first found enriched in a patient with systemic lupus erythematosus (Lerner and Steitz 1979), the wider protein family has since been described across all domains of life (Beggs 2005; Ma et al. 2005). Members include eukaryotic Lsm (Salgado-Garrido et al. 1999), Sm (Kambach et al. 1999) and SMN/Gemin proteins (Selenko et al. 2001; Ma et al. 2005), archaeal Lsm proteins (Collins et al. 2001), the bacterial protein Hfq (Schumacher et al. 2002) and a recently identified Lsm homolog of cyanophage origin (Das et al. 2009). Eukaryotic genomes can contain up to 16 Lsm and 7 Sm proteins (Albrecht and Lengauer 2004), yet 2-3 Lsm proteins are generally encoded in archaea (Collins et al. 2001; Toro et al. 2002; Mura et al. 2003) and only a single

form is evident in bacteria and cyanophage (Schumacher et al. 2002; Das et al. 2009).

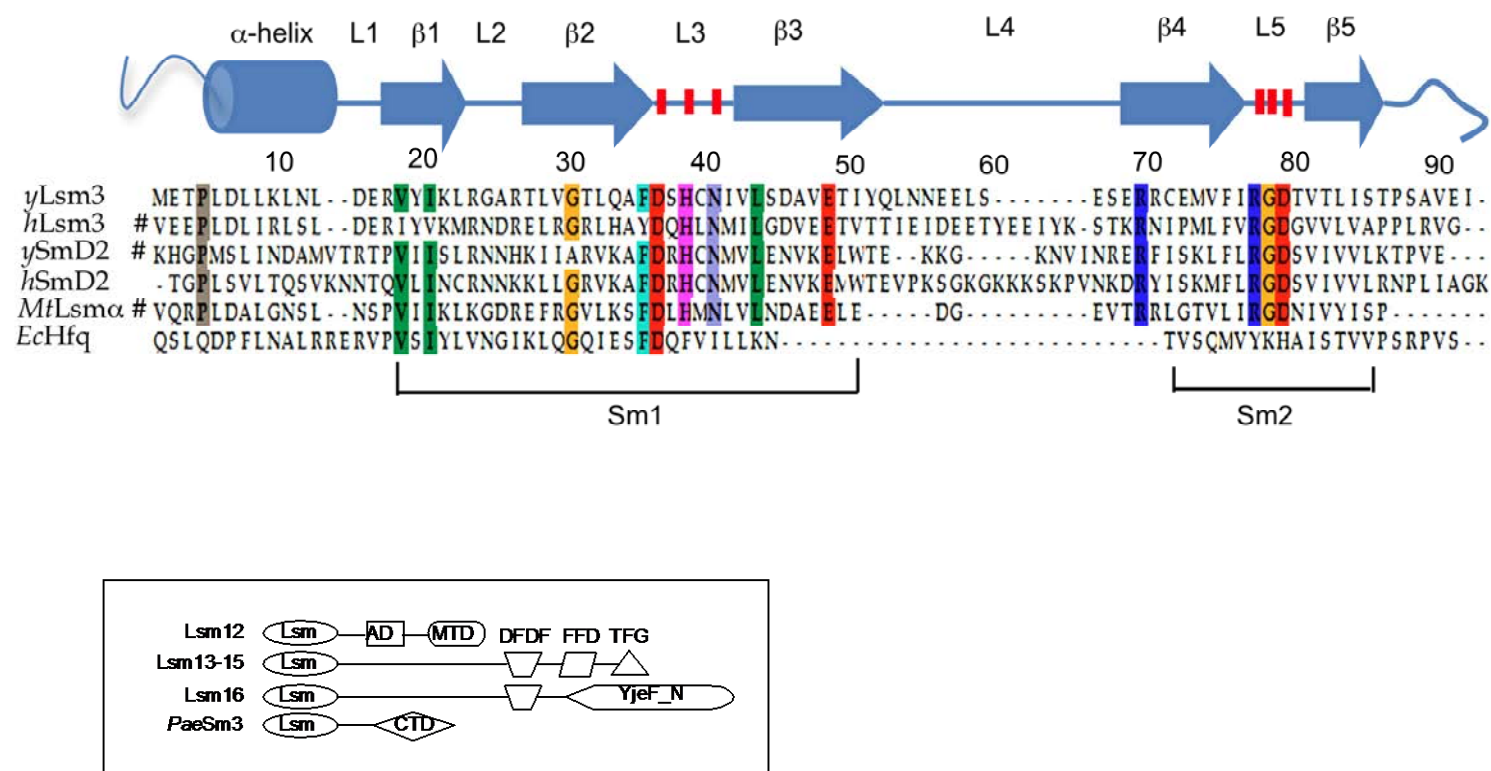
A characteristic feature of the Lsm proteins is their natural tendency to form ring-shaped quaternary complexes, each of a precise composition related to cellular location and RNA target (Beggs 2005; Spiller et al. 2007). In prokaryotes and archaea, homomeric complexes of six or seven Lsm protomers appear to be functional, whilst discrete heteromeric assemblies of seven distinct Lsm proteins are found in eukaryotes. The individual Lsm proteins vary in size from 8-25 kDa (78-240 amino acids); representative sequences are depicted in Figure 1.2. Within each, a bipartite consensus sequence (designated Sm1 and Sm2 motifs) can be identified. These motifs arise from strands  $\beta 1$ - $\beta 3$  and  $\beta 4$ - $\beta 5$  of the core  $\beta$ -sheet structure, respectively. A variable stretch of residues between these conserved segments is created by a surface-exposed interconnecting loop (Kambach et al. 1999; Collins et al. 2001).

The N- and C-terminal tail regions of each Lsm sequence are often highly charged and differ markedly between members; these are considered to provide contact points for additional protein or RNA interactions (Reijns et al. 2008; Reijns et al. 2009; Weber et al. 2010). In the case of the eukaryotic Lsm1 and Lsm4 proteins, these tail segments are notably elongated.

The most highly conserved sequence segments across the Lsm family include specific amino acid sidechains implicated in RNA-binding. These are localised to two specific loop features, as outlined in Figure 1.2. For archaeal and eukaryotic Lsm proteins, sequence motifs Asp-x- $\phi$ - $\phi$ -Asn ( $\phi$  = hydrophobic) and Arg-Gly-(Asp) (Kambach et al. 1999; Collins et al. 2001; Toro et al. 2001) are characteristic of loops L3 and L5, respectively.

In bacterial Hfq, these RNA-binding segments occur as Asp-x- $\phi$ - $\phi$ - $\phi$  (L3) and Tyr-Lys-His (L5) (Schumacher et al. 2002). For this bacterial ortholog, a highly conserved Gln residue on the N-terminal  $\alpha$ -helix is also implicated in RNA-binding (Schumacher et al. 2002).

Overall, the bacterial protein Hfq shows little sequence conservation with its archaeal and eukaryotic orthologs, yet the archaeal and eukaryotic Lsm proteins share some limited sequence similarity (>20 %). The following Lsm-Sm protein paralogs are identifiable: Lsm1-SmB, Lsm2-SmD1, Lsm3-SmD2, Lsm4-SmD3, Lsm5-SmE, Lsm6-SmF, Lsm7-SmG, Lsm8-SmB (Fromont-Racine et al. 2000). These specific sequence relationships suggest the eukaryotic Lsm proteins to have evolved from a common archaeal ancestor in two waves (Khusial et al. 2005; Veretnik et al. 2009). A first gene duplication event likely created eight distinct Lsm proteins, from which later evolved the Sm protein group. The diversity of biological activities of Lsm proteins compared to their more specialised Sm counterparts supports this two-step evolution model (Beggs 2005; Khusial et al. 2005).



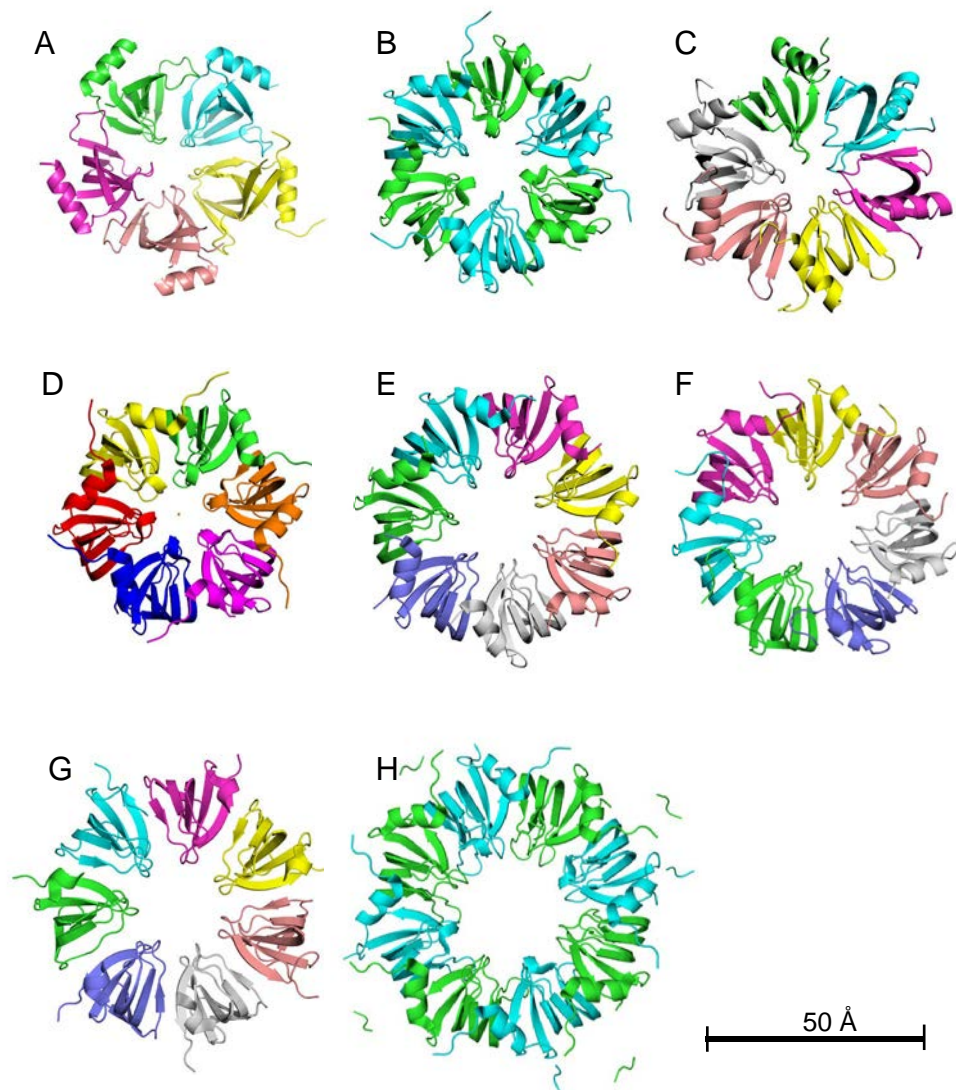
**Figure 1.2]** Structure-based Lsm protein sequence alignment. Sequences displayed are for *S. cerevisiae* Lsm3 (*yLsm3*; PDB 3BW1), *H. sapiens* Lsm3 (*hLsm3*), *S. cerevisiae* SmD2 (*ySmD2*), human SmD2 (*hSmD2*; PDB 1B34), *M. thermoautotrophicum* Lsm $\alpha$  (*MtLsm $\alpha$* ; PDB 1I81) and *E. coli* Hfq (*EcHfq*; PDB 1HK9). Shaded residues represent areas with  $\geq 80\%$  sequence identity. Secondary structure assignment is based on the crystal structure of *yLsm3* (Naidoo et al. 2008). Red bars indicate conserved residues implicated in RNA binding. # indicates additional truncated residues not displayed. Boxed insert shows organisation of other Lsm multidomain proteins: AD, anticodon binding domain; MTD, methyl transferase domain; DFDF, DFDF-x(7)-F containing domain; FFD, Y-x-K-x(3)-FFD-x(IL)-S containing motif; TFG, [RKH]-x(2-5)-E-x(0-2)-[RK]-x(3-4)-[DE]-TFG containing domain. CTD, C-terminal domain2.

The presence of up to three Lsm proteins in archaea, as well as an Hfq-like protein in archaeal *M. jannaschii*, further supports a common ancestor of eukaryotic and archaeal Lsm proteins (Fischer et al. 2011).

A few multidomain proteins incorporating Lsm components have been observed (summarised, Figure 1.2). Lsm12 includes t-RNA and methyltransferase domains (Albrecht and Lengauer 2004), and Lsm13, Lsm14 and Lsm15 all contain a central DFDF-x(7)-F domain (Albrecht and Lengauer 2004; Anantharaman and Aravind 2004). Lsm16 features a remarkably disrupted Lsm variant (lacking both the N-terminal  $\alpha$ -helix and a complete  $\beta$ 4 strand) in addition to FDF and YjeF-N domains (Albrecht and Lengauer 2004; Tritschler et al. 2007). This protein is suggested to be dimeric in solution (Ling et al. 2008). The archaeal protein *PaeSm3* contains an Lsm-like domain in addition to a C-terminal domain of unknown function adopting an  $\alpha/\beta$ -fold (Mura et al. 2003).

### 1.1.6 Structures of Lsm protein ring complexes

Crystal structures of Lsm and Sm proteins from diverse sources today provide many high-resolution views of the ring morphology of their assemblies. As shown in Figure 1.3, Lsm rings have been observed to range between 58-75 Å in diameter and to contain a central pore of 6-15 Å. Some crystal structures solved to date (Table 1.2) have been obtained in the presence of specific RNA partners. The recent solving of the human U1-snRNP structures containing the Sm assembly bound together with U1



**Figure 1.3** | Selected crystal structures solved for Lsm assemblies. A) Pentamer, cyanophage ECX21941 (PDB 3BY7) 60 Å ring, 9 Å pore. B) Hexamer, of *C. parvum* Lsm5 (PDB 3PGG) 60 Å ring, 10 Å pore. C) Hexamer, *S. aureus* Hfq (PDB 1KQ1) 65 Å ring, 11 Å pore. D) Hexamer, *A. fulgidus* Sm2 (PDB 1LJO) 58 Å ring, 6 Å pore. E) Hexamer, *A. fulgidus* Sm2 (PDB 1LJO) 58 Å ring, 6 Å pore. F) Heptamer, *A. fulgidus* Sm1 (PDB 1I4K) 65 Å ring, 13 Å pore. G) Heptamer, *M. thermoautotrophicum* Lsm $\alpha$  (PDB 1I81) 65 Å ring, 10-15 Å pore. F) Heptamer, *S. cerevisiae* SmF (PDB 1N9R) 65 Å ring, 10-15 Å pore. H) Octamer, *S. cerevisiae* Lsm3 (PDB 3BW1) 75 Å ring, 15 Å pore.

**Table 1.2|** Crystal structures solved for Lsm assemblies (to 2010)

	Protein <sup>a</sup>	PDB ID	Resolution (Å)	Organism	Reference
Hexameric	HsSmD3B	1D3B	2.00	<i>H. sapiens</i>	Kambach et al. 1999
	SaHfq	1KQ1	1.55	<i>S. aureus</i>	Schumacher et al. 2002
	SaHfq*	1KQ2	2.71	<i>S. aureus</i>	Schumacher et al. 2002
	AfSm2	1LJO	1.95	<i>A. fulgidus</i>	Toro et al. 2002
	EcHfq	1HK9	2.15	<i>E. coli</i>	Sauter et al. 2003
	PaHfq	1U1S	1.60	<i>P. aeruginosa</i>	Nikulin et al. 2003
	PaHfq	1U1T	1.90	<i>P. aeruginosa</i>	Nikulin et al. 2003
	MjSm	2QTX	2.50	<i>M. jannaschii</i>	Nielsen et al. 2007
	CpLsm5	3PGG	2.14	<i>C. parvum</i>	Vedadi et al. 2007
	AHfq	3HFN	2.31	<i>Anabena sp.</i>	Boggild et al. 2009
	EcHfq*	3GIB	2.40	<i>E. coli</i>	Link et al. 2009
	SHfq	3HFO	1.30	<i>Synchocystis sp.</i>	Boggild et al. 2009
	PaH57THfq	3INZ	1.70	<i>P. aeruginosa</i>	Moskaleva et al. 2010
	PaH57AHfq	3M4G	2.05	<i>P. aeruginosa</i>	Moskaleva et al. 2010
	BsHfq	3HSB	2.20	<i>B. subtilis</i>	Someya et al. 2010 <sup>3</sup>
Heptameric	MtLsm $\alpha$	1I81,	2.00,	<i>M.</i>	Collins et al. 2001
		1MGQ	1.70	<i>thermoautotrophicum</i>	
	PaeSm1	1I8F	1.75	<i>P. aerophilum</i>	
	AfSm1	1I4K	2.50	<i>A. fulgidus</i>	
	AfSm1*	1I5L	2.75	<i>A. fulgidus</i>	Toro et al. 2001
	MtLsm $\alpha$	1JBM	1.85	<i>M.</i>	Mura et al. 2003b
				<i>thermoautotrophicum</i>	
	PaeSm1	1JRI	1.75	<i>P. aerophilum</i>	
	PaeSm1	1LNx	2.05	<i>P. aerophilum</i>	
	PabSm1	1H64	1.90	<i>P. abysii</i>	Thore et al. 2003
	PabSm1*	1M8V	2.60	<i>P. abysii</i>	Thore et al. 2003
	PaeSm3	1M5Q	2.00	<i>P. aerophilum</i>	Mura et al. 2003a
	PaeSm1	1LOJ	1.90	<i>M.</i>	Mura et al. 2003b
				<i>thermoautotrophicum</i>	
	ScSmF	1N9R	2.80	<i>S. cerevisiae</i>	
	ScSmF	1N9S	3.50	<i>S. cerevisiae</i>	
Other	SsSm1	1TH7	1.68	<i>S. solfataricus</i>	Kilic et al. 2005
	U1-snRNP*	3CW1	5.49	<i>H. sapiens</i>	Pomeranz Krummel et al. 2009
	U1-snRNP*	3PGW	4.40	<i>H. sapiens</i>	Weber et al. 2010
	CphLsm	3BY7	2.60	<i>Cyanophage</i>	Das et al. 2009
	ScLsm3	3BW1	2.50	<i>S. cerevisiae</i>	Naidoo et al., 2008
	PfuQ8TZN2 <sup>b</sup>	1YCY	2.80	<i>P. furiosus</i>	Huang et al. 2004 <sup>c</sup>

<sup>a</sup>Proteins are named by the first letters of the species, followed by the type of protein. Asterisked entries indicate structures solved in the presence of RNA.

<sup>b</sup>Hypothetical protein adopting an Lsm fold.

<sup>c</sup>Structure deposited without supporting publication.

snRNA and proteins U1-70K and U1-A have been significant and exciting advances (Pomeranz Krummel et al. 2009; Weber et al. 2010). These provide the first molecular detail of L/Sm rings bound to the highly intertwined protein-RNA network within RNP complexes.

Within the various Lsm ring assemblies, each protomer occurs as a highly bent five-stranded antiparallel  $\beta$ -sheet overlaid in most cases by an N-terminal  $\alpha$ -helix (Figure 1.4A). The pronounced twist of the  $\beta$ -sheet aligns strand  $\beta 5$  against  $\beta 1$ , so forming an SH3-type barrel loosely related to the OB-fold (Kambach et al. 1999; Collins et al. 2001). Strands  $\beta 4$  and  $\beta 5$  each present on opposite ends of the module, so providing interaction sites for adjacent Lsm subunits via  $\beta 4$ - $\beta 5'$  pairing (Figure 1.4). Stacking of five to eight protomers in such a manner ultimately results in the formation of the toroid assembly characteristic of all Lsm assemblies (Figure 1.4).

Within this ring organisation, the N-terminal amphipathic  $\alpha$ -helices of each Lsm component are gathered across one face of the toroid, from which also project the unstructured N- and C-terminal extensions. The opposite face of the ring, named the distal face, is predominantly composed of residues of the variable loop L4 segments. All the Lsm ring structures (across eukarya, archaea and bacteria) reveal clusters of positively charged residues lining the internal pore, as well as pronounced positive elements on the distal face (Toro et al. 2001; Brennan and Link 2007; Naidoo et al. 2008).

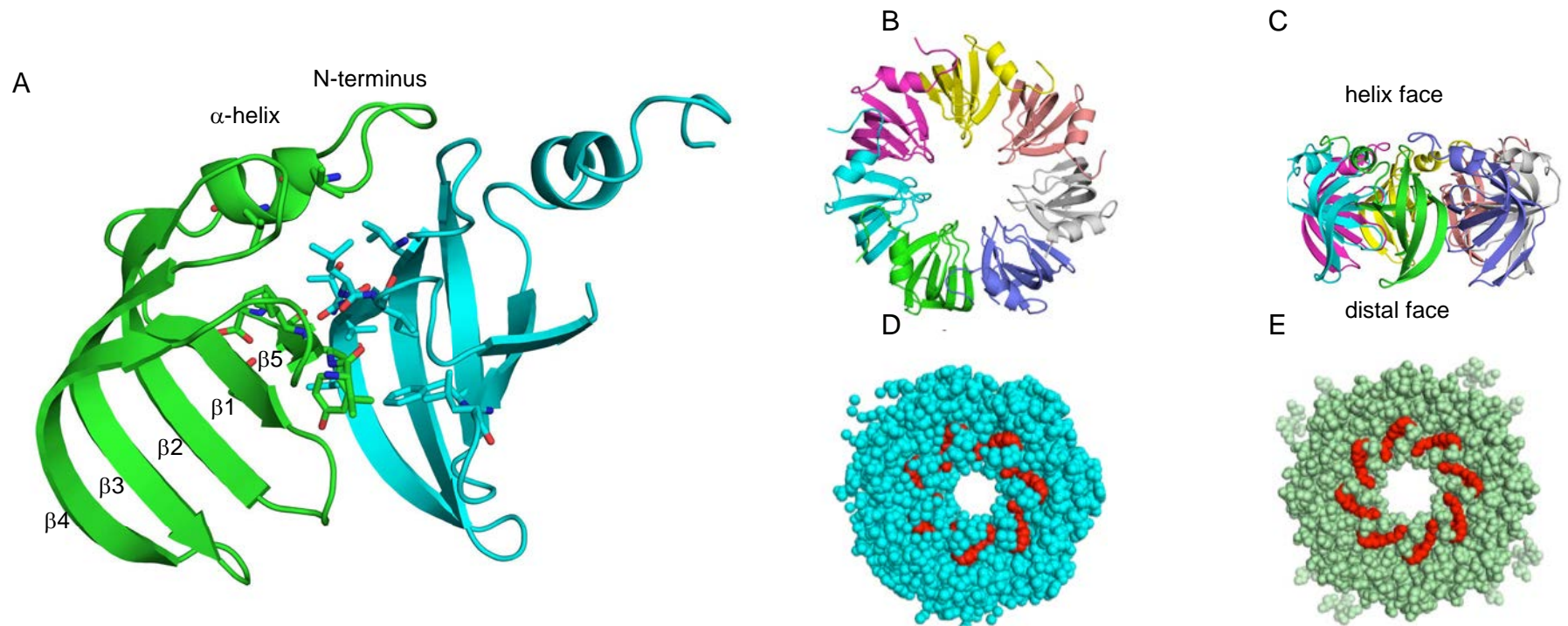


The body of structural data adds to biochemical understanding concerning L/Sm-RNA interactions, and distinct RNA sites within the protein oligomer. These include i) a binding site within the lumen of the ring, ii) an external contact site on the helix face and iii) residues located on the distal face of the complex (Figure 1.4). The first of these sites engages residues from loops L3 and L5, contributed from all Lsm components to create a nucleotide-binding pocket running around the inner rim (Weber et al. 2010). The specific architecture and repeated circular location of these specific, highly conserved, sidechains enables one nucleotide base to be bound per L/Sm protomer. Crystal structures of archaeal and bacterial Lsm complexed with RNA clearly show the oligonucleotides to be threaded around this rim of the toroid (Toro et al. 2001; Schumacher et al. 2002). Each binding “slot” allows specific base stacking to a hydrophobic sidechain of loop L3, as well as contact with the signature Arg residues of loop L5 and H-bonding with Asn residues (strand  $\beta$ 4). Further electrostatic contacts (involving conserved Asp (strand  $\beta$ 2), Arg (loop L5) and Gly (loop L5) residues) enhance the stability of the Lsm-RNA complex (Toro et al. 2001). Figure 1.5 displays these relevant binding interactions for U<sub>5</sub> within the lumen site of archaeal A/Sm1.

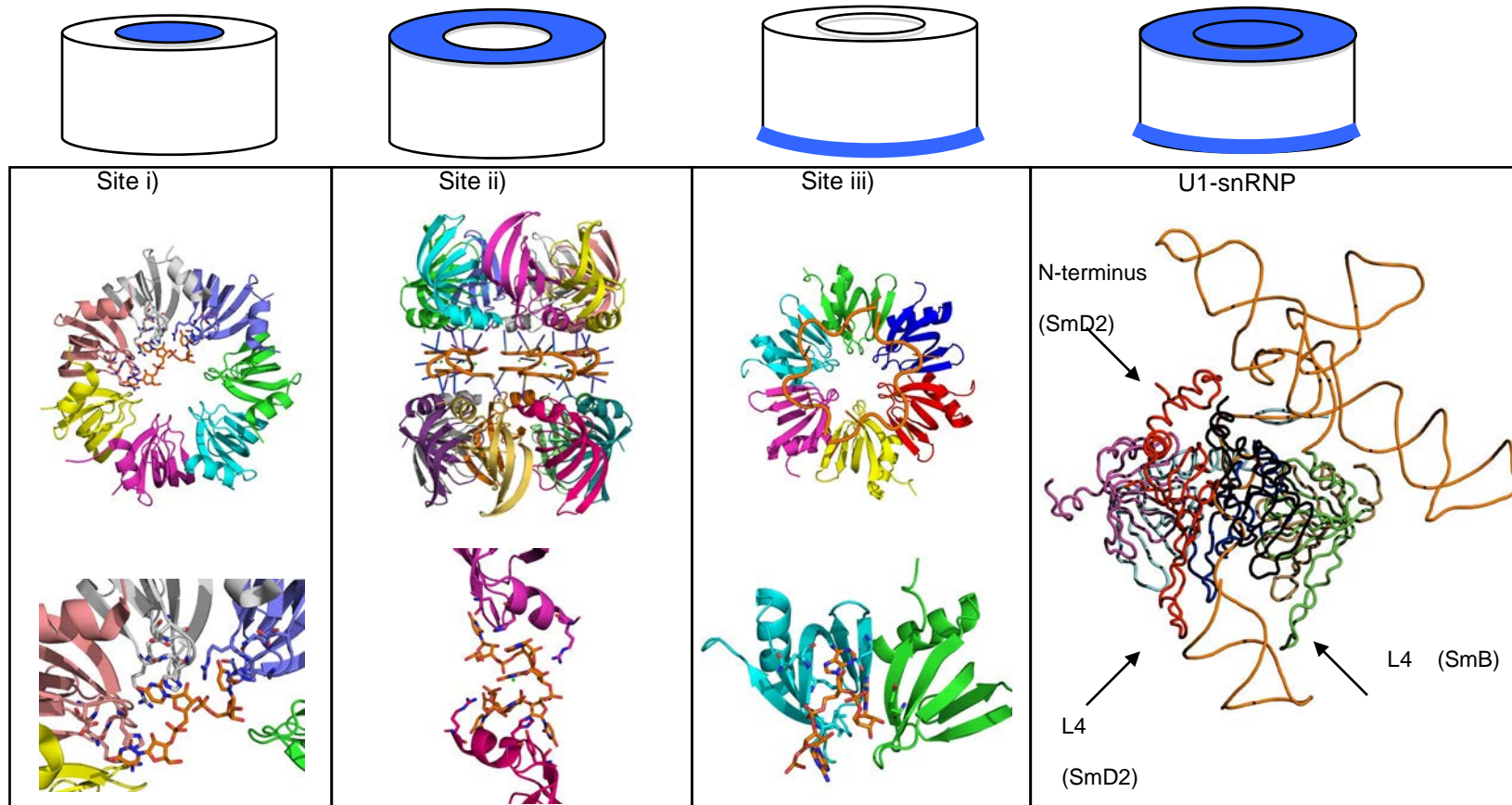
An external contact site for RNA at the helix face of the Lsm toroid (site ii) is suggested by the crystal structure of *Pa*Sm1 bound with U<sub>7</sub> oligonucleotide (Thore et al. 2003). In this case, each of two sandwiched Lsm rings engage two nucleotides at the N-terminal  $\alpha$ -helix (Arg, His) and strands  $\beta$ 2 (Tyr) via base stacking and H-bonding.

A third distinct RNA-binding site (iii) is likely unique to the bacterial Hfq assembly, and its tripartite form has been detailed in the crystal structure of Hfq bound to poly(A) RNA (Link et al. 2009). The protein Hfq engages poly(A) sequences on its distal face via specific residues exposed from strands  $\beta 2$  and  $\beta 4$ . There is, however, no evidence for poly(A) binding by eukaryotic Lsm proteins. In the structure of the Hfq/RNA complex, RNA contacts include electrostatic interactions from Lys (strand  $\beta 2$ ) and Gln (strand  $\beta 4$ ) sidechains, as well as stacking of bases between Tyr, Leu (strand  $\beta 2$ ) and Leu and Ile (strand  $\beta 2'$ ) of adjacent subunits. It is in this region of the toroid that sequence variability of the loop L4 across the Lsm family results in non-conservation of distal face chemistry, so explaining the unique binding properties of Hfq.

Within the crystal structures of the human U1-snRNP complex, multiple RNA interactions made by the ring of Sm proteins include binding sites i) and ii) outlined above (Weber et al. 2010). However, the U1-snRNP structure also clearly demonstrates the role of the Sm sequence extensions and loop regions as additional interaction sites, particularly the C-terminal extensions of SmD3 and SmB. In the lumen of the toroid (i.e. site i), snRNA threads to stack single nucleotides of the Sm site against the key loop L3 and L5 residues, notably the aromatic sidechains. From the helix face of the ring are projected residues of the N-terminal  $\alpha$ -helix and loop L3 of SmD2, forming an external contact site (reminiscent of site ii) that guides the snRNA into the ring pore.



**Figure 1.4|** Lsm fold and quaternary structure. Ribbon diagrams of *MtLsmα* (A, B; PDB 1I81) are displayed. A) Dimer interface of *MtLsmα*. Chain A is represented in green, chain B in blue. Residues involved in hydrophobic packing at the dimer interface (Chain A: Ile27, Val77, Tyr78 of chain A; Chain B: Leu 30, Phe36, Leu66, Val69, Ile71) are shown in stick representation. B-C) Top and side view of heptameric *MtLsmα*. D) Homo-heptameric *MtLsmα* (PDB 1I81) E) homooctameric yeast Lsm3 (PDB 3BW1). Space filled models highlight in red conserved residues implicated in RNA binding: Asp in  $\beta 2$ , Asn in L3, Arg and Gly in L5.



**Figure 1.5]** Three general sites for RNA binding within specific examples of Lsm complexes. Site i) AfSm1 (PDB 1I5L) bound to U<sub>5</sub> RNA viewed from helix face. Site ii) Two PaSm1 (PDB 1M8V) heptamers are bridged by a uridine heptamer. Site iii) EChfq (PDB 3GIB) bound to poly(A) viewed from distal face. U1-snRNP) Figure includes side view of the Sm-core of the human U1 snRNP structure (PDB 3PGW).

Residues from the loop L2 regions of SmD1 and SmD2 appear to guide RNA out from the Sm ring. Protruding beyond the distal face, residues of the elongated L4 loops of SmD2 and SmB provide another important interaction point to clamp and secure a stem-loop of the snRNA.

The majority of crystal structures of Lsm obtained to date portray the hexa- and heptameric protein assemblies that correspond to fully functional homomeric or heteromeric protein groupings. It is, for instance, assumed that complexes of SmD1-SmD2, SmD3-SmB and SmE-SmF-SmG can exist independently in the cytoplasm, yet rearrange into mixed heptamers in the presence of RNA during snRNP formation (Peng and Gallwitz 2004). However, a few crystal structures suggest that other compositions, e.g. pentamers and octamers, may be stable for eukaryotic Lsm proteins (Naidoo et al. 2008; Das et al. 2009). While it is currently not clear if these organisations are peculiar to recombinant preparations of the Lsm family, they suggest possibilities for a variety of multimeric assemblies *in vivo*. Our own interaction studies indicate that Lsm assemblies may be relatively dynamic in solution, providing capacity to engage in alternative protein partnerships and stable groupings (Sobti et al. 2010).

## 1.2 Scope of the thesis

The heptameric complexes Lsm1-7 involved in cytoplasmic mRNA degradation and Lsm2-8 required for mRNA splicing and processing of ncRNAs in the nucleus are the two best-characterised Lsm complexes in eukarya. Both complexes share six out

of seven Lsm subunits, yet, the exchange of a single protein, Lsm1 against Lsm8, has striking effects on not only localisation but also function and RNA target of the assembly. This suggests specific functions of individual Lsm proteins in mixed complexes. Two recently solved crystal structures of the human U1 snRNP (Pomeranz Krummel et al. 2009; Weber et al. 2010) containing the homologous Sm-complex bound to U1 snRNA provided the first structural evidence for a role of specific Lsm proteins in a heteromeric complex in RNA recognition. Other roles of individual Lsm proteins in heteromeric complexes including P-body assembly and nuclear assembly of Lsm complexes have been suggested (Reijns et al. 2008; Reijns et al. 2009). However, despite their ubiquitous role in RNA processing events, little is known about target discrimination and RNA-binding mechanisms by eukaryotic Lsm complexes. Hence, a model system is required to probe the biological complexity of these assemblies.

Lsm protein complexes are studied by a variety of researchers worldwide using different approaches. Research groups from the University of Edinburgh and the University of the Health Sciences, Bethesda, successfully extracted native heteromeric Lsm complexes from yeast cultures and were able to probe their interactions with protein and RNA partners (Chowdhury and Tharun 2009; Reijns et al. 2009).

A second strategy to produce mixed Lsm complexes suitable for structural and functional studies includes co-expression of individual components. An approach featuring co-expressed dimer and trimer combinations of Lsm proteins based on subcomplexes seen in the homologous Sm complex was successfully applied by a

research group from the Paul Scherer Institute in Zurich to express these subcomplexes in *E. coli* (Zaric et al. 2005).

The protein structure group at Macquarie University has developed a unique approach to study mixed Lsm protein complexes using polyproteins. Lsm polyproteins feature two Lsm proteins connected by a flexible linker sequence allowing co-expression of simplified dual Lsm complexes. Previous studies established that when individually expressed, yeast Lsm3 forms discrete stable octamers (Naidoo et al. 2008). Lsm polyprotein complexes though assemble into a mixture of different solution states (Sobti et al. 2010).

In this thesis, I aim to investigate if complexes of these polyproteins can be isolated as stable multimeric forms capable of mimicking complex heteromeric Lsm ring structures *in vivo*. Functional and structural characterisation of these complexes will be attempted to aid our understanding of Lsm complexes *in vivo*.

Preparations of eukaryotic Lsm complexes are known to form a diversity of different quaternary structures. Individually expressed, Lsm3 assembles into octamers. Preparation of heteromeric complexes by co-expression or using our polyprotein approach results in mixtures of different oligomeric species including trimers, hexamers, octamers, nonamers and 16mers (Zaric et al. 2005; Sobti et al. 2010). Hence, in this study, the solution behaviour of expressed Lsm polyproteins is of interest, in particular if Lsm polyproteins can form discrete stable assemblies.

Lsm complexes have a characteristic ring-shaped quaternary structure. Thus, to be useful as mimics of native Lsm complexes, Lsm polyproteins are required to form mixed ring structures in solution. Consequently, polyprotein preparations need to be tested for the presence of ring quaternary structures.

Simplified heteromeric Lsm ring assemblies would provide valuable tools in probing the contribution of individual Lsm proteins on RNA target discrimination by heteromeric Lsm complexes *in vivo*. It would be of great interest if Lsm polyprotein complexes can be utilised to reveal specific binding determinants for RNA related to individual Lsm proteins.

Two crystal structures of the Lsm homologous Sm complex are available, providing structural insights in RNA recognition by Lsm complexes. However, these crystal structures were solved at 5.5 Å and 4.4 Å resolution, respectively, and high resolution structures are exclusively available for homomeric Lsm assemblies mostly of archaeal and bacterial origin. Thus, to improve our understanding of eukaryotic Lsm biology, there is a need for structural information on mixed eukaryotic Lsm complexes. Hence, one aim of this work will be to test if Lsm polyproteins can provide appropriate candidates for structure determination by X-ray crystallography.



In summary, specific aims of my thesis were:

- 1) Solution-state characterisation of simplified heteromeric Lsm complexes to determine their individual oligomeric state,
- 2) Determination of conditions to form discrete stable Lsm polyprotein complexes,
- 3) Probing the actual nature of quaternary structures formed in solution,
- 4) Elucidation of interaction properties of these simplified complexes to examine specific binding determinants within the complex interactions of *in vivo* species, and
- 5) Testing if the polyprotein system can provide appropriate candidates for crystallisation of mixed Lsm assemblies.



## 2 Materials and Methods

### 2.1 Materials

#### 2.1.1 Reagents

All reagents used in this work were of analytical grade or higher. General reagents are summarised in Table 2.1. Purified water from a MilliQ system (Millipore) was used throughout.

#### 2.1.2 Growth media and buffers

Luria-Bertani (LB) medium (Sambrook and Russell 2001): tryptone (10 g/L), yeast extract (5 g/L) and NaCl (10 g/L) were dissolved in purified water and autoclaved.

ZYP-rich medium (Studier 2005): To prepare 1 L of ZYP-rich media, ZY media (925 ml),  $\text{MgSO}_4$  (1 ml, 1 M), 50x 5052 media (20 ml) and 20x NPS (50 ml) solution were combined. Individual components ZY, 50x 5052 and 20x NPS were prepared as follows: i) ZY media: tryptone (10 g/L), yeast extract (5 g/L) were weighed in and dissolved in purified water (925 ml); ii) 50x 5052:  $(\text{NH}_4)_2\text{SO}_4$  (66 g),  $\text{KH}_2\text{PO}_4$  (136 g),  $\text{Na}_2\text{HPO}_4$  (142 g) were added to purified water (900 ml); iii) 20x NPS: glycerol (250 g), glucose (25 g) and  $\alpha$ -lactose (100 g) were sequentially dissolved in purified water (730 ml). ZY, 50x 5052 and 20x NPS media were autoclaved prior to use.

**Table 2.1|** Reagents used in this work listed with commercial sources

Reagent	Supplier
acetic acid, glacial	BDH
acetone	BDH
acrylamide/Bis solution	Bio-Rad
agar (bacteriological)	Astral
ampicillin	Astral
APS (ammonium peroxodisulphate)	BDH
benchmark protein ladder	Invitrogen
Bio-Rad Protein Assay	Bio-Rad
bromophenol blue	Progen
chloramphenicol	Astral
chymotrypsin	Promega
citric acid	Sigma-Aldrich
cobalt chloride	Sigma-Aldrich
coomassie brilliant blue	BDH
dimethyl sulfoxide	Sigma-Aldrich
dithiothreitol (DTT)	BDH
DNase 1	Roche
ethylenediamine tetraacetic acid	Spectrum
ethanol, absolute	Fronine
formic acid	Sigma-Aldrich
glucose	Astral
glycerol	Astral
HEPES	Astral
hydrochloric acid	BDH
imidazole	Sigma-Aldrich
iodoacetamide	Sigma-Aldrich
IPTG (Isopropyl-B-D-thiogalactopyranoside)	ICN Biomedicals
isopropanol	BDH
lactose	BDH
lysozyme	Astral
magnesium chloride	BDH
magnesium formate	Hampton Research
methanol	Biolab Scientific
nitric acid	BDH
paraffin-liquid	BDH
pepstatin A	Sigma-Aldrich
polyethylene glycol 3500	Hampton Research
polyethylene glycol 6000	Hampton Research
polyethylene glycol 8000	Hampton Research
protease inhibitor cocktail	Sigma-Aldrich
RNase A	Roche
RNaseZap®	Ambion
sodium acetate	BDH
sodium chloride	Astral
sodium citrate tribasic dihydrate	Fluka
sodium dodecyl sulphate	Amresco

**Table 2.1** (cont.)

<b>Reagent</b>	<b>Supplier</b>
sodium formate	Ajax
sodium hydroxide pellets	Astral
surfactant P20	GE-Healthcare
TEMED	Bio-Rad
thrombin protease	Sigma-Aldrich
Tris	Astral
trypsin	Promega
urea	BDH
yeast extract	Oxoid

Crystal screening suites	
Additive Screen	Hampton Research
Natrix screen	Hampton Research
Peg/Ion	Hampton Research
Salt Rx	Hampton Research
JCSG Core suites I-IV	QIAGEN
JCSG + suite	QIAGEN
Nextal classics I	QIAGEN
Nextal classics II	QIAGEN
pH clear I – II	QIAGEN
AmSO <sub>4</sub> suite	QIAGEN

SOC medium (Sambrook and Russell 2001): tryptone (20 g/L), yeast extract (5 g/L), NaCl (0.58 g/L), KCl (0.18 g/L), MgCl<sub>2</sub> (0.95 g/L), MgSO<sub>4</sub> (1.20 g/L) were combined and were dissolved in purified water and autoclaved. After autoclaving, filter sterilised glucose solution (1 M) was added to a final concentration of 2 % (v/v).

Minimal medium for selenomethionine labelling: prepackaged kits of selenium-enriched defined growth media (M9 SeMet, Shanghai Medicilon) were used as per manufacturer's instructions.

Liquid growth media were autoclaved or filter sterilised and stored at 4 °C before use. For solid media, 1.5 % agar was added prior to autoclaving. For plasmid

maintenance, ampicillin (100 mg/ml) and chloramphenicol (25 mg/ml) were added to growth media. Compositions of growth media are summarised in Table 2.2.

**Table 2.2|** Composition of growth media and buffers employed in protein purification

Buffer	Composition
LB medium	tryptone (10 mg/ml), yeast extract (5 mg/ml), NaCl (10 g/L)
ZY medium	tryptone (10 g/L), yeast extract (5 g/L)
50 x 5052	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (66 g/L), KH <sub>2</sub> PO <sub>4</sub> (136 g/L), Na <sub>2</sub> HPO <sub>4</sub> (142 g/L)
20 x NPS	glycerol (250 g), glucose (25 g), $\alpha$ -lactose (100 g)
SOC medium	tryptone (20 mg/ml), yeast extract (5 mg/ml), NaCl (0.58 g/L), KCl (0.95 g/L), MgCl <sub>2</sub> (1.20 g/L), glucose (3.6 g/L)
bacterial lysis buffer	Tris (20 mM, pH 8.0), NaCl (1 M), imidazole (10 mM), glycerol (10 % v/v)
buffer A	Tris buffer (20 mM, pH 8.0), NaCl (400 M), imidazole (10 mM), glycerol (10 % v/v)
buffer B	Tris buffer (20 mM, pH 8.0), NaCl (400 M), imidazole (500 mM), glycerol (10 % v/v)
buffer C	Tris buffer (20 mM, pH 8.0), NaCl (400 M), glycerol (10 % v/v)
buffer D	Tris buffer (20 mM, pH 8.0), NaCl (200 mM)

### 2.1.3 Bacterial strains plasmids and RNA

The bacterial strains used in this work including their genotypes and sources are listed in Table 2.3. Hfq-deletion strains (MRE5, C41  $\Delta$ hfq) used in this study were a gift from Professor Beggs (University of Edinburgh).

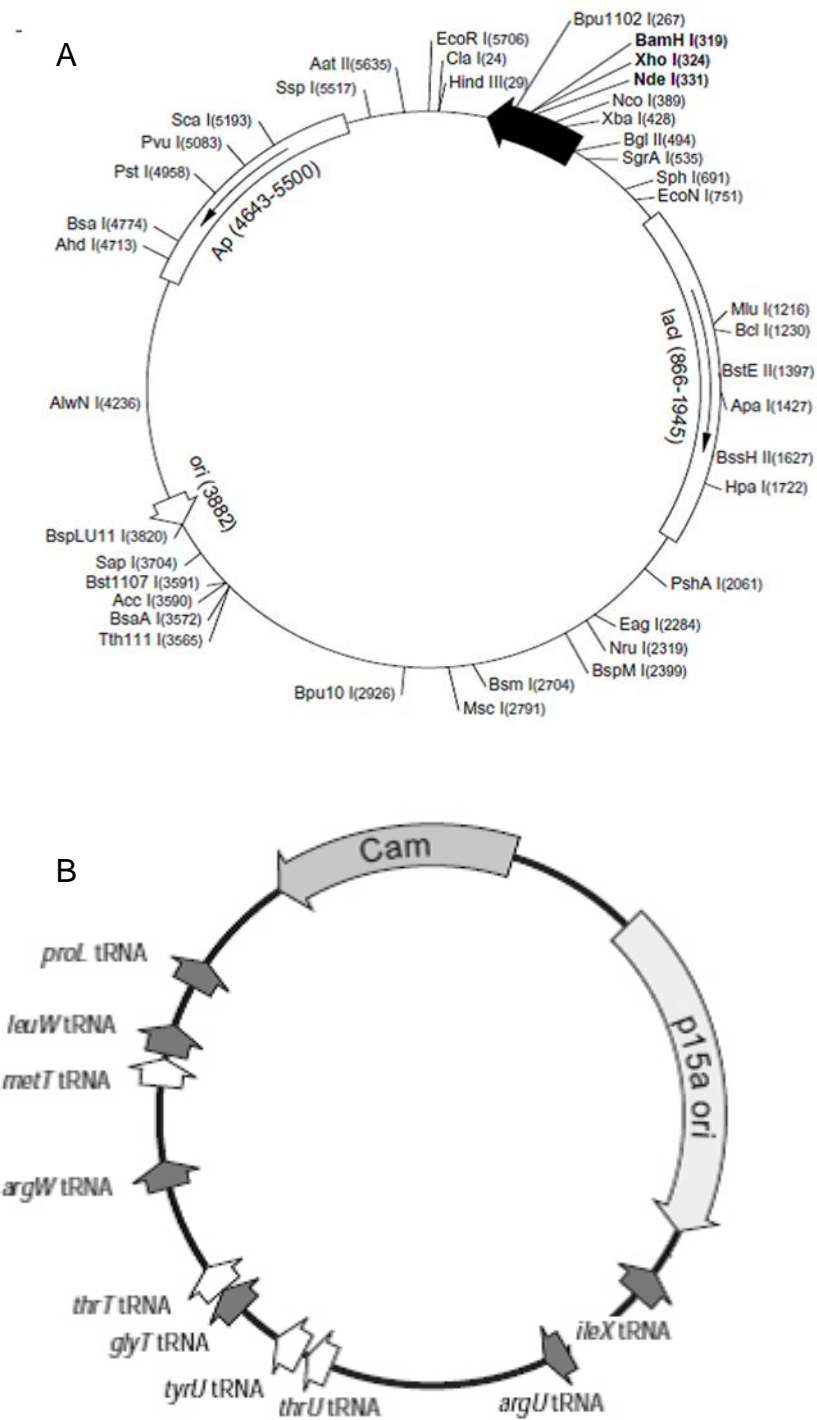
**Table 2.3|** Bacterial strains used to generate protein material

Strain	Genotype	Source
TOP10	<i>F<sup>-</sup> mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZ ΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (Str<sup>R</sup>) endA1 nupG</i>	Invitrogen
Fusion-blue	<i>endA1, hsdR17 (rK12<sup>-</sup>, mK12<sup>+</sup>), supE44, thi-1, recA1, gyrA96, relA1, lac F'[proA+B<sup>+</sup>, lacIqZΔM15::Tn10(tetR)]</i>	Merck
BL21 (DE3) pLysS	<i>F- ompT gal dcm lon hsdSB(rB<sup>-</sup> mB<sup>-</sup>) λ(DE3) pLysS(cmR)</i>	Sigma-Aldrich
BL21 Rosetta 2 (DE3)	<i>F- ompT hsdSB (rB<sup>-</sup> mB<sup>-</sup>) gal dcm (DE3) pRARE23 (Cam<sup>R</sup>)</i>	Merck
MRE5 (C41 Δhfq)	<i>F ompT gal dcm lon hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) λ(DE3) Δhfq (and an uncharacterized mutation Miroux et al., 1996)</i>	Gift from Prof. Beggs <sup>a</sup>

<sup>a</sup> University of Edinburgh, Wellcome Trust Centre for Cell Biology

A pET15B vector (Novagen), designed to achieve high levels of expression from a T7 RNA polymerase promoter (Studier and Moffatt 1986) was used for expression of Lsm polyproteins. This vector confers ampicillin resistance to the transformed cells and encodes sequences for an additional N-terminal hexahistidine sequence, followed by a thrombin cleavage site prior to the inserted sequence (Figure 2.1 A). Three cloning sites are incorporated at the DNA level. The target gene was inserted between *NdeI* and the *BamHI* site.

