REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
20-12-2015	Ph.D. Dissertation	-

4. TITLE AND SUBTITLE
INVESTIGATION OF LSM PROTEINS AS SCAFFOLDS IN BIONANOTECHNOLOGY

5a. CONTRACT NUMBER
W911NF-12-1-0554

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Akshita Wason

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Canterbury
20 Kirkwood Ave
Ilam
-

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
61663-CH.8

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Self-assembling materials have gained attention in the field of nanotechnology due to their potential to be used as building blocks for fabricating complex nanoscale devices. The biological world is abundant with examples of functional self-assembling biomolecules. Proteins are one such example, found in a variety of geometries and shapes. This research is focussed on the use of ring-shaped self-assembling proteins, called Lsm proteins, as componentary for applications in bionanotechnology. Lsm proteins were used because of their spontaneous association into stable rings, tolerance to mutations, and affinity to RNA. This thesis primarily focussed on the

15. SUBJECT TERMS
Lsm, tecton, nanotechnology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Juliet Gerrard
UU	UU	UU	UU		19b. TELEPHONE NUMBER +64-364-2987

Report Title

INVESTIGATION OF LSM PROTEINS AS SCAFFOLDS IN BIONANOTECHNOLOGY

ABSTRACT

Self-assembling materials have gained attention in the field of nanotechnology due to their potential to be used as building blocks for fabricating complex nanoscale devices. The biological world is abundant with examples of functional self-assembling biomolecules. Proteins are one such example, found in a variety of geometries and shapes. This research is focussed on the use of ring-shaped self-assembling proteins, called Lsm proteins, as componentary for applications in bionanotechnology. Lsm proteins were used because of their spontaneous association into stable rings, tolerance to mutations, and affinity to RNA. This thesis primarily focussed on the thermophilic Lsm? (from Methanobacterium. thermoautotrophicum) that assembles as heptameric rings.

The oligomeric state of the heptameric protein, and hence the diameter of its central cavity, was manipulated by judiciously altering appropriate residues at the subunit interface. Lsm? presented a complex set of interactions at the interface. Out of the mutations introduced, R65P yielded a protein for which SEC and SAXS data were consistent with a hexameric state. Moreover, key residues, L70 and I71, were identified that contribute to the stability of the toroid structure.

Covalent linking of rings provided nanotubular structures. To achieve this, the surface of the Lsm? ring scaffold was modified with Cys residues. This approach led to the formation of novel Lsm? nanotubes approximately 20 nm in length. Importantly, the assembly could be controlled by changing the redox conditions. As an alternative method to manipulate the supramolecular assembly, His6-tags were attached at the termini of the Lsm? sequence. The higher-order organisation of the constructs was influenced by the position of the His6-tag. The N-terminally attached His6-tag version of Lsm? showed a metal-dependent assembly into cage-like structures, approximately 9 nm across. This organisation was highly stable, reproducible, and reversible in nature.

The results presented in this thesis aid the understanding of generating complex nanostructures via in vitro self-assembly. The Lsm? rings were assembled into higher-order architectures at the quaternary level by employing protein engineering strategies. Future work is necessary to functionalise these supramolecular structures; however, this study confirms the potential role of Lsm? proteins as a molecular building block in bionanotechnology.

INVESTIGATION OF LSM PROTEINS AS SCAFFOLDS IN BIONANOTECHNOLOGY

A thesis submitted in partial fulfilment of the requirements for the

Degree of Doctor of Philosophy in Biochemistry

2014

Akshita Wason

University of Canterbury

Dedicated to the memory of Nikhita

Acknowledgements

A lot of hard work and sacrifice went into the preparation of this thesis and many people deserve to be thanked for their unfailing support and guidance. First and foremost to my supervisor at University of Canterbury, Prof. Juliet Gerrard for giving me this opportunity to work on this project and for her constant encouragement, guidance and support and not giving up on me at my lowest. At the same time, my heartfelt thank you to Assoc. Prof. Bridget Mabbutt for providing the starting materials of the project and for her patience, criticism, knowledge and faith in me.

I also sincerely thank my team of associate supervisors, Dr. Grant Pearce and Dr. Renwick Dobson, for their inputs and ideas that determined that direction of this project. A big round of thanks to Dr. Luigi Sasso, Dr. Céline Valéry, Dr. Moritz Lassé, and Dr. Hironori Suzuki for their daily advice. All this work could not be possible without our lab manager Jackie Healy, thank you for always ensuring constant supply of everything in the lab and for regular trips for TEM. Thank you to the members of MacDiarmid Institute and Biomolecular Interactions Centre for providing the opportunities and funds to support my PhD.

A special thank you to the lab members at Macquarie University, especially Francesca, for her excellent lab skills, support and regular trips for hot chocolate. Meghna, Jens, Bhumika, Heather for making me feel welcome at my various visits. My fellow colleagues at UC, Amy Philips, Amy Yewdall, Jeremy, Rishi, Arvind, Jen, Eric, Kat and Rachel for their lively discussions and giving me company during the good and bad times.

My non-university friends thank you for making Christchurch a home away from home. Tanmay, Sitara, Rishi, Dhairesh things would have been really tough without you all being there to support me through the difficult times. Thanks must also go to Mahima, Padmaja, Sush and Nilesh, for opening their hearts and home to me. The end journey would have been very difficult without your support. Aayushi and Nirokshana, even though we are in three corners of the world, a hearty thank you for the good times back then and in advance for the furure.

Lastly, sincere thanks to my parents back home. I have missed them dearly and could not have finished this journey without their encouragement. Thank you for your unfailing support, selfless love, and unshakable belief in my abilities.

Table of contents

List of Abbreviations

List is based on the abbreviations accepted by JMB

Å	angstrom
A_{280}	absorbance at 280 nm
A. fulgidus	*Archaeoglobus fulgidus*
Arg	arginine
Asn	asparagine
Au NPs	gold nanoparticles
BMOE	bis-maleimidoethane
BSA	bovine serum albumin
°C	degree Celsius
CD	circular dichroism
CV	column volume
Cys	cysteine
d	derivative
Da	Daltons
dH_2O	distilled water
D_{max}	maximal particle diameter
DMSO	dimethyl sulfoxide
dn/dc	refractive index increment
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ε_{280}	extinction coefficient
g	grams
GA	glutaraldehyde
Glu	glutamic acid
Gly	glycine
GST	glutathione S-transferase
h	hour

HcpI	protein component of type IV secretion system in *P. aeruginosa*
HEPES	N-2-hydroxyethlpiperazine-N'-2-ethane sulphonic acid
Hfq	bacterial Lsm protein
His	histidine
$I(q)$	q dependent Intensity
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_{av}	size exclusion distribution coefficient
kDa	kiloDaltons
L	litre
LB	Luria Bertani
LDS	lithium dodecyl sulfate
Lsm	Sm-like
Lys	lysine
µg	micrograms
µL	microlitre
µM	micromolar
mL	millilitre
mM	millimolar
min	minute
MOPS	3-morpholinopropanesulfonic acid
Mt Lsmα	*M. thermoautotrophicum* Lsm
MW	molecular weight
MWCO	molecular weight cut-off
nm	nanometres
NPs	nanoparticles
NTA	nitrilotriacetic acid
OD_{600}	optical density at 600 nm
P(r)	electron distance distribution function
P. aeruginosa	*Pseudomonas aeruginosa*
P. aerophilum	*Pyrobaculum aerophilum*
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer saline
PDB	Protein Data Bank

psi	pounds per square inch
R_g	radius of gyration
RNA	ribonucleic acid
RNP	ribonucleoprotein
rpm	revolutions per minute
s	second
q	scattering vector
S200	Superdex 200 matrix
SAXS	small angle X-ray scattering
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
Ser	serine
SLS	static light scattering
SmAPs	Sm-like archaeal proteins
SOC	super optimal broth with glucose added
S. pombe	*Schizosaccharomyces pombe*
S. aureus	*Staphylococcus aureus*
TBP	titanium-binding peptide
TEM	transmission electron microscopy
TMV	tobacco mosaic virus
TRAP	tRNA attenuation protein
Tris	tris(hydroxymethyl)aminoethane
U	unit
UV	ultraviolet
v/v	unit volume per unit volume
V_o	void volume
w/v	unit weight per unit volume
w/w	unit weight per unit weight
ZYM-5052	autoinduction media

Abstract

Self-assembling materials have gained attention in the field of nanotechnology due to their potential to be used as building blocks for fabricating complex nanoscale devices. The biological world is abundant with examples of functional self-assembling biomolecules. Proteins are one such example, found in a variety of geometries and shapes. This research is focussed on the use of ring-shaped self-assembling proteins, called Lsm proteins, as componentary for applications in bionanotechnology. Lsm proteins were used because of their spontaneous association into stable rings, tolerance to mutations, and affinity to RNA. This thesis primarily focussed on the thermophilic Lsmα (from *Methanobacterium. thermoautotrophicum*) that assembles as heptameric rings.

The oligomeric state of the heptameric protein, and hence the diameter of its central cavity, was manipulated by judiciously altering appropriate residues at the subunit interface. Lsmα presented a complex set of interactions at the interface. Out of the mutations introduced, R65P yielded a protein for which SEC and SAXS data were consistent with a hexameric state. Moreover, key residues, L70 and I71, were identified that contribute to the stability of the toroid structure.

Covalent linking of rings provided nanotubular structures. To achieve this, the surface of the Lsmα ring scaffold was modified with Cys residues. This approach led to the formation of novel Lsmα nanotubes approximately 20 nm in length. Importantly, the assembly could be controlled by changing the redox conditions. As an alternative method to manipulate the supramolecular assembly, His_6-tags were attached at the termini of the Lsmα sequence. The higher-order organisation of the constructs was influenced by the position of the His_6-tag. The N-terminally attached His_6-tag version of Lsmα showed a metal-dependent assembly into cage-like structures, approximately 9 nm across. This organisation was highly stable, reproducible, and reversible in nature.

The results presented in this thesis aid the understanding of generating complex nanostructures via *in vitro* self-assembly. The Lsmα rings were assembled into higher-order architectures at the quaternary level by employing protein engineering strategies. Future work is necessary to functionalise these supramolecular structures; however, this study confirms the potential role of Lsmα proteins as a molecular building block in bionanotechnology.

1 Chapter one

Introduction

1.1 Context

A role has been recognised for a number of protein assemblies in the thriving new field of bionanotechnology, where their unique quaternary geometries may be used in a number of ways; for example, as nanotubes or biocontainers. This thesis examines a self-assembling homo-oligomeric ring-shaped protein that is a part of a family of RNA chaperones. The protein was used as a building block to construct new bionanostructures. The design and solution-based structural studies have been carried out using a combination of experimental approaches such as site-directed mutagenesis, size exclusion chromatography coupled static light scattering and small angle X-ray scattering. The nanostructures obtained were examined with transmission electron microscopy.

This chapter provides a background on self-assembling protein systems, strategies used to generate novel nanostructures, and how these unique supramolecular organisations may fit into the field of bionanotechnology.

1.2 Background

The field of nanoscience holds promise to deliver the next generation of advanced technology [1; 2]. Despite being a broad field, it has witnessed tremendous growth over the past few years. It builds upon principles and expertise from diverse fields such as physics, chemistry, material science, biology and engineering. A sub-branch, bionanotechnology, is a highly interdisciplinary field that aims at developing methods to fabricate nanoscale structures with biological and abiological components that respond to their external environment [3]. There are

many examples of biological entities, such as oligonucleotides, proteins, viral capsids and cells, used for fabrication of new biomaterials [4]. Such biological components have been successfully incorporated in biosensors, bionanoarrays and nanoparticle (NP) conjugates [5]. As more and more characterisation and visualisation techniques become available, we can expect to see an increase in the variety of biomolecules used to generate functional nanostructures.

Nucleic acids such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) have been used extensively to create complex 3D structures. This success is largely attributed to their simple Watson-Crick base pairing and ease of chemical synthesis. Peptides, too, have attracted a lot of attention for fabricating new nanostructures [6]. Despite numerous examples of DNA and peptide structures, functionalisation remains a big hurdle [7]. DNA structures are also difficult to control in three dimensions [8]. Proteins, on the other hand, offer diversity in three dimensional architecture and flexibility towards modification by genetic engineering and chemical methods. They already possess complex architectures and a variety of three dimensional shapes such as rings, tubes and cages that can be used as components to create nanomachines [9]. Therefore, proteins are increasingly playing an important role as building blocks in the field of bionanotechnology and multicomponent protein systems are increasingly being used in synthetic biology [10]. They exhibit some very valuable properties such as assembly into complex supramolecular structures, interactions with specific ligands and sequence dependent function. The challenge is learning to control these interactions *in vitro* to engender useful functionality.

1.3 Bionanotechnology- an interface between biology and physical sciences

Bionanotechnology combines knowledge from life sciences and physical sciences to fabricate responsive devices that can contribute to the development of drug delivery systems, sensors and "smart" biomaterials. The field of nanotechnology has witnessed reasonable progress in fabrication of devices by lithographic techniques, but to construct a device at nanoscale still remains a challenge [11]. Thus, a biomimetic approach towards designing a nanoscale device is fast gaining momentum where principles of biological systems such as molecular self-assembly are replicated and applied to synthesise artificial nanostructures.

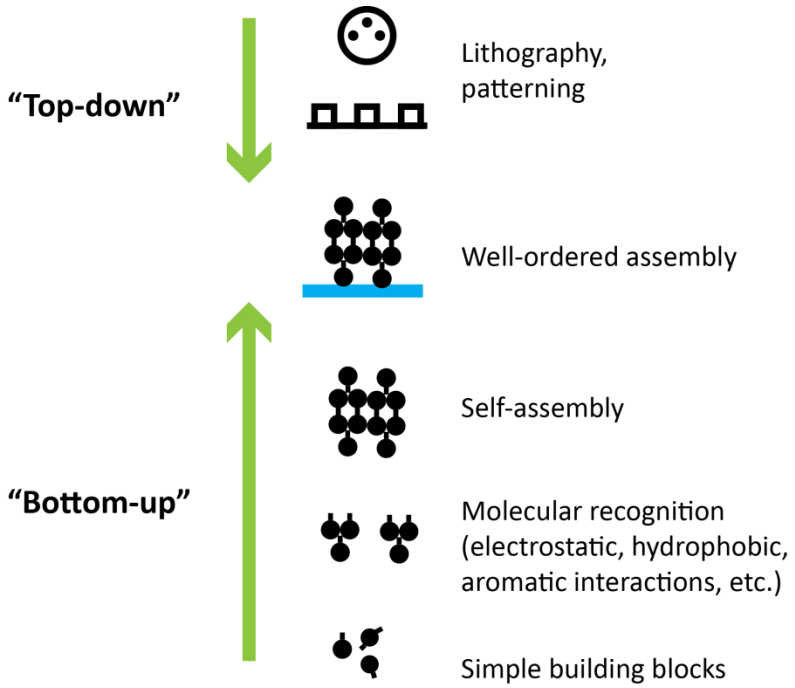


Figure 1.1| *A comparison of the "top-down" and "bottom-up" approaches applied to manufacturing of nanomaterials. The "bottom-up" approach is based on the interactions of simple building blocks which can self-assemble into well-ordered complex structures. Schematic adapted from Gazit (2007)* [12].

1.3.1 Self-assembly

Self-assembly is the phenomenon by which molecules spontaneously associate to form ordered systems through a combination of non-covalent interactions [11]. As shown in **Figure 1.1**, the relevant associations take place through hydrogen bonding, van der Waals interactions, aromatic stacking, and in some cases metal coordination, leading to the formation of stable assemblies with a defined shape and architecture [13]. Harnessing these interactions in a biomolecular self-assembling system offers promise for the development of bionanomaterials with unique properties.

1.3.1.1 Self-assembly in biology

Biological systems abound with examples of hierarchical self-assembly. From a school of fish to a bacterial swarm, all these processes involve some kind of distinct pattern formation and dynamic self-assembly [11]. Cellular processes such as mitosis, DNA replication, protein transcription and translation and protein signalling are facilitated by the phenomenon of self-assembly, making it one of the most fundamental processes to support life [14]. It involves an interplay of associations among molecules such as nucleic acids, proteins and oligosaccharides. Virtually all functions in a living cell are carried out by intricate protein machineries that assemble *in vivo* through transient interactions. These highly precise and regulated interactions give rise to molecular machines in the cells such as ATP synthase and ribosomes that are capable of performing highly sophisticated functions [15]. Viruses are supramolecules based on assembly of protein modules and nucleic acid, and form different shapes in a variety of sizes [16]. The tobacco mosaic virus (TMV) self-assembles into rod-like structures with more than 2000 protein structural units assembling around an RNA molecule to form a 300 nm long cylinder [17]. These nanoscale cylinders can be reconstituted if broken into their individual components. TMV molecules represent an excellent example of a robust

biological self-assembled system. Such biological organisations serve as an inspiration and a platform to artificially synthesise molecular structures *in vitro* based on the principles of hierarchical self-assembly [18].

1.3.1.2 In vitro self-assembly

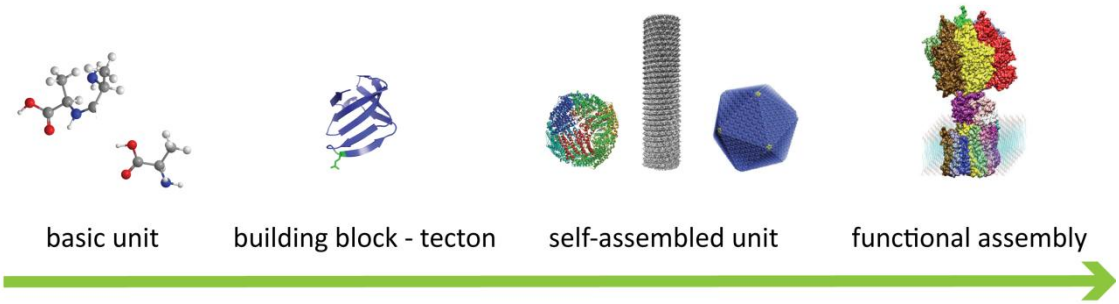
There are two ways to obtain nanostructures. The "top-down" approach essentially is breaking down or decomposing large structural systems to their constituent parts. In nanotechnology, this involves using large external devices to create nanoscale structures and includes processes such as nanolithography (**Figure 1.1**). "Bottom-up" approaches involve the interaction of atoms and molecules to self-assemble into large complex structures by means of various intermolecular interactions. It is generally believed that the "bottom-up" approach is relatively energy efficient and, in principle, cheaper for mass production of nanoscale devices [19]. The study of self-assembly can undoubtedly lead to the discovery of intriguing protein nanosystems that can have interesting applications in the synthesis of novel biomaterials and biosensors (or nanobiosensors).

In order to use biomolecules in conjunction with material science, the biological entities have to be manipulated beyond their physiological parameters. Lipids, carbohydrates, and peptides have all been explored as building blocks to create supramolecular materials [13]. Significant success has been seen for peptide systems with a self-assembling propensity which have been used in hydrogel biomaterials [20], tissue engineering, and 3D cell culture [21]. Nucleic acids offer a relatively simple mechanism of interaction and have been used for diagnostic and biosensing applications [22], but lack the ability to be functionalised and form a variety of complex 3D structures [23].

Algorithm-based design of organised supramolecular structures generates a large library of nanomaterials with distinct configurations and allows for a large combination of energetically favourable structures to be tested [24]. Huge strides have been made by this approach to generate structures with atomic level precision [25]. Baker et al. have successfully generated a variety of cage-like architectures from simple trimeric protein building blocks by using the RosettaDesign algorithm and extending it for modelling multiple protein interface interactions [26]. This approach can give rise to a variety of morphologies by using building blocks of different symmetries, combined in geometrically specific ways [27].

Nearly every protein forms transient or permanent complexes with metal ions to facilitate their structural or functional role [28]. With new methods in protein engineering, it is possible to engineer individual metal binding sites to create a metal-induced self-assembly event. For example, the monomeric component of microtubules, tubulin, crystallises as sheets on coordination with Zn^{2+} [29]. In another instance, disulfide bonds were engineered on a monomeric protein in order to mimic its quaternary structure to a cryptand-like complex with Zn-binding capacity [30]. Specific sites on the protein were targeted to generate large-ordered macromolecular structures on controlled addition of metal to the protein solution. The same protein was also induced to assemble as very stable protein sheets and arrays on addition of Zn^{2+} [31]. Thus, with various strategies of protein engineering, either by re-engineering existing quaternary assemblies or using algorithm-based strategies, a variety of functional geometries can be generated (**Figure 1.2**).

A



B

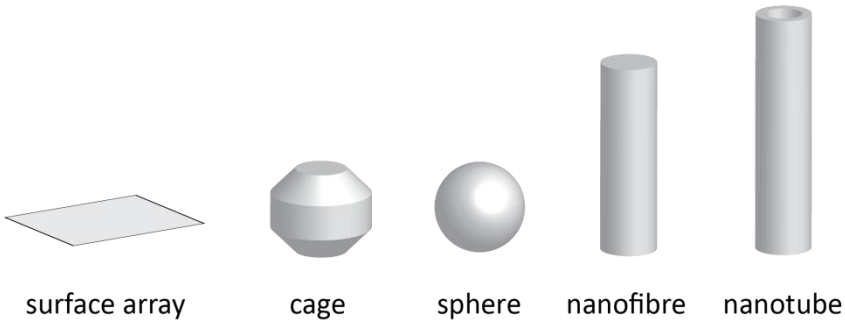


Figure 1.2| *The process of self-assembly. (A) Hierarchical organisation of simple building blocks into ordered functional assemblies found in vivo. Adapted from Howorka et al. (2007)* [32]*. (B) A schematic representation of common nanostructures. Adapted from Chopra et al. (2007)* [33].

1.4 Nucleic acid as building blocks

Oligonucleotides have been successfully employed in the creation of nucleic acid-based nanoscale assemblies including 3D nano objects, crystalline arrays and nanogrids [34]. They can be synthesised rapidly and reliably and their 3D shape can be controlled by specific oligonucleotide sequence motifs. Thus, the architecture of nucleic acid-based nanostructures can be controlled with relative ease. DNA is a simple molecule with respect to its constituent building blocks, the four deoxyribonucleotides, through which it encodes the entire genomic information of the cell. A single strand of oligonucleotide has the potential to fold into a predictable and stable 2D or 3D architecture, making DNA an excellent molecule for construction of nanoscale structures. Two-dimensional structures were created from single

stranded DNA, using a four-arm junction as the basic structural unit. These 2D objects were further reconfigured to yield three-dimensional structures and curved objects [36]. In another example, fragments of nucleic acids were used as modular building blocks to fabricate robust cuboid structures with interior cavities and tunnels [37].

Focus has now shifted towards investigating RNA as an equivalent building block for nanostructures. Recent studies show that by incorporating a K-turn motif in RNA that allows it to bend at angle of 60^{o}, the geometry of the molecule can be controlled. An equilateral triangle shaped RNA molecule was successfully synthesised that was stabilised at the corners by an RNA binding protein (L7Ae) (**Figure 1.3**). The repertoire of such ribonucleoprotein (RNP) structures can be increased by incorporating and experimenting with many other characterised RNP motifs to generate novel nanoarchitectures [35]. Strategies to functionalise these structures are beginning to emerge, but are at present limited to DNAzymes and ribozymes [38]. Due to the paucity of chemical groups available, functionalisation of nucleotide-based structures is a challenge, limiting their use as a molecular tecton.

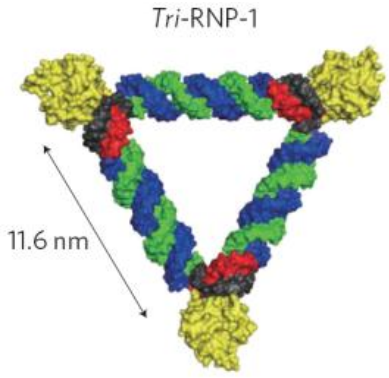


***Figure 1.3**| RNA-based nanostructure. An RNP equilateral triangle [35].*

1.5 Proteins and peptides as functional building blocks

Proteins and peptides are appealing for the synthesis of nanoscale devices because of their versatility in function and variety in their 3D structures. They can be made cheaply and quickly either by synthetic means or recombinant methods. Peptide sequences with certain motifs exhibit self-assembling properties and associate to form sophisticated structures. Extensive work has been carried out with coiled-coil motifs [39], amyloid-like structures comprised of β-strands and peptide amphiphiles [40]. Peptide supramolecular chemistry has witnessed immense progress over the years. Self-assembling peptides have been explored to create complex nanoarchitectures like nanoribbons, nanospheres, and nanotubes (**Figure 1.2B**). Peptide-based nanofibres have been applied in 3D cell culture and tissue engineering [41]. Although peptides have a more modular and robust nature, they lack the chemical complexity and hence, functionality, exhibited by proteins and are comparatively more expensive to synthesise at a large scale.

