

Lsm proteins as ring tectons

By Dominic Logel, BSc

A thesis submitted in partial fulfilment of the requirements for the degree of
Masters of Research (MRes)

Department of Chemistry and Biomolecular Sciences

Macquarie University

9th October 2015

Tables of Content

| | |
|---|-----------|
| Acknowledgements | i |
| Abbreviations | ii |
| Abstract | iii |
| Chapter 1: Form and architecture of Lsm proteins | |
| 1.1 Approaches for proteins in bionanotechnology | 1 |
| 1.1.2 Ring proteins as tectons | 3 |
| 1.2 Lsm Proteins | 5 |
| 1.2.1 Lsm proteins provide scaffold for RNA metabolism | 5 |
| 1.2.2 Sequence and structure of Lsm proteins | 5 |
| 1.3 Fused Lsm polyproteins as tectons | 9 |
| 1.3.1 Eukaryotic Lsm rings present multiple stable ring organisations | 9 |
| 1.4 Scope of this thesis | 19 |
| Chapter 2: Materials & Methods | |
| 2.1 Materials | 12 |
| 2.1.1 Reagents | 12 |
| 2.1.2 Bacterial strains and plasmids | 12 |
| 2.1.3 Growth media and buffers | 12 |
| 2.2 Molecular Biology Procedures | 13 |
| 2.2.1 Storage of bacterial strains | 13 |
| 2.2.2 Plasmid isolation from <i>E. coli</i> | 13 |
| 2.2.3 Preparation of MRE5 Δ Hfq- expression cells | 13 |
| 2.2.4 Bacterial transformation for protein expression | 14 |
| 2.3 Protein Expression Protocols | 14 |
| 2.3.1 Small scale expression of Lsm polyproteins | 14 |
| 2.3.2 Large scale purification of protein | 14 |
| 2.3.2.1 <i>Expression via Auto Induction</i> | 14 |
| 2.3.2.2 <i>Expression via IPTG Induction</i> | 15 |
| 2.3.2.3 <i>Cell lysis</i> | 15 |
| 2.3.2.4 <i>Immobilised metal affinity chromatography (IMAC)</i> | 15 |
| 2.3.2.5 <i>GST purification of GST-tagged proteins</i> | 16 |
| 2.3.2.6 <i>Size exclusion chromatography (SEC)</i> | 16 |
| 2.4 Protein Analysis Methods | 17 |
| 2.4.1 Protein electrophoresis | 17 |
| 2.4.2 Protein concentration | 18 |
| 2.4.3 Molecular size by SEC | 18 |
| 2.4.4 Multi-angle laser light scattering | 19 |
| 2.5 Protein crystallisation | 20 |
| 2.5.1 Sparse-matrix screening of crystalline conditions | 20 |
| 2.5.2 X-ray diffraction screening | 21 |
| 2.5.2 Crystal optimisation strategies | 22 |

Chapter 3: Characterisation of Lsm polyproteins as potential tectons

| | |
|---|----|
| 3.1 Lsm polyproteins modelled from yeast Lsm rings | 24 |
| 3.2 Preparation of recombinant Lsm polyproteins | 26 |
| 3.3 Differences between Lsm[1+4] and Lsm[4+1] | 29 |
| 3.4 Biophysical characterisation of Lsm[5+6] | 29 |

Chapter 4: Crystallisation of an engineered Lsm α tecton

| | |
|---|----|
| 4.1 An engineered form of Lsmα | 31 |
| 4.2 R65PLsmα in solution | 31 |
| 4.3 Crystallisation of Lsmα R65P | 36 |
| 4.3.1 Crystallisation | 36 |
| 4.3.2 Crystal optimisation | 39 |
| 4.3.3 Crystal harvesting and cryopreservation | 41 |

Chapter 5: Conclusion

| | |
|--|----|
| 5.1 An evaluation of different Lsm polyproteins | 43 |
| 5.2 Crystal structure of R65PLsmα | 43 |

References

Acknowledgements

The past year has been the most challenging and most rewarding year of my life with its share of up and downs. Primarily I would like to thank my supervisor, A/Prof Bridget Mabbutt for guiding me through my project and for providing me with my interest in protein engineering when she taught her third year unit on the subject. Most importantly I would like to thank her for teaching me how to write a proper thesis. Secondly, I would like to thank Dr Bhumika Shah and Francesca Manea for training me this year. This year would not have been possible without their guidance. I would like to thank the other members of my group, Heather and Jacob, for helping me out whenever I required it. I would also like to thank Dr Louise Brown for convincing me to do my masters in the first place by believing in me and persuading me that I had it in me to complete this year. I would have never believed I was capable of this without her faith.

I would like to thank Dr Juanita Phang and Dr Mohan Bhadbhade, both from the University of New South Wales, for their assistance in operating the x-ray beamline at the UNSW Department of Physics for the diffraction screening of my crystals.

I would like to give an incredibly heartfelt thanks to my three closest friends, Katie, Leah, and Keith. I would not have been able to finish this year without your support when I needed it most. My life would be that much poorer without you. Finally, I'd like to thank my fellow MRes students, Mike, Emily, and Ben, for the day beers, Mexican food, and burgers.

List of abbreviations

Styles adopted in this work are according to Instructions for Authors in the Journal of Molecular Biology.

| | |
|-------------------|--|
| A ₂₈₀ | absorbance at 280 nm |
| DMSO | dimethyl sulfoxide |
| Hfq | bacterial Lsm protein |
| hsp60 | heat shock protein 60 |
| IMAC | immobilised metal affinity chromatography |
| IPTG | isopropyl β -d-1-thiogalactopyranoside |
| JSCG | Joint Centre For Structural Genomics |
| K _{av} | size exclusion distribution coefficient |
| LB | Luria-Bertani media |
| Lsm | like-sm |
| Lsm[1+4] | polyprotein composed of fused lsm1 and lsm4 |
| Lsm[5+6] | polyprotein composed of fused lsm5 and lsm6 |
| MALLS | multi-angle laser light scattering |
| MCSG | Midwest Centre for Structural Genomics |
| MPD | 2-methyl-2, 4-pentadiol |
| OD ₆₀₀ | optical density at 600 nm |
| PAGE | polyethylene glycol gel electrophoresis |
| PDB | protein data bank |
| PEG | poly ethylene glycol |
| PEG MME | poly ethylene glycol methyl ether acrylate |
| SAXS | small angle x-ray scattering |
| SDS | sodium dodecyl sulfate |
| SEC | size exclusion chromatography |
| SP1 | stable protein 1 |
| TEMED | <i>n,n,n',n'</i> -tetramethyl-ethane-1,2-diamine |
| TRAP | <i>trp</i> RNA-binding attenuation protein |
| V _o | void volume of chromatography material |

Declaration

Where appropriate, work done in collaboration with other groups or individuals has been acknowledged. Outside of these contributions, the material in this thesis is entirely my own work, and to the best of my knowledge is original. No part of this thesis has been submitted for a higher degree to any other university or institution. I consent to my thesis being made available for photocopy and loan.

Dominic Logel
Department of Chemistry and Biomolecular Sciences
Macquarie University
Sydney, NSW 2109
Australia
October 9th 2015

Abstract

Bionanotechnology is beginning to exploit naturally occurring proteins as building blocks for creating nanoscale materials and complex devices. One example of interest is the Lsm family of proteins, which spontaneously assemble *in vivo* either as homomeric (archaea or prokarya), or heteromeric (eukarya) oligomers comprising six or seven Lsm protomers which play a role in RNA metabolism. At Macquarie, several Lsm proteins have been re-engineered to form novel artificial RNA-binding rings.

In this thesis I explore biophysical characteristics of new versions of yeast-derived tectons: utilising polyproteins Lsm[1+4] and Lsm[5+6]. For the first time, the Lsm[5+6]₄ tecton has been cleanly produced, and its stability evaluated, whilst pure Lsm[1+4] proved to be difficult to obtain.

This thesis also describes the first steps towards a crystal structure of a ring assembly formed by R65PLsm α . This mutant has been proposed as a variant of the heptameric Lsm α ₇, possibly comprising a hexameric organisation. I successfully generated many crystal forms, which led on to provide diffraction of good quality. Results from structure solving allowed definition of the correct number of components, which was not able to be determined from biophysical solution methods.

Chapter 1: Form and architecture of Lsm proteins

1.1 Approaches for proteins in bionanotechnology

Bionanotechnology today seeks to mine nature's rich source of molecular systems for useful pre-existing constructs amenable to engineering, one purpose being to create novel scaffolds or biologically responsive devices [1-4]. These may include molecular cages, filaments, and porous materials [1, 5].

Crucial to the creation of artificial bio-inspired nanostructures are the constituent parts, also known as tectons [6]. Tectons are the fundamental building blocks required to construct a desired assembly, piece by piece. Whilst nucleotides and peptides have been successfully used as tectons for nanofabrication, creating structures such as DNA origami and hydrogels respectively [7-10], there are benefits to using proteins as tectons in nanotechnology. These utilise the natural features of proteins, such as their native tertiary or quaternary architecture, and easily modified chemistry [11-13]. As protein sequences can be easily modified, changes can be made to encourage higher order assembly, enhance specific functions, and introduce novel functionalities [14-17].

New developments in molecular sciences have benefitted from both the increased affordability of synthetic biology and the growing accuracy of computational design tools for engineering protein interactions. This has led to new approaches to protein engineering increasing its speed and efficiency [6, 15, 18, 19]. As of writing, the Protein Data Bank [20] has 112,722 characterised structures available to the research community. This allows for mining of protein structures for potential starting tectons for the rational engineering of novel functional assemblies where a detailed structure is already defined [21].

Currently there are two main processes for constructing nanomaterials, bottom up and top down. For a bottom-up approach, near atomic level structures are fabricated via molecular synthesis, polymerisation, or colloidal aggregation to create precise nanoscale structures whilst sacrificing long range ordering. Recent improvements in *in silico* design tools [22-24] have led to the design of a highly symmetrical assemblies through novel interface oligomerisation [25]. An alternative top-down process might utilise lithographic techniques

to order materials on a larger scale whilst sacrificing the resolution of the bottom-up approach [26, 27].

A bridging strategy to combining the benefits of both these approaches would be able to incorporate molecular self-assembly [28], a natural process often driving protein organisation. Molecular self-assembly incorporates interaction between molecular partners through non-covalent interactions (e.g. hydrogen bonding, or van der Waals forces) [26]. By exploiting self-assembly as part of the tecton design process, native symmetry and assembly can be utilised to remove the need to manually engineer this level of complex assembly, as demonstrated in Fig 1.1 [1]. Examples of this have been able in the construction of protein assemblies where multiple interfaces, such as β strands [29], have been engineered within a protein to allow its interaction with different proteins on difference interfaces [19].

Self-assembly is the driving force behind protein quaternary structure, also known as oligomerisation, which occurs between protein-protein interfaces which are networked together via multiple weak intermolecular forces [30].

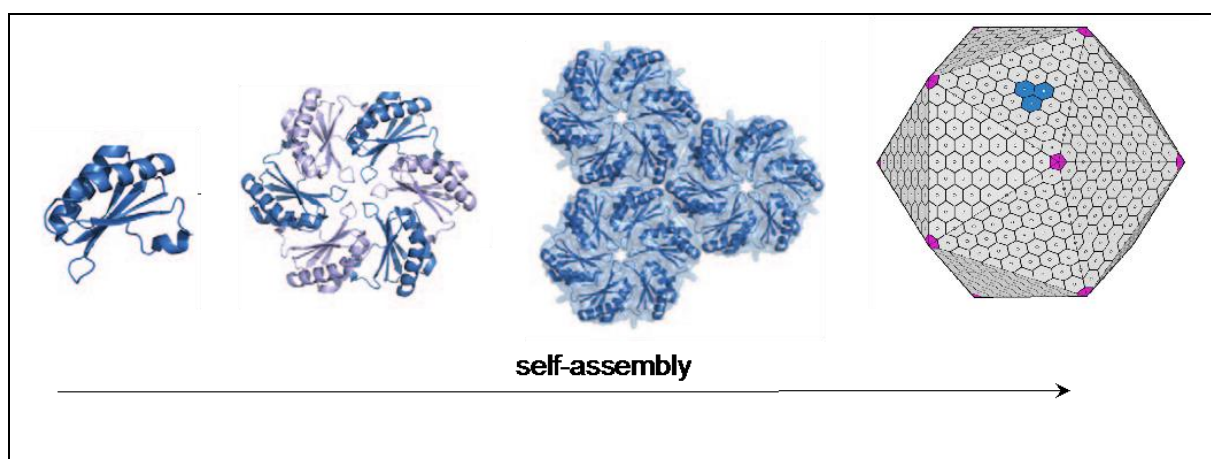


Fig 1.1| Natural quaternary organisation of Bacterial Microcompartment Organelles (BMC) monomers.

Self-assembly pathway of BMC shell protein monomers (PDB 3NWG) occurs through the corresponding interfaces self-assembling into increasingly higher order architectures to make large highly organised supramolecular structures [31]

Although no universal explanation exists for the occurrence of multimeric proteins, recent investigations have determined potential explanations. These explanations link oligomerisation to both protein stability and functionality, as oligomerisation has been seen to occur between protein faces which expose a hydrophobic core. Thus, an oligomerisation of these surfaces may act to protect the structural integrity of the protein's active site existing within the junction of two protomers, or to impart increased stability in complex with other protomers [32, 33]. The stability and activity imparted by oligomerisation is highly advantageous to protein nanotechnology as these features provide useful tectons for functionalisation and fabrication [18, 34]. An example of this was where two natural proteins, KDPGal and FkpA, were fused into a single protein which self-assembled into a large porous cube-shaped protein assembly through the original proteins native oligomerisation behaviours [29].

1.1.2 Ring proteins as tectons

Whilst success has been seen in engineering protein cages and arrays using proteins [19, 24, 35], protein of ring morphology have also been extensively investigated as tectons [6, 8, 11, 14, 15, 18, 23, 36-42]. Ring proteins inherently contain rotational symmetry resulting from repeating subunits, and four unique faces which can be differentially modified (see Fig 1.2).

An example of using ring proteins as a starting point for engineering is the work carried out by the Heddle group on the *trp* RNA-binding attenuation protein (TRAP) [41]. TRAP rings are a naturally-occurring thermostable protein found in *Bacillus subtilis* which self-assembles into an highly stable homo-undecameric ring assembly. Based on the pre-existing crystal structure for the protein, *Miranda et al.* was able to incorporate Cys residues at two locations on upper and lower ring faces. In an oxidative environment, the Cys residues formed disulphide linkages forming coaxial stacks of multiple tectons. This resulted in nanotubes measuring from several hundred nanometres to one micrometre (see Fig 1.2).

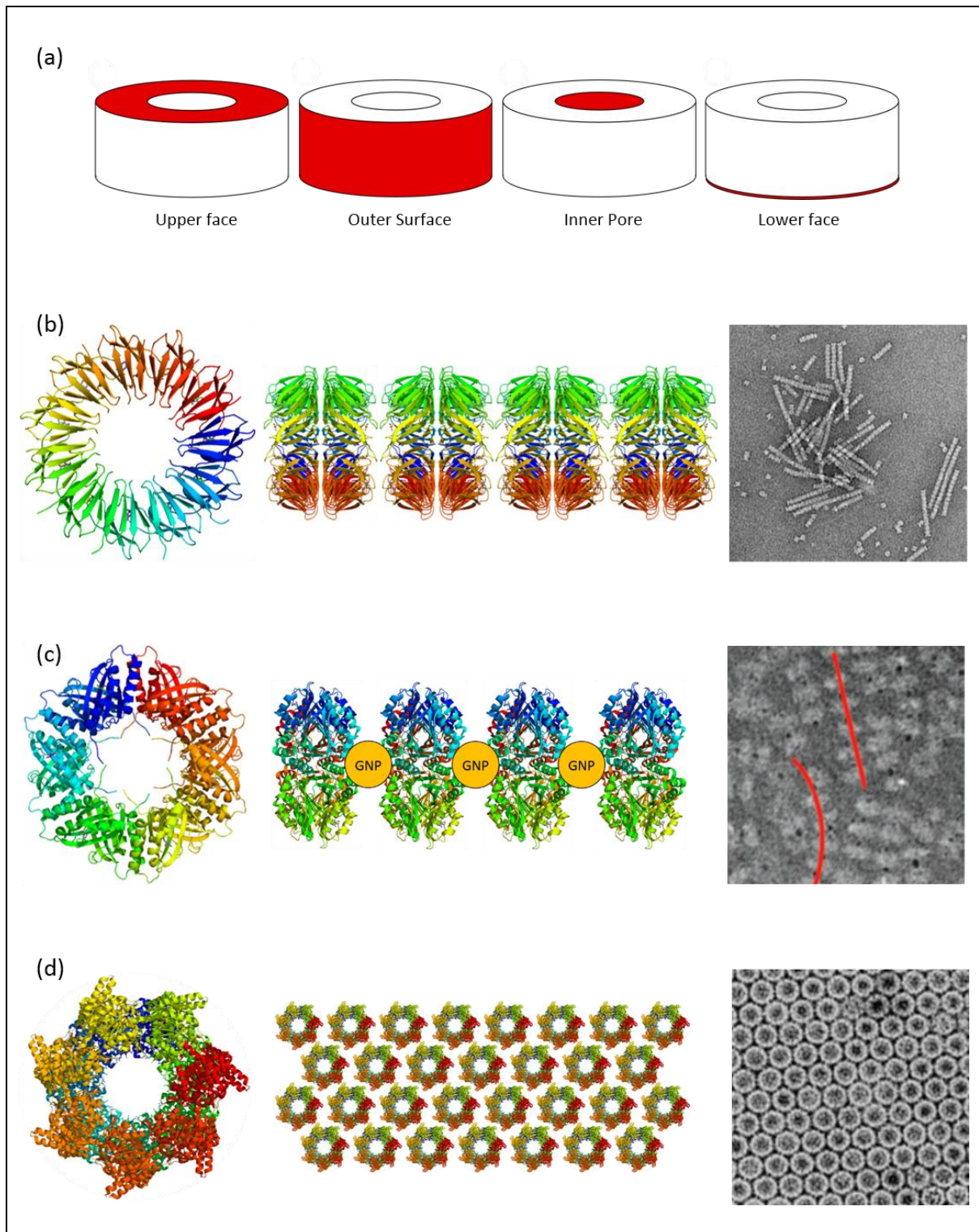


Figure 1.2| Examples of ring tecton supramolecular assemblies.