Since *LSM* genes 1-4 contain up to 13 rare codons, the pRARE plasmid (Merck) encoding for rare tRNAs was isolated from BL21 Rosetta 2 cells (DE3, Merck) for later transformation into expression hosts. This construct allows enhanced expression of genes isolated from organisms with AT or GC rich genomes with a corresponding codon bias (Baca and Hol 2000). Supplied codons include rare tRNAs required for the translation of Arg, Ile, Gly, Leu and Pro residues. The construct carries native



**Figure 2.1** Plasmid maps for A) *E. coli* expression vector pET15B (Mierendorf and Yeager 1984). B) pRARE plasmid for the expression of rare tRNAs (Baca and Hol 2000).



tRNA promoters and a chloramphenicol resistance marker. A map of this vector is displayed in Figure 2.1 B.

RNA oligonucleotides were supplied in HPLC purified and desalted form (Dharmacon).

## **2.2 Methods**

### **2.2.1 Preservation of bacterial strains**

For a period of up to two weeks, bacterial cultures were stored on solid media plates at 4 °C. Glycerol stocks were prepared in order to store *E. coli* cells indefinitely. 10 ml of LB medium containing the appropriate antibiotics were inoculated. Cells were grown over night at 37 °C. The resulting cell suspension was pelleted by centrifugation (3000 x g, 10 min, 4 °C) and resuspended in growth medium containing 25 % glycerol. This suspension was transferred to cryo-tubes (Greiner) and stored at -80 °C.

### **2.2.2 Cloning of Lsm polyproteins**

In previous work, the genes encoding Lsm1, Lsm2, Lsm3 and Lsm4 were isolated and amplified by Meghna Sobti (Sobti et al. 2010). The LSM4 and LSM1 genes and the LSM2 and LSM3 genes were covalently linked utilising partially overlapping primer based-PCR (Te'o et al. 2000). The resulting dual gene cassettes were cloned into a pET15B vector, which carries an N-terminal His<sub>6</sub>-tag followed by a thrombin

cleavage site in frame with the inserted genes. As a consequence, gene products will be expressed from this vector with an additional cleavable N-terminal affinity-tag. Furthermore, the pET15B system places the genes of interest under control of a T7  $\Phi 10$  promoter and thus requires T7 RNA polymerase for transcription (Studier and Moffatt 1986; Rosenberg et al. 1987). The use of bacterial DE3 expression strains carrying the T7 RNA polymerase gene under control of the lac operon allows the tight regulation of recombinant protein expression using lactose and lactose analogues like IPTG (Studier and Moffatt 1986; Rosenberg et al. 1987). Glycerol stocks of Fusion-blue and B121 Rosetta 2 cells containing polycistronic Lsm gene cassettes were made and used for starter cultures in this study.

### **2.2.3 Plasmid isolation from *E. coli***

Lsm expression plasmids and pRARE constructs were isolated from bacterial cells using a commercial kit (QIAprep Spin Miniprep Kit, QIAGEN) according to the manufacturer's instructions. Isolated plasmids were collected in supplied elution buffer and stored at -20 °C.

### **2.2.4 Preparation of chemically-competent cells**

In the case of the MRE5 cell line competent cells were prepared in the laboratory as follows (Inoue et al. 1990): An LB-agar plate was spread with MRE5 cells and grown overnight (37 °C). Single colonies from the plate were used to inoculate LB media (5 ml) containing MgSO<sub>4</sub> (20 mM) and shaken (250 rpm, 37 °C, 12 h). The resulting culture was transferred to fresh LB media (250 ml) containing MgSO<sub>4</sub> (20

mM) and was incubated (200 rpm, 23 °C) until an OD<sub>600</sub> of 0.4-0.6 was reached. Cells were sedimented by centrifugation (3000 x g, 10 min, 4 °C) and resuspended in ice-cold sterile TB buffer (80 ml). The suspension was incubated on ice (10 min) prior to sedimenting the cells as before. Cell pellets were resuspended in an ice-cold solution (20 ml) KCl<sub>2</sub> (250 mM), CaCl<sub>2</sub> (15 mM), MnCl<sub>2</sub> (55 mM) and PIPES buffer (10 mM, pH 6.7). DMSO (1.5 ml) was added and the solution was incubated on ice (10 min) prior to flashfreezing of 200 µl aliquots in liquid nitrogen. Resulting chemically-competent cells were stored at -80 °C.

### **2.2.5 Transformation for plasmid propagation**

For propagation and maintenance of plasmid DNA, pet15B vectors containing *Lsm* gene constructs and pRARE plasmids were transformed into commercial chemically competent cells (Top10, Invitrogen; Fusion-Blue, Clontech) cells. Plasmid DNA (2 ml) was added to thawed cells (50 ml) and mixed gently. After cooling on ice, cells were heatshocked (45 s, 42 °C) and again incubated on ice (2 min). SOC medium (450 ml) was added and cells were recovered with shaking at 37 °C (1.5 h, 140 rpm). An aliquot (50 ml) of this cell suspension was spread on a pre-warmed LB-agar plate containing the appropriate antibiotics. The remainder was spread on a second LB-agar plate. Both plates were incubated at 37 °C (12 h) and single colonies from these cultures were isolated and restreaked on LB-agar plates for a second incubation at 37 °C (12 h). Scrapings from the resulting cultures were used to prepare glycerol stocks or for plasmid isolation (Section 2.1.3).

## 2.2.6 Transformation for protein expression

pET15B vectors containing Lsm expression cassettes and pRARE constructs containing rare tRNA genes were co-transformed into chemically-competent MRE5 (Section 2.1.3) cells. Aliquots (2 ml) of both Lsm expression plasmids and pRARE were added to thawed MRE5 cells (50 ml) and were transformed as described in Section 2.2.5. Glycerol stocks of the expression host carrying the plasmids were made and used to inoculate expression cultures.

## 2.2.7 Expression of Lsm polyproteins

To exclude contamination of Lsm polyprotein preparations with the bacterial Lsm paralogue Hfq, the Hfq-deficient cell line MRE5 (Section 2.1.3) was used for protein expression. Up to five clones from each transformation were screened for protein expression. Single colonies were restreaked on LB-agar plates containing ampicillin (100 µg/ml) and chloramphenicol (25 mg/ml) and incubated at 37 °C (12 h). Resulting cultures were used to prepare glycerol stocks and to optimise the expression protocols with regards to temperature and induction method in small-scale (2 ml) expression trials.

### 2.2.7.1 *Small-scale expression trials*

For small-scale expression screening, two methods of induction were utilised, IPTG induction or auto-induction, allowing the expression of recombinant proteins under control of the lac operon (Baneyx 1999):

- i) IPTG induced protein expression

For IPTG induction, glycerol stocks or scrapings from LB-agar plates were used to inoculate LB medium (2 ml). Suspension cultures were grown in conical tubes (50 ml) to an optical density at 600 nm ( $OD_{600}$ ) of 0.4-0.6 before IPTG was added (0.1 mM or 1 mM). To assess effects of induction method and temperature on protein expression levels, cultures were subsequently incubated at (250 rpm) at two different temperatures (16 °C, 24 h; 25 °C, 12 h).

#### ii) Protein expression by auto-induction

Similar to IPTG induced cultures glycerol stocks or scrapings from LB-agar plates were used to inoculate ZYP-rich medium (2 ml). These cultures were incubated (250 rpm) at two temperatures (16 °C, 30 h; 25 °C, 24 h).

Cells were harvested by centrifugation (3000 x g, 10 min, 4 °C). To identify optimal expression conditions, samples (1 ml) were taken, spun down (3000 x g, 10 min) and analysed by SDS-PAGE.

### **2.2.7.2 Large scale protein expression**

Analogous to small-scale trials, IPTG induction and auto-induction were utilised in large scale:

#### i) IPTG induced expression

Bacteria were spread on LB-agar plates containing ampicillin (100 mg/ml) and chloramphenicol (25 mg/ml) and incubated at 37 °C (12 h). These starter cultures were transferred into baffled Erlenmeyer-flasks (2 L) containing LB medium (1 L). Bacterial suspension cultures were grown to an  $OD_{600}$  of 0.4-

0.6 at 37 °C (250 rpm) before IPTG (0.1-1 mM) was added and the cells were incubated at 25 °C (250 rpm, 12 h).

ii) Protein expression by auto-induction

Glycerol stocks or scrapings from LB-agar plates were used to inoculate ZYP-rich medium (0.5 L). Protein expression using the auto-induction method (Studier 2005) was performed in ZYP-rich medium at 25 °C (250 rpm, 24 h).

Cells were harvested by centrifugation (3000 x g, 10 min, 4 °C) and resuspended in bacterial lysis buffer (Table 2.2) prior to storage at -80 °C.

### **2.2.7.3 Protein expression for L-selenomethionine incorporation**

For preparation of SeMet derivatised proteins, M9 SeMet media kits (Shanghai Medicillon) were used according to the manufacturers instructions. Cultures (1 L) were shaken (37 °C, 250 rpm) in baffled flasks (2 L). At an OD<sub>600</sub> of ~1.2, the temperature was reduced to 20 °C and protein expression was induced by addition of IPTG (1 mM). Following incubation over night, the bacteria were harvested by centrifugation (3000 x g, 10 min, 4 °C).

### **2.2.8 Protein extraction**

Protein extracts were prepared using BugBuster reagent (Novagen), for small-scale expression trials, or high-pressure homogenisation using an EmulsiFlex<sup>®</sup>-C3 (Avestin) homogeniser of cell pellets in bacterial lysis buffer (Table 2.2) containing RNase A (100 µg/ml), DNase I (10 µg/ml), protease inhibitor cocktail (1 % v/v), and

lysozyme (1 mg/ml) for large-scale expression. For the former, cell pellets were mixed with the BugBuster reagent as per manufacturers instructions. Soluble and insoluble fractions were separated by centrifugation (10000 x g, 10 min, 20 °C). For high-pressure homogenisation cell pellets were resuspended in bacterial lysis buffer (40 ml) and passed through the instrument twice at 15000 PSI. Cell lysate was cleared by centrifugation (20000 x g, 20 min, 4 °C) and subsequent filtration using a 0.2 µm syringe filter.

## **2.2.9 Purification of recombinant Lsm polyproteins**

### **2.2.9.1 Chromatography equipment, media and columns**

Chromatography operations were performed on an Äkta Explorer system (GE-Healthcare). Columns (1 ml) prepacked with Ni-sepharose matrix (HisTrap HP, GE-Healthcare) were used for immobilised metal affinity chromatography (IMAC) purification of His<sub>6</sub>-tagged proteins. For analytical size exclusion chromatography, Superdex 200 10/300 GL (GE-Healthcare) columns (24 ml) and for preparative scale HiLoad 16/60 Superdex 200 pg (GE-Healthcare) columns (125 ml) were employed, respectively.

### **2.2.9.2 IMAC purification of Lsm polyproteins**

Recombinant His<sub>6</sub>-tagged proteins were purified by immobilised metal affinity chromatography (IMAC). Cleared cell lysate was loaded (1 ml/min) on a prepacked IMAC column pre-equilibrated in buffer A (Table 2.2). The ratio of cell lysate to affinity matrix volumes was chosen such that the amount of recombinant protein exceeded the dynamic binding capacity of the column (40 mg/ml). This ensured the

high purity of Lsm protein preparations required for functional and structural studies. Following lysate application, adsorbed proteins were washed for 60 column volumes (cv) with buffer A containing 40 mM imidazole to remove loosely bound contaminants. Elution was performed using a step gradient from 40 mM to 500 mM imidazole (buffer C, Table 2.2). Elution fractions from affinity purification were pooled and subjected to size exclusion chromatography. A typical elution profile for IMAC purification is shown in Figure 2.2.

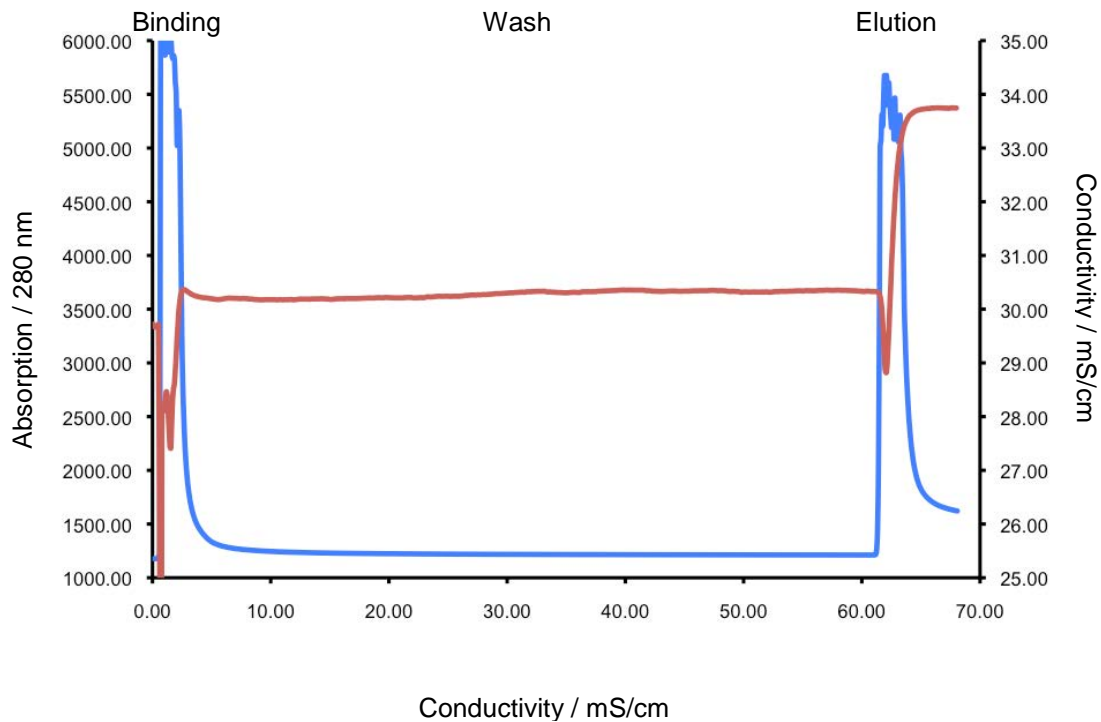
### **2.2.9.3 Preparative size exclusion chromatography (SEC)**

All Lsm protein preparations were further purified by preparative SEC to ensure a uniform oligomeric distribution of the preparations. Lsm samples (~2 ml) were applied to a HiLoad 16/60 Superdex 200 pg column in buffer C (1 ml/min). Eluted protein fractions were collected as 0.5 ml fractions and relevant fractions were pooled.

### **2.2.10 Protein concentration and storage**

Purified proteins were concentrated to > 10 mg/ml using centrifugal ultrafiltration devices (Millipore) with a molecular weight cut-off of 10 kDa. Protein concentrations were estimated spectrophotometrically using theoretical extinction coefficients of individual proteins at 280 nm (from ExPASy proteomics server) from amino acid composition (Pace et al. 1995) or using a Bradford Assay (Bio-Rad Protein assay). Absorbance ratios at 280 nm and 260 nm were calculated to assess the purity of Lsm protein preparations.





**Figure 2.2|** Exemplary IMAC purification of the polyprotein Lsm[4+1]. Absorbance at 280 nm is displayed in blue. The red line represents the conductivity.

$A_{260}/A_{280}$  ratios can serve to estimate the purity of protein preparations. On average, the extinction coefficients of nucleic acids at wavelengths of 260 nm are twice their extinction coefficient at 280 nm, while the absorption of pure protein at 260 nm is approximately 0.6 times its absorption at 280 nm (Warburg and Christian 1941; Glasel 1995). Therefore,  $A_{260}/A_{280}$  ratios of  $\sim 2.0$  indicate pure nucleic acid, while ratios of  $\sim 0.6$  suggest pure protein.  $A_{260}/A_{280}$  ratios are relatively robust against deviations from standard extinction coefficients of individual proteins (Glasel 1995) and therefore allow an easy assessment of the presence of nucleic acids in protein preparations. Approximate concentrations of nucleic acid and protein in a mixture can be obtained from Warburg-Christian tables.

UV-absorptions (220-350 nm) were recorded on a Cary 100 UV-Visible spectrophotometer (Varian) or a Nandrop 1000 instrument (Thermo Scientific). For Bradford assays, a standard curve was made using BSA (0.2, 0.4, 0.6, 0.8, 1.0 mg/ml). BSA samples were mixed (20 ml) with dye reagent (1 ml) and the absorbance at 595 nm was recorded. The resulting absorbances were plotted against protein concentrations and a linear fit was used to determine the concentrations of Lsm samples.

For storage, protein samples were transferred to thin-walled PCR tubes (30 µl aliquots) and were flash frozen in liquid nitrogen. Frozen samples were stored at -80 °C.

## **2.2.11 Protein analysis**

### **2.2.11.1 Protein electrophoresis**

Protein samples were visualised by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) at various stages of protein preparation according to Sambrook and Russell (2001). Prior to loading samples on the gel, they were mixed in a 1:1 ratio with 2 x loading dye and boiled (2 min). The gels used consisted of a 15 % polyacrylamide separating gel and a 5 % polyacrylamide stacking gel. Applied samples were separated for 45 min (200 V) in a Tris-glycine buffer system (Laemmli 1970). Following electrophoresis, gels were incubated in fixing solution (10 min, Table 2.4). Protein bands were visualised with staining solution containing

Coomassie Brilliant Blue G-250 (0.5 %) and destained with acetic acid (10 %). Table 2.4 summarises compositions of buffers and solutions utilised in SDS-PAGE.

**Table 2.4|** SDS-PAGE buffer compositions

Buffer	Composition
2 x loading dye	SDS (4 % w/v), glycerol (20 % w/v), DTT (200 mM), Tris buffer (100 mM, pH 6.8) and bromophenol blue (0.2 % w/v)
running buffer	Tris buffer (25 mM), glycine (250 mM), SDS (10 % w/v)
fixing solution	ethanol (50 % v/v), acetic acid (10 % v/v)
staining solution	coomassie brilliant blue (0.25 % w/v), ethanol (10 % v/v), acetic acid (10 % v/v)
destaining solution	acetic acid (10 % v/v)

#### **2.2.11.2      *Analytical size exclusion chromatography***

For analytical SEC, Lsm proteins were injected (0.5 ml/min) on a Superdex 200 10/300 GL (GE-Healthcare) column (24 ml) that was pre-equilibrated in the appropriate buffer. The separation range of the chosen matrix is 10-600 kDa. In order to estimate the molecular weight of SEC fractions,  $K_{av}$  values were calibrated using Gel Filtration LMW and HMW Calibration Kits (GE-Healthcare). Calibration was performed in three subsequent runs. In the first run, Blue Dextran 2000 (2000 kDa) was injected to determine the void volume. Following the determination of the void volume, protein standards were injected in two consecutive runs containing i) a mixture of ferritin (440 kDa), conalbumin (75 kDa), carbonic anhydrase (29 kDa) and RNase A (13.7 kDa); and ii) a mixture of aldolase (158 kDa), ovalbumin (43 kDa), RNase A (13.7 kDa) and aprotinin (6.5 kDa). The log of the molecular weight of the

calibration proteins was plotted over the  $K_{av}$  values of the standard proteins. The values were calculated from:

$$K_{av} = \frac{V_E - V_0}{V_C - V_0} \quad (2-1)$$

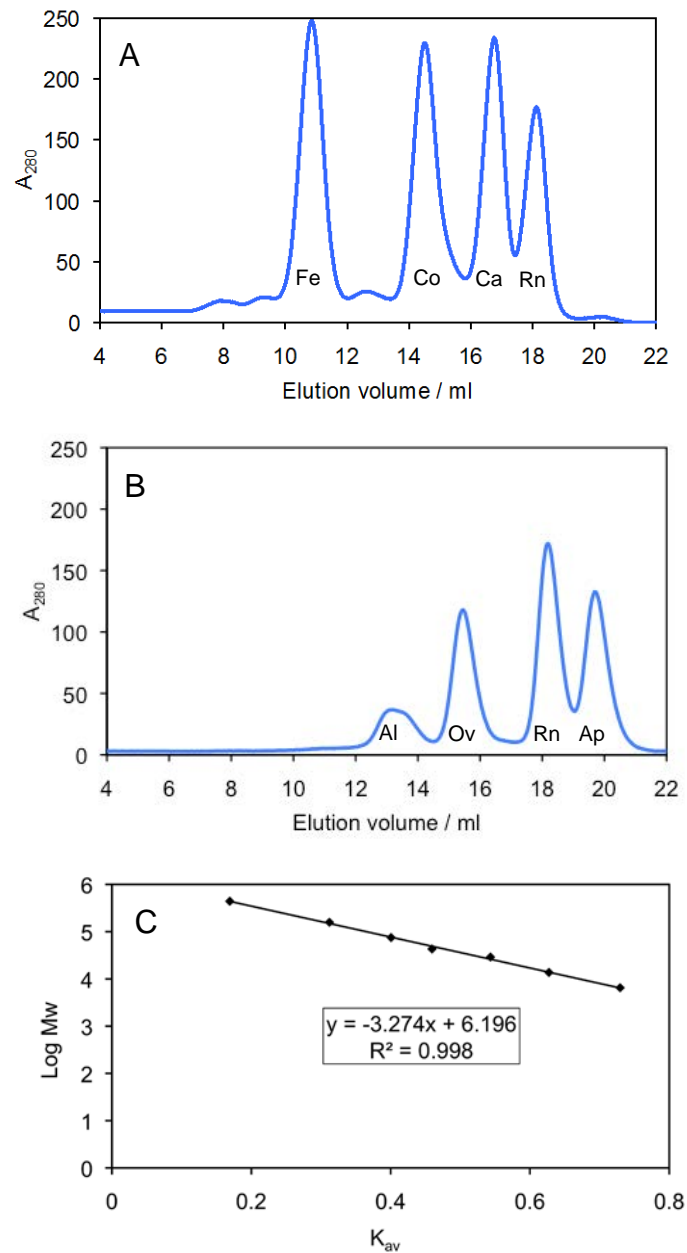
where  $V_E$  is the recorded elution volume,  $V_0$  the void volume (8.3 ml) and  $V_c$  the column volume (24 ml).

Molecular weights of proteins of unknown mass were approximated from a linear fit to this calibration curve. Figure 2.3 shows the resulting graphs.

#### **2.2.11.3      *SEC coupled to multi angle laser light scattering*** **(SEC-MALLS)**

For accurate mass determination independent of protein shape, SEC-MALLS was carried out. An analytical SEC column was placed in line with a triple-angle static light scattering detector (MiniDAWN TREOS, Wyatt) and a refractive index detector (RID-10A, Shimadzu). Data analysis was performed using Astra 5.4.3 (Wyatt). Prior to molecular weight determination of sample proteins, the triple detectors were normalised using monomeric solutions of Ovalbumin (3 mg/ml) or RNase A (3 mg/ml).

Molecular weight determination via SEC-MALLS using the Astra method relies on the Rayleigh relationship (Slotboom et al. 2008):

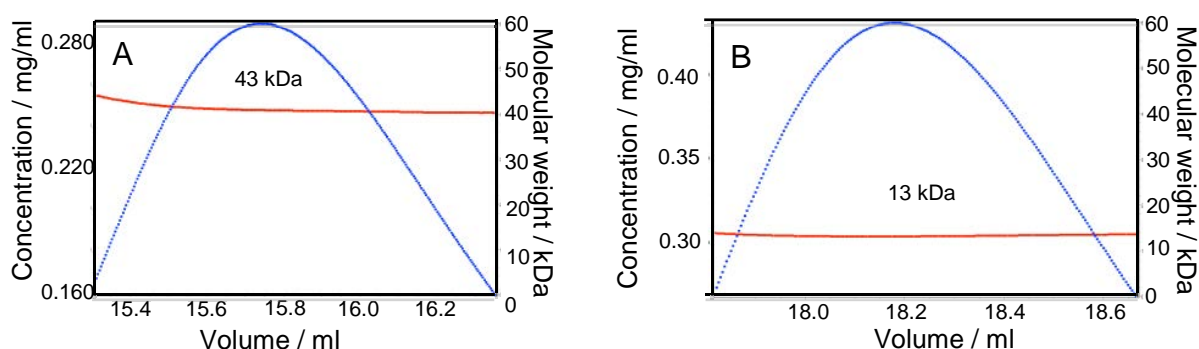


**Figure 2.3|** Superdex 200 10/300 GL (GE-Healthcare) calibration. Two sets of calibration proteins were run in Tris buffer (20 mM, pH 8.0), NaCl (400 mM), glycerol (10 % v/v) at 0.5 ml/min. A) Ferritin (Fe), conalbumin (Co), carbonic anhydrase (Ca), RNase A (Rn). B) aldolase (Al), ovalbumin (Ov), RNase A (Rn), aprotinin (Ap). C) Plot of log of the molecular weight of standard proteins over corresponding  $K_{av}$  values.

$$\Delta LS = \left(\frac{I_{\theta}}{I_0}\right)_{solution} - \left(\frac{I_{\theta}}{I_0}\right)_{buffer} = K \left(\frac{dn}{dc}\right)^2 M_w C \quad (2-2)$$

wherein  $I_{\theta}/I_0$  describes the ratio of the intensities of scattered light at angle  $\theta$ ,  $K$  an instrument calibration constant,  $C$  the concentration of the macromolecule in solution,  $M_w$  the molecular weight of the macromolecule and the  $dn/dc$  the refractive index increment.

In the above-described setup,  $\Delta LS$  and  $C$  can be readily determined from the light scattering and refractive index detectors. Hence, with knowledge of the exact  $dn/dc$  of the protein, its molecular weight can be determined. In general,  $dn/dc$  values of 0.185-0.190 are assumed for proteins in aqueous buffers (Wen et al. 1996). However, due to individual buffer compositions,  $dn/dc$  values of proteins can deviate significantly from these standard values (Slotboom et al. 2008; Liu et al. 2011). As a consequence, protein  $dn/dc$  values were determined experimentally in buffer C (Table 2.2) by adjusting  $dn/dc$  values until correct masses were obtained for the standard proteins. Figure 2.4 shows SEC-MALLS data obtained for calibration proteins Ovalbumin (43 kDa) and RNase A (13.7 kDa) using an optimised  $dn/dc$  of 0.165 in buffer C (Table 2.2). Masses obtained for the calibration proteins are within a 5 % error margin of the expected molecular weight. Accordingly, a  $dn/dc$  of 0.165 was used for the determination of molecular masses of all proteins in this solvent.



**Figure 2.4|** Determination of refractive index increments of proteins in buffer C (Table 2.2). Monodisperse standard proteins Ovalbumin (43 kDa) and RNase A (13.7 kDa) were utilised to empirically determine  $dn/dc$  value of protein in buffer C. Optimised  $dn/dc$  values allowed determination of the molecular weight of the calibration proteins within 5 % of expected masses. Red lines represent the molecular weight corresponding to each point in the chromatogram. A) SEC-MALLS of Ovalbumin yields a molecular weight of 41 kDa. B) SEC-MALLS of RNase A yields a molecular weight of 13 kDa.

#### 2.2.11.4 Small angle X-ray scattering (SAXS)

SAXS experiments were performed in collaboration with Dr. Grant Pearce (University of Canterbury, Christchurch) on the SAXS/WAXS beamline at the Australian Synchrotron (Melbourne, Australia). A monochromatic X-ray beam with a wavelength of  $1.03320 \text{ \AA}$  was used. The range of momentum transfer  $s = 4\pi/\lambda \sin\theta$  was adjusted to  $0.010 \leq s \leq 0.618 \text{ \AA}^{-1}$  for experiments on this beamline. All protein samples were adjusted to 5 mg/ml concentration and cleared by centrifugation ( $10000 \times g$ , 2 min). Subsequently, protein samples were injected in 100  $\mu$ l aliquots on an analytical SEC column (Superdex 200 5/150 GL,  $V_c = 3 \text{ ml}$ , F

= 0.2 ml/min) in line with SAXS data collection. For RNA-binding studies using SAXS, RNA ( $U_{10}$ ) was incubated with protein at a twofold excess over protein concentration (5 mg/ml; 53 mM). Excess RNA was removed from samples on the Superdex 200 5/150 GL matrix utilised prior to SAXS data collection.

SAXS data were collected at room temperature using an exposure time of 2 s per image. Averaged scattering from all buffers was recorded and used for background subtraction. Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> samples were measured in buffer C, while Lsm[2+3]<sub>8</sub> was analysed in buffer D. Data reduction and background subtraction were performed using the SAXSID15D software (Cookson et al. 2006). The resulting data were processed using the ATSAS package (Konarev et al. 2003). Guiner approximations were carried out using PRIMUS (Konarev et al. 2003). Electron distribution functions were calculated in GNOM (Svergun 1992) and normalised with respect to  $p(r)$  by division of all data points through their respective  $p(r)$  maxima. Kratky plots of all samples were calculated manually by plotting  $I(s) * s^2$  over  $s$ . For comparison, Kratky plots were normalised with respect to  $I(s) * s^2$  by division of the absolute  $I(s) * s^2$  values of the polyprotein scattering datasets by their respective maximum values. For determination of molecular envelopes, DAMMIN (Svergun 1999) was employed. Eight models were calculated for each Lsm sample using spacegroup P1. Resulting models were averaged and sorted using DAMAVER (Volkov and Svergun 2003). Solution scattering was evaluated using CRY SOL (Svergun et al. 1995).



## 2.3 Lsm-RNA interaction studies

### 2.3.1 Surface plasmon resonance (SPR)

Measurements were performed on a BIAcore 2000 instrument (GE-Healthcare) using streptavidin coated sensor chips (GE-Healthcare) at 25 °C. All buffers were prepared RNase free and were degassed using RNase free syringes (Terumo) and 0.2 µm syringe filters (sartorius stedim). 2'-O-bis(2-acetoxyethoxy)methyl protected, biotinylated RNA oligonucleotides were purchased from Dharmacon. Prior to biosensor analysis, RNA oligonucleotides were deprotected by incubation in acetate-TEMED buffer at 60 °C for 2 h. Deprotection buffer was subsequently removed in a vacuum centrifuge. Deprotected RNA was diluted in buffer D containing 0.05 % surfactant P20 (GE-Healthcare), 62.5 µg/ml BSA and immobilised to a density of 25 RU on the sensor chip at a flow rate of 10 µl/min.

Lsm protein complexes were injected (2 min) at 25 °C in different concentrations (0.1-10 mM) over blank and RNA coated flow cells as duplicates in random order at a flow rate of 50 µl/min using the kinject function. All experiments were performed as duplicates. Regeneration of the chip surface was achieved by injection (2 min) of 2 M NaCl. For experimental design and data analysis, the BIACORE 2000 Control Software (GE-Healthcare) and the BIAevaluation 3.0.2 program (GE-Healthcare) were used. Data processing included subtraction of reference channel signals from the responses of channels with immobilised RNA, curve alignment and curve fitting. Sensorgrams are presented as normalised responses by division of all datapoints by the molecular weight of the analyte.

### 2.3.2 Isothermal titration calorimetry (ITC)

ITC experiments were conducted on a VP-ITC instrument (MicroCal) stationed at the Biomolecular Interaction Centre (University of Canterbury, Christchurch) following training and advice by Dr. Richard Hutton (University of Canterbury, Christchurch). Temperature was kept constant at 25 °C and stirrer speed was set to 307 rpm. For data analysis, the VPViewer2000 (MicroCal) and Origin 7 SR4 v7.0552 (Origin) programs were used. All sample buffers were prepared in DEPC treated purified water to eliminate any RNase activity. For the same reasons, all test tubes used were incubated at 180 °C (12 h) prior to RNA-binding experiments. Furthermore, reaction cell and syringe were washed with RNaseZAP® (Ambion) to remove RNase activity. A thorough wash with RNase free water and degassed sample buffer further ensured the removal of the decontaminant. Protein samples were characterised and buffer exchanged by SEC prior to ITC analysis. Protein concentrations were adjusted to 10-30 µM and 1.4 ml of the resulting protein solution were filled into the reaction chamber. 600 µl of a 100-200 µM RNA solution were introduced into the injection syringe. Following baseline stabilisation, 36 injections of 7.5-10 µl RNA solution were carried out.

## **2.4 Crystallographic studies of Lsm polyproteins**

### **2.4.1 Crystallisation of Lsm polyproteins**

Flash frozen aliquots of Lsm polyproteins were used for crystallisation experiments. Protein concentrations were adjusted to 10-30 mg/ml. Following thawing of an aliquot, samples were cleared by centrifugation (10000 x g, 5 min). Initial crystal screening was carried out in sitting drop format in 96 well INTELLI-PLATEs (Art Robbins Instruments) using a Phoenix Liquid Handling System (Art Robbins Instruments). 50-70 µl of precipitant solutions from commercial crystal screening kits (Table 2.1) were filled into the reservoirs and 200-300 nl of protein solution were mixed with precipitant in a 1:2, 1:1 or 2:1 ratio. Plates were sealed and kept at room temperature and examined for crystallogenesis in 2-5 day intervals for 2 months.

Optimisation of crystallisation conditions was performed in grid screens using the four corner method (Hennessy et al. 2009). These optimisation trials were carried out as sitting drop experiments in 96 well format as described above and in hanging drop format in pregreased 24 well VDX plates (Hampton research). For hanging drop experiments, 1-2 µl of protein solution were mixed with precipitant solution in 1:1, 1:2 or 2:1 ratios on a siliconised cover slide. 500 µl of crystallant solution were added to the reservoir. Plates were incubated at room temperature and checked at regular intervals.

For crystallisation screens including Tacsimate<sup>TM</sup>, saturated stock solutions of the crystallant were prepared (1.83 M malonic acid, 0.25 M ammonium citrate tribasic,

0.12 M succinic acid, 0.3 M DL-malic acid, 0.4 M sodium acetate trihydrate, 0.5 M sodium formate, 0.16 M ammonium tartrate dibasic = 100 % Tacsimate<sup>TM</sup>) and subsequently diluted. To extend the buffering range of the mixture above a pH of 6.5, Tris (0.1 M) was added.

In addition, microseeding was employed to obtain new crystal forms and to increase the success rate of crystallisation experiments. For microseeding experiments, existing crystals were harvested and transferred into 20-50  $\mu$ l of stabilising solution. Stabilising solution contained the reservoir solution from the hit condition with a 10-20 % increased precipitant concentration and 5-10 mg/ml protein. Generation of a seed stock was achieved by vortexing crystals with plating beads (Sigma-Aldrich) and subsequent serial dilution. These stocks were used to introduce crystal seeds in drops that did not show signs of precipitation or crystal formation two weeks after setting up the trays. Crystal seeds were transferred by dipping a horsehair into seed stocks followed by dipping or streaking the hair through the protein drop. Outcomes of seeding experiments were assessed every 1-3 days.

#### **2.4.2 Crystallographic data collection**

Crystals obtained for Lsm polyproteins were screened for X-ray diffraction by myself with the assistance of Dr. Stephen Harrop (University of New South Wales, Sydney) on a Rigaku RU200 Cu (Rigaku) rotating anode generator in line with a Mar345 image plate at University of New South Wales.

Crystals were picked up from mother liquor using nylon microfibre loops (Hampton Research) ranging from 0.025-1 mm diameter and were flash cooled in liquid nitrogen. Frozen crystals were mounted in a cryostream on the beamline at 100 K. Diffraction images at  $\Phi$ -angles of 0 ° and 90 ° with exposure times of 1 s were recorded and the diffraction quality of the crystals was judged by resolution, ice ring formation and the shape and intensity of the observed reflections.

All crystals were subjected to cryogenic screening. This procedure was employed to empirically determine the optimal additive and its concentration that protects the crystals from damage in the freezing process and prevents the formation of ice rings. Cryoprotective agents that were screened include: co-solvents (e.g. glycerol and ethyleneglycole, polyethyleneglycol), sugars (e.g. glucose, sucrose and xylitol), alcohols (e.g. ethanol, methanol), oils (e.g. paratone-N, parafine oil) and the precipitant present in the mother liquor. Crystals were exposed to these cryoprotectants by sequential soaking in mother liquor with increasing cryoprotectant concentrations.

Data collection from Lsm polyprotein crystals was carried out on the Micro Crystallography Beamline (MX2) at the Australian Synchrotron, Victoria, Australia, with the assistance of Dr. Harrop. This beamline features tunable wavelengths and a highly focused beam of 37 x 32  $\mu\text{m}$  allowing data collection from small crystals with a maximal signal to noise ratio. For data acquisition, the Blu-Ice program (McPhillips et al. 2002) was used. Crucial system parameters including the optimal  $\phi$ -range and detector distances, and exposure time for data collection were determined for each protein crystal from images at 0 ° and 90 °. The recorded images

were processed in the MOSFLM program (Leslie 1992) in order to obtain estimates of the unit cell parameters and the likely space group of the crystal. In addition, the optimal orientation of the crystal in the beam for data collection was decided from the Strategy option integrated in the software. Direct beam coordinates were calibrated by beamline scientists prior to the experiment. These were verified using the Lawrence Berkeley Lab Indexing Toolbox (LABELIT, Computational Crystallography Initiative). Parameters used for Lsm polyprotein crystals are summarised in Table 2.5.

**Table 2.5|** Data collection parameters for Lsm[4+1]<sub>4</sub> crystals

Parameter	Value
detector distance (mm)	450
beamstop (mm)	50
wavelength ( $\lambda$ ; Å)	0.954
exposure time (s)	5
$\Delta\phi$ (°)	1

### 2.4.3 Crystallographic data processing

Diffraction images were indexed and integrated using the MOSFLM package (Leslie 1992). This process aims to produce a set of indices (hkls) and their associated intensities as well as crystal unit cell parameters. In a first step (autoindexing), MOSFLM estimates the crystal orientation and unit cell dimensions. These unit cell parameters are refined and the mosaicity is determined in a second step. In a third step, the position and intensities of the Bragg reflections are estimated.

Integrated diffraction data were further analysed using the Collaborative Computational Project No. 4 (CCP4) software suite (Winn et al. 2011) or the Python-based Hierarchical ENvironment for Integrated Xtallography (PHENIX) (Adams et al. 2010). Knowledge of the correct Laue group is absolutely essential in crystallographic data processing. As a consequence, initial estimates for the Laue group from MOSFLM were verified using POINTLESS (Evans 2006).

Due to a number of experimental factors, the intensities of reflections that were recorded in the diffraction experiment are not uniform. Factors causing this will include slow variation in incident-beam intensities, varying illuminated volumes, absorption in the primary and secondary beam direction or radiation damage. Hence, a data reduction step is utilised in which intensities are merged and scaled and an internally consistent dataset is produced. This is achieved by minimising the differences between the intensities of redundant, symmetry related reflections using the SCALA tool in CCP4 (Evans 2006). SCALA also allows an evaluation of the quality of the diffraction data. From SCALA analysis, the maximal resolution to which the datasets could be processed was identified using a resolution cut-off of  $\langle I/\sigma(I) \rangle \geq 1.3$ . For  $I/\sigma(I) \leq 1.3$ , the signal to noise ratio was judged too poor for data processing. The extent of radiation damage and internal consistency of the datasets were assessed from relative B factors and  $R_{\text{merge}}$ .

Protein crystals contain a significant amount of solvent, which in general is found between 27 and 78 % (Matthews 1968). If the molecular weight of the protein is known, the solvent content can be used to determine the number of molecules that fit in the asymmetric unit of a given space group and unit cell. The solvent content of

Lsm[4+1] crystals was analysed using the cell content analysis tool in CCP4 (Matthews 1968; Kantardjieff and Rupp 2003).

#### 2.4.4 Phasing of diffraction data

The diffraction experiment provides us with a pattern of reflections. The intensities at any given point (hkl) of the diffraction pattern are the result of the interference from waves diffracted by electrons in the crystal lattice. Waves are described by an amplitude and a phase angle. Depending on their phase angle, the diffracted waves interfere constructively or destructively and their sum determines the detected reflections. Hence, the observed diffraction pattern can mathematically be expressed as the Fourier transform of the electron density. As a consequence, we can use an inverse Fourier transform to calculate the electron density at a given point xyz in the asymmetric unit (Taylor 2003):

$$\rho(xyz) = \frac{1}{V} \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i h x + k y + l z)$$

(2-3)

where V describes the volume of the unit cell,  $F_{hkl}$  the structure factor amplitude and  $\alpha$  the phase angle.

The unit cell volume and the amplitudes of the structure factors can be directly obtained from the recorded reflections, however, information regarding the phases is lost. Hence, phases need to be determined experimentally or molecular replacement



must be employed to orient a homology model in the experimental unit cell. The oriented model can then be translated relative to the origin of the experimental unit cell (Taylor 2003). This is what is referred to as the phase problem. In this study, the phase problem was tackled by molecular replacement.

Molecular replacement attempts to place a search model in the asymmetric unit of the crystal. Search models for phasing were identified with the Basic Local Alignment Tool (BLAST) (Altschul et al. 1990) implemented in NCBI. Search models were edited using the programs CHAINSAW (Stein 2008) and PHENIX Sculptor (Bunkoczi and Read 2011) based on pairwise sequence alignments with the Lsm4 and Lsm1 target sequences. A varying number of these models was placed in the asymmetric unit using the PHASER program (McCoy et al. 2007) in CCP4. Positioning a search model in the asymmetric unit requires a six-dimensional search that covers three dimensions of rotational space and three dimensions of translation. PHASER aims to determine the orientation of the search model and its position in the asymmetric unit through maximum likelihood probability theory and multivariate statistics. Molecular replacement using PHASER was performed using either the automated molecular replacement function or separate searches for rotation and translation functions.

### 3 Solution behaviour of Lsm polyproteins

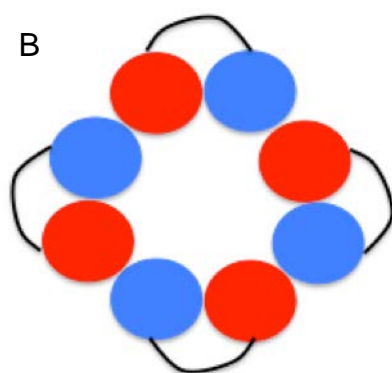
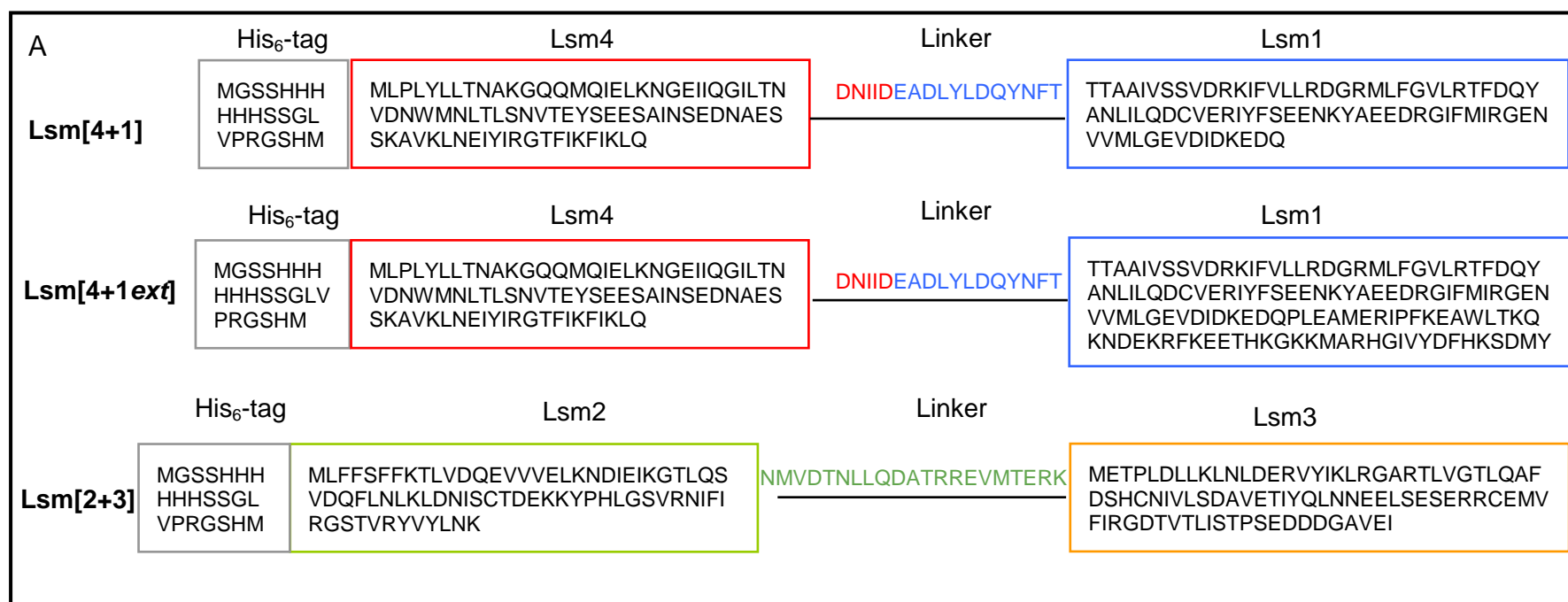
In eukarya, Lsm proteins form heteromeric complexes of heptameric composition. The heptamer containing Lsm proteins 1-7 differs in RNA target, function and cellular localisation from that made up from Lsm 2-8. Our current understanding of the composition and assembly of heteromeric Lsm complexes is shaped by knowledge concerning the Sm protein system. The heptameric Sm complex is likely to assemble sequentially about pre-mRNA from three preformed subcomplexes: SmD1/D2, SmD3/B, and SmF/E/G (Raker et al. 1996; Beggs 2005). This suggests that corresponding dimer and trimer complexes of Lsm, namely Lsm2/3, Lsm4/1, and Lsm6/5/7, are very likely to possess some inherent stability and are also worthy of study as independent modules. This hypothesis is supported by the successful preparation of stable recombinant assemblies of these Lsm complexes, in octa-, tetra- and trimeric forms by co-expression (Zaric et al. 2005).

When expressed individually, recombinant Lsm products are often obtained in low yield or not at all (Zaric et al. 2005; Sobti 2008). Those Lsm proteins successfully produced tend to aggregate or form stable homomeric complexes (Naidoo et al. 2008). Thus, preparation of mixed Lsm assemblies has in the past required extensive refolding steps. Coexpression of Lsm complexes has significantly improved expression levels and solubility (Zaric et al. 2005) and allowed the preparation of some discrete heteromeric complexes.