Proteins are formed by a modular approach, wherein subunits interact and assemble into well defined supramolecular architecture, through both permanent and transient bonds [42; 43]. Thus, these biomolecules offer an inherent complexity and are capable of performing a fleet of functions in biology. The quaternary structures formed by self-assembling proteins have the potential to be designed and manipulated to form structures that can be useful in nanotechnology. However, the principles of protein folding are still not fully understood, which makes predictable protein design difficult [44]. As mentioned in **Section 1.3.1.2**, algorithm-based protein design as an *in vitro* self-assembly strategy is now beginning to witness routine success. Multi-component protein nanomaterials were generated using such a strategy [45].

The level of structural complexity increases with an increase in the number of components involved. The number of interactions that drive the assembly offer various levels of control over the process, thereby providing precision and directionality to construct nanostructures. There are some proteins which exhibit excellent self-assembling properties and can take the shape of cages, rings and tubes.

The first protein architecture to be studied as a nanostructure was a cage protein. Ferritin is a 24 subunit ubiquitous intracellular protein that stores iron in a microcrystalline non-toxic form and releases it in a controlled manner [46]. Small molecules such as metal ions can diffuse into the interior cavity via acidic channels and accumulate, resulting in the formation of NPs. The dimensions of the NPs are controlled by the size of the cage structure. Such a protein cage architecture was used as a 'container' for synthesis of cobalt oxide minerals (Co_3O_4 and Co(O)OH) and iron oxide [47].

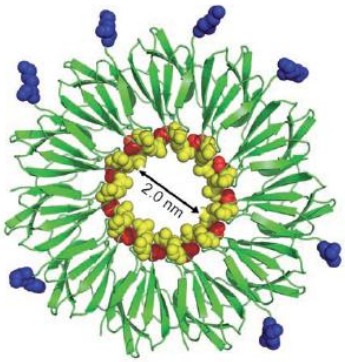
The bacterial tRNA attenuation protein (TRAP) assembles as an 11-mer ring with a pore diameter of approximately 2.0 nm [48]. It is involved in the regulation of the *trp* operon encoding for enzymes involved in tryptophan synthesis in bacteria [49]. This protein is heat-stable and reasonably tolerant to mutations, making it a very convenient protein to work with *in vitro*. Heddle et al. constructed a modified protein ring by linking three to four monomers which resulted in the formation of a toroid assembly with 12 subunits rather than an 11-mer as found naturally [50]. The authors further created nanodot binding modules by replacing an unconserved arginine (Arg) residue in the inner pore with a cysteine (Cys) residue, thus artificially introducing a thiol moiety in the pore of protein. With only one Cys residue present, the authors were able to bind gold NPs in the inner pore. This protein was further functionalised by attaching a titanium-binding peptide (TBP) on the termini, thereby immobilising these protein rings on an inorganic surface [51] (**Figure 1.4A**). With further

modification to the protein ring, this group also demonstrated that TRAP protein can assemble as nanotubes approximately 1 µm in length through disulfide and hydrophobic interactions [52] (**Figure 1.4B**).

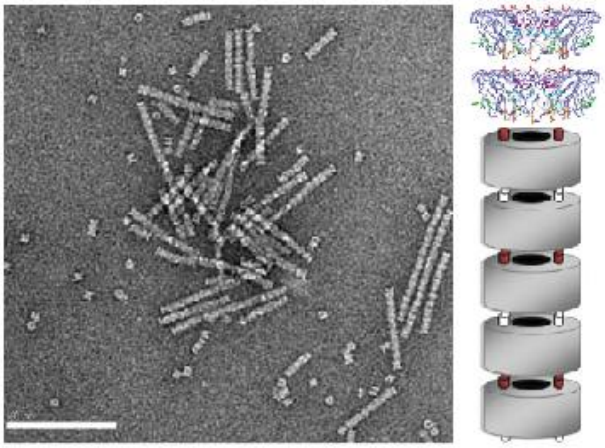
Stable protein 1 (SP1) is a highly thermostable protease resistant homo-oligomer that forms a dodecamer, first isolated from *Aspen populus* [53]. In a series of experiments, SP1 was shown to be a potential candidate for use as a building block to generate nanostructures. The enzyme glucose oxidase (GOx) was linked to the SP1 subunit that subsequently assembled as a nanotube with GOx protruding out [54]. It was further shown to form regular arrays and interact with gold NPs, successfully forming gold nanowires [55]. More recently, a gold NP conjugate of SP1 linked to GOx enzyme was layered on an electrode surface by means of dithiol bridging units [56]. The SP1-NPs hybrid was also successfully used as a component of a basic logic circuit [57].

Another type of protein architecture with growing interest is protein tubes, as they can form containers for controlled drug release and also have applications in electronics. [58; 59; 60]. A recent example of such application is HcpI protein from *Pseudomonas aeruginosa*. HcpI is a small 17.4 kDa protein that is a secreted component of the type IV secretion system. It adopts a homo-hexameric ring-shaped structure with an outer diameter of 9 nm and inner diameter of 4.0 nm [61]. By introducing Cys residues on the top and bottom surface of the ring, the protein nanotube was stabilised by disulfide bonds (**Figure 1.4C**). The tube was capped with HcpI muteins that had an attached epitope, thus creating an enclosed elongated cavity which could function as a nanocapsule [62]. Biophysical characterisation of the HcpI protein showed that its quaternary structure can be assembled and disassembled without disrupting the secondary structure by using detergents, thereby providing a control over its assembly process. This can be further exploited for fabrication of nanocontainers [63].

A



B



C

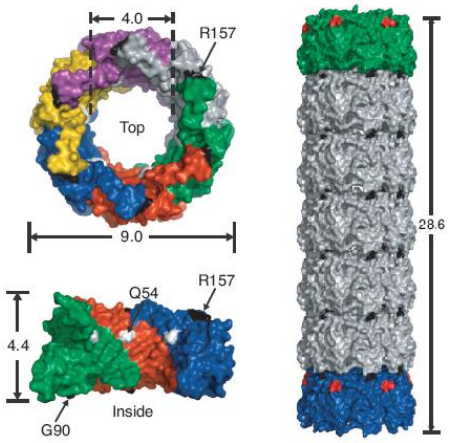


Figure 1.4| *(A) Crystal structure of TRAP protein showing the C-terminal region in blue and Arg residue in yellow* [51]*. (B) TEM of mutant TRAP protein tubes in presence of a weak reducing agent* [52]*. (C) Hcp1 protein genetically modified to assemble into tubules* [62]*.*

Well-ordered arrays can also be formed by capturing inorganic materials into the central cavity of such protein rings. An example illustrating this is HSP60 protein from *Sulfolobus shibatae*, in which the chaperone proteins form a barrel-shaped structure. By utilising the reactive thiol group of the incorporated Cys residues, gold NPs as well as semiconductor materials such as CdSe–ZnS were attached to the central cavity [64].

Another strategy used to alter the functions of self-assembled protein structures is to incorporate peptides that have different binding capabilities towards inorganic materials. There have been extensive studies on peptides that bind to metals such as gold [65] and silver [66]. Filamentous temperature sensitive protein Z (FtsZ) served as a template for fabrication of inorganic materials. By incorporating peptides with established affinities towards gold, silver and nickel ions, protein hybrid/inorganic materials were synthesised [67].

NPs can also be used as a template for assembly of protein architectures. Proteins with a His_6-tag bind to Au-NPs derivitised with nitrilotriacetic acid (NTA), thereby assembling as monomers, dimers, trimers, and spherical shells that can be controlled by varying the NPs size [68]. In another example, the assembly TMV coat protein was controlled by attachment of a His_6-tag which caused it to assemble into disks, tubes, arrays and fibres [17].

From the examples of protein tectons mentioned above, the criteria required for a successful building block is spontaneous assembly into its quaternary form, a robust nature, stability over a range of buffer conditions, and potential for functionalisation. Although a plethora of proteins are available to be studied, their inherent complexity brings with it a new level of difficulty to manipulate them into useful architectures. Hence, the use of proteins as tectons is not a trivial task. This thesis focuses on one such protein system, Lsm (like-Sm) proteins that show various levels of self-assembly, exhibit oligomeric plasticity and have thermostable

versions available. These features are essential to adapt Lsm proteins as tectons for further downstream applications in nanotechnology.

1.6 Lsm proteins

Lsm proteins derive their name from the previously characterised Sm family of proteins because of their structural similarity [69]. Sm proteins are a group of seven polypeptides that are found associated with several small nuclear RNAs (snRNA). These form heteroheptameric complexes on binding to a uracil-rich consensus region of snRNAs [70]. These ribonucleoprotein complexes are involved in mRNA splicing and have functions in post-transcriptional mRNA modification in eukaryotes [71].

In eukaryotes, Lsm proteins assemble as heteroheptameric assemblies and assist in the RNA processing events in the nucleus and cytoplasm. Similar to Sm proteins, Lsm complexes were named Lsm1 to Lsm8. Out of these polypeptides, Lsm2 to Lsm8 were identified to be associated with U6 snRNA in the nucleus and found to be necessary for stability of the complex [72]. On the other hand, Lsm1-7 is a cytoplasmic heteroheptameric complex which is involved in mRNA degradation events. Lsm proteins have also been shown to be involved in processing of tRNAs, small nucleolar RNAs (snoRNAs) and pre-rRNAs, basically acting as a chaperone protein that facilitates efficient association of pre-tRNA with their substrates [73]. **Figure 1.5** shows the location of Lsm proteins in the cell and the various complexes formed by Lsm polypeptides in association with ribonucleoproteins (RNPs).

Lsm proteins are also present in Archaea, forming homomeric complexes [74]. It is thought that archaeal Lsm proteins represent a primitive form of Sm proteins of eukaryotic snRNPs. These protein groups may have arisen through gene duplication events followed by acquiring their distinct function and cellular location [75]. It is inferred that Lsm homologues in Archaea bind to small RNA and tRNA [76]. Proteins with structural similarity to Sm proteins are also found

in Bacteria. The first protein to be identified was Hfq in *Escherichia coli*, which forms a homo-hexameric ring complex [77]. It was found to be involved in a number of cellular processes including the regulation of translocation of mRNA and its degradation [78].

A

B

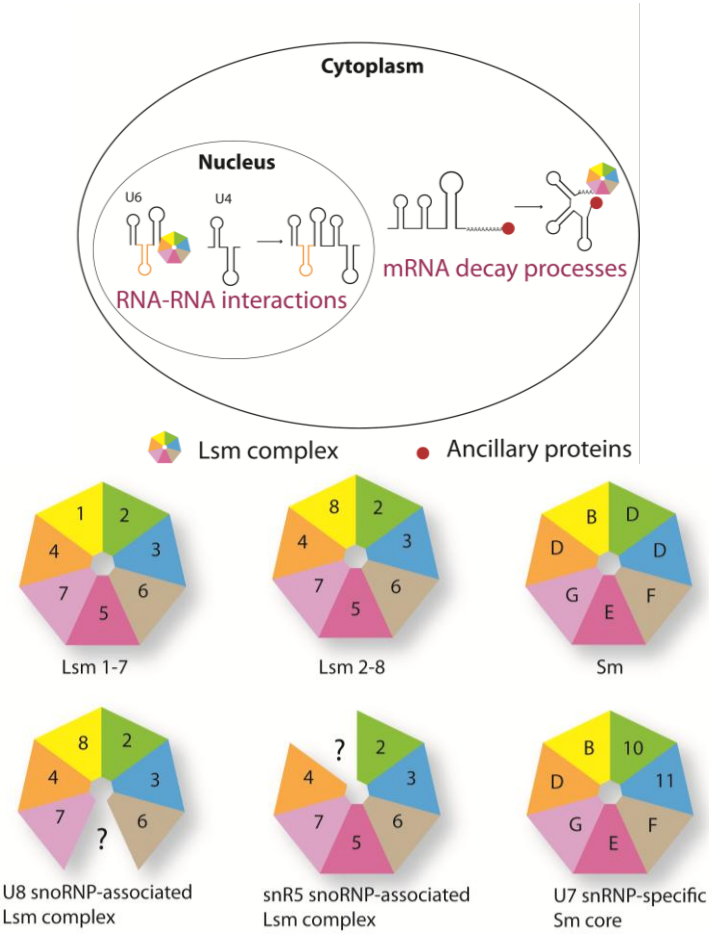


Figure 1.5| *Schematic showing location of Lsm proteins and their basic cellular functions. (B) Schematic of heteroheptamers of Lsm and Sm found in yeast and vertebrate cells [79]. Depictions of U8 snoRNA-associated complex and snR5 snoRNP where the complex could be a hexamer [80]. U7 snRNP Sm core depicting interaction of Sm and Lsm monomers [81]. Image adaped from Wilusz et al* (2005)[79].

1.7 Sequence and structure of Lsm monomers

The Lsm polypeptide sequence stretches for approximately 80 amino acids with the molecular weight of proteins varying from 8-25 kDa. The sequences are depicted in **Figure 1.6**, within each, a bipartite consensus sequence (designated Sm1 and Sm2 motifs) can be identified. These motifs arise from strands β1-β3 and β4-β5 of the core β-sheet structure, respectively. A variable stretch of residues between these conserved segments is created by a surface-exposed interconnecting loop [82]. The common feature of the Lsm monomer is five strongly bent anti-parallel β-strands capped by a short N-terminal α-helix [83]. Out of the five β-strands, the first and fifth are shorter in length as compared to the second, third and fourth, which are longer and curved. The Sm1 motif is comprised of an N-terminal α-helix and β1-, β2- and β3- strands followed by a variable loop sequence, whereas the Sm2 motif consists of β4- and β5- strands [84]. The core of the Lsm monomer is hydrophobic in nature and in order to bury this core, the Lsm monomers interact at the β4/β5 interface of adjacent subunits, thereby forming a closed ring-shape architecture (**Figure 1.7C**). The interface between the subunits involves β-sheet extension hydrogen bonding, with reciprocal interaction of β4 of one subunit with β5 of the neighbouring subunit. The resulting assembly is a highly stable ring structure with one face consisting of N-terminal α-helices and the opposite face containing the variable loop4 region (**Figure 1.7B**).

The Lsm β-sheet is a highly curved feature of the Lsm fold, making the monomers nearly elliptical or U-shaped in cross-section. The polypeptide backbone can adopt this curvature due to specific Glycine (Gly) residues serving as pivot points to support the shape of the β-sheet. A characteristic feature of the monomer structure is the variation in loop4. This loop links β3- and β4-strands and varies from just a few residues in bacterial and archaeal

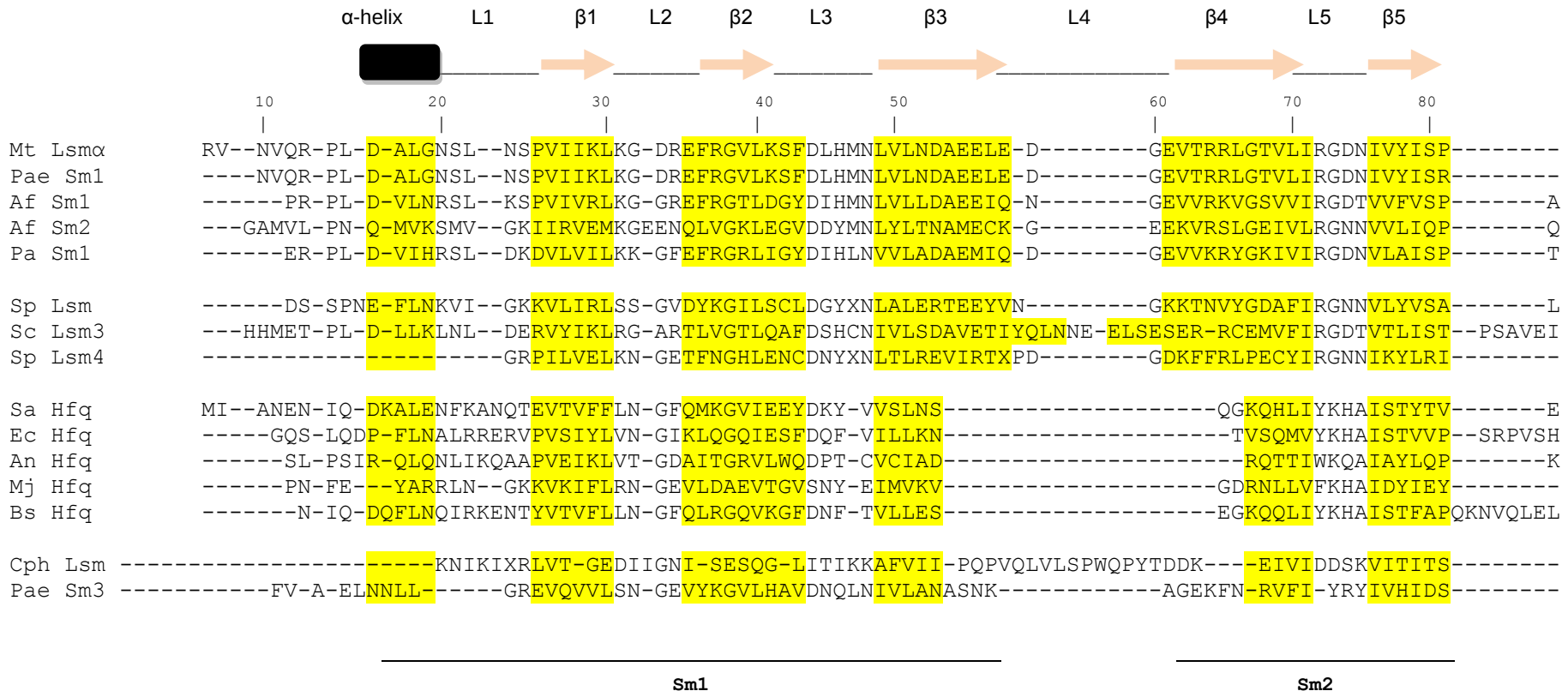
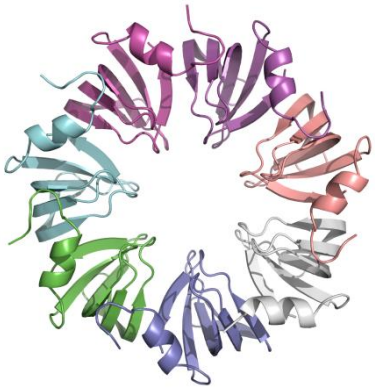


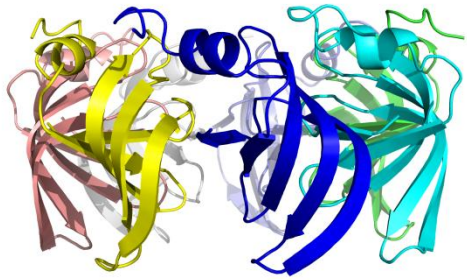
Figure 1.6| *Structure-based alignment of Lsm protein sequences. Lsm proteins are from the following species M. thermoautotrophicum (Mt), A. fulgidus (Af), Pa (P. abyssi), P. aerophilum (Pae), S. cerevisiae (Sc), S. pombe (Sp), S. aureus (Sa), E. coli (Ec), Anabena sp. (An), M. janaschii (Mj), B. subtilis (Bs) cyanophage (Cph). Secondary structure assignment is based on the crystal structure of Mt Lsmα* [85].

homologs to several residues in eukaryotic homologs. The orientation of the residues towards the outside on loop4 make them prominent structural features on the distal face or loop face of the ring scaffold (**Figure 1.7B**). The amino acid variation also modulates the RNA-binding properties of the Lsm ring. These features make loop4 a strong target for engineering. Thus, the Lsm cyclic oligomer provides four surfaces i.e. helix face, loop face, central pore, and edge of ring, that can be altered to influence its higher order organisation.

A



B



C

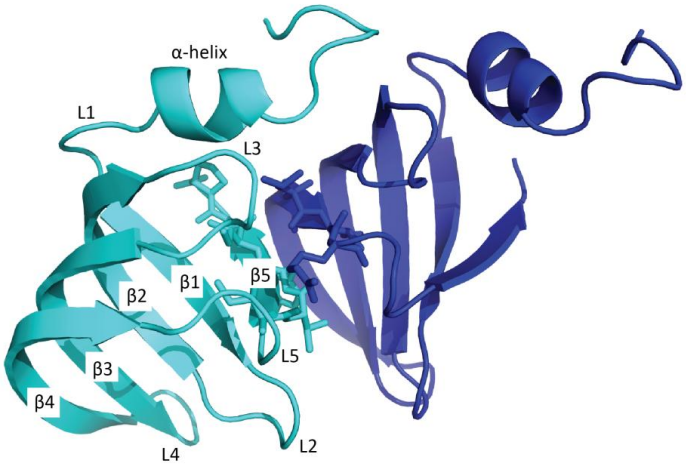


Figure 1.7| *The Lsm fold. Ribbon structure of Lsmα from Methanobacterium thermoautotrophicum (Mt Lsmα, PDB ID 1I81)* [74] *showing the self-assembled heptameric ring structure. (A) Top view. (B) Side view. (C) Interaction of β-strands from two Mt Lsmα subunits within the heptameric complex, Chain A is represented in cyan and Chain B in blue. Residues involved in hydrophobic packing at the dimer interface (Chain A: Ile27, Val77, Tyr78 of chain A; Chain B: Leu 30, Phe36, Leu66, Val69, Ile71) are shown in stick representation.*

1.8 Oligomeric plasticity of Lsm rings

Several Lsm proteins have been crystallised, providing high-resolution views of the ring morphology of the assembled molecules. As shown in **Figure 1.8**, the Lsm rings range from 6 nm to 8 nm in outer diameter and contain a central cavity of 0.6 nm to 1.5 nm. Since the inherent function of Lsm protein is to bind RNA, some crystal structures have been solved in the presence of the nucleic acid, which reveals the precise residues involved in RNA binding. This is an added advantage for functionalising the molecular tecton with nucleic acid binding properties.

Despite the conservation in the amino acid sequence and the three dimensional structures, Lsm proteins exhibit a high degree of oligomeric plasticity. The majority of crystal structures of Lsm are hexameric and heptameric protein assemblies, although the recent discovery of pentameric and octameric forms, have broadened the spectrum of quaternary structures of Lsm proteins. While these forms are prepared through recombinant means, it does raise the possibility of such oligomers existing *in vivo* as fully functional forms, further suggesting the presence of a variety of multimeric forms. Studies have shown that Lsm assemblies may be relatively dynamic in solution, providing capacity to engage in alternative protein partnerships and other stable groupings [86].

Although the physicochemical and stereochemical cause of this plasticity is unknown, it makes the Lsm system an interesting candidate for use as a molecular building block in nanotechnology. Its series of oligomeric states (n = 5,6,7,8...) provides various levels at which the molecule can be tuned. For example, to obtain different pore sizes, the inter-subunit interface can be engineered to potentially alter the oligomeric state of the ring scaffold. Thus, the oligomeric plasticity is an attractive quality of Lsm proteins for targeting it as a modular building block.