(a) Ring proteins provide four unique faces which can be differentially modified. (b) Trp RNA-binding attenuation protein (TRAP) ring upper and lower faces incorporating Cys residues lead to extended nanotubes in redox conditions [41] (PDB ID 1QAW [36]). (c) Stable protein 1 (SP1) rings incorporating His residues within the inner pore of the ring formed GNP-mediated nanowires [40]. (PDB ID: 1TR0 [37]). (d) Heat shock protein 60 (hsp60) reengineered to relocate terminal tails from the interior of the ring to the exterior, permitting the conjugation of inorganic nanomaterials [35, 43]. (PDB ID: 4AAU [44]).

Another example of utilising ring proteins as tectons was research carried out on Stable Protein 1 (SP1), a thermostable protein found in *Populus tremula*, which forms a homododecameric ring [40]. SP1 ring organisation is different to the TRAP ring tecton; as ring faces from two SP1 rings oligomerise to form a co-axially stacked double ring. SP1 was rationally designed by exploiting the structural positioning information within the available crystal structure. Residues located along the interior of the ring pore were selected for mutagenesis into histidines, creating metal binding sites, facilitating metal ion coordination within the tecton pore. When treated with gold nanoparticles (GNP) in solution, the tectons sequestered the GNP to form nanowire chains constituted of alternating SP1 and GNPs (see Fig 1.2). In this study, *Medalsy et al.* was able to easily tune the distance between the tecton and the GNP by truncating thirty amino acids from the subunit N-terminal, reducing the distances present in the protein-nanodot packing site [40].

Heat shock protein 60 (hsp60) is another commonly engineered target due to its ability to form octadecameric rings from two co-axially stacked nonameric rings [43]. The protein had been previously engineered to form filaments through apical stacking and 2D crystal arrays through equatorial interactions [42, 45]. *Paavola et al.* further engineered circular permuted forms by relocating the terminal tails of the protein from the ring interior, to the ring exterior. From this, no loss in the ability to form 2D crystal arrays was noted (see Fig 1.2). The relocation of the terminal tails permitted additional proteins, such as enhanced yellow fluorescent protein to be genetically fused to the now exposed terminal tails [43].

1.2 Lsm Proteins

1.2.1 Lsm proteins provide scaffold for RNA metabolism

The Lsm family of proteins is found across all three domains of life and plays a key role in various RNA metabolic processes [46-48]. In prokaryotes and archaea, this family is represented by a single gene oligomerising to form ring assemblies, such as [Lsm α]₇ in *Methanobacterium thermoautotrophicum* and [Hfq]₆ in *Escherichia coli*. These are depicted on Fig 1.3. In eukaryotes, multiple genes encode Lsm proteins which forms heteromeric rings, typically in heptamers (see Fig 1.3). [49, 50]. However, recent investigations expressing Lsm3 have seen the formation of octameric ring organisations, previously unseen in Lsm systems (Fig 1.3) This suggests a degree of plasticity in ring organisation and the possibility of pore size tuning without performing equatorial interfaces engineering [50]. In organisms such as

Saccharomyces cerevisiae, different types of ring assemblies have been described, incorporating seven mixed Lsm gene products named Lsm[1-7] and Lsm[2-8].

The primary biological role of Lsm rings is to sequester RNA for modification. This binding is driven by residues found in specific loops of each individual, as seen in Fig 1.3 [51]. Multiple weak intermolecular forces, such as π - π stacking, hydrogen bonds, and cation- π bonds between the nucleotide and protein sidechains provide specificity to the RNA-protein interaction [52, 53]. The tuning of the RNA binding has also been observed in studies focusing on the bacterial analogue, Hfq [54, 55]. Previous studies demonstrate that the RNA sequences which exhibit the strongest interaction with Lsm are those rich in sequential uracil bases. Indeed, in some cases, the binding of RNA may also be necessary to the formation of the quaternary Lsm ring [52, 56, 57].

1.2.2 Sequence and structure of Lsm proteins

The Lsm quaternary ring assembly is an incredibly robust structure, due to its rigid β -propeller organisation. The primary sequences of Lsm proteins from *Saccharomyces cerevisiae* and archaeal Lsm from *M. thermoautotrophicum* are aligned in Fig 1.4 demonstrating conserved regions across the Lsm proteins. This fold is imparted from a continuous anti-parallel β sheet of five strands (reminiscent of an OB fold), where the protomer N-terminal α helices overlay one another throughout the ring [49, 50, 53, 58]. The secondary structure is indicated in Fig 1.4. The protomers oligomerise with their adjacent neighbour through multiple weak interactions occurring on the β 4 strand of one protomer, and the β 5 of its neighbour [49, 50, 53]. The N-terminal α helix overlaying the adjacent protomer is a consistent feature across the ring face, allowing the naming of this ring face as the “helix face”. The predominant feature of the opposing face of the ring is a highly variable unstructured loop sequence; thus this face is known as the “loop face” (Fig 1.4) [53]. High variability is also located in the N and C terminals of the protein, for example in Lsm1 where an extended C-terminal α helix is described [49].

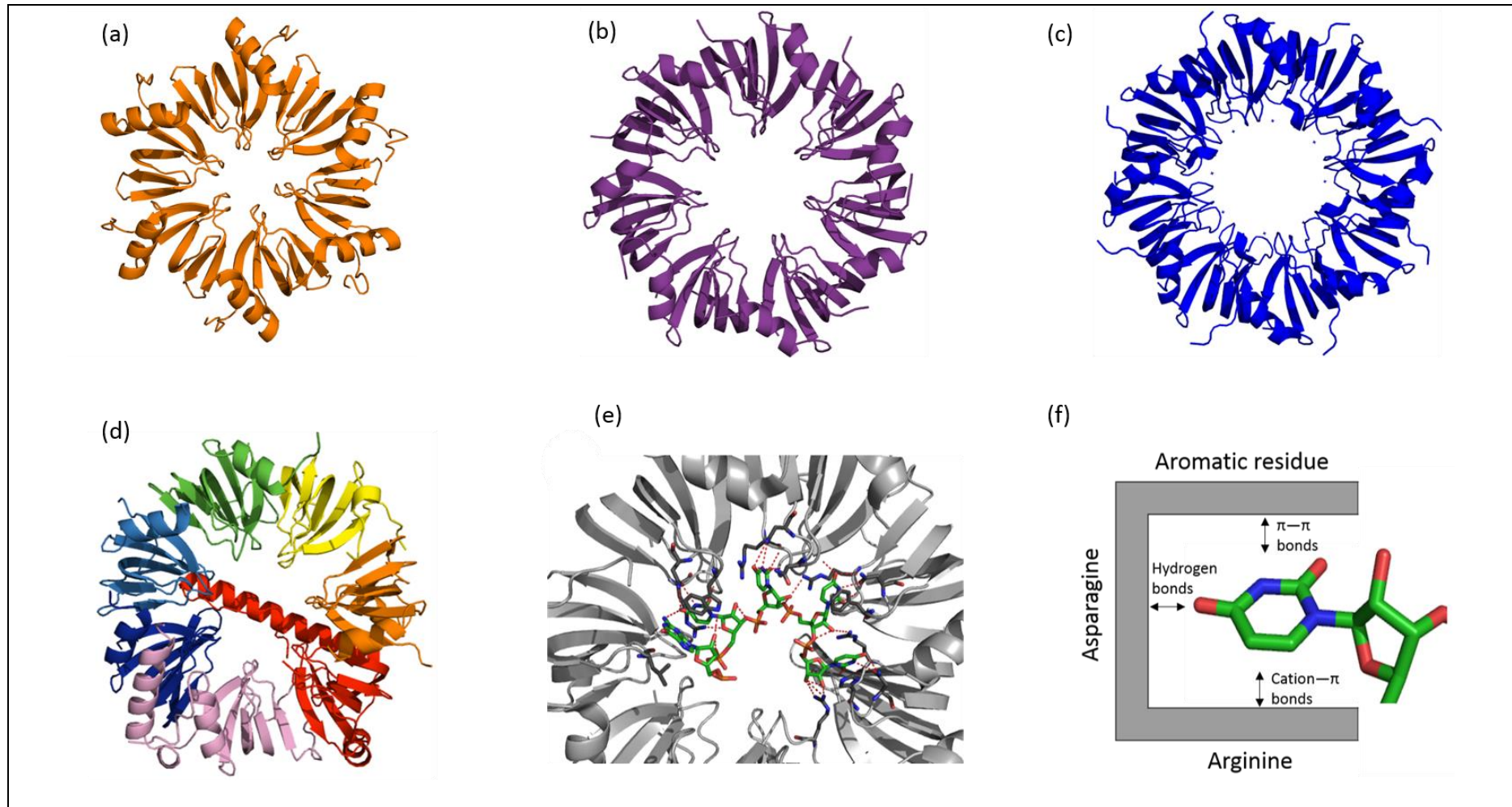


Fig 1.3| Lsm rings form diverse range of ring assemblies and present RNA binding specificity through loops internal to the ring pore.

(a) Hfq (*E. coli*) assembles as a homo-hexamer (PBD ID: 3QHS [59]). (b) Lsm α (*M. thermoautotrophicum*) assembles as a homo-heptamer (PBD ID: 1MGQ [53]). (c) Lsm[1-7] (*S. cerevisiae*) assembles as a hetero-heptamer (PBD ID: 4C92 [49]). (d) Recombinant Lsm3 (*S. cerevisiae*) assembles as a homo-octamer (PBD ID: 3BW1 [50]). (e) RNA bind to specific residues within the Lsm ring pore. (f) Schematic outlining Uracil specificity is imparted by three residues within the active site (PBD ID: 4M7A). Adapted from [60].

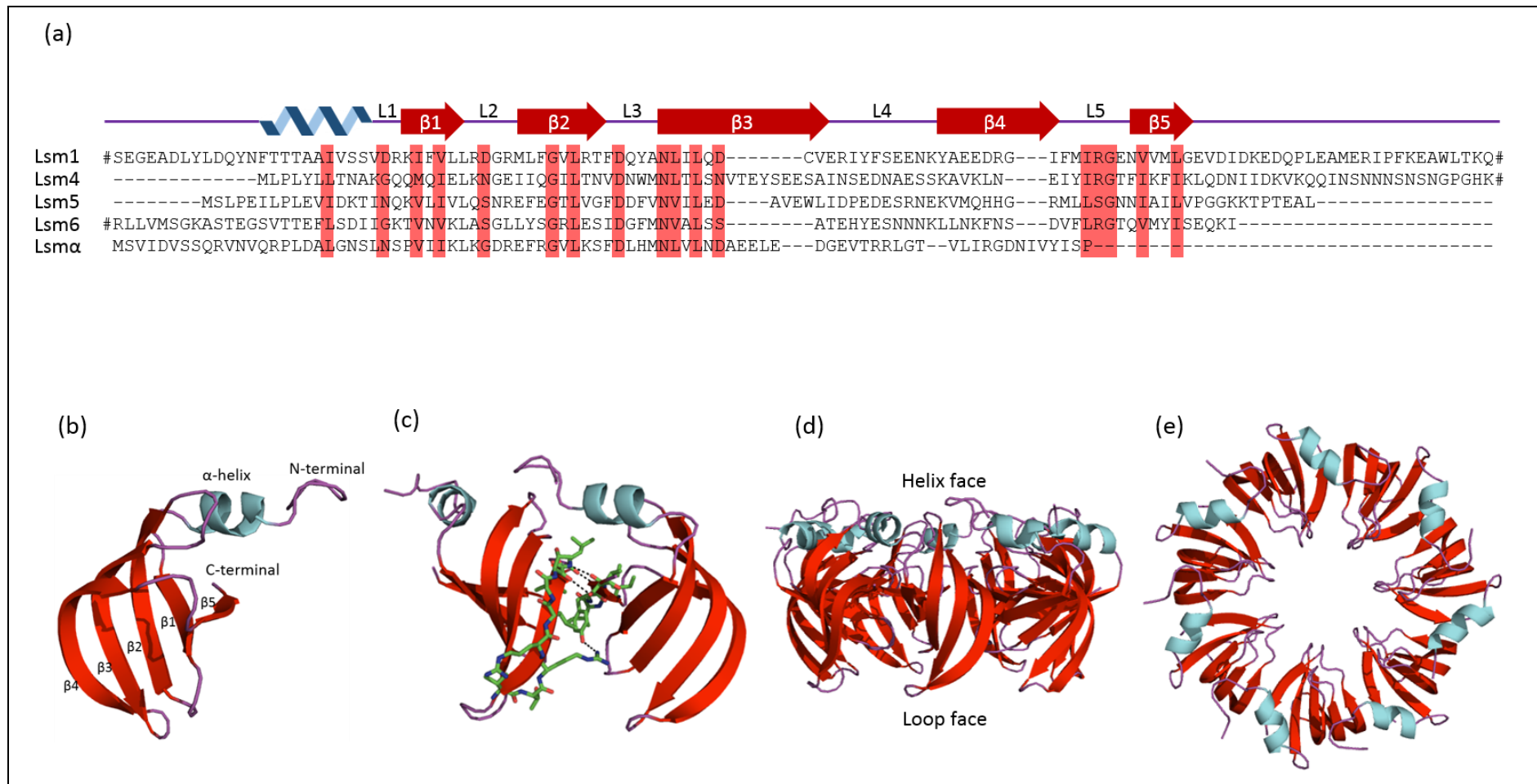


Fig 1.4| Sequence and structure of Lsm proteins.

(a) Sequence alignment of Lsm proteins from *S. cerevisiae* (Lsm1-6), and *M. thermoautotrophicum* (Lsm α). # indicates truncated sequence and conserved sequences shown in red. (b) Lsm proteins consist of an N-terminal α helix and anti-parallel β -strands. (c) Adjacent Lsm protomers interact through residues on the β_4 and β_5 protomer strands, providing high right stability. (d) Multiple protomers interact to form a ring assembly with distinct faces, the helix face, and the loop face. (e) Lsm rings show β propeller fold. PDB ID: 1MGQ [53]

1.3 Fused Lsm polyproteins as tectons

1.3.1 Eukaryotic Lsm rings present multiple stable ring organisations

A series of Lsm polyproteins were designed several years ago by Dr Meghna Sobti [61] by focusing on stable ring assemblies observed in nature as a template [62]. Dr Sobti genetically fused stable ring partners, so transforming two or more protomers into polyprotein dimer or trimer combinations. The rationale behind this technique was to force a predictable order into the formation of Lsm rings, thus manufacturing tectons with a simplified symmetry, e.g. AB_n or ABC_n . From this design principle, a suite of polyproteins were successfully created and expressed in solution: $Lsm[4+1]_4$, $Lsm[1+4]_3$, $Lsm[2+3]_4$, and $Lsm[5+6]_4$ (see Fig 1.5).

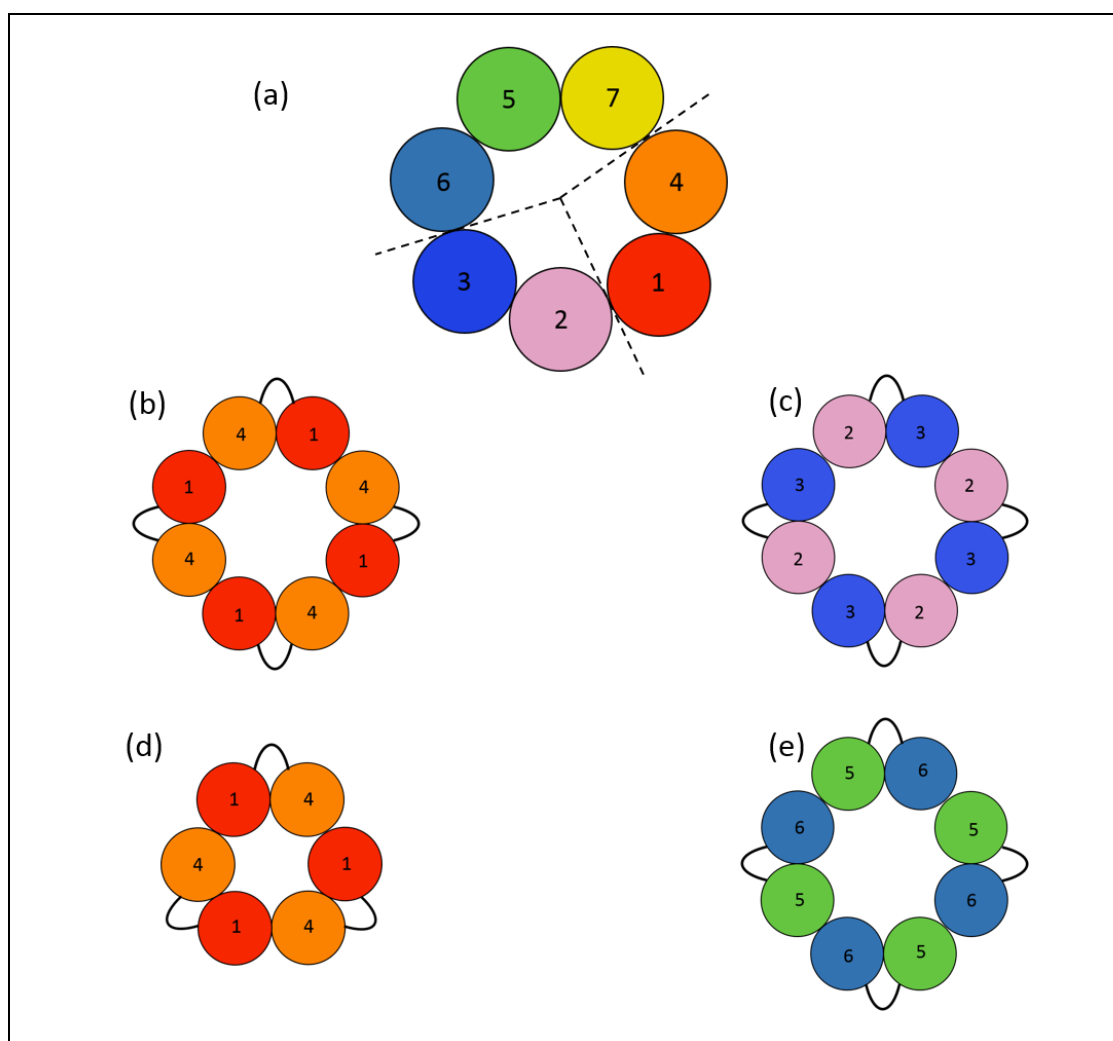


Fig 1.5 | Polyproteins were designed from adjacent pairings of natural *S. cerevisiae* Lsm[1-7] ring.