The protein structure group at Macquarie University has been investigating an alternative production route for mixed Lsm complexes by preparation of polyprotein versions of Lsm[2+3] and Lsm[4+1] (Sobti et al. 2010). When utilised as baits for affinity pull-down experiments within yeast lysate, appropriate groups of Lsm proteins are captured. These results demonstrate that dynamic rearrangements of Lsm complexes occur *in vivo*, possibly as discrete Lsm sub-complexes and also establish the functional integrity of polyprotein forms (Sobti et al. 2010).

In order to pursue structural and functional characterisation of the Lsm sub-complexes, large quantities of highly pure protein must be prepared *in vitro*, often a major bottleneck in structural studies (Chayen and Saridakis 2008). To meet this demand, I have continued to develop and optimise the production of Lsm polyproteins. Figure 3.1 outlines this approach in which Lsm gene products are fused by a covalent linker sequence utilising natural sequence segments. N- and C-termini from the flexible linker sequences are presumed not to interfere with the integrity and correct folding of the individual component Lsm domains (Figure 3.1). The N-terminal sequence comprises a hexaHis affinity tag coupled with a protease cleavage site (not utilised in my preparations). This sequence provides for rapid purification but adds some charged residues additional to the native Lsm fold, and may modulate solution behaviour. However, a recent study on the archaeal relative *HvLsm $\alpha$*  incorporating an identical N-terminal sequence demonstrated native-like RNA binding (Fischer et al 2011).

The polyprotein approach for Lsm guarantees stoichiometric expression of the individual components to assemble coherent oligomeric forms. To obtain insights



**Figure 3.1|** Dual Lsm polyprotein constructs. A) Sequences of Lsm polyproteins Lsm[4+1], Lsm[4+1ext] and Lsm[2+3]. B) Proposed representation of polyprotein quaternary assemblies in this case for Lsm[4+1]. (Lsm4-red, Lsm1-blue). Circular form based on observed crystal structure of Lsm[3]<sub>8</sub> (Naidoo et al. 2008).

into their oligomerisation behaviour and to ultimately identify conditions that enable the isolation of stable monodisperse polyprotein complexes, I have performed extensive solution-state characterisation using biophysical methods. These probes will guide our efforts to ensure correct folding and assembly of the Lsm polyprotein complexes.

### 3.1 Preparation of recombinant Lsm polyprotein complexes

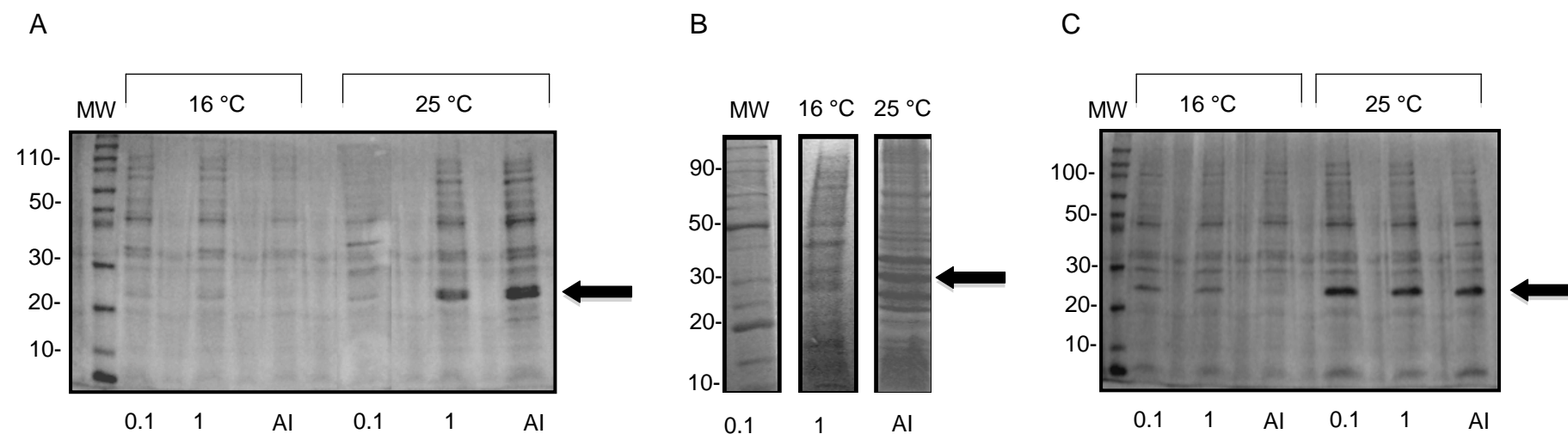
N-terminally His<sub>6</sub>-tagged versions of the polyproteins Lsm[4+1], Lsm[4+1*ext*] and Lsm[2+3] were prepared in bacterial cells. Early preparations showed incorporation of the bacterial Lsm paralogue Hfq. To prevent this contamination, the Hfq deficient *E. coli* strain MRE5 was used for the production of these polyproteins. The gene sequences of *LSM* 1-4 contain 9-13 rare tRNA codons. Hence, Lsm expression constructs were co-transformed with a plasmid encoding for rare tRNAs. Ampicillin resistance conferred by the vector carrying the Lsm genes and chloramphenicol resistance encoded on the pRARE plasmid allowed selection for cells containing both plasmids.

Following the generation of Lsm expression cell lines, levels of polyprotein expression were optimised with respect to temperature and induction method in small-scale expression trials (2 ml). Protein expression levels at 16 and 25 °C were analysed as were various induction methods: 0.1 mM IPTG, 1 mM IPTG and auto-induction. For the polyprotein Lsm[4+1], cultures incubated for 12 - 24 h at 25 °C and induced with either 1 mM IPTG or via auto-induction showed good product

yield. Figure 3.2 shows the SDS-PAGE gels for these conditions with a band at ~25 kDa, consistent with the polyprotein mass. As highest yield was observed for auto-induction cultures, auto-induction at 25 °C was chosen as the standard expression method for Lsm[4+1]. Expression of the related polyprotein construct Lsm[4+1*ext*] via auto-induction is also optimal at 25 °C. Expression of polyprotein Lsm[2+3] at 25 °C is not critically dependent on the induction method (Figure 3.2).

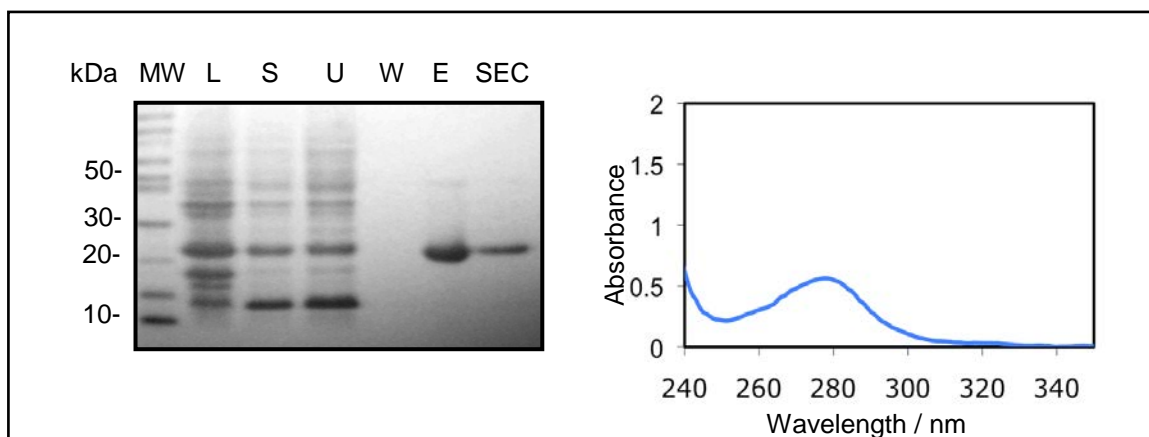
Thus, for ease of handling, auto-induction was used for the preparation of Lsm[2+3]. By these means, general production yields of 30 mg/ml, 20 mg/ml and 15 mg/ml for Lsm[4+1], Lsm[4+1*ext*] and Lsm[2+3] were routinely obtained.

Due to the presence of an N-terminal His<sub>6</sub>-tag, high purity of all products was readily obtained through the use of IMAC purification, followed by a SEC polishing step to separate oligomeric components. SDS-PAGE analysis of single species of Lsm[4+1], Lsm[4+1*ext*] and Lsm[2+3] yields bands at ~25, ~30, and ~25 kDa, respectively (Figure 3.3) and indicates good chemical purity. The high quality of these polyprotein preparations is further ensured by absorbance scans (Figure 3.3) (Warburg and Christian 1941).  $A_{260}/A_{280}$  absorbance ratios of ~0.6 in each case indicate the absence of nucleic acid contaminants in my protein preparations.

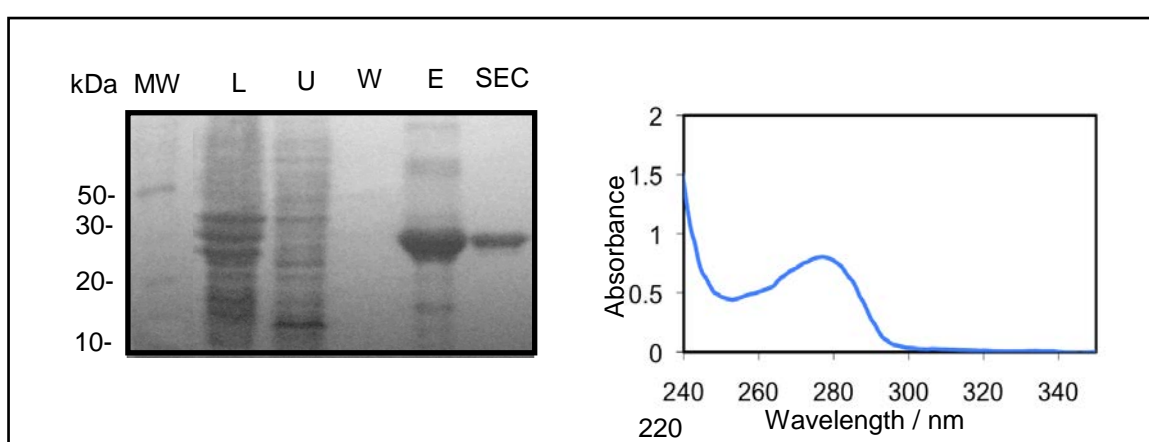


**Figure 3.2** | SDS-PAGE of soluble fractions from expression trials for Lsm polyproteins. A) Lsm[4+1], B) Lsm[4+1exl], C) Lsm[2+3]. Three induction methods and two temperatures were compared. Induction methods are indicated below the lanes; 0.1 mM IPTG (0.1), 1 mM IPTG (1), auto-induction (AI). Expression cultures are grown overnight for IPTG induction and 24 h for auto-induction. Arrows indicate polyprotein bands.

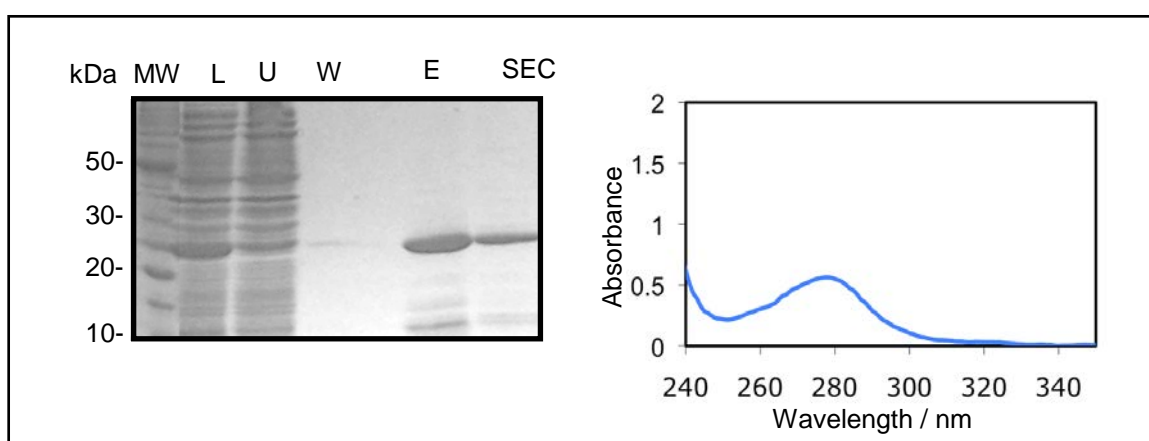
A



B



C



**Figure 3.3|** Purification of Lsm polyproteins. A) Lsm[4+1], B) Lsm[4+1exf], C) Lsm[2+3]. SDS-PAGE gels show samples taken from crude lysate (L), soluble lysate fractions (S), unbound fractions (U), wash fractions (W), IMAC elution fractions (E), SEC purified samples (SEC). Molecular weight markers are indicated by MW. UV-traces display the absorbance of purified Lsm polyprotein samples in the range of 220-350 nm.

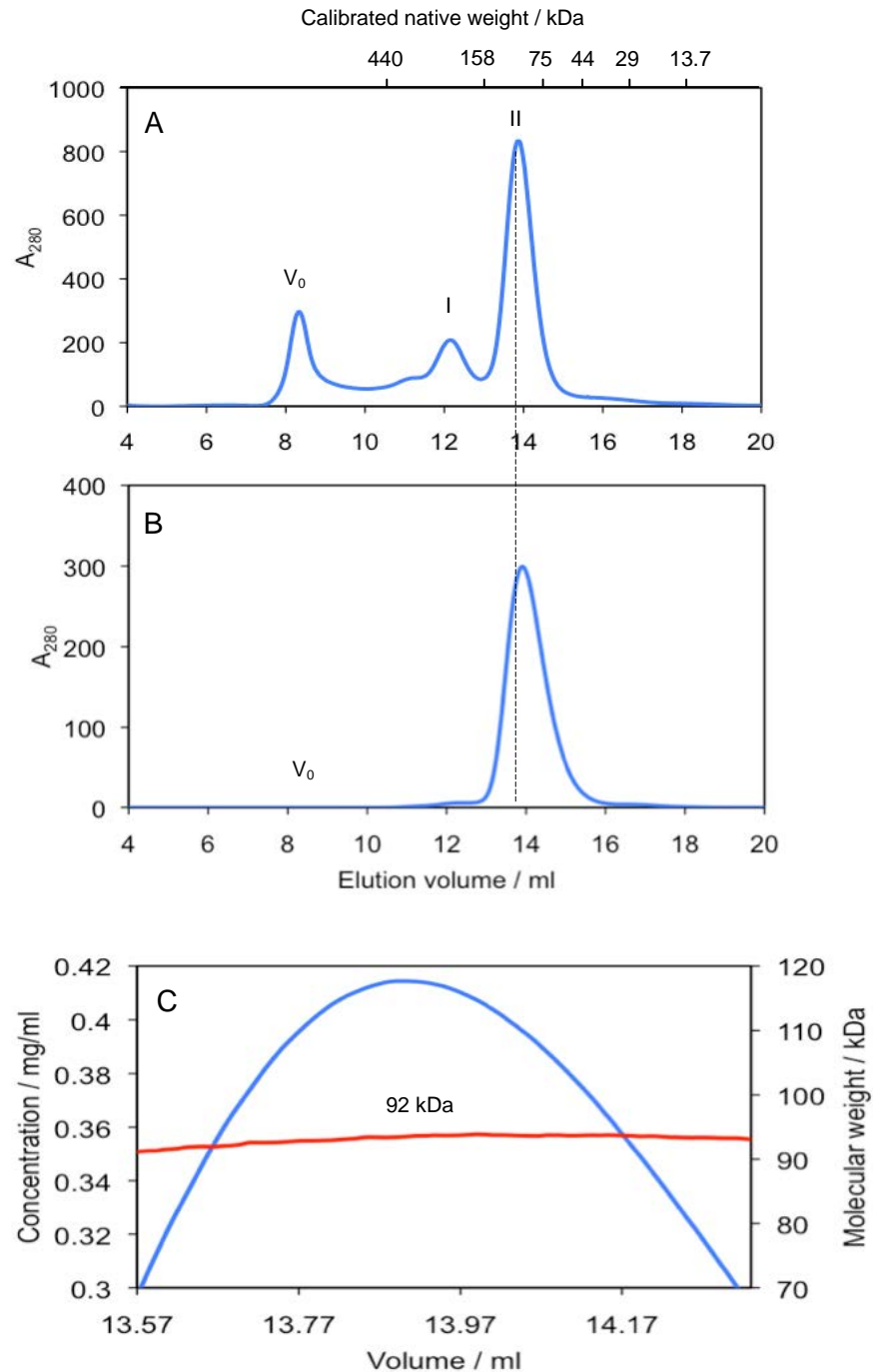


### 3.1 Quaternary structures of Lsm polyproteins in solution

#### 3.2.1 Polyprotein Lsm[4+1]

Following purification, samples of Lsm[4+1] were subjected to analytical SEC in a standard Tris buffer (20 mM, pH 8.0) to ascertain the quaternary structure for the solution complex. In a moderate level of salt (400 mM NaCl), as well as glycerol (10 % v/v), three distinct protein species can be identified. Figure 3.4 shows these as three distinct peaks by SEC corresponding to masses of 220 kDa (fraction I), 100 kDa (fraction II) and species > 1300 kDa ( $V_0$ ). These values are consistent with discrete octameric and tetrameric Lsm complexes (fractions I and II, respectively), in solution, as well as large aggregate species. The formation of discrete octameric and tetrameric complexes is reminiscent of stacked and single ring structures previously defined by the crystal structure of Lsm3 (Naidoo et al. 2008). These assemblies of Lsm[4+1] will be referred to as Lsm[4+1]<sub>4</sub> and Lsm[4+1]<sub>8</sub>.

In Tris buffer with 400 mM NaCl and glycerol, this polyprotein appears to exist as ~ 85 % Lsm[4+1]<sub>4</sub> and 15 % Lsm[4+1]<sub>8</sub> (proportions assessed by peak integration of SEC traces). These two species do not kinetically interconvert, as isolation of individual fractions and subsequent reapplication to the SEC column results in a single form only. Figure 3.4 demonstrates the purity of a Lsm[4+1]<sub>4</sub> sample prepared in this manner: the rechromatographed fraction yields a symmetric single peak, with no signs of any redistribution into multiple species. Subsequent SEC-MALLS analysis reveals a monodisperse average mass of 92 kDa over the peak, clearly

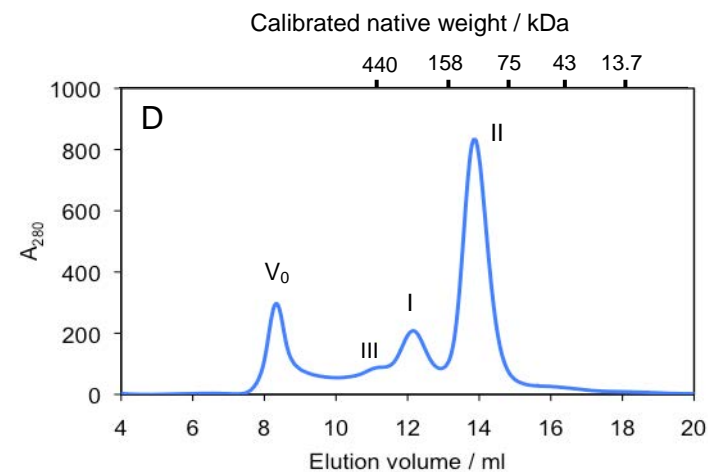
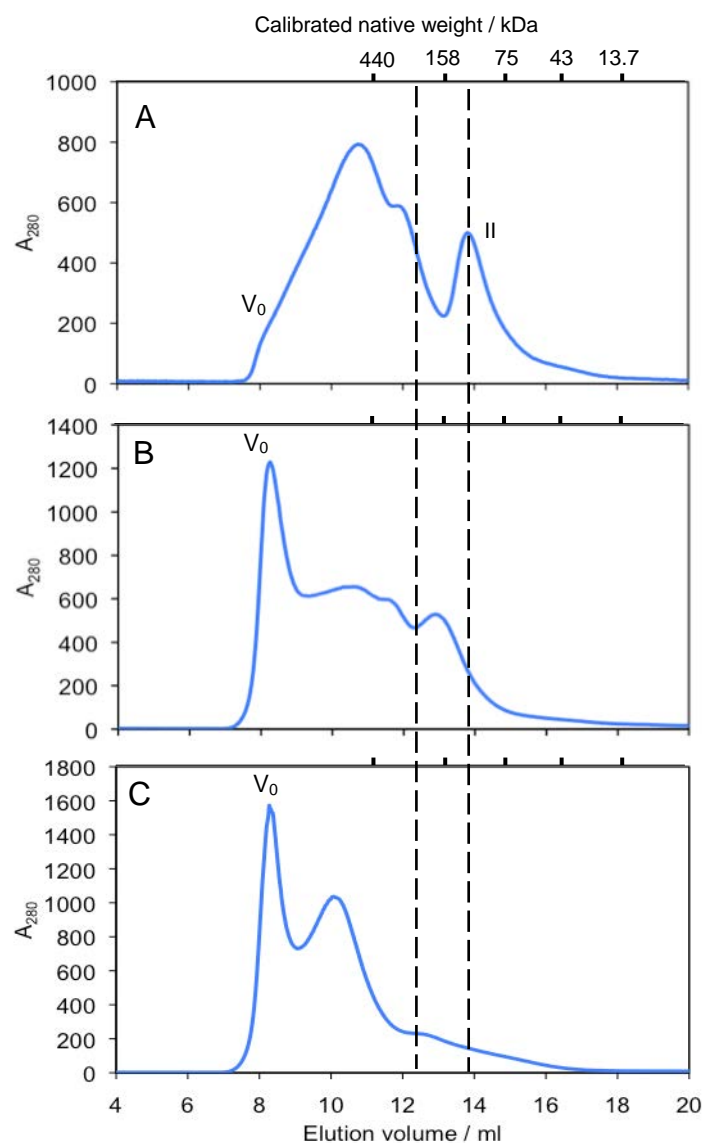


**Figure 3.4|** Solution state of the polyprotein Lsm[4+1] analysed by analytical SEC and SEC-MALLS in Tris buffer (20 mM, Tris pH 8.0) with 400 mM NaCl, 10 % v/v glycerol. A) IMAC-purified Lsm[4+1] separated on a Superdex 200 10/300 GL column at 0.5 ml/min. B) Re-chromatographed fraction II material. C) SEC-MALLS analysis of isolated fraction II. The molecular weight of each point across the chromatogram is indicated as red line.

confirming a single stable complex of tetrameric composition (theoretical tetramer mass = 92.21 kDa).

The solution behaviour of Lsm[4+1] preparations and their mixed components was further probed in a search for stabilising conditions. Dielectric and ionic strength effects on affinity-purified polyprotein were probed.

Comparison of Lsm[4+1] samples purified in standard Tris buffer conditions (20 mM, pH 8.0 with 400 mM NaCl) demonstrates that the oligomerisation distribution of the polyprotein is markedly affected by glycerol. A comparison of panels A and D in Figure 3.5 demonstrates that without the polyol many fractions are present. These are consistent with multiple oligomeric forms of 200 kDa and above. Approximately 20 % of the material appears to arise from Lsm[4+1]<sub>4</sub> (fraction II) only. However, with the addition of glycerol (at 10 % v/v), the distribution of these higher molecular weight species is dramatically altered. While some large species are observed to pass in the void volume, the majority of material (85 %) elutes as fraction II, i.e. attributable to Lsm[4+1]<sub>4</sub>. Only a smaller proportion corresponds to Lsm[4+1]<sub>8</sub> (fraction II), and a discrete 20-mer (fraction III) may be evident within the SEC trace (corresponding to ~450 kDa) at a very low level.



**Figure 3.5]** SEC profiles of Lsm[4+1] preparations on Superdex 200 10/300 GL column at 0.5 ml/min. Traces are obtained in Tris buffer (20 mM, pH 8.0) with additional components. A) NaCl (400 mM), B) NaCl (200 mM), C) NaCl (50 mM), D) NaCl (400 mM), glycerol (10 % v/v). Dashed line corresponds to native mass elution for Lsm[4+1]<sub>8</sub> and Lsm[4+1]<sub>4</sub> (named fractions I and II). Possible Lsm[4+1]<sub>20</sub> (III).

Solvent ionic strength was also found to strongly affect the oligomerisation behaviour of Lsm[4+1]. Samples containing intermediate levels of salt (200 mM NaCl; Figure 3.5 B) show a significant loss in the amount of material present as Lsm[4+1]<sub>4</sub>, and a very different and broad mass distribution of species. This is altered yet again as the salt concentration is further reduced to 50 mM NaCl. In this case, the average molecular weight of the Lsm[4+1] species present is close to 640 kDa, possibly corresponding to a 28-mer, and a significant amount of material also elutes at V<sub>0</sub> (Figure 3.5 C).

Table 3.1 summarises the trend for the stabilisation by salt on the tetrameric form of Lsm[4+1]. This solution form is stabilised at high salt concentrations. The trend to larger, possibly stacked species with lower salt indicates the Lsm stacking mechanism to be driven by electrostatic forces. Two types of multimerisation have been seen in Lsm crystal structures engaging either face-to-face or head-to-tail stacking of Lsm rings. Weakened intermolecular forces due to an increase in counter ion concentration explains the preference for smaller single-ring complexes at high salt concentrations.

**Table 3.1|** Estimated tetramer content of Lsm[4+1] and Lsm[2+3] preparations<sup>a</sup>

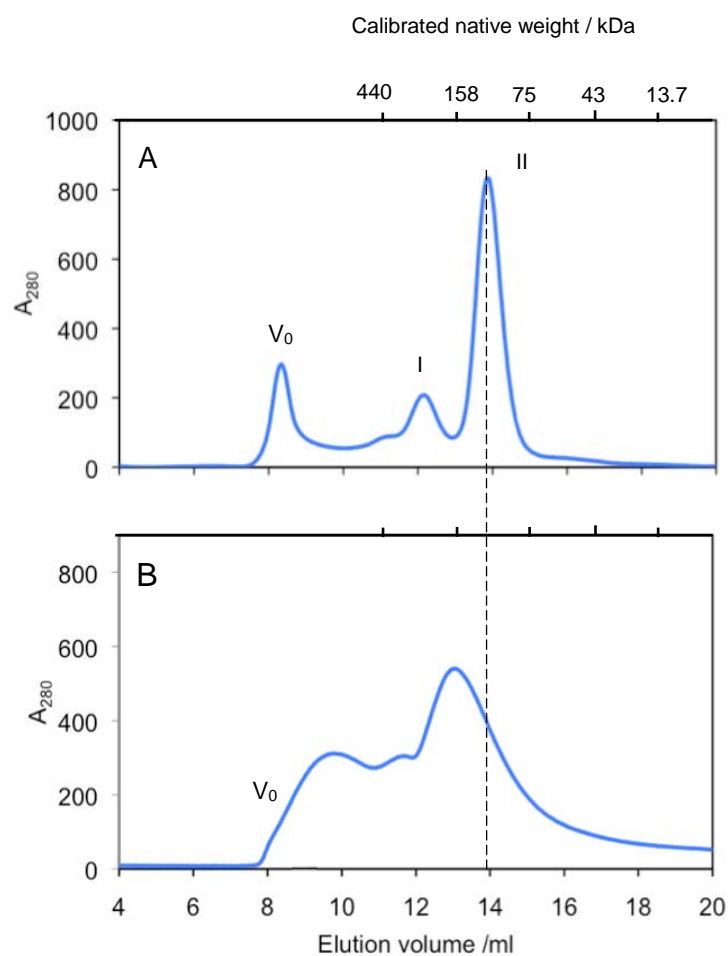
Buffer additives <sup>b</sup>	Lsm[4+1] <sub>4</sub> content (%)	Lsm[2+3] <sub>4</sub> content (%)
400 mM NaCl, 10 % glycerol	85	70
400 mM NaCl	20	60
200 mM NaCl	0	40
50 mM NaCl	0	< 40

<sup>a</sup> relative proportions of oligomers assessed by peak integration using UNICORN 4.12

<sup>b</sup> components additional to Tris buffer (20 mM, pH 8.0)

Face-to-face stacking of two Lsm rings has been observed to cause close packing of the His residues comprising the affinity tag at the N-terminus (Naidoo et al. 2008). Evidence for any multimerisation mediated by these residues is provided noting the pH dependence of the oligomeric composition. Lsm[4+1] samples analysed by SEC in MES buffer (20 mM, pH 6.0) containing both NaCl (400 mM) and glycerol (10 % v/v) display a large number of high mass species (Figure 3.6). This indicates an enhanced formation of larger Lsm multimers with a protonated rather than neutral His imidazole sidechain. Such a change in the oligomeric distribution is consistent with electrostatically-mediated ring stacking. Similar pH effects on the oligomeric state have also been observed for Lsm3 and Lsm[2+3] in previous solution studies (Naidoo et al. 2008; Sobti et al. 2010).

Overall, in order to prepare the most stable solutions of tetrameric Lsm[4+1], I have found best conditions to be at high salt and pH. Accordingly, for further biophysical work on this protein the solvent was Tris buffer (20 mM, pH 8.0) with NaCl (400 mM) and glycerol (10 % v/v).



**Figure 3.6]** pH dependency of the oligomeric state of Lsm[4+1]. SEC traces of Lsm[4+1] run on a Superdex 200 10/300 GL column. A) pH 8.0 (20 mM Tris, 400 mM NaCl, 10 % glycerol), B) pH 6.0 (20 mM MES buffer, 400 mM NaCl, 10 % v/v glycerol). Lsm[4+1]<sub>4</sub> assemblies are indicated by a dashed line.

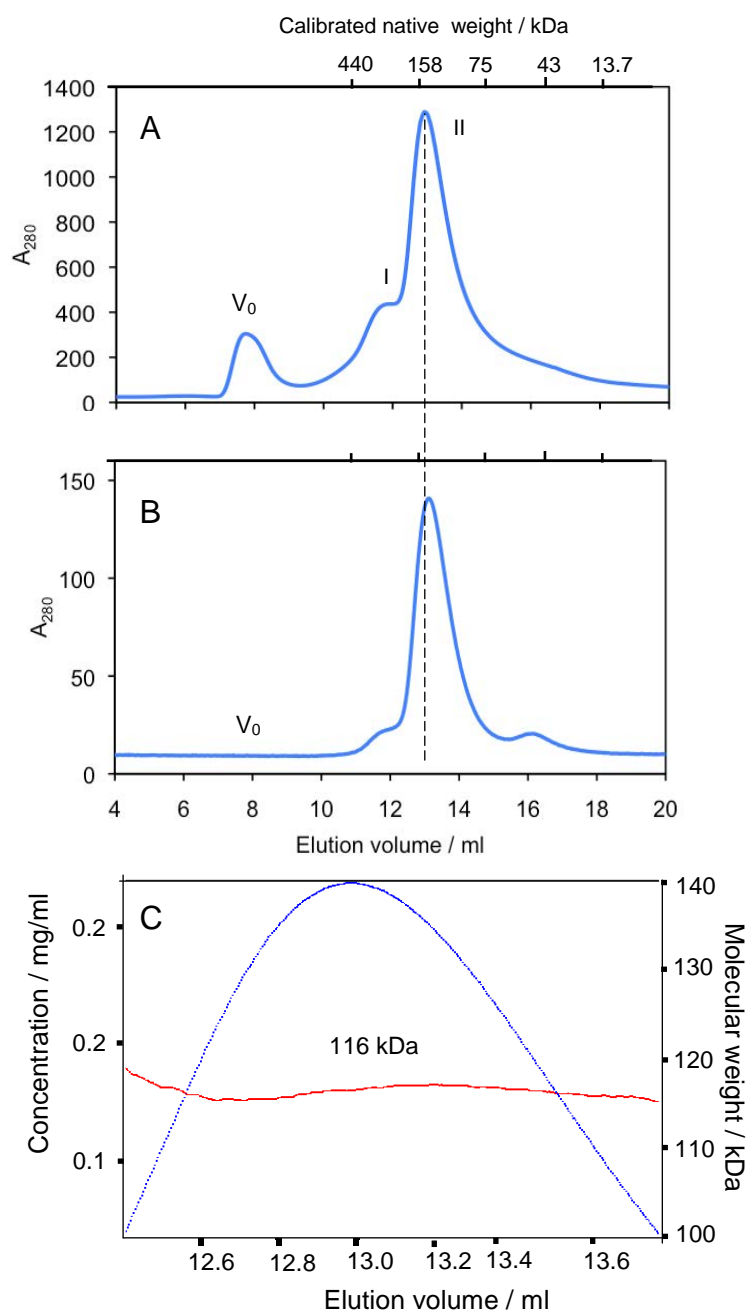
### 3.2.2 Polyprotein Lsm[4+1 $_{ext}$ ]

This variant of the polyprotein Lsm[4+1] incorporates the full C-terminal region of Lsm1, i.e. containing an additional 50 amino acids to Lsm[4+1] (Figure 3.1). The added residues are highly charged and are considered to be outside any folded domain. This species is of interest, as the flexible and charged C-terminus is likely to serve as recruitment point for RNA and protein partners, as seen for Sm proteins (Weber et al. 2010). A highly charged exposed C-terminal segment, attached at the proximal face of any Lsm toroid, may also elevate stacking of rings.

In my standard Tris buffer conditions, Lsm[4+1 $_{ext}$ ] can be seen by SEC to be possibly more diverse in its oligomeric forms than Lsm[4+1] under the same conditions (Figure 3.7).

While the main peak (at 140 kDa) likely corresponds to a tetrameric composition, shoulders are present, indicating octameric even monomeric groupings. Material rechromatographed from fraction II is shown by static light scattering to be monodisperse with a mass average of 116 kDa consistent with Lsm[4+1 $_{ext}$ ]<sub>4</sub> (116.64 kDa). A peak corresponding to a mass of 30 kDa is detected in reinjected material, confirming the formation of a small percentage of monomeric species over time. As a consequence, it can be reasoned that this polyprotein version is more dynamic in quaternary structure than Lsm[4+1].





**Figure 3.7]** Solution state characterisation of the polyprotein Lsm [4+1ext] using SEC coupled to static light scattering. Purified protein samples are analysed by analytical SEC and SEC-MALLS in 20 mM Tris pH 8.0, 400 mM NaCl, 10 % glycerol. A) IMAC purified Lsm[4+1] separated on a Superdex 200 10/300 GL gel filtration column at 0.5 ml/min. B) Re-chromatographed fraction II. C) SEC-MALLS analysis of isolated fraction II. A red line represents the molecular weight of each point in the chromatogram.

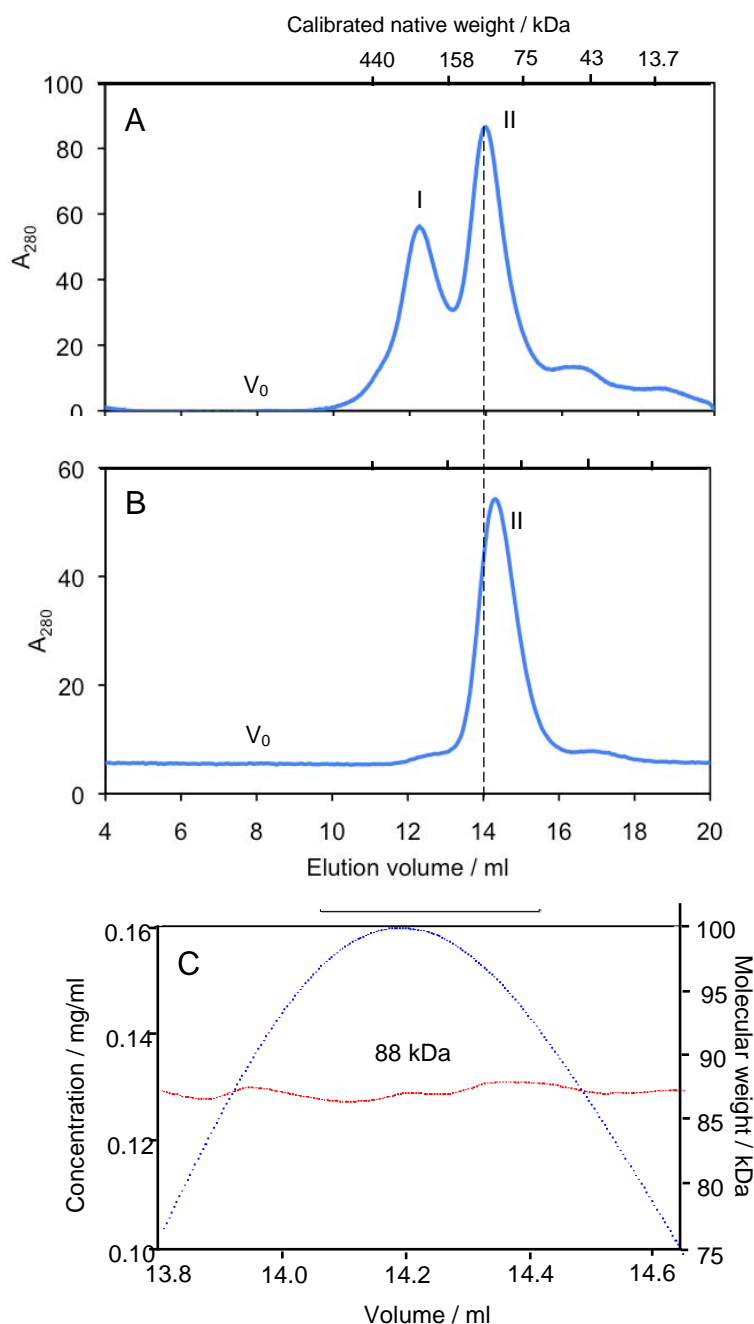
### 3.2.3 Polyprotein Lsm[2+3]

This polyprotein has been studied previously in this laboratory (Sobti et al. 2010). Using my optimal Tris buffer/salt/glycerol system, only clear tetra and octameric Lsm[2+3] assemblies occur in solution (Figure 3.8). As outlined in Table 3.1, the ratio of octameric to tetrameric forms is approximately 30:70 %, indicating an elevated preference for octamer formation relative to Lsm[4+1]. The two Lsm[2+3]<sub>4</sub> and Lsm[2+3]<sub>8</sub> fractions can be stably isolated and do not interconvert. Subsequent SEC-MALLS analysis of Lsm[4+1]<sub>4</sub> material reveals an average molecular weight of 88 kDa, consistent with a tetrameric organisation (93.88 kDa).

The increased stacking propensity for rings of Lsm[2+3] may arise from differences in the linker sequences:

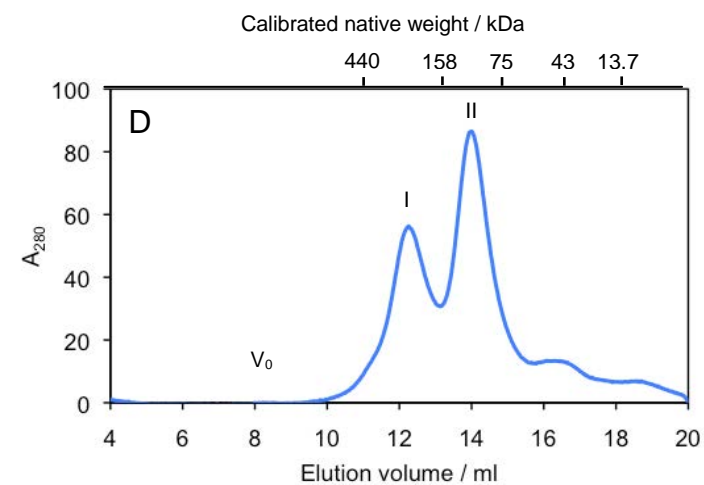
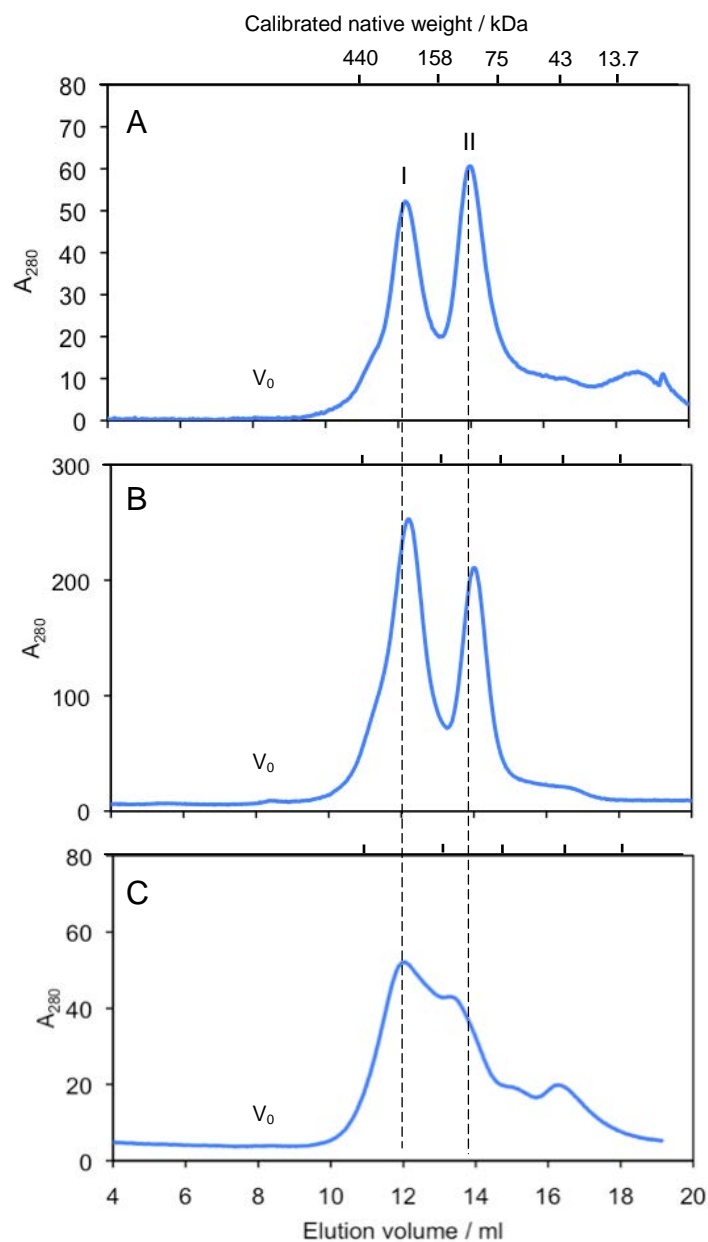
- DNIIDEADLYLDQYNFT for Lsm[4+1] and
- NMVDTNLLQDATRREVMTERKMET

which likely lie along the proximal face of each single toroid of the Lsm tetramer. Certainly, this segment is far larger in Lsm[2+3], possibly providing increased interaction points which can be utilised in face-to-face packing for an octamer .



**Figure 3.8]** Lsm[2+3] SEC profile. Purified protein samples are applied to a Superdex 200 10/300 GL column and run in Tris buffer (20mM, pH 8.0) containing 400 mM NaCl and 10 % glycerol at 0.5 ml/min. A) IMAC purified polyprotein separated via SEC showing two distinct fractions. B) Isolated and re-chromatographed fraction II from A. C) SEC-MALLS of isolated fraction II. A red line represents the molecular weight distribution.

Similar to Lsm[4+1], the effects of glycerol and the NaCl concentration on the solution state of the polyprotein were probed (Figure 3.9). As summarised in Table 3.1 the same trend of stabilisation of tetrameric polyprotein species by glycerol and high salt concentrations is observed for Lsm[2+3]. This is consistent with the proposed mechanism of electrostatically mediated polyprotein stacking. Reduction of the ionic strength to 50 mM NaCl seems to have a severe effect on the oligomerisation behaviour of the protein: total absence of tetrameric assemblies and the formation of a wide size distribution including monomeric forms are observed. Monomer formation in low salt conditions indicates that electrostatics may even override the H-bond stability of the  $\beta$ -propeller of the toroid.



**Figure 3.9]** SEC profiles of Lsm[2+3] at various solvent conditions on a Superdex 200 10/300 GL column at 0.5 ml/min. Traces are obtained in Tris buffer (20 mM, pH 8.0) with additional components. A) NaCl (400 mM), B) NaCl (200 mM), C) NaCl (50 mM), D) NaCl (400 mM), glycerol (10 % v/v). Dashed lines highlight octameric and tetrameric assemblies.



### 3.3 Conclusion

Using suitable solvent conditions, I could isolate stable tetramers of Lsm polyproteins by SEC. Reinjection of isolated tetramers on an analytical SEC column results in a single oligomeric species and SEC-MALLS confirms tetrameric composition. SEC results demonstrate that the oligomeric state of Lsm polyproteins depends on the solvent compositions, in particular on the presence of glycerol, the ionic strength and the pH of the solvent.

Selective stabilisation of Lsm[4+1]<sub>4</sub> by glycerol was observed. Glycerol can induce conformational changes towards a more compact and ordered state increasing protein stability (Prieu et al. 1996; Scharnagl et al. 2005) and reduce protein flexibility (Knubovets et al. 1999). Glycerol has been observed to reduce partial unfolding of proteins (Kornblatt et al. 1993; Mishra et al. 2007) and to inhibit protein aggregation (Chi et al. 2003). Underlying mechanisms include:

- i) preferential hydration of volume increments by direct interaction of glycerol and protein via electrostatic contacts (Scharnagl et al. 2005; Vagenende et al. 2009), and
- ii) non-specific steric exclusion of glycerol (Scharnagl et al. 2005; Vagenende et al. 2009).

Both mechanisms provide an explanation for glycerol-induced stabilisation of tetrameric Lsm polyprotein complexes. In order to retain a thermodynamically favourable situation, glycerol is preferentially excluded from the protein surface. This is achieved by a minimisation of the protein-solvent interface through tight

packing and compaction of the protein resulting in the aggravation of intermolecular contacts and a less dynamic conformation.

Observed salt effects on Lsm oligomer distribution suggest electrostatic stacking of polyprotein complexes. This may be due to

- i) Face-to-face stacking of tetramers. Packing interactions in the crystal structure of Lsm3 include face-to-face stacking of two octameric rings mediated mainly via the N-termini (Naidoo et al. 2008). They include a salt bridge between Glu2 and Lys9 of two interacting octamers, and contacts between the imidazole rings of His5 and His6 of the N-terminal His<sub>6</sub>-tag. His1-4 have not been resolved in the crystal structure, which could be due to the flexibility of the tag. However, they are likely to make additional contacts across the octamer interface.
- ii) Head-to-tail stacking of tetramers. Stacking of archaeal *MtLsm* $\alpha$  has been observed to occur via electrostatic contributions of acidic residues in loop L4 (EDGE) and exposed basic and polar residues (R) in the unstructured N-terminus and the N-terminal  $\alpha$ -helix (NRD) of the opposing subunit (Mura et al. 2003).

Observed solvent effects on the interaction of polyprotein complexes are consistent with both proposed mechanisms.