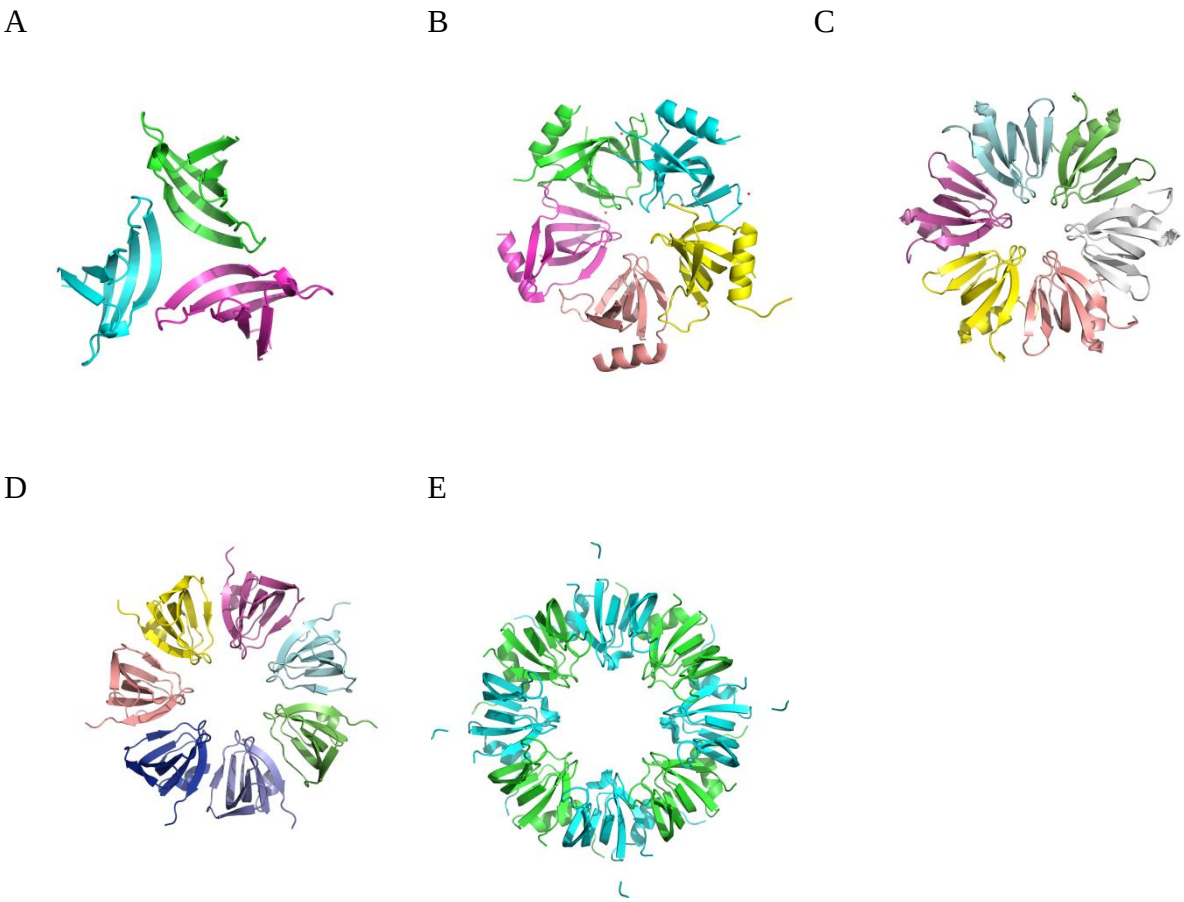


Figure 1.8| *Selected crystal structures solved for Lsm assemblies. (A) Trimer adopted by Lsm4 fragment, S. pombe (PDB ID 4EMH). (B) Pentamer, cynanophage (PDB 3BY7) 6.0 nm ring, 0.9 nm pore. (C) Hexamer, E. coli Hfq (PDB 1HK9) 6.5 nm ring, 1.0 nm pore. (D) Heptamer, M. thermoautotrophicum Lsmα (PDB 1I81) 6.5 nm ring, 1.5 nm pore. (E) Octamer, S. cerevisiae Lsm3 (PDB 3BW1) 7.5 nm ring, 1.5 nm pore.*

1.9 Higher order structures

Lsm proteins also form higher order oligomers assembling either as a 14-mer (in the case of the heptamer) or 16-mer (in the case of the octamer) by stacking two rings together. They may interact between helical faces or between loop to loop stack or by helix to loop stacking [87; 88]. Certain Lsm proteins also show fibrogenesis in the absence of RNA [89]. Although both bacterial and archaeal Lsm have been shown to form fibril structures, biophysical characterisation indicates different hierarchical assembly of the two. The fibre models for *E. coli* Hfq indicate contacts between the hexameric protein rings mediating formation of a protofilament [89] in a multilayer arrangement where each layer is made up of six hexameric rings. Alternatively, the crystal structure of *Mt* Lsmα reveals a tubular arrangement of homo-heptamers via head to tail stacking [90] in a directional manner. **Figure 1.9** shows the various quaternary structures and higher order organisations adopted by different Lsm proteins. The exact biological role of these polymers, if any, remains elusive. However, they are ideally suited as self-assembling proteins for bionanotechnology.

1.10 Yeast Lsm protein – Lsm3

Lsm proteins found in yeast systems assemble as heteromeric complexes. Although recent discoveries of homomeric complexes in the case of yeast SmF, Lsm3 and Lsm5 from *Cerevisiae parvum*, indicate that other oligomeric organisations of Lsm proteins are possible in non-native conditions. The Lsm proteins in yeast exist as complexes of heteroheptamers. However, subunit three alone, Lsm3, assembles *in vitro* as a stable ring of eight units [87]. Lsm3 exhibits many typical features of Lsm proteins, but there are marked changes in the geometry of the β4/β5 interface between two subunits, which is believed to play a vital role in determining the organisation of the Lsm3 ring structure [87]. This relationship between the sub-

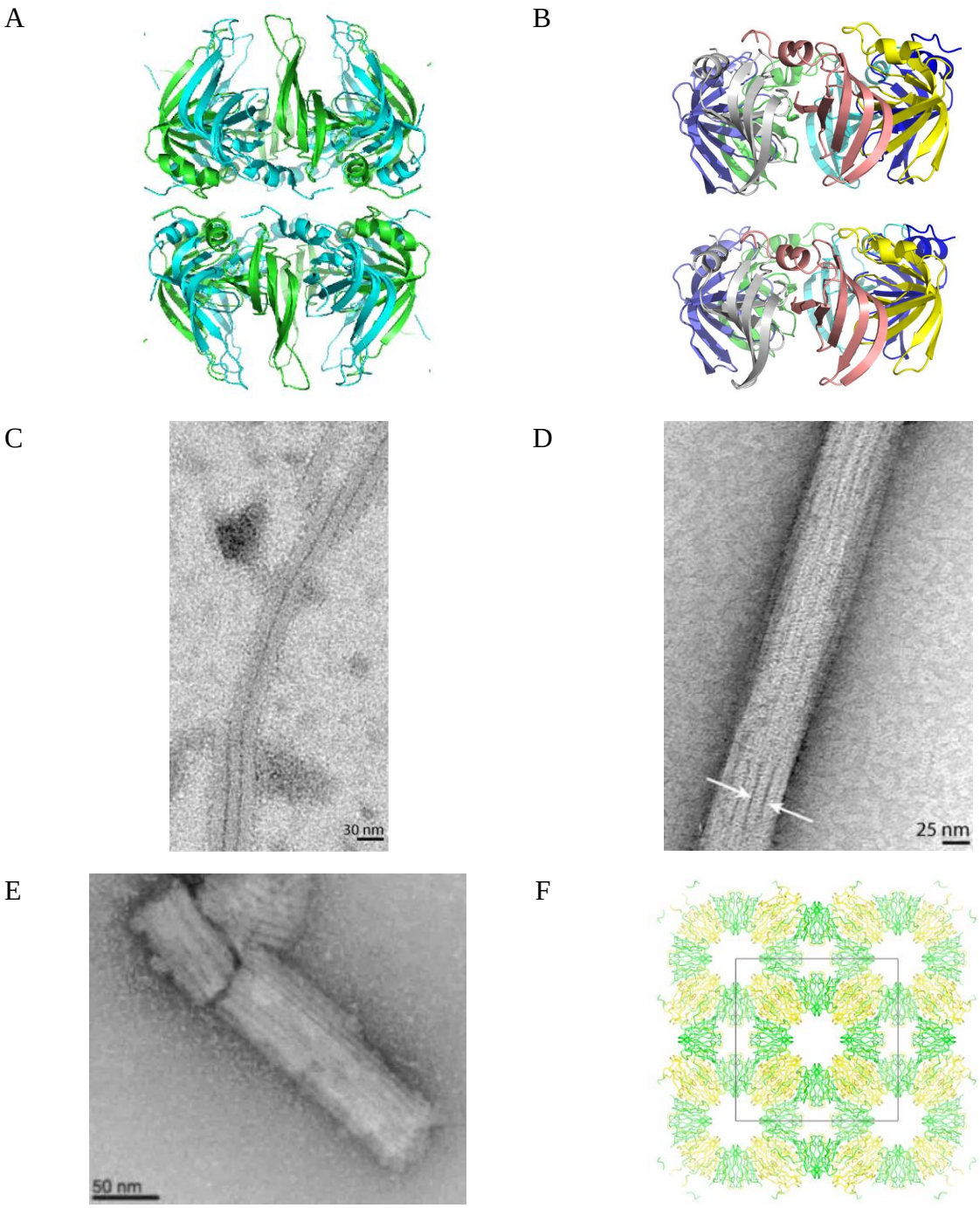


Figure 1.9| *Higher order quaternary structures adopted by Lsm proteins. (A) Face-to-face stacking seen in Lsm3 (PDB ID 3BW1) crystal* [87]*. (B) Head-to-tail stacking observed in Mt Lsmα crystal packing (PDB ID 1I81). (C) Transmission EMs of E. coli Hfq showing bundles of unbranched fibres* [91]*. (D) TEM showing polymerisation Mt Lsmα into polar fibrils* [89]*. (E) Well ordered bundles formed by Pae Lsm* [91]*. (F) Crystal lattice of Lsm3 showing interaction of the rings.*

unit interface angle and the number of subunits forming the ring has also been highlighted by the crystal structure of the Lsm5/6/7 complex of *Shizosaccharomyces pombe* [72].

The crystal structure of Lsm3 shows the presence of coaxial stacks in which the helical faces of two rings interact by hydrogen bonds and salt bridges, giving rise to a 16-mer. Furthermore, the C-terminus of Lsm3 interacts with β-strands of other rings, as seen from the crystal structure, giving rise to arrays [87]. The C-terminal residues of Lsm3 extend as an additional stretch of β-strand within the Lsm3 subunits, which runs antiparallel to the twisted β3/L4/β4 ribbon from a neighbouring octamer, resulting in the formation of hydrogen bonds with residues of the β3-strand. This insertion results in formation of an array in the crystal lattice of Lsm3.

1.11 Archaeal Lsm proteins

Sm homologs were discovered in Archaea through sequence analysis. Although no definite biological role is assigned to archaeal Lsm proteins, they have been shown to interact with RNA molecules *in vitro*, implicating a biological role in RNA metabolism and mediating RNA protein interactions [76; 92]. Archaeal Lsm proteins do not possess an elongated N-terminus, C-terminus or loop region. Sm1 and Sm2 motifs essentially occupy most of the sequence, as opposed to other eukaryotic Lsm proteins that have extended end terminal residues [74; 83]. An atypical structure has been reported in *Pyrobaculum aerophilum,* in which the Sm domain extends up to 60 residues at the C-terminus and it assembles as a 14-mer [88]. *Mt* Lsmα, from thermophile *M. thermoautotrophicum,* consists of an 81 amino acid polypeptide sequence, with each monomer consisting of a highly hydrophobic core due to the highly curved nature of β-sheets. It self-assembles as a toroid with seven identical subunits having a central pore diameter of 1.0-1.5 nm [74]. *Mt* Lsmα does contain an unusually long N-terminal sequence when compared with other archaeal Lsm proteins with the α-helix starting

at residue 13. These extended N-terminal residues are not essential for ring formation [74], making this region an ideal site for further modification by genetic or chemical means. For the purpose of this thesis, *Mt* Lsmα will be referred to as Lsmα in subsequent sections.

1.12 Significance and potential outcomes

The field of bionanotechnology is a thriving new area of research, fast creating a niche in life sciences and nanotechnology. A bottom-up approach for fabrication of nanodevices offers more control and precision on the nanometre scale. Producing complex nanoscale devices presents many challenges, as the components of such devices require special properties. Biomolecules offer unique features and advantages from the assembly standpoint. With self-assembling biological systems, it is now possible to create elaborate patterns and arrays on surfaces with better control over the dimension and position of the components [93]. To customise the fabrication of molecular biomaterials, it is necessary to understand the process of self-assembly of the biomolecular units into their higher order states. By choosing a protein system as a module for the development of molecular tectons, this thesis aims to provide a new protein tecton, as well as useful insight into the assembly process and present strategies to control the supramolecular organisation of these protein nanocomponents.

1.13 Scope of the thesis

Lsm proteins readily form stable ring structures with specific binding capacity for RNA. With different oligomeric states present, available crystal structures, and thermostable versions to choose from, Lsm proteins could be customised into new componentary suitable for fabrication of new biomaterial. Therefore, Lsm proteins were chosen as potential candidates to fabricate novel biological nanostructures, with its readily accessible sites for addition of functional moieties and potential to tune the pore size.

In this project, Lsm proteins have been investigated as building blocks to create novel nanostructures. The study focused primarily on a thermophilic protein from *M. thermoautotrophicum*, which assembles as a heptamer. The Lsm proteins were well studied in biology due to their involvement in a plethora of RNA processing events in the cell. Atomic level resolution was provided by the crystal structures to elucidate the precise structural features of Lsm proteins. Although Sm/Lsm proteins have been studied for their biochemical and evolutionary structural properties, their characteristic ring-shaped architecture lends itself to the fabrication of novel nanostructures that, to date, had not been explored.

The specific aims of the research experiments were:

i. To examine the behaviour of Lsm3 and Lsmα and determine the system with properties better suited for use as a tecton.
ii. To carry out site-directed mutagenesis to explore the interface between the monomers of the Lsmα ring module.
iii. To engineer the cyclic oligomers into higher order organisations and determine their structure by physical methods.
iv. To characterise the nature of the quaternary structure of the discrete species engineered and the interface muteins formed in solution.

1.14 Thesis overview

Chapter 2 outlines the experimental procedures used throughout this body of work.

Chapter 3 details the design and strategies used for investigation of Lsmα subunit interface. It discusses the mutations generated and the potential change they induced in the oligomeric state of Lsmα. The chapter also discusses the trial of Lsm3 protein from a eukaryotic source.

Chapter 4 presents the strategy to induce higher order forms of the circular oligomer as nanotubes and discusses methods to control its organisation to achieve a new hierarchical level of assembly.

Chapter 5 focuses on the surface modification of the Lsmα scaffold, detailing the biophysical characterisation of the variants generated. It entails the investigation of the Lsmα supramolecular structure and its suitability to be used as a molecular tecton.

Chapter 6 summarises the major findings of this thesis, and future directions to be pursued from this work.

1.15 References

1. Bucke, C. (2005). Bionanotechnology—lessons from nature. . *Journal of Chemical Technology and Biotechnology* **80**, 964-965.
2. Drexler, K. E. (1992). Nanosystems-Molecular machinery, manufacturing, and computation. *John Wiley & Sons, Inc.*
3. Niemeyer, C. M. & Mirkin, C. A. (2004). Nanobiotechnology: concepts, applications and perspectives. *John Wiley & Sons, Inc.*
4. Kohli, P. & Martin, C. R. (2005). Smart nanotubes for biotechnology. *Current Pharmaceutical Biotechnology* **6**, 35-47.
5. van Rijn, P. & Boker, A. (2011). Bionanoparticles and hybrid materials: tailored structural properties, self-assembly, materials and developments in the field. *Journal of Materials Chemistry* **21**, 16735-16747.
6. Cui, H., Webber, M. J. & Stupp, S. I. (2010). Self-assembly of peptide amphiphiles: From molecules to nanostructures to biomaterials. *Peptide Science* **94**, 1-18.
7. Jaeger, L. & Chworos, A. (2006). The architectonics of programmable RNA and DNA nanostructures. *Current Opinion in Structural Biology* **16**, 531-543.
8. Rothemund, P. W. K. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302.
9. Heddle, J. G. (2008). Protein cages, rings and tubes: useful components of future nanodevices? *Nanotechnology, Science and Applications* **1**, 67-78.
10. Grünberg, R. & Serrano, L. (2010). Strategies for protein synthetic biology. *Nucleic Acids Research* **38**, 2663-2675.
11. Whitesides, G. M. & Grzybowski, B. (2002). Self-assembly at all scales. *Science* **295**, 2418-21.
12. Gazit, E. (2007). Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chemical Society Reviews* **36**, 1263-1269.
13. Mendes, A. C., Baran, E. T., Reis, R. L. & Azevedo, H. S. (2013). Self-assembly in nature: using the principles of nature to create complex nanobiomaterials. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* **5**, 582-612.
14. Elemans, J. A. A. W., Rowan, A. E. & Nolte, R. J. M. (2003). Mastering molecular matter. Supramolecular architectures by hierarchical self-assembly. *Journal of Materials Chemistry* **13**, 2661-2670.

15. Nakamoto, R. K., Baylis Scanlon, J. A. & Al-Shawi, M. K. (2008). The rotary mechanism of the ATP synthase. *Archives of Biochemistry and Biophysics* **476**, 43-50.
16. Rong, J., Niu, Z., Lee, L. A. & Wang, Q. (2011). Self-assembly of viral particles. *Current Opinion in Colloid & Interface Science* **16**, 441-450.
17. Bruckman, M., Soto, C., McDowell, H., Liu, J., Ratna, B., Korpany, K., Zahr, O. & Blum, A. (2011). Role of hexahistidine in directed nanoassemblies of tobacco mosaic virus coat protein. *ACS Nano* **5**, 1606-1622.
18. Channon, K., Bromley, E. H. C. & Woolfson, D. N. (2008). Synthetic biology through biomolecular design and engineering. *Current Opinion in Structural Biology* **18**, 491-498.
19. Whitesides, G. M. & Boncheva, M. (2002). Beyond molecules: self-assembly of mesoscopic and macroscopic components. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 4769-4774.
20. Mahler, A., Reches, M., Rechter, M., Cohen, S. & Gazit, E. (2006). Rigid, self-assembled hydrogel composed of a modified aromatic dipeptide. *Advanced Materials* **18**, 1365-1370.
21. Lakshmanan, A., Zhang, S. & Hauser, C. A. (2012). Short self-assembling peptides as building blocks for modern nanodevices. *Trends in Biotechnology* **30**, 155-165.
22. Krishnan, Y. & Simmel, F. C. (2011). Nucleic acid based molecular devices. *Angewandte Chemie International Edition* **50**, 3124-3156.
23. Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F. & Shih, W. M. (2009). Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414-418.
24. Saven, J. (2010). Computational protein design: Advances in the design and redesign of biomolecular nanostructures. *Current Opinion in Colloid & Interface Science* **15**, 13-30.
25. Mills, J. H., Khare, S. D., Bolduc, J. M., Forouhar, F., Mulligan, V. K., Lew, S., Seetharaman, J., Tong, L., Stoddard, B. L. & Baker, D. (2013). Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *Journal of the American Chemical Society* **135**, 13393-13399.

26. King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O. & Baker, D. (2014). Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108.
27. King, N. P. & Lai, Y.-T. (2013). Practical approaches to designing novel protein assemblies. *Current Opinion in Structural Biology* **23**, 632-638.
28. Song, W. J., Sontz, P. A., Ambroggio, X. I. & Tezcan, F. A. (2014). Metals in protein–protein interfaces. *Biophysics* **43**, 409-431.
29. Gaskin, F. & Kress, Y. (1977). Zinc ion-induced assembly of tubulin. *Journal of Biological Chemistry* **252**, 6918-6924.
30. Medina-Morales, A., Perez, A., Brodin, J. D. & Tezcan, A. F. (2013). *In vitro* and cellular self-assembly of a Zn-binding protein cryptand via templated disulfide bonds. *Journal of the American Chemical Society* **135**, 12013-12022.
31. Brodin, J. D., Carr, J. R., Sontz, P. A. & Tezcan, F. A. (2014). Exceptionally stable, redox-active supramolecular protein assemblies with emergent properties. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2897-2902.
32. Howorka, S. (2007). Creating regular arrays of nanoparticles with self-assembling protein building blocks. *Journal of Materials Chemistry* **17**, 2049-2053.
33. Chopra, N., Gavalas, V. G., Bachas, L. G. & Hinds, B. J. (2007). Functional one-dimensional nanomaterials: applications in nanoscale biosensors. *Analytical Letters* **40**, 2067-2096.
34. Linko, V. & Dietz, H. (2013). The enabled state of DNA nanotechnology. *Current Opinion in Biotechnology* **24**, 555-561.
35. Ohno, H., Kobayashi, T., Kabata, R., Endo, K., Iwasa, T., Yoshimura, S. H., Takeyasu, K., Inoue, T. & Saito, H. (2011). Synthetic RNA-protein complex shaped like an equilateral triangle. *Nature Nanotechnology* **6**, 116-120.
36. Han, D., Pal, S., Yang, Y., Jiang, S., Nangreave, J. & Liu, Y. (2013). DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412-1415.
37. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. (2012). Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177-1183.
38. Liu, Y., Lin, C., Li, H. & Yan, H. (2005). Aptamer-directed self-assembly of protein arrays on a DNA nanostructure. *Angewandte Chemie International Edition* **117**, 4407-4412.

39. Mason, J. M. & Arndt, K. M. (2004). Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* **5**, 170-176.

40. Valéry, C., Artzner, F. & Paternostre, M. (2011). Peptide nanotubes: molecular organisations, self-assembly mechanisms and applications. *Soft Matter* **7**, 9583-9594.

41. Rajagopal, K. & Schneider, J. P. (2004). Self-assembling peptides and proteins for nanotechnological applications. *Current Opinion in Structural Biology* **14**, 480-486.

42. Main, E. R. G., Lowe, A. R., Mochrie, S. G. J., Jackson, S. E. & Regan, L. (2005). A recurring theme in protein engineering: the design, stability and folding of repeat proteins. *Current Opinion in Structural Biology* **15**, 464-471.

43. Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O. & Schreiber, G. (2005). The modular architecture of protein–protein binding interfaces. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 57-62.

44. Grueninger, D., Treiber, N., Ziegler, M. O. P., Koetter, J. W. A., Schulze, M. S. & Schulz, G. E. (2008). Designed protein-protein association. *Science* **319**, 206-209.

45. King, N., Sheffler, W., Sawaya, M., Vollmar, B., Sumida, J., André, I., Gonen, T., Yeates, T. & Baker, D. (2012). Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171-1175.

46. Liu, X. & Theil, E. C. (2005). Ferritins: dynamic management of biological iron and oxygen chemistry. *Accounts of Chemical Research* **38**, 167-175.

47. Allen, M., Willits, D., Young, M. & Douglas, T. (2003). Constrained synthesis of cobalt oxide nanomaterials in the 12-subunit protein cage from *Listeria innocua*. *Inorganic Chemistry* **42**, 6300-6305.

48. Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X. & Gollnick, P. (1999). Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**, 235-242.

49. Babitzke, P. (2004). Regulation of transcription attenuation and translation initiation by allosteric control of an RNA-binding protein: the *Bacillus subtilis* TRAP protein. *Current Opinion in Microbiology* **7**, 132-139.

50. Heddle, J. G., Yokoyama, T., Yamashita, I., Park, S. Y. & Tame, J. R. H. (2006). Rounding up: Engineering 12-membered rings from the cyclic 11-Mer TRAP. *Structure* **14**, 925-933.

51. Heddle, J. G., Fujiwara, I., Yamadaki, H., Yoshii, S., Nishio, K., Addy, C., Yamashita, I. & Tame, J. R. H. (2007). Using the ring shaped protein TRAP to capture and confine gold nanodots on a surface. *Small* **3**, 1950-1956.

52. Miranda, F. F., Iwasaki, K., Akashi, S., Sumitomo, K., Kobayashi, M., Yamashita, I., Tame, J. R. H. & Heddle, J. G. (2009). A self-assembled protein nanotube with high aspect ratio. *Small* **5**, 2077-2084.

53. Wang, W. X., Dgany, O., Wolf, S. G., Levy, I., Algom, R., Pouny, Y., Wolf, A., Marton, I., Altman, A. & Shoseyov, O. (2006). Aspen SP1, an exceptional thermal, protease and detergent resistant self-assembled nano particle. *Biotechnology and Bioengineering* **95**, 161-168.

54. Heyman, A., Levy, I., Altman, A. & Shoseyov, O. (2007). SP1 as a novel scaffold building block for self-assembly nanofabrication of submicron enzymatic structures. *Nano Letters* **7**, 1575-1579.

55. Medalsy, I., Dgany, O., Sowwan, M., Cohen, H., Yukashevska, A., Wolf, S. G., Wolf, A., Koster, A., Almog, O. & Marton, I. (2008). SP1 protein-based nanostructures and arrays. *Nano Letters* **8**, 473-477.

56. Frasconi, M., Heyman, A., Medalsy, I., Porath, D., Mazzei, F. & Shoseyov, O. (2011). Wiring of redox enzymes on three dimensional self-assembled molecular scaffold. *Langmuir* **27**, 12606-12613.

57. Medalsy, I., Klein, M., Heyman, A., Shoseyov, O., RemacleF, Levine, R. D. & Porath, D. (2010). Logic implementations using a single nanoparticle-protein hybrid. *Nature Nanotechnology* **5**, 451-457.