(a) Heptameric complex $Lsm[1-7]$ as organised in *S. cerevisiae* serves as a template for the design of Lsm polyprotein rings [61, 62]. (b) $Lsm[4+1]_4$ [63]. (c) $Lsm[2+3]_4$ [63]. (d) $Lsm[1+4]_3$ [61]. $Lsm[5+6]_4$ [61].

A well-characterised polyprotein, Lsm[4+1], self-assembles from repeating polyproteins to form tetrameric rings in an AB₍₄₎ pattern, as seen in Fig 1.5. Two populations are observed in solution, corresponding to Lsm[4+1]₄ and two stacked tectons, 2Lsm[4+1]₄. This tecton was further engineered to covalently bond via helix and loop face Cys residues to form nanotubes and to also form protein cages via Ni(II) sequestration by His₆ tags [63]. These examples show the potential of Lsm tectons as biologically responsive nanoparts.

1.4 Scope of this thesis

By using ring proteins as a natural building block for tecton development, many concatenation techniques can be implemented to induce the new organisations desired for bionanotechnology. This is easily done in a rational manner if structural information of the protein, such as assembly size and crystal structures, is known before engineering is carried out.

Whilst Lsm[4+1] and Lsm[2+3] showed some success as novel tectons, two polyproteins originally designed by Dr Sobti [61] have yet to be extensively characterised [61]. The first of these was Lsm[1+4] and initial characterisation determined that the polyprotein assembles into two populations in solution, Lsm[1+4]₃ and monomeric Lsm[1+4]. Another polyprotein yet to be extensively characterised is Lsm[5+6]. Initial biophysical characterisation of this assembly, performed by Dr Sobti, discovered a series of higher-order assemblies. Once isolated, the largest forms were observed to interconvert over time towards the smaller assemblies, indicating that a Lsm[5+6] tecton may be unstable [61].

Recently, as part of an internationally funded program with Dr Akshita Wason and Dr Juliet Gerrard (University of Canterbury, New Zealand), significant construction has been conducted with Lsm α from *M. thermoautotrophicum* as tectons [64]. In addition to fabricating nanotubes and protein cages, a panel of mutants designed to alter ring organisations were developed. R65PLsm α is of interest as it appeared to perturb the native intersubunit interfaces. Dr Wason's subsequent biophysical investigation was consistent with a reduced ring organisation of six protomers in comparison to the archaeal heptameric assembly.

This thesis focuses on two main goals. The first of these is to investigate Lsm polyproteins constructs, Lsm[1+4] and Lsm[5+6], to determine their feasibility as novel tectons. This will permit the expansion of the current suite of Lsm tectons, diversifying the range of nanostructures possible from Lsm materials. Significant work has already been

carried out to characterise and optimise the existing Lsm tectons by Ms Francesca Manea, therefore comparisons will be drawn between the polyproteins investigated here and previous work performed to detect novel organisations and features [63].

The second goal of this thesis is to further investigate R65PLsm α by performing mutant assembly size characterisation and initiating crystallographic studies. X-ray crystallography may be used to determine a highly accurate atomic structure for the mutant to conclusively determine the number of protomers constituting its ring and to detect any novel interfaces contained within its interior.

Chapter 2: Materials & Methods

2.1 Materials

2.1.1 Reagents

All chemical reagents used were of analytical grade or higher, and were obtained through mainstream suppliers (Astral, VWR International, Sigma Aldrich). Purified water from a MilliQ system (Millipore) was used throughout experimentation.

2.1.2 Growth media and buffers

Growth media were prepared with components listed in table 1.1 and autoclaved before use [65, 66]. Yeast extract was obtained from Thermo Scientific, NSW. Buffer was pH adjusted following probe calibration with standards (Scientifix, VIC), and if the buffer was utilised on an ÄKTA, they were filter-sterilised before use. To ensure plasmid stability, all growth media were supplemented with ampicillin and chloramphenicol (Astral Scientific, NSW). Agar plates were prepared by the addition of 1.5% bacteriological-grade agar to Luria-Berthani (LB) broth. Buffers and growth media were stored at 4°C, with the exception of SOC media and M9 salts (stored at -20°C).

2.1.3 Bacterial strains and plasmid

Two bacterial cell lines were used for plasmid propagation, Stellar (Clontech, VIC) and Rosetta (Merck, VIC). These were transformed with pET15b and pGEX-4T-2, which conferred resistances to ampicillin and chloramphenicol respectively, as a selectable marker. For protein expression two cell lines were used: BL21(DE3) pLysS (Promega, NSW) [67] for the expression of rare codons, and a *Hfq*-deficient MRE5 strain (gifted by Jean Beggs, University of Edinburgh). Lsm polyprotein genes were previously prepared by Dr Meghna Sobti (Macquarie University) [61], and the Lsm α mutant provided by Dr Akshita Wason (University of Canterbury) [64].

Table 1.1 | Composition of growth media^a

| Buffer or media type ^b | Components |
|-----------------------------------|---|
| M9 Salts | KH ₂ PO ₄ (15 g), Na ₂ HPO ₄ •7H ₂ O (64 g), NaCl (2.5 g), NH ₄ Cl (5 g) |
| 50x5052 | glycerol (250 g), glucose (25 g), α-lactose (100 g) |
| 20xNPS | (NH ₄) ₂ SO ₄ (66 g), KH ₂ PO ₄ (126 g), Na ₂ HPO ₄ (142 g), pH 6.5 |
| SOC | tryptone (20 g), yeast extract (5 g), NaCl (0.58 g), KCl (0.18 g), MgCl ₂ (0.95 g), MgSO ₄ , (1.20 g) glucose (20% w/v). |
| LB | tryptone (10 g), yeast extract (5 g), NaCl (5 g) |
| ZY | tryptone (10 g), yeast extract (5 g) |
| ZYP-rich | MgSO ₄ (1 ml), 50x5052 media (20 ml), 20x NPS media (50 ml), ZY media (925 ml) |

^a quantities are specified per litre^b prepared as described in Sambrook and Studier [65, 66]

2.2 Molecular Biology Procedures

2.2.1 Storage of bacterial strains

Glycerol stocks of *E. coli* plasmids were shaken in LB media (37 °C, 14 h). Cultures were centrifuged (2200 *g*, 4 °C, 10 min) and the pellet resuspended in M9 Salts (750 µl) and 50% glycerol (750 µl), and stored at -80 °C [65].

2.2.2 Plasmid isolation from *E. coli*

Plasmids were isolated through use of a commercial isolation kit (QIAprep Spin Miniprep Kit, Qiagen) according to the manufacturer's instructions. Isolated plasmids were collected in Tris-HCl buffer, pH 8.0 (30 µl, 10 mM) and stored at -20 °C.

2.2.3 Preparation of MRE5Δ*Hfq*- Expression Cells

Competent MRE5Δ*Hfq*- *E. coli* cells were prepared following the protocol of Inoue *et al.* [68]. MRE5 cells were grown overnight on non-selective LB agar plates, and single colonies used to inoculate a starter culture (5 ml LB media, 20 µM MgSO₄). The starter culture was grown overnight (37 °C) and used to inoculate large scale LB broth (500 ml LB media, 40 µM MgSO₄) in a 2 l baffled flask. This was shaken (25 °C) until an optical density at 600 nm (OD₆₀₀) reading of 0.4-0.6 was achieved. Cells were chilled on ice, centrifuged (3000 *g*, 4 °C, 10 min), and the pellet washed in ice-cold TB buffer (80 ml) and again centrifuged. Recovered cells

were resuspended in TB buffer (20 ml) with DMSO (1.5 ml). The material was chilled (0°C, 10 min) and snap frozen in liquid N₂ in 200 µl aliquots for storage (-80 °C).

2.2.4 Bacterial transformation for protein expression

For Lsm polyproteins, pET15b encoding AmpR, and pRARE encoding tRNAs for rare codons and CAM resistance, were co-transformed into MRE5Δ*Hfq*- cells. For generation of Lsmα, pGEX-4T-2 encoding AmpR was transformed into BL21 (DE3) pLysS cells.

Plasmid DNA (2 µl) was mixed with thawed competent cells (50 µl) and chilled on ice (30 min). The cells were heat shocked to 42 °C (40 s) and chilled on ice (2 min). SOC media (450 µl) preheated to 37°C was added and the cells shaken. Transformants were streaked onto pre-warmed plates of selective LB. Successfully colonies were re-streaked on identical antibiotic selective plates to confirm antibiotic selection.

2.3 Protein Expression Protocols

2.3.1 Small-scale expression of Lsm polyproteins

Transformed MRE5Δ*Hfq*- cells in ZYP-rich media (5 ml) were grown overnight (25 °C). Recovered cell pellet (14,000 *g*, 1 min) was snap frozen, then thawed at room temperature. The pellet was resuspended in a Tris cell lysis buffer (500 µl), containing Tris-HCl (20 mM) pH7.5, NaCl (150 mM), Na₂EDTA (1 mM), EGTA (1 mM), Triton-X 100 (0.1% v/v), Na₄P₂O₇ (2.5 mM), β-glycerophosphate (1 mM), Na₃VO₂ (1 mM), leupeptin (1 µg/ml), and rotated at 4 °C for 1.5 h and the remaining lysate centrifuged (14,000 *g*, 15 min). Aliquots (25 µl) were taken throughout procedure for visualisation via sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

2.3.2 Large-scale purification of recombinant protein

2.3.2.1 Auto induction of recombinant expression

A starter culture (5 ml LB media) containing ampicillin (100 mg/ml), and chloramphenicol (25 mg/ml) was used to inoculate ZYP-rich media (500 ml) (table 1.1) and grown at 25 °C (24 h). Pellets were recovered by centrifugation (4000 *g*, 4 °C, 20 min). The pellets were pooled and resuspended in Tris (20 mM, pH 8.0) buffer, with 400 mM NaCl, (40 ml) and frozen for storage (-80 °C).

2.3.2.2 Expression via IPTG Induction

An LB starter culture (5 ml) containing ampicillin (100 mg/ml) was used to inoculate LB media (500 ml) and grown at 37 °C until OD₆₀₀ reading of 0.4-0.6 (2-3 h). Cultures were inoculated with isopropyl β-d-1-thiogalactopyranoside (IPTG) (100 μl, 27% w/v) and grown overnight (25 °C) before centrifugation (4000 *g*, 4°C, 20 min). The pellets pooled and resuspended in PBS at pH 7.5 (40 ml) and stored in at -80 °C.

2.3.2.3 Cell Lysis

The pellets were thawed and a lysis solution containing protease inhibitor (500 μl), lysozyme (500 μl), RNase (80 μl), and DNase (80 μl) was added. Lysis was achieved with sonication (S-2500 Branson Digital Sonifier, 70 s, 10 s on/off cycles, 60% amplitude). The lysate was centrifuged (11,000 *g*, 4 °C, 1 h) before being passed through a 0.22 μm filter tip.

2.3.2.4 Immobilised metal affinity chromatography (IMAC)

A common protein purification technique exploits the metal binding attributes of His-rich regions of protein through the incorporation of a His6 affinity tag to a protein terminus. IMAC utilises this chemistry by coordinating the His6 affinity tags with a bound metal ion, commonly Ni(II). This tightly binds the target protein until the binding reactions is outcompeted through free imidazole molecules [69].

Cell lysate was passed through a pre-packed Ni-Sepharose column (His-Trap, GE Lifesciences) equilibrated with Tris (20 mM) buffer with 400 mM NaCl and 50 mM imidazole on with a peristaltic pump operating at room temperature. Due to a maximum column absorption efficiency of 40 mg/ml, the cell lysate was set to recycle in an attempt to saturate the column with target protein. Once the matrix was loaded with absorbed protein, elution was carried out under high pressure at 1 ml/min on an ÄKTA Explorer (GE Lifesciences) and the reaction monitored via protein absorbance at 280 nm (A₂₈₀). Tris buffers with 50 mM and 500 mM imidazole were used and set to wash the IMAC column for 30 min at 50 mM, to remove non-specifically bound protein, until a single step increase of imidazole of 500 mM was used to outcompete the absorbed His-rich protein which eluted off the column. Typical profile is shown in Fig 3.1

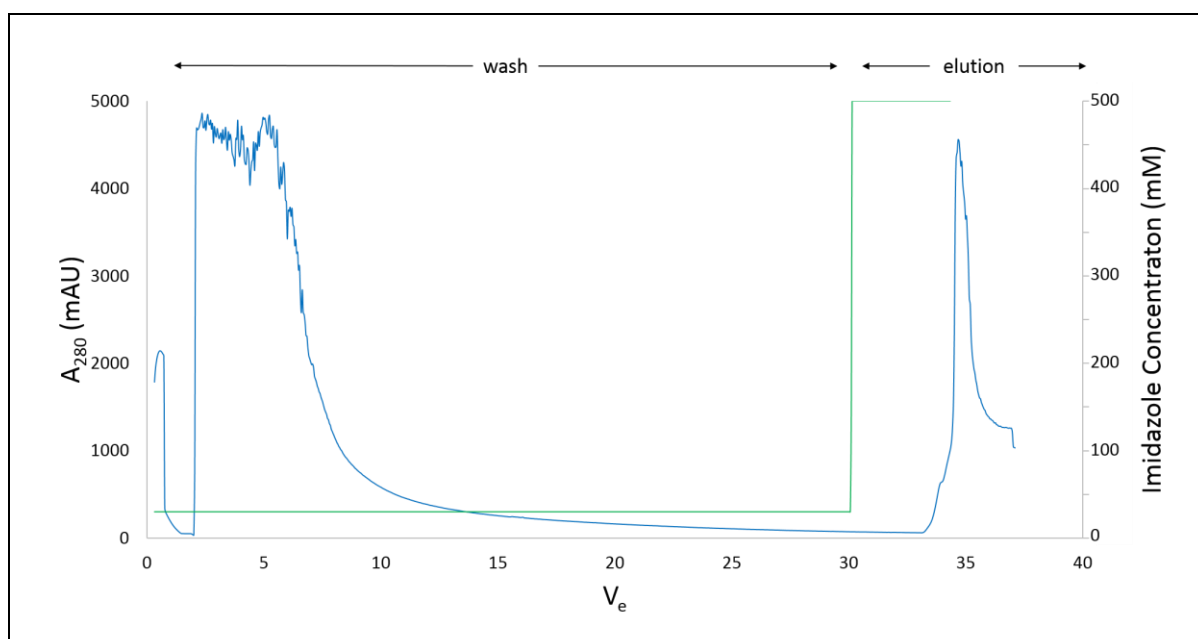


Fig 3.1| IMAC purification typically conducted in this work

Chromatography progress of protein is monitored by A_{280} (blue) and concentration of imidazole (green). The wash step uses 50 mM imidazole to remove non-specifically bound proteins and the elution step increases the concentration to 500 mM to outcompete and elute the target protein. All steps carried out in Tris (20 mM) buffer, pH 8.0 with 400 NaCl at 1 ml/min

2.3.2.5 Purification of GST-tagged proteins

Filtered lysate was applied to Glutathione Sepharose beads (GS4B), prepared according to the manufacturer's instructions, (GE Lifesciences), and gently agitated (4 °C, 1 h) to allow absorption. The slurry (3 ml) was transferred to a gravity-flow casing (1.5 ml) and washed with 23 ml of PBS, pH 7.5. The beads were washed with 3 ml of PBS, pH 7.5 with 0.1% Triton X 100 and PBS, pH 7.5 with excess NaCl (1 M) each. This was followed with a 7.5 ml PBS wash. The mixture was treated with 9% v/v thrombin (1000 U diluted in 1 ml dH₂O, Sigma Aldrich) in PBS solution and left to incubate (room temperature, 20 h). The beads were washed with 15 ml PBS, pH 7.5 and protein was collected.

2.3.2.6 Size Exclusion Chromatography (SEC)

Protein samples (5 ml) from affinity chromatography procedures were injected onto a Superdex 200 matrix on a HiLoad preparative column 16/600 (GE Lifesciences) operating on an AKTÄ Explorer system (GE Healthcare) at 1 ml/min. Superdex 200 matrix was used to assess oligomeric profile of Lsm proteins to allow for separation of different assembly populations, as previously characterised Lsm proteins were observed to assemble between 20 kDa and 200 kDa, the range effectively covered by this matrix. Protein was collected in 0.5 ml factions and

concentrated in a membrane centrifuge filtration unit (Amicon) before being stored in at -80 °C.

2.4 Protein Analysis Methods

2.4.1 Protein electrophoresis

SDS-PAGE is an analytical technique which denatures proteins in the presence of the anionic SDS detergent. This allows for the separation of multimeric protein complexes into their monomeric forms, permitting the visualisation of the relative molecular weight of each component [70]. Relative molecular weights are determined by running the SDS coated proteins in a polyacrylamide gel, where larger molecular weight proteins will pass through the pores at a reduced speed compared to lower molecular weight proteins. With Lsm polyprotein assemblies, this usually results in a profile displaying a single protein band at approximately 20 kDa.

Protein samples were visualised by SDS-PAGE (15% separating and 5% stacking gels) and molecular sizes compared with commercial protein ladder (Novex, Life-technologies). Buffers and solutions used are listing in Table 1.2. Protein samples were mixed with running buffer (1:1 ratio) and then with 2x loading dye (1:1 ratio) and boiled (98 °C, 20 min) on a PCR thermocycler (Bio-Rad). Electrophoresis was carried out at 150 V (10 min), then 100 V until completion in Tris/glycine running buffer (see table 1.2). Gels were fixed, and stained, for 10 min each and de-stained until gel background was clear.