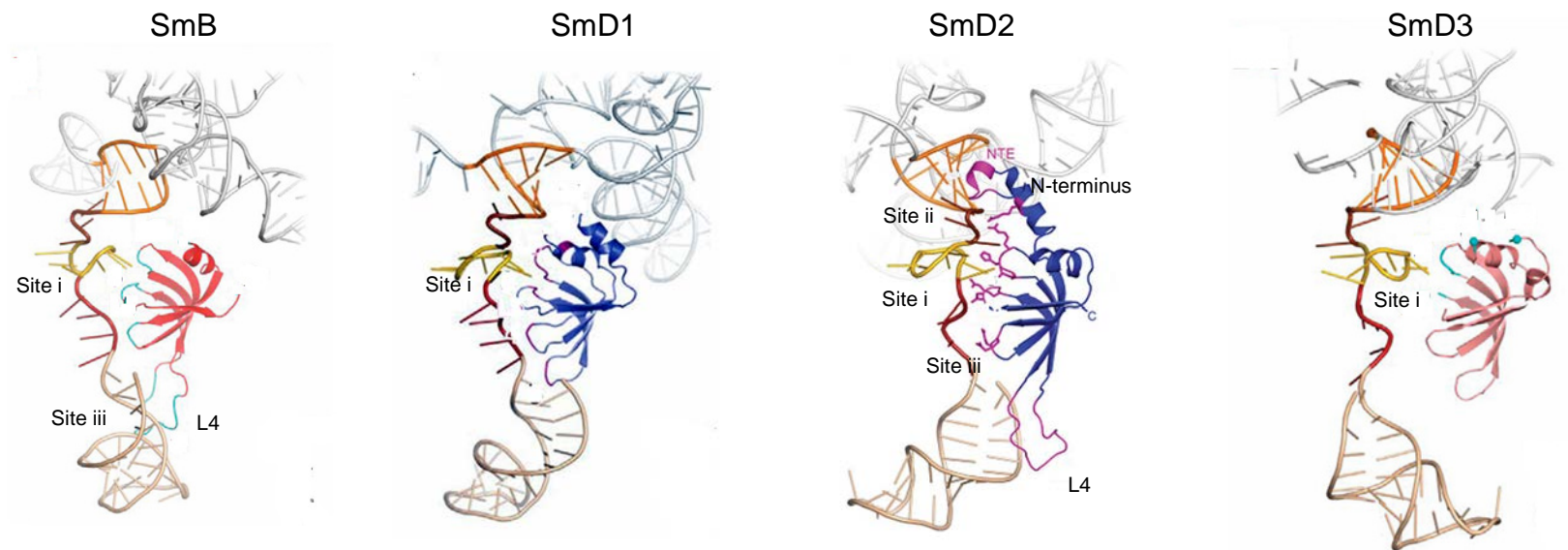


## 4 Biophysical characterisation of Lsm complexes and their RNA interactions

Our engineered Lsm polyprotein complexes provide valuable probes to determine the affinity of specific Lsm components towards RNA. Within the crystal structures of L/Sm proteins complexed with RNA solved to date (see Section 1.3), some details of the RNA-binding sites can be seen. We have classified three distinct interactions sites (i)-(iii) located about the Lsm toroid, and its extensions (Section 1.4). In the crystal structure of the U1 snRNP (Weber et al. 2010), U1 snRNA is channeled through the central pore (site i) with additional RNA contacts provided by extended termini and loop L4 regions of individual Sm component proteins (Figure 4.1).

Gel-shift assays conducted with U-rich RNA indicated that polyproteins Lsm[2+3]<sub>4</sub>, Lsm[4+1]<sub>4</sub> and octameric Lsm[3] have differential RNA-binding capacities (Sobti et al. 2010). RNA trialed in these affinity studies included the oligonucleotides A<sub>2</sub>U<sub>5</sub>, U<sub>10</sub>, (AU)<sub>5</sub>, and A<sub>10</sub>, for which Lsm proteins are thought to have some affinity (Achsel et al. 1999; Raker et al. 1999; Chowdhury et al. 2007; Link et al. 2009). Interestingly, in this gel-shift study, the octameric complex Lsm[3]<sub>8</sub> did not display any detectable level of RNA binding.

In this chapter, I investigate more detailed aspects of Lsm-RNA binding events utilising a variety of solution state biophysical methods.



**Figure 4.1|** Interactions of Lsm 1, 2,3 and 4 Sm paralogues SmB, SmD1, SmD2 and SmD3 with U1 snRNA. Figures adapted from (Weber et al. 2010).

## 4.1 Small Angle X-ray Scattering (SAXS)

From previous solution state characterisation it is evident that polyproteins Lsm[4+1] and Lsm[2+3] form stable tetrameric and octameric complexes. Physical parameters are lacking however, as whether these assemblies are arranged as single and stacked ring structures similar to octameric Lsm3. Accordingly, I utilised the solution-based technique of SAXS to pursue structural characterisation of my polyprotein complexes. Toroids can be readily distinguished from other aggregates and shapes by their X-ray scattering patterns, and the technique is therefore highly suitable for probing the dimensions and forms of Lsm complexes.

### 4.1.1 Theory of Small Angle X-ray Scattering

The electric field  $E(r,t)$  of X-rays scattered by an electron can be described by the Thompson relation:

$$E(r,t) = r_0 \frac{\sin \Psi}{r} E(t) \quad (4-1)$$

wherein  $r_0$  describes the electron radius,  $\Psi$  the polarisation angle,  $r$  the distance between the observer and the scattering event.

In an experimental setup, the amplitude of the wave is usually detected as intensity,  $I$ , which can be related to the scattering angle ( $2\theta$ ) for the incident beam intensity  $I_0$ :

$$I(2\theta) = r_0^2 \left( \frac{1 + \cos^2(2\theta)}{2} \right) \frac{1}{r^2} I_0 \quad (4-2)$$

In analogy to scattering by a single electron, the scattering of assemblies of electrons (such as provided by macromolecules) is given by the sum of all electromagnetic waves scattered by each electron in the molecule. Distances between the scatterers ( $\Delta$ ) give rise to a phase difference ( $\varphi$ ) between the waves:

$$\varphi = \mathbf{s} \cdot \Delta \quad (4-3)$$

$$|\mathbf{s}| = 4\pi \sin\theta / \lambda \quad (4-4)$$

with the momentum transfer vector ( $\mathbf{s}$ ) and the wavelength  $\lambda$ .

For two atoms in a molecule,  $\Delta$  can be regarded as a fixed distance. Thus, the amplitudes of the scattered waves can be summed up with an appropriate phase shift by Fourier transformation.

Scattering from a randomly oriented particle in solution results in spherically symmetric intensities in all possible directions (isotropic scattering). This will include contributions from the macromolecule, the sample buffer and the capillary used in data collection. Subtraction of buffer and capillary signals from collected data yields the scattering of the macromolecule, which is a function of the momentum transfer vector ( $\mathbf{s}$ ). For a homogenous sample, the scattering curve  $I(\mathbf{s})$  results from the spherically-averaged electron distribution  $p(r)$  of the macromolecule:

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty srI(s) \cdot \sin(sr) ds \quad (4-5)$$

and can be computed as

$$I(s) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin(sr)}{sr} dr \quad (4-6)$$

where  $D_{\max}$  is the maximal distance within the particle.

At low resolution, the intensities of scattered X-rays can be approximated by the Guinier relation (Mertens and Svergun 2010)

$$I(s) = I(0) \exp\left(-\frac{1}{3} R_g^2 s^2\right) \quad (4-7)$$

where  $R_g$  is the radius of gyration of the particle. Generally, this approximation is assumed to be valid for a range of  $R_g * s < 1.3$ .

$R_g$  can also be extracted from the  $p(r)$  (Koch and Svergun 2003)

$$R_g^2 = \frac{\int r^2 p(r) dr}{2 \int p(r) dr} \quad (4-8)$$

For well behaved samples, a plot of  $\ln I(s)$  over  $s^2$  results in a linear relation from which  $I(0)$  and  $R_g$  can be extracted. Deviations from linearity indicate nonideal samples and may influence further data processing steps (Mertens and Svergun 2010) (Figure 4.3).

Indirect Fourier transform (Glatter 1977), a mathematical method allowing the construction and evaluation of trial electron distribution functions ( $p(r)$ ) against experimental scattering data, provides the electron pair distribution function  $p(r)$ . This function provides a histogram of distances between electron pairs in the scattering particles. It can be regarded as the SAXS equivalent of the Patterson function in X-ray crystallography (Putnam et al. 2007).

At low angles, the scattering intensities of folded macromolecules follow Porods law (Putnam et al. 2007):

$$I(s) \propto s^{-4} \quad (4-9a)$$

However, at high scattering angles, this relation does not hold and the equation must be expanded to describe the fractal degrees of freedom (df) (Putnam et al. 2007):

$$I(s) \propto s^{-df} \quad (4-9b)$$

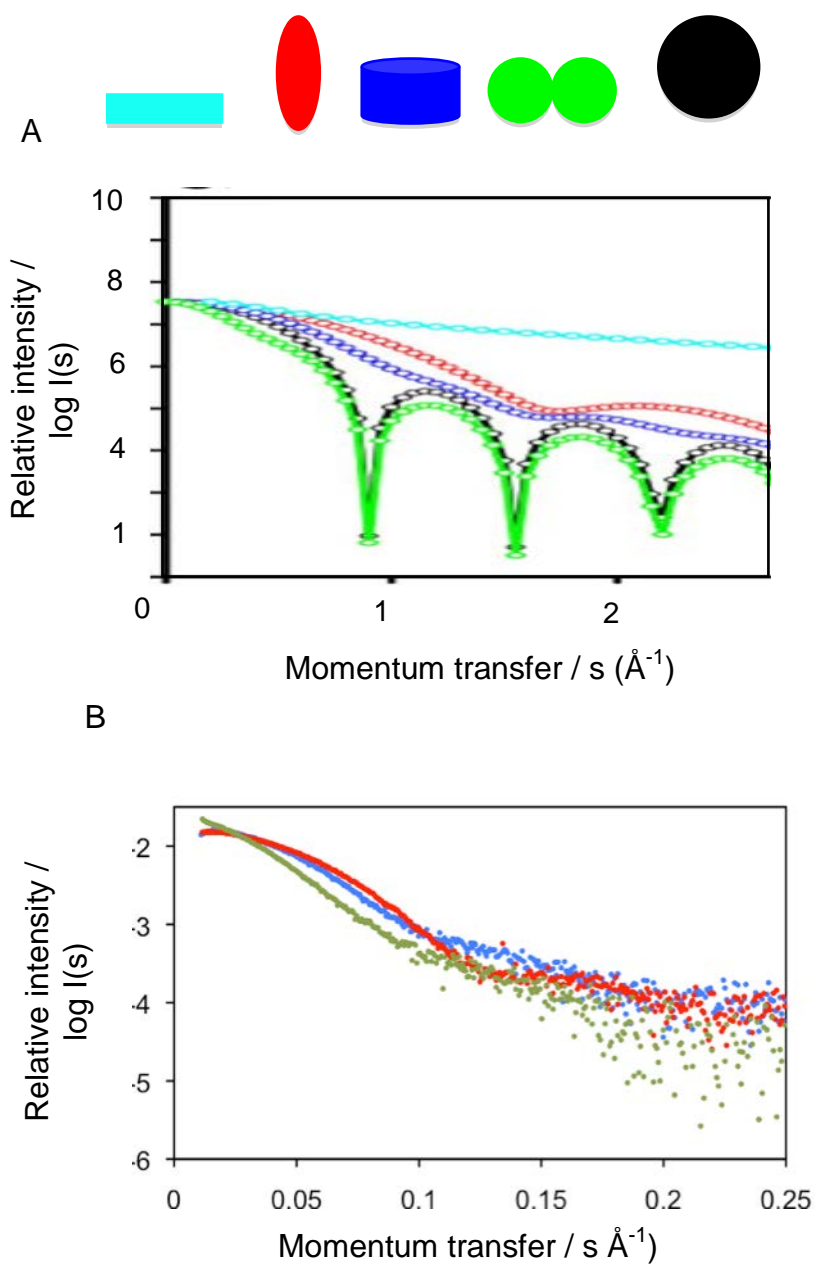
Due to the random orientation of unfolded coils, the scattering interference patterns of these cancel out and intensities decay with a df that is significantly different from folded globular proteins (5/3 and 4, respectively) (Putnam et al. 2007). Because of this phenomenon, a Kratky representation of  $I(s)$  over  $s^2$  can reveal the degree of unfolding. Folded proteins typically yield a symmetric peak at low angles followed by a plateau at high angles. Unfolded domains lack this clear peak and display an increase of  $I(s)$  with  $s^2$  (Doniach 2001; Putnam et al. 2007) (Figure 4.5).

### 4.1.2 Scattering by Lsm polyprotein complexes

Having prepared stable samples of Lsm[4+1]<sub>4</sub>, Lsm[2+3]<sub>4</sub> and Lsm[2+3]<sub>8</sub>, these could be analysed using the SAXS beamline at the Australian Synchrotron. In the Melbourne setup available to users, it is possible to collect SAXS data in-line with an SEC system. Thus, protein preparations can be separated into their component oligomeric fractions and individually investigated. Figure 4.2 compares the theoretical scattering intensities of typical geometrical shapes with the experimental scattering data collected for Lsm polyproteins. To ensure monodispersity, all samples are injected on a Superdex 200 5/150 GL gel filtration column in line with the SAXS beamline. Data were continuously collected as samples eluted from the column. Scattering patterns from Lsm polyproteins are radially averaged to derive one-dimensional scattering curves and scattering contributions of the buffer were subtracted from the presented curves.

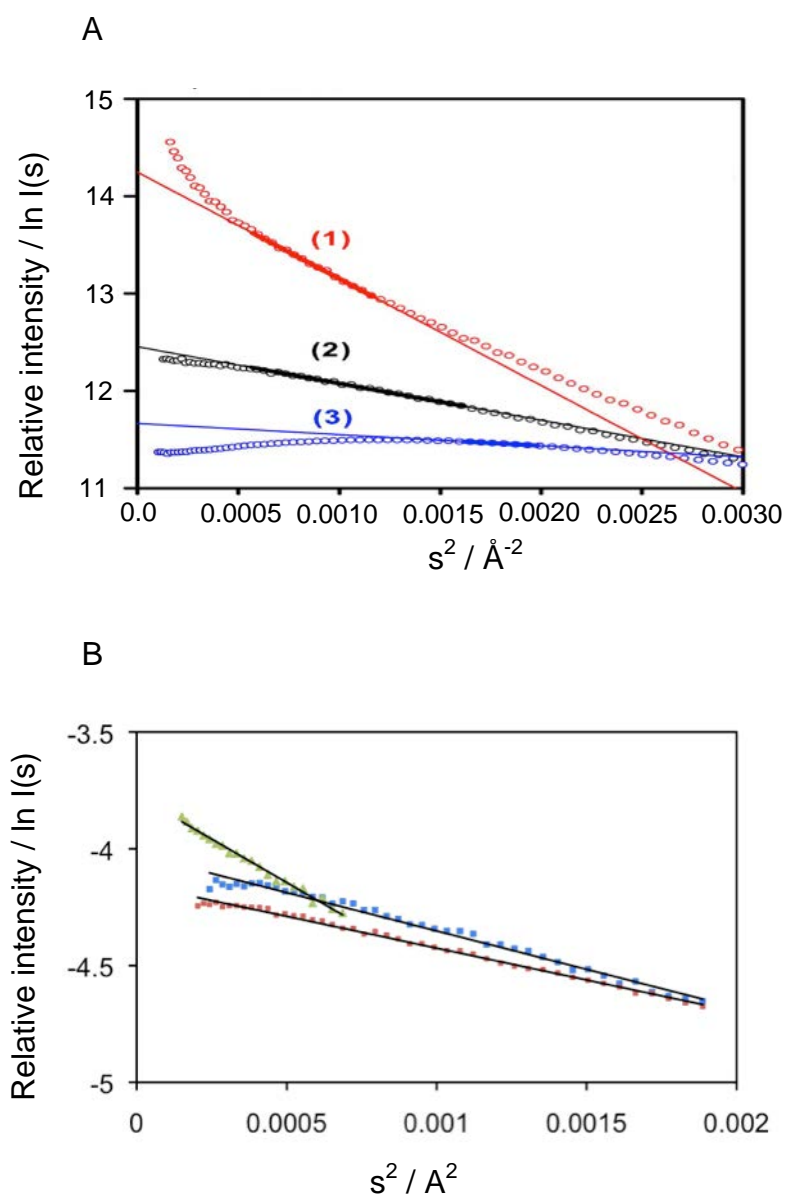
Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> assemblies yield scattering curves reminiscent of cylindrical particles, while Lsm[2+3]<sub>8</sub> seems to be slightly more elongated.

To assess data quality and to allow more precise definition of molecular dimensions, Guinier plots were calculated (Figure 4.3). The fact that scattering of all the polyprotein complexes yields linear Guinier plots and that scattering data did not show any upward curvature at low angles allows us to exclude sample aggregation. For the sample of Lsm[4+1]<sub>4</sub>, a small degree of inner-particle repulsion is indicated by a slight intensity decrease of the seven first data points. Seven data points were consequently removed from these data. Overall, however, good linearity is observed, signifying data quality is sufficient for further processing and shape determination.



**Figure 4.2|** Scattering curves of Lsm polyproteins and typical geometric shapes. A) Scattering curves for typical geometric shapes long rod (cyan), prolate (red), cylinder (blue), dumbbell (green) and sphere (black). Figure adapted from Mertens and Svergun (2010). B) Recorded buffer subtracted scattering curves for Lsm[4+1]<sub>4</sub> (blue), Lsm[2+3]<sub>4</sub> (red) and Lsm[2+3]<sub>8</sub> (green).





**Figure 4.3** Guinier plots of Lsm polyproteins and BSA samples. A) BSA samples showing aggregation (1), good data (2) and inner particle repulsion (3). Figure adapted from Mertens and Svergun (2010). B) Guinier fits derived for tetrameric Lsm[4+1]<sub>4</sub> (blue), Lsm[2+3]<sub>4</sub> (red) and Lsm[2+3]<sub>8</sub> (green). Intensities are scaled according to concentration and to  $I_0$ .

The slopes of the Guinier plots yield  $R_g$  values in reciprocal space (equation 4-7). Results obtained were as follows: for tetrameric complexes of Lsm[4+1] and Lsm[2+3]  $R_g^{\text{rec}}$  are calculated as  $31.9 \pm 0.52 \text{ \AA}$ ,  $29.2 \pm 0.4 \text{ \AA}$ , respectively. The octameric species analysed, namely Lsm[2+3]<sub>8</sub>, yielded scattering consistent with  $R_g^{\text{rec}}$  of  $46.8 \pm 1.82 \text{ \AA}$ . A comparison of all listed parameters is listed in Table 4.1.

**Table 4.1|** Geometric parameters derived from scattering data of Lsm polyproteins

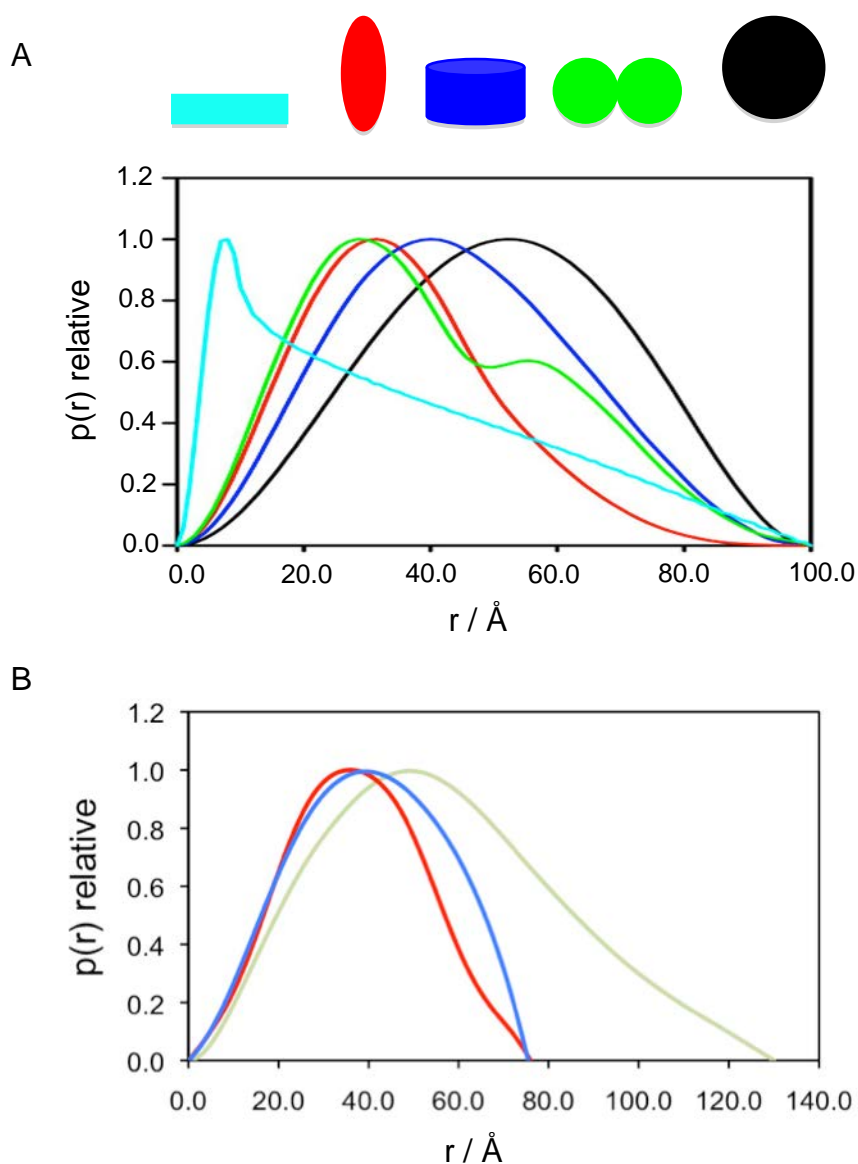
Parameter <sup>a</sup>	Lsm[4+1] <sub>4</sub> <sup>b</sup>	Lsm[2+3] <sub>4</sub> <sup>b</sup>	Lsm[2+3] <sub>8</sub> <sup>c</sup>	Lsm[2+3] <sub>4</sub> + U <sub>10</sub> <sup>b</sup>
$R_g^{\text{rec}}$ (Å)	$31.9 \pm 0.52$	$29.2 \pm 0.4$	$46.8 \pm 1.82$	$29.5 \pm 0.2$
$R_g^{\text{real}}$ (Å)	$30.7 \pm 0.05$	$28.4 \pm 0.08$	$44.4 \pm 0.35$	$28.4 \pm 0.06$
$D_{\text{max}}$ (Å)	75.5	75.5	135.0	77.5
$\chi^2$	$0.81 \pm 0.002$	$0.49 \pm 0.001$	$0.41 \pm 0.002$	$0.44 \pm 0.001$
NSD	$0.53 \pm 0.02$	$0.50 \pm 0.04$	$0.67 \pm 0.03$	$0.56 \pm 0.05$

<sup>a</sup>  $R_g^{\text{rec}}$ , reciprocal space radius of gyration derived from Guinier approximation;  $R_g^{\text{real}}$ , real space radius of gyration from distance distribution function;  $D_{\text{max}}$ , maximum particle diameter;  $\chi^2$ , averaged discrepancies of experimental scattering data and eight independent *ab initio* models; NSD, averaged normal spatial discrepancies for eight independent *ab initio* models.

<sup>b</sup> Data obtained in Tris buffer (20 mM, pH 8.0) with NaCl (400 mM) and glycerol (10 % v/v)

<sup>c</sup> Data obtained in Tris buffer (20 mM, pH 8.0) with NaCl (200 mM)

Indirect Fourier transformation of collected data (GNOM) produced smooth bell-shaped electron-pair distribution functions ( $p(r)$ ) for all complexes. For tetrameric groupings, i.e. polyproteins Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub>, real space radii of gyration ( $R_g^{\text{real}}$ ) of  $30.7 \pm 0.05 \text{ \AA}$  and  $28.4 \pm 0.08 \text{ \AA}$ , and maximal particle dimensions of 75.5 Å, respectively were obtained (Figure 4.4).  $R_g$  values calculated (CRY SOL) from the crystal structures of hexameric Hfq (PDB 1HK9), heptameric Lsm $\alpha$  (PDB 1I81) and octameric Lsm3 (PDB 3BW1) yielded 23.6, 25.4 and 29.8 Å, respectively. The value obtained for Lsm[3]<sub>8</sub> is consistent with the dimensions of polyproteins Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub>.

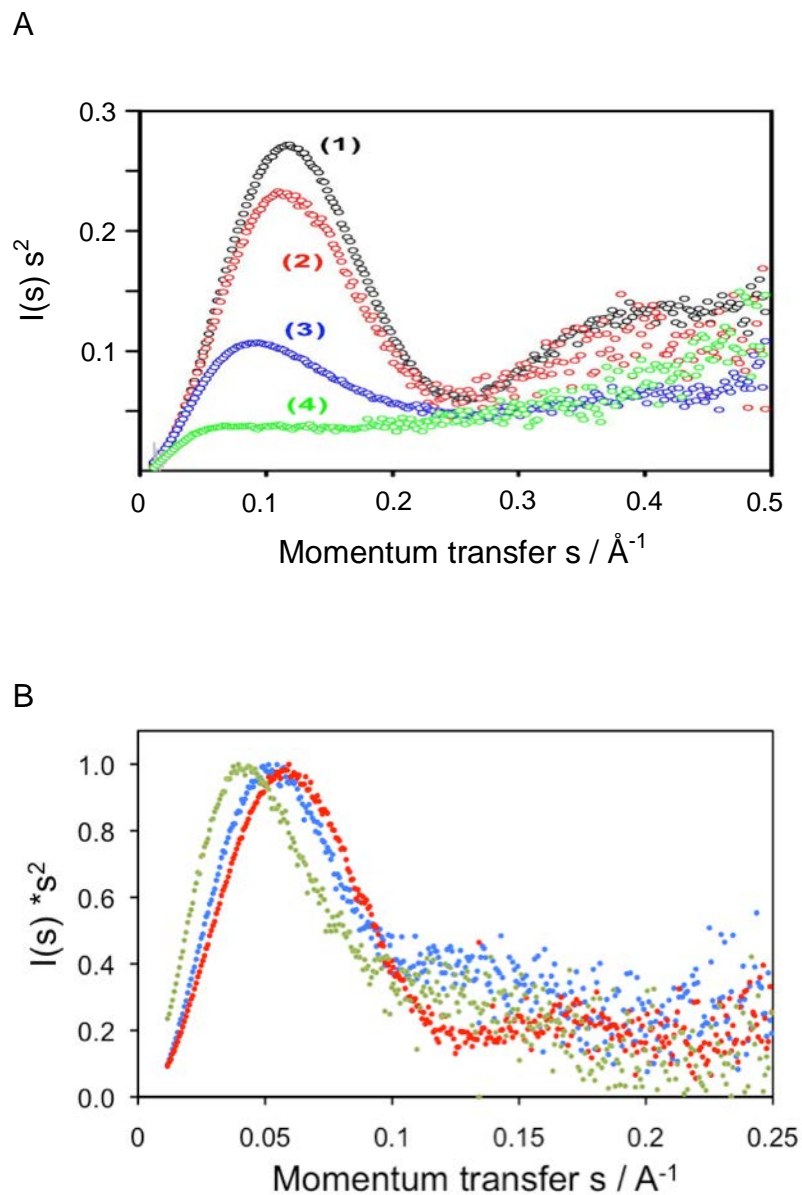


**Figure 4.4|** Electron distribution functions of Lsm polyproteins and typical geometric shapes. A) Long rod (cyan), prolate (red), cylinder (blue), dumbbell (green) sphere (black). Figure adapted from Mertens and Svergun (2010). B) Lsm[4+1]<sub>4</sub> (blue), Lsm[2+3]<sub>4</sub> (red) and Lsm[2+3]<sub>8</sub> (green). Traces were normalised to unity at their maxima similar to (Nakel et al. 2010).

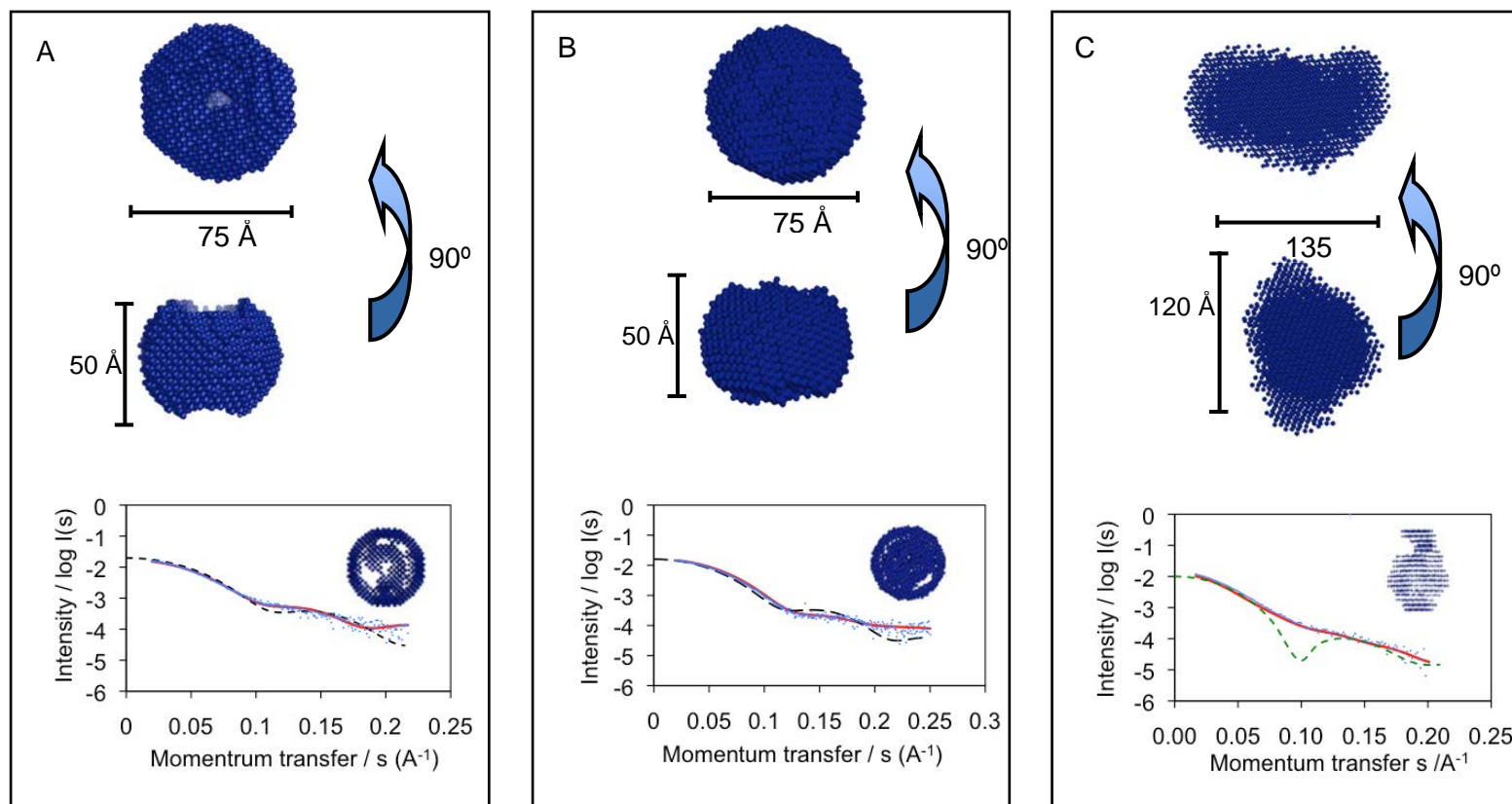
The  $p(r)$  function obtained for  $\text{Lsm}[2+3]_4$  shows a slightly narrower distribution, possibly due to a slightly more compact conformation.  $\text{Lsm}[2+3]_8$  displays an elongated distribution function yielding an  $R_g^{\text{real}}$  value of 44.4 Å consistent with a  $D_{\text{max}}$  of 135 Å. The real space radii of gyration are in good agreement with the reciprocal space data derived from the Guinier fit.

It is possible to detect from these data that all polyproteins exist in a folded globular state. All samples yield Kratky plots with a clear peak at low  $s$  values (Figure 4.5). No systematic signal increase is apparent and the peak at low angles is followed by a plateau at higher  $s$  values. For comparison, all three Kratky plots are normalised with respect to  $I(s) * s^2$ .

A more detailed picture concerning the molecular forms within my samples can be obtained from the  $p(r)$  distributions, from which *ab initio* calculations (DAMMIN) lead to three-dimensional molecular envelopes.  $\text{Lsm}[4+1]_4$  and  $\text{Lsm}[2+3]_4$  yield disc shaped bead models each with lateral diameters of 75 Å and a height of 50 Å. In the case of  $\text{Lsm}[4+1]_4$ , a central pore can be resolved within this solution structure as seen in Figure 4.6. Crystal structures of Lsm complexes vary in complex diameters between 60-75 Å (Moll et al. 2011). Using the previously determined crystal structure of the ring shaped  $\text{Lsm}[3]_8$  complex, a diameter for this simple octamer can be determined as 75 Å. Thus, the observed particle dimensions for the tetrameric Lsm polyproteins appear to be consistent with this molecular form.



**Figure 4.5]** Kratky plots of Lsm polypeptide and lysozyme scattering data. Kratky plots are normalised to unity at  $I(s) s^2$ . A) Folded (1), partially unfolded in 8M urea (2), partially unfolded at 90 °C, unfolded (4) lysozyme in 8 M urea at 90 °C. Figure adapted from Mertens and Svergun (2010). B) Lsm[4+1]<sub>4</sub> (blue), Lsm[2+3]<sub>4</sub> (red) and Lsm[2+3]<sub>8</sub> (green).



**Figure 4.6** | Lsm polyprotein solution structures. Sorted and averaged bead models are displayed on top. Experimental scattering data (blue) is fitted to simulated curves of the best bead models (red) and to the theoretical scattering pattern calculated from the crystal structure of a single Lsm[3]<sub>8</sub> ring (PDB 3BW1, black) A) or two stacked Lsm[3]<sub>8</sub> rings (green). A) Lsm[4+1]<sub>4</sub> B) Lsm[2+3]<sub>4</sub> C) Lsm[2+3]<sub>8</sub>.

Octameric Lsm[2+3]<sub>8</sub> yields a more elongated shape consistent with its scattering profile and p(r) function. *Ab initio* shape reconstruction yields a long cylindrical shape. However, the dimensions obtained for the Lsm[2+3] octamer do not agree with two stacked tetrameric rings. As a consequence, it can not be excluded that Lsm[2+3]<sub>8</sub> adopts a solution structure different from two stacked rings.

Molecular envelopes are calculated via simulated annealing without imposing symmetry constraints on the particle shape using space group P1. Eight bead models have been generated in this way and have subsequently been averaged to obtain a final model. Obtained models can be fitted to the experimental data with average  $\chi^2$  values of 0.8, 0.5 and 0.4 for Lsm[4+1]<sub>4</sub>, Lsm[2+3]<sub>4</sub> and octameric Lsm[2+3]<sub>8</sub>, respectively. Figure 4.6 demonstrates the agreement of the best models obtained for the polyproteins with the experimental data. In addition, the average spatial discrepancy (NSD) between the input models is determined as 0.53, 0.56 and 0.67. This parameter allows the quantitative comparison of the agreement between three-dimensional shapes of the same resolution (Kozin and Svergun 2001). For perfectly superimposed objects, the NSD tends to 0. It exceeds 1, if systematic differences are found between the models, hence generally values < 1 are expected for SAXS envelopes. This suggests that the determined molecular envelopes match the experimental data and individual reconstructions used for averaging are highly similar. Structural parameter of Lsm polyproteins determined by SAXS are summarised in Table 4.1.

Scattering data and SAXS envelopes of Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> are compared to a theoretical scattering curve derived from the crystal structure of octameric Lsm[3]<sub>8</sub>

(PDB 3BW1) (Figure 4.6). The scattering data fit the Lsm[3]<sub>8</sub> derived curve with  $\chi^2$  values of 0.97 and 0.90, respectively, indicating a very good fit. However, visual inspection of the curves indicates some discrepancies between X-ray scattering from Lsm[2+3]<sub>4</sub> and theoretical scattering data from Lsm[3]<sub>8</sub>. Comparison of scattering from Lsm[2+3]<sub>8</sub> with a scattering curve derived from two stacked Lsm3 octamers (PDB 3BW1) yields a  $\chi^2$  value of 1.64 and visual inspection confirms a poor fit. Thus, Lsm3 makes a poor model for Lsm[2+3]<sub>8</sub>.

SAXS data suggest Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> form compact, globular assemblies of ring or disc-like morphologies in solution. The dimensions of these particles (75 Å diameter, 50 Å wide) are compatible with the crystal structure of Lsm[3]<sub>8</sub> (Naidoo et al. 2008) and thereby indicate similar conformations of the polyproteins.

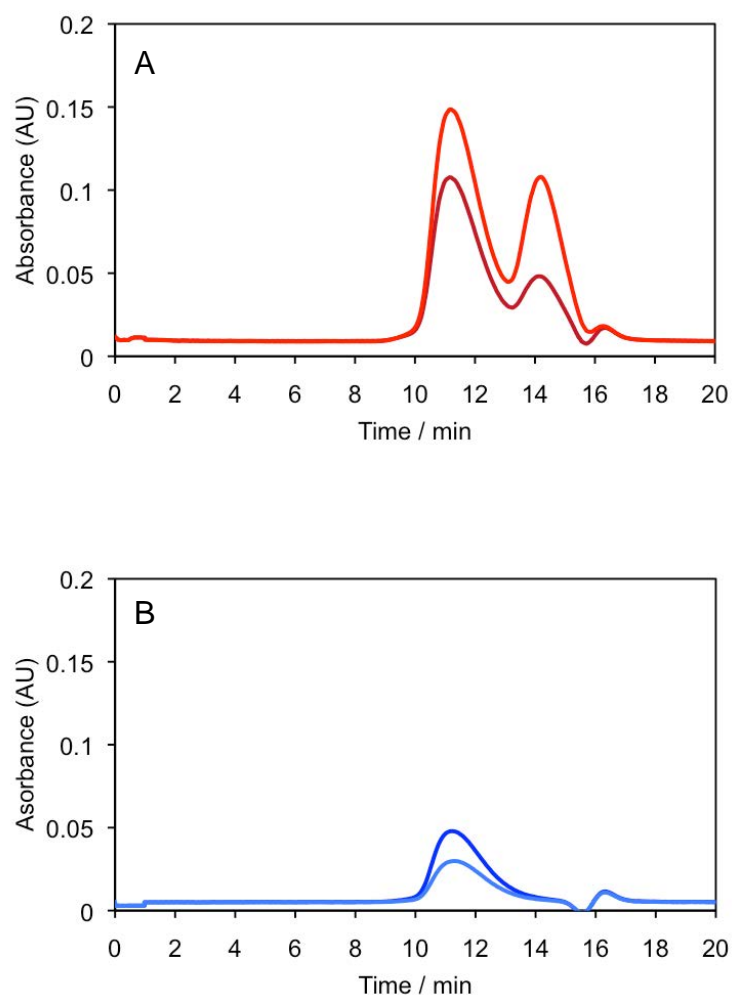
#### 4.1.3 Solution shape of Lsm[2+3]<sub>4</sub> with RNA

By preparing tetrameric Lsm[2+3] in the presence of U<sub>10</sub> (by incubation for 30 min), it is possible to collect SAXS data for comparison with the unbound form of the protein. I demonstrate in later sections of this chapter that this RNA species engages in significant binding with Lsm[2+3]<sub>4</sub>.

Figure 4.7 shows a comparison of the SEC chromatograms for Lsm[2+3]<sub>4</sub> with and without U<sub>10</sub>. In Tris buffer (20 mM, pH 8.0) containing moderately high levels of NaCl (400 mM) and glycerol (10 % v/v), the polyprotein itself produces a single species with an A<sub>260</sub>/A<sub>280</sub> ratio of 0.6 which is indicative of pure protein (Warburg and Christian 1941). Two peaks are obtained for Lsm[2+3]<sub>4</sub> with U<sub>10</sub>. The first peak



elutes at an elution time consistent with the mass of free Lsm[2+3]<sub>4</sub>, but displaying  $A_{260}/A_{280}$  of 1.4.



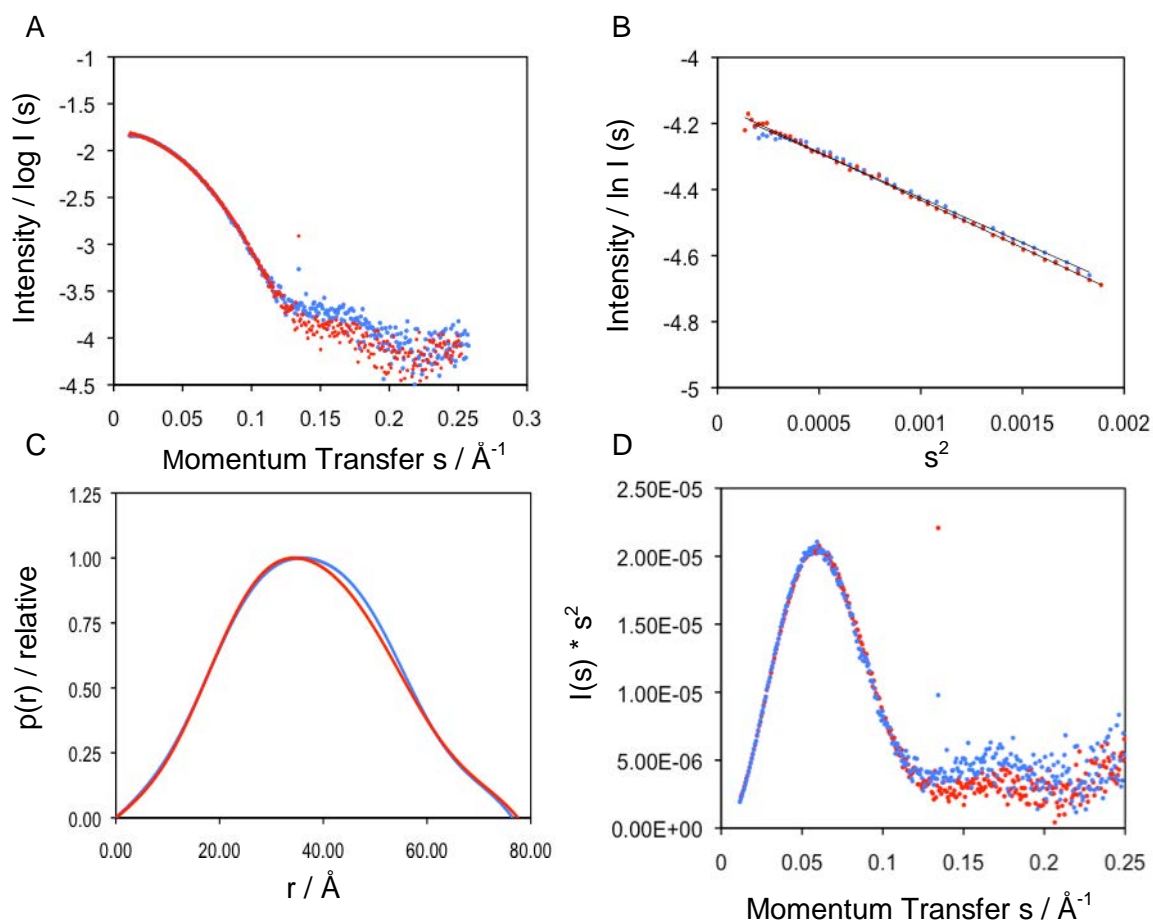
**Figure 4.7** | Size separation of Lsm polypeptide complexes prior to SAXS data collection. A) Lsm[2+3]<sub>4</sub> preincubated with U<sub>10</sub> in 1:2 ratio prior to SEC, B) Lsm[2+3]<sub>4</sub>. SEC is performed on a Superdex 200 5/150 GL column ( $V_c = 3$  ml) at 0.2 ml/min in Tris buffer (20 mM, pH 8.0) containing NaCl (400 mM) and glycerol (10 % v/v). Sample absorbances recorded at wavelengths of 260 nm and 280 nm are displayed in blue for Lsm[2+3]<sub>4</sub> ( $A_{260}$  —,  $A_{280}$  —) and in red for the Lsm[2+3]<sub>4</sub>-RNA complex ( $A_{260}$  —,  $A_{280}$  —).

A second fraction is attributed to excess RNA due to its long elution time and high  $A_{260}/A_{280}$  ratio (2.5). An  $A_{260}/A_{280}$  ratio of 1.4 obtained for peak I (Figure 4.7) clearly demonstrates the presence of protein and RNA in this fraction.

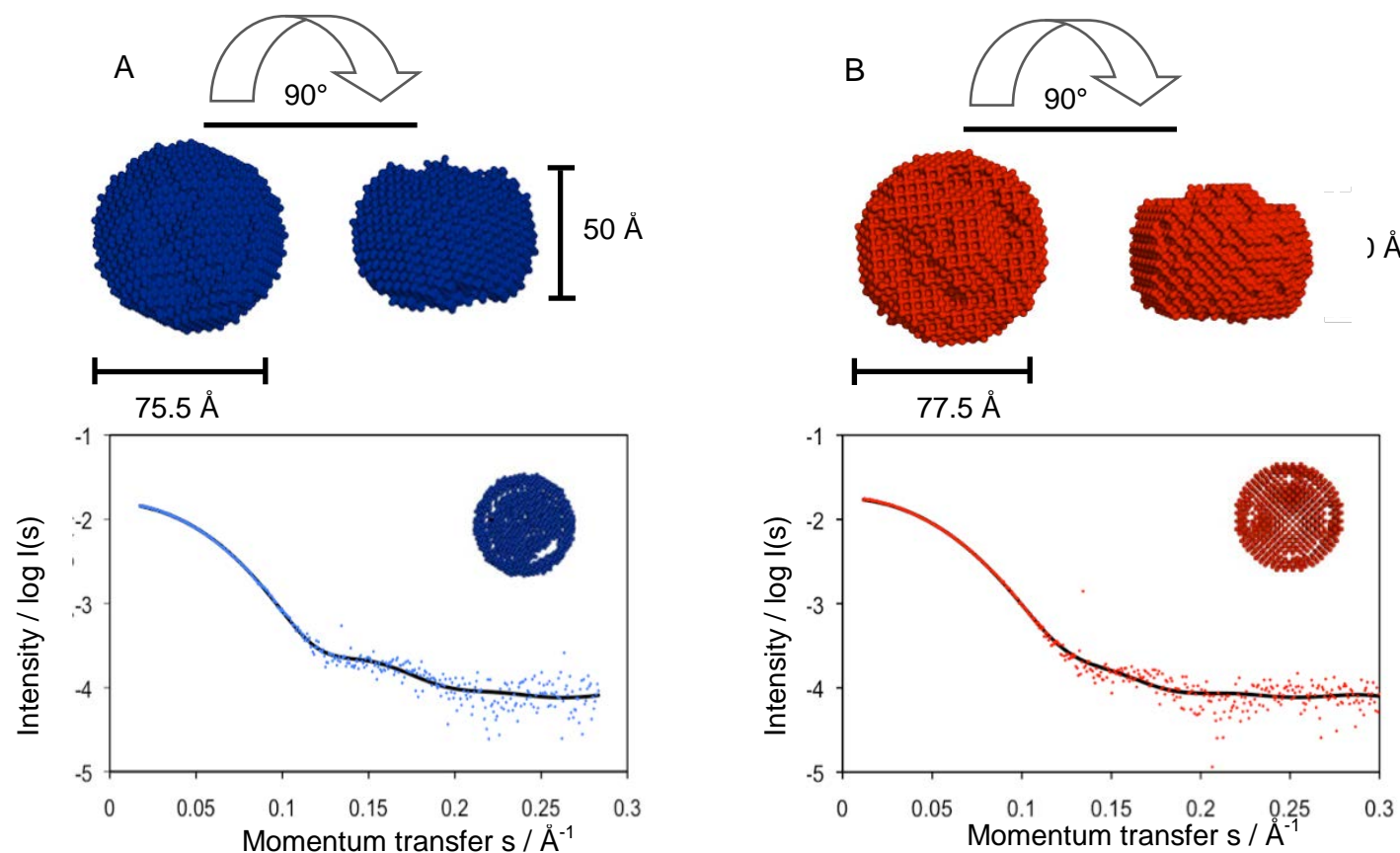
Protein and protein-RNA complexes of Lsm[2+3]<sub>4</sub> overall show very similar SAXS patterns (Figure 4.8). The radially averaged scattering curves of both samples are reminiscent of a disc shaped or cylindrical particle. No up or downward curvature of the scattering data is visible at lower angles in Guinier plots. Thus, sample aggregation or inner-particle repulsion in the RNA complex are excluded (compare Figure 4.3). The  $R_g$  for the RNA-complex is extracted from the slope of the Guinier plot as  $29.5 \pm 0.2 \text{ \AA}$ , only marginally greater than the value ( $29.2 \text{ \AA}$ ) for the protein itself. From the bell shaped  $p(r)$  function a maximal particle diameter of  $77.5 \text{ \AA}$  is estimated, an  $R_g^{\text{real}}$  of  $28.4 \pm 0.06 \text{ \AA}$  is determined. These are again only slightly larger than the dimensions of the free Lsm complex ( $D_{\text{max}} = 75.5 \text{ \AA}$ ;  $R_g = 28.4 \text{ \AA}$ ). Figure 4.8 displays an overlay of both  $p(r)$  functions.