58. Hou, S., Wang, J. & Martin, C. R. (2005). Template-synthesized protein nanotubes. *Nano Letters* **5**, 231-234.

59. Audette, G. F., van Schaik, E. J., Hazes, B. & Irvin, R. T. (2004). DNA-binding protein nanotubes: Learning from nature's nanotech examples. *Nano Letters* **4**, 1897-1902.

60. Nuraje, N., Banerjee, I. A., MacCuspie, R. I., Yu, L. & Matsui, H. (2004). Biological bottom-up assembly of antibody nanotubes on patterned antigen arrays. *Journal of the American Chemical Society* **126**, 8088-8089.

61. Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., Goodman, A. L., Joachimiak, G., Ordoñez, C. L. & Lory, S. (2006). A virulence locus

of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526-1530.

62. Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y. & Mougous, J. D. (2008). *In vitro* self-assembly of tailorable nanotubes from a simple protein building block. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3733-3738.
63. Schreiber, A., Zaitseva, E., Thomann, Y., Thomann, R., Dengjel, J., Hanselmann, R. & Schiller, S. M. (2011). Protein yoctowell nanoarchitectures: assembly of donut shaped protein containers and nanofibres. *Soft Matter* **7**, 2875-2878.
64. McMillan, R. A., Paavola, C. D., Howard, J., Chan, S. L., Zaluzec, N. J. & Trent, J. D. (2002). Ordered nanoparticle arrays formed on engineered chaperonin protein templates. *Nature Materials* **1**, 247-252.
65. Tamerler, C., Duman, M., Oren, E. E., Gungormus, M., Xiong, X., Kacar, T., Parviz, B. A. & Sarikaya, M. (2006). Materials specificity and directed assembly of a gold-binding peptide. *Small* **2**, 1372-1378.
66. Li, H., Siu, K. M., Guevremont, R. & Le Blanc, J. Y. (1997). Complexes of silver (I) with peptides and proteins as produced in electrospray mass spectrometry. *Journal of the American Society for Mass Spectrometry* **8**, 781-792.
67. Ostrov, N. & Gazit, E. (2010). Genetic engineering of biomolecular scaffolds for the fabrication of organic and metallic nanowires. *Angewandte Chemie International Edition* **49**, 3018-3021.
68. Hu, M., Qian, L., Briñas, R. P., Lymar, E. S. & Hainfeld, J. F. (2007). Assembly of nanoparticle–protein binding complexes: from monomers to ordered arrays. *Angewandte Chemie International Edition* **119**, 5203-5206.
69. Achsel, T., Stark, H. & Lührmann, R. (2001). The Sm domain is an ancient RNA-binding motif with oligo (U) specificity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3685-3689.
70. Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. (2010). Functional organization of the Sm core in the crystal structure of human U1 snRNP. *The EMBO Journal* **29**, 4172-4184.
71. Pomeranz, K. D. A., Oubridge, C., Leung, A. K., Li, J. & Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458**, 475-480.

72. Mund, M., Neu, A., Ullmann, J., Neu, U. & Sprangers, R. (2011). Structure of the LSm657 complex: An assembly intermediate of the LSm1–7 and LSm2–8 rings. *Journal of Molecular Biology* **414**, 165-176.

73. Kufel, J., Allmang, C., Petfalski, E., Beggs, J. & Tollervey, D. (2003). Lsm proteins are required for normal processing and stability of ribosomal RNAs. *Journal of Biological Chemistry* **278**, 2147-2156.

74. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M. G. & Mabbutt, B. C. (2001). Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs1. *Journal of Molecular Biology* **309**, 915-923.

75. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989). Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 4569-4573.

76. Fischer, S., Benz, J., Späth, B., Maier, L. K., Straub, J., Granzow, M., Raabe, M., Urlaub, H., Hoffmann, J. & Brutschy, B. (2010). The archaeal Lsm protein binds to small RNAs. *Journal of Biological Chemistry* **285**, 34429.

77. Sauter, C., Basquin, J. & Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Research* **31**, 4091-4098.

78. Valentin Hansen, P., Eriksen, M. & Udesen, C. (2004). The bacterial Sm like protein Hfq: a key player in RNA transactions. *Molecular Microbiology* **51**, 1525-1533.

79. Wilusz, C. J. & Wilusz, J. (2005). Eukaryotic Lsm proteins: lessons from bacteria. *Nature Structural & Molecular Biology* **12**, 1031-1036.

80. Tomasevic, N. & Peculis, B. A. (2002). Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence. *Molecular and Cellular Biology* **22**, 4101-4112.

81. Pillai, R. S., Will, C. L., Lührmann, R., Schümperli, D. & Müller, B. (2001). Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1 like protein. *The EMBO Journal* **20**, 5470-5479.

82. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J. & Nagai, K. (1999). Crystal structures of two Sm protein

complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**, 375-387.

83. Khusial, P., Plaag, R. & Zieve, G. W. (2005). LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends in Biochemical Sciences* **30**, 522-528.
84. Hermann, H., Fabrizio, P., Raker, V., Foulaki, K., Hornig, H., Brahms, H. & Lührmann, R. (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *The EMBO Journal* **14**, 2076-2088.
85. Collins, B. M. (2001). Structural studies of archaeal and yeast Sm/Lsm proteins (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW.*
86. Sobti, M., Cubeddu, L., Haynes, P. A. & Mabbutt, B. C. (2010). Engineered rings of mixed yeast Lsm proteins show differential interactions with translation factors and U-Rich RNA. *Biochemistry* **49**, 2335-2345.
87. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R., Curmi, P. M. G. & Mabbutt, B. C. (2008). Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: implications for Lsm ring organisation and recruitment. *Journal of Molecular Biology* **377**, 1357-1371.
88. Mura, C., Phillips, M., Kozhukhovsky, A. & Eisenberg, D. (2003). Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4539-4544.
89. Arluison, V., Mura, C., Guzman, M. R., Liquier, J., Pellegrini, O., Gingery, M., Regnier, P. & Marco, S. (2006). Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. *Journal of Molecular Biology* **356**, 86-96.
90. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M. & Eisenberg, D. (2003). The oligomerization and ligand binding properties of Sm like archaeal proteins (SmAPs). *Protein Science* **12**, 832-847.
91. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M. & Eisenberg, D. (2003). The oligomerization and ligand- binding properties of Sm- like archaeal proteins (SmAPs). *Protein Science* **12**, 832-847.
92. Mura, C., Randolph, P. S., Patterson, J. & Cozen, A. E. (2013). Archaeal and eukaryotic homologs of Hfq. *RNA Biology* **10**, 636-651.

93. Yamashita, I., Kirimura, H., Okuda, M., Nishio, K., Sano, K. I., Shiba, K., Hayashi, T., Hara, M. & Mishima, Y. (2006). Selective nanoscale positioning of ferritin and nanoparticles by means of target-specific peptides. *Small* **2**, 1148-1152.

2 Chapter two

Materials and methods

2.1 Introduction

This chapter begins with an overview of the materials used in the experiments. A wide variety of techniques from varied disciplines have been used ranging from general microbiology to small angle X-ray scattering at the Australian Synchrotron. The laboratory texts authored by Maniatis et al. [1] were invaluable for routine manipulation of DNA plasmids and bacterial strains. G E Healthcare prescribed protein purification techniques were used as a guide to design the protein purification protocols. Also used were the previous methods developed in the Protein Structure lab, Macquarie University, Sydney.

2.2 Materials and equipment

2.2.1 Reagents

Unless otherwise stated, chemicals were purchased from Aldrich Chemicals or Sigma Chemical Company Ltd. Media for growing bacterial cultures were purchased from Life Technologies (Christchurch, New Zealand). General reagents used are summarised in **Table** 2.1. Purified water from a MilliQ system (Millipore) was used throughout. Sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gels and protein ladders were purchased from Life Technologies. Column chromatography media were purchased from G E Healthcare Life Sciences. The centrifuges used were Eppendorf 5810R (fixed angle rotor F-34-6-38) and Sorvall RC 6 plus (rotor F10S6x500Y).

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

Item	**Supplier**
acetic acid, glacial	Ajax Finechem
agar (bacteriological)	Life Technologies
ampicillin	Life Technologies
ammonium sulfate	Astral
Benchmark protein ladder	Life Technologies
bromophenol blue	Progen
calcium chloride	BDH
chloramphenicol	BDH
Coomassie brilliant blue	BDH
dimethylformamide	Astral
dithiothreitol	BDH
DNase I	Astral
glucose	Astral
glycerol	AppliChem
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	Astral
hydrochloric acid	Ajax Finechem
imidazole	AppliChem
isopropyl-β-D-thiogalactopyranoside (IPTG)	Astral
lactose	Merck
lysozyme	Astral
magnesium sulfate	AppliChem
potassium chloride	BDH
potassium phosphate	Astral
protease inhibitor cocktail	Sigma-Aldrich
RNase A	Astral

sodium acetate	Ajax Finechem
sodium chloride	Ajax Finechem
sodium dodecyl sulfate	AppliChem
sodium hydroxide pellets	BDH
sodium phosphate	Astral
tetramethylethylenediamine	Bio-Rad
tris(hydroxymethyl)aminomethane	Astral
bacto-tryptone	Oxoid
urea	Merck
yeast extract	Life Technologies

2.2.2 Growth media and buffers

Luria-Bertani (LB) – To 1 L dH_2O was added bacto-tryptone (10 g), yeast extract (5 g) and NaCl (10 g) and sterilised by autoclaving. LB plates were prepared by using 15 g/L agar.

Autoinduction media – Bacto-tryptone (10 g) and yeast extract (5 g) was added to 925 mL dH_2O. The media were sterilised by autoclave following which $MgSO_4$ (1 M, 1 mL), 50x5052 media (20 mL) and 20xNPS (50 mL) were added (all autoclaved separately prior to addition). [50x5052 media – glycerol (250 g), glucose (25 g), α-lactose (100 g) in 730 mL dH_2O]; [20xNPS pH 6.75 – $(NH_4)_2SO_4$ (66 g), KH_2PO_4 (136 g), Na_2HPO_4 (142 g) added to 900 mL ddH_2O].

SOC media – Bacto-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) were added to 1 L dH_2O. 20% w/v glucose was added after autoclaving.

Table 2.2| Composition of growth media.

Buffer	**Composition**
LB medium	bacto-tryptone (10 g/L), yeast extract (5 g/L), NaCl (10 g/L)
ZY medium	bacto-tryptone (10 g/L), yeast extract (5 g/L)
50x5052	glycerol (250 g), glucose (25 g), α-lactose (100 g)
20xNPS	$(NH_4)_2SO_4$ (66 g/L), KH_2PO_4 (136 g/L), Na_2HPO_4 (142 g/L)
SOC medium	bacto-tryptone (20 mg/mL), yeast extract (5 mg/mL), NaCl (0.58 g/L), KCl (0.95 g/L), $MgCl_2$ (1.20 g/L), $MgSO_4$ (1.20 g/L), glucose (3.6 g/L)

Table 2.3| Buffers used at various stages of protein purification.

Process	**Buffer**	**Composition**
Immobilised metal affinity chromatography	Buffer A	20 mM Tris/HCl, pH 8.0, 200 mM NaCl, 2% glycerol.
	Buffer B	20 mM Tris/HCl, pH 8.0, 200 mM NaCl, 2% glycerol, 10 mM imidazole.
	Buffer C	20 mM Tris/HCl, pH 8.0, 200 mM NaCl, 2% glycerol, 500 mM imidazole.
	lysis buffer	20 mM Tris/HCl, pH 8.0, 200 mM NaCl, 2% glycerol, DNase (10 µg/mL), RNase (100 µg/mL), lysozyme (1 mg/mL), protease inhibitor cocktail.
GST-tagged protein purification	Buffer D	PBS, pH 7.3 (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3).
	Buffer E	50 mM Tris/HCl, 10 mM reduced glutathione, pH 8.0.
	Buffer F	PBS, pH 7.3 (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3), 1 mM TCEP.
	Buffer G	PBS, pH 8 (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3).
	Wash buffer 1	PBS, pH 7.3, 0.1 % Triton X-100.
	Wash buffer 2	PBS, pH 7.3, 1 M NaCl.

2.2.3 Expression hosts and plasmids

The bacterial strains trialled and used in this work have been listed in **Table 2.4**. Hfq-deficient strain, (named MRE5) was a gift to the Macquarie University Protein Structure Group from Jean Beggs (University of Edinburgh).

Table 2.4| Bacterial strains used to generate protein material.

***E. coli* strain**	**Use**	**Source**
BL21	expression host	Life Technologies
BL21 (DE3)	expression host	Life Technologies
BL21 (DE3) pLysS	expression host	Macquarie University
TOP10	plasmid isolation	Life Technologies
MRE5Δhfq	expression host	Jean Beggs

Antibiotics used in the media were filter sterilised and stored as 10x stock solutions at -20 °C. The final working concentrations (**Table 2.5**) were used for the selection of an appropriate bacterial strain.

Table 2.5| List of antibiotics used as selection markers for the bacterial strains.

Antibiotic	**Stock (mg/mL)**	**Working (µg/mL)**
Ampicillin	100	100
Chloramphenicol	30	30
Kanamycin	50	50

The yeast Lsm3 gene was cloned in vector pCL774 by Nishen Naidoo [2], Protein Structure Group, Macquarie University, and transformed in BL21(DE3) pLysS strain of *E. coli*. For the purpose of this study, the material was obtained as a bacterial glycerol stock and stored at -80 °C. The Lsmα full length gene was synthesised by Epoch Biolabs (Sugar Land, Texas, USA). The genes were cloned into pET24a and pGEX-4T-2 vectors, containing a His_6-tag and GST-tag, respectively. The full length gene and the other sequences used to generate Lsmα constructs are shown in **Table 2.6**.

Table 2.6| Sequences of Lsm proteins used for this work.

Gene	Number of amino acids	Sequence
yeast Lsm3	96	**MHHHHHH**METPLDLLKLNLDERVYIKLRGARTLVGTLQAFDSHC NIVLSDAVETIYQLNNEELSESERRCEMVFIRGDTV TLISTPSEDDDGAVEI
Lsmα	83	MSVIDVSSQRVNVQRPLDALGNSLNSPVIIKLKGDR EFRGVLKSFDLHMNLVLNDAEELEDGEVTRRLGTVLIRGDNIVY ISP
His_6- Lsmα	99	**MH_6ENLYFQGGS**MSVIDVSSQRVNVQRPLDALGNSLNSPVIIKL KGDREFRGVLKSFDLHMNLVLNDAEELEDGEVTRRLGTVLIRGD NIVYISP
GST cleaved Lsmα	85	**GS**MSVIDVSSQRVNVQRPLDALGNSLNSPVIIKLKGDR EFRGVLKSFDLHMNLVLNDAEELEDGEVTRRLGTVLIRGDNIVY ISP

2.2.4 Chromatography equipment, media and columns

Chromatography operations were performed on an ÄKTA Explorer system (G E Healthcare). Hi-Trap Chelating columns (1 mL, G E Healthcare) and HisTrap columns (1 mL, G E Healthcare) were used for immobilised metal affinity chromatography (IMAC) precharged with different metal ions (**Section 2.5.2**). For analytical and preparative size exclusion chromatography, Superdex 200 GL 10/300 (G E Healthcare) columns (24 mL) and HiLoad 16/600 Superdex 200 pg (G E Healthcare) columns (120 mL) were employed, respectively.

2.3 Molecular biology

2.3.1 Storage of bacterial strains

Glycerol stocks were prepared in order to store *E. coli* strains indefinitely. The cells were grown in 10 mL of LB media containing the appropriate antibiotics for 16 h at 37 °C. 0.5 mL of resulting cell suspension was mixed with 0.5 mL of sterilised 50% glycerol. The cells were snap frozen in liquid nitrogen and stored at -80 °C.

2.3.2 Plasmid isolation from *E. coli*

Lsmα expression plasmids and pRARE constructs were isolated from bacterial cells using a commercial kit (QIAprep Spin Miniprep kit, QIAGEN) [3] according to the manufacturer's instructions. The plasmids were collected in the supplied elution buffer and stored at -20 °C in 20 µL aliquots.

2.3.3 Competent cell preparation

A modified Inoue method [4] was used to prepare competent cells for this work. A 5 mL starter culture (5 mL LB with 20 mM $MgSO_4$) was grown overnight with shaking at 180 rpm at 37 °C. This starter culture was used to inoculate 250 mL of LB broth containing 20 mM $MgSO_4$ in a 2 L conical flask. The culture was incubated with shaking at 180 rpm at 26 °C for 4 h until the optical density at 600 nm (OD_{600}) reached 0.5. After cooling on ice for 10 min, the cells were gently centrifuged at 4000 rpm, 10 min, 4 °C. The supernatant was discarded and the cells resuspended in ice-cold TB buffer (80 mL, containing 250 mM KCl_2, 15 mM $CaCl_2$, 10 mM PIPES pH 6.7, 55 mM $MnCl_2$) and centrifuged as before and the supernatant removed. The cells were resuspended in ice cold TB buffer (20 mL) for a second time with addition of 1.5 mL DMSO followed by incubation on ice for 10 min. Aliquots of 50 µL were made and were used immediately or snap frozen in liquid nitrogen before storing at -80 °C.

2.3.4 Bacterial transformation for plasmid propagation

Plasmid DNA was propagated and maintained in commercially available *E. coli* TOP10 cells (Life Technologies). 2 µL of plasmid DNA was added to a thawed 50 µL aliquot of competent cells and incubated on ice for 30 min. The cells were subjected to heat shock in a digital heat block at 42 °C for precisely 35 seconds after which they were immediately transferred to ice. 250 µL of room temperature SOC media was added and incubated at 37 °C for 60 min with shaking at 140 rpm. 50 µL of the culture was spread on a pre-warmed LB

agar plate containing appropriate antibiotics and incubated at 37 °C for 12-16 h. The colonies from the resulting cultures were used to prepare glycerol stocks or for subsequent plasmid preparation (**Section 2.3.2**).

2.4 Protein expression

2.4.1 Transformation for protein expression

pET24a and pGEX-4T-2 vectors containing the Lsmα gene was co-transformed with pRARE plasmid into a commercially available BL21(DE3) strain. Aliquots (2 μL) of Lsmα expression plasmids and pRARE were added to the thawed bacterial cells and were transformed as described in **Section 2.3.4**. The glycerol stocks of the expression strains were made (**Section 2.3.1**) and used to inoculate expression cultures.

2.4.2 Small-scale expression for screening of transformed colonies

To determine the optimum incubation time and temperature for protein expression, small scale expression trials were carried out. Two methods of induction were trialled using isopropyl β-D-1-thiogalactopyranoside (IPTG) and auto-induction allowing the expression of the recombinant proteins [5].

2.4.2.1 IPTG induced protein expression

Colonies from LB agar plates were used to inoculate LB media (10 mL) containing appropriate antibiotics. Suspension cultures were grown in Falcon tubes (50 mL) to an OD_{600} of 0.4-0.6 before IPTG was added (0.2 mM or 1 mM). To assess the effect of temperature on protein expression, the cultures were subsequently incubated at 26 °C or 37 °C, with shaking at 180 rpm. 500 μL samples were taken at various time points and centrifuged (12,000 rpm, 5 min) and the supernatant removed. 80 μL of the SDS-PAGE buffer was added to the pellets which were subsequently boiled for 5 min to lyse the cells. 10 μL was loaded onto an SDS-PAGE gel (**Section 2.6.3**).

2.4.2.2 Protein expression by auto-induction

In a similar way to the IPTG induction protocol, colonies from LB agar plates were used to inoculate auto-induction media (10 mL) containing antibiotics and grown at 26 °C and 37 °C with shaking at 180 rpm. Samples were collected as described above.

2.4.3 Large-scale expression of protein

For the production of recombinant proteins, cells were streaked on LB agar plates containing appropriate antibiotics and incubated overnight at 37 °C. A single colony was used to inoculate 10 mL LB broth (with antibiotics) and incubated overnight (37 °C, 180 rpm) as a starter culture.

2.4.3.1 IPTG induced protein expression

The starter culture was used to inoculate 200-400 mL of LB media in a 2 L baffled flask. Following incubation at 37 °C with shaking at 180 rpm until the OD_{600} reached 0.4-0.6, protein expression was induced with 0.2 mM IPTG. The culture was transferred to 26 °C and cells grown overnight with shaking at 180 rpm. Cells were further harvested by centrifugation at 4000 rpm for 10 min and re-suspended in 20 mL of buffer B. The cell suspension was used directly to purify the recombinant protein or else snap frozen in liquid nitrogen and stored at -80 °C.

2.4.3.2 Protein expression by auto-induction

A starter culture was used to inoculate 500 mL of ZY-rich media (**Table 2.2**) which was then incubated at 26 °C with shaking at 180 rpm for 22 h. Cells were further harvested by centrifugation at 4000 rpm for 10 min and re-suspended in 20 mL of buffer B. The cell suspension was used directly to purify the recombinant protein or else snap frozen in liquid nitrogen and stored at -80 °C.

2.5 Protein purification

2.5.1 Protein extraction

2.5.1.1 Sonication

Sonication uses a small metal probe oscillating at ultrasonic frequency generating localised low pressure and membrane disruption through cavitation. Lysis was carried out using a Sonicator 3000 (Misonix) equipped with a Microprobe (Misonix) for volumes less than 5 mL, and a standard probe for larger volumes (30 mL). The cell suspension was prepared in a 50 mL plastic beaker and placed in an ice slurry to prevent excessive heating of the sample. The cells were sonicated at 60% amplitude with a pulse rate of 3 s on and 10 s off for a total of 20 min.

2.5.1.2 Higher pressure homogenisation

Cell lysis by homogenisation is achieved when a sample is passed through a needle valve at high pressure causing cell disruption by shear stresses and decompression as cells return to normal atmospheric pressure. Volumes larger than 30 mL were lysed by passing the sample through a Microfluidics M-110P cell disruptor at 17,000 psi at 4 °C. The cycle was repeated three times to ensure complete lysis.

The cell debris from the lysed samples was cleared by centrifugation at 15,000 rpm for 25 min at 4 °C. The supernatant contained the soluble recombinant protein, which was subjected to downstream chromatography steps for purification.

2.5.2 Immobilised affinity chromatography for His_6-tagged proteins (IMAC)

IMAC was used to purify the His_6-tagged proteins using the pre-packed column mentioned in **Section 2.2.4**. The clarified lysate was filtered through a 0.22 μm syringe filter and applied to a 1 mL Hi-Trap Chelating column pre-equilibrated in buffer B [10 column volumes (CV)] using a Gilson peristaltic pump. Where appropriate, a HisTrap (G E Healthcare) column was used with the same protocol. The column was washed with 10 CV of buffer C to remove any unbound protein. Pure protein was eluted with a gradient set to 100% buffer C on an ÄKTA Explorer system (G E Healthcare) and the eluted fractions stored at 4 °C or snap frozen and stored at -80 °C for later use. All flow rates were maintained at 1 mL/min. To optimise the affinity chromatography, different transition metals were trialled to check the maximum affinity of the proteins for purification. The Hi-Trap Chelating column was charged with 100 mM $CoSO_4$, 100 mM $ZnSO_4$, 100 mM $NiSO_4$ and 100 mM $CuSO_4$ for each trial. The sample fractions were assessed by SDS-PAGE (**Section 2.6.3**).

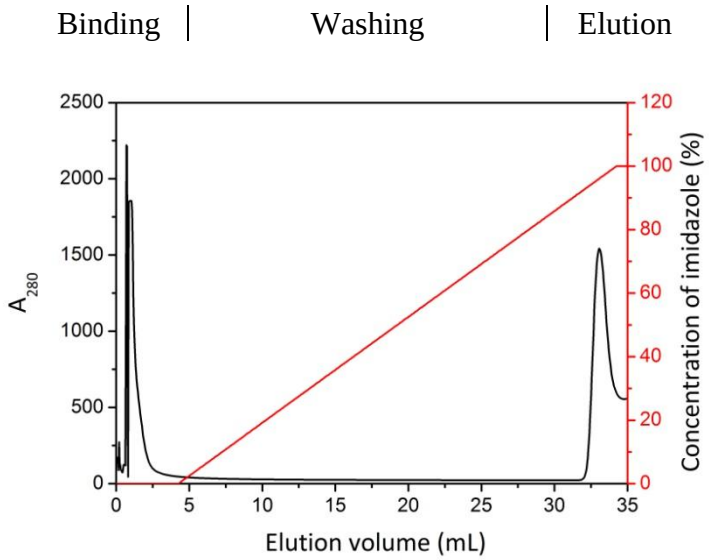


Figure 2.1| *Typical trace of immobilised metal affinity chromatography. The binding, washing and elution steps are shown above the graph. Absorbance at 280 nm (mAU) is shown in black and concentration of imidazole is in red.*

2.5.3 Purification of GST-fusion proteins

Batch purification was employed to purify the GST-tagged proteins. The binding of GST-tagged proteins depends on the size, conformation and concentration of the protein in the sample being loaded. Binding of GST to glutathione is also flow dependent, and lower flow rates often increase the binding capacity. Therefore, the increased incubation time in a batch purification mode increases the binding of the recombinant protein to the matrix for efficient separation from other soluble protein in the cell lysate.