Table 1.2: Buffers and solutions for SDS page

| Buffer | Composition |
|----------------------|--|
| Running Buffer | Tris (25 mM), glycine (250 mM), SDS (10% w/v) |
| 2x loading dye | SDS (4% w/v), glycerol (20% v/v), DTT (200 mM), Tris (100 mM) pH 6.8, bromophenol blue (0.2% w/v) |
| Fixing solution | ethanol (50% v/v), acetic acid (10% v/v) |
| Staining solution | Coomassie Brilliant Blue Dye (BDH Laboratory Supplies) (0.25% w/v), ethanol (10% v/v), acetic acid (10% v/v) |
| De-staining solution | acetic acid (10% v/v) |

2.4.2 Protein concentration

To determine accurate concentration, purified samples were diluted appropriately into standard buffers and the A_{280} measured. Known protein characteristics, generated from the ExPasy ProtParam webtool [71] which provided extinction coefficients for Lsm[1+4] (17,420), Lsm[5+6] (9970), and R65PLsm α (1490), were used to determine the protein concentration using the Beer-Lambert Law.

2.4.3 Molecular size by SEC

A 10/300 GL Superdex 200 column operating at 1 ml/min in Tris (20 mM) buffer with 400 mM NaCl was calibrated with known protein standards (GE-Healthcare): thyroglobin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), RNase A (13.7 kDa), aprotin (6.5 kDa). Blue dextran was used to determine the void volume of the column. The elution volumes (V_e) of standards were used to generate partition coefficients (K_{av}):

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

Where V_e represents the elution volume, V_o represents the void volume, and V_t represents the total column volume, 24 ml.

The linear relationship between the K_{av} and logMW values generated a straight line equation of $y=3.30x+3.22$ with an R^2 of 0.9923 (Fig 3.2). This allows the determination of unknown masses from V_e values on this same matrix.

Purified Lsm proteins underwent analytical SEC on a Superdex 200 10/300 GL column equilibrated with either Tris or PBS buffer. The V_e of the subsequent peak was converted into a K_{av} and molecular weight estimated. Samples were collected and rerun to determine oligomeric stability.

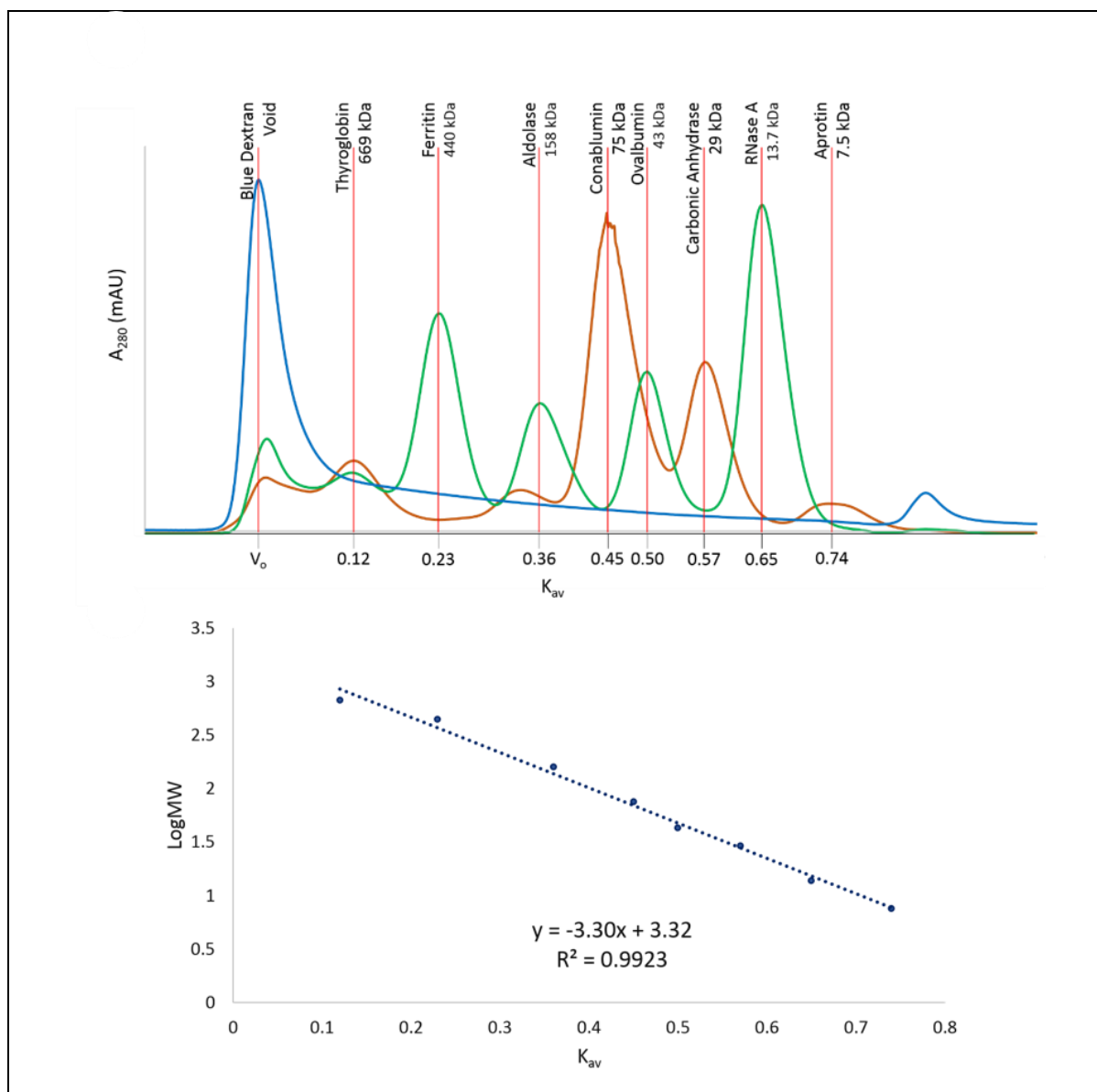


Fig 1.2| Calibration of Superdex 200

Overlaid SEC traces of the used calibration protein standards showing V_e to generate the K_{av} on the 24 ml column. Calibration of logMW with K_{av} . The equation generated was $y = -3.30x + 3.32$

2.3.4 Multi-angle laser light scattering

Multi-angle laser light scattering (MALLS) [72] can yield highly accurate mass estimation of proteins in solution independent of the protein shape by measuring the size of particles based on its light scattering properties. Molecular weight determination is dependent on Rayleigh scattering:

$$\Delta LS (I_{\theta} / I_0)_{\text{solution}} - (I_{\theta} / I_0)_{\text{buffer}} = K (dn / dc)^2 M_w \cdot C$$

Where I_{θ} / I_0 = ratio of intensities of scattered light at angle θ , K = calibration constant, C = protein concentration (g/ml), M_w = molecular mass of protein, and dn/dc = refractive index increment

Protein concentration can be accurately determined from light scattering and refractive index detectors, this allows for an accurate mass estimation to be determined, assuming the refractive index values are accurate. An optimised dn/dc value for Lsm proteins of 0.165 was previously determined (Moll, 2011) and was employed for all mass estimations in this work.

Isolated protein species were re-chromatographed on the Superdex 200 10/300 GL column in series with a triple-static light scattering (MiniDAWN TREOS, Wyatt) and refractive index detector (RID-10A, Shimadzu). All columns and detectors were pre-equilibrated with either Tris (20 mM) buffer, pH 8.0 with 400 mM NaCl, or PBS buffer, pH 7.5 prior to run.

2.5 Protein crystallisation

2.5.1 Sparse-matrix screening of crystallisation conditions

The aim of crystallisation is to slowly reduce the solubility of a specific target protein to encourage crystallogenesis within supersaturated conditions (Fig 1.3) [73]. Successful crystallogenesis and optimisation is dependant on multiple factors, including protein purity, concentration, crystallisation additives (salt, precipitant), pH, and physical parameters (drop size, temperature, diffusion format, and protein:precipitant ratios) [74, 75].

In this work, preparations of R65PLsm α were subjected to crystallisation screens. These utilised protein at 10 mg/ml with and without RNA (U₄CGU₄) in a 1:1 ratio. This was carried out using three screening kits for the RNA free preparations; MCSG1, MCSG2, both provided by the Midwest Centre for Structural Genomics (MSCG), and JCSG+ provided by the Joint Centre for Structural Genomics (JCSG); and two screening kits with RNA preparations; MCSG1, and JCSG+. This provided 288 conditions and 192 conditions for the crystallisation, respectively, to occur. Initially, conditions were screened via vapour diffusion in 96 well plates, using a sitting drop method, via the automated Phoenix Liquid Handling System (Art Robbins Instruments), with a 0.5 μ l drop protein:precipitant ratio of 1:1 and a 50 μ l mother liquor reservoir. The plates were sealed and monitored regularly over 3 months for crystallogenesis. Heatmaps for the screening plates were developed using a scoring system to assess the level of crystallogenesis noted. A score of 1 corresponds to light precipitate, 2 to heavy precipitate, 3 to phase separation of the protein and mother liquor, 4 to crystalline material or micro-crystals, and 5 to crystals large enough to harvest.

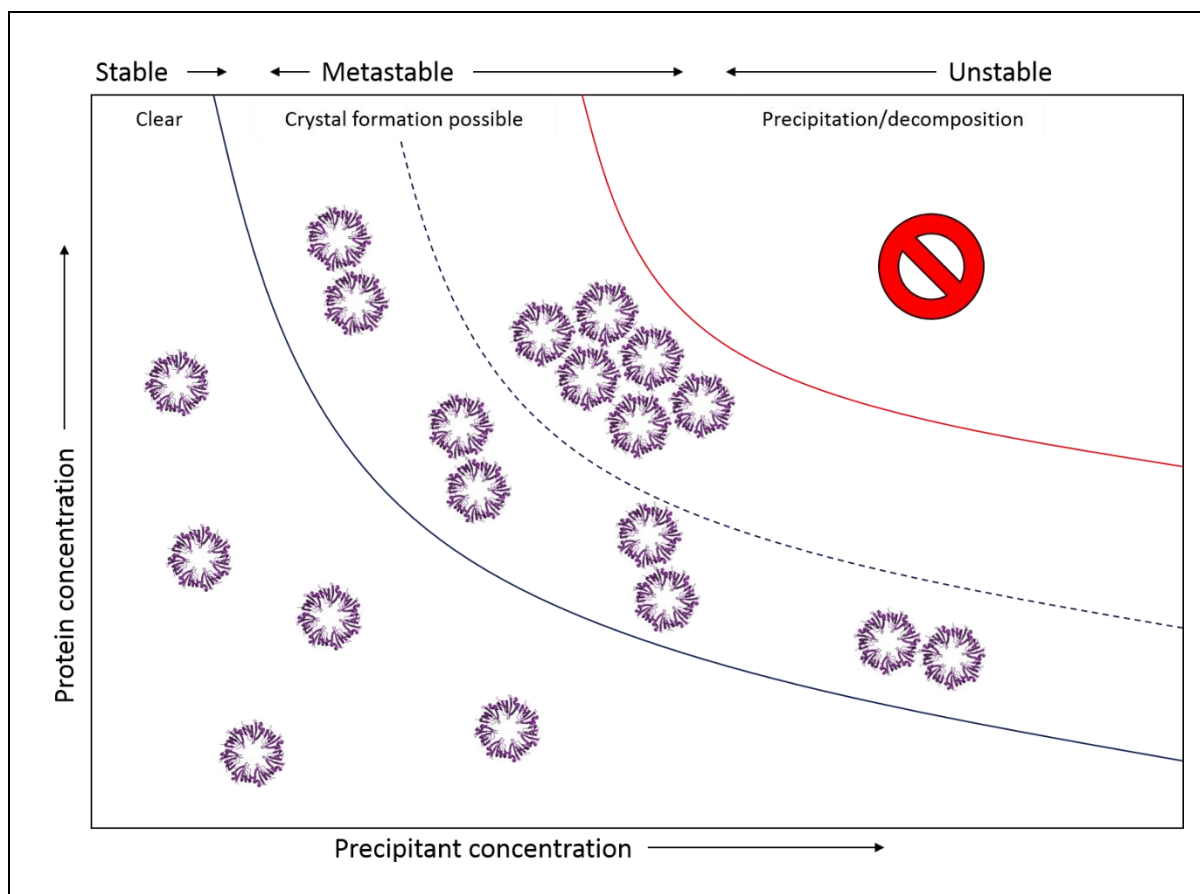


Fig 1.3 | Nucleation zones in protein crystallography

The solution will remain clean until reaching metastability where either heterologous nucleation (microcrystals) or homologous nucleation (stable crystallisation nuclei) occurs. If the solution exceeds the solubility of the protein, then it will precipitate out of solution and degrade. Adapted from [76].

2.5.2 X-ray diffraction screening

Crystals obtained for [R65PLsm α]₆ were X-ray screened at the University of New South Wales (UNSW) under the supervision of Dr Juanita Phang (UNSW), and Dr Mohan Bhadbhade (UNSW) on the in-house Rigaku Ru-200 (Rigaku) rotating anode generator in line with a Mar345 image plate. Crystals were harvested by Dr Bhumika Shah using nylon microfiber loops (Hampton Research and Mitegen) ranging from 0.025-1 mm diameter and flash cooled in liquid N₂. Dr Shah mounted these crystals in a cryostream on the beamline at 100 °K. Diffraction images at Φ -angles of 0° and 90° with exposure times of 600 s were recorded and the crystals assessed on their identity, resolution, ice ring content, and intensity.

2.5.2 Crystal optimisation strategies

Crystals exhibiting high resolution from x-ray screening were selected for further optimisation using a common grid screening methodology [77] and varying protein:precipitant ratios. A 24 well format was implemented using both sitting drop and hanging drop vapour diffusion methods with a mother liquor reservoir of 500 μ l, this was carried out at room temperature and the plates were regularly monitored over 1 month. Optimisation of crystallisation conditions focused on two factors, salt molarity and solution pH. Protein:precipitant ratios were 1:1, 2:1, 3:1, and 1:2, in addition to this, the size of the droplet was altered and ranged between 1-4 μ l. The matrixes implemented are shown in table 1.3 and 1.4. Crystal hits were harvested and stored in liquid nitrogen by Dr Shah for future data collection.

Table 1.3/ Crystal optimisation matrix 1 investigating protein: precipitant ratios.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|------------------------|------------------------|----------------------|----------------------|------------------|------------------|
| A | prot 0.5 M.L. 0.5 | prot 0.5 M.L. 0.5 | prot 1 M.L. 1 | prot 1 M.L. 1 | prot 2 M.L. 2 | prot 2 M.L. 2 |
| B | prot 0.5 M.L. 0.25 | prot 0.5 M.L. 0.25 | prot 1 M.L. 0.5 | prot 1 M.L. 0.5 | prot 2 M.L. 1 | prot 2 M.L. 1 |
| C | prot 0.75 M.L. 0.25 | prot 0.75 M.L. 0.25 | prot 1.5 M.L. 0.5 | prot 1.5 M.L. 0.5 | prot 3 M.L. 1 | prot 3 M.L. 1 |
| D | prot 0.25 M.L. 0.5 | prot 0.25 M.L. 0.5 | prot 0.5 M.L. 1 | prot 0.3 M.L. 1 | prot 1 M.L. 2 | prot 1 M.L. 2 |

Prot = protein
M.L. = mother liquor
All values in µl

Table 1.4/ Crystal optimisation matrix 2 investigating salt molarity and solution pH

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|--------------------------------------|--|---|---|--|--------------------------------------|
| A | S1 500 S2 0 S3 0 S4 0 | S1 400 S2 100 S3 0 S4 0 | S1 300 S2 200 S3 0 S4 0 | S1 200 S2 300 S3 0 S4 0 | S1 100 S2 400 S3 0 S4 0 | S1 0 S2 500 S3 0 S4 0 |
| B | S1 333.5 S2 0 S3 166.5 S4 0 | S1 266.5 S2 66.5 S3 133.5 S4 33.5 | S1 200 S2 133.5 S3 100 S4 66.5 | S1 133.5 S2 200 S3 66.5 S4 100 | S1 66.6 S2 266.5 S3 33.5 S4 133.5 | S1 0 S2 333.5 S3 0 S4 166.5 |
| C | S1 166.5 S2 0 S3 333.5 S4 0 | S1 133.5 S2 33.5 S3 266.5 S4 66.5 | S1 100 S2 66.5 S3 200 S4 133.5 | S1 66.5 S2 100 S3 133.5 S4 200 | S1 33.5 S2 133.5 S3 66.5 S4 266.5 | S1 0 S2 166.5 S3 0 S4 333.5 |
| D | S1 0 S2 0 S3 500 S4 0 | S1 0 S2 0 S3 400 S4 100 | S1 0 S2 0 S3 300 S4 200 | S1 0 S2 0 S3 200 S4 300 | S1 0 S2 0 S3 100 S4 400 | S1 0 S2 0 S3 0 S4 500 |

All values in µl

Chapter 3: Characterisation of Lsm polyproteins as potential tectons.

3.1 Lsm polyproteins modelled from yeast Lsm rings

Previously [52], a series of novel polyproteins were constructed from concatenated pairs of Lsm subunits. Within each polyprotein, Lsm elements were paired according to their adjacent subunit in the mixed Lsm[1-7] ring assembly known to occur in *S. cerevisiae* (section 1.3.2). Characterisation on the polyproteins, Lsm[4+1] and Lsm[2+3], as well as on the supramolecular assemblies they form determined the presence of stable ring organisations amenable as tectons. [63]. These rings have been explored and have been highly functional as tectons to form supramolecular assemblies, such as nanotubes and protein clusters. In this chapter, I investigate two less-studied polyproteins, Lsm[1+4] and Lsm[5+6], to determine their stability and biophysical characteristics in solution.

In Fig 3.1, I outline the full composition of the two investigated polyproteins, and contrast them with the well-characterised polyprotein, Lsm[4+1] by Dr Sobti (Protein Structure Group, Macquarie University). All three polyproteins are prefixed by a His₆ affinity tag, shown in black in Fig 3.1. Linkers composed of native sequences were utilised, rather than disordered linkers constituted from glycine and serine residues. Previous work has show that this approach produced more stable ring assemblies [78]. For the polyprotein Lsm[5+6], initial biophysical characterisation determined four distinct assemblies when separated on SEC: corresponding to Lsm[5+6]₂, as well as Lsm[5+6]₄, and its paired form 2Lsm[5+6]₄, and species larger than 600 kDa.

Lsm[1+4] was designed to be structurally similar to Lsm[4+1] but with three two key differences. This was inverting the order of the subunits thus relocating the His₆ affinity tag onto Lsm1 subunit instead of Lsm4, and the length of the natural linker sequence between the two subunits within the polyprotein. Whereas the Lsm[4+1] design incorporated a natural fourteen amino acid linker, the Lsm[1+4] design only incorporated a natural nine amino acid linker constructed of seven residues from the C-terminal of Lsm1 and two residues from the N-terminal of Lsm4.