From Guinier and  $p(r)$  analysis the dimensions of Lsm[2+3]<sub>4</sub> with RNA are found slightly increased to the protein complex itself indicating the formation of a protein-RNA complex. Both samples display very similar  $p(r)$  functions reminiscent of globular shape, suggesting RNA binding does not involve a pronounced conformational change.

Kratky plots of both samples overlay perfectly and display a sharp peak at low angles followed by a plateau at higher angles indicating similar globular folds. Hence, it can



**Figure 4.8** | SAXS comparison of tetrameric Lsm[2+3] with the same protein preincubated with U<sub>10</sub>. Data were recorded at the SAXS/WAXS beamline at the Australian Synchrotron (Melbourne, Victoria). Red traces represent data from the Lsm-RNA complex, blue traces data recorded for Lsm[2+3]<sub>4</sub>. A) Forward scattering intensities. B) Guinier approximation. C) Distance distribution functions. For comparison distance distribution functions are normalised to unity at their maxima similar to Nakel et al. (2010). D) Kratky representation of scattering data.



**Figure 4.9** | *Ab initio* models of Lsm[2+3]<sub>4</sub> (blue) and a Lsm[2+3]<sub>4</sub>-U<sub>10</sub> complex (red). Averaged models from DAMMIN are displayed on top. Fits of the best envelope obtained to the experimental data are displayed below. A) Lsm[2+3]<sub>4</sub>;  $\chi^2 = 0.49$ . B) Lsm[2+3]<sub>4</sub> complexed with U<sub>10</sub>;  $\chi^2 = 0.44$ .

be reasoned that RNA-binding did not induce any profound conformational changes in Lsm[2+3]<sub>4</sub>.

A molecular envelope was determined for the Lsm-RNA complex from the  $p(r)$  functions by simulated annealing. Eight disc-shaped models were calculated and averaged yielding a final model that matches experimental data with an average  $\chi^2$ -value of 0.49 (Table 4.1). Figure 4.9 demonstrates the agreement of the best models obtained for Lsm[2+3]<sub>4</sub> and the protein with U<sub>10</sub> with experimental data.

## **4.2 Surface plasmon resonance studies using immobilised RNA**

### **4.2.1 Theory of surface plasmon resonance**

Biomolecular interactions can be monitored in realtime by surface plasmon resonance (SPR). SPR is observed when a polarised light beam hits a thin metal film between two transparent media at a critical angle that creates total internal reflection (Harrick 1967). A general experimental setup is depicted in Figure 4.10. At the metal interface, photons from the incident beam are absorbed and their energy converted into an evanescent wave that propagates into the medium of lower refractive index (Harrick 1967; Kretschmann and Raether 1968; Katsamba et al. 2002). At a distinct angle ( $\theta$ , the SPR angle), this energy transfer is maximal and is detected as a loss in intensity of the reflected beam. The value of the SPR angle depends on the refractive index of the two media forming the interface.

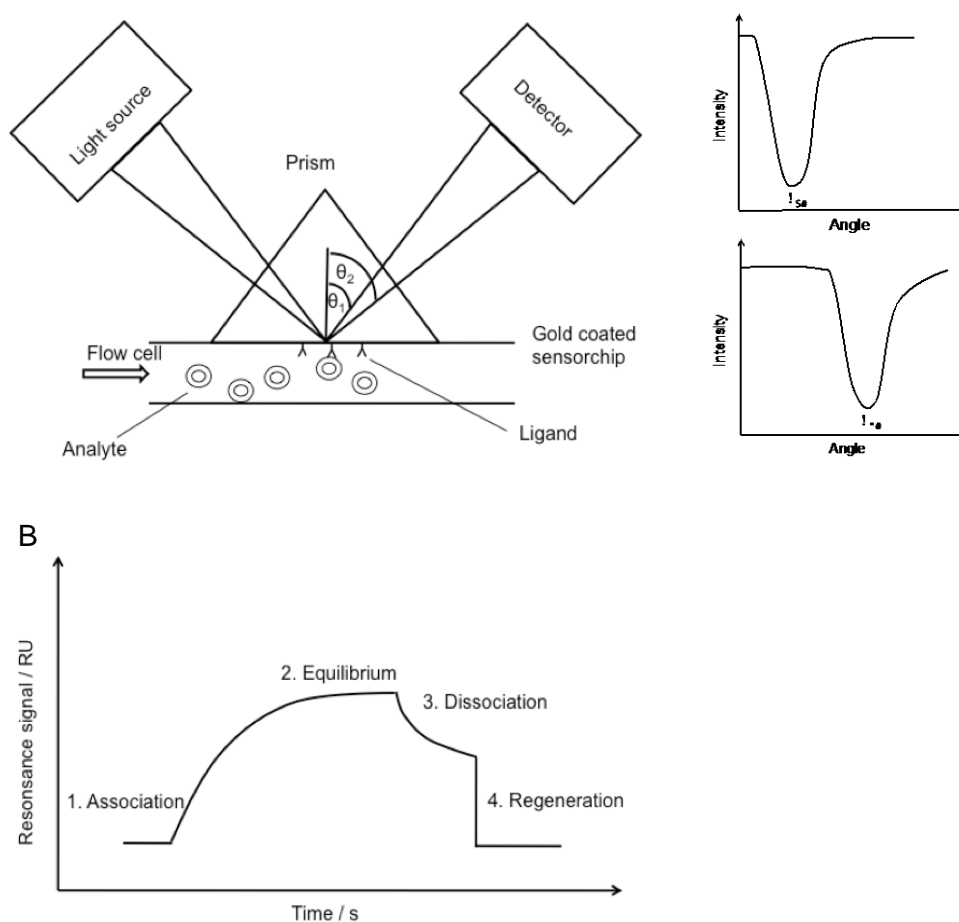
In an SPR-based interaction experiment, a ligand is immobilised in a flow cell at the interface of two transparent media. Following introduction of analyte, binding to the ligand changes the refractive index of the media by an amount proportional to the mass of bound analyte (Katsamba et al. 2002). Continuous detection of the SPR angle over time results in a sensorgram characterised by association and dissociation. In special cases when the analyte concentration is sufficient and the injection time long enough, an equilibrium phase may be reached. A regeneration step removes any analyte that remains bound after the dissociation phase and brings the signal back to baseline.

In contrast to techniques such as electromobility shift assays or ITC, which monitor an interaction at equilibrium, SPR can be used to monitor the kinetics of protein-RNA binding processes. Therefore, association and dissociation rates can be determined in addition to the overall equilibrium constant.

In an ideal case where the rate of mass transport of an amount [A] of analyte to an amount [B] of ligand is much higher than the association rate of analyte and ligand, the binding reaction can be described by (Karlsson 1994):



where  $k_a$  and  $k_d$  describe association and dissociation rate constants and [AB] is the amount of the resulting complex.



**Figure 4.10|** Basic principle of SPR-based interaction experiments adapted from Katsamba et al. (2002). A) A wedge shaped beam of light is directed through a prism onto a gold coated interface formed by a flow cell and its glass wall. At distinct angles greater than a critical angle, the beam is reflected from the surface and transfers energy into the gold film. This energy transfer is maximal at a specific reflection angle ( $\theta_1$ ), the SPR-angle proportional to the mass immobilised on the surface of the flow cell. A detector records the intensity of the reflected beam as a function of the reflection angle and the SPR angle is identified as a sharp dip in intensity. As a consequence of analyte binding to ligand that is immobilised on the sensorchip surface, the SPR-angle changes ( $\theta_2$ ). B) A plot of the SPR-angle over time produces a sensorgram displaying characteristic association, equilibrium and dissociation phases.

For the equilibrium association ( $K_A$ ) and dissociation ( $K_D$ ) constants follows:

$$K_A = \frac{[AB]}{[A][B]} = \frac{k_a}{k_d} = \frac{1}{K_D} \quad (4-11)$$

The kinetics of complex formation then follow a pseudo-first-order reaction

(O'Shannessy and Winzor 1996):

$$\frac{d[AB]}{dt} = k_a[A][B] - k_d[AB] \quad (4-12)$$

In a SPR experiment, parameters  $[AB]$ ,  $[A]$  and  $[B]$  are replaced by the binding response  $R$  (in response units RU), the analyte concentration  $C$  and the concentration of free ligand  $R_{\max} - R$ :

$$\frac{dR}{dt} = k_a C(R_{\max} - R) - k_d R \quad (4-13)$$

where  $R_{\max}$  is the maximum response possible for the RNA surface.

SPR has been successfully applied to describe protein-RNA interactions in a variety of studies. Examples include the characterisation of the interaction of spliceosomal U1-A with hairpin II of U1 snRNA (Katsamba et al. 2001) and elucidation of an mRNA binding signature motif in T-cell restricted intracellular antigen 1 related protein (TIAR) (Kim et al. 2007). In these studies, RNA was immobilised in the flow cell for subsequent kinetic measurements, as the detected SPR signals then



responded to the large change resulting from protein binding. In addition, commercial sensorchips are generally coated with a negatively charged dextran matrix that may disturb kinetic measurement of nucleic acid analytes due to repulsion effects from the surface. Finally, RNA oligomers can easily be modified to carry terminal extensions allowing their immobilisation to the surface chemistry of the flow cell.

Protein-RNA interaction experiments using SPR require careful preparation to exclude potential pitfalls including steric hindrance, surface effects, non-specific binding, mass-transport limitations, and poor sample quality (Katsamba et al. 2002). Steric hindrance and surface effects can be limited by spacer sequences between the dextran surface and the sequence of interest. Addition of bovine serum albumin (BSA) and surfactants (P20) in sample buffers and the use of blank reference cells minimise non-specific binding, while mass-transport effects can be reduced by the use of high flow rates and low ligand immobilisation levels.

#### **4.2.2 SPR of Lsm polyproteins**

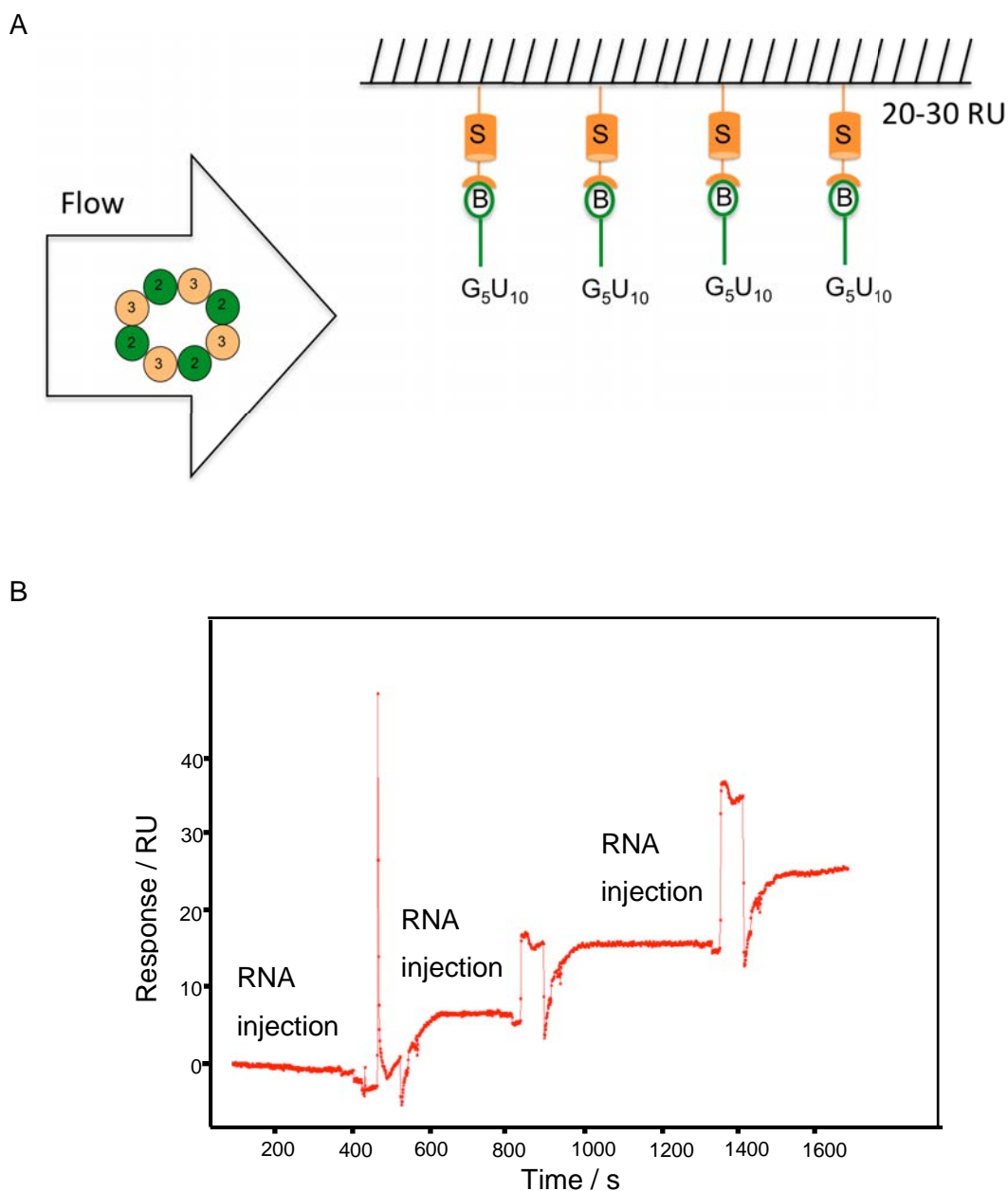
For real-time Lsm-RNA interaction experiments, a U-rich RNA sequence ( $U_{10}$ ) was chosen based on previous electro mobility shift assays (Sobti et al. 2010). The RNA oligomer contained a 5' biotin group to allow immobilisation to a streptavidin coated sensor chip surface and a spacer consisting of five guanosine nucleotides connecting the  $U_{10}$  sequence to the biotin group. This provides sufficient length for the interaction of one Lsm subunit with one nucleotide and to minimise steric hindrance

and surface effects. A schematic outline of the SPR setup utilised for this study is shown in Figure 4.11.

Prior to immobilisation, RNA was deprotected and diluted (2 nM) in Tris buffer (20 mM, pH 8.0) containing NaCl (200 mM) and surfactant P20 (0.05 %) and injected over the flow cell sufficient to generate a low density (25 RU) surface. Figure 4.10 demonstrates the successful immobilisation of RNA on a sensor chip.

Three Lsm complexes of different compositions were analysed: i) Lsm[2+3]<sub>4</sub>, ii) Lsm[3]<sub>8</sub> and iii) Lsm[4+1]<sub>4</sub>. Lsm complexes were injected at concentrations of 0.1–10 µM surrounding a  $K_D$  of 450 nM obtained in gelshift experiments with tetrameric Lsm[2+3] and U<sub>10</sub> RNA (Sobti et al. 2010). Signals obtained from blank flow cells were subsequently subtracted from signals of RNA surfaces. Since the absolute SPR-response is proportional to the molecular weight of the analyte, all traces displayed here are normalised with respect to analyte molecular weight (Figure 4.12).

Injection of Lsm[2+3]<sub>4</sub> over the low density RNA surface results in a very steep association curve converging to an equilibrium state at 10 µM concentration. At lower protein concentrations, no equilibrium is reached. However, steep association phases are detected for 0.5-1 µM concentration. At 0.1 µM concentration, the slope of the association phase is significantly reduced. Only slow signal decay is observed following the injections of buffer over the flow cells.



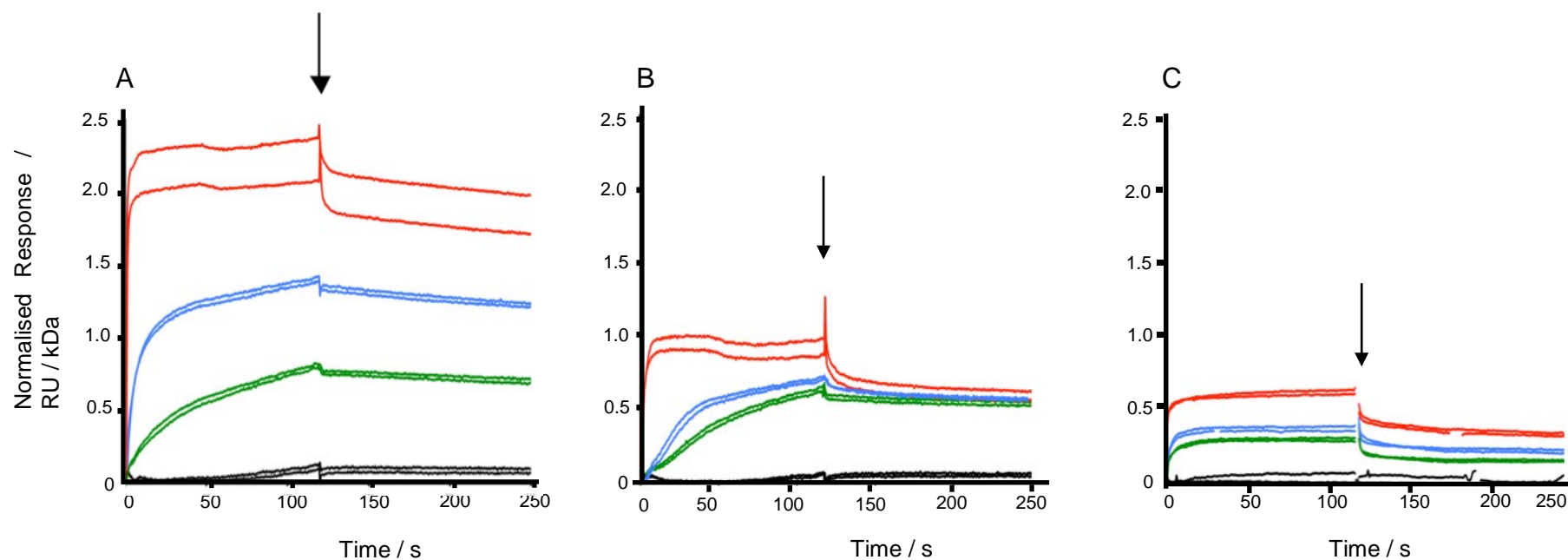
**Figure 4.11|** Setup of Lsm-RNA surface plasmon resonance experiments. A) Schematic diagram of Lsm-RNA binding experiments. 5' biotinylated RNA (G<sub>5</sub>U<sub>10</sub>) is immobilised in a streptavidin coated flow cell. Following RNA immobilisation, purified Lsm polyproteins are injected over the sensorchip surface. S (streptavidin), B (biotin). B) Realtime detection of RNA immobilisation levels on the chip.

The slopes observed for association and dissociation curves suggest a high association rate constant along with a very low dissociation constant, consistent with the formation of a stable protein-RNA complex.

Lsm[3]<sub>8</sub> shows very similar behaviour to Lsm[2+3]<sub>4</sub>. High protein concentrations (10  $\mu$ M) display a rapid formation of an equilibrium response and very steep and short association phases. Reduction of the protein concentration results in a less pronounced slope of the association phases and the curves do not reach equilibrium. Dissociation phases are characterised by very shallow slopes. Analogous to Lsm[2+3]<sub>4</sub>, this indicates the formation of a stable protein-RNA complex.

By contrast, all trialled concentrations of Lsm[4+1]<sub>4</sub> produce sensorgrams that converge to an equilibrium response within seconds after injection consistent with rapid association as well as dissociation rates. These results indicate that all of the assayed Lsm complexes readily interact with the U-rich RNA oligonucleotide.

Interestingly, different equilibrium responses ( $R_{eq}$ ) are recorded for the individual Lsm complexes at 10  $\mu$ M concentration. The highest  $R_{eq}$  values are generated by Lsm[2+3]<sub>4</sub> followed by Lsm[3]<sub>8</sub> and then Lsm[4+1]<sub>4</sub>. Since SPR-traces are normalised with respect to the molecular weight of the individual complexes, these signals directly relate to different surface concentrations of the Lsm complexes at equilibrium. This can be explained with either different binding stoichiometries between the Lsm complexes and RNA or with different overall affinities. Assuming



**Figure 4.12** SPR traces of Lsm proteins with  $G_5U_{10}$ . Experiments with Lsm[2+3]<sub>4</sub> and Lsm[3]<sub>8</sub> are conducted in Tris buffer (20 mM, pH 8.0) containing intermediate concentrations (200 mM) of NaCl, BSA (62.5 mg/ml) and surfactant P20 (0.05 %). Lsm[4+1]<sub>4</sub> is analysed using moderately high concentrations of NaCl (400 mM). SPR signals are normalised with respect to the molecular weight of the Lsm complexes. Temperature and flow rate for protein injections are kept constant at 25 °C and 50  $\mu$ l/min. Injected protein concentrations are indicated by coloured lines ( — 10  $\mu$ M; — 1  $\mu$ M; — 0.5  $\mu$ M; — 0.1  $\mu$ M ) A) Lsm[2+3]<sub>4</sub>, B) Lsm[3]<sub>8</sub>, C) Lsm[4+1]<sub>4</sub>. Arrows depict injection stops.

a simple 1:1 interaction between the Lsm complexes and the RNA oligomer as is generally found for the RNA interactions of these complexes (Urlaub et al. 2001; Weber et al. 2010; Moll et al. 2011), we find that Lsm[2+3]<sub>4</sub> has the highest affinity for G<sub>5</sub>U<sub>10</sub> RNA followed by Lsm[3]<sub>8</sub> and then Lsm[4+1]<sub>4</sub>.

In order to quantify the underlying binding kinetics, it is in some cases possible to determine affinity rate constants from SPR data by non-linear regression using a global analysis (Karlsson 1994). In this analysis, data from several experiments sharing common variables (e.g. same analyte in different concentrations) are fitted simultaneously to obtain one set of parameters (rate constants, surface binding capacity) describing the interaction (Katsamba et al. 2002). Experimental curves are overlayed with simulated curves resulting from the fitted parameters and the goodness-of-fit is estimated by a  $\chi^2$ -test.

Global fitting of the data was attempted, however, the sensorgrams obtained for Lsm complexes did not fit a global 1:1 binding model as indicated by poorly overlayed curves and high  $\chi^2$ -values. Thus, reliable association and dissociation rate constants could not be determined for the specific Lsm-RNA complexes probed here. Though, all sensorgrams obtained for Lsm[4+1]<sub>4</sub> with G<sub>5</sub>U<sub>10</sub>-RNA produced equilibrium responses. If equilibrium responses for several analyte concentrations can be detected, steady-state analysis becomes possible (Myszka et al. 1998; Kortt et al. 1999). At equilibrium, association and dissociation rates are identical, the complex concentration and the related SPR-signal ( $R_{eq}$ ) are constant. This can be described by:

$$\frac{dR}{dt} = k_a C(R_{\max} - R_{eq}) - k_d R_{eq} = 0 \quad (4-14)$$

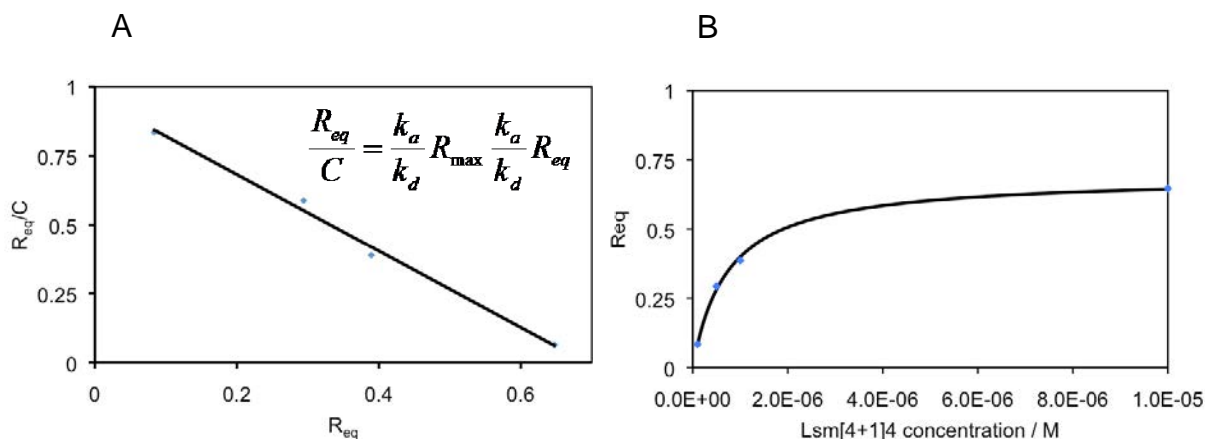
where  $k_a$  is the association rate constant,  $C$  the analyte concentration,  $R_{\max}$  the maximum binding capacity of the RNA surface and  $k_d$  the dissociation rate constant (Karlsson 1994).

On rearrangement, this relationship can be expressed as:

$$\frac{R_{eq}}{C} = \frac{k_a}{k_d} R_{\max} - \frac{k_a}{k_d} R_{eq} \quad (4-15)$$

Hence, the slope of a plot of  $R_{eq}/C$  against the equilibrium response  $R_{eq}$  provides  $k_a/k_d$ , i.e. the reciprocal of  $K_D$ , the equilibrium dissociation constant.

From a linear fit to this plot for Lsm[4+1]<sub>4</sub>, the relationship  $R_{eq}/C = -1.39 (R_{eq}) = 88.47$  is obtained (explained variance  $R^2 = 0.998$ ) and an equilibrium constant of  $K_D = 730$  nM is determined (Figure 4.13 A). Analysis by non-linear regression results in an identical  $K_D$  (Figure 4.13 B). The determined affinity constant, however, is significantly weaker than affinities determined for the interaction of other Lsm protein complexes with other RNA species (Table 4.2).



**Figure 4.13|** Steady state analysis of Lsm[4+1]<sub>4</sub> binding to G<sub>5</sub>U<sub>10</sub> as monitored by SPR in Tris buffer (20 mM, pH 8.0) containing NaCl (400 mM), BSA (62.5 mg/ml) and surfactant P20 (0.05 %). A) Scatchard plot for the binding data. The resulting relationship is  $R_{eq}/C = -1.39 (R_{eq}) + 88.47$  with an explained variance ( $R^2$ ) of 0.998. B) Non-linear regression of the binding data. The binding curve fits the experimental data with a  $\chi^2$ -value of 1.4.

**Table 4.2|** Selected interactions of Lsm proteins with RNA.

Lsm complex <sup>a</sup>	RNA	K <sub>D</sub>	Method	Reference
ScLsm[2+3] <sub>4</sub> <sup>b</sup>	U <sub>10</sub>	34 nM	ITC	<i>This thesis</i>
ScLsm[4+1] <sub>4</sub> <sup>b</sup>	G <sub>5</sub> U <sub>10</sub>	730 nM	SPR	<i>This thesis</i>
ScLsm[2+3] <sub>4</sub> <sup>b</sup>	U <sub>10</sub>	453 nM	Gelshift	(Sobti et al. 2010)
HsLsm1-7	U6 snRNA (ncRNA)	150-274 nM	Gelshift	(Licht et al. 2008)
ScLsm1-7	MFA2 (mRNA)	200 nM	Gelshift	(Chowdhury et al. 2007)
HvLsm $\alpha$	U <sub>30</sub>	72 nM	Gelshift	(Fischer et al. 2011)
EcHfq	DsrA (ncRNA)	21 nM, 94nM <sup>c</sup>	Gelshift	(Mikulecky et al. 2004)

<sup>a</sup> Proteins are named by the first letters of the species, followed by the type of protein.

<sup>b</sup> [ ]<sub>4</sub> indicates tetrameric Lsm complexes.

<sup>c</sup> Two affinity constants were observed due to two sequential binding events.



### 4.3 Probing Lsm-RNA interactions by isothermal titration calorimetry (ITC)

#### 4.3.1 Isothermal titration calorimetry theory

ITC allows measurement of affinities between macromolecules in solution in contrast to surface-based SPR measurements. The advantage of ITC is the absence of any surface effects on the interaction partners as well as free accessibility of both molecules. In ITC, the heat changes associated with binding reactions are detected. Usually, this is done by titrating one of the reaction partners into a concentrated solution of the other molecule (Falconer and Collins 2011). Each titrant injection  $i$  releases the heat  $Q_i$  (Recht et al. 2008; Feig 2009):

$$Q_i = nF[A]\Delta HV \quad (4-16)$$

with the stoichiometry  $n$ , the fraction  $F$  of sites on  $A$  bound by  $B$ , the reaction enthalpy  $\Delta H$  and the cell volume  $V$ .

For a simple 1:1 binding interaction following relation (4-1) with a  $K_A$  according to (4-2) this relation can be expressed as:

$$Q_i = \frac{n[A]\Delta HV}{2} \left( 1 + \frac{[B]}{n[A]} + \frac{1}{nK_A[A]} - \sqrt{\left( 1 + \frac{[B]}{n[A]} + \frac{1}{nK_A[A]} \right)^2 - \frac{4[B]}{n[A]}} \right) \quad (4-17)$$

Fitting of this equation to the experimental data by non-linear regression allows the determination of  $n$ ,  $K_A$  and  $\Delta H$  (Wiseman et al. 1989). From these parameters, a full thermodynamic profile can be determined in addition to equilibrium constants by calculating the free energy change ( $\Delta G$ ) from :

$$\Delta G = -RT \ln K_a \quad (4-18)$$

$$\Delta G = \Delta H - T\Delta S \quad (4-19)$$

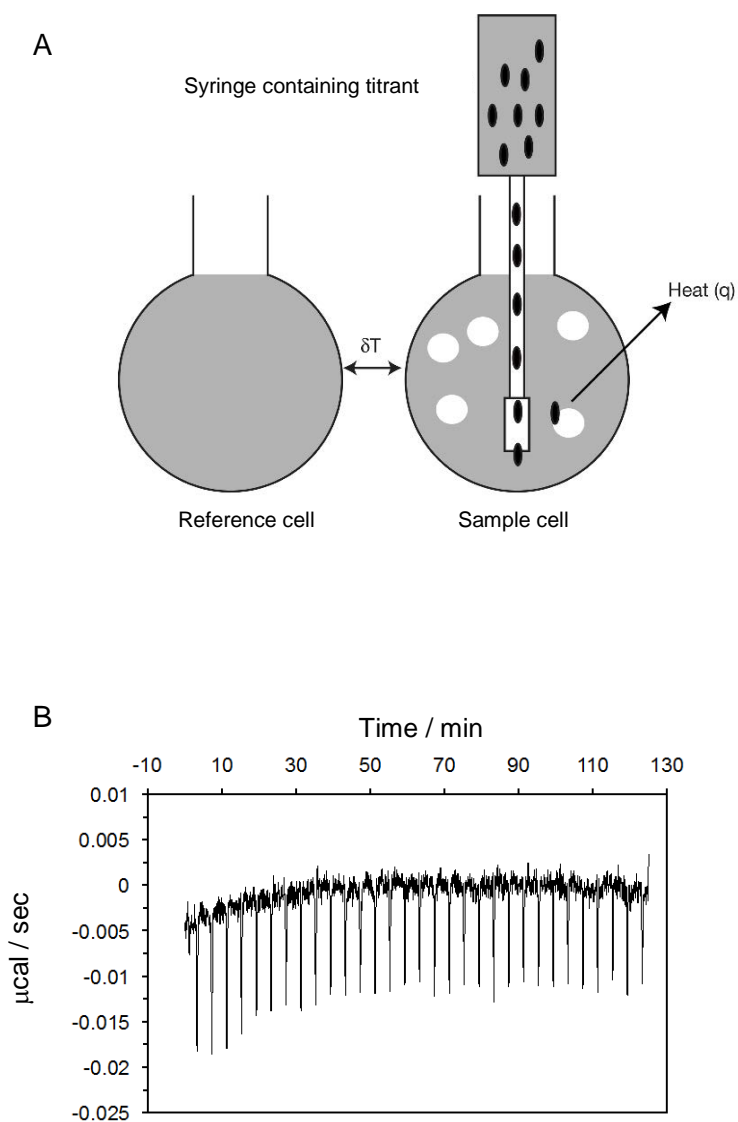
where  $R$  is the universal gas constant (1.9858775 cal/mol K),  $T$  the temperature in K and  $\Delta S$  the change in entropy.

#### 4.3.2 ITC experiments with Lsm polyproteins and RNA

The experimental setup involved an adiabatic chamber hosting a sample and a reference cell with an approximate volume of 1.4 ml inside the calorimeter (Figure 4.14). Defined amounts of ligand solution are introduced into the sample cell using a motorised syringe stirrer. Which of the interacting molecules is used as the titrant or the titrate generally does not influence the experiment. However, the titrant concentration is typically chosen 10 - 30 times higher than the tirate concentration in the sample chamber (Feig 2009). As a consequence, for protein-RNA studies it can be advantageous to place RNA in the injection syringe to avoid protein aggregation at high sample concentrations. Sample concentration is an important factor to consider prior to an ITC experiment. A factor  $c$ :

$$c = n K_A [A] \quad (4-20)$$

was described by Wiseman et al. (1989). This factor should lie between 10 and 100 for the determination of reliable  $K_A$  values and allows the estimation of a starting concentration for the analyte. Steady-state analysis of the interactions of Lsm[4+1]<sub>4</sub> with U<sub>10</sub> by SPR yields a  $K_D$  of 730 nM. From these data, a concentration range between  $[A]_1 = 7.3$  and  $[A]_2 = 73$   $\mu$ M is calculated from equation (4-21) using  $c$  values of  $c_1 = 10$  and  $c_2 = 100$ . In previous electro mobility shift assays, an affinity constant of  $K_D = 450$  nM was obtained for tetrameric Lsm[2+3] and U<sub>10</sub> (Sobti et al. 2010). Similarly, boundary conditions for the analyte concentration  $[A]_1$  and  $[A]_2$  are determined as 4.5 and 45  $\mu$ M. Based on these estimates Lsm polyproteins were adjusted to 10  $\mu$ M concentration, well within the determined concentration range for ITC analysis.

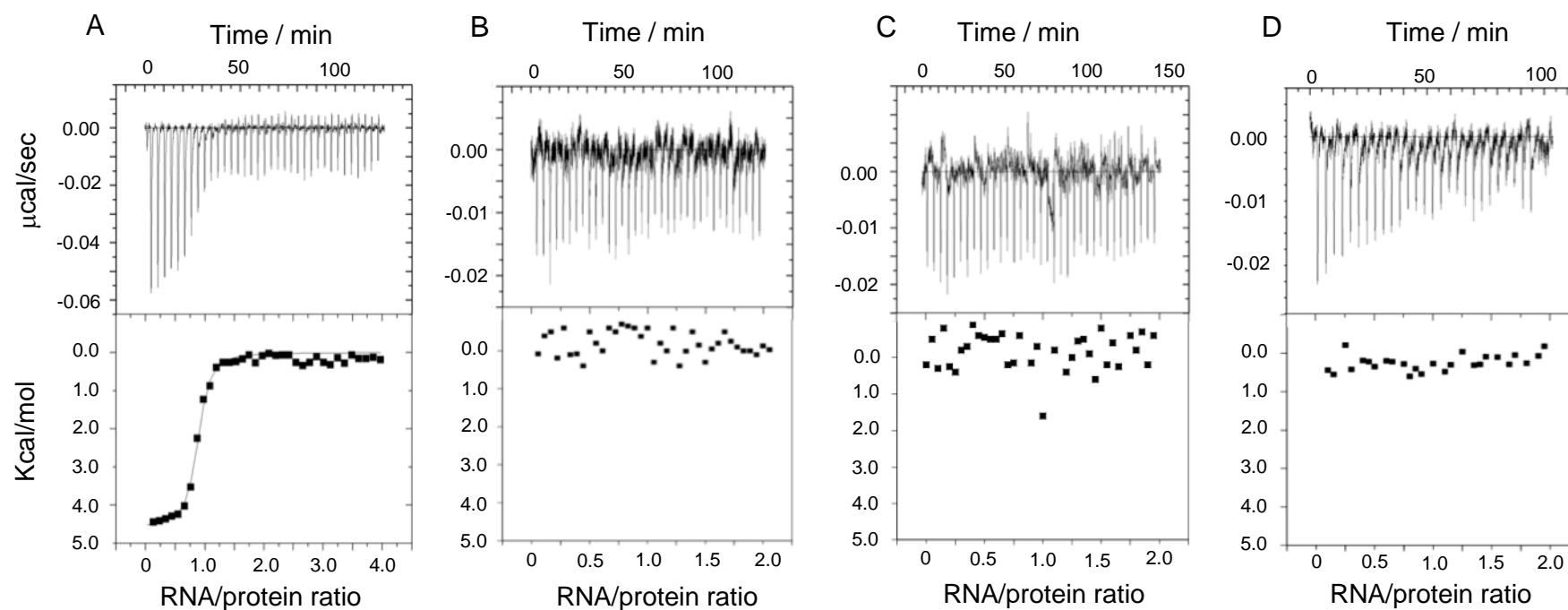


**Figure 4.14** | Schematic view of the ITC setup. A) Figure adapted from Recht et al. (2008). Temperatures of a reference cell and a sample cell located in an adiabatic chamber are continuously compared. Injection of titrant molecules into the sample cell containing a titrate solution results in heat changes  $q$  relative to the reference cell. B) Heat changes recorded for the injections (10  $\mu\text{l}$ ) of RNA (200  $\mu\text{M}$ ) into buffer solution not containing protein at 25  $^{\circ}\text{C}$ .

Lsm[2+3] and Lsm[4+1] polyprotein complexes of tetra and octameric composition and octameric Lsm3 were isolated by SEC and flash frozen prior to ITC analysis. The Lsm complexes were placed in the reaction chamber of the ITC instrument, while RNA at a concentration of 100-200  $\mu$ M was filled in the injector. The signal from negative controls of RNA titrated in sample buffer without protein (Figure 4.14) is subtracted from all following binding curves.

Sequential injections of RNA into Lsm polyprotein solutions produce the heat changes that are displayed in Figure 4.15. The heat changes observed over the course of the RNA titration form a sigmoidal curve for Lsm[2+3]<sub>4</sub> proposing complex formation between RNA and protein. For Lsm[4+1]<sub>4</sub>, and Lsm[3]<sub>4</sub> and Lsm[2+3]<sub>8</sub> only a scatter of heat changes corresponding to  $0 \pm 0.7$  kcal/mol ( $\pm 1.5$  for Lsm[4+1]) is observed. These likely arise from dilution effects only and not due to macromolecular interactions.

Considering the kinetic data obtained from SPR, it is somewhat surprising that no significant heat changes are observed for the titration of U<sub>10</sub> into tetrameric Lsm[4+1]<sub>4</sub> and the other complexes tested. A possible explanation is the nature of the experiments conducted. During SPR analysis RNA was immobilised and hence may behave differently to free RNA in solution in ITC. Furthermore RNA oligomers utilised for SPR experiments incorporated a biotin group for immobilisation followed by a G<sub>5</sub> stretch to minimise steric hindrance. Complexes Lsm[4+1]<sub>4</sub> and Lsm[3]<sub>8</sub> may inherit binding determinants for G<sub>5</sub> resulting in the observed differences between SPR and ITC



**Figure 4.15** Thermal parameters detected by ITC for Lsm with U<sub>10</sub> (25 °C). Displayed are heat changes for addition of 200 μM RNA into a reaction chamber (V = 1.4 ml) containing 10 μM solutions of A) Lsm[2+3]<sub>4</sub>, B) Lsm[2+3]<sub>8</sub>, C) Lsm[4+1]<sub>4</sub>, and D) Lsm[3]<sub>8</sub>.

data. This is consistent with a role of individual Lsm proteins in RNA recognition as suggested by the crystal structures of the human U1 snRNP (Pomeranz Krummel et al 2009, Weber et al 2010). The absence of heat changes observed for the Lsm[2+3]<sub>8</sub> fraction can be explained by an obstruction of the major site for RNA recognition (site I, Figure 1.5) due to the stacking interaction of two protein rings.

Consistent with the SPR data, Lsm[2+3]<sub>4</sub> is found to have the highest affinity for U<sub>10</sub> in ITC studies. Further, these findings are in good agreement with RNA-binding data on these Lsm complexes from gel-shift experiments (Sobti et al. 2010).

In order to quantify the affinity of Lsm[2+3]<sub>4</sub> towards U<sub>10</sub>, a curve is fitted to the data using a 1:1 binding model as described in section 4.2.1. This yields an equilibrium association constant of  $K_A = 29 \cdot 10^6 \pm 6.2 \cdot 10^6 \text{ M}^{-1}$  and an equilibrium dissociation constant  $K_D$  of  $34 \pm 15 \text{ nM}$ . The stoichiometry of the reaction is determined as  $0.84 \pm 0.01$ , which is an acceptable value for a 1:1 binding interaction and thus demonstrates that the Lsm ring binds one U<sub>10</sub> molecule. The reaction enthalpy for complex formation is estimated as  $\Delta H^0 = -4591 \pm 91.6 \text{ cal/mol K}$ . At a temperature of 25 °C, the entropic contribution can be estimated from equations (4-18) and (4-19). From these relations, first  $\Delta G$  is determined as  $\Delta G^0 -10187.3 \text{ cal/mol}$  and subsequently the entropy change  $\Delta S^0$  is calculated as:

$$\frac{\Delta H^0 - \Delta G^0}{T} = \frac{-4591 + 10187.3 \cdot \text{cal}}{298.15 \cdot \text{mol} \cdot \text{K}} = 18.8 \frac{\text{cal}}{\text{mol} \cdot \text{K}}$$

For the entropic contribution to the reaction follows  $-T\Delta S = -298 \text{ K } 18.8 \text{ cal/mol K} = 5602.4 \text{ cal/mol}$ . This suggests a binding interaction that is driven by approximately equal enthalpic and entropic contributions. Binding experiments at a range of different temperatures would be required to elucidate the exact thermodynamic parameters of this reaction. Table 4.3 summarises kinetic and thermodynamic parameters determined for the interaction of Lsm[2+3]<sub>4</sub> and U<sub>10</sub>.

**Table 4.3|** ITC-derived thermodynamic parameters for the interaction of tetrameric Lsm[2+3]<sub>4</sub> and U<sub>10</sub>.

Parameter	Value
K <sub>A</sub>	$30 \times 10^6 \pm 6.2 \times 10^6 \text{ M}^{-1}$
K <sub>D</sub>	$34 \pm 15 \text{ nM}$
N	$0.84 \pm 0.01$
$\Delta H^0$	$-4.6 \pm 0.091 \text{ kcal/mol}$
$\Delta S^0$	$0.02 \text{ kcal/mol K}$
$\Delta G^0$	$-10.2 \text{ kcal/mol}$

## 4.4 Conclusion

Tetrameric Lsm polyproteins form cylindrical or ring shaped quaternary structures in solution. Molecular envelopes obtained for those assemblies agree well with experimental data and dimensions (75 x 50 Å) are consistent with the crystal structure of Lsm[3]<sub>8</sub> (Naidoo et al. 2008).

Using two different methodologies (SPR and ITC), the RNA binding behaviour of a variety of Lsm complexes was characterised and it could be shown that Lsm protein



complexes possess different affinities towards U-rich RNA. These studies provide the first realtime and thermodynamic data for the interaction of Lsm complexes with RNA.

Analysed complexes Lsm[2+3]<sub>4</sub>, Lsm[4+1]<sub>4</sub> and complex Lsm[3]<sub>8</sub> form ring structures containing eight protomers in solution (Section 4.1.1 and (Naidoo et al. 2008)). As outlined in Chapter 1.1.6, Lsm toroids contain an RNA binding site (site I) in the lumen of the ring formed by highly conserved residues in loops L3 and L5. However, depending on their individual composition, Lsm polyprotein complexes display differential affinities for U-rich RNA oligonucleotides. These findings are in line with earlier studies using electro mobility shift assays (Sobti et al. 2010) and suggest the presence of distinct binding determinants of individual Lsm proteins for U-rich RNA outside site I.

We suggest target RNA discrimination by individual Lsm proteins requires residues beyond the Lsm folded domains, specifically in N-terminal sequence extensions and likely in the variable loop regions L4 and C-terminal sequence extensions. Such a mechanism of RNA recognition by individual Lsm proteins is consistent with the recently solved crystal structure of the U1 snRNP, where these sequence extensions in particular N-terminal extensions and loop L4 regions have been observed to contain RNA binding residues (Figure 4.15) implicated in guiding RNA in and out of the Sm-ring (Weber et al. 2010). N-terminal sequence extensions of SmD2 are observed to make contact to U1 snRNA via Arg and Lys residues embedded N-terminal sequence extension preceding the  $\alpha$ -helix (Figure 4.15). Lsm3 is the Lsm paralogue of SmD2 (Section 1.1.5), however, it lacks the N-terminal sequence

extension that is responsible for SmD2 specific RNA interactions. Yet, in the polyprotein Lsm[2+3], the C-terminal extension of Lsm2 is covalently linked to the Lsm3 N-terminus. This generates an artificial loop region (NMVDTNLLQDATRREVMTERKMET) upstream of the N-terminal  $\alpha$ -helix of Lsm3 containing potential RNA binding Arg and Lys residues (Figure 4.16). These additional residues may be responsible for the increased affinity of Lsm[2+3] towards U-rich RNA.

An RNA-binding function of the Lsm[2+3] linker is consistent with studies demonstrating the contribution of sequence extensions to the activity of Lsm complexes *in vivo*. These include: i) P-body assembly of Lsm1-7 (Reijns et al. 2008), ii) nuclear localisation of Lsm2-8 (Reijns et al. 2009), and iii) cytoplasmic localisation of Lsm1-7 (Reijns et al. 2009).