Glutathione Sepharose 4B (G E Healthcare) was prepared by centrifugation at 600 rpm for 2 min and the storage buffer removed by decanting the supernatant. The resin was equilibrated by re-suspending the pelleted resin in buffer D, re-pelleting by centrifugation at 600 rpm for 2 min and removing the supernatant. This process was repeated thrice before adding buffer G to the pelleted resin. Approximately 2 mL of the prepared slurry was incubated with soluble lysate, extracted from 400 mL of bacterial culture (**Section 2.5.1**), for 90 min at room temperature with gentle shaking. The slurry was then centrifuged at 600 rpm for 5 min to pellet the resin with bound protein. The supernatant was discarded and the pelleted resin was gently transferred to a PD-10 gravity column (G E Healthcare). Unbound protein was removed by washing with wash buffer 1 and wash buffer 2 (5 CV). The slurry was equilibrated with buffer E before adding 80 µL of thrombin (1 U) for 16 h at room temperature. The cleaved protein was eluted in buffer D (3 CV). Generally the eluted proteins were further purified by preparative SEC (**Section 2.5.4**) using buffer G. The scheme of purification using glutathione Sepharose 4B is shown in **Figure 2.2**, with on-column cleaving of the GST-tag using thrombin.

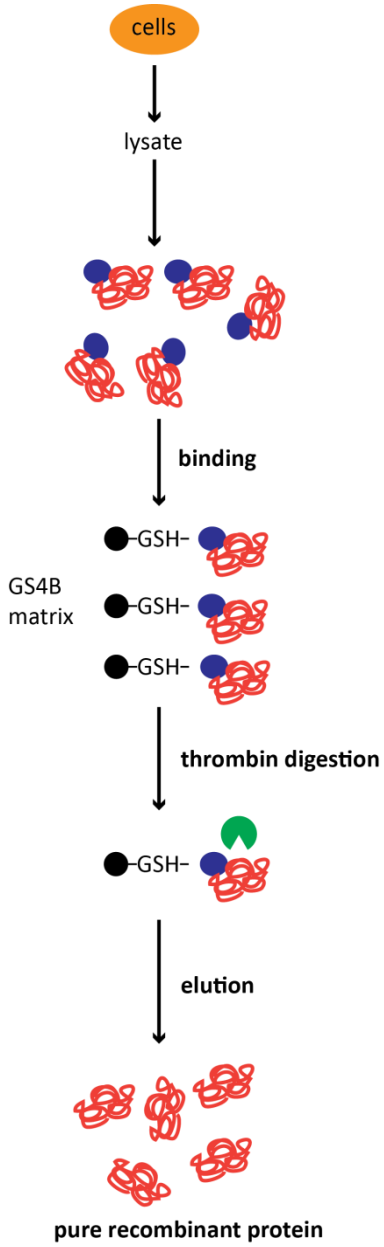


Figure 2.2| *Schematic of affinity purification using GST-tag. The GST fused recombinant protein was bound to the glutathione matrix and was digested by thrombin yielding pure protein after elution. This purification scheme was utilised for the purification of Lsmα and its muteins.*

2.5.4 Preparative size exclusion chromatography (SEC)

All the Lsm protein samples were further subjected to preparative SEC in order study the oligomeric distribution of the preparations. Lsm samples (~3 mL) were applied to a HiLoad 16/600 Superdex 200 pg column in appropriate buffer at a flow rate of 1 mL/min. The eluted protein fractions were collected as 0.5 mL fractions and relevant fractions were pooled.

2.6 Protein analysis

2.6.1 Protein concentration

Purified protein samples were analysed by UV absorption at 280 nm on the NanoDrop 2000. Extinction coefficients of the protein samples were calculated from amino acid sequence using the ExPasy online bioinformatics server [6] and were used to determine their concentration using the relation:

$$\text{Concentration} = (A_{280}/\varepsilon_{280}) * l$$

Where A_{280} is the absorption at 280 nm in AU, ε_{280} is the extinction coefficient at 280 nm (M^{-1} cm^{-1}), and l is the path length (1 mm in case of NanoDrop 2000).

2.6.2 Buffer exchange

(i) Buffer exchange by dialysis

Protein samples were transferred into dialysis tubing (10,000 MWCO), and dialysed against the required buffer. Sample to dialysis solution ratios would typically be 1:2000. Dialysis was carried out at 4°C with stirring to increase the rate of exchange. Typically, this would be repeated to achieve complete exchange.

(ii) Buffer exchange by chromatography

5 mL HiTrap Desalting columns (G E Healthcare) were purchased pre-packed with Sephadex G-25 Superfine cross-linked dextran. These employ the same principles as gel filtration to separate molecules with a mass larger than 5 kDa from those with a mass below 1 kDa, thereby separating proteins from buffer salts. The column was connected to the ÄKTA Explorer system, and equilibrated with 3 column volumes of the required buffer at 2 mL/min. The sample was loaded onto the column at a maximum volume of 1 mL. Protein was eluted at 2 mL/min using the required buffer, 0.5 mL fractions collected and the elution of protein monitored by UV absorbance at 280 nm.

(iii) Buffer exchange by dialysis devices

For faster buffer exchange of smaller volumes of samples (< 1 mL), Slide-A-Lyzer Dialysis Devices (Thermo Fisher Scientific) were used with a 7000 MWCO. 100-500 µL of the protein sample was pipetted into the dialysis cup partially submerged in the appropriate buffer. Dialysis was carried out at room temperature or 4°C, with stirring to increase the rate of exchange.

2.6.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is the most common technique to check the purity of protein samples. It completely denatures proteins in the presence of an ionic detergent, SDS, breaking the hydrogen bonds and coating the entire length of protein giving it a negative charge. This makes the migration of protein through the gel uniform. In the presence of reducing agents, dithiothreitol (DTT) or β-mercaptopethanol, the disulfide bonds are completely disrupted, thus completely denaturing the protein molecule. The samples were normally heated to 70 - 90 °C for 5 min before loading onto the gel. For SDS-PAGE, 3 µL of Novex® Sharp protein ladder (Life Technologies) was used as a molecular weight marker. NuPAGE® 4-12 % Bis-

Tris gels were run according to manufacturer's recommendations in a NuPAGE® gel electrophoresis box at room temperature. The protein samples were mixed with reducing agent and 4xlithium salt of dodecyl sulfate (LDS) sample buffer or 2xTris-Gly sample buffer depending upon the nature of gel to be run. The samples were typically boiled for 5 min when dealing with thermostable protein. The gels were stained with Simply blue stain using a three step protocol. The gels were placed in distilled water after electrophoresis and microwaved for 20 s to remove the SDS present in the running buffer. It was placed in stain for 1 h, until protein bands appeared. To destain the gel, it was placed in distilled water for 1 h.

Table 2.7| Buffer composition used in gel electrophoresis.

Solution	**Content**
MES running buffer	50 mM Tris-MES, pH 7.6, 0.1% SDS, 1 mM EDTA.
LDS loading dye	564 mM Tris base, 424 mM Tris/HCl, 40% glycerol, 8% LDS, 2.04 mM EDTA, 0.88 mM Coomassie Blue G-250, 0.7 mM phenol red, pH 8.5.
Reducing agent	1 M β-mercaptoethanol.
Tris-Gly loading dye	126 mM Tris/HCl pH 6.8, 20% glycerol, 0.005% bromophenol blue.
Simply Blue stain	80 mg Coomassie Brilliant Blue (CBB) dissolved in 1 L MilliQ water with 35 mM HCl.

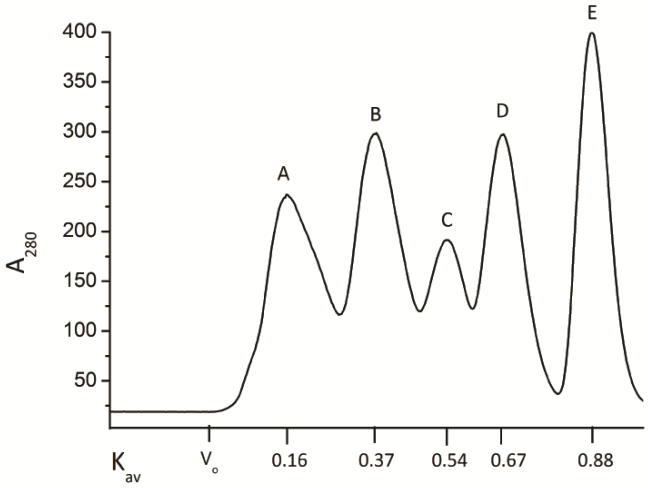
2.6.4 Analytical size exclusion chromatography

Size exclusion chromatography (SEC) separates macromolecules according to their hydrodynamic volume which is defined by the Stokes radius [7]. The column consists of porous polymer beads designed to have pores of different sizes. The particles with smaller hydrodynamic volumes equilibrate into these pores when a mobile phase is passed through the column. Therefore, smaller particles would have a longer path length than larger particles and would be separated along the length of the column as their flow would be retarded [8].

Although a low resolution technique, SEC is frequently used as a final step in protein purification and for discerning between high and low molecular weight components. It is an important means to assess the homogeneity of the sample as aggregated material can be easily separated from a folded protein. Most relevant to the present piece of work, is the application of SEC to determine the oligomeric state of complexes as they can be resolved in native conditions in which the macromolecular interactions are conserved [9; 10]. Additional advantages of SEC are its easy usage and no requirement for specialised equipment.

SEC has been successfully used for estimation of protein molecular weights, which involves plotting curves relating molecular weights and gel filtration behaviour of known standard globular proteins [11]. An empirical method of plotting the log of molecular weight ($\log M_w$) against the partition coefficient, K_{av}, gives a linear graph over the middle part of the working range of the column (**Figure 2.3B**). Comparison of an unknown protein elution volume with this linear plot gives an estimate of the molecular weights [12].

A



B

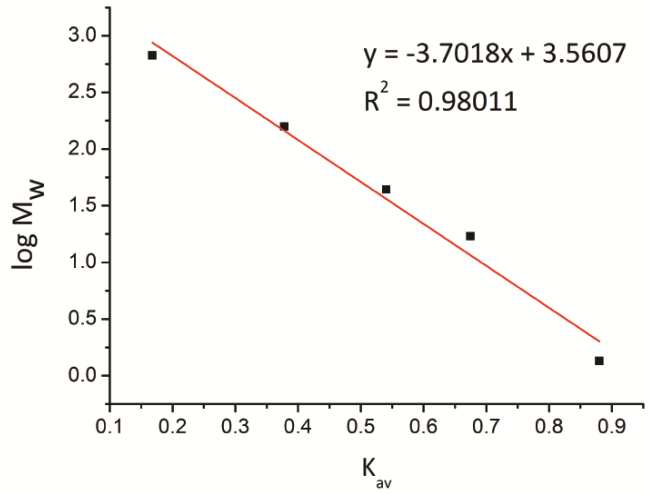


Figure 2.3| *Calibration of Superdex 200 10/300 GL size exclusion column using BioRad Gel filtration standards. (A) Fractionation of a mixture of protein standards in a calibration run for molecular weight estimation on Superdex 200 GL 10/300 at 0.5 mL/min in 20 mM Tris, pH 8.0, 150 mM NaCl, 2% glycerol. The peaks on the chromatograph correspond to A-thyroglobulin (670 kDa), B-γ-globulin (158 kDa), C-ovalbumin (44 kDa), D-myoglobin (17 kDa) and E-vitamin B_{12} (1.3 kDa). (B) Linear fit of K_{av} of the standards versus the log of molecular weights ($logM_w$).*

2.6.4.1 Calibration curves

For analytical SEC, 200 µL aliquots of proteins were injected at a flow rate of 0.4 mL/min, on a Superdex 200 GL 10/300 (G E Healthcare) column (24 mL) that was pre-equilibrated in the appropriate buffer. The optimum separation range of the chosen matrix was 10-600 kDa [13]. In order to estimate the molecular weight of the separated fractions, K_{av} values were calibrated using a Gel Filtration Calibration kit (BioRad). Calibration was performed in two subsequent runs. In the first run, 0.2 mL of 2mg/mL Blue Dextran 2000 (2000 kDa) was injected to determine the void volume of the column. Following the determination of the void volume, 150 µL of the BioRad calibration standard was injected as per the manufacturer's instructions. The calibration standard consisted of a mixture of thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B_{12} (1.3 kDa). The log of molecular weight was plotted over the K_{av} values of standard proteins. The values were calculated from:

$$K_{av} = \frac{V_e - V_o}{V_c - V_o} \quad \textbf{(2-1)}$$

Where V_e is the elution volume, V_o the void volume of column (7.9 mL) and V_c is the column volume (24 mL). The molecular weights of proteins of unknown mass were approximated from a linear fit to this calibration curve. **Figure 2.3B** shows the resulting calibration curve. All absorbance for the protein samples were measured at wavelength 280 nm in mAU.

2.6.5 SEC in line with static light scattering (SEC-SLS)

Static light scattering (SLS) is a technique whereby an absolute molecular mass of a protein sample in solution may be experimentally determined through exposure to a low intensity laser light. The intensity of the scattered light is measured as a function of angle and the data are analysed to calculate the molecular mass. Along with quality control of samples, SLS is a precise method to determine the oligomeric state of a protein sample. Since the light scattering and concentration are measured for each of the eluting fractions, the mass and size can be determined independently of the elution position. This is particularly important for novel protein architectures which may elute at positions distant from that predicted by the calibration for the SEC column.

The weight-average molecular weight [14] is given by:

$$M = K' \frac{(LS)}{(RI)} \qquad \textbf{(2-2)}$$

Where *K'* is given by:

$$K' = \frac{K_{RI}}{K_{LS}\left(\frac{dn}{dc}\right)} \qquad \textbf{(2-3)}$$

K_{RI} and K_{LS} are instrument calibration constants, LS is the light scattering signal, RI is the refractive index signal and dn/dc is the refractive index increment and is taken as 0.186 mL/g for globular proteins [14].

A Superdex 200 GL 10/300 column (G E Healthcare) was connected to a Viscotek 302-040 Triple Detector GPC/SEC system (Malvern Instruments Ltd.), operated at 28 °C, and equilibrated with appropriate buffer at 0.4 mL/min for a minimum of three column volumes

until a stable baseline was obtained. 110 µL aliquots of protein samples were loaded onto the column. Absolute molecular weight, radius of hydration and size distributions were calculated using the refractive index (RI), intrinsic viscosity, and right-angle light scattering (RALS) measurements calibrated against bovine serum albumin (BSA) (66.5 kDa, Sigma), which was run at the beginning and end of each sample sequence. Analysis was carried out using OmniSEC software as per the instructions.

2.6.6 Circular dichroism spectroscopy (CD)

UV circular dichroism spectroscopy was used to analyse the secondary structure of proteins. In proteins, the optically active groups are the amide bonds of the peptide backbone, and the aromatic side chains. When these peptide chromophores interact with circularly polarised light, the secondary structure motifs of proteins generate a specific CD spectrum, which is used to analyse the conformation of protein in solution [15; 16]. Circular dichroism spectra were generally recorded from 200 - 260 nm in buffer G. Protein samples were diluted sufficiently so as to keep the high tension voltage of the photomultiplier detector below 600 (usually 1:500 dilutions). Data were collected on a Jasco J-815 spectropolarimeter in a 2 mm cuvette with a 2 second response time. For each experiment, ten spectra were summed and averaged to maximise signal-to-noise. Spectra were analysed by Jasco Spectra Manager software (Version 1.52).

2.6.7 Small angle X-ray scattering (SAXS)

Small angle X-ray scattering (SAXS) is a technique to carry out structural characterisation of biological macromolecules in solution at relatively low resolution (1-3 nm). It provides information about the conformation and overall structure of large assemblies (**Figure 2.4A**). Compared to X-ray crystallography and nuclear magnetic resonance (NMR), SAXS offers some clear advantages [17]; (i) All experiments are carried out in solution. (ii) Large complex macromolecular assemblies can be analysed. (iii) Sample preparation is fast and simple. (iv) The technique is compatible with other biophysical methods. With synchrotron sources readily available, computational and instrumentation advances, SAXS is fast gaining importance as a reliable technique to study large organised protein assemblies [18]. In this thesis, SAXS was employed to extract geometric properties such as radius of gyration (R_g) and maximal particle diameter (D_{max}) to obtain shape and assembly information of the macromolecular assemblies.

2.6.7.1 Theory of small angle X-ray scattering

In solution scattering, the signal from all orientations of the target molecules relative to one another and the experimental apparatus, are averaged together. Solution scattering is continuous and radially symmetric (isotropic). Being essentially a contrast method, the scattering signal in SAXS is derived from the difference in the average electron density, $\Delta\rho(r)$, and bulk solvent, $\rho(s)$ [19], presented by the relation:

$$\Delta\rho(r) = \rho(r) - \rho(s) \qquad \textbf{(2-4)}$$

Data are collected on a buffer blank and on a sample. Subtraction of observed scattering yields the signal from the scattering due to the macromolecule.

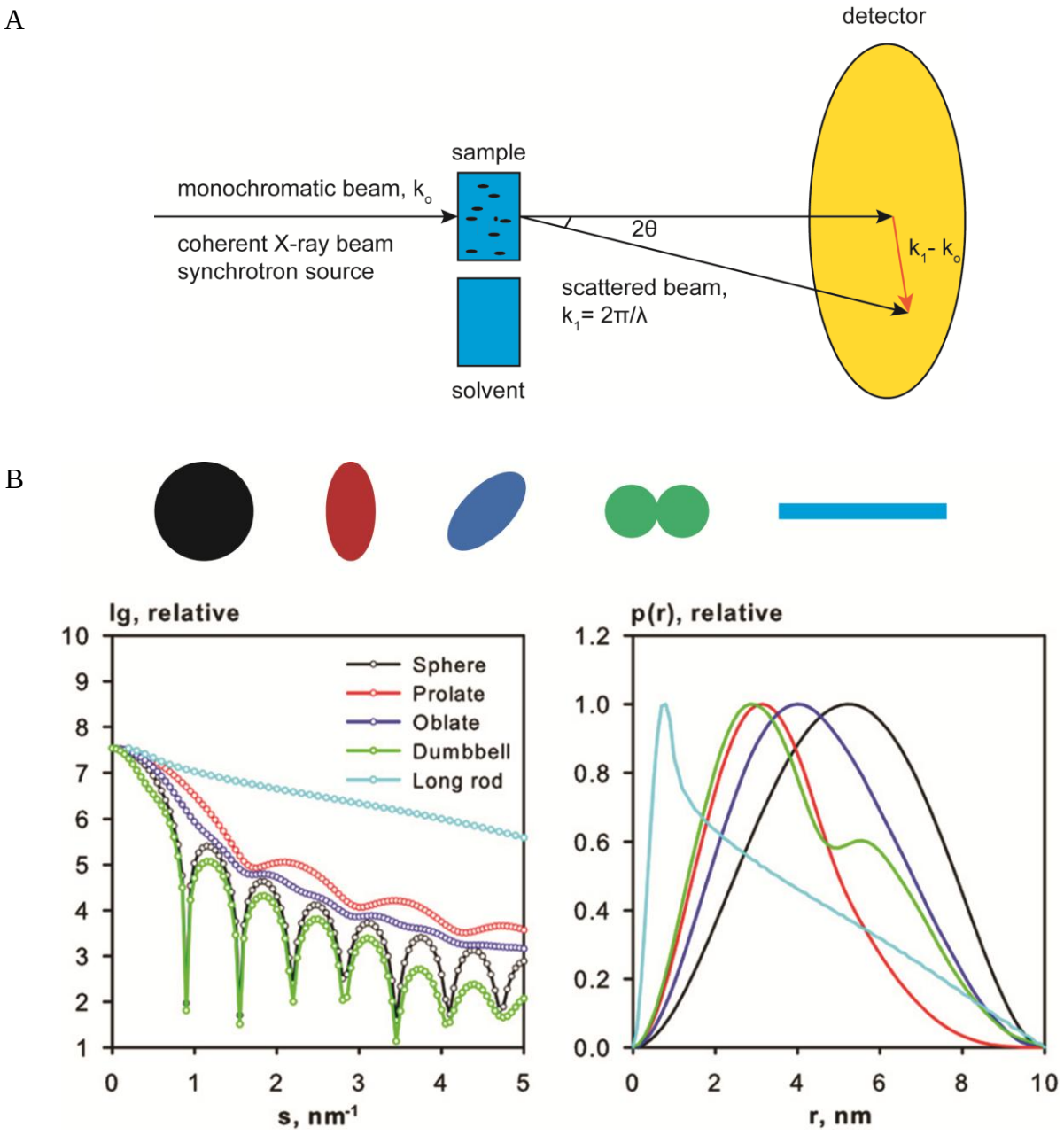


Figure 2.4| *(A) Schematic representation of small angle X-ray scattering. Adapted from Mertens and Svergun 2010. (B) Scattering intensities and distance distribution functions, P(r) calculated for typical geometric shapes. Solid shapes (black), prolate ellipsoid (red), oblate ellipsoid (blue), two domain (green) and long rod (cyan). Diagram taken from Mertens and Svergun, 2010.*

SAXS exploits coherent X-ray scattering given by the Thompson relation [20]:

$$E(r,t) = -r_o(\sin\psi/r)E(t) \quad \textbf{(2-5)}$$

wherein r_o decribes the electron radius, ψ polarisation angle, r the distance between the observer and the scattering event.

The angle between the wave vectors of incident and scattered waves is called the scattering angle (2θ) (**Figure 2.4A**) and is related to the intensity of Thompson scattering for an incident beam of intensity I_o by:

$$I(2\theta) = r_0^2\left(\frac{1+\cos^2 2\theta}{2}\right)\frac{I_o}{r^2} \quad \textbf{(2-6)}$$

Scattering from an assembly of electrons (in the case of macromolecules) is given by the sum of the waves originating from each electron in the given volume. Separation between the scatterers gives rise to a phase difference ϕ, given by:

$$\phi = q\Delta \quad \textbf{(2-7)}$$

$$|q| = 4\pi\frac{\sin\theta}{\lambda} \quad \textbf{(2-8)}$$

where |q| is the momentum transfer vector and wavelength λ.

The pair distribution function *P(r)* is also called the pair density distribution function. It is a radially averaged autocorrelation function calculated through a Fourier transform of the scattering curve and provides direct information about the distances between electrons in the scattering particles in the sample [21]. The *P(r)* plot is also an indicator of good quality data. Typically the *P(r)* function is zero at $r = 0$ and at $r \geq D_{max}$. If such a condition is not satisfied,

it may suggest the presence of unfolded proteins and aggregate formation. *P(r)* plots also give a fairly good indication of the overall shapes of the molecules (**Figure 2.4B**).

For a homogenous solution, the scattering intensity over the entire range, *q*, results from the spherically averaged electron distribution *P(r)* of the macromolecule [22]:

$$P(r) = \frac{\mathrm{r}}{2\pi}\int_0^\infty I(q)q\,\sin(qr)\,\mathrm{dq} \qquad \textbf{(2-9)}$$

Since, theoretically for a homogenous sample, the $I(q)$ is measured only for a finite number of points rather than [0,∞], it is preferable to compute *P(r)* by inverse transformation over the interval [0,D_{max}], and hence get [20]:

$$I(q) = \int_0^{Dmax} P(r)\frac{\sin\,(qr)}{qr}\mathrm{d}r \qquad \textbf{(2-10)}$$

Guinier approximation helps describe the scattering at low resolution [19].

$$I(q) = I(0)\exp\left[-\frac{1}{3}R_G^2\mathrm{q}^2\right] \qquad \textbf{(2-11)}$$

Where R_G is the radius of gyration of the particle.

R_G and $I(0)$ can be extracted from the Guinier plot of $\log(I(0))$ against q^2, which gives a straight line. Generally for globular proteins, Guinier approximation is valid over the range of $q\,R_G < 1.3$ [19]. Guinier plots are also a means of quality control of the sample, as a non-linearity in the Guinier range can indicate elongated samples and requires careful analysis of data.