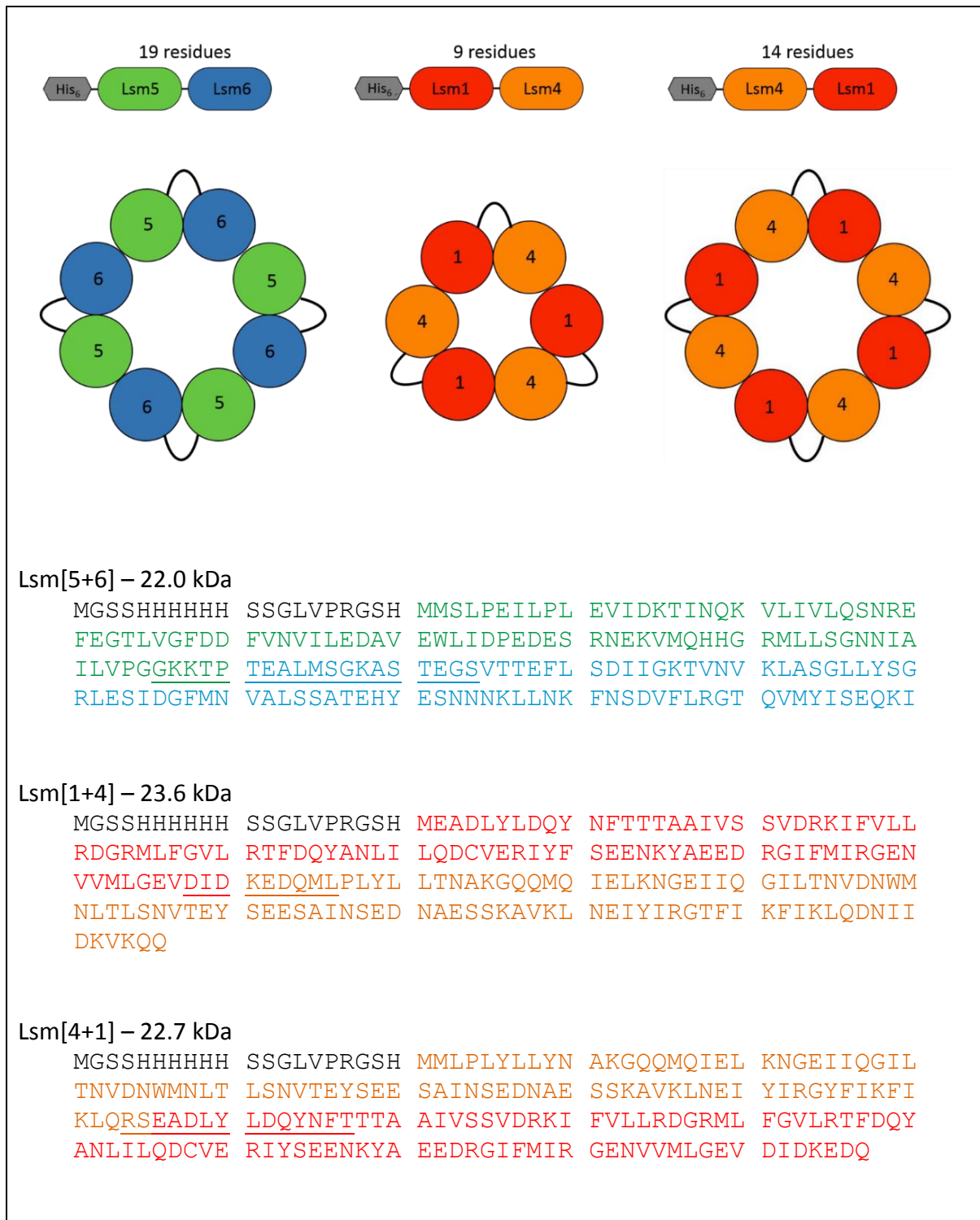


Fig 3.1 | Proposed ring organisation and sequence of Lsm polypeptides

Proposed ring assemblies for Lsm[5+6]₄ and Lsm[1+4]₃ as determined by Dr Meghna Sobti [61] in comparison to Lsm[4+1]₄, characterised by Francesca Manea [63]. The primary sequence for each polypeptide protomer is coloured separately. N-terminal His₆ sequences are shown in black, and the linker connecting each Lsm protomer underlined.

Initial investigation of Lsm[1+4] revealed that this induced a size assembly difference with the assemblies observed corresponding to Lsm[1+4]₃, differing from Lsm[4+1]₄.

For polyprotein Lsm[1+4], Dr Sobti found two protein populations corresponding Lsm[1+4]₃, and Lsm[1+4]. This suggest the predominant species was monomeric, which has not been previously seen with Lsm polyproteins. When analysed by Dr Sobti, no interconversion between the monomer and trimer populations was observed suggesting it was unlikely that the monomer population resulted from decaying Lsm[1+4]₃ rings. This polyprotein assembly profile differed from what was observed for the synonymous variant, Lsm[4+1] which was observed to form highly stable tetrameric rings. The key differences between these two assemblies are detailed in Fig 3.1. The two assemblies are effectively identical apart from the rearrangement of the original subunits. As the subunits are now reversed, this changes the components in the C-terminal to N-terminal fusion linker.

3.2 Preparation of recombinant Lsm polyproteins

Recombinant production of the Lsm polyproteins employed an autoinduction (AI) method followed by IMAC purification (see sections 2.3.3.1 and 2.3.3.4). This method was employed as these constructs contained an N-terminal His₆ affinity tag and can be successfully purified via IMAC procedures [61, 63, 64]. Protein samples obtained were assessed by SDS-PAGE, as shown in Fig 3.2.

For both Lsm[1+4] and Lsm[5+6], both productions displayed bands at the predicted size, 23.6 kDa and 22.0 kDa respectively, in both the soluble and protein eluted from IMAC fractions, suggesting that the protein was both soluble and bound to the affinity matrix. This is further supported as the insoluble fractions did not appear to contain recombinant protein at the predicted 20-25 kDa size suggesting that the two polyproteins were soluble in solution.

In order to polish the separation of different quaternary assemblies in solution [50, 61, 63, 64], a preparative SEC step was used to assess the population distribution and effect their separation. For both polyprotein products, SDS-PAGE indicated recovery of desired products, namely a 25 kDa band corresponding to an Lsm polyprotein.

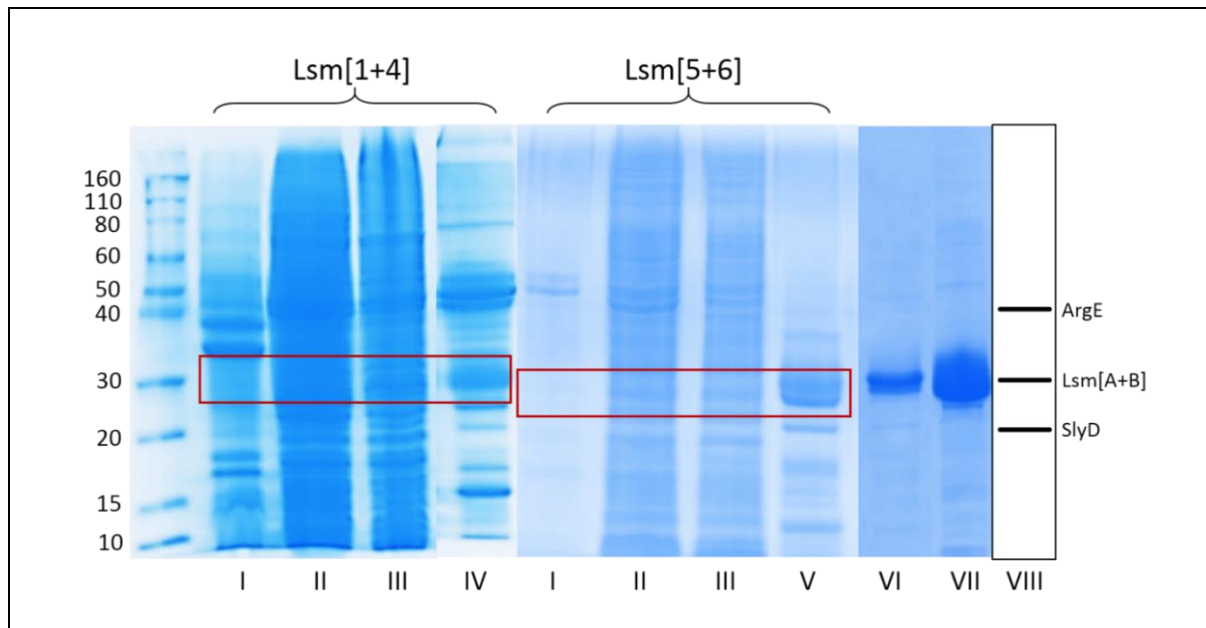


Fig. 3.2 | SDS-PAGE profile of Lsm polyprotein material prepared in this study

I: insoluble fraction from host cell. II: soluble fraction from host cell. III: IMAC eluate fraction. IV Lsm[1+4] purification following SEC. V: Lsm[5+6] purification following SEC. VI: Lsm[4+1] [63]. VII: Lsm[2+3] [63]. VIII: predicted sizes of Lsm polyproteins and His-rich *E. coli* IMAC contaminants [79]. Ladder sizes indicated in kDa.

In the case of preparations of Lsm[1+4], a high degree of contamination ranging from 60 kDa to 10 kDa was consistently found in samples analysed following SEC, which produced chromatographic traces with a large number of peaks between 80 kDa and 20 kDa.

However, preparations of Lsm[5+6] appeared somewhat purer, with some low molecular weight contaminants seen below the predicted polyprotein. SDS-PAGE profiles for successful tectons, Lsm[4+1] and Lsm[2+3] [63], are included in Fig 3.2. These present isolated bands corresponding to 20 kDa with no contaminations detected. This indicates that the two Lsm polyproteins I investigated were potentially expressed at reduced yields compared to Lsm[4+1] and Lsm[2+3], as with less quantity of absorbed recombinant product, other host His-rich contaminants are not crowded out and bind to the affinity matrix.

Unlike the previously characterised Lsm polyproteins, the Lsm polyproteins I investigated did not produce pure samples. Sample contamination by host *E. coli* proteins was most pronounced in Lsm[1+4]. It is well documented that IMAC-based protocols will recover His-rich proteins from *E. coli* such as ArgE (42.3 kDa) and SlyD (20.8 kDa), both of which are known to adsorb to Ni^{2+} containing matrix, which was present in the implemented IMAC

protocol [79, 80]. Another common IMAC contaminant is Hfq, but this was not present in this purification as Hfq deficient cell lines were utilised for production. Lsm[5+6] does not appear to contain this higher molecular weight host proteins, thus the multiple bands observed after purification might suggest protein degradation.

To reduce any non-specifically bound His-rich host contaminants proteins, the IMAC protocol was altered to increase imidazole concentration in Lsm[1+4] from 50 mM to 500 mM over a 30 min time gradient instead of the standard single concentration increase standardly used. This was undertaken from the assumption that His₆ tagged protein would display tightest binding to the IMAC matrix with the host contaminants less tightly adsorbed. By slowly increasing the imidazole concentration, it was predicted that the less tightly adsorbed proteins would elute first leaving the target polyprotein bound until a higher imidazole concentration was reached. Protein eluted off the matrix at 230 mM and fractions were taken every 0.5 ml. Fig 3.3 shows the SDS-PAGE profiles of eluted protein, with contaminants eluting between 230 mM and 260 mM imidazole. A protein band corresponding to Lsm[1+4] was seen at 280 mM, suggesting that an imidazole gradient could be used as a purification technique although very low polyprotein concentration was seen and some contaminants remained. As Lsm[5+6] displayed material of interest as potential tectons, this polyprotein was pursued for biophysical characterisation.

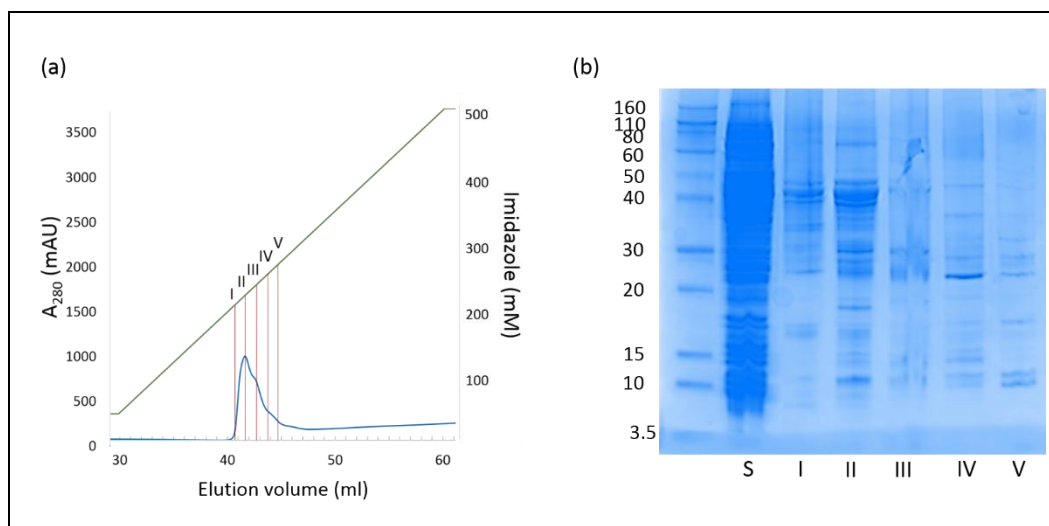


Fig 3.3 | Imidazole concentration dependant purification of Lsm[1+4]

S: soluble material of host cell. I: eluted protein at 230 mM imidazole. II: eluted protein at 240 mM imidazole. III: eluted protein at 255 mM imidazole. IV: eluted protein at 281 mM imidazole. V: eluted protein at 290 mM imidazole. Ladder indicates molecular size in kDa.

3.3 Differences between Lsm[1+4] and Lsm[4+1]

Previous investigation by Francesca Manea [78] determined that the composition of the genetic linker influences the final oligomeric assembly as increasing the charge for Lsm[4+1] lead to an increased proportion of Lsm[4+1]₄ over 2Lsm[4+1]₂. Linker length also plays a role in ring assembly, suggesting that the reduced length of only nine amino acids was not long enough to be favourable to ring formation as steric interference may have occurred.

The main organisation difference between Lsm[4+1] and Lsm[1+4] is the inversion of subunits. Due to a lack of atomic structure, it is currently unknown what order the Lsm[4+1] arrange themselves in solution. Therefore, the high stability of Lsm[4+1] may be a result of adjacent polyprotein assembling along the natural Lsm4-Lsm1 interface, with the non-natural interface, i.e. the natural interface between Lsm4-Lsm7 and Lsm1-Lsm2, being internal to each polyprotein and structural welded via the intersubunit linker. As this stability is not observed for Lsm[1+4] [61], it is possibly that the polyprotein subunit is reversed and the non-natural interface connects adjacent polyproteins, resulting in the dramatically reduced stability observed.

3.4 Biophysical characterisation of Lsm[5+6]

Purified samples of polyprotein Lsm[5+6] were subjected to analytical SEC to assess protein species distribution and the masses of quaternary assemblies. Superdex 200 matrix, known to separate assembly sizes previously observed for Lsm polyproteins [61, 63, 64], was used and equilibrated in Tris buffer containing 400 mM NaCl to promote hydrophilic interactions to assist natural Lsm interface association, to promote the formation of Lsm ring assemblies.

At high ionic strength (400 mM), two populations were evident in solution whose elution was consistent with masses of 90 kDa and 45 kDa, (peaks I and II in Fig 3.4). When compared to predicted masses of 22 kDa for Lsm[5+6], these might be attributed to Lsm[5+6]₄ (88 kDa), and Lsm[5+6]₂ (44 kDa). The purity of the sample was supported by the SDS-PAGE analysis which showed no proteins at 45 kDa and 90 kDa. Under these relatively high salt conditions, the dimeric population appears predominant constituting 70% of the population. Thus, this polyprotein does not appear to form a stable tecton, whether made up of dimer or tetramer combinations. The observed tetrameric and dimeric populations were isolated during analytical SEC and independently reapplied to the analytical column. These

populations were seen to be kinetically stable as no interconversion between the tetrameric and dimeric populations was observed and the assemblies eluted at sizes identical to the mixed solution.

To probe the nature of subunit assembly forces driving these complexes, the Lsm[5+6] material was investigated in solution at low ionic strength (200 mM NaCl). In this case, the material completely disassociated to predominantly monomeric Lsm[5+6] (Peak III, 20 kDa) whilst retaining a small dimeric state (peak II, 45 kDa). This indicates that stable ring tectons cannot be formed by this material

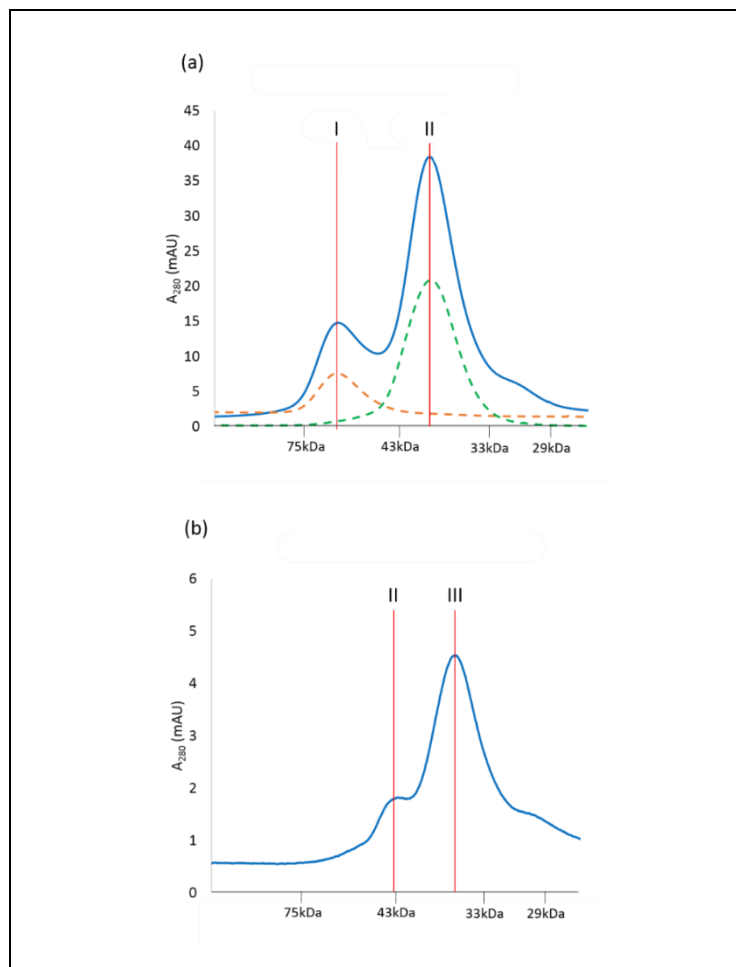


Fig 3.4 | Salt dependency SEC profile of Lsm[5+6] (pH 8.0)

(a) Lsm[5+6] at 400 mM NaCl on Superdex 200. Isolated populations reappplied to same matrix are shown in orange and green. (b) Profile of Lsm[5+6] at 200 mM NaCl. Flow rate was 0.5 ml/min. Estimated peak sizes are: I = 90 kDa, II = 45 kDa, and III = 20 kDa.