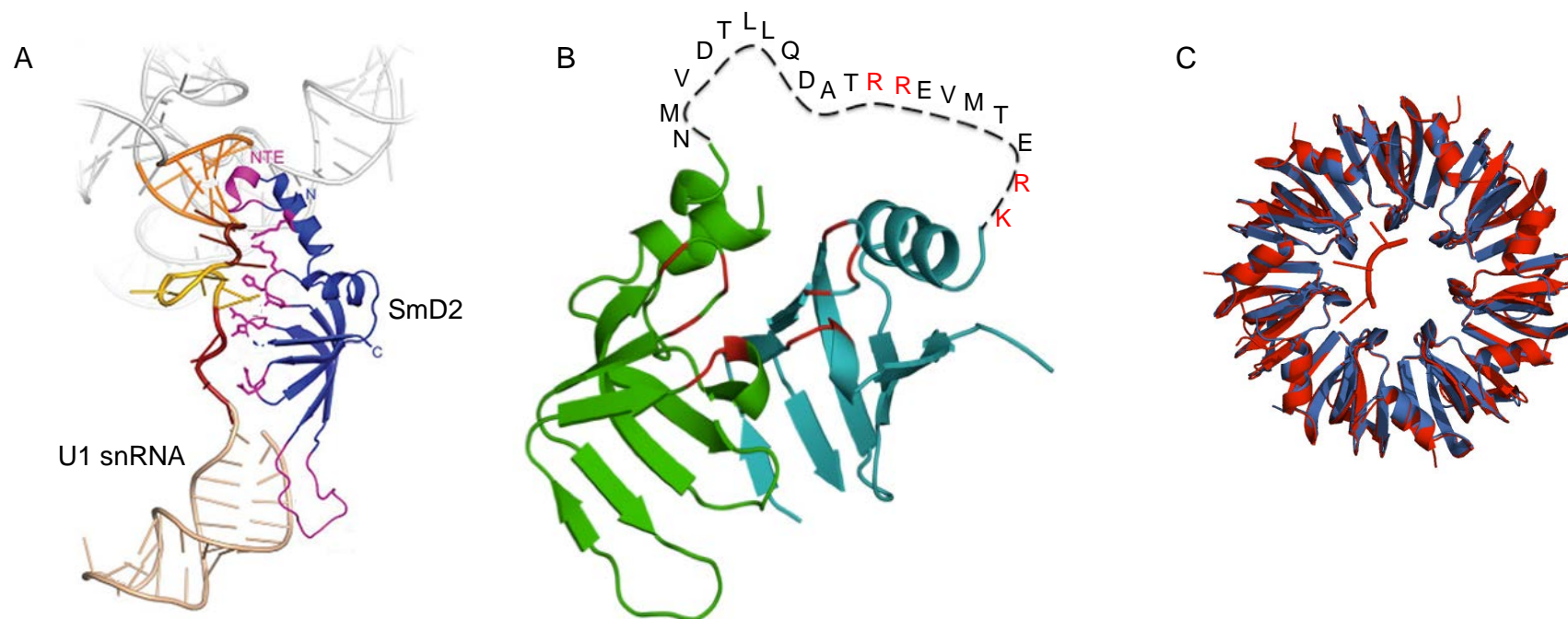
SAXS data further indicate that Lsm polyproteins form ring structures and RNA-binding does not result in a conformational change in solution. Given a 1:1 binding interaction as generally assumed for Lsm complexes and their RNA targets (Khusial et al. 2005; Moll et al. 2011) and indicated by ITC experiments, RNA only constitutes approximately 3 % of the scattering mass. Consequently, the data primarily reflects conformational changes of the protein upon RNA binding (Gupta et al. 2010).

Global structural parameters  $R_g$  and  $D_{max}$  as well as distance distribution functions and Kratky plots of free and RNA bound Lsm[2+3]<sub>4</sub> are highly similar proposing RNA-binding does not induce significant conformational changes. Comparable

changes in global structural parameters,  $p(r)$  functions and Kratky plots were found for the interaction of splicing factor SF1 and the branch point sequence of its target mRNA (Gupta et al. 2010). This prompted the authors to exclude any conformational changes of SF1 upon RNA binding. For the complex of Lsm[2+3]<sub>4</sub>, this may indicate most of the RNA is bound within the pore and consequently does not significantly influence the dimensions of the complex. It is well established that Lsm complexes are able to bind

one nucleotide per Lsm subunit in the lumen of the ring (Toro et al. 2001; Urlaub et al. 2001; Weber et al. 2010).

Crystal structures of archaeal Lsm proteins bound to short RNA oligonucleotides show only minor changes to the rigid Lsm scaffold on RNA binding (Toro et al. 2001; Thore et al. 2003). *AfSm1* and a complex of *AfSm1* with U<sub>5</sub> overlay almost perfectly with a corresponding RMSD of 0.74 Å, while *PaSm1* and a complex of *PaSm1* and U<sub>5</sub> can be superposed with an RMSD of 0.56 Å (Figure 4.16). Our SAXS data are consistent with the conservation of the rigid Lsm scaffold upon RNA binding and further suggests ring structures formed from Lsm polyproteins bind RNA in a native like manner. Hence, it is likely Lsm[2+3]<sub>4</sub> binds eight uracil residues in the lumen of the ring (site I) and the remaining two nucleotides are stabilised by Arg and Lys residues from the linker sequence connecting Lsm2 and Lsm3 domains.



**Figure 4.16** RNA binding by L/Sm proteins. A) RNA contacts made by SmD2 as found in the crystal structure of the human U1 snRNP (Weber et al. 2010). Figure adapted from Weber et al. (2010). In addition to conserved RNA binding residues located in loop regions L3 and L5, elements in N-terminal sequence extensions, and loop regions L2 and L4 are observed to form RNA contacts. Positively charged residues (Lys6, Lys8, Arg 19) in the N-terminal sequence extension implicated in RNA binding are highlighted in red. B) Model of the dimer interface of the polyprotein Lsm[2+3] based on the crystal structure of the SmD1D2 dimer (Kambach et al. 1999). The linker sequence between the Lsm2 C-terminus and the Lsm3 N-terminus is displayed as dashed line. Conserved RNA binding residues in the protein structure and potential RNA binding residues in the linker sequence of Lsm[2+3] are highlighted in red. C) Superposition of the crystal structures of A/Sm1 (blue) and A/Sm1 bound to U<sub>5</sub> (red) (Toro et al. 2001). The two structures can be aligned with an RMSD of 0.74.

## 5 Crystallographic studies of Lsm polyproteins

The three-dimensional structure of a protein ultimately defines its biological form and function. Thus, to fully understand the biochemical activity of a protein (or a protein complex), it is essential to have knowledge of its molecular structure. As outlined in Section 1.2, a number of crystal structures have been determined for Lsm protein complexes deriving from archaeal, bacterial and eukaryotic sources. However, at the time of writing, no high-resolution structure of a heteromeric Lsm complex exists. The protein structure group at Macquarie University successfully solved the crystal structure for yeast Lsm3 in 2008 (Naidoo et al. 2008). My work has shown Lsm polyprotein forms to be useful mimics of the heteromeric complexes (Chapters 3 and 4). Therefore, with the aim of providing the first atomic detail for the assembly of a heteromeric Lsm protein complex, the polyproteins Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> were subjected to crystallographic studies.

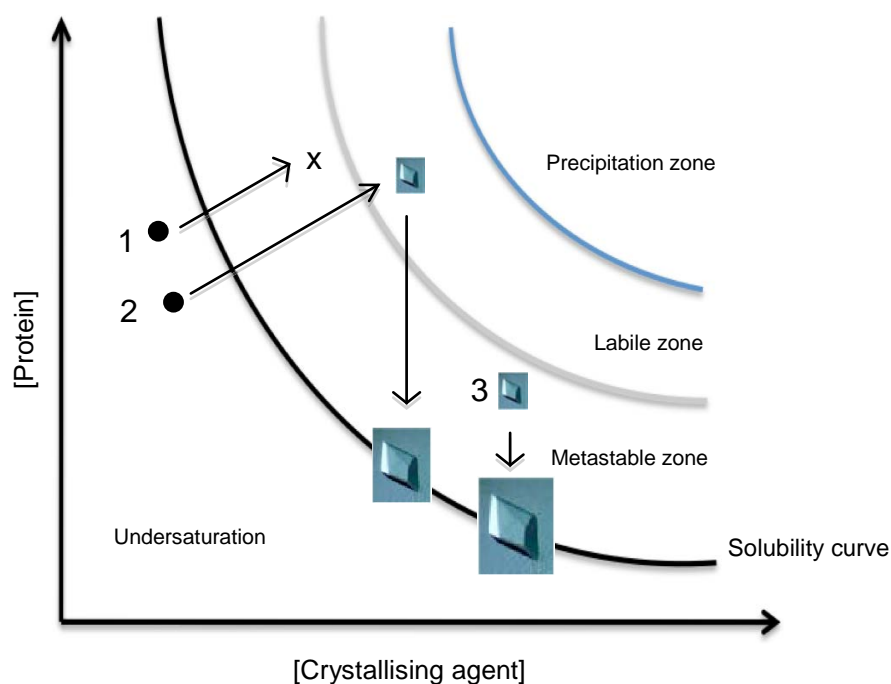
Crystallisation of a protein requires concentrating a protein above its solubility limit, to supersaturation. The most common method to achieve supersaturation is vapour diffusion. In this technique, a mixture of protein and a precipitant is allowed to equilibrate with a larger reservoir solution containing a higher concentration of the precipitant by diffusion of water from the protein solution to the reservoir. Concentrating a protein above its solubility limit in this way leads to supersaturation. Supersaturation can be divided into three zones (Figure 5.1) (Bergfors 2003):

- i) metastable zone: In the metastable zone, protein is supersaturated but not enough so for spontaneous nucleation to occur. This zone provides optimal growth conditions for crystals.
- i) labile zone: The labile zone allows spontaneous nucleation to take place and crystals can grow from formed nuclei.
- ii) precipitation zone: In the precipitation zone, the protein is so concentrated it will aggregate as amorphous precipitate.

Crystal formation at supersaturated conditions occurs in two phases: a nucleation phase and a growth phase. The spontaneous nucleation of protein crystals is thought to take place in the labile zone, while optimal conditions for the growth of large crystals are found in the metastable zone (Luft and DeTitta 1999; Bergfors 2003).

## 5.1 Crystallisation screening

Preparations of the polyproteins Lsm[4+1] and Lsm[2+3] result in two solution states that were separated by SEC (Chapter 3). These correspond to tetrameric and octameric assemblies consistent with single and possibly stacked Lsm rings observed in the crystal structure of yeast Lsm3 (Naidoo et al. 2008). The predominant forms of the polyproteins are Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> which can be isolated in large quantities. Following fractionation, tetrameric polyprotein fractions were concentrated to 20-30 mg/ml and flash frozen in liquid nitrogen.

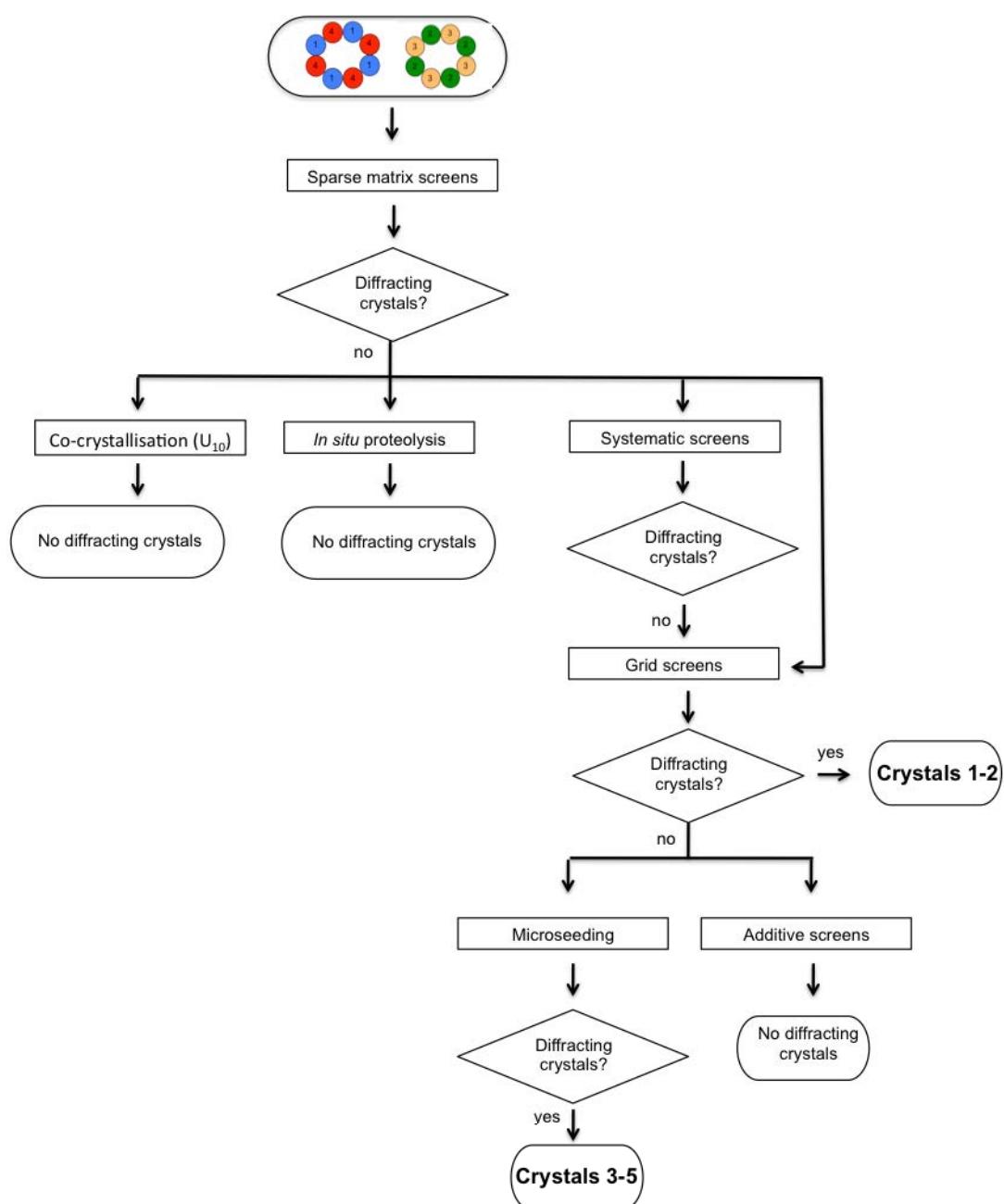


**Figure 5.1|** Two-dimensional theoretical protein phase diagram for a vapour-diffusion experiment. Adapted from Luft and DeTitta (1999). The phase diagram is divided in four zones: undersaturation, metastable zone, labile zone and precipitation zone. 1) Equilibration of a protein crystallant mixture against a crystallant reservoir reaches metastable zone and no crystals are formed. 2) Equilibration of a protein crystallant mixture against a crystallant reservoir reaches labile zone. Nuclei are formed and these grow to large crystals. 3) Microseeding. Nuclei are introduced into metastable zone and large crystals grow from these.

Isolated Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> are stable and monodisperse. SAXS data further indicate ring morphology for these oligomers. In addition, it has been found that high NaCl concentrations and the presence of glycerol stabilise tetrameric complexes of the polyproteins (Chapter 3). Most crystallisation screens utilise high salt or polyol concentrations to supersaturate the protein. Such conditions likely further stabilise Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> and make them promising targets for crystallisation studies.

Crystallisation screening was performed in vapour diffusion experiments. To reduce the amount of protein required for each individual experiment and to permit screening of a large number of conditions, a robotic liquid handling system was used to aliquot solutions for crystallisation trials. This system allowed the quick setup of sitting drop vapour diffusion experiments with as little as 200-300 nl of protein solution. This allowed screening of more than 1000 conditions with varying protein to precipitant ratios. All obtained crystals were screened for optimal cryogenic conditions and diffraction on an in-house diffractometer setup (University of New South Wales, Sydney). Diffraction experiments were performed at 100 K. For non-diffracting crystals a freeze/thaw process was incorporated to encourage crystal annealing. Figure 5.2 shows a flow diagram of our crystallisation strategy leading to the formation of five diffracting crystals of Lsm[4+1]<sub>4</sub>. These crystals will be referred to as crystals 1-5 in the following section.





**Figure 5.2|** Flow diagram of crystallisation process for Lsm polyproteins.

Initial screening was performed using five sparse-matrix suites providing 384 conditions (QIAGEN) to cover a maximal portion of chemical space. These screens are a compilation of precipitant combinations that previously enabled the successful crystallisation of a variety of proteins (Newman et al. 2005). For samples of Lsm[4+1]<sub>4</sub>, crystalline material was obtained across a variety of precipitant conditions. These varied from sandy precipitate, spherulites to microcrystals and small crystals. Specific crystallants identified from five sparse-matrix screens producing Lsm[4+1]<sub>4</sub> crystals include:

A) magnesium chloride (0.2 M), HEPES buffer (0.1 M, pH 7.5), PEG 400 (30 % v/v),

B) magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (10 % v/v),

C) magnesium formate (0.3 M), Bis-Tris buffer (0.1 M, pH 5.5), and

D) PEG 400 (2 % v/v), HEPES buffer (0.1 M, pH 7.5), ammonium sulfate (2 M).

Microcrystals and crystals obtained from conditions A-D are displayed in Figure 5.3.

As microcrystals and crystals from conditions A-D were too small and did not diffract beyond 20 Å resolution upon mounting on the University of New South Wales in-house system. The absence of intense reflections at high resolutions typical for the small lattices of salt crystals served to exclude the formation of salt crystals. Further optimisation experiments including additive screens with RNA, *in situ* proteolysis, systematic screens and grid screens were undertaken.

Similar to Lsm[4+1]<sub>4</sub>, the polyprotein Lsm[2+3]<sub>4</sub> was subjected to crystal screening. However, no crystalline material was obtained in sparse-matrix screens for Lsm[2+3]<sub>4</sub> in a concentration range of 8-30 mg/ml. Consequently, more extensive screening strategies including co-crystallisation with U<sub>10</sub>, *in situ* proteolysis and systematic screens of crystallants were carried out.

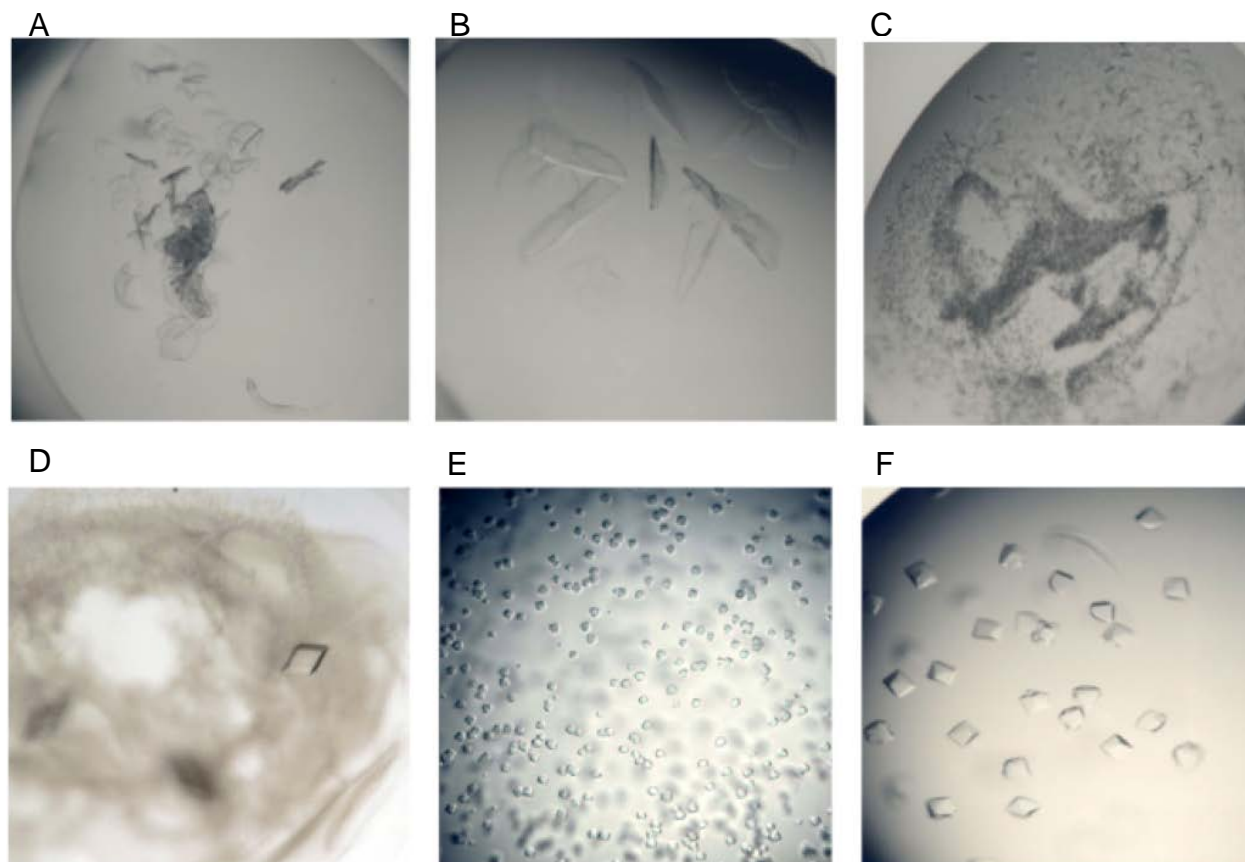
## 5.2 More extensive crystallisation strategies

### 5.2.1 Co-crystallisation with RNA

With the knowledge that Lsm polyproteins have high to moderate affinity for U<sub>10</sub> and can form stable complexes with the oligonucleotide (Chapter 4), all sparse-matrix screens were repeated in the presence of U<sub>10</sub>. In all cases, a variety of protein RNA ratios (1:2, 1:1, 2:1) was trialled. However, no diffraction quality crystals were obtained.

### 5.2.2 *In situ* proteolysis

*In situ* proteolysis has gained increasing attention as a rescue strategy for proteins that are recalcitrant to crystallisation (Dong et al. 2007; Wernimont and Edwards 2009). The rationale behind this approach is the generation of trimmed and stable protein domains through proteolytic cleavage of N- and C-terminal sequence extensions (and occasionally, loop regions) of the target proteins. For a large number (339) of proteins of both eukaryotic and prokaryotic origin that failed to produce diffraction quality crystals, the Structural Genomics Consortium demonstrated that



**Figure 5.3|** Lsm[4+1] Microcrystals and crystals from a variety of sparse-matrix screens. A) magnesium chloride (0.2 M), HEPES buffer (0.1 M, pH 7.5), PEG 400 (30 % v/v), B) magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (10 % v/v) C) Magnesium formate (0.3 M), Bis-Tris buffer (0.1 M, pH 5.5), D) PEG 400 (2 % v/v), HEPES (0.1 M, pH 7.5), ammonium sulfate (2 M). Crystals illustrated obtained from additional PEG and ammonium sulfate systematic optimisation E) Magnesium acetate (0.2 M), and PEG 3350 (20 % v/v) F) Trisodium citrate (0.1 M), ammonium sulfate (0.5 M), lithium sulfate (1 M). No diffraction pattern was obtained with characteristics of inorganic salts for any of those crystals.

the addition of trace amounts of protease leads to the formation of crystals that allowed data collection for 16 % of the targets (Dong et al. 2007; Wernimont and Edwards 2009).

In the two polyproteins utilised in this work, the N- and C-terminal sequence extensions beyond the folded domains of the component Lsm domains have already been removed. However, the flexible linker regions connecting the fused Lsm domains may have sufficient flexibility (Chapter 3) to interfere with the crystallisation process.

For the Lsm[4+1] construct, the linker sequence (DNIIDEADLYLDQYNFT) contains Tyr and Phe residues at positions 10 and 16, respectively. These residues are both favourable to chymotrypsin cleavage. In order to determine an appropriate amount of chymotrypsin for cleavage, Lsm [4+1]<sub>4</sub> preparations were subjected to varying protease concentrations for 24 h (20 °C). Examination of protein sizes observed in SDS-PAGE (Figure 5.4) identified that samples (20 mg/ml) treated with 20-200 ng/ml chymotrypsin contained two major protein species corresponding to masses of 10-15 kDa. These are consistent with individual Lsm4 (12 kDa) and Lsm1 (11 kDa) proteins being released through cleavage. Subsequently, *in situ* proteolysis was performed at a 1:10000 protein:chymotrypsin ratio. Crystals were successfully grown in the presence of the protease in base condition D PEG 400 (2 % v/v), HEPES buffer (0.1 M, pH 7.5), ammonium sulfate (2 M) (Figure 5.4). However, no diffraction was observed from these crystals, despite extensive optimisation attempts, including crystal annealing and screening of multiple cryogenic agents (polyols, oils and salts).

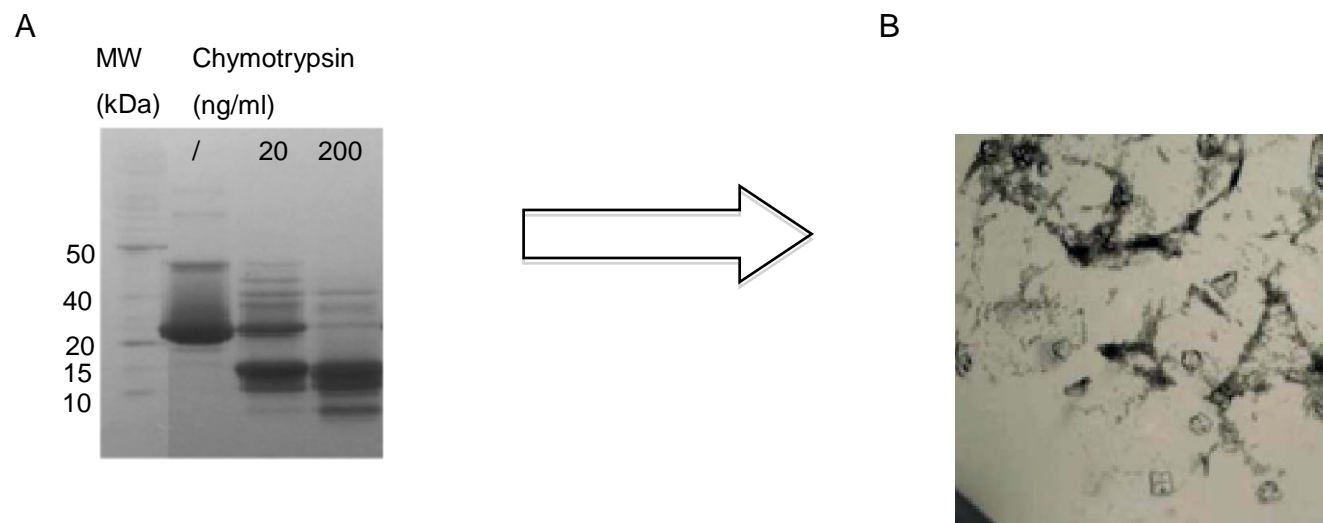
The linker sequence of Lsm[2+3] contains Arg residues at positions 13, 14 and 20 and Lys residues at position 21 likely allowing trypsin cleavage of the linker sequence. Hence, *in situ* proteolysis of Lsm[2+3]<sub>4</sub> with trypsin was carried out at protein:protease ratios of 1:100 - 1:10000. Like sparse-matrix screens, these trials did not yield crystalline material.

### 5.2.3 Systematic screens

The majority of crystal yielding conditions identified for Lsm[4+1]<sub>4</sub> in sparse-matrix screens contained magnesium ions and PEG. Thus, systematic screens with suites containing combinations of cations and polyethyleneglycols were carried out. Small crystals were observed in a solution containing magnesium acetate (0.2 M) and PEG 3350 (20 % v/v) (condition E in Figure 5.3). Due to the small size of these crystals, they required further optimisation in grid screens.

Additionally, an ammonium sulfate based crystallant solution has been found to produce Lsm[4+1]<sub>4</sub> crystals in sparse matrix screen (Figure 5.3). Hence, ammonium sulfate based precipitants were systematically screened. This approach resulted in small non-diffracting crystals from trisodium citrate (0.1 M), ammonium sulfate (0.5 M) and lithium sulfate (1 M) (condition F in Figure 5.3), which were subjected to further optimisation in grid screens.

Similarly, Lsm[2+3]<sub>4</sub> was subjected to systematic screening of crystallants. However, despite probing over 1000 conditions including PEG-, salt- and pH-based screens, no crystals could be grown for this construct.



**Figure 5.4** | *In situ* proteolysis trials]. A) SDS-PAGE of Lsm[4+1]<sub>4</sub> samples that were exposed (24 h) to different protease concentrations. B) Lsm[4+1]<sub>4</sub> microcrystals formed in the presence of chymotrypsin in PEG 400 (2 % v/v), HEPES buffer (0.1 M, pH 7.5), ammonium sulfate (2 M), chymotrypsin (2 µg/ml).

A large difference between constructs Lsm[4+1] and Lsm[2+3] is within the linker sequences connecting the folded Lsm units, corresponding to DNIIDEADLYLDQYNFT and NMVDTNLLQDATRREVMTERKMET residues for Lsm[4+1] and Lsm[2+3], respectively. The increased linker length in the Lsm[2+3] construct may confer more flexibility and thus interfere with the crystallisation process.

#### **5.2.4 Grid screens**

Once promising crystallisation conditions are identified, these are generally refined in grid screens, typically as evenly stepped gradients of the key precipitant or the pH (Hennessy et al. 2009). Grid screens were set up manually with drop sizes of 1-4  $\mu$ l over 0.5 ml reservoirs. Crystallant gradients were generated using the four corner method (Hennessy et al. 2009) from serial dilution of four stock solutions containing minimum and maximum concentrations of the components being varied in the screen. In this way, base conditions A-F identified in sparse-matrix and systematic screens were optimised and diffracting crystals (1-5) were obtained for some crystallant combinations.

Optimisation of many variables including chemical (e.g. precipitant concentration, pH, and cryoprotection) and kinetic (e.g. reservoir composition, temperature and drop ratios) parameters based on conditions A, C, E and F did not yield diffraction from these crystal forms. However, iteration of D and B resulted in crystals that diffracted to resolutions of 5 Å and 4.3 Å, respectively.



Base condition D was optimised in grid screens by fine tuning of co-solvent (1-8 % PEG 400 v/v), pH (7-8) and precipitant (1.5-3 M ammonium sulfate) and eventually reproducibly formed hexagonal crystals (Figure 5.5). These crystals showed plate morphology (0.2 x 0.2 mm) and displayed diffraction to a maximal resolution of 5 Å on the in-house setup in PEG 400 (5 % v/v), HEPES buffer (0.1 M, pH 7.4) and ammonium sulfate (2 M). Given this relatively low resolution of diffraction, many strategies were employed in an attempt to improve crystal quality from these samples. Further optimisation of cryogenic conditions, the protein concentration (10-30 mg/ml) or the precipitant-to-protein drop ratio (1:2, 1:1, 2:1) failed to improve these hexagonal plates.

Base condition B was optimised by stepwise variation of PEG 8000 (7-12 % v/v), magnesium chloride (0-0.5 M) and the pH (6.8-7.4) and resulted in half-moon-shaped crystal plates (Figure 5.6). These were grown in magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0) and PEG 8000 (8 % v/v). By screening numerous cryoprotective agents for these crystals, optimal cryogenic protection and diffraction was achieved through sequential transfer of the crystals into reservoir solution containing increasing PEG 400 concentrations (5 % increments) (Figure 5.7). A final concentration of 20 % (v/v) PEG 400 was sufficient to prevent ice ring formation and to allow an improved diffraction to 4.3 Å. I tested over 40 crystals obtained from this condition and flashfroze the 20 best diffracting ones for data collection at the Australian Synchrotron. Two of these crystals were suitable for data collection and will be referred to as crystals 1 and 2 in Section 5.3.

**A**

↓ Ammonium sulfate (1.5-2.5 M)  
HEPES buffer (0.1M, pH 7.2-7.8)

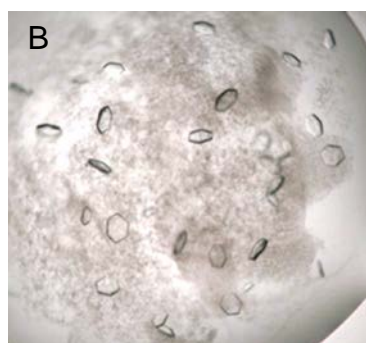
		Ammonium sulfate concentration / M						
		1.5	1.7	1.9	2.1	2.3	2.5	
HEPES (0.1 M)	pH 7.2							PEG 400 (2%)
	pH 7.4		X	X	X			
	pH 7.6		X	X				
	pH 7.8							

↓ PEG 400 (1-6 % v/v)  
HEPES buffer (0.1M, pH (7.4-7.7))

		PEG 400 (%)						
		1	2	3	4	5	6	
HEPES (0.1 M)	pH 7.4		X	X	X	X	X	Ammonium sulfate (2 M)
	pH 7.5		X	X	X	X		
	pH 7.6			X	X			
	pH 7.7			X				

↓ Ammonium sulfate (1.6-2.1 M)  
PEG 400 (2-5 % v/v)

		Ammonium sulfate (%)						
		1.6	1.7	1.8	1.9	2.0	2.1	
PEG 400 (%)	2	X	X	X	X	X	X	HEPES pH 7.4 (0.1 M)
	3	X	X	X	X	X	X	
	4	X	X	X	X	X	X	
	5	X	X	X	X	X	X	



**Figure 5.5|** Optimisation of crystallant conditions using grid screens A) Optimisation based on the condition ammonium sulfate (2 M), HEPES buffer (0.1 M, pH 7.5) and PEG 400 (2 % v/v). B) Hexagonal crystals obtained in refined condition PEG 400 (5 % v/v), HEPES buffer (0.1 M, pH 7.4), PEG 8000 (8 % v/v). Conditions leading to crystals are indicated by X, best diffracting crystals are indicated by **X**.

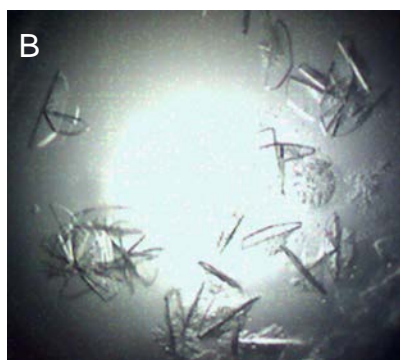
A

Tris buffer (0.1M, pH 6.8-7.4),  
PEG 8000 (7-12 % v/v)

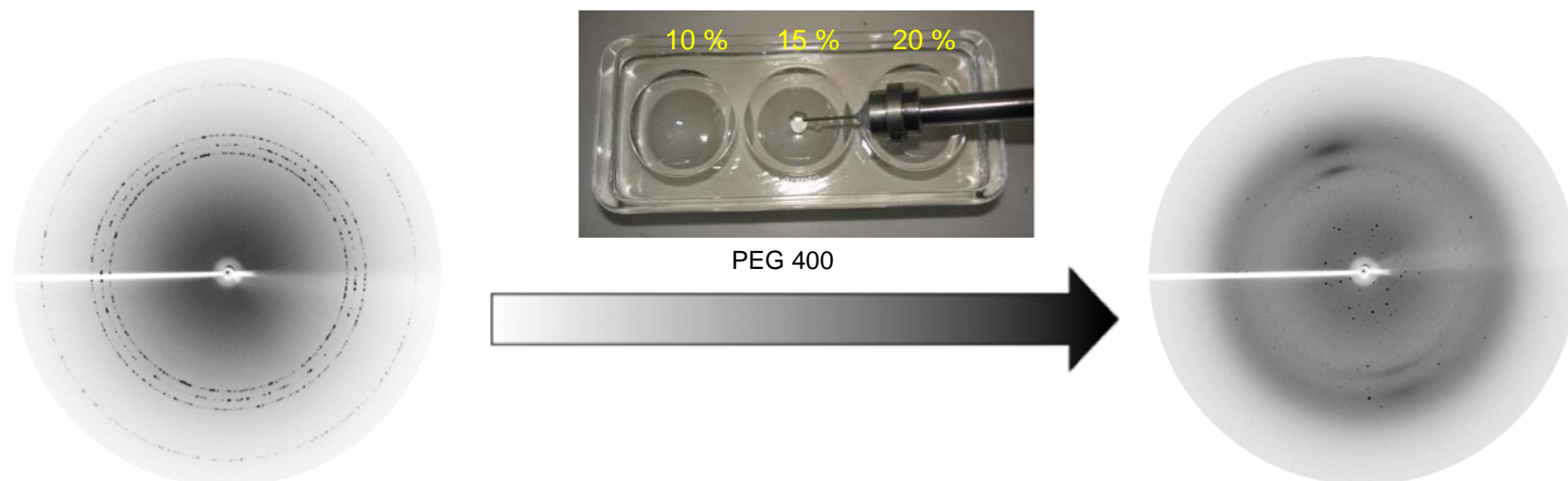
		PEG 8000 (%; v/v)						MgCl <sub>2</sub> (0.2 M)
		7	8	9	10	11	13	
Tris (0.1 M)	pH 6.8		X	X	X	X		
	pH 7.0		X	X	X			
	pH 7.2		X	X	X			
	pH 7.4		X	X	X			

MgCl<sub>2</sub> (0.0-0.5 M),  
PEG 8000 (8-11% v/v)

		MgCl <sub>2</sub> (M)						Tris pH 7.0 (0.1 M)
		0	0.1	0.2	0.3	0.4	0.5	
PEG 8000 %	8		X	<b>X</b>	X			
	9		X	X	X			
	10		X	X				
	11		X	X				



**Figure 5.6|** Optimisation of crystallant conditions for the base condition magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (10 % v/v) (A). Half-moon-shaped crystals grown in magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (8 % v/v). Conditions leading to crystals are indicated by X, best diffracting crystals are indicated by **X**.



**Figure 5.7|** Optimisation of cryogenic conditions for Lsm[4+1]<sub>4</sub> crystals grown in magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0) and PEG 8000 (10 % v/v). Sequential transfer (5 % increments) of crystals into reservoir solution containing increasing PEG 400 concentrations provided cryoprotection.

Despite screening a large number of Lsm[4+1]<sub>4</sub> samples, no crystals were obtained displaying diffraction beyond 4 Å. As a consequence, I attempted to improve crystal quality by screening additives (Cudney et al. 1994) and microseeding (Bergfors 2003).

### 5.2.5 Searches of additives to assist crystal quality

Grid screens surrounding base conditions A-F were set up in the presence of a variety of additives to improve my crystals of Lsm[4+1]<sub>4</sub>. The types of additives used and their concentrations are summarised in Table 5.1. All of these screens were performed with and without U<sub>10</sub> (0.5:1, 1:1, 1:0.5 ratio). Additives were mixed with reservoir solution to the final concentrations given in Table 5.1. The resulting reservoir solution was subsequently mixed 1:1 with 1-2 µl of Lsm[4+1]<sub>4</sub> in Tris buffer (20 mM, pH 8.0), NaCl (400 mM) and glycerol (10 % v/v).

Of the six crystal-yielding conditions identified for the polyprotein in previous screens, four contained magnesium salts. Thus, multivalent cations were incorporated into additive screens. Magnesium chloride addition did not improve the quality of Lsm [4+1]<sub>4</sub> crystals.

**Table 5.1|** Additives screened for Lsm[4+1]<sub>4</sub>

Additive	Classification	Drop concentration
barium chloride	multivalent cation	5-10 mM
cadmium chloride	multivalent cation	5-10 mM
calcium chloride	multivalent cation	5-10 mM
cobaltous (III) chloride	multivalent cation	5-10 mM
magnesium chloride	multivalent cation	5-10 mM
manganese (III) chloride	multivalent cation	5-10 mM
strontium chloride	multivalent cation	5-10 mM
yttrium (III) chloride	multivalent cation	5-10 mM
zinc chloride	multivalent cation	5-10 mM
iron (III) chloride	multivalent cation	5-10 mM
chromium chloride	multivalent cation	5-10 mM
praesodium chloride	multivalent cation	5-10 mM
6-aminohexanoic acid	linker	1.5-3% (v/v)
1, 5 diaminopentane dihydrochloride	linker	1.5-3% (v/v)
1, 6 diaminohexane	linker	1.5-3% (v/v)
1, 8 diaminooctane	linker	1.5-3% (v/v)
glycine	linker	50-100 mM
glycyl-glycyl-glycine	linker	15-30 mM
taurine	linker	5-10 mM
betaine hydrochloride	linker	5-10 mM
sodium malonate	salt	50-100 mM
TCEP hydrochloride	reducing agent	5-20 mM
DTT	reducing agent	5-20 mM
EDTA	chelating agent	5-10 mM
ethylene glycole	polyol	1.5-3% (v/v)
glycerol	polyol	1.5-3% (v/v)

However, the addition of cobaltous chloride (2.5-20 mM) to lithium sulfate based condition F improved the size of the crystals (0.1 x 0.1 mm) previously obtained in this base condition (Figure 5.8). Diffraction to 5 Å resolution was observed in-house for these crystals following soaking in high concentrations of lithium sulfate (2 M) for cryoprotection. The eleven best diffracting crystals from this condition were flash frozen and stored in liquid nitrogen for synchrotron diffraction screening.

The Lsm polyprotein preparations themselves may contain trace concentrations of divalent cations either from growth media or as a result from nickel leaching during IMAC purification. Hence, effects of metal ion removal were examined in additive screens including EDTA (5-10 mM). For these experiments, EDTA was added to preparations of Lsm[4+1]<sub>4</sub> prior to grid screens with conditions A-F. The affinity of EDTA towards Ni<sup>2+</sup> is significantly higher than for Mg<sup>2+</sup> (Smith et al. 1985) contained in crystallant solutions. Thus, EDTA would preferentially chelate remaining Ni<sup>2+</sup> over Mg<sup>2+</sup>. Furthermore, at the utilised concentrations (5-10 mM), EDTA would not significantly reduce the Mg<sup>2+</sup> concentration (0.2 M) in the crystallisation solutions. However, no crystals were grown in the absence of divalent cations.

Certain small molecules have the capacity to crosslink protein molecules via either electrostatic or hydrophobic interactions and have been shown to maintain or create a crystal lattice (McPherson and Cudney 2006). The power of these small molecule linkers has been demonstrated by the crystallisation of 85 % of a panel of proteins and viruses from only two crystallisation conditions (PEG 3350 and Tacsimate<sup>TM</sup>) by incorporating different linkers (McPherson and Cudney 2006). Similarly, mixtures of the salts of neutralised organic acids (e.g. Tacsimate<sup>TM</sup>, Hampton Research) have been repeatedly noted as highly successful crystallisation reagents due to their high ionic strength and their capacity to act as electrostatic linkers between protein molecules (McPherson 2001; McPherson and Cudney 2006).

Preparations of Lsm[4+1]<sub>4</sub> were subjected to small molecule linkers in additive screens. The types of linkers and their concentrations are summarised in Table 5.1.

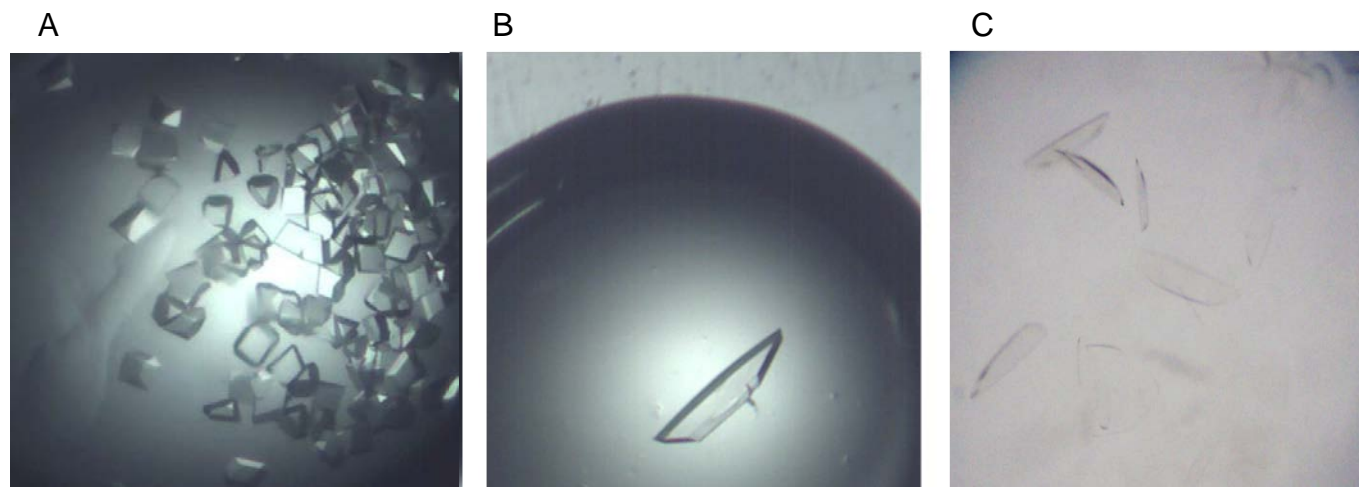
No changes in crystal morphology or diffraction were observed in these screens. However, Lsm[4+1]<sub>4</sub> co-yielded larger and thicker crystals (0.4 x 0.8 mm, Figure 5.8) with U<sub>10</sub> (1:1 ratio) in the presence of the small molecule linker taurine (10 mM) compared to crystals produced in the base condition lacking RNA and linkers (magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (8 % v/v)). These crystals, however, did only diffract to 5 Å resolution.

My solution studies of the Lsm polyprotein complexes demonstrated their selective stabilisation as a tetrameric state by high ionic strength and glycerol (10 % v/v). This indicated that polyols such as glycerol and ethylene glycol might be utilised as additives within crystallisation screens. However, no effects on the crystallisation of Lsm[4+1]<sub>4</sub> were observed in the presence of ethylene glycol and additional glycerol.

### 5.2.6 Microseeding

Microseeding uncouples the two crystal formation phases of nucleation (labile zone) and crystal growth (metastable zone) to promote larger and more ordered crystals. This is achieved by manually introducing nuclei into a metastable protein solution (Figure 5.1). For Lsm[4+1]<sub>4</sub>, the seeding procedure was performed according to the method of Bergfors (2003). Crystals obtained from magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.4) and PEG 8000 (8 % v/v) were transferred to a stabilising solution containing reservoir solution with a 10-20 % (v/v) increased precipitant (PEG 8000) concentration and 5-10 mg/ml protein. The suspended crystals were crushed by vortexing and transferred to clear crystallisation drops, which had





**Figure 5.8|** Crystals for Lsm[4+1]<sub>4</sub> obtained from additive screens, *in situ* proteolysis and microseeding. A) Additive screens with added cobaltous chloride (0.1 M) in trisodium citrate (0.1 M), ammonium sulfate (0.5 M), lithium sulfate (1 M), cobaltous chloride (0.1 M). B) Additive screen with added U<sub>10</sub> to magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (8 % v/v), U<sub>10</sub> RNA (equimolar). C) Crystals 3-5 obtained from microseeding in Tacsimate™ (44 % v/v), Tris buffer (0.1 M, pH 7.0).

equilibrated over 2-7 days. Gradual refinement of the crystallant conditions (44 % Tacsimate<sup>TM</sup> (v/v), 0.1 M Tris pH 7-8) and the cryogenic conditions (20 % glycerol v/v) resulted in crystal formation

Crystals grown in Tacsimate<sup>TM</sup> (44 % v/v) and Tris buffer (0.1 M, pH 7.0) are shown in Figure 5.8. These diffracted to 3.5 Å in-house. As a consequence, a variety (21 crystals) of Lsm[4+1]<sub>4</sub> crystals were flash frozen and stored in liquid nitrogen for subsequent synchrotron data collection. This set of crystals included crystals 3-5 from which complete native datasets could be collected as discussed later in this chapter.

## 5.2.7 Heavy atom derivatives of Lsm[4+1]<sub>4</sub>

### 5.2.7.1 SeMet-Lsm[4+1]<sub>4</sub>

Since the Lsm[4+1] polyprotein sequence contains a total of eight Met residues, the substitution of these with SeMet should provide a strong anomalous signal for any subsequent diffraction experiment of a prepared crystal. Hence, a SeMet derivative of Lsm[4+1]<sub>4</sub> was prepared. Replacement of Met with SeMet was achieved through protein expression in a defined growth medium supplemented with SeMet and inhibitory amino acids to suppress Met synthesis in the expression strain.