I(*0*), intensity at *q*=0, depends on the square of the number of electrons (molecular weight) hence SAXS is particularly useful to determine the assembly of proteins [17]. Calculation of molecular weight from I(0) is less susceptible to inter particle correlations than extrapolation of low *q* data. *P(r)* based I(0) method has the following relation.

$$I(0) = 4\pi \int_{0}^{Dmax} P(r)dr \quad \textbf{(2-12)}$$

At low angles, the scattering intensities of folded macromolecules follow Porod's law [17]

$$I(q) \propto q^{-4} \quad \textbf{(2-13)}$$

However at higher angles, this relation does not hold true and must be expanded to describe the fractal degrees of freedom (df) [17]:

$$I(q) = q^{-\mathrm{df}} \quad \textbf{(2-14)}$$

SAXS is an ideal method to study unfolded domains in proteins. The Kratky plot which is $I(q)$ against q^2 can reveal the degree of unfolding depending on the shape of the plot. For a folded protein, the graph yields a peak shaped like a parabola at low angles and a plateau at higher angles. Unfolded domains clearly lack this peak [23].

SAXS experiments were performed at the SAXS/WAXS beamline at the Australian Synchrotron (Melbourne, Australia). A monochromatic X-ray beam with a wavelength of 1.03320 Å was used. The range of momentum transfer $q = 4\pi / \sin\theta$, was adjusted to $0.010 \leq q \leq 0.618$ Å^{-1} for experiments on this beamline. A Dectris-Pilatus detector was used to record scattering patterns (1 m, 170 mm x 170 mm, effective pixel size, 172 x 172 μm). All the protein samples were adjusted to 2 mg/mL and cleared either by centrifugation (12000 rpm, 2 min) or by 0.22 μm syringe filter. A Superdex 200 5/150 GL SEC column was used in line with SAXS collection data. 100 μL aliquots of protein were injected into the SEC column and the scattering data collected as each fraction eluted from the column.

SAXS data were collected at room temperature using an exposure time of 2 s per image. Averaged scattering from the buffer was recorded and used for background subtraction. Native Lsmα samples and muteins were measured in buffer G and $[H_6\text{Lsm}\alpha]_7$ was analysed in buffer A. Data reduction and subtraction were performed using ScatterBrain (Australian Synchrotron) software [24]. The resulting data were processed using the ATSAS [25] package. Guinier approximations were carried out using PRIMUS [26]. Electron distribution functions were calculated in GNOM [27] and normalised with respect to *P(r)* by division of all data points through their respective *P(r)* maxima. For determination of molecular envelopes, GASBOR [23] was employed. Comparison of the solution scattering of the experimental with that of the theoretical scattering was evaluated using CRYSOL [28].

2.6.8 Negative stain transmission electron microscopy (TEM)

Transmission electron microscopy uses a high energy electron beam transmitted through a thin sample to image and analyse the microstructure of materials. TEM is now a standard technique to visualise macromolecular assemblies [29] and supramolecular complexes in biological systems [30]. Negative stain TEM utilises heavy metal salts to coat the protein molecules, thereby creating differential scattering of electrons. This causes a phase contrast forming an image of the material on the grid [31] (**Figure 2.5**). Negative stain TEM has been utilised in this project to study the novel structures formed by Lsmα.

Carbon-coated Formvar 200-mesh copper grids (ProSciTech, Australia) were deposited successively for one minute each onto 12 μL drops of (i) protein samples, (ii) water (3 times) and (iii) a 2% w/w uranyl acetate solution in water. Filter paper (Whatman #1) was used to remove the excess liquid from the grids, which were then left to dry for a few hours before electron microscopy observations. The grids were examined with a FEI Morgagni 268D transmission electron microscope operating at 80 kV, with magnifications up to 180k. Micrographs were captured using a SIS/Olympus Megapixel III digital camera mounted above the phosphor screen.

The EMAN v1.9 software suite [32; 33] was used to generate a two-dimensional average molecular image of the generated complexes from 447 individual particle images, selected from 30 micrographs. Particles were interactively selected using boxer; autoboxed particles were visually checked and poor quality images were manually deleted. To reduce high-and low-frequency noise components, a 2 nm to 20 nm band-pass filter was applied to the raw images, which were subsequently normalised. Parameters of individual micrographs were determined and corrected for using EMAN v1.9 methodologies. Reference-free class

averages were generated in 8 rounds of iterative refinement by using the refine2d.py routine in EMAN v1.9.

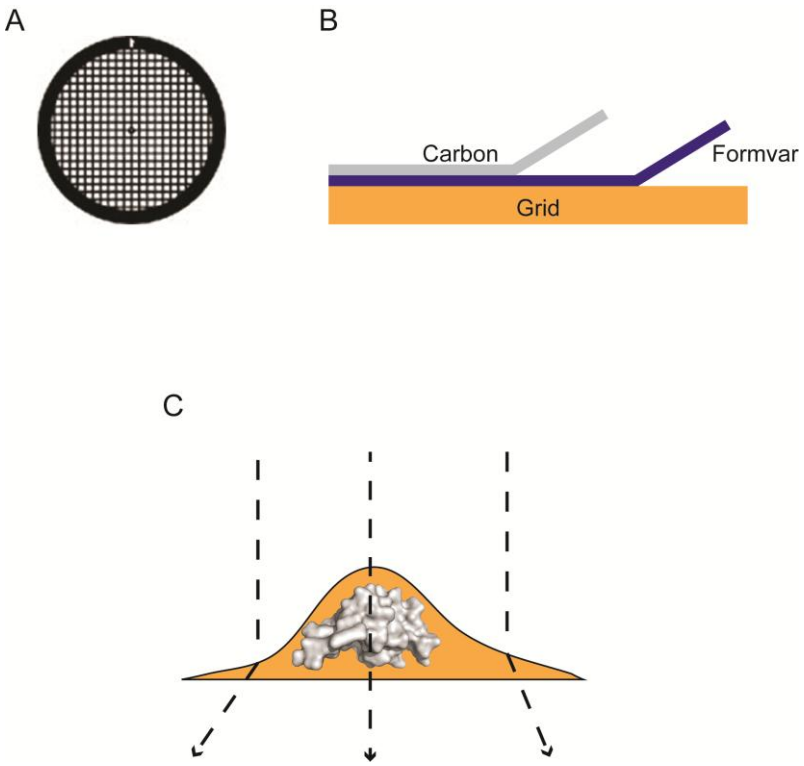


Figure 2.5| *Diagram showing components used in negative stain transmission electron microscopy. (A) Copper grid 200 mesh square. (B) Transverse section of grid showing carbon and formvar layers. (C) Electron scattering through protein sample stained with heavy metal salts.*

2.7 References

1. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). Molecular cloning: a laboratory manual, Vol. 545, *Cold Spring Harbor Laboratory Cold Spring Harbor, NY*.
2. Nishen, N. (2006). Molecular structures of the RNA-binding Lsm proteins (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
3. Qiagen. (2012). *QIAGEN® plasmid purification handbook, Qiagen*.
4. Inoue, H., Nojima, H. & Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-28.
5. Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10**, 411-421.
6. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*, pp. 571-607. Springer.
7. Erickson, H. P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological Procedures Online* **11**, 32-51.
8. Striegel, A., Yau, W. W., Kirkland, J. J. & Bly, D. D. (2009). Modern size-exclusion liquid chromatography: practice of gel permeation and gel filtration chromatography. *2nd edit, John Wiley & Sons*.
9. Kunji, E. R., Harding, M., Butler, P. J. G. & Akamine, P. (2008). Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. *Methods* **46**, 62-72.
10. Folta-Stogniew, E. (2006). Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. In *New and Emerging Proteomic Techniques*, pp. 97-112. Springer.
11. Fraenkel-Conrat, H., Harris, J., Levy, A. & Glick, D. (1955). Methods of Biochemical Analysis. *Vol. 2, Interscience, New York*, 359.
12. Andrew, P. (1970). Estimation of molecular size and molecular weights of biological compounds by gel filtration. *Methods of Biochemical Analysis **18***, 1-53.
13. Kågedal, L., Engström, B., Elelgren, H., Lieber, A.-K., Lundström, H., Sköld, A. & Schenning, M. (1991). Chemical, physical and chromatographic properties of

Superdex 75 prep grade and Superdex 200 prep grade gel filtration media. *Journal of Chromatography A* **537**, 17-32.

14. Wen, J., Arakawa, T. & Philo, J. S. (1996). Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Analytical Biochemistry* **240**, 155-166.
15. Greenfield, N. J. (2007). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876-2890.
16. Greenfield, N. J. (1996). Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Analytical Biochemistry* **235**, 1-10.
17. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**, 191-285.
18. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole Ii, F. L., Tsutakawa, S. E., Jenney Jr, F. E., Classen, S., Frankel, K. A. & Hopkins, R. C. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature Methods* **6**, 606-612.
19. Svergun, D. I. & Koch, M. H. J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* **66**, 1735-1782.
20. Petoukhov, M. V. & Svergun, D. I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Current Opinion in Structural Biology* **17**, 562-571.
21. Koch, M. H., Vachette, P. & Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly Reviews of Biophysics* **36**, 147-227.
22. Mertens, H. D. T. & Svergun, D. I. (2010). Structural characterisation of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology* **172**, 128-141.
23. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* **80**, 2946-2953.

24. Kirby, N., Mudie, S. & Hawley, A. (2013). A low-background-intensity focusing small-angle X-ray scattering undulator beamline. *Journal of Applied Crystallography* **46.6**, 1670-1680.
25. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Applied Crystallography* **40**, 223-228.
26. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282.
27. Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495-503.
28. Svergun, D., Barberato, C. & Koch, M. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773.
29. Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. *Annual Review of Biophysics and Biomolecular Structure* **31**, 303-319.
30. Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galán, J. E. & Aizawa, S. (1998). Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602-605.
31. Brenner, S. & Horne, R. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta* **34**, 103-110.
32. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology* **128**, 82-97.
33. Ludtke, S. J. (2010). 3-D structures of macromolecules using single-particle analysis in EMAN. In *Computational Biology, pp. 157-173. Springer*.

3 Chapter three

Investigation of the Lsmα interfaces using site directed mutagenesis

3.1 Introduction

Self-assembling biomolecules have formed a niche as molecular building blocks with potential bionanotechnological applications [1; 2]. In these applications, the modular nature and the ability to control the oligomeric state of ring-shaped proteins are desirable. The change in the size of the cavity provides additional control over the dimensions and morphology of the nanostructures obtained. The inner pore lining also provides a malleable site to chemically engineer functional groups that would enhance the functionality of the supramolecular structures. For this thesis, change in the pore size of Lsmα was investigated by altering the relevant amino acid residues involved in the interactions at the subunit interface of each monomer.

As described in Chapter one, Lsm proteins are found in a number of oligomeric states in different species. The subunit interface was engineered to tune the ring diameter. The first part of this chapter deals with the preparation and isolation of recombinant Lsm proteins and the muteins. The second part discusses the solution studies with comparison to the wild type protein.

3.2 Lsm systems investigated for change in pore diameter

The modular nature of a self-assembling system provides a platform to control and modify the individual subunits in order to generate a biological nanostructure [3]. The oligomeric plasticity exhibited by the Lsm proteins (as discussed in **Section 1.8**) make them an attractive target for manipulation and control over the pore size by altering the number of subunits interacting to form a ring-shaped structure. Two Lsm proteins from different species were looked into for altering their oligomeric states to obtain varying pore size of the ring.

The Lsm3 polypeptide from *S. cerevisiae* assembles as an octameric ring when expressed using recombinant methods [4]. Eukaryotic Lsm proteins generally adopt a heptameric quaternary state, it was therefore of interest to study the Lsm3 interface. The second system chosen was a thermophilic and highly stable protein, the Lsmα system (from *M. thermoautotrophicum*). **Figure 3.1** shows the various crystal structures of ring-shaped proteins that have been studied for their potential nanotechnological applications [5; 6] and have been modified to generate useful nanostructures [7]. The Lsm scaffold presents one of the smallest ring diameters to work with, adding a new dimension to the nanostructures generated.

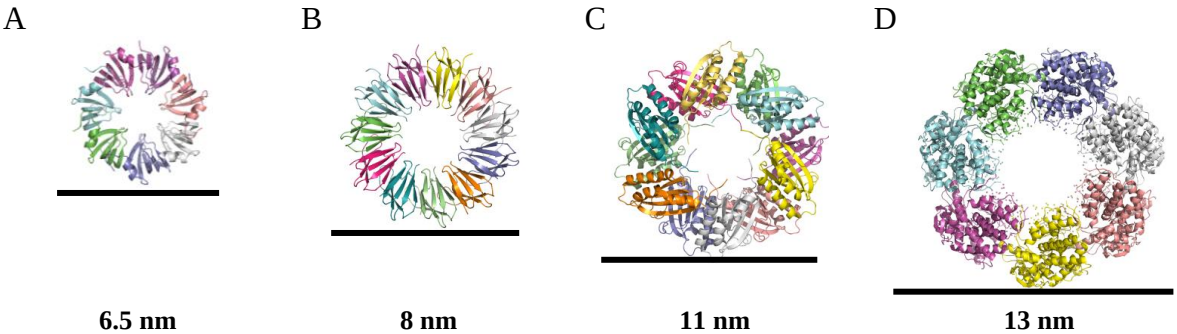


Figure 3.1| *Crystal structures of the ring-shaped proteins. (A) Lsmα (PDB ID 1I81) 6.5 nm ring [8]. (B) TRAP protein (PDB ID 1QAW) 8 nm ring [9]. (C) Stable protein 1 (PDB ID ITR0) 11 nm ring [10]. (D) GroEL (PDB ID 1GRL) ~ 13 nm ring [11]. All proteins shown approximately to scale.*

3.3 Yeast Lsm3 system

An inspection of the Lsm3 subunit interface from X-ray crystallographic data shows that the presence of an extra hydrogen bond and an absence of aromatic residues results in opening of the ring and hence, accommodation of a larger number of monomers in the ring complex [12]. From the structural alignment of Lsm proteins (**Figure 1.6**), it is evident that Lsm3 has a longer loop4 region compared to other Lsm proteins. This has been implicated in protein binding functions [12]. In addition, the C-terminal tail of each Lsm3 polypeptide engages in β-sheet interactions across the rings [4] (**Figure 1.9F**). To investigate its assembling properties, the expression and solubility of Lsm3 was trialled to obtain sufficient quantities of the protein.

3.3.1 Yeast Lsm3 expression system

An N-terminally His_6-tagged version of yeast Lsm3, cloned in plasmid pCL774, was obtained as a bacterial (*E. coli*, BL21(DE3) pLysS) glycerol stock, from the Protein Structure Group, Macquarie University (**Section 2.2.3**). The stock was used to inoculate a starter culture and incubated overnight at 37 °C, 180 rpm. This was diluted into fresh media (4 mL into 400 mL) and grown to an OD_{600} of 0.5-0.6 (**Section 2.4.3**). The protein expression was induced by addition of IPTG (0.2 mM) and the cells were harvested by centrifugation after growth at 26 °C for 4 h. The pellet was resuspended in buffer B and protein extracted by high pressure homogenisation (**Section 2.5.1.2**). The lysate was centrifuged at 15,000 rpm, 4 °C for 25 min to remove the cell debris. The supernatant containing the soluble protein was filtered (0.22 μm) and applied to the affinity columns.

Adapting protocols established previously [4], a HisTrap column (G E Healthcare) pre-charged with Ni^{2+} was used to purify Lsm3. The clarified cell lysate was directly loaded onto the matrix using a peristaltic pump and the elution step carried out by increasing the imidazole concentration to 500 mM in a gradient mode (**Section 2.5.2**). The His_6-tagged Lsm3 protein failed to bind to the HisTrap matrix since most of it is seen in the flowthrough and wash fractions, as assessed by SDS-PAGE (**Figure 3.2A**). There was no protein that eluted off the IMAC column (**Figure 3.2B**). Hence, its purification failed using this column.

A Hi-Trap Chelating system (G E Healthcare) pre-charged with Ni^{2+} was then used to purify Lsm3. The clarified lysate was loaded onto the column, pre-equilibrated with buffer B (**Section 2.5.2**). Following washing with Tris-based buffer in high salt at pH 8.0, bound Lsm3 was eluted by increasing the imidazole concentration to 350 mM (70% of the gradient) (**Figure 3.3B**). The differential binding of Lsm3 to the affinity matrix could arise due to the presence of iminodiacetic acid (IDA) ligand (Hi-Trap Chelating matrix) as opposed to NTA ligand.

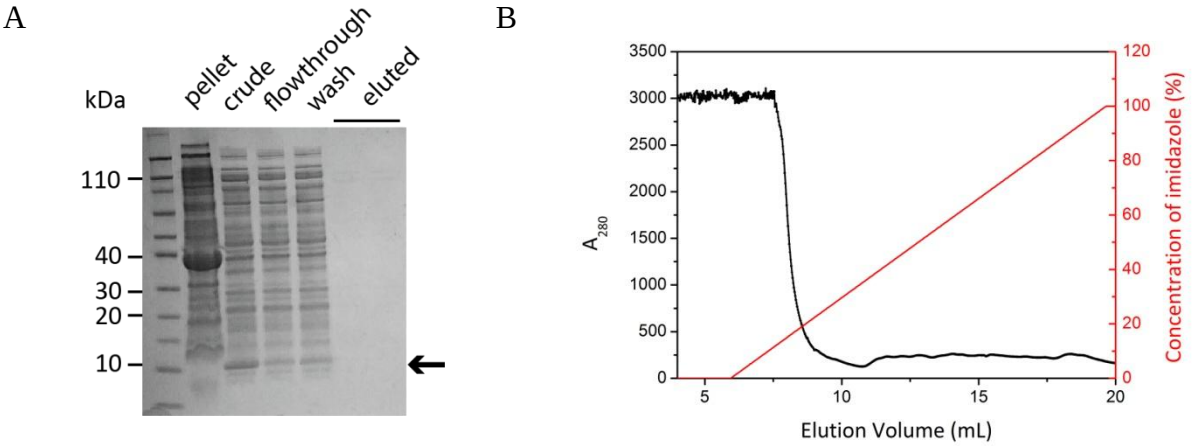


Figure 3.2| *(A) SDS-PAGE of Lsm3 showing stages of IMAC purification by HisTrap column (G E Healthcare). Arrow indicates the absence of monomeric form in the elution fraction corresponding to ~ 10 kDa. (B) IMAC chromatography trace. Absorbance at 280 nm is shown in black and concentration of imidazole is in red.*

(HisTrap column matrix). Lsm3 was obtained as a highly pure sample with a yield of 40 mg from 400 mL of culture. The SDS-PAGE gel shows the complete dissociation of the protein into its monomer after boiling for 5 min. A large band was obtained at ~ 10 kDa corresponding to the monomeric form of Lsm3 (**Figure 3.3A**). The purified sample was collected and subjected to analytical size exclusion chromatography (SEC).

The trace obtained by analytical SEC equilibrated with Tris/NaCl buffer pH 8.0 (**Figure 3.4**), shows that the majority of the affinity purified sample, eluted at the void volume (fraction I), indicating aggregation. A small peak (fraction II) was obtained at K_{av} corresponding to ~40 kDa. Thus, the affinity purified Lsm3 sample was mostly aggregated in solution. The sample also showed signs of visible precipitation after a few hours post-purification when stored at 4 °C.

Since Lsm3 was prone to aggregation during purification it was deemed to be not stable enough to act as a tecton and no further work was done on this protein.

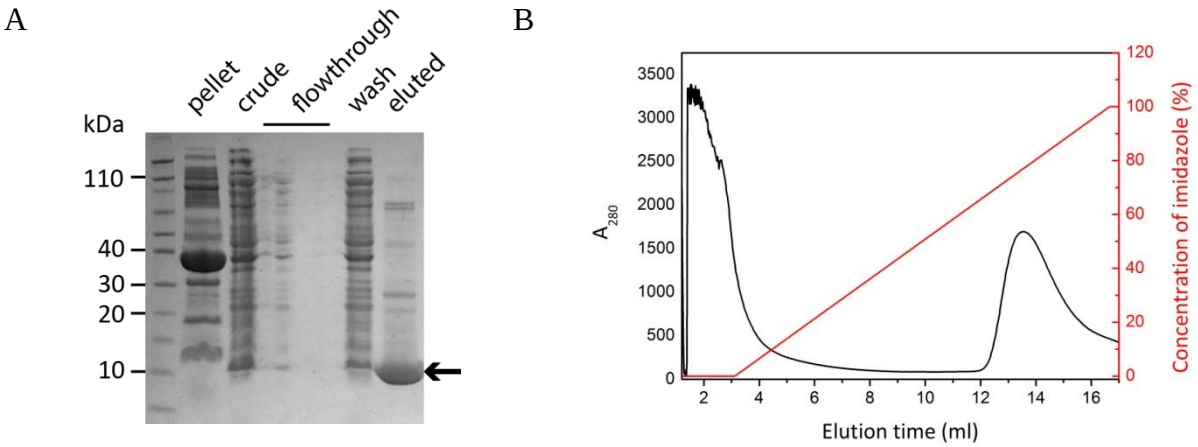


Figure 3.3| *(A) SDS-PAGE of Lsm3 showing stages of IMAC purification by HiTrap Chelating column (G E Healthcare). Arrow indicates the monomeric form of Lsm3 in the elution fraction. (B) IMAC chromatography trace. Absorbance at 280 nm is shown in black and concentration of imidazole is in red.*

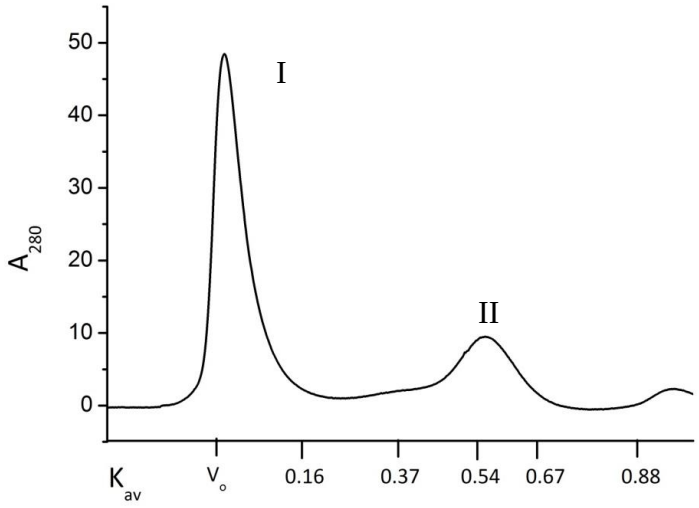


Figure 3.4| *Analytical SEC of Lsm3. The sample was run in Tris/NaCl buffer pH 8.0 at a flow rate of 0.4 mL/min on a Superdex 200 GL 10/300 (G E Healthcare) column.*

3.4 Design of Lsmα muteins

Muteins were designed to investigate residues involved in controlling the pore size and obtain a toroid of a different oligomeric state. To achieve this, residues were chosen that were considered important for the hydrogen bonding and hydrophobic interactions between the subunits of the Lsmα heptameric ring [13]. In particular the residues constituting the β4- and β5-strands, known to be involved in ring formation were targeted. The mutations were chosen to structurally conserve the Lsm fold per monomer and simultaneously perturb the hydrogen bonding network between the monomers. For each of the muteins generated, it was predicted that the toroid formation would be altered and result in a different oligomeric state of the protein during the *in vivo* assembly of the Lsmα monomers.

Previous work on Lsm proteins from different sources have shown that different sub-complexes can be generated by covalently linking different combinations of Lsm polypeptides. Heterodimeric and heterotrimeric sub-complexes were reconstituted from the

components of human Lsm1-7 and Lsm2-8 *in vitro* [14]. More recent work shows the formation of varying oligomeric forms adopted by the subunits of Lsm protein from *S. pombe* [15]. To our knowledge, to date, no work has been carried out investigating the inter-subunit interactions between Lsmα monomers, which may hold answers to the oligomeric plasticity of Lsm proteins in general [16]. The following sections detail the generation of the site specific mutations, their expression and purification and solution state characterisation in comparison to the wildtype heptameric Lsmα.

Another strategy employed to alter the pore size of Lsmα was the generation of chimeric proteins. Whole sections of the β4- and β5-strands of Lsmα were replaced with appropriate sections from the *E. coli* Hfq protein. It was hypothesised that the nature of interactions at the subunit interface could be altered by the exchange of segments between a hexameric toroid and a heptameric toroid. The generation and expression of these chimeric proteins is also presented in this chapter.