Chapter 4: Crystallisation of an engineered Lsm α tecton

4.1 An engineered form of Lsm α

The homomeric Lsm ring assembly from *M. thermoautotrophicum* has been extensively investigated to assess its feasibility as a supramolecular assembly tecton. It is known to consist of seven protomers and possess high thermostability, structural stability, and longevity. Previous engineering, performed by Dr Akshita Wason [64], successfully added non-native covalent linkages to engineer nanotubes and nanocages via metal ion coordination.

An extension of Dr Wason's project investigated the possibility of tuning Lsm tecton ring and pore diameter by influencing the number of protomers constituting the final assembly. Lsm protomers primarily self-assemble through hydrophobic interactions stabilised by a hydrogen bonding network connecting a $\beta 4$ β -strand and the $\beta 5$ β -strand of the neighbouring protomer [81]. This occurs for all Lsm assemblies whether they be hexameric, heptameric, or octameric. Dr Wason's design hypothesis was to mutate residues critical to the interface formed between protomers, without disrupting the integrity of the Lsm protomer fold. A series of interface residues found to interface between protomers, engaged in hydrogen bonds, salt bridges, or hydrophobic interactions, were selected for mutagenesis generally to alanine or proline.

When comparing the interfacing $\beta 4$ - $\beta 5$ residues between a hexamer (Hfq), a heptamer (Lsm α), and octamer (Lsm3), shown in Fig 4.1, high conservation in the $\beta 5$ β -strand was noted whilst the $\beta 4$ β -strand demonstrated some variation. This informed a hypothesis that modifying the $\beta 4$ β -strand may result in modified intersubunit angles between protomers.

R65Lsm α was originally selected for mutagenesis as although it plays no direct role in the interface, it was hypothesised that the mutation would alter the bend in the $\beta 4$ β -strand, perturbing the native $\beta 4$ - $\beta 5'$ interactions. As R65PLsm α appeared to be consistent with a hexameric ring morphology from biophysical data generated by Dr Wason, which detected assemblies corresponding to six subunits via small angle x-ray scattering (SAXS) and SEC,

instead of the unmodified heptameric protein, it is likely that the β 4 bend was indeed altered, thereby altering the hydrogen binding network between the protomers [64].

In my thesis, I took the opportunity to test this hypothesis by further validating the mutant assembly in solution form and take it onto crystallographic study. An atomic model provided by x-ray crystallography will conclusively outline the number of protomer components and their potentially novel interface interactions in the ring assembly.

4.2 R65PLsm α in solution

The construct utilised for prior crystallographic study of Lsm α [53] incorporated an N-terminal GST fusion tag for purification purposes. This production route was therefore also used by Dr Wason for generation of R65PLsm α . This contrast with the His₆ affinity tags implemented in chapter 3, which would have simplified purification, but was required to prevent the large protein cages formed from His₆-Lsm α sequestering Ni(II) material being formed [64]. Therefore to ensure fabrication of single ring tectons, GST fusions were prepared for the mutant of interest.

Recombinant production of R65PLsm α employed IPTG induction (see sections 2.3.3.2, 2.3.3.5). Fig 4.2 shows SDS-PAGE analysis of my samples at various stages of their purification. Following cell lysis a broad suite of proteins were present, with an intense band observed at approximately 30 kDa. This is likely to be the recombinant product GST-R65PLsm α , as it matches the predicted molecular weight at 35 kDa. Following GST cleavage with thrombin and a polishing step of preparative SEC, three protein bands only are present corresponding to 40 kDa, 30 kDa, and 10 kDa. The most intense, at 10 kDa, signifies a good yield of expressed and cleaved R65PLsm α (predicated mass of 9 kDa). The weakest band, at 30 kDa, is most likely the result of thrombin contamination, itself a 30 kDa protein. Knowing the desired product to an aggregate species of ~60 kDa in solution, an attempt to eliminate thrombin was conducted by arresting protein collection on preparative SEC at a elution volume corresponding to a 45 kDa cutoff, meaning all protein below that size was excluded from pooled samples. This successfully removed the 30 kDa band, confirming that thrombin contamination had occurred. The 40 kDa band is a common feature in all purifications of Lsm α proteins, seen in both GST and His₆ purifications, and is hypothesised to be a multimer of Lsm α brought on by the high thermostability of the protein ring [64]. It is highly unlikely to be un-cleaved GST-

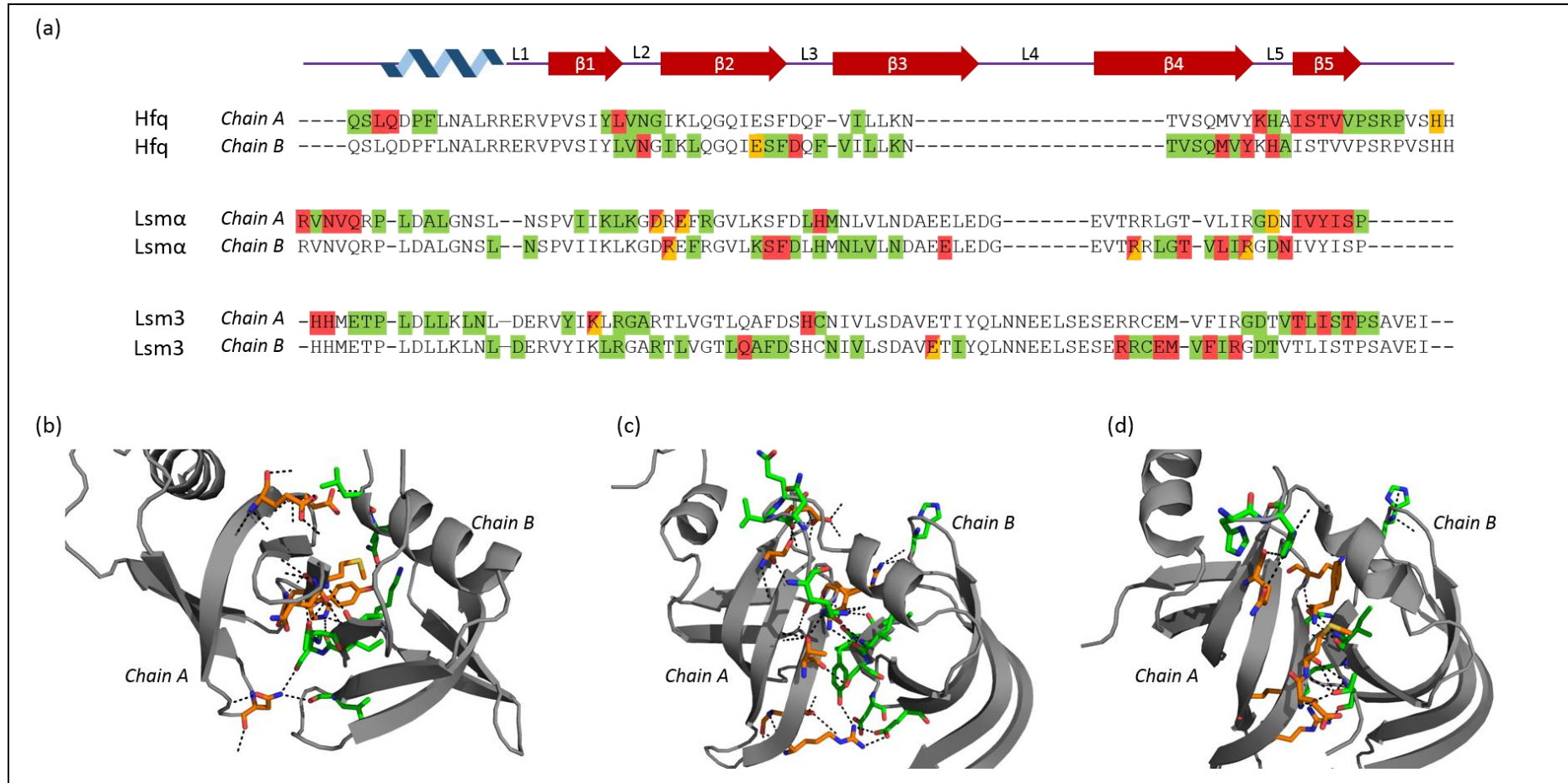


Fig 4.1 | Lsm protein subunit interfaces

(a) Structural alignment of Hfq, $\text{Lsm}\alpha$, and Lsm3. Interacting residues shown in green, hydrogen binding residues in red, and salt bridging residues in yellow. (b) Residues responsible for hydrogen binding and salt bridges on $\beta 4$ - $\beta 5'$ hexameric Hfq interface. (PDB ID: 3QHS [82]). (c) Residues responsible for hydrogen binding and salt bridges on $\beta 4$ - $\beta 5'$ heptameric $\text{Lsm}\alpha$ interface. (PDB ID: 1MGQ [53]). (d) Residues responsible for hydrogen binding and salt bridges on $\beta 4$ - $\beta 5'$ octameric Lsm3 interface. (PDB ID: 3BW1 [50]). All interface data generated from PDBePISA [83].

R65PLsm α as the Lsm α components would assemble in solution leading to high molecular species at approximately 210 kDa in solution. This was not observed during preparative SEC. Whilst native PAGE should allow for the determination of assembly size, practice with Lsm α has found them to not yield accurate sizes.

To assess the solution state of the recovered R65PLsm α , purified samples were introduced to SEC matrix to ascertain assembly size, where a single species, corresponding to 60 kDa, was observed. R65PLsm α , along with prepared Lsm α , were run individually on analytical SEC, and the two assemblies were seen to elute at slightly different elution points, corresponding to approximately 65 kDa and 55 kDa, the predicted sizes of a Lsm α heptamer and hexamer respectively. This allowed for a comparison between Lsm α and R65PLsm α demonstrating that the mutant appeared to possess a smaller assembly mass, as shown in Fig 4.3. This result corresponded with biophysical data observed by Dr Wason in her project [64]. However, determining a mass comparison between Lsm α and R65PLsm α from the analytical SEC was difficult to achieve as only a 9 kDa difference exists between the two assemblies. The predicted elution values for the two assemblies on the utilised size exclusion column was 15.3 ml for the heptamer, and 16.0 ml for the hexamer. Therefore differentiating between the two assemblies was problematic as the difference of 9 kDa is difficult to observe via SEC techniques. To determine a more accurate molecular weight for the variant R65PLsm α , I performed MALLS as this solution-based technique has the potential to more precisely determine correct assembly masses. The readout obtained from this data is shown in Fig 4.3.

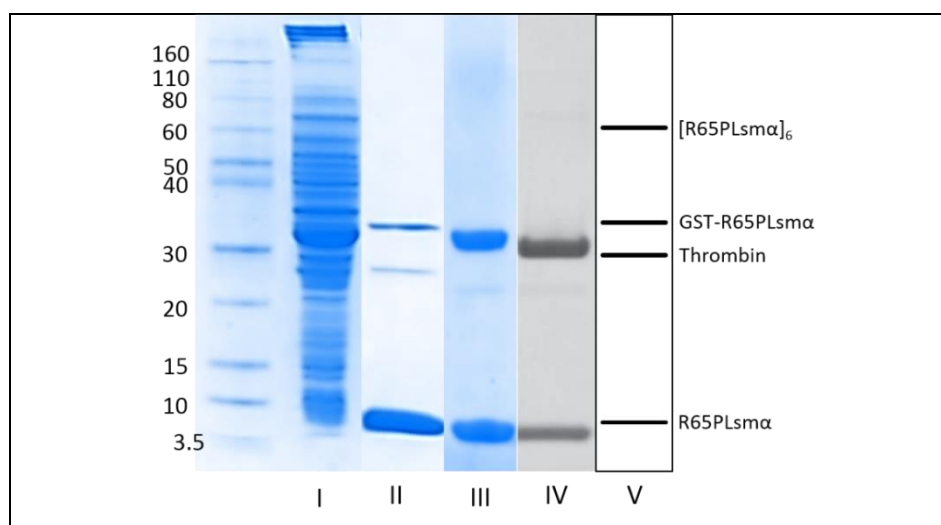


Fig 4.2| SDS PAGE analysis of samples of R65PLsm α production steps

(I) lysate (II) post initial preparative SEC (III) post thrombin excision SEC (IV) R65PLsm α as produced by Dr Akshita Wason [64] (V) Theoretical molecular weights of proteins involved in the purification protocol. Molecular weight ladder shown in kDa.

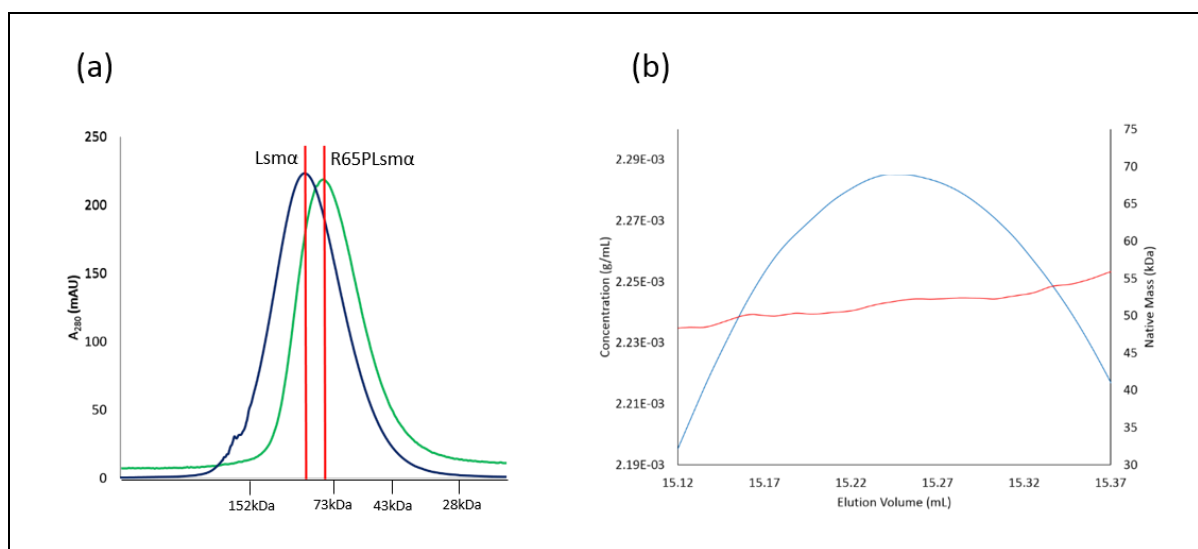


Fig 4.3 | Biophysical characterisation of R65PLsm α through size exclusion and light scattering

(a) Lsm α (blue) and R65PLsm α (green) eluting on analytical SEC (b) SEC-MALLS of R65PLsm α where the absorbance at 280nm is shown in blue and the molecular weight distribution shown in red. Both experiments were run at 0.5 ml in a 24 ml column and in PBS, pH 7.5.

Whilst a molecular weight of 53.3 kDa was detected, this was found with a 23% margin of error, as although a single population was identified, refractive index measurements could not precisely determine a single solution mass. As such, this margin of error fell within the size of a heptameric assembly (see Fig 4.4). This reconfirmed the limitations of solution based mass determination as difficulties are found in differentiating a 9 kDa difference. As such, a more accurate method was required.

4.3 Crystallisation of R65PLsm α

I continued to pursue structural definition of R65PLsm α via x-ray crystallography to confirm at atomic level the nature of the assembled form of this complex. First steps involved commercial sparse matrix screens, based on highly successful crystallisation conditions determined by the MCSG and JCSG [84, 85]. This is a first pass of a broad range of conditions to observe formation of crystalline material [75]. Previous studies demonstrated that the Lsm

family displays a preferred RNA binding on the helix face of the ring with uracil rich sequences of RNA (section 1.2.3) [52, 60]. Investigation by the Protein Structure Group at MQ utilised isothermal titration calorimetry (ITC) determined that U₄CGU₄ RNA demonstrated the tightest binding with Lsm (Moll, Littlejohn, and Mabbutt, unpublished). As such, this RNA sequence was included as a binding partner in co-crystallisation scenarios.

4.3.1 Crystallisation

For the initial screening, 96 well plates in a sitting drop format were used in an attempt to slowly crystallise R65PLsm α , with and without an RNA partner (U₄CGU₄). The top 200 crystallisation conditions from MCSG (MCSG1 and MCSG2), and JCSG (JCSG+) were used for R65PLsm α . These conditions are known to easily reduce the solubility of a target protein until supersaturation and crystallogenesis is achieved [85]. For R65PLsm α with U₄CGU₄ RNA, screening was conducted with the top 100 conditions from MCSG and JCSG (MCSG1 and JCSG+). The plates were monitored under a microscope over three months and crystal growth recorded. These plates were scored so as to note the nature and growth of crystalline material (see Fig 4.4). Wells were assigned a value of 1-5 where 1 = light precipitant, 2 = heavy precipitant, 3 = phase separation, 4 = crystalline material, 5 = crystals.