Preparations of SeMet-Lsm[4+1]<sub>4</sub> showed an increased propensity to aggregate in crystallisation screens compared to its native counterpart. Despite screening with over 1000 conditions in sparse matrix format, no crystalline material was obtained. Precipitation of the protein during screening was found to be alleviated in high

concentrations of reducing agents DTT (10 mM) or TCEP (> 2 mM). Systematic grid screens either based on leads successful for native Lsm[4+1]<sub>4</sub>, or on conditions that generated promising forms of precipitate in sparse-matrix screens were not successful.

The use of *in situ* proteolysis with chymotrypsin (20 ng/ml) in Bicine buffer (0.1 M, pH 8.5), PEG 6000 (20 % v/v) successfully allowed microcrystals to form. These are shown in Figure 5.8 but were too small for diffraction to be attempted. Attempts to increase their size by refining crystallant conditions in grid screens were unsuccessful.

Microseeding was, to a certain extent, successful in generating crystal forms of Se-Met[4+1]<sub>4</sub>. The half-moon-shaped protein crystals obtained in Tacsimate<sup>TM</sup> (52 % v/v) at pH 6.5 are shown in Figure 5.9. Seed stocks were generated from native Lsm[4+1]<sub>4</sub> preparations, and nuclei were subsequently cross-seeded into crystallisation trays containing droplets of SeMet-Lsm[4+1]<sub>4</sub>.

Optimisation of cryogenic conditions for crystals established that 20 % (v/v) glycerol was most suitable. However, despite screening more than 20 crystals, diffraction could only be detected to 7-8 Å on the in-house diffractometer. To obtain the best possible diffraction from these crystals, the six best diffracting crystals were flash frozen and stored in liquid nitrogen for further synchrotron diffraction screening.



**Figure 5.9** Crystals obtained for SeMet and iodide derivatives of Lsm[4+1]<sub>4</sub>. A) Se-Met microcrystals obtained from *in situ* proteolysis in Bicine buffer (0.1 M, pH 8.5), PEG 6000 (20 % v/v), chymotrypsin (20 ng/ml), B) Crystals obtained from cross-seeding with native Lsm[4+1]<sub>4</sub> seeds in Tacsimate<sup>TM</sup> (52 % v/v), Tris buffer (0.1 M, pH 6.5). C) Crystals from I3C derivatised Lsm[4+1]<sub>4</sub> grown in Magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 6.5), PEG 8000 (12 % v/v).

#### **5.2.7.2 Iodide derivatives of Lsm[4+1]<sub>4</sub>**

As an alternative source of heavy atom derivatives, Lsm[4+1]<sub>4</sub> solutions were screened in the presence of the compound 5-amino-2,4,6-triiodoisophthalic acid (I3C, Hampton Research). The chemical structure essentially causes an iodide triangle to be formed within any resulting co-crystalline material, suitable for eventual single wavelength anomalous dispersion (SAD) or single isomorphous replacement plus anomalous scattering (SIRAS) phasing. For Lsm[4+1]<sub>4</sub>, co-crystallisation with I3C yielded small half-moon-shaped crystals (Figure 5.9) similar to those previously observed for native protein in a PEG 8000-based condition. Addition of PEG 400 at 17.5 % (v/v) to these iodide-derived crystals was sufficient to achieve cryoprotection as indicated by the absence of ice rings in diffraction screening. Using these optimised conditions, diffraction was observed from I3C-Lsm[4+1]<sub>4</sub> crystals, but diffraction quality corresponded to 5 Å resolution. Five of these crystals were stored in liquid nitrogen for synchrotron diffraction screening.

### **5.3 Crystallographic studies**

#### **5.3.1 Diffraction from Lsm[4+1]<sub>4</sub> crystals**

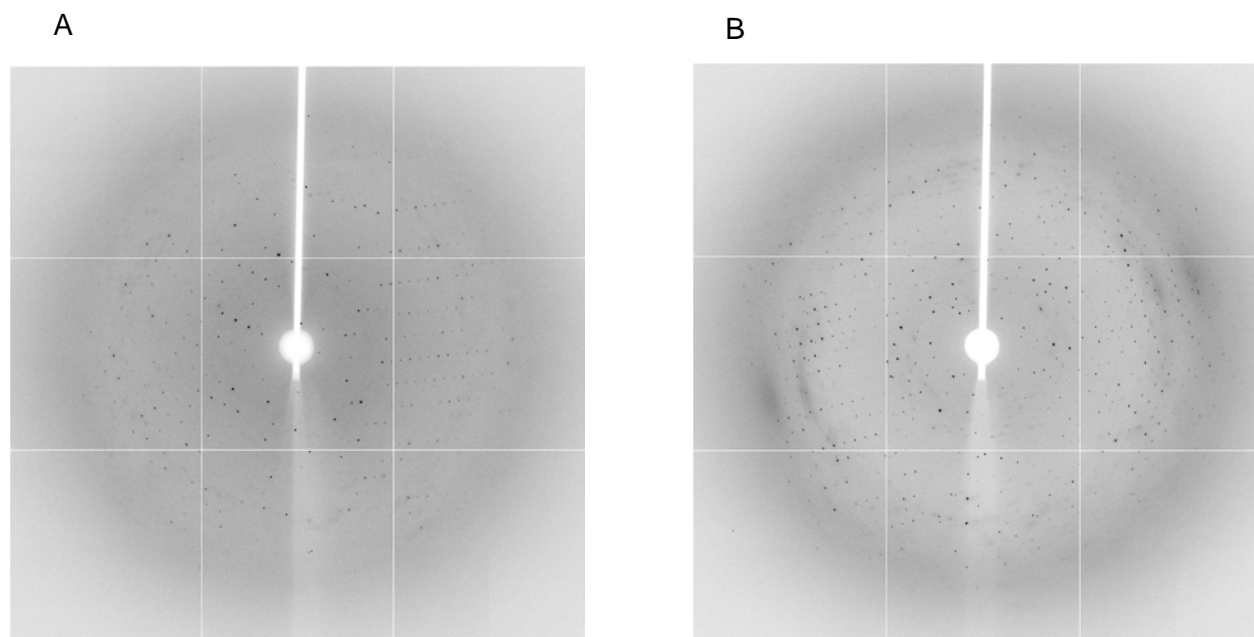
Common features of all the crystals obtained for Lsm[4+1]<sub>4</sub> are their small size and low diffraction resolution and weak reflection intensities. Thus, successful diffraction data collection required the use of a high intensity beam, as provided by a synchrotron source. The micro crystallography (MX2) beamline at the Australian

Synchrotron provides a finely focused beam (30 x 26  $\mu\text{m}$ ) of the necessary brightness and intensity.

A total of 52 promising native crystals, six SeMet crystals and five I3C crystals were screened for diffraction on the MX2 beamline at the Australian Synchrotron. These included:

- i) 20 crystals grown in magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0-7.4), PEG 8000 (8-10 % v/v),
- ii) 11 crystals from trisodium citrate (0.1 M), ammonium sulfate (0.5 M), lithium sulfate (1 M) and cobaltous chloride (10 mM) and
- iii) 21 crystals formed in Tacsimate<sup>TM</sup> (40-60 % v/v) and Tris buffer (0.1 M, pH 7.0-8.0)
- iv) 6 SeMet crystals grown in Tacsimate<sup>TM</sup> (52 % v/v) and Tris buffer (0.1 M, pH 6.5)
- v) 5 crystals from magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 6.5), PEG 8000 (12 % v/v)

Systematic test exposures along the length of these 52 Lsm[4+1]<sub>4</sub> crystals were performed to isolate areas with best diffraction properties. From these, images at  $\Phi$  angles of 0° and 90° were collected as shown in Figure 5.10. Processing of these images of Lsm[4+1]<sub>4</sub> (in MOSFLM) identified crystals belonging to either space group H3 or the symmetry-related H32 group (Table 5.2). Thus, data were collected over 60° and 180° wedges to create complete datasets for either symmetry group H3



**Figure 5.10** | Test exposures of Lsm[4+1]<sub>4</sub> crystal 3 (Table 5.3) grown in Tacsimate™ (44 % v/v), Tris buffer (0.1 M, pH 7.0) taken at A)  $\Phi = 0^\circ$  and B)  $\Phi = 90^\circ$ . Crystal 3 diffracted to  $\sim 3$  Å resolution. Data were collected on the MX2 beamline at the Australian Synchrotron.

or H32. Table 5.3 summarises experimental parameters and strategies utilised for the datasets collected for five separate Lsm[4+1]<sub>4</sub> crystals. These crystals derived from two crystallisation conditions, one utilising sets of crystals 1 and 2 (0.2 M MgCl<sub>2</sub>, 0.1 M Tris buffer pH 7.0, 8 % v/v PEG 8000) and the other crystals 3-5 (44% v/v Tacsimate<sup>TM</sup>, 0.1 M Tris buffer pH 7.0).

A total of six SeMet-Lsm[4+1]<sub>4</sub> crystals and five I3C crystals could be mounted and screened for diffraction. Resolutions of 7 Å and 5 Å were obtained for SeMet-Lsm[4+1]<sub>4</sub> and I3C-Lsm[4+1]<sub>4</sub>, respectively. To accurately measure anomalous signals, especially for weakly diffracting crystals, high data redundancy is required (Dauter and Adams 2001; Debreczeni et al. 2003; Wagner et al. 2006). However, radiation damage evident from quickly declining resolution during data collection prevented collection of a complete and highly redundant dataset.

**Table 5.2|** Unit cell dimensions and space group estimates<sup>a</sup>

No.	Score <sup>b</sup>	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Space group
6	93	161.6	105.8	97.1	89.9	118.4	89.9	C2
5	90	96.5	96.6	106.2	90.0	123.0	113.6	P1
4 <sup>c</sup>	6	105.8	106.2	224.6	90.2	90.0	119.8	H3, H32
		96.5	96.5	97.1	66.5	66.6	66.6	R3, R32
3	4	161.8	106.4	96.5	90.0	118.5	90.3	C2
2	1	161.6	105.8	97.1	90.1	118.4	90.1	C2
1	0	96.5	96.6	97.1	66.6	66.5	66.4	P1

<sup>a</sup> Program used was MOSFLM

<sup>b</sup> Penalty function score: the lower score (< 20) the better

<sup>c</sup> Solution selected for data collection and processing



### 5.3.2 Processing of acquired diffraction data

The five collected datasets obtained on native Lsm[4+1]<sub>4</sub> crystals were indexed and integrated in space group H3 and H32. These initial space groups were estimated based on the lattice geometry (Table 5.2). Spacegroups were subsequently confirmed following further analysis of unmerged integrated intensities (using the POINTLESS routine (Evans 2006)). This approach takes the symmetry of the diffraction pattern into account and thus provides a more reliable space group estimate. The most likely solutions are indicated by a scoring function and a total probability for the solution is determined. For the five Lsm[4+1]<sub>4</sub> datasets space group H32 was suggested in this way as a unique solution with a probability of 99.6-99.8 %.

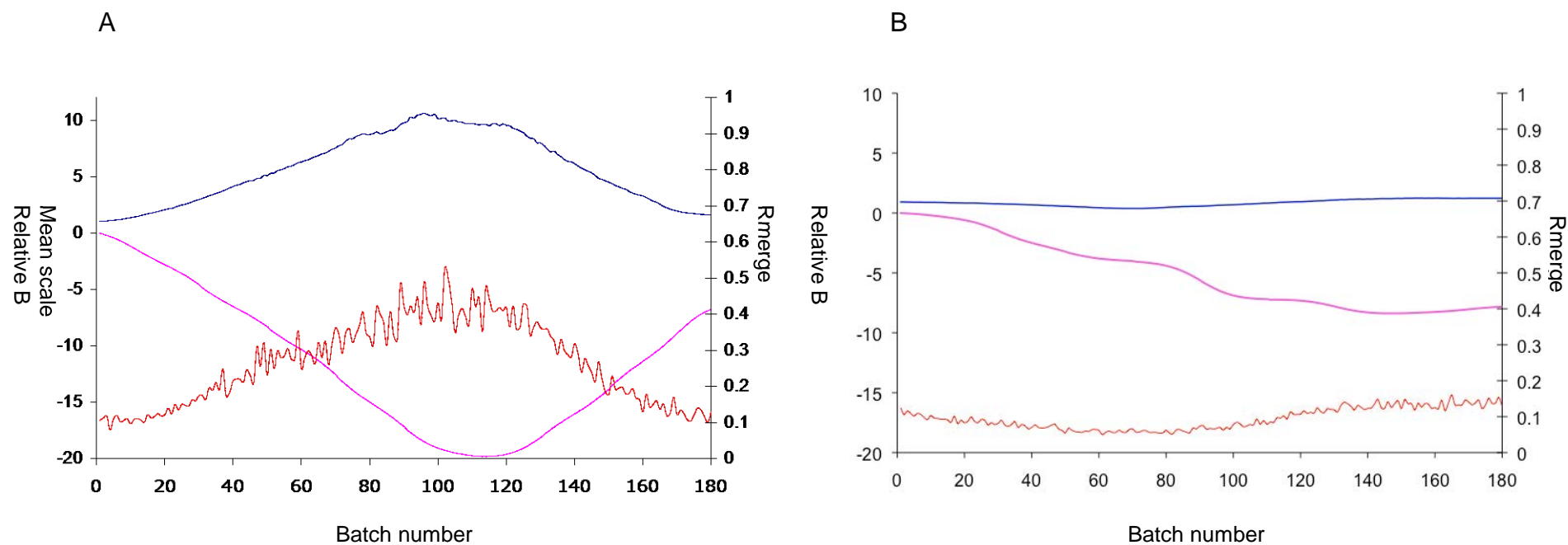
Following data integration and space group confirmation, intensities were scaled, merged and assessed for quality. Plots of intensity scales, relative B-factors and  $R_{\text{merge}}$  over the image number are shown for datasets from crystals 2 and 3 in Figure 5.11.

Scale- and relative B-factors are correction factors used to describe changes in the intensities of symmetry related reflections and radiation damage, respectively (Evans 2006). In an optimal case, these parameters should be uniform and especially relative B-factors below -10 should be treated with suspicion (Evans 2006).  $R_{\text{merge}}$  is a measure of the discrepancies between the intensities of symmetry related reflections (Evans 2006). Consequently, plots of scales, relative B-factors and  $R_{\text{merge}}$  over the image number can be used to identify radiation damage and bad parts of a dataset.

**Table 5.3|** Lsm[4+1]<sub>4</sub> Data collection parameters

Crystal growth conditions	Crystal	Detector distance (mm)	Beamstop (mm)	Exposure time (s)	Beam attenuation (%)	$\Delta\Phi$ (°)	$\Phi$ -range (°)
Magnesium chloride (0.2 M), Tris buffer (0.1 M, pH 7.0), PEG 8000 (8 % v/v)	1	450	50	2	50	1	0-180
	2	450	50	2	20	1	0-180
Tacsimate <sup>TM</sup> (44 % v/v), Tris buffer (0.1 M, pH 7.0)	3	450	50	5	80	1	0-180
	4	450	50	5	80	1	136-316
							136-316 <sup>a</sup>
	5	420	50	5	80	1	90-150
							150-210 <sup>a</sup> 210-270 <sup>a</sup>

<sup>a</sup> Beam was moved to a different location on the same crystal for additional data collection.



**Figure 5.11** | Diffraction data quality assessment. Two datasets are compared regarding their mean scales (blue), the relative B-factor (purple) and the  $R_{\text{merge}}$  (red) over the image number. Datasets shown were collected from A) crystal 2 (Table 5.3) and B) crystal 3 on the MX2 beamline at the Australian Synchrotron.



Data from crystals 1 and 2 showed strong variation in diffraction indices (intensity scales, relative B-factors,  $R_{\text{merge}}$ ) resulting in incomplete datasets. This can likely be attributed to the half-moon-shaped crystal morphology or poor centering of the crystal in the beam. Thus, these datasets were not utilised for analysis, as more stable diffraction was observed for crystals 3-5 of my  $\text{Lsm}[4+1]_4$  material.

The best dataset suitable for structure determination was obtained from  $\text{Lsm}[4+1]_4$  crystal 3 and yielded a resolution of 2.8 Å. Figure 5.11 shows an evaluation of this dataset. Uniform scales, a moderate decrease of the relative B-factors ( $> -8$ ) and uniform  $R_{\text{merge}}$  ( $< 15\%$ ) values were observed indicating a good quality dataset with minimal radiation damage. Crystals 4 and 5 produced similar diffraction quality datasets.

In order to proceed with structure determination utilising this diffraction data, it was necessary to determine the most appropriate resolution cut-off. By established convention (Evans 2006) only data for which the average variance weighted intensities ( $I/\sigma(I)$ ) are  $\geq 2.0$  should be considered. For the dataset of  $\text{Lsm}[4+1]_4$  crystal 3, the resolution cut-off was determined in this way as 3.0 Å. Full crystallographic data collection statistics for this native dataset of  $\text{Lsm}[4+1]_4$  are summarised in Table 5.4. At this resolution, Wilson plots showed a dip at 5 Å resolution and a linear portion between 4.5 Å - 3 Å. This is characteristic for protein crystals.

**Table 5.4**| Selected crystallographic statistics for Lsm[4+1]<sub>4</sub>

<b>Data collection</b>	<b>Crystal 3</b>
Space group	H32
Resolution	3.00 Å
Outer shell	3.16-3.00 Å
Unique reflections (outer shell)	10134 (1463)
Completeness (outer shell)	99.9 (100) %
<I/sigma (I)>	16.7 (3.5)
Multiplicity	10.7 (11.0)
Rmerge <sup>a</sup> (outer shell)	0.09 (0.778)

$$a \sum_h \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l \langle I_h \rangle$$

### 5.3.3 Molecular replacement procedures

At the outset, due to failure to obtain crystals of any derivatised material in the time available, molecular replacement remained an appropriate route to structure determination of Lsm[4+1]<sub>4</sub>.

The Lsm protein fold is highly conserved (see Chapter 1) and there are a variety of crystal structures now available for several archaeal, bacterial and eukaryotic forms (Chapter 1). These include a structure for an octameric organisation of yLsm3 determined recently by the protein structure group at Macquarie University (Naidoo et al. 2008). In addition, quaternary assemblies of Lsm proteins as penta-, hexa-, hepta- and octamers can be observed in other crystal structures. My light scattering and SEC data had very clearly established that the solution form of the polypeptide Lsm[4+1]<sub>4</sub> used as crystal source was tetrameric consisting of eight individual Lsm

domains. Consequently, the coordinates of octameric Lsm3 (PDB 3BW1) appeared to be very appropriate as an input model for molecular replacement. The Lsm3 sequence displays 34 % sequence identity to Lsm1.

Additionally, separate search models for Lsm4 were generated from coordinates of *P. aerophilum* Sm3 (30 % sequence identity; PDB 1M5Q), *A. fulgidus* Sm2 (30 % sequence identity; PDB 1LJO), *H. sapiens* SmB (29 % sequence identity; PDB 1D3B), *C. parvum* Lsm5 (28 % sequence identity; PDB 3PGG) and *S. cerevisiae* Lsm3 (26 % sequence identity; PDB 3BW1). For Lsm1 models, coordinates from *P. aerophilum* Sm1 (39 % sequence identity, PDB 1I8F), *H. sapiens* SmB (33 % sequence identity, PDB 1D3B) and *M. thermoautotrophicum* Lsm $\alpha$  (34 % sequence identity, PDB 1I81) were used as input models in addition to Lsm3 (Table 5.5). Within each search model, side chains that did not align with the target sequence (Lsm1 or Lsm4) were deleted or mutated. Figure 5.12 shows examples of monomeric and dimeric search models derived from the crystal structure of the *HsSmD3B* dimer.

To develop a search strategy for molecular replacement using these models, the number of polyprotein molecules in the asymmetric unit (ASU) was estimated. When the unit cell volume of a crystal and its solvent content are known, the number of protein molecules in the ASU can be determined assuming a partial specific volume of 0.74 cm<sup>3</sup>/g for proteins (Matthews 1968). For a typical protein crystal, the solvent content is found between 23 and 78 % (Matthews 1968) and can be determined from the Matthews coefficient, a measure of the crystal volume per unit of protein molecular weight. Modern solvent analysis tools use a probability calculator to rate

**Table 5.5|** Search models for molecular replacement of Lsm[4+1]<sub>4</sub>

<b>Lsm4</b>	<b>Lsm1</b>	<b>Lsm[4+1]</b>
<i>PaeSm3</i> (30 %; PDB 1M5Q)	<i>PaeSm1</i> (39 %; PDB 1I8F)	<i>HsSmD3B</i> dimer (31 %; PDB 1D3B)
<i>AfSm2</i> (30 %; PDB 1LJO)	<i>MtLsma</i> (34 %; PDB 1I81)	<i>ScLsm3</i> (30 %; PDB 3BW1)
<i>HsSmB</i> (29 %; PDB 1D3B)	<i>ScLsm3</i> (34 %; PDB 3BW1)	
<i>CpLsm5</i> (28 %; PDB 3PGG)	<i>HsSmD3</i> (33 %; PDB 1D3B)	
<i>ScLsm3</i> (26 %; PDB 3BW1)		
<sup>a</sup> 1, 2, 4,6 and 8 copies of the monomeric models placed in the asymmetric unit		
<sup>b</sup> 1, 2, 3 and 4 copies of the dimeric models placed in the asymmetric unit		

possible solvent contents and the corresponding number of molecules in the asymmetric unit (Kantardjieff and Rupp 2003).

For Lsm[4+1]<sub>4</sub> crystal 3, solvent content analysis was carried out using the molecular weight of a single polyprotein unit (23 kDa) to calculate the Matthews coefficient. This yields four possible solutions with Matthews coefficients between 1.3 and 5.3 Å<sup>3</sup>/Da corresponding to 1-4 polyprotein molecules in the ASU (Table 5.6). Out of these results, the most likely solution with a probability of 94 % is obtained for a Matthews coefficient of 2.7 Å<sup>3</sup>/Da corresponding to a solvent content of 54 % and 2 molecules in the ASU. Consequently, molecular replacement searches for 1, 2, 3 and 4 Lsm[4+1] molecules in the ASU were conducted. This further translates to a search for 1, 2, 4, 6 and 8 molecules in the case of monomeric search models.

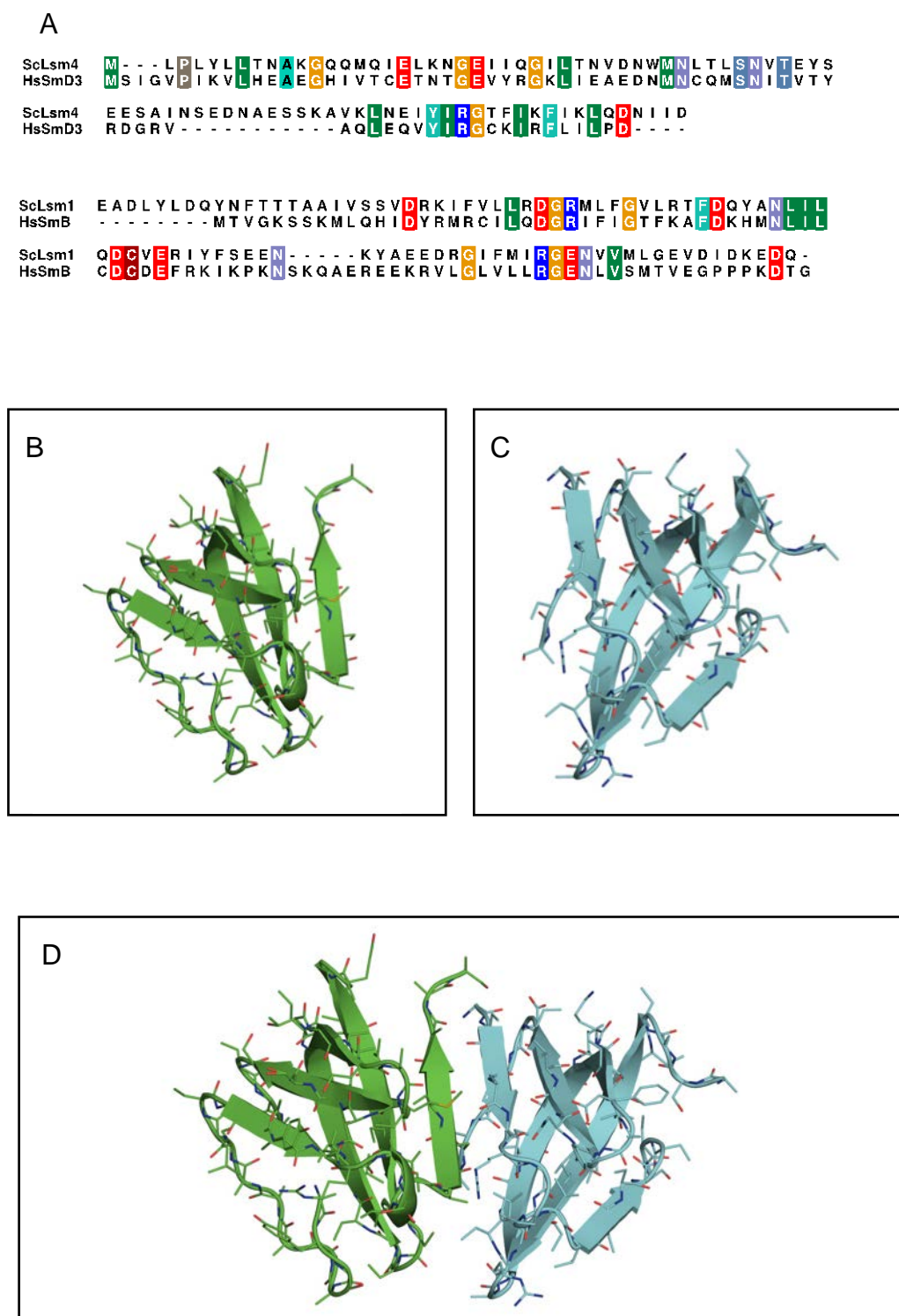


**Table 5.6** Solvent content of Lsm[4+1]<sub>4</sub> crystal 3 in space group H32

Subunits / asymmetric unit	Matthews coefficient	Solvent content (%)	Probability
1	5.33	76.92	0.01
2	2.66	53.84	0.94
3	1.78	30.76	0.04
4	1.33	7.68	0.00

The limited resolution of the diffraction data (3 Å) and the close structural homology between the Lsm4 and the Lsm1 domains in the polyprotein made molecular replacement searches a challenging task. Numerous attempts to place i) 1-4 dimeric and ii) 1-8 monomeric models into the ASU did initially not yield a convincing solution.

However, using a novel, currently undocumented, algorithm for search model editing based on a pairwise sequence alignment, that is embedded in the PHENIX SCULPTOR tool (Bunkoczi 2011), Dr. Airlie McCoy (University of Cambridge) generated a dimeric model for Lsm[4+1] (Figure 5.12) derived from the coordinates of human SmD3B (PDB 1D3B) yielding a single molecular replacement solution. The search model lacks the N-terminal  $\alpha$ -helix and variable loop region L4 and residues that did not align with Lsm4 and Lsm1, respectively, were mutated or deleted. Using this model, Dr. McCoy successfully placed two dimers in the ASU. This molecular replacement solution was characterised by a log likelihood gain (LLG) of 266, a rotation function Z-score (RFZ) of 4.5 and a translation function Z-score (TFZ) of 11.9 indicating an excellent solution. Figure 5.13 displays the C- $\alpha$  trace of the suggested molecular replacement solution. Electron density maps of the



**Figure 5.12|** Lsm[4+1]<sub>4</sub> models based on the SmD3B dimer. A) pairwise alignment of ScLsm4 and ScLsm1 sequences against HsSmD3 and HsSmB used for model design. B) Lsm4 model based on the coordinates of HsSmD3 (PDB 1D3B). C) Lsm1 model generated from HsSmB (PDB 1D3B). D) Lsm[4+1] model designed using coordinates from the HsD3B structure (PDB 1D3B.)

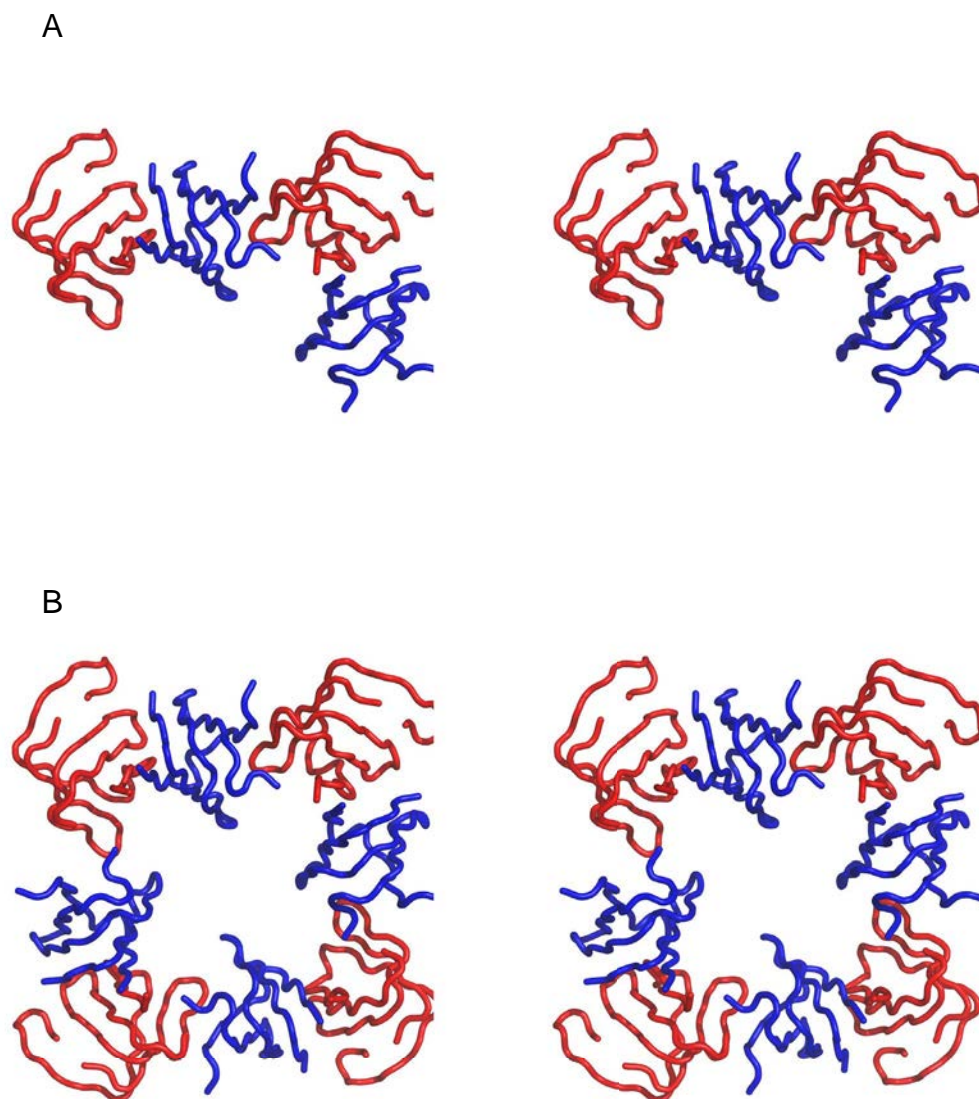
$\beta$ -sheet regions of Lsm4 and Lsm1 from the suggested solution are shown in Figure 5.14.

Dr. McCoy was able to refine the suggested molecular replacement solution for crystal 3 to obtain an R-factor of 40 %. Rigid body movement was observed against a second dataset (from crystal 4) and it did not refine.

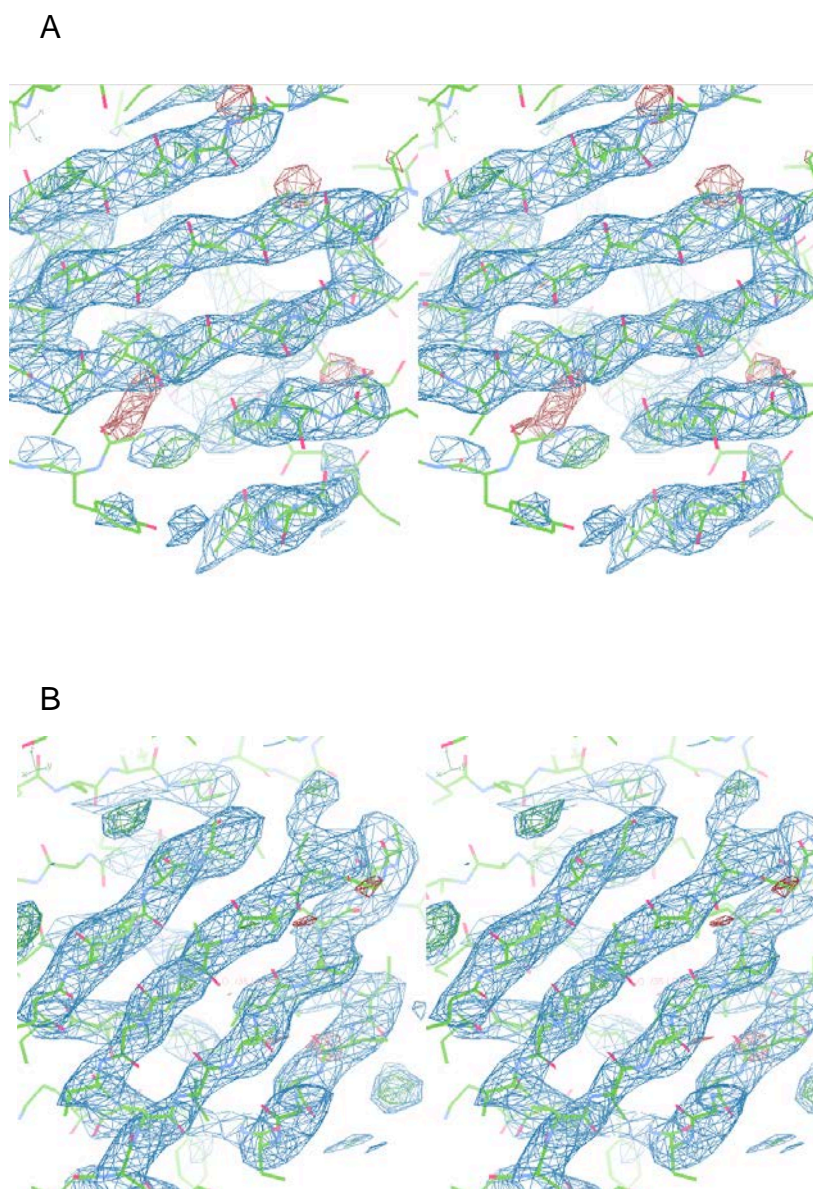
The coordinate file from molecular replacement shows two different types of interfaces between the Lsm4 and Lsm1 subunits. One interface is formed by pairing of strand  $\beta$ 4 of

Lsm4 with strand  $\beta$ 5 of Lsm1. This type of interaction is commonly observed between individual Lsm proteins in complex form (Collins et al. 2001; Beggs 2005; Naidoo et al. 2008). However, in addition the  $\beta$ 5 strand of the same Lsm4 subunit pairs with the  $\beta$ 2 strand of a second Lsm1 chain. Such an interface has not been described in the literature before and it remains unclear if this can be attributed to the covalent connection of the Lsm4 C-terminus to the Lsm1 N-terminus via an unstructured linker sequence or to incorrect packing of two dimer units in the molecular replacement solution.

A cuboid, consisting of four symmetry related dimers, can be generated through application of the crystallographic 2-fold rotation axes in spacegroup H 32 (Figure 5.13). To validate the obtained solution, the coordinate file was used to calculate a



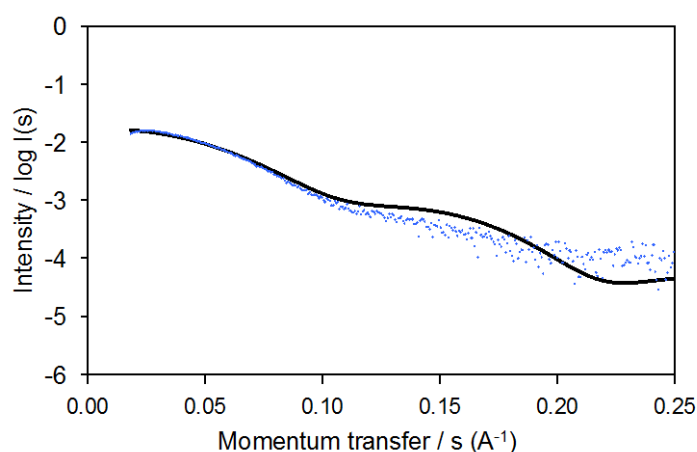
**Figure 5.13|** C- $\alpha$  trace of suggested molecular replacement solution using the PDB coordinates 1D3B (human Sm D3B) as search model in space group H32. Lsm4 subunits are displayed in red, Lsm1 subunits are showed in blue. A) Stereoview of two dimers placed in the ASU. B) Search for symmetry related molecules within  $\pm$  one unit cell and 8 Å reveals octameric organisation.



**Figure 5.14|** Electron density maps from suggested molecular replacement solution. A) Stereoview of the  $\beta$ -sheet region of Lsm4, B) stereoview of the  $\beta$ -sheet region of Lsm1. Electron density is represented as blue mesh contoured at 1.5  $\sigma$ .

theoretical X-ray scattering curve. The resulting curve was compared to solution X-ray scattering from Lsm[4+1]<sub>4</sub>. Both curves could only be aligned with a  $\chi^2$  value of 2.8 suggesting a poor fit. Visual inspection of the alignment confirms poor agreement (Figure 5.15).

Considering the low resolution and the distant model used for molecular replacement, the atypical interactions observed between the Lsm subunits together with the formation of a cuboid rather than a ring structure and the poor fit to solution scattering from Lsm[4+1]<sub>4</sub> indicate that the obtained solution may be a false positive. Further refinement and possibly anomalous diffraction data will be necessary to validate the molecular replacement solution and to solve the crystal structure of Lsm[4+1]<sub>4</sub>.



**Figure 5.15** | Solution X-ray scattering from Lsm[4+1]<sub>4</sub>. Scattering intensities from LSm[4+1]<sub>4</sub> (blue) are overlaid with a theoretical scattering curve derived from the coordinate file of a suggested molecular replacement solution for Lsm[4+1]<sub>4</sub> diffraction data.

## 6 Conclusion

Members of the Lsm superfamily are RNA chaperones and therefore indispensable for RNA processing events in all domains of life. Various members of this protein family are involved in mRNA-splicing, mRNA degradation, maturation of diverse ncRNAs and a variety of disease states. Recent progress in the field also highlights functions of these proteins in the pathogenicity of multidrug resistant *S. aureus* strains and viral replication. Despite their importance in cellular RNA regulation, disease states and host-pathogen interactions, little is known about the mechanism of RNA-binding and target discrimination by heteromeric eukaryotic Lsm protein complexes.

Therefore, a model system is needed to probe binding determinants of Lsm complexes for protein and RNA partners. Such a model system using polyproteins has been established in the Protein Structure group at Macquarie University (Sobti et al. 2010). Lsm polyproteins feature individual Lsm proteins covalently connected by unstructured natural sequence extensions, and allow for the simultaneous expression of two Lsm proteins as heteromeric complexes.

Polyprotein methods have previously been successfully employed for the expression and structure determination of macromolecular complexes. Examples include the *trp* RNA-binding attenuation protein (TRAP) complex structure in which up to four monomers were covalently linked with three alanine residues (Heddle et al. 2006). In another example, the seven subunits of the chaperone complex GroEL were covalently linked by Ala- and Glyrich linkers and successfully expressed (Farr et al.

2000). These published examples utilise simplified peptide segments to ensure full flexibility of linkers. Our Lsm complex differs in this respect by employing linker segments comprising the natural highly charged N- and C-terminal sequence extensions thought to be unstructured in nature.

The aims of this project were to characterise the oligomeric state of Lsm polyproteins and ultimately identify conditions to produce discrete stable complexes. To validate whether polyprotein complexes are suitable to mimic heteromeric complexes of eukaryotic Lsm proteins *in vivo*, ring formation in solution needed to be confirmed. The interaction properties of polyproteins with RNA had to be investigated to examine specific binding determinants within mixed Lsm complexes. Finally, it was also of interest if these Lsm polyproteins served as suitable candidates for protein crystallisation.

## 6.1 Solution behaviour of Lsm polyproteins

Lsm polyproteins were characterised in solution by SEC, demonstrating the formation of multiple oligomeric forms. Oligomerisation was found to be solvent dependent, and large quantities of highly pure and stable tetrameric groupings of Lsm[4+1] and Lsm[2+3] were subsequently isolated. I demonstrated that the presence of glycerol, high ionic strength and high pH values favour these discrete tetrameric assemblies of Lsm polyproteins. Larger oligomers (> 200 kDa) are formed in low salt conditions in the absence of glycerol, indicating electrostatically-mediated multimerisation of Lsm can occur, possibly via (i) helix-to-helix face stacking, (ii) helix-to-loop face stacking or (iii) loop-to-loop face stacking of rings.



- i) Helix-to-helix face stacking of two Lsm protein rings has been observed in the previously solved crystal structure of Lsm[3]<sub>8</sub> (Naidoo et al. 2008). Packing interactions include a salt bridge (Glu2 and Lys9), and contacts between the imidazole rings of His5 and His6 of the N-terminal His<sub>6</sub>-tag.
- ii) Helix-to-loop face stacking of Lsm rings via electrostatic contributions of acidic residues in loops L4 (EDGE) and exposed charged and polar residues in the unstructured N-terminus (R) and the N-terminal  $\alpha$ -helix (NRD) has been observed for archaeal Lsm protein complexes (Mura et al. 2003). *M. thermoautotrophicum* Lsm $\alpha$  (MtLsm $\alpha$ ) was found to produce nanofibril structures, incorporating head-to-tail stacking of heptameric rings in low ionic strength buffers.
- iii) Loop-to-loop face stacking has also been observed within the structure of yeast SmF (Collins et al. 2003). The intermolecular stacking interactions utilised loop L4 residues.

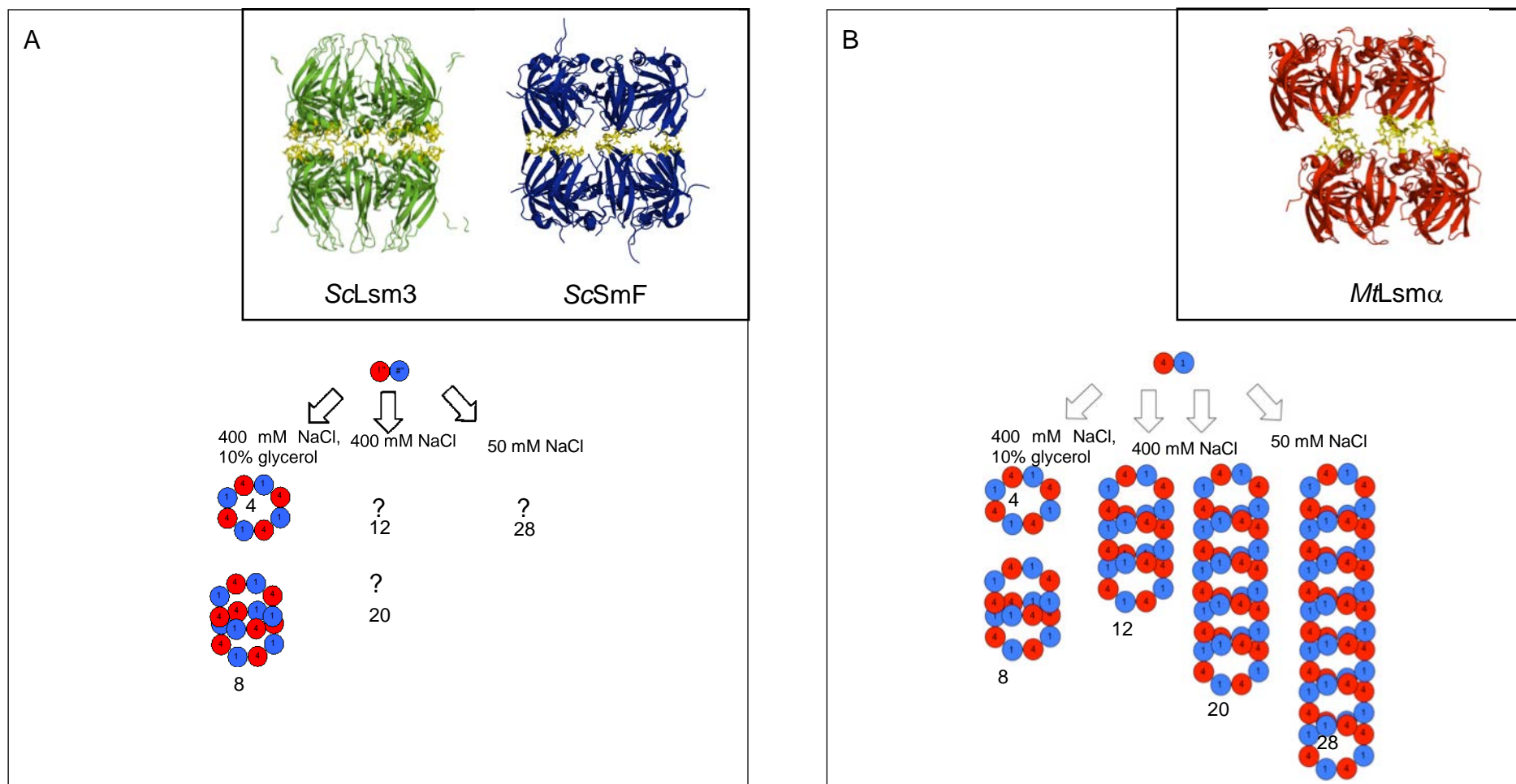
Two of these interaction types, i) and iii) are “symmetrical” in nature; i.e. they engage the same binding interfaces to engage two rings. This type of interaction does is self limiting and therefore not provide for regular polyimetric formation. However, the “asymmetrical” interaction (ii) of helix-to-loop face utilises two distinct binding interfaces. If repeated many times, such stacking could give rise to a large polymer. The oligomeric molecule generated by this type of interaction would in theory only be limited by the pool of available protein rings and mechanical strain. Furthermore, should a polymer engage more than just on type of interaction it is possible to obtain

large multimers through repeated combinations. Figure 6.1 provides a schematic for some proposed Lsm[4+1]<sub>4</sub> oligomeric stacking events explaining detected solution masses.

An important outcome of affinity pull-down experiments conducted by Meghna Sobti demonstrated that Lsm polyproteins are able to recruit specific Lsm partner proteins directly from yeast lysate (Sobti et al. 2010). This indicates that Lsm complexes are more dynamic than previously thought *in vivo*, possibly able to swap components in and out of their (different) oligomeric structures. My results confirm this feature, demonstrating that Lsm complexes are relatively pliable in oligomeric state, taking up distinct forms under varied solvent conditions. Relatively high concentrations of RNA localised within the cell (or nucleus) may thus likely provide the suitable ionic conditions which promote the simple and most stable ring groupings of Lsm proteins needed for their direct interaction prior to processing.

## 6.2 Evidence for ring structures in solution

Solution-based structural analysis using SAXS demonstrated Lsm polyproteins Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> form quaternary assemblies that are cylindrical in shape. In the case of Lsm[4+1]<sub>4</sub>, the determined molecular envelope even displays a central pore, confirming a ring morphology. Dimensions of the solution models obtained for Lsm[4+1]<sub>4</sub> and Lsm[2+3]<sub>4</sub> are consistent with the crystal structure of Lsm[3]<sub>8</sub>. This shows that simplified heteromeric Lsm ring structures with the potential to mimic *in vivo* Lsm complexes have been successfully formed by these polyproteins.