3.5 Structure of Lsmα cyclic oligomer complex

As mentioned in **Section 1.7**, Lsm proteins have the propensity to assemble into ring-shaped structures. Lsmα assembles as a cyclic oligomer of seven identical subunits, with an overall diameter of 6.5 nm and pore diameter of 1.5 nm [8]. Within the heptameric ring the individual monomers interact via hydrogen bond pairing of the β4- and β5-strands coupled with hydrophobic and hydrophilic interactions. There is formation of the left handed propeller-like structure as each monomer interacts with its adjoining neighbours via β-sheet extension, whereby the β4-strand residues are hydrogen bonded to the β5-strand of the neighbouring monomer [17].

In the subunit A/subunit B interface, the complex is stabilised by hydrophobic contacts as well as hydrophilic interactions and a salt bridge between Glu-35 (subunit A) and Arg-64 (subunit B). There are two hydrophobic pockets enclosed within each interface that extend to all seven subunits, resulting in the formation of the heptameric ring. The first pocket consists of Ile-27, Val-77 and Tyr-78 of chain A, and Leu-30, Phe-36, Leu-66, Val-69 and Ile-71 of chain B. A second hydrophobic pocket is formed between side chains of the amphipathic α-helix and hydrophobic core of chain A, with Val-50 and Leu-70 of chain B [17]. Enclosure of these hydrophobic cores of all component monomers results in the formation of a stable cyclic oligomer.

The loop2 (Lys-31 and Gly-32), loop3 (Leu-45 and His-46) and loop5 (Gly-73 and Asp-74) of all seven subunits line the inner hole of the Lsmα ring, while the outside edge of the ring is formed by the curved β-sheets of each protein chain. This arrangement produces two distinct faces for the complex, with the N-terminal helices covering one face, and the loop4 dominating the opposite surface. These faces will be hereby referred to as the helix side and the loop4 side, respectively.

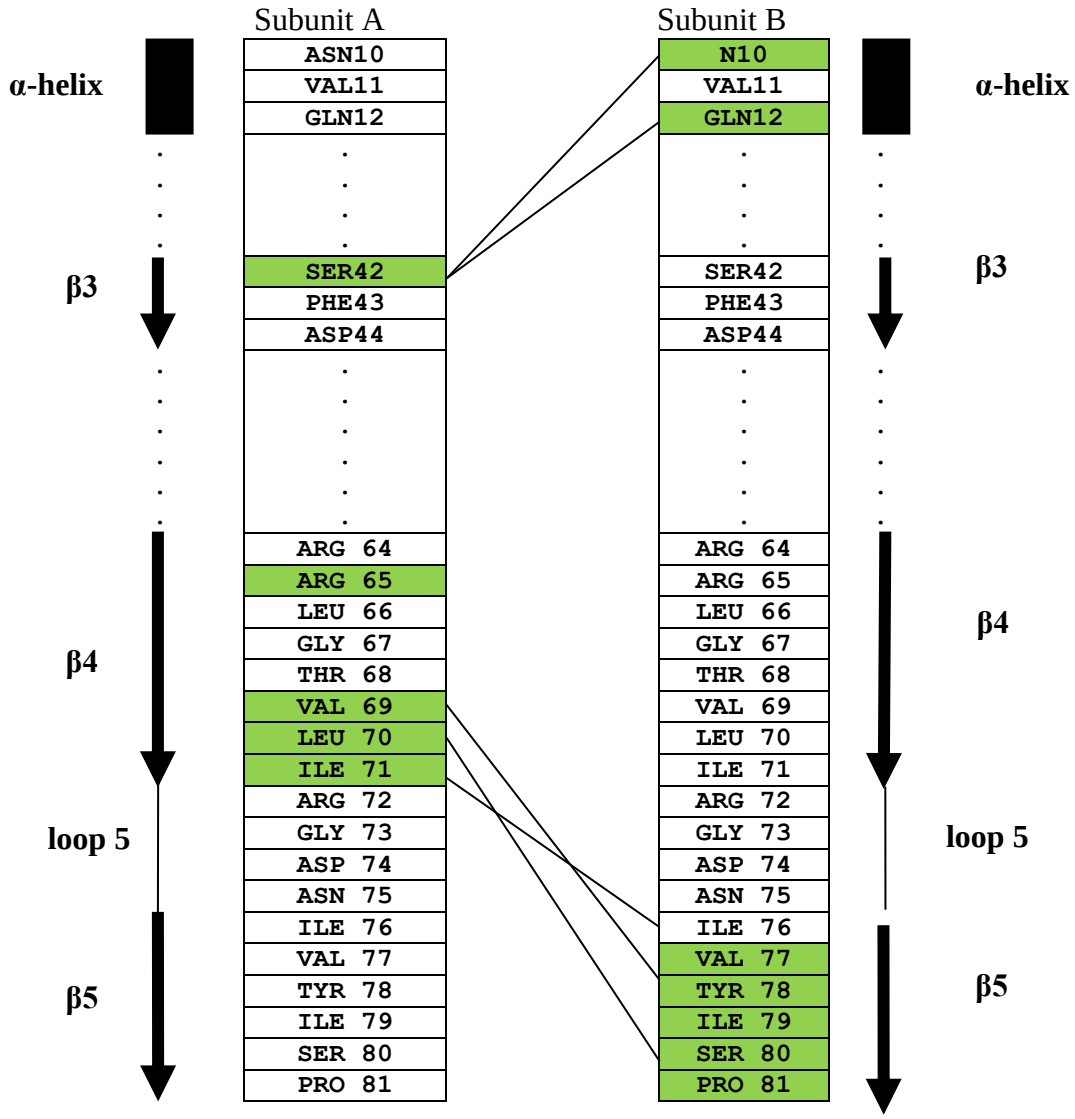


Figure 3.5| *Interface regions of Lsmα.. β4- and β5-strands (shown in green) were selected to be mutated. Black lines denote the hydrogen bonds between the residues. Dotted areas represent the regions in the Lsmα fold not considered for any change.*

3.6 Generation of Lsmα interface mutations

The residues involved in the subunit interface were determined through analysis of the interface carried out using the PDBePISA [18] online bioinformatics tool which lists the interactions by identifying the residues that become buried when two subunits come together and analysing the proximity and bonding potential of these residues. It is a useful tool for the exploration of macromolecular interfaces.

Analysis of the Lsmα interface in PDBePISA revealed an intricate network of hydrogen bonds and hydrophobic interactions. Essentially, the most important and consequently the region probed were the β4-strand and β5-strand of each monomer, which consist of at least five hydrogen bonds as listed in **Table 3.1**. Two residues located at the α-helix, N10 and Q12, were also chosen as they form hydrogen bond with S42 of the adjacent subunit. The residues were chosen from this list and mutated to alanine to eliminate the bond formation. Since the nature of interactions at the monomer interface is very complex, single residues were chosen to be mutated thereby altering the overall pattern of hydrogen bond formation to a relatively small extent.

Table 3.1| Residues involved in hydrogen bond formation at β4/β5 interface*.

	Subunit A	**Subunit B**
1	N10, Q12	S42
2	L70	I79
3	R72, N75	V77
4	E56	Y78
5	T68	S80

* Evaluated in PDBePISA.

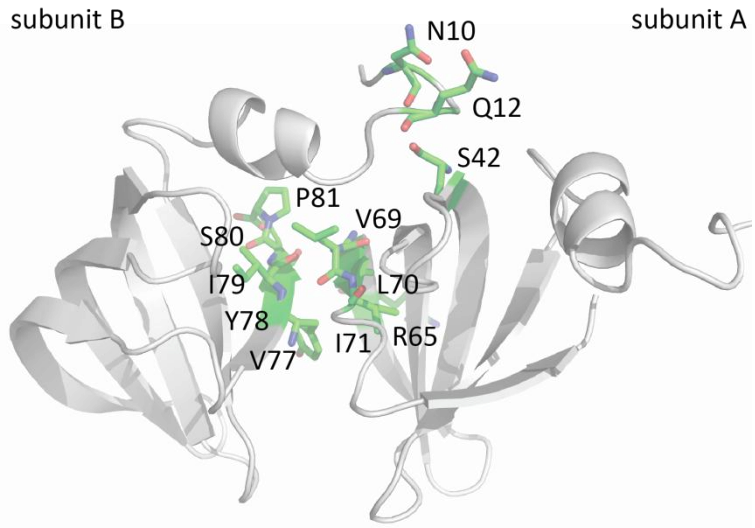


Figure 3.6| *Ribbon representation of the residues mutated at the subunit interface of Lsmα.*

Previous work on Lsmα has attempted to identify the RNA-binding residues and elucidate the biological function of Lsm proteins including Lsmα [19; 20]. Intensive work on the residues involved at the subunit interface was lacking in these studies. Hence the residues spanning the β4- and β5-strands were each mutated to alanine. Alanine screening has been used to identify hotspots in protein in protein interfaces and successfully disrupt interactions [21; 22]. Mutations to alanine are effectively a side chain deletion, as the hydrogen bonding potential, charge and the steric properties of the residues are largely eliminated. Single residue alanine substitutions were therefore introduced to the Lsmα subunit interface at positions involved in direct hydrogen bonding (**Figure 3.5**).

Archaeal Lsm proteins exhibit a certain degree of oligomeric plasticity. Lsm proteins from *A. fulgidus* belonging to two different subfamilies, Sm1 and Sm2, assemble as a heptameric and a hexameric complex, respectively [23]. The authors explain this difference to arise from the absence of a set of electrostatic interactions constituted by a highly conserved negatively charged residue in the case of Sm2 archaeal proteins. Thus, if this electrostatic interaction is eliminated in the case of Lsmα, it was predicted to assemble as a hexameric complex. However, due to the involvement of this residue in RNA binding, it would have diminished

the functionality of the Lsmα scaffold and further downstream applications involving RNA binding would have become problematic. In addition to altering the inter-subunit interface, residues located at the helix faces were also chosen to probe possible formation of ring-to-ring association.

The mutations made to the Lsmα ring scaffold can be generalised with respect to the location on the ring surface and on the structural elements as follows:

(i) Helix face/loop face

N10 and Q12 residues are involved in interactions with adjacent subunits as described in **Table 3.2**.

(ii) β4-strand

The β4-strand is comprised of eleven residues in the case of Lsmα, out of which four were chosen to be mutated to alanine. These residues either make up the hydrophobic core or are involved in the hydrogen bonding at the monomeric interface. The muteins generated were R65P, V69A, L70A and I71A. R65 is located along the edge of the Lsmα scaffold and is solvent accessible. Changing it to a proline was predicted to alter the bend in the strand and hence influence the angle between the monomers.

Although archaeal Lsm proteins assemble as heptameric complexes, there is evidence of a hexamer formation in the case of Af Sm2 (Lsm protein from species *A. fulgidus*) [23]. As seen from the structural alignment, for Af Sm2, positions 70 and 71 of the β4-strand consist of valine in place of leucine and isoleucine, respectively (**Figure 1.7**). It was hypothesised that by altering the side chain to the methyl group of alanine, the hydrophobic interactions would decrease at the interface causing a change in assembly of monomers, thereby altering the oligomeric state of the Lsmα ring.

Table 3.2| Summary of the mutations made to the Lsmα scaffold*.

Region	Mutation	Structure/function role
α-helix	N10A	Asparagine is a polar interfacing residue involved in hydrogen bond formation with Ser at position 42 of the adjacent monomer thereby contributing to the formation of cyclic oligomer.
	Q12A	Similar to N10, it is an interfacing polar residue which forms hydrogen bond with S42 of adjacent monomer.
β4-strand	R65P	R65 faces towards the solvent.
	V69A	Valine forms part of the hydrophobic pocket driving the oligomeric assembly of the ring scaffold.
	L70A	Leucine takes part in hydrogen bond formation with I79 and forms part of the hydrophobic core of the interface.
	I71A	I71 is an interfacing hydrophobic residue.
β5-strand	V77A	Hydrophobic residue involved in hydrogen bond with N75 and R72 of the corresponding adjacent subunit.
	Y78A	Involved in hydrogen bond formation with E56 and forms part of the hydrophobic pocket.
	I79A	Hydrogen bond formation with L70 and part of hydrophobic core.
	S80A	Penultimate residue of C-terminus involved in hydrogen bond formation with T68.
	P81T	Terminates the C-terminal tail. Engineered to tyrosine to enable hydrogen bonding capacity.

* Interfacing residues which are involved in hydrogen bond formation as evaluated in PDBePISA[18].

(iii) β5-strand

The β5-strand is comprised of seven residues from the structural alignment. The residues changed to alanine were V77, Y78, I79 and S80. The Lsmα sequence is characterised by an unusually short C-terminal and terminated by a proline. Therefore, it was hypothesised that if more hydrogen bonds are included at the C-terminus, the result would be a different molecular oligomer. Hence, the last residue, proline, was mutated to tyrosine (P81T) to mimic the C-terminus of the octameric Lsm3, and potentially induce the formation of an octameric Lsmα ring.

3.7 Expression and purification of Lsmα muteins

To study the impact of the site specific single mutations on the quaternary structure of the Lsmα ring, the samples were prepared recombinantly as GST-fusion products. Although using an affinity peptide is easier and more efficient to produce a recombinant protein, it was found that the His_6-tag influences the quaternary organisation of the Lsmα ring scaffold. This is discussed in detail in Chapter 5. The GST-tag has been shown to protect against intracellular protease cleavage and otherwise stabilise the recombinant protein [24]. The tag can be easily cleaved with a site specific protease and the recombinant protein purified using glutathione affinity matrix [25]. This also provided access to a protein that was definitively without a His-tag, an important control protein for the work in Chapter 5.

The wildtype Lsmα and the mutated genes were cloned into pGEX-4T-2 vector and ordered from Epoch Life Sciences (Sugar Land, Texas, USA). A conventional strain of *E. coli* BL21, was transformed with the appropriate vectors and small scale expression trials were carried out (**Section 2.4.2**). Protein levels in the cells were assessed by SDS-PAGE analysis. The protein expression was induced by IPTG (0.2 mM) and the cells grown for 16 h post-

induction at 26 °C with shaking at 180 rpm (**Section 2.3.4**). All the muteins showed good level of expression for large scale preparation. As outlined in **Section 2.5.1**, the samples were prepared for purification by the glutathione Sepharose affinity method.

Pure protein samples of all the muteins were obtained using a glutathione Sepharose matrix followed by thrombin cleavage, using a purification scheme as described in **Section 2.5.3**. The filtered cell lysate was incubated with the pre-washed glutathione affinity matrix required for binding of the fusion protein to the matrix. Wash steps were included to ensure the removal of any loosely bound contaminant proteins. The matrix was poured into a gravity flow column and incubated with thrombin for efficient cleavage of the GST-tag from the protein of interest. Typical yields for all the muteins were around 5-10 mg per 400 mL of culture. Pure protein was eluted in PBS buffer and fractions were stored at 4 °C. Purity was assessed by SDS-PAGE (**Figure 3.7**) and further analysed by preparative SEC and SEC-SLS.

Thermostable proteins present an interesting case when examined by SDS-PAGE. Due to their stability towards heat, it is difficult to completely dissociate them by conventional methods employed for protein denaturation. Lsmα and its muteins had these issues. As seen in **Figure 3.7**, most muteins along with wildtype Lsmα dissociate into the monomer (~ 9 kDa), a trimeric state (~ 25 kDa) in certain cases, a tetrameric state (~ 30-40 kDa), and for some muteins, a very faint band was also observed at ~ 55-60 kDa, corresponding to the heptameric non-denatured ring. The incomplete dissociation was likely caused by the presence of the complex network of hydrogen bonds and hydrophobic interactions at the subunit interface, part of which is detailed in **Table 3.2.** However, it could also indicate a change in the structure caused by the mutations at the interface or a different oligomeric state

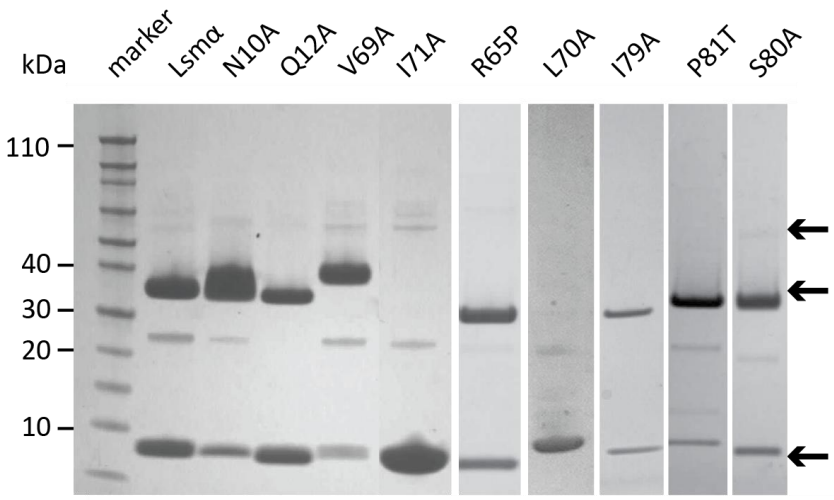


Figure 3.7| *SDS-PAGE gel of purified Lsmα with muteins. Most muteins show similar bands as Lsmα as depicted by arrows. The arrows correspond to the monomeric (~9 kDa), tetrameric (~30-40 kDa) and the non-dissociated (~60 kDa) states.*

of the dissociated species, giving rise to different dissociation products. As observed from the gel, there are variations in the tetrameric species ranging from 30 kDa for R65P and I79A, to 40 kDa for V69A. These could be attributed to the difference in the oligomeric state of the dissociated product of the muteins caused by the respective mutations.

It is interesting to point out that among the mutations studied, L70A and I71A showed a near complete dissociation into the monomeric state, with complete absence of the tetrameric species, when compared with the other muteins (**Figure 3.7**). For L70A, these results were consistent with the solution studies as will be discussed in **Section 3.9.1**. Both L70 and I71 are involved in hydrogen bonding and form a part of the hydrophobic pocket at the interface (**Table 3.2**). These results indicate that among the list of the residues mutated, these two residues seem to contribute to the stability of the Lsmα ring by constituting an important interaction at the subunit interface.

3.8 Characterisation of Lsmα

3.8.1 Solution state of Lsmα

In order to characterise the oligomeric state of Lsmα after cleaving with thrombin, the affinity-purified sample was subjected to analytical SEC and SEC-SLS in PBS buffer at pH 8.0. These findings were corroborated with the previous solution state analysis of Lsmα prepared by similar means at the Protein Structure Group, Macquarie University [17]. Lsmα had a homogenous elution profile at a K_{av} 0.46 consistent with its single-ring molecular weight of ~ 63 kDa (**Figure 3.8A**). The monomeric mass of Lsmα calculated from its amino acid composition is 9.173 kDa (including the extra Gly and Ser residues at the N-terminal from the pGEX plasmid). The analytical SEC results indicate that the Lsmα cyclic oligomer is composed of seven subunits and is consistent with the previous results [8]. **Figure 3.8B** shows the bands obtained on SDS-PAGE after being treated with SDS and being boiled for 5 min. The monomeric band is obtained at ~ 9 kDa and stable oligomers at ~ 36 kDa and ~ 60 kDa. The presence of stable oligomer bands indicates that the Lsmα structure can tolerate boiling in the presence of the detergent SDS. The Lsmα heptameric oligomer does not completely break down into its monomeric subunits being a thermostable protein and is further stabilised by the complex network of hydrogen bonds and hydrophobic interactions at the protomeric interfaces as noted in **Section 3.5**. These stability results are consistent with the thermophilic nature of the host species *M. thermoautotrophicum*.

SEC-SLS analysis of fraction I revealed that the purified sample is homogenous and no higher order molecular species are formed (**Figure 3.8C**). It is shown to be monodisperse with a mass of 63 kDa. Thus, Lsmα remains as a single toroidal complex in solution and there is no association of the heptameric rings into higher-order structures.

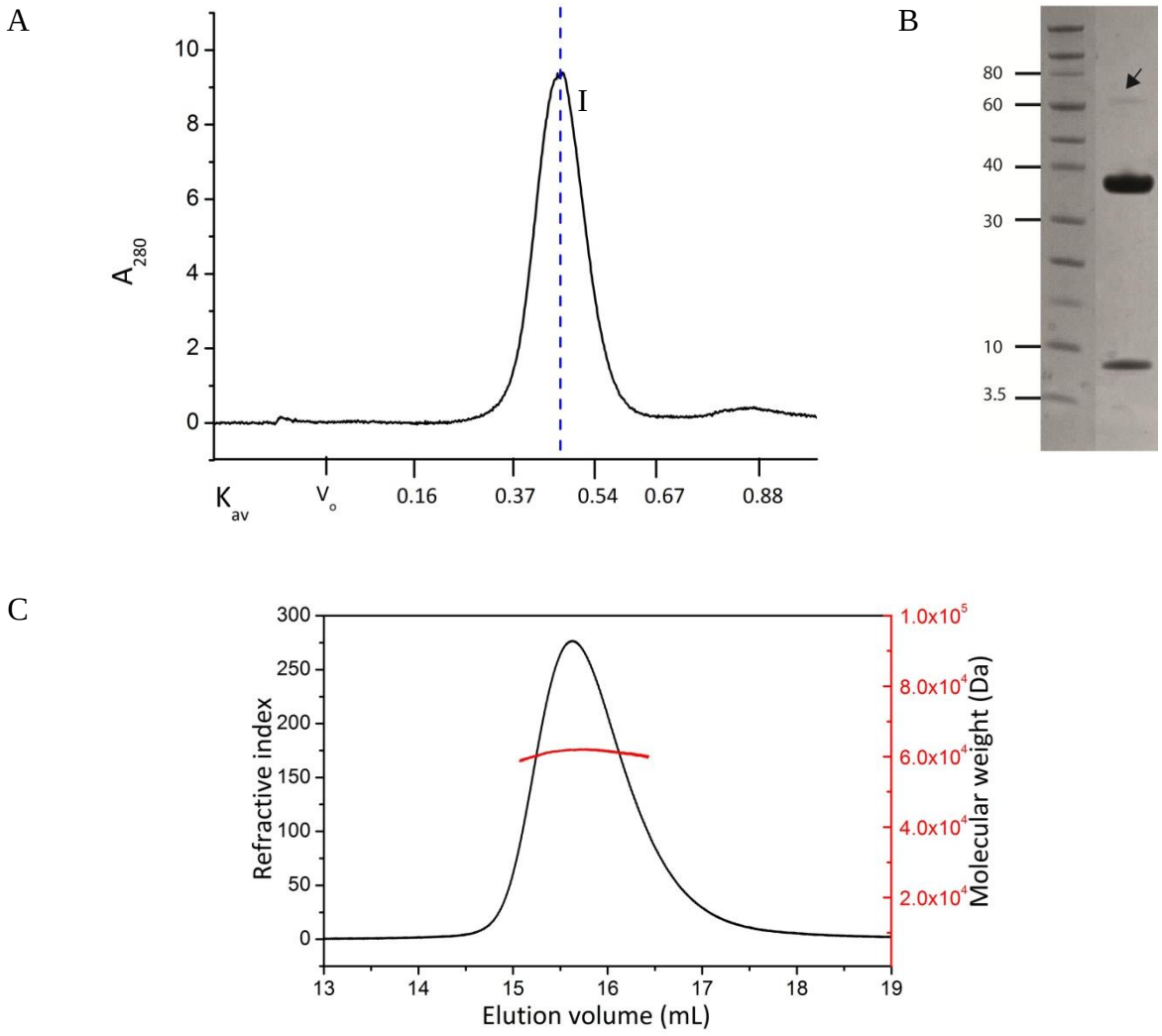
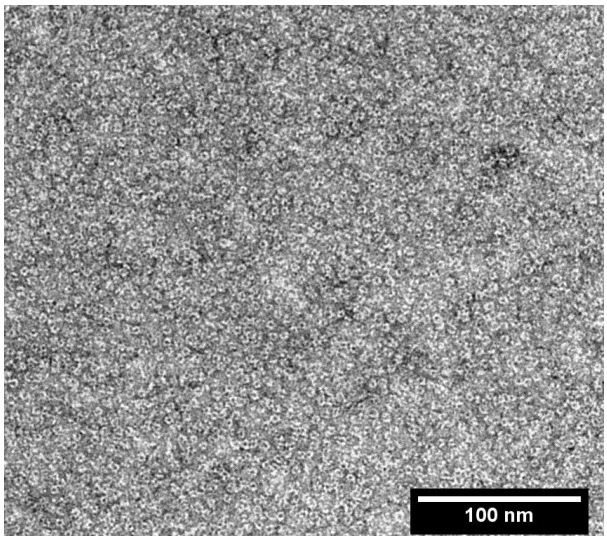


Figure 3.8| *(A) SEC chromatograph of Lsmα in PBS pH 8.0. (B) SDS-PAGE gel of pure Lsmα after boiling for 5 min. Arrow indicates the non-dissociated heptameric ring. (C) SEC-SLS of Lsmα in PBS pH 8.0 at a flow rate of 0.4 mL/min.*

3.8.2 Transmission electron microscopy of Lsmα rings

Transmission electron microscopy (TEM) of Lsmα revealed individual rings of outer diameter of ~ 6.5 nm and inner pore diameter of ~ 2 nm which corresponds well with the crystal structure dimensions. A histogram plot of diameter reveals a narrow distribution from 5 nm to 8 nm with the majority of the rings at 6 nm (**Figure 3.9**). This kind of distribution could be attributed to the different angular positions of the rings on the TEM grid [26]. TEM also confirms the absence of any higher order structures under these conditions.