After the full 3 month period, five specific conditions, three from MCSG2 and 2 from JCSG+ showed crystal growth and the size of these crystals varied between 20-100 μ m. These conditions were: 0.2 M potassium chloride, 0.05 M HEPES:NaOH pH 7.5, 35% pentaerythritol propoxylate; 0.2 M ammonium acetate, 0.1 M HEPES pH 7.5, 45% 2-methyl-2, 4-pentadiol (MPD), 0.2 M magnesium chloride, 0.1 M HEPES:NaOH pH 7.5, 30% polyethylene glycol (PEG) 400; 0.2 M magnesium chloride hexahydrate, 0.2 M Tris pH 8.0, 20% PEG 8000, and 0.2 M potassium acetate, 20% PEG 3350, as shown in Table 4.2. The crystals displayed an obvious pH condition trend as three of the five conditions were at pH 7.5 and one at pH 8.5. Additionally, all these five conditions contained metal salts with a valence of +1 or +2. Select crystals are shown Fig 4.5.

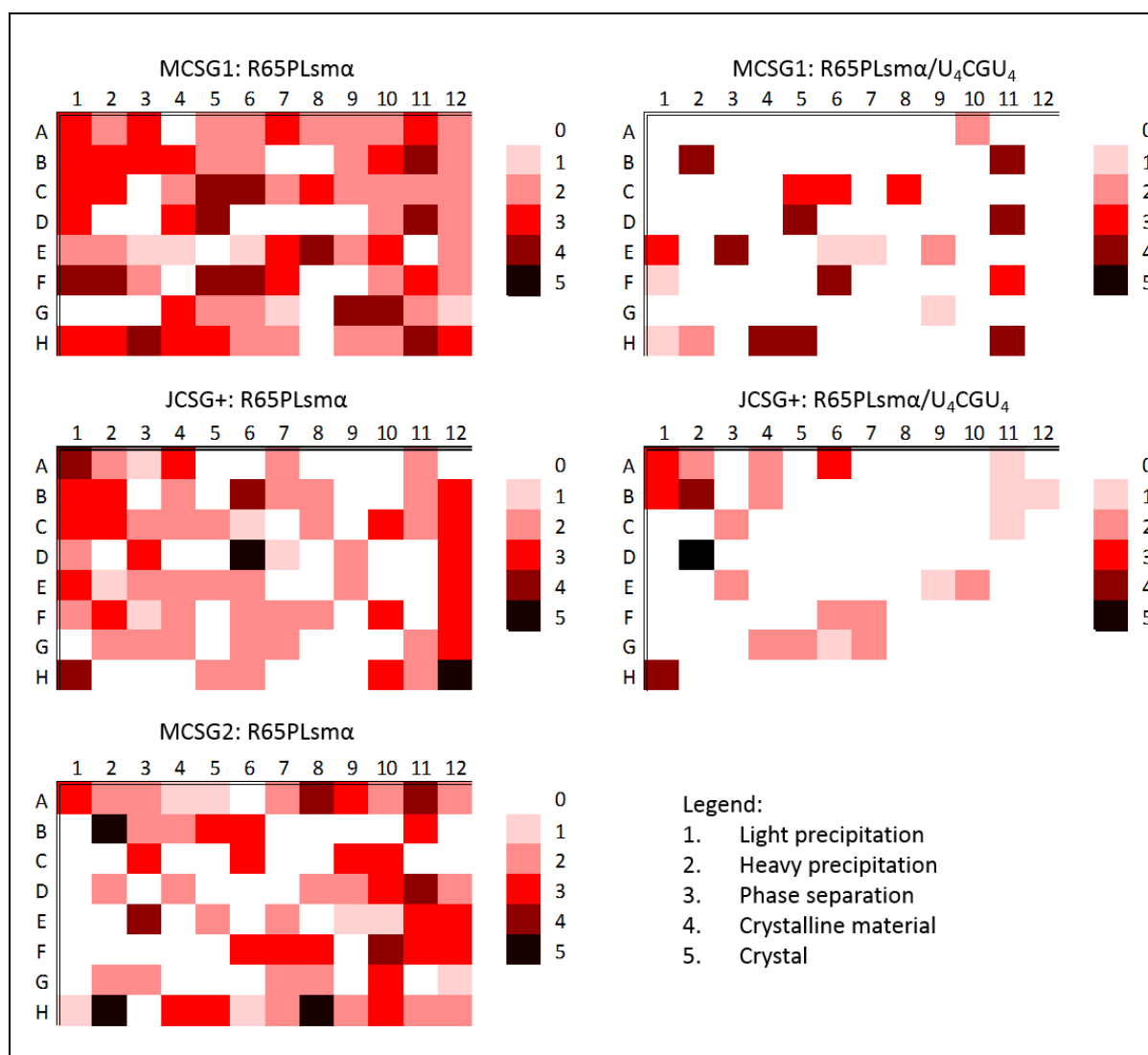


Fig 4.4| Crystallisation heat maps of R65PLsm α and R65PLsm α /U₄CGU₄

Five crystal hits were observed for R65PLsm α (black) in MCSG2 (H2, H8, and B2) and JCSG+. For R65PLsm α /U₄CGU₄, one crystal was observed (black) in JCSG+ (D2).

Significantly less crystals were observed for R65PLsm α /U₄CGU₄ as only one condition grew crystals large enough to harvest (see table 4.1), although 12 conditions were seen with crystalline material. Lower levels of precipitation and phase separation was also detected across the plates suggesting they were undersaturated with R65PLsm α /U₄CGU₄. A similar trend to the non-RNA plates was observed as the single hit contained a magnesium salt and a pH of 7.5.

Table 4.1 | Conditions of crystals screened at the UNSW home source

| Without RNA | | | | | | |
|-------------|-------|------|--------------------------------------|--------------------------|---------------------------------|-------------|
| Crystal | Plate | Well | Salt | Conditions Buffer | Precipitant | Diffraction |
| A | MCSG2 | H2 | 0.2 M potassium chloride | 0.05 M HEPES:NaOH pH 7.5 | 35% pentaerythritol propoxylate | 2.4 Å |
| B | JCSG+ | H12 | 0.2 M ammonium acetate | 0.1 M HEPES pH 7.5 | 45% MPD | 3.0 Å |
| C | MCSG2 | H8 | 0.2 M magnesium chloride | 0.1 M HEPES:NaOH pH 7.5 | 30% PEG 400 | 3.7 Å |
| D | JCSG+ | D6 | 0.2 M magnesium chloride hexahydrate | 0.1 M Tris pH 8.5 | 20 % w/v PEG 8000 | 8 Å |
| E | MCSG2 | B2 | 0.2 M potassium acetate | | 20% PEG 3350 | >8 Å |

| With <i>U₄CGU₄</i> RNA | | | | | | |
|--|-------|----|--------------------------|--------------------|--------------|-------|
| F | JCSG+ | D2 | 0.2 M magnesium chloride | 0.1 M Hepes pH 7.5 | 30% PEG 8000 | >12 Å |

Full list of sparse matrix screening condition available on suppliers website:

MCSG (<http://www.microlytic.com/content/mcsg-suite>) (last accessed 7th October 2015)

JCSG (<http://www.moleculardimensions.com/applications/upload/MD1-40.pdf>) (last accessed 7th October 2015)

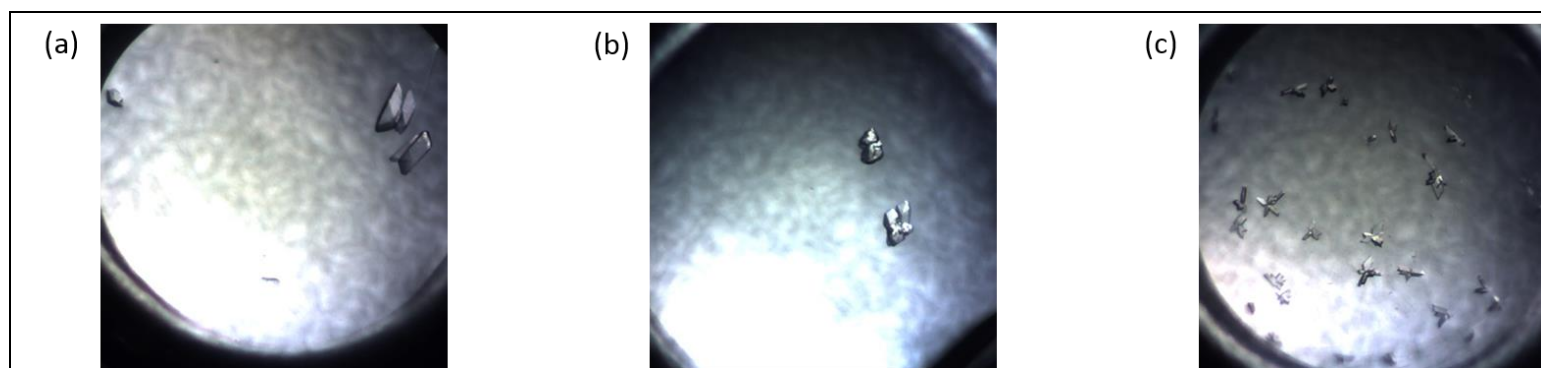


Fig 4.6 | Select crystal images from crystallisation screening

(a) 0.2 M potassium chloride, 0.05 M HEPES:NaOH pH 7.5, 35% pentaerythritol propoxylate. (b) 0.2 M ammonium acetate, 0.1 M HEPES pH 7.5, 45% MPD (c) 0.2 M magnesium chloride, 0.1 M HEPES:NaOH pH 7.5, 30% PEG 400. Individual image designations match crystal naming in table 4.2

4.3.2 Crystal Optimisation

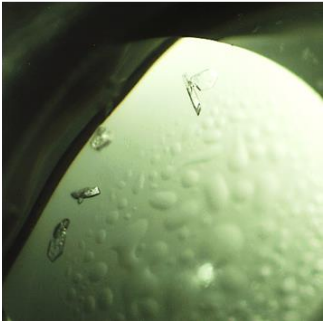

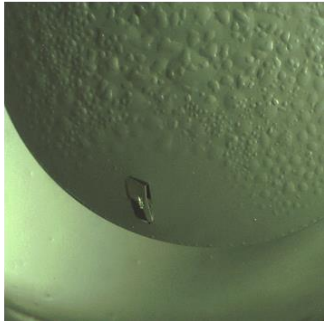
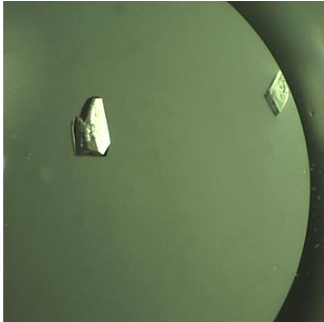

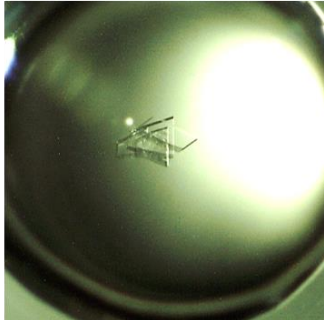
The six crystals were screened at the UNSW home source (see section 2.5.2) by Dr Shah. Once placed into the cryostream, images were taken at 0° and 90° to assess the extent of crystal diffraction and potential space groups within the crystal allowing for a data collection strategy to be determined. For the 5 crystals observed for R65PLsm α , the three crystals at pH 7.5 all diffracted to a resolution close to 3 Å (crystals A, B, and C – table 4.1), with the highest resolution observed at 2.4 Å, as compared to 8 Å for the pH 8.5 crystal (crystal G). The RNA co-crystallisation crystal (crystal F) diffracted (>12 Å) as well as showing signs of extensive radiation damage.

Conditions from the sparse matrix screens which diffracted to a high resolution of ~3.5 Å under initial screening (A, B, and C) were selected for optimisation to generate improved crystals, diffracting to a higher resolution. Optimisation was carried out employing three main approaches; varying the protein:precipitant ratios, optimising the crystallisation cocktail conditions to trial different salt molarities and pH conditions, and utilising both sitting drop and hanging drop vapour diffusion formats.

For condition A, large crystals were seen across the majority of protein:precipitant ratios trialled in both the sitting drop and hanging drop formats. The crystals were seen in protein:precipitant ratios of 1:1, 2:1, and 3:1, crystal images are shown in table 4.2. These crystals shared a similar flat plate morphology to what was observed in the initial screen (see Fig 4.5), and appeared within a week of plate set up. Overall, differing protein:precipitant ratios seemed to have no overall effect in crystallogenesis, as crystals were seen in almost all ratios. However, no crystals were observed when the components, salt molarity and pH, were varied, indicating that the standard conditions were more suitable for crystallogenesis. Crystal growth was consistent between sitting drop and hanging drop formats.

Very few crystals were observed for conditions B and C as only a single crystal for each was observed in optimisation trials. These were a large and regular hexagonal crystal for condition B and a series of long needle-like crystals for condition C. These were both observed when the crystallisation conditions, salt molarity and pH, were varied.

Table 4.2 | Successful crystal hits of condition A following optimisation

| | | Protein:Precipitant Ratios | | |
|--------------|-----|---|-----|---|
| | | 1:1 | 2:1 | 3:1 |
| Sitting Drop | (a) |  | (b) |  |
| | (c) |  | (d) |  |
| Hanging Drop | (e) |  | (f) |  |

All crystallisation conditions were 0.2 M potassium chloride, 0.05 M HEPES:NaOH pH 7.5, 35% pentaerythritol propoxylate

4.3.3 Crystal harvesting and cryopreservation

The crystals were harvested on a nylon microfiber loop by Dr Shah and rapidly quenched in liquid N₂ to limit radiation damage occurring under the x-ray beamline. If submersion in the N₂ is not carried out quickly enough, the formation of ice crystals within the looped crystal droplet occurs. This can be remedied through the addition of a cryoprotectant, which prevents the formation of crystalline ice within the looped droplet. Typical cryoprotectants added are PEG, glycerol, or a salt which creates a high salt environment in the droplet [86].

When the crystal from condition A (table 4.1) was exposed to the home source x-ray beamline at UNSW diffraction at 3 Å was observed along with persistent ice rings present (Fig 4.6). As a result, a cryoprotectant was added to remove the ice rings as well as to prevent radiation damage. Dr Shah achieved this by increasing the concentration of pentaerythritol propoxylate, the cryoprotectant already found in the condition, from 35% to 50% by soaking the crystal in a mixed solution of mother liquor and cryoprotectant at a 1:1 ratio for 30 s to 5 min. This successfully removed the ice rings, allowing the crystals to diffract at higher resolution of 2.4 Å (see Fig 4.6). For crystals from conditions B and C, no ice rings were observed in screening and therefore did not require the addition of a cryoprotectant.

All crystals observed in both the initial screens and optimisation screens were harvested and stored in liquid N₂ for shipping to the Australian Synchrotron for data collection scheduled for the 2nd October 2015.

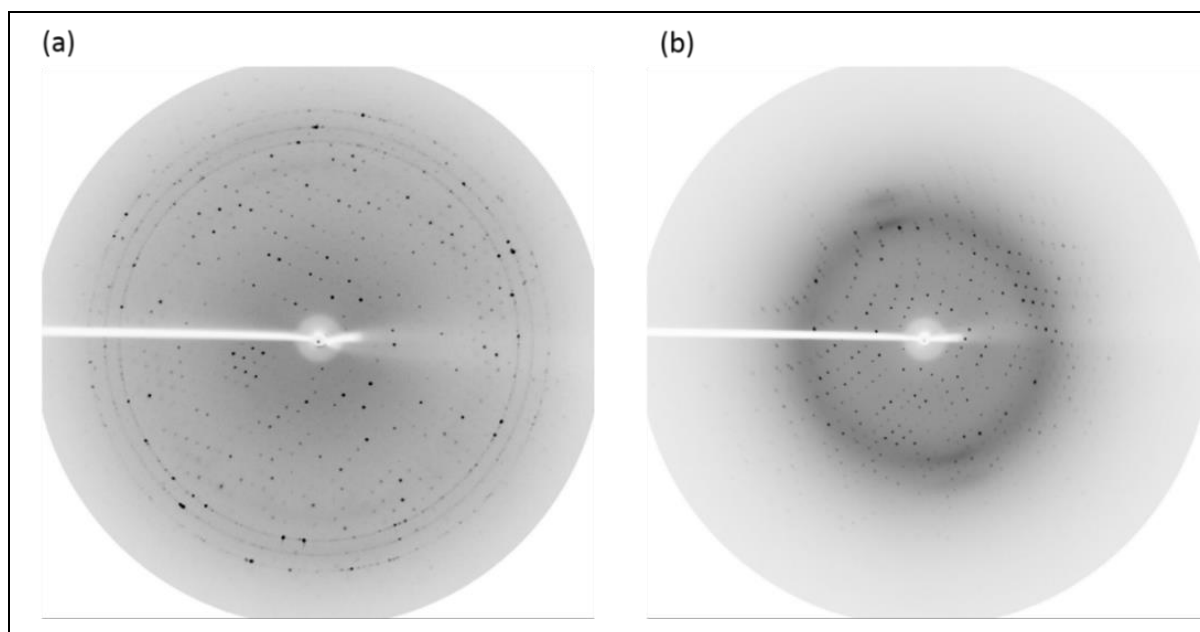


Fig 4.6| Crystal A diffraction patterns before and after cryopreservation

(a) Initial screening of condition A screening detected persistent ice rings within the looped droplet, limiting diffraction resolution to 3 Å. Detector distance at 250 mm. (b) Treatment with addition pentaerythritol propoxylate removed ice rings, increasing diffraction resolution to 2.4 Å. Detector distance at 200 mm.

Chapter 5: Conclusion

5.1 An evaluation of different Lsm polyproteins

Some organisations of paired Lsm polyproteins have proved successful as tectons for nanofabrication. Two less-characterised polyproteins, Lsm[1+4] and Lsm[5+6], have here been investigated to determine whether they are appropriate additions to the current repertoire of dual polyproteins.

For Lsm[5+6], two protein populations were observed in solution. These corresponded to Lsm[5+6]₄ and Lsm[5+6]₂. The conclusion was made that Lsm[5+6] has less potential as a ring tecton as compared to previously investigated polyproteins, Lsm[4+1] and Lsm[2+3], due to lack of a cohesive hydrophobic interaction. The lack of hydrophobic cohesion could be explained by the polyprotein components and their corresponding natural interfaces as the inspiration for this assembly comes from a natural observed hexamer constituted of Lsm5, Lsm6, and Lsm7 [62]. Any Lsm[5+6] complexes may not be stable due to the lack of Lsm7 which may provide a crucial interface within the assembly. Polyprotein versions of a Lsm[5+6+7] trimer were attempted but were insoluble in solution. If Lsm7 is crucial for the assembly of a hexamer, a stable assembly could be possible if the co-expression of three polyproteins, Lsm[5+6], Lsm[7+5], and Lsm[6+7], was attempted as this could recreate the hexameric assembly order seen in nature.