**Figure 6.1** Schematic diagram of multimer formation of Lsm polyproteins in various solvent conditions. Numbers given refer to the number of Lsm[4+1] polyprotein molecules detected in solution. A) Possible face-to-face or tail-to-tail stacking of Lsm rings. Inset shows face-to-face stacking observed in the crystal structure of ScLsm[3]<sub>8</sub> (PDB 3BW1) and tail-to-tail stacking in the crystal structure of ScSmF (PDB 1N9S). B) Possible head-to-tail stacking of Lsm rings. Inset displays head-to-tail stacking seen in the crystal structure of *MtLsmα* (PDB 1JRI ).

Lsm[2+3]<sub>8</sub> was found to be significantly more elongated than both Lsm[2+3]<sub>4</sub> and Lsm[4+1]<sub>4</sub>, and determination of its molecular envelope in solution yielded a long cylinder. The dimensions of the calculated model, however, does not agree with the simple geometry of two stacked Lsm[3]<sub>8</sub> complexes. Hence, other stacked forms of the Lsm rings need to be considered to fully explain the solution species of Lsm.

### 6.3 RNA binding by Lsm polyproteins

Eukaryotic Lsm proteins are known to interact with a plethora of different RNA species. Consensus RNA sequences recognised by these complexes include U-rich sequences at the 3' end of oligoadenylated mRNA (Chowdhury et al. 2007) and ncRNA (Beggs 2005). Despite their common recognition of U-rich stretches on the RNA target sequence, the composition of the hetero-heptameric Lsm complexes determines their RNA target. An example of this target discrimination is the tight interaction ( $K_D = 150\text{-}274\text{ nM}$ ) of the Lsm2-8 complex with U6 snRNA, while the Lsm1-7 complex does not interact with U6 snRNA (Beggs 2005; Licht et al. 2008).

RNA binding by Lsm complexes occurs at three sites (i-iii) on the toroid (Moll et al. 2011). Site i includes highly conserved residues located in loop regions L3 and L5 presented in the lumen of the Lsm ring (Figure 1.5 A). Site ii is formed by N-terminal sequence extensions and elements of the  $\alpha$ -helices on the proximal face (Figure 1.5 B). Site iii comprises loop regions L2 and L4 (Figure 1.5 C).

While slightly increased particle dimensions were observed upon U<sub>10</sub> binding with Lsm[2+3]<sub>4</sub>, my analysis of SAXS parameters indicates no structural rearrangements occur in the protein complex upon RNA binding. This is consistent with a highly stable protein scaffold provided by the Lsm ring system, little altered by its substrate (Toro et al. 2001; Thore et al. 2003; Weber et al. 2010).

Interactions of Lsm polyprotein complexes with U-rich RNA probed by SPR and ITC revealed differential RNA binding affinities of Lsm polyprotein complexes towards U<sub>10</sub>. Lsm[2+3]<sub>4</sub> displayed the highest affinity ( $K_D = 34$  nM), similar to affinities determined for homoheptameric complexes of archaeal *HvLsm $\alpha$*  (72 nM) (Fischer et al. 2011) and homohexameric bacterial *EcHfq* (21 nM, 94 nM) (Mikulecky et al. 2004) complexes towards short RNA oligomers (Table 4.2).

Despite the fact that residues responsible for RNA binding in the lumen of the ring are highly conserved, reduced affinities were observed for Lsm[4+1]<sub>4</sub> and Lsm[3]<sub>8</sub>. The relatively elevated affinity of Lsm[2+3]<sub>4</sub> is thus highly likely to originate in the composition of the linker sequence itself. For this polyprotein, the linker sequence is rich in positively charged Arg and Lys residues, and may form an RNA interaction surface reminiscent of site ii (Figure 4.15).

This RNA-binding behaviour of Lsm[2+3]<sub>4</sub> strongly indicates the major importance of the unstructured and highly varied sequence segments of individual Lsm proteins for target discrimination. These Lsm sequence extensions have also been shown to be determinants for P-body assembly (Reijns et al. 2008), nuclear localisation (Reijns et al. 2009) and RNA binding by heteromeric Lsm complexes (Weber et al.

2010). Further interaction studies utilising Lsm polyproteins will thus need to incorporate additional chemistry from the natural elongated sequences if they are to better mimic specific RNA (and possibly protein) interactions.

## **6.4 Crystallisation of Lsm polyproteins**

The process of protein structure determination by crystallisation encounters three major bottlenecks. First, large amounts of highly pure and stable protein are required; second, conditions for the formation of diffraction-quality crystals must be found; third, the crystallographic phase problem needs to be solved using either molecular replacement or experimental methods. The emergence of recombinant DNA technology has generally greatly simplified the production and purification of sufficient amounts of protein for subsequent crystallisation. Yet, the assembly of multicomponent protein complexes - such as Lsm protein complexes - remains a challenging task throughout structural biology.

I have demonstrated here that high quality preparations of Lsm polyproteins yielding large amounts of protein (> 15 mg/L culture) can now be readily achieved using optimised conditions (Section 3.1). Further, I have demonstrated that the resulting protein samples form discrete monodisperse, stable and globular multimers of tetrameric composition in solution. Having overcome the first rate-limiting step in protein structure determination, the polyproteins Lsm[4+1] and Lsm[2+3] provide excellent starting material for crystallisation experiments.

Lsm polyprotein preparations were repeatedly subjected to a barrage of contemporary crystallisation procedures in my study. The ensuing lack of crystals for Lsm[2+3]<sub>4</sub> is likely due to the flexibility of the linker sequence connecting Lsm domains in this construct. Of differing chemistry, many different crystal forms have successfully been obtained for Lsm[4+1]<sub>4</sub>. Two crystallisation conditions produced crystals that diffracted to a resolution of ~3 Å, indicating the polyprotein to be an appropriate candidate for crystallographic studies. To our knowledge, these are the first diffracting crystals ever to be grown for a heteromeric eukaryotic Lsm complex.

Co-crystallisation of Lsm[4+1]<sub>4</sub> with RNA (U<sub>10</sub>), did not yield crystals that diffracted beyond ~5 Å. My RNA-binding experiments (Chapter 4) indicated only a weak affinity of Lsm[4+1]<sub>4</sub> for this particular RNA oligonucleotide. Hence, a dynamic equilibrium between the polyprotein and the RNA may be preventing quality crystal growth.

Five native datasets were collected and analysed for Lsm[4+1]<sub>4</sub>. Three of these were of sufficient quality for structure determination by molecular replacement. Accordingly, molecular replacement searches using a variety of models were conducted. A single solution containing two Lsm[4+1] molecules in the ASU was obtained. However, non-native β-strand interactions were observed at the interfaces formed between Lsm4 and Lsm1 subunits, suggesting the solution to be incorrect. In addition, generation of symmetry-related molecules indicated unusual packing of four Lsm[4+1] molecules into a cuboid rather than a ring quaternary structure.

A theoretical X-ray scattering curve calculated from this cuboid molecular replacement solution demonstrated a poor fit to the observed Lsm[4+1]<sub>4</sub> solution scattering. Thus, the unusual subunit packing, as well as the poor fit to solution X-ray scattering data indicate the existing molecular replacement solution to be incorrect. Further refinement and possibly diffracting crystals of polyproteins derivatised with heavy atoms will likely be required for full structure determination of Lsm[4+1]<sub>4</sub>. However, it can currently not be excluded that the obtained solution is correct and Lsm[4+1]<sub>4</sub> packs as a cuboid in this crystal form.

## 6.5 Future perspectives and outlook

My solution and interaction studies demonstrated that Lm polyproteins form stable ring complexes in solution suitable for biochemical and structural studies. Differential affinities of Lsm polyprotein towards U-rich oligonucleotides were detected by SPR and ITC and highly charged unstructured sequence extensions were suggested to be involved in RNA interactions and possibly electrostatic stacking of Lsm polyproteins. Diffracting crystals of the polyprotein complex Lsm[4+1]<sub>4</sub> were grown and several native datasets were collected to ~3 Å resolution.

Using the existing polyprotein constructs, the contribution of individual Lsm proteins to the interactions of heteromeric Lsm complexes *in vivo* can be probed. It would be of interest to reveal and display specific epitopes responsible for the functions of individual Lsm proteins. Generation of chimeric constructs using site-directed mutagenesis to exchange specific sequence elements of Lsm proteins may allow specific mapping of epitopes for RNA and protein partners.



The linker region of Lsm[2+3] was suggested to act as an RNA binding site similar to site ii observed in native Lsm complexes. The generation of chimeric constructs could be used to validate this effect. Transplantation of the Lsm[2+3] linker sequence to the Lsm[4+1] construct should result in a significant increase in affinity towards U<sub>10</sub>. Further interaction studies with Lsm polyproteins containing N- and C-terminal sequence extensions could reveal the contribution of these elements to the functions of individual Lsm proteins.

In addition, improvements of sample preparations may allow better diffracting crystals and enable the determination of the first high-resolution structure of a mixed Lsm complex. One avenue to further optimise the Lsm polyprotein samples would be re-engineering of the linker sequences. Due to the characteristic Lsm-fold, linkers connecting C- and N-termini of Lsm proteins are presumed to be located on the surface of the proximal face (Figure 4.16). Since Lsm ring stacking seems to be driven by electrostatic forces, replacing the highly charged natural linkers with Gly and Ala content may provide enhanced stability.

Furthermore,, the use of different linker lengths (currently optimised to minimise strained domain interactions (Sobti et al. 2010) may reduce the flexibility of polyproteins and hence improve diffraction properties of Lsm crystals.

An alternative strategy would be the integration of cleavage sites within the linker sequences. These could take the form of self-cleavage elements, or recognition sites for highly specific proteases (e.g. TEV). This would allow stoichiometric co-

expression of Lsm proteins followed by linker removal to reduce possible strain and inherent flexibility.

In this study, I provided insights into the factors driving the oligomerisation behaviour of Lsm proteins and their RNA interaction properties. Unstructured N- and C-terminal sequence extensions of individual Lsm proteins were suggested to confer specific functions to individual Lsm proteins, emphasising the importance of specific Lsm proteins for the functions of eukaryotic assemblies. Hence, the results obtained from simplified Lsm complexes in this study aid our understanding of Lsm complex assembly *in vivo*.

## 7 References

- Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm and R. Luerhmann (1999). "A doughnut-shaped heteromer of human Sm-like proteins binds to the 3' end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro." *EMBO J.* **18**(20): 5789-5802.
- Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 213-21.
- Afonyushkin, T., B. Vecerek, I. Moll, U. Blasi and V. R. Kaberdin (2005). "Both RNase E and RNase III control the stability of sodB mRNA upon translational inhibition by the small regulatory RNA RyhB." *Nucleic Acids Res* **33**(5): 1678-1689.
- Albrecht, M. and T. Lengauer (2004). "Novel Sm-like proteins with long C-terminal tails and associated methyltransferases." *FEBS Lett* **569**(1-3): 18-26.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." *J Mol Biol* **215**(3): 403-410.
- Anantharaman, V. and L. Aravind (2004). "Novel conserved domains in proteins with predicted roles in eukaryotic cell-cycle regulation, decapping and RNA stability." *BMC Genomics* **5**(1): 45.
- Baca, A. M. and W. G. Hol (2000). "Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli." *Int J Parasitol* **30**(2): 113-118.
- Balbo, P. B. and A. Bohm (2007). "Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis." *Structure* **15**(9): 1117-1131.
- Baneyx, F. (1999). "Recombinant protein expression in Escherichia coli." *Curr Opin Biotechnol* **10**(5): 411-421.
- Beggs, J. D. (2005). "Lsm proteins and RNA processing." *Biochemical Society Transactions* **33**: 439-501.
- Bergfors, T. (2003). "Seeds to crystals." *J Struct Biol* **142**(1): 66-76.

- Boeck, R., B. Lapeyre, C. E. Brown and A. B. Sachs (1998). "Capped mRNA degradation intermediates accumulate in the yeast *spb8-2* mutant." *Mol Cell Biol* **18**(9): 5062-5072.
- Bonnerot, C., R. Boeck and B. Lapeyre (2000). "The two proteins Pat1p (Mrt1p) and Spb8p interact in vivo, are required for mRNA decay, and are functionally linked to Pab1p." *Mol Cell Biol*. **20**(16): 5939-46.
- Bouveret, E., G. Rigaut, A. Shevchenko, M. Wilm and B. Seraphin (2000). "A Sm-like protein complex that participates in mRNA degradation." *EMBO J* **19**(7): 1661-1671.
- Bregues, M., D. Teixeira and R. Parker (2005). "Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies." *Science* **310**(5747): 486-489.
- Brennan, R. G. and T. M. Link (2007). "Hfq structure, function and ligand binding." *Curr Opin Microbiol* **10**(2): 125-133.
- Brow, D. A. (2002). "Allosteric cascade of spliceosome activation." *Annu Rev Genet* **36**: 333-360.
- Bunkoczi, G. and R. J. Read (2011). "Improvement of molecular-replacement models with Sculptor." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 303-312.
- Chable-Bessia, C., O. Meziane, D. Latreille, R. Triboulet, A. Zamborlini, A. Wagschal, J. M. Jacquet, J. Reynes, Y. Levy, A. Saib, Y. Bennasser, and M. Benkirane (2009). 'Suppression of HIV-1 replication by microRNA effectors', *Retrovirology*, **6**(26).
- Chayen, N. E. and E. Saridakis (2008). "Protein crystallization: from purified protein to diffraction-quality crystal." *Nat Methods* **5**(2): 147-153.
- Chi, E. Y., S. Krishnan, T. W. Randolph and J. F. Carpenter (2003). "Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation." *Pharm Res* **20**(9): 1325-1336.
- Chowdhury, A., J. Mukhopadhyay, and S. Tharun (2007b). 'The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs', *RNA* **13**(7): 998-1016.
- Chowdhury, A. and S. Tharun (2009). "Activation of decapping involves binding of the mRNA and facilitation of the post-binding steps by the Lsm1-7-Pat1 complex." *RNA* **15**(10): 1837-1848.

- Coller, J. M., M. Tucker, U. Sheth, M. A. Valencia-Sanchez, and R. Parker (2001). 'The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes', *RNA* **7**(12): 1717-27.
- Collins, B. M., S. J. Harrop, G. D. Kornfeld, I. W. Dawes, P. M. G. Curmi, and B. C. Mabbutt (2001). 'Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs', *J Mol Biol* **309**(4): 915-23.
- Collins, B. M., L. Cubeddu, N. Naidoo, S. J. Harrop, G. D. Kornfeld, I. W. Dawes, P. M. G. Curmi, and B. C. Mabbutt. (2003), 'Homomeric Ring Assemblies of Eukaryotic Sm Proteins Have Affinity for both RNA and DNA', *J Biol Chem*, **278**(May 9): 17291-98.
- Conte, N., E. Charafe-Jauffret, B. Delaval, J. Adelaide, C. Ginestier, J. Geneix, D. Isnardon, J. Jacquemier, and D. Birnbaum (2002). 'Carcinogenesis and translational controls: TACC1 is down-regulated in human cancers and associates with mRNA regulators', *Oncogene* **21**(36): 5619-30.
- Cookson, D., N. Kirby, R. Knott, M. Lee and D. Schultz (2006). "Strategies for data collection and calibration with a pinhole-geometry SAXS instrument on a synchrotron beamline." *J Synchrotron Radiat* **13**(Pt 6): 440-444.
- Crick, F. (1979). "Split genes and RNA splicing." *Science* 204(4390): 264-271.
- Cudney, R., S. Patel, K. Weisgraber, Y. Newhouse and A. McPherson (1994). "Screening and optimization strategies for macromolecular crystal growth." *Acta Crystallogr D Biol Crystallogr* **50**(Pt 4): 414-423.
- Das, D., P. Kozbial, H. L. Axelrod, M. D. Miller, D. McMullan, S. S. Krishna, P. Abdubek, C. Acosta, T. Astakhova, P. Burra, D. Carlton, C. Chen, H. J. Chiu, T. Clayton, M. C. Deller, L. Duan, Y. Elias, M. A. Elsliger, D. Ernst, C. Farr, J. Feuerhelm, A. Grzechnik, S. K. Grzechnik, J. Hale, G. W. Han, L. Jaroszewski, K. K. Jin, H. A. Johnson, H. E. Klock, M. W. Knuth, A. Kumar, D. Marciano, A. T. Morse, K. D. Murphy, E. Nigoghossian, A. Nopakun, L. Okach, S. Oommachen, J. Paulsen, C. Puckett, R. Reyes, C. L. Rife, N. Sefcovic, S. Sudek, H. Tien, C. Trame, C. V. Trout, H. van den Bedem, D. Weekes, A. White, Q. Xu, K. O. Hodgson, J. Wooley, A. M. Deacon, A. Godzik, S. A. Lesley, and I. A. Wilson (2009). 'Crystal structure of a novel Sm-like protein of putative cyanophage origin at 2.60 Å resolution', *Proteins* **75** (2): 296-307.
- Dauter, Z. and D. A. Adamiak (2001). "Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy." *Acta Crystallogr D Biol Crystallogr* **57**(Pt 7): 990-995.

- Debreczeni, J. E., G. Bunkoczi, B. Girmann and G. M. Sheldrick (2003). "In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case." *Acta Crystallogr D Biol Crystallogr* **59**(Pt 2): 393-395.
- Diez, J., M. Ishikawa, M. Kaido and P. Ahlquist (2000). "Identification and characterization of a host protein required for efficient template selection in viral RNA replication." *Proc Natl Acad Sci U S A* **97**(8): 3913-3918.
- Dong, A., X. Xu, A. M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G. V. Avvakumov, A. Bochkarev, L. G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Jr. Finerty, S. Flodin, A. Flores, S. Graslund, M. Hammerstrom, M. D. Herman, B. S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H. W. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J. R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng, and H. Zhu (2007). 'In situ proteolysis for protein crystallization and structure determination', *Nat Methods*, **4** (12), 1019-21.
- Doniach, S. (2001). "Changes in biomolecular conformation seen by small angle X-ray scattering." *Chem Rev* **101**(6): 1763-1778.
- Eddy, S. R. (2001). "Non-coding RNA genes and the modern RNA world." *Nat Rev Genet* **2**(12): 919-929.
- Evans, P. (2006). "Scaling and assessment of data quality." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1): 72-82.
- Falconer, R. J. and B. M. Collins (2011). "Survey of the year 2009: applications of isothermal titration calorimetry." *J Mol Recognit* **24**(1): 1-16.
- Fantappie, L., M. M. Metruccio, K. L. Seib, F. Oriente, E. Cartocci, F. Ferlicca, M. M. Giuliani, V. Scarlato, and I. Delany (2009). 'The RNA chaperone Hfq is involved in stress response and virulence in *Neisseria meningitidis* and is a pleiotropic regulator of protein expression', *Infect Immun* **77**(5): 1842-53.
- Farr, G. W., K. Furtak, M. B. Rowland, N. A. Ranson, H. R. Saibil, T. Kirchhausen and A. L. Horwich (2000). "Multivalent binding of nonnative substrate proteins by the chaperonin GroEL." *Cell* **100**(5): 561-573.
- Feig, A. L. (2009). "Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry." *Methods Enzymol* **468**: 409-422.

- Fischer, S., J. Benz, B. Spath, L. K. Maier, J. Straub, M. Granzow, M. Raabe, H. Urlaub, J. Hoffmann, B. Brutschy, T. Allers, J. Soppa, and A. Marchfelder (2011). 'The archaeal Lsm protein binds to small RNAs', *J Biol Chem* **285**(45): 34429-38.
- Franze de Fernandez, M. T., L. Eoyang and J. T. August (1968). "Factor fraction required for the synthesis of bacteriophage Qbeta-RNA." *Nature* **219**(5154): 588-590.
- Fraser, M. M., P. M. Watson, M. M. Fraig, J. R. Kelley, P. S. Nelson, A. M. Boylan, D. J. Cole, and D. K. Watson (2005). 'CaSm-mediated cellular transformation is associated with altered gene expression and messenger RNA stability', *Cancer Res* **65**(14): 6228-36.
- Fromont-Racine, M., A. E. Mayes, A. Brunet-Simon, J. C. Rain, A. Colley, I. L. Dix, Decourty, N. Joly, F. Ricard, J. D. Beggs, and P. Legrain (2000). 'Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins', *Yeast* **17**(2): 95-110.
- Galao, R. P., A. Chari, I. Alves-Rodrigues, D. Lobao, A. Mas, C. Kambach, U. Fischer, and J. Diez (2010). 'LSm1-7 complexes bind to specific sites in viral RNA genomes and regulate their translation and replication', *RNA* **16** (4): 817-27.
- Garneau, N. L., J. Wilusz, and C. J. Wilusz (2007). "The highways and byways of mRNA decay." *Nat Rev Mol Cell Biol* **8**(2): 113-26.
- Glasel, J. A. (1995). "Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios." *Biotechniques* **18**(1): 62-63.
- Glatter, O. (1977). "A new Method for the Evaluation of Small-Angle Scattering Data." *J. Appl. Cryst.* **10**: 415-421
- Gottesman, S. and G. Storz (2010). "Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations." *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a003798.
- Gupta, A., J. L. Jenkins and C. L. Kielkopf (2010). "RNA induces conformational changes in the SF1/U2AF65 splicing factor complex." *J Mol Biol* **405**(5): 1128-1138.
- Harrick, N. (1967). International Reflection Spectroscopy, Interscience.
- He, W. and R. Parker (2001). "The yeast cytoplasmic LsmI/Pat1p complex protects mRNA 3' termini from partial degradation." *Genetics* **158**(4): 1445-1455.
- Heddle, J. G., T. Yokoyama, I. Yamashita, S. Y. Park and J. R. Tame (2006). "Rounding up: Engineering 12-membered rings from the cyclic 11-mer TRAP." *Structure* **14**(5): 925-933.

- Hennessy, D. N., B. Narayanan and J. M. Rosenberg (2009). "Automatic implementation of precise grid screens: the four-corners method." *Acta Crystallogr D Biol Crystallogr* **65**(Pt 9): 1001-1003.
- Inoue, H., H. Nojima and H. Okayama (1990). "High efficiency transformation of *Escherichia coli* with plasmids." *Gene* **96**(1): 23-28.
- Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Luhrmann, J. Li, and K. Nagai (1999). 'Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs', *Cell* **96**(3): 375-87.
- Kantardjieff, K. A. and B. Rupp (2003). "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." *Protein Sci* **12**(9): 1865-1871.
- Karaduman, R., P. Dube, H. Stark, P. Fabrizio, B. Kastner and R. Luhrmann (2008). "Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSsm complex as revealed by electron microscopy." *RNA* **14**(12): 2528-2537.
- Karaduman, R., P. Fabrizio, K. Hartmuth, H. Urlaub, and R. Luhrmann (2006). 'RNA structure and RNA-protein interactions in purified yeast U6 snRNPs', *J Mol Biol* **356**(5): 1248-62.
- Karaduman, R., P. Dube, H. Stark, P. Fabrizio, B. Kastner, and R. Luhrmann (2008). 'Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSsm complex as revealed by electron microscopy', *RNA* **14**(12): 2528-37.
- Karlsson, R. (1994). "Real-time competitive kinetic analysis of interactions between low-molecular-weight ligands in solution and surface-immobilized receptors." *Anal Biochem* **221**(1): 142-151.
- Katsamba, P., D. Myszka and I. A. Laird-Offringa (2001). "Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA." *J Biol Chem* **276**(24): 21476-21481.
- Katsamba, P. S., S. Park, and I. A. Laird-Offringa (2002), 'Kinetic studies of RNA-protein interactions using surface plasmon resonance', *Methods* **26**(2): 95-104.
- Kelley, J. R., J. M. Brown, M. M. Frasier, P. L. Baron, C. W. Schweinfest, J. N. Vournakis, D. K. Watson, and D. J. Cole, (2000). 'The cancer-associated Sm-like oncogene: a novel target for the gene therapy of pancreatic cancer', *Surgery* **128**(2): 353-60.
- Kelley, J. R., M. M. Fraser, C. W. Schweinfest, J. N. Vournakis, D. K. Watson and D. J. Cole (2001). "CaSm/gemcitabine chemo-gene therapy leads to



- prolonged survival in a murine model of pancreatic cancer." *Surgery* **130**(2): 280-288.
- Khusial, P., R. Plaag, and G. W. Zieve (2005). 'LSm proteins form heptameric rings that bind to RNA via repeating motifs', *Trends Biochem Sci* **30**(9): 522-8.
- Kim, H. S., Y. Kuwano, M. Zhan, R. Jr. Pullmann, K. Mazan-Mamczarz, H. Li, N. Kedersha, P. Anderson, M. C. Wilce, M. Gorospe, and J. A. Wilce (2007a). 'Elucidation of a C-rich signature motif in target mRNAs of RNA-binding protein TIAR', *Mol Cell Biol* **27**(19) 6806-17.
- Knubovets, T., J. J. Osterhout, P. J. Connolly and A. M. Klibanov (1999). "Structure, thermostability, and conformational flexibility of hen egg-white lysozyme dissolved in glycerol." *Proc Natl Acad Sci U S A* **96**(4): 1262-1267.
- Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." *J Appl Crystallogr* **36**(5): 1277-1282.
- Kornblatt, J. A., M. J. Kornblatt, G. H. Hoa and A. G. Mauk (1993). "Responses of two protein-protein complexes to solvent stress: does water play a role at the interface?" *Biophys J* **65**(3): 1059-1065.
- Kortt, A. A., E. Nice and L. C. Gruen (1999). "Analysis of the binding of the Fab fragment of monoclonal antibody NC10 to influenza virus N9 neuraminidase from tern and whale using the BIAcore biosensor: effect of immobilization level and flow rate on kinetic analysis." *Anal Biochem* **273**(1): 133-141.
- Kozin, M. and D. I. Svergun (2001). "Automated matching of high- and low-resolution structural models." *J Appl Crystallogr* **34**: 33-41.
- Kretschmann, E. and H. Raether (1968). *Zeitschrift f. Naturforschung* 23.
- Kufel, J., C. Allmang, E. Petfalski, J. Beggs and D. Tollervey (2003a). "Lsm Proteins are required for normal processing and stability of ribosomal RNAs." *J Biol Chem* **278**(4): 2147-2156.
- Kufel, J., C. Allmang, L. Verdone, J. Beggs and D. Tollervey (2003b). "A complex pathway for 3' processing of the yeast U3 snoRNA." *Nucleic Acids Res* **31**(23): 6788-6797.
- Kufel, J., C. Allmang, L. Verdone, J. D. Beggs and D. Tollervey (2002). "Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p." *Mol Cell Biol* **22**(14): 5248-5256.
- Kufel, J., C. Bousquet-Antonelli, J. D. Beggs and D. Tollervey (2004). "Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex." *Mol Cell Biol* **24**(21): 9646-9657.

- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**(5259): 680-685.
- Lease, R. A. and S. A. Woodson (2004). "Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA." *J Mol Biol* **344**(5): 1211-1223.
- Lefebvre, S., L. Burglen, S. Reboullet, O. Clermont, P. Burlet, L. Viollet, B. Benichou, C. Cruaud, P. Millasseau, M. Zeviani (1995), "Identification and characterization of a spinal muscular atrophy-determining gene." *Cell* **80**(1): 155-65.
- Lerner, M. R. and J. A. Steitz (1979). "Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus." *Proc Natl Acad Sci U S A* **76**(11): 5495-5499.
- Leslie, A. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**.
- Licht, K., J. Medenbach, R. Luerhmann, C. Kambach and A. Bindereif (2008). "3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins." *RNA* **14**(8): 1-7.
- Lilley, D. M. (2005). "Structure, folding and mechanisms of ribozymes." *Curr Opin Struct Biol* **15**(3): 313-323.
- Ling, S. H., C. J. Decker, M. A. Walsh, M. She, R. Parker and H. Song (2008). "Crystal structure of human Edc3 and its functional implications." *Mol Cell Biol* **28**(19): 5965-5976.
- Link, T. M., P. Valentin-Hansen and R. G. Brennan (2009). "Structure of Escherichia coli Hfq bound to polyribadenylate RNA." *Proc Natl Acad Sci U S A* **106**(46): 19292-19297.
- Liu, J., D. C. Yin, Y. Z. Guo, X. K. Wang, S. X. Xie, Q. Q. Lu and Y. M. Liu (2011). "Selecting temperature for protein crystallization screens using the temperature dependence of the second virial coefficient." *PLoS One* **6**(3): e17950.
- Liu, Y., N. Wu, J. Dong, Y. Gao, X. Zhang, C. Mu, N. Shao, and G. Yang (2010), 'Hfq is a global regulator that controls the pathogenicity of Staphylococcus aureus', *PLoS One* **5**(9).
- Livny, J. and M. K. Waldor (2007). "Identification of small RNAs in diverse bacterial species." *Curr Opin Microbiol* **10**(2): 96-101.
- Luft, J. R. and G. T. DeTitta (1999). "A method to produce microseed stock for use in the crystallization of biological macromolecules." *Acta Crystallogr D Biol Crystallogr* **55**(Pt 5): 988-993.

- Luhrmann, R., B. Kastner and M. Bach (1990). "Structure of spliceosomal snRNPs and their role in pre-mRNA splicing." *Biochim Biophys Acta* **1087**(3): 265-292.
- Lyng, H., R. S. Brovig, D. H. Svendsrud, R. Holm, O. Kaalhus, K. Knutstad, H. Oksefjell, K. Sundfor, G. B. Kristensen, and T. Stokke (2006). 'Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer', *BMC Genomics* **7**, 268.
- Ma, Y., J. Dostie, G. Dreyfuss and G. D. Van Duyne (2005). "The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure." *Structure* **13**(6): 883-892.
- Majdalani, N., C. K. Vanderpool and S. Gottesman (2005). "Bacterial small RNA regulators." *Crit Rev Biochem Mol Biol* **40**(2): 93-113.
- Mas, A., I. Alves-Rodrigues, A. Noueiry, P. Ahlquist and J. Diez (2006). "Host deadenylation-dependent mRNA decapping factors are required for a key step in brome mosaic virus RNA replication." *J Virol* **80**(1): 246-251.
- Matthews, B. W. (1968). "Solvent content of protein crystals." *J Mol Biol* **33**(2): 491-497.
- Mattick, J. S. (2001). "Non-coding RNAs: the architects of eukaryotic complexity." *EMBO Rep* **2**(11): 986-991.
- Mattick, J. S. and I. V. Makunin (2006). "Non-coding RNA." *Hum Mol Genet* **15** Spec No 1: R17-29.
- McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." *J Appl Crystallogr* **40**(Pt 4): 658-674.
- McPherson, A. (2001). "A comparison of salts for the crystallization of macromolecules." *Protein Sci* **10**(2): 418-422.
- McPherson, A. and B. Cudney (2006). "Searching for silver bullets: an alternative strategy for crystallizing macromolecules." *J Struct Biol* **156**(3): 387-406.
- McPhillips, T. M., S. E. Phillips, H. J. Chiu, A. E. Cohen, A. M. Deacon, P. J. Ellis, E. Garman, A. Gonzalez, N. K. Sauter, R. P. Phizackerley, S. M. Soltis, and P. Kuhn, (2002). "Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines." *J Synchrotron Radiat* **9**(Pt 6): 401-6.
- Mertens, H. D. and D. I. Svergun (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." *J Struct Biol* **172**(1): 128-141.

- Mierendorf, R. and N. Yeager (1984). "The pET system: Your choice for Expression." *in* *Novations*.
- Mikulecky, P. J., M. K. Kaw, C. C. Brescia, J. C. Takach, D. D. Sledjeski, and A. L. Feig (2004). "Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. " *Nat Struct Mol Biol* **11**(12): 1206-14.
- Mishra, R., R. Bhat and R. Seckler (2007). "Chemical chaperone-mediated protein folding: stabilization of P22 tailspike folding intermediates by glycerol." *Biol Chem* **388**(8): 797-804.
- Moll, I., D. Leitsch, T. Steinhauser and U. Blasi (2003). "RNA chaperone activity of the Sm-like Hfq protein." *EMBO Rep* **4**(3): 284-289.
- Moll, J. M., M. Sobti and B. C. Mabbutt (2011). The Lsm proteins: Ring architectures for RNA capture. *RNA processing*, (ed. Grabowski P.) *InTech*
- Morita, T., K. Maki and H. Aiba (2005). "RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs." *Genes Dev* **19**(18): 2176-2186.
- Morita, T., Y. Mochizuki and H. Aiba (2006). "Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction." *Proc Natl Acad Sci U S A* **103**(13): 4858-4863.
- Mura, C., A. Kozhukhovskiy, M. Gingery, M. Phillips and D. Eisenberg (2003). "The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs)." *Protein Sci* **12**(4): 832-847.
- Mura, C., M. Phillips, A. Kozhukhovskiy and D. Eisenberg (2003). "Structure and assembly of an augmented Sm-like archaeal protein 14-mer." *Proc Natl Acad Sci U S A* **100**(8): 4539-4544.
- Myszka, D. G., M. D. Jonsen and B. J. Graves (1998). "Equilibrium analysis of high affinity interactions using BIACORE." *Anal Biochem* **265**(2): 326-330.
- Naidoo, N., S. J. Harrop, M. Sobti, P. A. Haynes, B. R. Szymczyna, J. R. Williamson, P. M. G. Curmi, and B. C Mabbutt (2008). "Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: implications for Lsm ring organisation and recruitment. " *J Mol Biol* **377**(5): 1357-71.
- Nakel, K., S. A. Hartung, F. Bonneau, C. R. Eckmann and E. Conti (2010). "Four KH domains of the *C. elegans* Bicaudal-C ortholog GLD-3 form a globular structural platform." *RNA* **16**(11): 2058-2067.

- Newman, J., D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart, and A. Perrakis (2005). "Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy." *Acta Crystallogr D Biol Crystallogr* **61**(Pt 10): 1426-31.
- Nissan, T., P. Rajyaguru, M. She, H. Song and R. Parker (2010). "Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms." *Mol Cell* **39**(5): 773-783.
- Noueiry, A. O., J. Diez, S. P. Falk, J. Chen and P. Ahlquist (2003). "Yeast Lsm1p-7p/Pat1p deadenylation-dependent mRNA-decapping factors are required for brome mosaic virus genomic RNA translation." *Mol Cell Biol* **23**(12): 4094-4106.
- O'Shannessy, D. J. and D. J. Winzor (1996). "Interpretation of deviations from pseudo-first-order kinetic behavior in the characterization of ligand binding by biosensor technology." *Anal Biochem* **236**(2): 275-283.
- Pace, C. N., F. Vajdos, L. Fee, G. Grimsley and T. Gray (1995). "How to measure and predict the molar absorption coefficient of a protein." *Protein Sci* **4**(11): 2411-2423.
- Parker, R. and U. Sheth (2007). "P bodies and the control of mRNA translation and degradation." *Mol Cell* **25**(5): 635-646.
- Patel, S. B. and M. Bellini (2008). "The assembly of a spliceosomal small nuclear ribonucleoprotein particle." *Nucleic Acids Res* **36**(20): 6482-6493.
- Peng, R. and D. Gallwitz (2004). "Multiple SNARE interactions of an SM protein: Sed5p/Sly1p binding is dispensable for transport." *EMBO J* **23**(20): 3939-3949.
- Pomeranz Krummel, D. A., C. Oubridge, A. K. Leung, J. Li and K. Nagai (2009). "Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution." *Nature* **458**(7237): 475-480.
- Pozzoli, U., M. Sironi, R. Cagliani, G. P. Comi, A. Bardoni and N. Bresolin (2002). "Comparative analysis of the human dystrophin and utrophin gene structures." *Genetics* **160**(2): 793-798.
- Priev, A., A. Almagor, S. Yedgar and B. Gavish (1996). "Glycerol decreases the volume and compressibility of protein interior." *Biochemistry* **35**(7): 2061-2066.
- Proudfoot, N. J., A. Furger and M. J. Dye (2002). "Integrating mRNA processing with transcription." *Cell* **108**(4): 501-512.

- Putnam, C. D., M. Hammel, G. L. Hura and J. A. Tainer (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." *Q Rev Biophys* **40**(3): 191-285.
- Rader, S. D. and C. Guthrie (2002). "A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation." *RNA* **8**(11): 1378-1392.
- Raker, V., K. Hartmuth, B. Kastner and R. Luhrmann (1999). "Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner." *Mol Cell Biol* **19**(10): 6554-6565.
- Raker, V., G. Plessel and R. Luerhmann (1996). "The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro." *EMBO J* **15**(9): 2256-2269.
- Recht, M. I., S. P. Ryder and J. R. Williamson (2008). "Monitoring assembly of ribonucleoprotein complexes by isothermal titration calorimetry." *Methods Mol Biol* **488**: 117-127.
- Reijns, M. A., R. D. Alexander, M. P. Spiller and J. D. Beggs (2008). "A role for Q/N-rich aggregation-prone regions in P-body localization." *J Cell Sci* **121**(Pt 15): 2463-2472.
- Reijns, M. A., T. Auchynnikava and J. D. Beggs (2009). "Analysis of Lsm1p and Lsm8p domains in the cellular localization of Lsm complexes in budding yeast." *FEBS J* **276**(13): 3602-3617.
- Rosen, J., M. He, C. Umbricht, H. R. Alexander, A. P. Dackiw, M. A. Zeiger and S. K. Libutti (2005). "A six-gene model for differentiating benign from malignant thyroid tumors on the basis of gene expression." *Surgery* **138**(6):1050-1056; discussion 1056-1057.
- Rosenberg, A. H., B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn and F. W. Studier (1987). "Vectors for selective expression of cloned DNAs by T7 RNA polymerase." *Gene* **56**(1): 125-135.
- Ryan, D. E., S. W. Stevens and J. Abelson (2002). "The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telostem region." *RNA* **8**(8): 1011-1033.
- Salgado-Garrido, J., E. Bragado-Nielsson, S. Kandels-Lewis and B. Seraphin (1999). "Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin." *EMBO J* **18**(12): 3451-3462.
- Sambrook, J. and D. Russell (2001). *Molecular Cloning Cold Spring Harbour Laboratory Press (Third edition)*.

- Scharnagl, C., M. Reif and J. Friedrich (2005). "Stability of proteins: temperature, pressure and the role of the solvent." *Biochim Biophys Acta* **1749**(2): 187-213.
- Scheller, N., L. B. Mina, R. P. Galao, A. Chari, M. Gimenez-Barcons, A. Noueiry, U. Fischer, A. Meyerhans, and J. Diez (2009), "Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates." *Proc Natl Acad Sci U S A* **106**(32), 13517-22.
- Schumacher, M., R. Pearson, T. Moller, P. Valentin-Hansen and R. Brennan (2002). "Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein." *EMBO J* **21**(13): 3546-3556.
- Schweinfest, C. W., M. W. Graber, J. M. Chapman, T. S. Papas, P. L. Baron and D. K. Watson (1997). "CaSm: an Sm-like protein that contributes to the transformed state in cancer cells." *Cancer Res* **57**(14): 2961-2965.
- Selenko, P., R. Sprangers, G. Stier, D. Buhler, U. Fischer and M. Sattler (2001). "SMN tudor domain structure and its interaction with the Sm proteins." *Nat Struct Biol* **8**(1): 27-31.
- Sheth, U. and R. Parker (2003). "Decapping and decay of messenger RNA occur in cytoplasmic processing bodies." *Science* **300**(5620): 805-808.
- Slotboom, D. J., R. H. Duurkens, K. Olieman and G. B. Erkens (2008). "Static light scattering to characterize membrane proteins in detergent solution." *Methods* **46**(2): 73-82.
- Smith R. M., A. E. Martell and R. J. Motekaitis (1985). Prediction of stability constants. I. Protonation constants of carboxylates and formation constants of their complexes with class a metal ions. *Inorganica Chim Acta* **9**(2):207-216.
- Sobti, M. (2008). Characterization of Lsm family of RNA-binding proteins. *Department of Chemistry & Biomolecular Sciences*, Macquarie University.
- Sobti, M., L. Cubeddu, P. A. Haynes and B. C. Mabbutt (2010). "Engineered rings of mixed yeast Lsm proteins show differential interactions with translation factors and U-rich RNA." *Biochemistry* **49**(11): 2335-2345.
- Spiller, M. P., K. L. Boon, M. A. Reijns and J. D. Beggs (2007). "The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA." *Nucleic Acids Res* **35**(3): 923-929.
- Stein, N. (2008). "CHAINSAW: A program for mutating pdb files used as templates in molecular replacement." *J Appl Crystallogr* **41**: 641-643.
- Storz, G., J. A. Opdyke and A. Zhang (2004). "Controlling mRNA stability and translation with small, noncoding RNAs." *Curr Opin Microbiol* **7**(2): 140-144.

- Streicher, K. L., Z. Q. Yang, S. Draghici and S. P. Ethier (2007). "Transforming function of the LSM1 oncogene in human breast cancers with the 8p11-12 amplicon." *Oncogene* **26**(14): 2104-2114.
- Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." *Protein Expr Purif* **41**(1): 207-234.
- Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." *J Mol Biol* **189**(1): 113-130.
- Sukhodolets, M. V. and S. Garges (2003). "Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq." *Biochemistry* **42**(26): 8022-8034.
- Svergun, D. I. (1992). "Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria." *J Appl Cryst* **25**: 495-503.
- Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." *Biophys J* **76**(6): 2879-2886.
- Svergun, D., C. Barberato, M. Malfois, V. Volkov, P. Konarev, M. Petoukhov and A. Shkumatov (1995). "CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates." *J Appl Crystallogr* **28**: 768-773.
- Taylor, G. (2003). "The phase problem." *Acta Crystallogr D Biol Crystallogr* **59**(Pt 11): 1881-1890.
- Te'o, V. S., A. E. Cziferszky, P. L. Bergquist and K. M. Nevalainen (2000). "Codon optimization of xylanase gene xynB from the thermophilic bacterium Dictyoglomus thermophilum for expression in the filamentous fungus Trichoderma reesei." *FEMS Microbiol Lett* **190**(1): 13-19.
- Tharun, S., W. He, A. E. Mayes, P. Lennertz, J. D. Beggs and R. Parker (2000). "Yeast Sm-like proteins function in mRNA decapping and decay." *Nature* **404**(6777): 515-518.
- Thore, S., C. Mayer, C. Sauter, S. Weeks and D. Suck (2003). "Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya." *J Biol Chem* **278**(2): 1239-1247.
- Toro, I., J. Basquin, H. Teo-Dreher and D. Suck (2002). "Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus." *J Mol Biol* **320**(1): 129-142.



- Toro, I., S. Thore, C. Mayer, J. Basquin, B. Seraphin and D. Suck (2001). "RNA binding in an Sm core domain: X-ray structure and functional analysis of an archeal Sm protein complex." *EMBO J* **20**(9): 2293-2303.
- Tritschler, F., A. Eulalio, V. Truffault, M. D. Hartmann, S. Helms, S. Schmidt, M. Coles, E. Izaurralde, and O. Weichenrieder (2007). "A divergent Sm fold in EDC3 proteins mediates DCP1 binding and P-body targeting." *Mol Cell Biol*, **27**(24): 8600-11.
- Tsui, H. C., H. C. Leung and M. E. Winkler (1994). "Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12." *Mol Microbiol* **13**(1): 35-49.
- Tucker, M., M. A. Valencia-Sanchez, R. R. Staples, J. Chen, C. L. Denis and R. Parker (2001). "The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae." *Cell* **104**(3): 377-386.
- Urlaub, H., V. A. Raker, S. Kostka and R. Luhrmann (2001). "Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure." *EMBO J* **20**(1-2): 187-196.
- Vagenende, V., M. G. Yap and B. L. Trout (2009). "Mechanisms of protein stabilization and prevention of protein aggregation by glycerol." *Biochemistry* **48**(46): 11084-11096.
- Vedadi, M., J. Lew, J. Artz, M. Amani, Y. Zhao, A. Dong, G. A. Wasney, M. Gao, T. Hills, S. Brokx, W. Qiu, S. Sharma, A. Diassiti, Z. Alam, M. Melone, A. Mulichak, A. Wernimont, J. Bray, P. Loppnau, O. Plotnikova, K. Newberry, E. Sundararajan, S. Houston, J. Walker, W. Tempel, A. Bochkarev, I. Kozieradzki, A. Edwards, C. Arrowsmith, D. Roos, K. Kain and R. Hui (2007). "Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms." *Mol Biochem Parasitol* **151**(1): 100-10.
- Veretnik, S., C. Wills, P. Youkharibache, R. E. Valas and P. E. Bourne (2009). "Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome." *PLoS Comput Biol* **5**(3): e1000315.
- Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." *J Appl Crystallogr* **36**: 860-864.
- Wagner, A., M. Pieren, C. Schulze-Briese, K. Ballmer-Hofer and A. E. Prota (2006). "Structure determination of VEGF-E by sulfur SAD." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 11): 1430-1434.
- Wahl, M. C., C. L. Will and R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." *Cell* **136**(4): 701-718.

- Wan, L., D. J. Battle, J. Yong, A. K. Gubitz, S. J. Kolb, J. Wang and G. Dreyfuss (2005). "The survival of motor neurons protein determines the capacity for snRNP assembly: biochemical deficiency in spinal muscular atrophy." *Mol Cell Biol* **25**(13): 5543-5551.
- Warburg, O. and W. Christian (1941). *Biochemische Zetischrift* **310**: 384-421.
- Wassarman, K. M., F. Repoila, C. Rosenow, G. Storz and S. Gottesman (2001). "Identification of novel small RNAs using comparative genomics and microarrays." *Genes Dev* **15**(13): 1637-1651.
- Weber, G., S. Trowitzsch, B. Kastner, R. Luhrmann and M. C. Wahl (2010). "Functional organization of the Sm core in the crystal structure of human U1 snRNP." *EMBO J* **29**(24): 4172-4184.
- Wen, J., T. Arakawa and J. S. Philo (1996). "Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions." *Anal Biochem* **240**(2): 155-166.
- Wen, Y. and A. J. Shatkin (1999). "Transcription elongation factor hSPT5 stimulates mRNA capping." *Genes Dev* **13**(14): 1774-1779.
- Wernimont, A. and A. Edwards (2009). "In situ proteolysis to generate crystals for structure determination: an update." *PLoS One* **4**(4): e5094.
- Will, C. L. and R. Luhrmann (2001). "Spliceosomal UsnRNP biogenesis, structure and function." *Curr Opin Cell Biol* **13**(3): 290-301.
- Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson (2011), 'Overview of the CCP4 suite and current developments', *Acta Crystallogr D Biol Crystallogr*, **67** (Pt 4), 235-42.
- Wiseman, T., S. Williston, J. F. Brandts and L. N. Lin (1989). "Rapid measurement of binding constants and heats of binding using a new titration calorimeter." *Anal Biochem* **179**(1): 131-137.
- Wolin, S. L. and T. Cedervall (2002). "The La protein." *Annu Rev Biochem* **71**: 375-403.
- Yan, Y., S. Rubinchik, A. L. Wood, W. E. Gillanders, J. Y. Dong, D. K. Watson and D. J. Cole (2006). "Bystander effect contributes to the antitumor efficacy of CaSm antisense gene therapy in a preclinical model of advanced pancreatic cancer." *Mol Ther* **13**(2): 357-365.
- Zaric, B., M. Chami, H. Remigy, A. Engel, K. Ballmer-Hofer, F. K. Winkler, and C. Kambach (2005), 'Reconstitution of two recombinant LSm protein complexes

reveals aspects of their architecture, assembly, and function', *J Biol Chem* **280**(16): 16066-75.