The other polyprotein investigated in this thesis, Lsm[1+4], proved difficult to cleanly produce. However, it may be alteration of purification protocols at IMAC level, or a follow through step with an ion-exchange SEC, may allow for less contaminated samples. My characterisation was limited by the quality of the samples at this point.

5.2 Crystal structure of R65PLsm α

On the 2nd of October, 2015, R65PLsm α crystals were screened at the Australian Synchrotron and data was collected from them by Dr Shah. From the collected 2.4 Å diffraction patterns, Dr Shah trialled molecular replacement solutions using monomeric atomic structures of Lsm α (PBD 1I31) [51] so as to reduce any bias to the computed structure. The molecular replacement solutions performed self-rotation analysis (POLARRFN program in the CCP4 suite [51, 87] with both six-fold and seven-fold internal symmetry.

Very surprisingly, the preliminary symmetry map, which details electron densities, was consistent with seven-fold internal symmetry (see Fig 5.1). This proves that the R65PLsm α assembly was heptameric in format, not the hexamer indicated by the solution-state work conducted by both myself and others. This further confirms the limitations of SEC and SAXS solution based biophysical characterisation as being insufficient for the differentiation of assemblies differing in 9 kDa. These limitations are that SEC allows the separation of globular biomolecules by size as they filter through porous matrix. This separation is dependent on differing Stoke radii between eluting particles, and as only a 14% size difference was predicted between a heptamer and hexamer, only a small difference in Stokes radius would have been present. SEC also relies on globular proteins, and their respective Stokes radii, as estimation standards, making mass estimations for non-globular proteins, such as Lsm α rings, difficult as these molecules would pass through the matrix in a manner not solely dependent on mass due to their non-globular morphology [88]. SAXS, on the other hand, detects the light scattering of x-rays to provide a size estimate of biomolecules at relatively low resolution. As such this can only provide an estimate of size differences between a 64 kDa heptamer and a 53 kDa hexamer.

As biophysical characterisation performed by myself and Dr Wason [64] detected a reduced assembly, further refinement of the generated atomic structure is required to determine if the mutation may have resulted in a compacted assembly, which may provide explains for the biophysical data acquired.

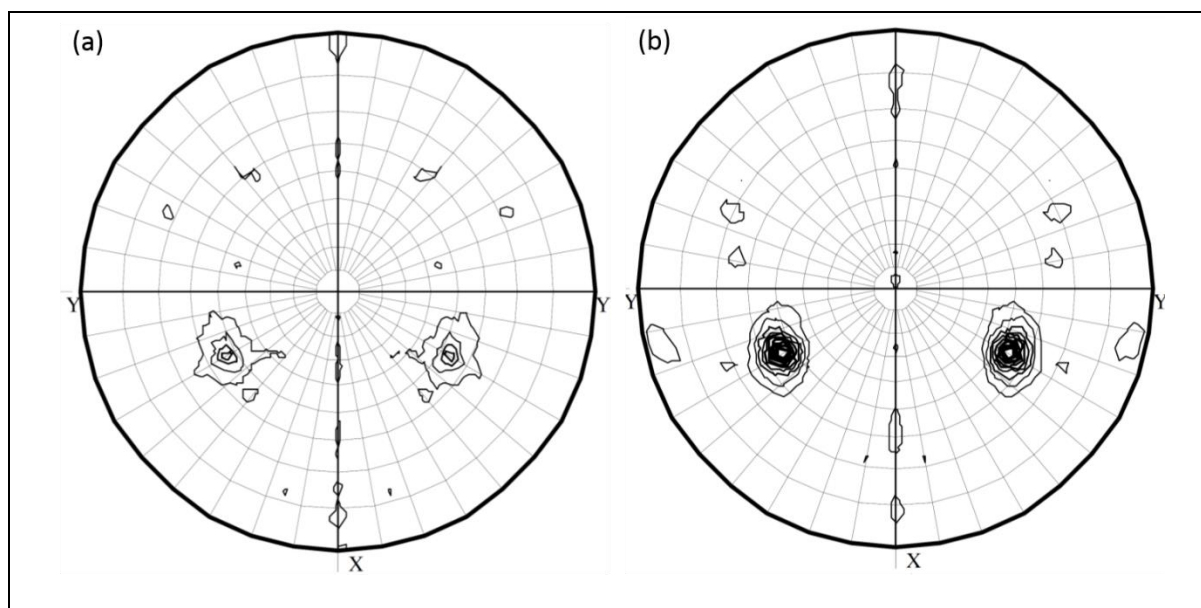


Fig 5.1 | Self-rotation analysis of R65PLsma electron densities

(a) molecular replacement of R65PLsma under six-fold symmetry, at Chi 60. (b) molecular replacement of R65PLsma under seven-fold symmetry, at Chi 51.4.

References:

- [1] Whitesides GM. The 'right' size in nanobiotechnology. *Nat Biotechnol.* 2003;21:1161-5.
- [2] Rothemund PW. Folding DNA to create nanoscale shapes and patterns. *Nature.* 2006;440:297-302.
- [3] Tenne R. Inorganic nanotubes and fullerene-like nanoparticles. *Nat Nanotechnol.* 2006;1:103-11.
- [4] Storm K, Halvardsson F, Heurlin M, Lindgren D, Gustafsson A, Wu PM, et al. Spatially resolved Hall effect measurement in a single semiconductor nanowire. *Nat Nanotechnol.* 2012;7:718-22.
- [5] Padilla JE, Colovos C, Yeates TO. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A.* 2001;98:2217-21.
- [6] Channon K, Bromley EH, Woolfson DN. Synthetic biology through biomolecular design and engineering. *Curr Opin Struct Biol.* 2008;18:491-8.
- [7] Saaem I, LaBean TH. Overview of DNA origami for molecular self-assembly. *Wiley Interdiscip Rev Nanomed Nanobiotechnol.* 2013;5:150-62.
- [8] Cui H, Webber MJ, Stupp SI. Self-assembly of peptide amphiphiles: from molecules to nanostructures to biomaterials. *Biopolymers.* 2010;94:1-18.
- [9] Simon RJ, Kania RS, Zuckermann RN, Huebner VD, Jewell DA, Banville S, et al. Peptoids: a modular approach to drug discovery. *Proc Natl Acad Sci U S A.* 1992;89:9367-71.
- [10] Ulijn RV. Molecular self-assembly: Best of both worlds. *Nat Nanotechnol.* 2015;10:295-6.
- [11] Heddle JG. Protein cages, rings and tubes: useful components of future nanodevices? *Nanotechnol Sci Appl.* 2008;1:67-78.
- [12] Ulijn RV, Woolfson DN. Peptide and protein based materials in 2010: from design and structure to function and application. *Chem Soc Rev.* 2010;39:3349-50.
- [13] Gerrard JA. Protein Nanotechnology: What Is It? In: Gerrard JA, editor. *Protein Nanotechnology*: Humana Press; 2013. p. 1-15.
- [14] Astier Y, Bayley H, Howorka S. Protein components for nanodevices. *Curr Opin Chem Biol.* 2005;9:576-84.
- [15] Howorka S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr Opin Biotechnol.* 2011;22:485-91.
- [16] Glover DJ, Giger L, Kim JR, Clark DS. Engineering protein filaments with enhanced thermostability for nanomaterials. *Biotechnol J.* 2013;8:228-36.
- [17] Smith BA, Hecht MH. Novel proteins: from fold to function. *Curr Opin Chem Biol.* 2011;15:421-6.
- [18] Lai YT, King NP, Yeates TO. Principles for designing ordered protein assemblies. *Trends Cell Biol.* 2012;22:653-61.
- [19] Sinclair JC. Constructing arrays of proteins. *Curr Opin Chem Biol.* 2013;17:946-51.
- [20] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28:235-42.
- [21] Evdokimov AG, Mekel M, Hutchings K, Narasimhan L, Holler T, McGrath T, et al. Rational protein engineering in action: the first crystal structure of a phenylalanine tRNA synthetase from *Staphylococcus haemolyticus*. *J Struct Biol.* 2008;162:152-69.

- [22] Boyle AL, Bromley EHC, Bartlett GJ, Sessions RB, Sharp TH, Williams CL, et al. Squaring the Circle in Peptide Assembly: From Fibers to Discrete Nanostructures by de Novo Design. *J Am Chem Soc.* 2012;134:15457-67.
- [23] Doll TAPF, Neef T, Duong N, Lanar DE, Ringler P, Müller SA, et al. Optimizing the design of protein nanoparticles as carriers for vaccine applications. *Nanomedicine: Nanotechnology, Biology and Medicine.* 2015;11:1705-13.
- [24] Mattheaei JF, DiMaio F, Richards JJ, Pozzo LD, Baker D, Baneyx F. Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations. *Nano Lett.* 2015;15:5235-9.
- [25] King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, et al. Accurate design of co-assembling multi-component protein nanomaterials. *Nature.* 2014;510:103-8.
- [26] Whitesides GM, Grzybowski B. Self-assembly at all scales. *Science.* 2002;295:2418-21.
- [27] Zemla J, Lekka M, Raczowska J, Bernasik A, Rysz J, Budkowski A. Selective protein adsorption on polymer patterns formed by self-organization and soft lithography. *Biomacromolecules.* 2009;10:2101-9.
- [28] Whitesides GM, Kriebel JK, Mayers BT. Self-assembly and nanostructured materials. *Nanoscale Assembly: Springer;* 2005. p. 217-39.
- [29] Lai Y-T, Reading E, Hura GL, Tsai K-L, Laganowsky A, Asturias FJ, et al. Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat Chem.* 2014;6:1065-71.
- [30] Ali MH, Imperiali B. Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry.* 2005;13:5013-20.
- [31] Yeates TO, Thompson MC, Bobik TA. The protein shells of bacterial microcompartment organelles. *Curr Opin Struct Biol.* 2011;21:223-31.
- [32] Devenish SR, Gerrard JA. The role of quaternary structure in (beta/alpha)₈-barrel proteins: evolutionary happenstance or a higher level of structure-function relationships? *Org Biomol Chem.* 2009;7:833-9.
- [33] Griffin MDW, Dobson RCJ, Pearce FG, Antonio L, Whitten AE, Liew CK, et al. Evolution of Quaternary Structure in a Homotetrameric Enzyme. *Journal of Molecular Biology.* 2008;380:691-703.
- [34] Lai Y-T, Cascio D, Yeates TO. Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science.* 2012;336:1129.
- [35] McMillan RA, Paavola CD, Howard J, Chan SL, Zaluzec NJ, Trent JD. Ordered nanoparticle arrays formed on engineered chaperonin protein templates. *Nat Mater.* 2002;1:247-52.
- [36] Chen X, Antson AA, Yang M, Li P, Baumann C, Dodson EJ, et al. Regulatory features of the trp operon and the crystal structure of the trp RNA-binding attenuation protein from *Bacillus stearothermophilus*. *J Mol Biol.* 1999;289:1003-16.
- [37] Dgany O, Gonzalez A, Sofer O, Wang W, Zolotnitsky G, Wolf A, et al. The structural basis of the thermostability of SP1, a novel plant (*Populus tremula*) boiling stable protein. *J Biol Chem.* 2004;279:51516-23.
- [38] Gazit E. Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem Soc Rev.* 2007;36:1263-9.
- [39] Grunberg R, Serrano L. Strategies for protein synthetic biology. *Nucleic Acids Res.* 2010;38:2663-75.
- [40] Medalsy I, Dgany O, Sowwan M, Cohen H, Yukashevskaya A, Wolf SG, et al. SP1 protein-based nanostructures and arrays. *Nano Lett.* 2008;8:473-7.

- [41] Miranda FF, Iwasaki K, Akashi S, Sumitomo K, Kobayashi M, Yamashita I, et al. A self-assembled protein nanotube with high aspect ratio. *Small*. 2009;5:2077-84.
- [42] Trent JD, Kagawa HK, Yaoi T, Olle E, Zaluzec NJ. Chaperonin filaments: the archaeal cytoskeleton? *Proc Natl Acad Sci U S A*. 1997;94:5383-8.
- [43] Paavola CD, Chan SL, Li Y, Mazzarella KM, McMillan RA, Trent JD. A versatile platform for nanotechnology based on circular permutation of a chaperonin protein. *Nanotechnology*. 2006;17:1171.
- [44] Clare DK, Vasishtan D, Stagg S, Quispe J, Farr GW, Topf M, et al. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*. 2012;149:113-23.
- [45] Koeck PJ, Kagawa HK, Ellis MJ, Hebert H, Trent JD. Two-dimensional crystals of reconstituted beta-subunits of the chaperonin TF55 from *Sulfolobus shibatae*. *Biochim Biophys Acta*. 1998;1429:40-4.
- [46] Beggs JD. Lsm proteins and RNA processing. *Biochem Soc Trans*. 2005;33:433-8.
- [47] Spiller MP, Boon KL, Reijns MA, Beggs JD. The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res*. 2007;35:923-9.
- [48] Weber G, Trowitzsch S, Kastner B, Luhrmann R, Wahl MC. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J*. 2010;29:4172-84.
- [49] Sharif H, Conti E. Architecture of the Lsm1-7-pat1 complex: a conserved assembly in eukaryotic mRNA turnover. *Cell reports*. 2013;5:283-91.
- [50] Naidoo N, Harrop SJ, Sobti M, Haynes PA, Szymczynska BR, Williamson JR, et al. Crystal Structure of Lsm3 Octamer from *Saccharomyces cerevisiae*: Implications for Lsm Ring Organisation and Recruitment. *Journal of molecular biology*. 2008;377:1357-71.
- [51] Collins BM, Cubeddu L, Naidoo N, Harrop SJ, Kornfeld GD, Dawes IW, et al. Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA. Crystal structure of an oligomeric complex of yeast SmF. *J Biol Chem*. 2003;278:17291-8.
- [52] Sobti M, Cubeddu L, Haynes PA, Mabbutt BC. Engineered rings of mixed yeast Lsm proteins show differential interactions with translation factors and U-rich RNA. *Biochemistry*. 2010;49:2335-45.
- [53] Collins BM, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC. Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. *J Mol Biol*. 2001;309:915-23.
- [54] Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *Embo j*. 2002;21:3546-56.
- [55] Brennan RG, Link TM. Hfq structure, function and ligand binding. *Curr Opin Microbiol*. 2007;10:125-33.
- [56] Achsel T, Stark H, Luhrmann R. The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc Natl Acad Sci U S A*. 2001;98:3685-9.
- [57] Peng Y, Curtis JE, Fang X, Woodson SA. Structural model of an mRNA in complex with the bacterial chaperone Hfq. *Proc Natl Acad Sci U S A*. 2014;111:17134-9.
- [58] Sauter C, Basquin J, Suck D. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res*. 2003;31:4091-8.
- [59] Beich-Frandsen M, Vecerek B, Sjoblom B, Blasi U, Djinovic-Carugo K. Structural analysis of full-length Hfq from *Escherichia coli*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2011;67:536-40.

- [60] Zhou L, Hang J, Zhou Y, Wan R, Lu G, Yin P, et al. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature*. 2014;506:116-20.
- [61] Sobti M. Characterisation of Lsm family of RNA-binding proteins. 2008.
- [62] Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, et al. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*. 1999;96:375-87.
- [63] Manea F. Engineering synthetic Lsm rings for applications in nanotechnology. 2015.
- [64] Wason A. Investigation of Lsm Proteins as Scaffolds in Biotechnology. 2014.
- [65] Sambrook J, Russel DW. *Molecular Cloning: a laboratory manual*. New York Cold Spring Harbor Laboratory Press (Third Edition). 2001.
- [66] Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005;41:207-34.
- [67] Rosano G, Ceccarelli E. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microbial Cell Factories*. 2009;8:41.
- [68] Inoue H, Nojima H, Okayama H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*. 1990;96:23-8.
- [69] Cheung RC, Wong JH, Ng TB. Immobilized metal ion affinity chromatography: a review on its applications. *Appl Microbiol Biotechnol*. 2012;96:1411-20.
- [70] Manz A, Pamme N, Iossifidis D. *Bioanalytical chemistry*: World Scientific; 2004.
- [71] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999;112:531-52.
- [72] Oliva A, Llabres M, Farina JB. Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering. *J Pharm Biomed Anal*. 2001;25:833-41.
- [73] McPherson A. *Crystallization of biological macromolecules*. New York Cold Spring Harbor Laboratory Press. 1999.
- [74] Bergfors T. Seeds to crystals. *J Struct Biol*. 2003;142:66-76.
- [75] Jancarik J, Kim S-H. Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of applied crystallography*. 1991;24:409-11.
- [76] Wooh JW, Kidd RD, Martin JL, Kobe B. Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallographica Section D: Biological Crystallography*. 2003;59:769-72.
- [77] Hennessy DN, Narayanan B, Rosenberg JM. Automatic implementation of precise grid screens: the four-corners method. *Acta Crystallogr D Biol Crystallogr*. 2009;65:1001-3.
- [78] Manea F. Control and fabrication of ring-forming protein modules. 2011.
- [79] Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta*. 2006;1760:1304-13.
- [80] Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC. Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol*. 2011;77:4634-46.
- [81] Zaric BL, Jovanovic VB, Stojanovic SD. Non-covalent interactions across subunit interfaces in Sm proteins. *J Theor Biol*. 2011;271:18-26.
- [82] Valentin-Hansen P, Eriksen M, Udesen C. MicroReview: The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Molecular Microbiology*. 2004;51:1525-33.
- [83] Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*. 2007;372:774-97.

- [84] Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, et al. Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography*. 2005;61:1426-31.
- [85] Newman J, Fazio VJ, Lawson B, Peat TS. The C6 web tool: a resource for the rational selection of crystallization conditions. *Crystal Growth & Design*. 2010;10:2785-92.
- [86] Rupp B. *Biomolecular Crystallography: principles, practice, and applications to structural biology*. Garland Science. 2010:392.
- [87] Collaborative Computational Project N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994;50:760-3.
- [88] Sheehan D. *Physical biochemistry: principles and applications*: John Wiley & Sons; 2009.