A



B

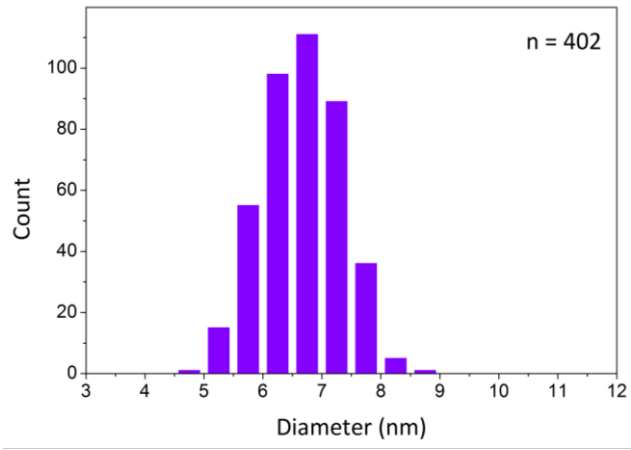


Figure 3.9| *TEM analysis of Lsmα. (A) Negative stain micrograph of well-distributed single rings of Lsmα. (B) Size distribution histogram collected for 402 particles.*

3.9 Characterisation of Lsmα muteins

3.9.1 Solution characteristics of muteins

The muteins were prepared as GST-fusion proteins and analysed by preparative SEC post thrombin-cleavage (as outlined in **Section 2.5.3**). The conditions of purification were similar to that of the wildtype Lsmα to ensure uniformity and valid comparison between them.

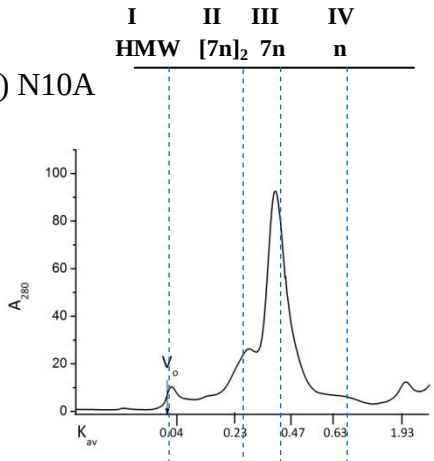
All muteins eluted with a mixture of oligomeric species, with the majority of the sample eluting at K_{av} ~ 0.4, which corresponds to a single ring form. Within the error of the preparative SEC, the oligomeric state ranged from hexameric to heptameric. The SEC chromatographs for all muteins, with their apparent molecular weights, are shown in **Figure 3.10** (see legend for molecular weight values) with comparison with the wildtype Lsmα (**Figure 3.10L**). In Y78A (**Figure 3.10H**), a significant proportion of the sample was aggregated compared to the other muteins and fraction III was spread over a larger elution volume with multiple peaks. This is inidicative of different oligomeric states that could not be separated by the S200 matrix.

In identical conditions, it was apparent that the single residue changes at the subunit interface changed the species distribution of Lsmα. Although most muteins showed a heterogeneous distribution on preparative SEC, it was evident that certain muteins had more pronounced species than others. L70A (**Figure 3.10F**) exhibited a mixture of species, with fraction IV eluting at K_{av} = 0.6, indicating a monomeric state. It was difficult to carry out further analysis on this fraction as it degraded within a few hours at 4 °C. Interestingly, as discussed in **Section 3.7**, SDS-PAGE of L70A (**Figure 3.7**), showed a complete dissociation into its monomer post-boiling and SDS treatment. This is indicative that the residue Leu-70 takes part in crucial interactions that contribute to the stability of the Lsmα ring scaffold.

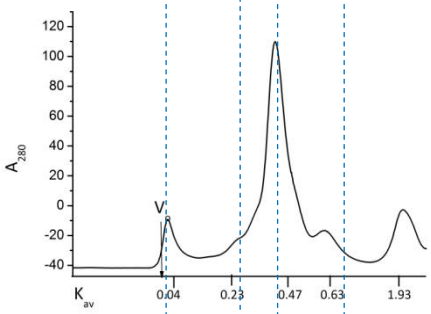
It was also observed that the majority of the muteins contained fraction II at K_{av} corresponding to ~ 120 kDa. This is suggestive of multimerisation of the rings into putative dimers, potentially caused by electrostatic interactions or perhaps suggestive of a different type of higher molecular weight species as reported in the case of *P. aerophilum* Sm3 [27]. Due to the limited resolving power of the S200 chromatography matrix used, it was difficult to isolate this fraction in sufficient quantity to carry out further analysis. For all muteins, fraction III was collected and subjected to SEC-SLS to determine the molecular weight and polydispersity of the proteins.

The preparative SEC provided an overview of the various oligomeric states adopted by Lsmα and its muteins. From **Figure 3.10**, a pattern appears, reflecting the distribution of oligomeric states in solution. Four distinct states seem to be common among the muteins analysed. A higher molecular weight (HMW) species eluted at the void volume, which was likely to be an aggregate or may represent an ordered form. A dimer of rings ($[7n_2]$), was observed for most samples, likely to be two rings associated. The majority of the muteins eluted at a molecular weight consistent with a single ring form. A small peak eluted at K_{av} corresponding to ~ 9 kDa which is consistent with the monomeric state of Lsmα, which is pronounced in the case of L70A. This state has not been otherwise observed for Lsm proteins.

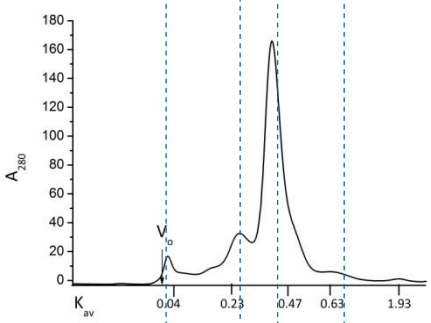
(A) N10A



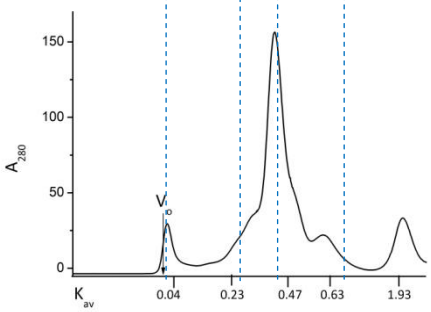
(B) Q12A



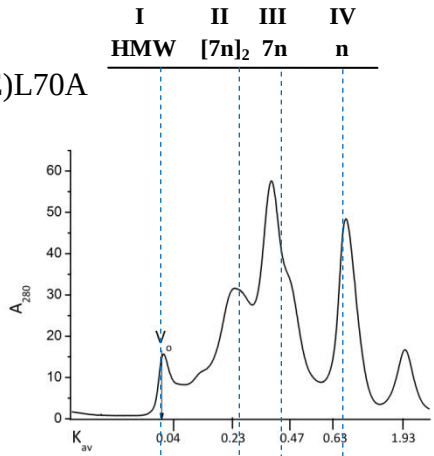
(C) R65P



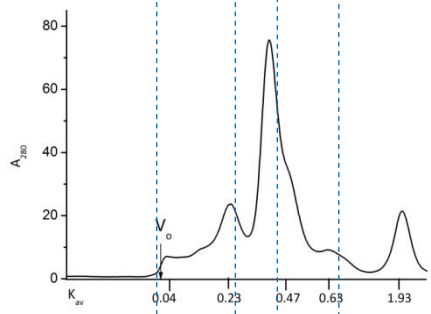
(D) V69A



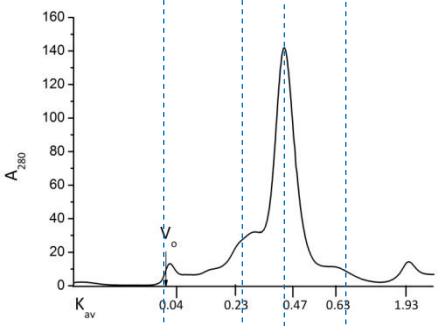
(E)L70A



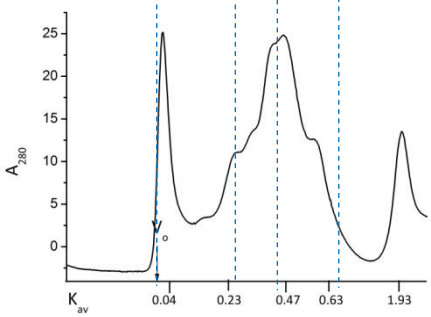
(F) I71A



(G) V77A



(H) Y78A



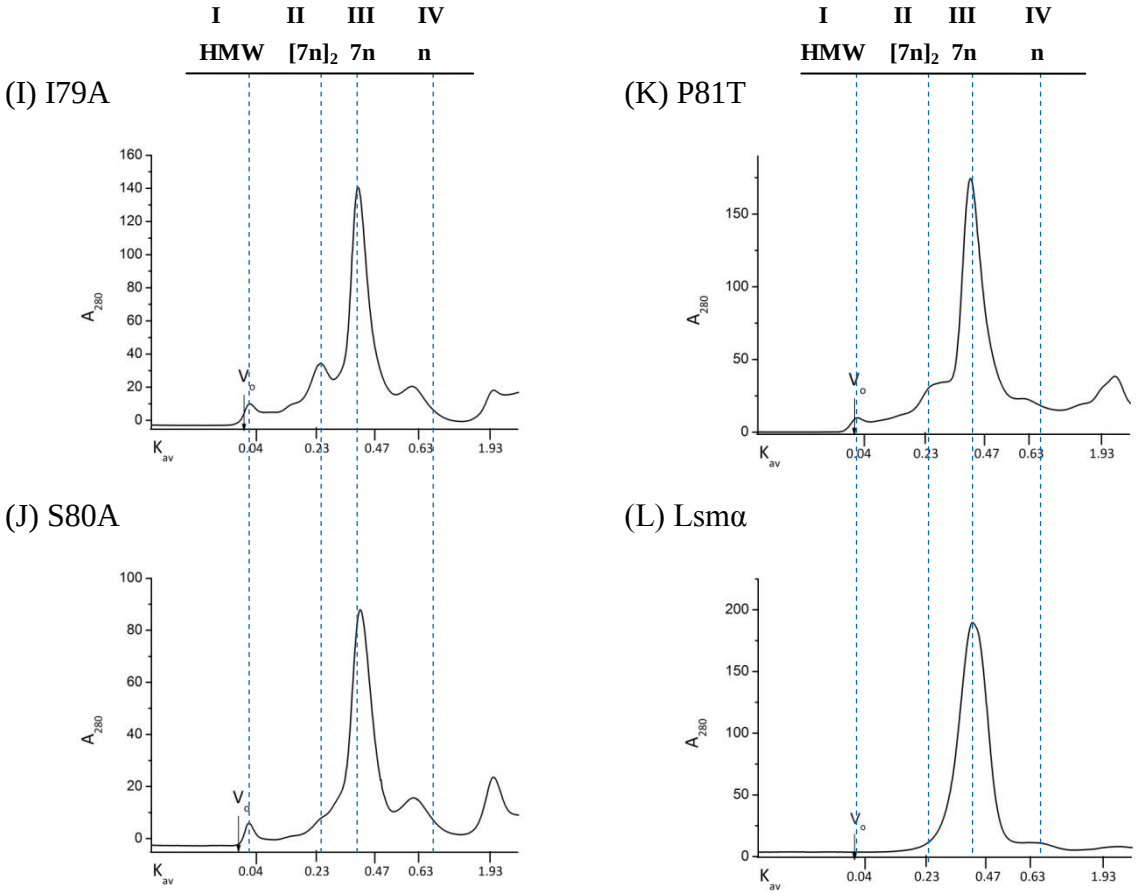
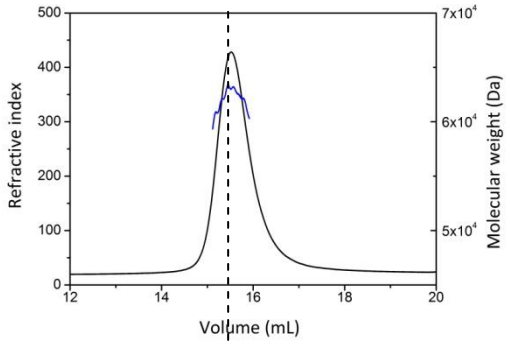


Figure 3.10| *Preparative SEC profiles of the single site directed muteins. All muteins were analysed in PBS pH 8.0 on a Superdex 200 16/600 column at a flow rate of 1 mL/min. Fraction numbers are indicated above. Dashed lines represent the higher molecular weight species (HMW), dimer of heptameric rings ([7n]$_2$), heptameric ring (7n), and the monomeric form (n). The molecular weight estimated from preparative SEC is indicated in bracket A – N10A (53 kDa); B – Q12A (51 kDa); C – R65P (55 kDa); D – V69A (51 kDa); E – L70A (52 kDa); F – I71A (55 kDa); G – V77A (50 kDa); H – Y78A (41 kDa); I – I79A (54 kDa); J – S80A (52 kDa); K – P81T (50 kDa); (L)-wildtype Lsmα (60 kDa).*

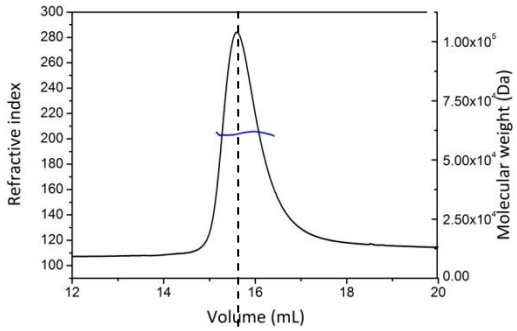
3.9.2 Static light scattering of muteins

SEC-SLS was performed on fraction III collected post preparative SEC. **Figure 3.11** shows the SEC-SLS results for the muteins in PBS pH 8.0. **Table 3.3** lists the molecular weight values calculated for the muteins taken at the midpoint of the eluted peak. Fraction III of the muteins from the preparative SEC eluted as single species, suggesting that the single ring form is stable and does not form higher molecular weight species. Except for the muteins N10A, Q12A, L70A, I71A and P81T, most of the samples showed a decreasing trend in the molecular weight (**Figure 3.11**). It may suggest a mixture of species that could not be separated due to the limitation of the S200 separating matrix. As the peaks do not show baselines, it was difficult to discern the decreasing molecular weight across the peak. The decreasing trend could also suggest that a hexameric species is present with the heptameric ring state. The typical range of the muteins lay between 61 kDa to 64 kDa, suggestive of heptameric states. R65P and V77A exhibited M_w range of 55 kDa to 57 kDa, which are closest to the theoretical molecular weight of a hexamer calculated from its amino acid sequence.

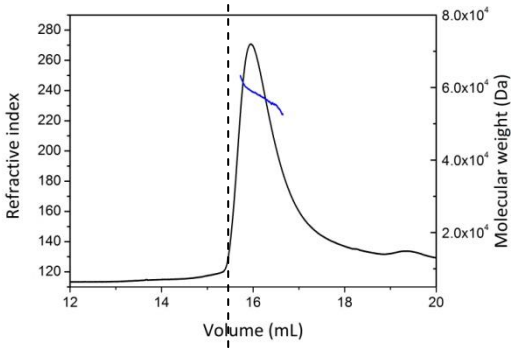
(A) N10A



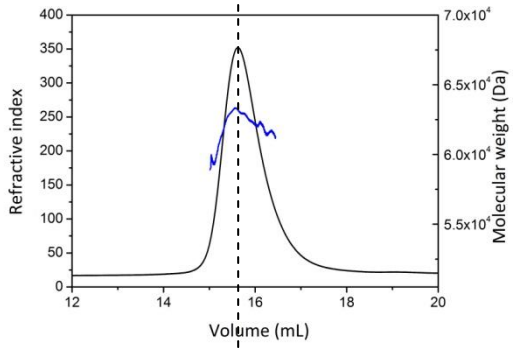
(B) Q12A



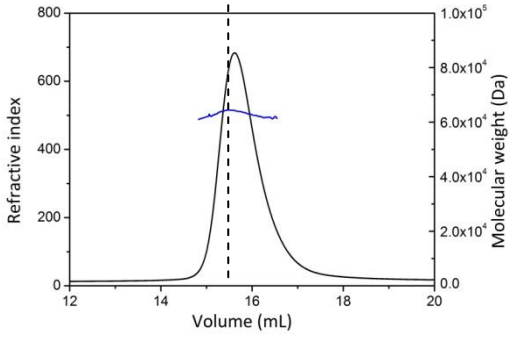
(C) R65P



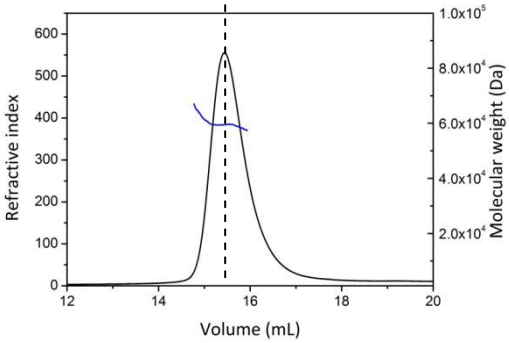
(D) V69A



(E) L70A

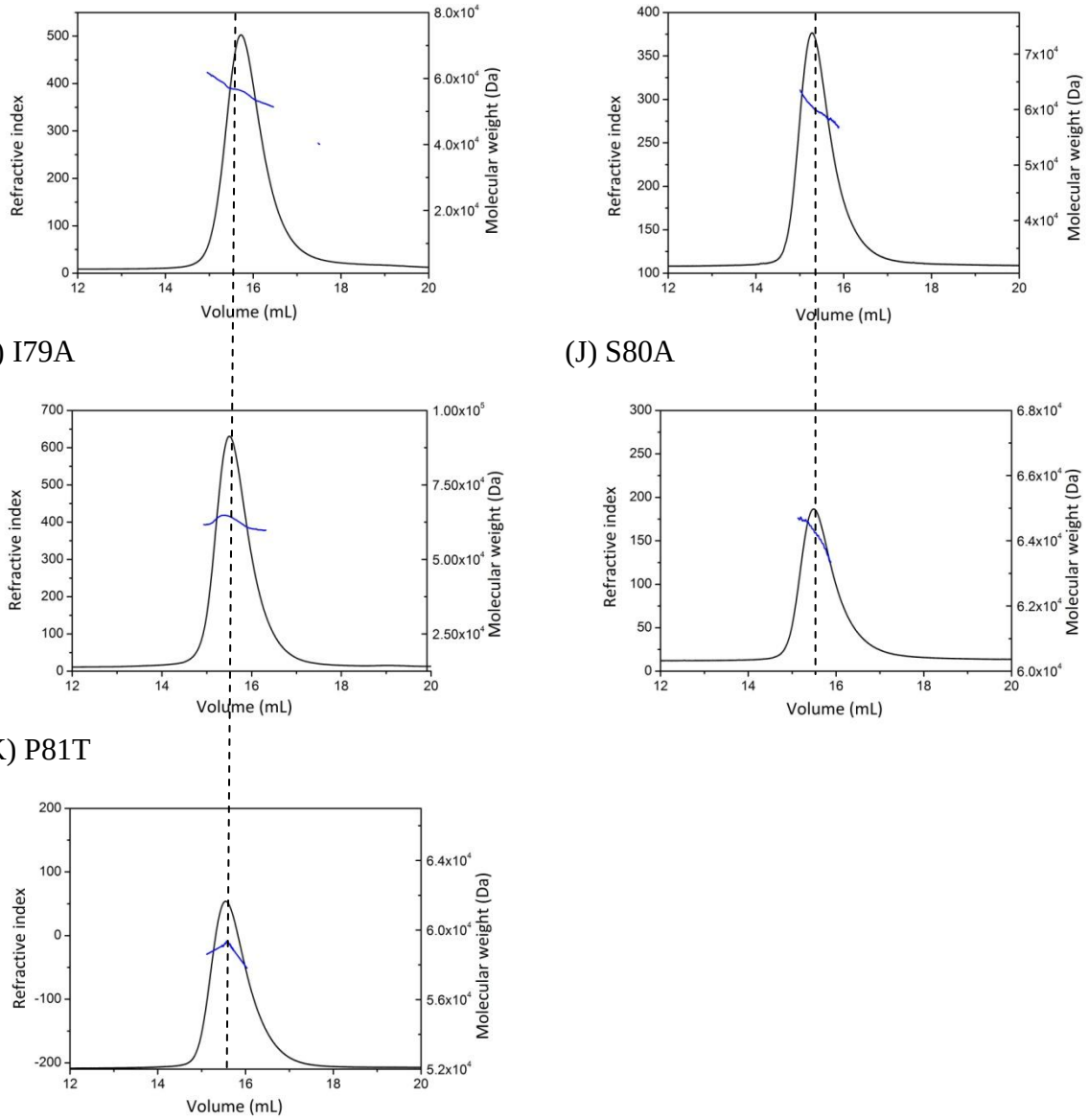


(F) I71A



(G) V77A

(H) Y78A



(I) I79A

(J) S80A

(K) P81T

Figure 3.11| *SEC-SLS of the muteins carried out in PBS pH 8.0 at a flow rate of 0.4 mL/min. The dashed line indicates the elution volume of the heptameric ring.*

Table 3.3| List of molecular weights (kDa) obtained for muteins in comparison to Lsmα

Sample	Molecular weight [a] (kDa)	Subunits[b]
Lsmα	62	6.8
N10A	62	6.9
Q12A	63	7.0
R65P	57	6.2
V69A	63	7.0
L70A	63	7.0
I71A	59	6.5
V77A	56	6.1
Y78A	61	6.7
I79A	64	7.0
S80A	64	7.0
P81T	59	6.5

All columns were pre-equilibrated with PBS pH 8.0.
[a]Superdex 200 GL 10/300 column was run at 0.4 mL/min. The M_w indicated was taken at the midpoint of the eluted peak.
[b]Calculated by dividing the molecular weight from SEC-SLS by theoretical monomeric weight of 9.1 kDa.

3.10 Small angle X-ray scattering by Lsmα and muteins

As mentioned in **Section 2.6.7**, the use of small angle X-ray scattering as a technique to study protein in solution is becoming well established. Due to the small wavelength of X-rays (< 1 nm), finer structural details can be resolved. SAXS gives a direct measurement of the average scattering by proteins in solution [28] and provides an accurate molecular weight calculation [29]. Since the data are collected in solution, potential matrix interactions are eliminated which can occur during chromatography.

Based on the molecular weight assessment by SEC-SLS, two muteins, R65P and V77A, were selected for SAXS analysis. Wildtype Lsmα and N10A were also subjected to SAXS as controls for the heptameric composition. V77A showed signs of visible precipitation during the transport of the protein sample and therefore, could not be analysed. Thus, Lsmα, N10A and R65P were analysed by SAXS.

Samples of wildtype Lsmα and the muteins were analysed using the SAXS beamline at the Australian Synchrotron. To ensure monodispersity, all samples were injected on a Superdex 200 5/50 gel filtration column in line with the SAXS beamline. Data were continuously collected as the samples eluted from the column (**Section 2.6.7**). The protein preparations' scattering intensities raw data were radially averaged to obtain one-dimensional scattering curves and scattering contributions of the buffer were subtracted from the sample scattering using the program ScatterBrain [30]. **Figure 3.12A** compares the experimental scattering intensities of the muteins with wildtype Lsmα.

The assemblies yielded scattering curves resembling elongated particles. To allow for a more precise definition of molecular dimensions, Guinier plots were derived using PRIMUS [31] (**Figure 3.12B**). The scattering of all the prepared samples yielded linear Guinier plots. The scattering did not show any upward curvature at low angles indicating absence of aggregation. Overall, however, good linearity signified data quality were sufficient for further processing and shape determination. The slopes of Guinier plots yield R_g values in reciprocal space (**Section 2.6.7.1**).

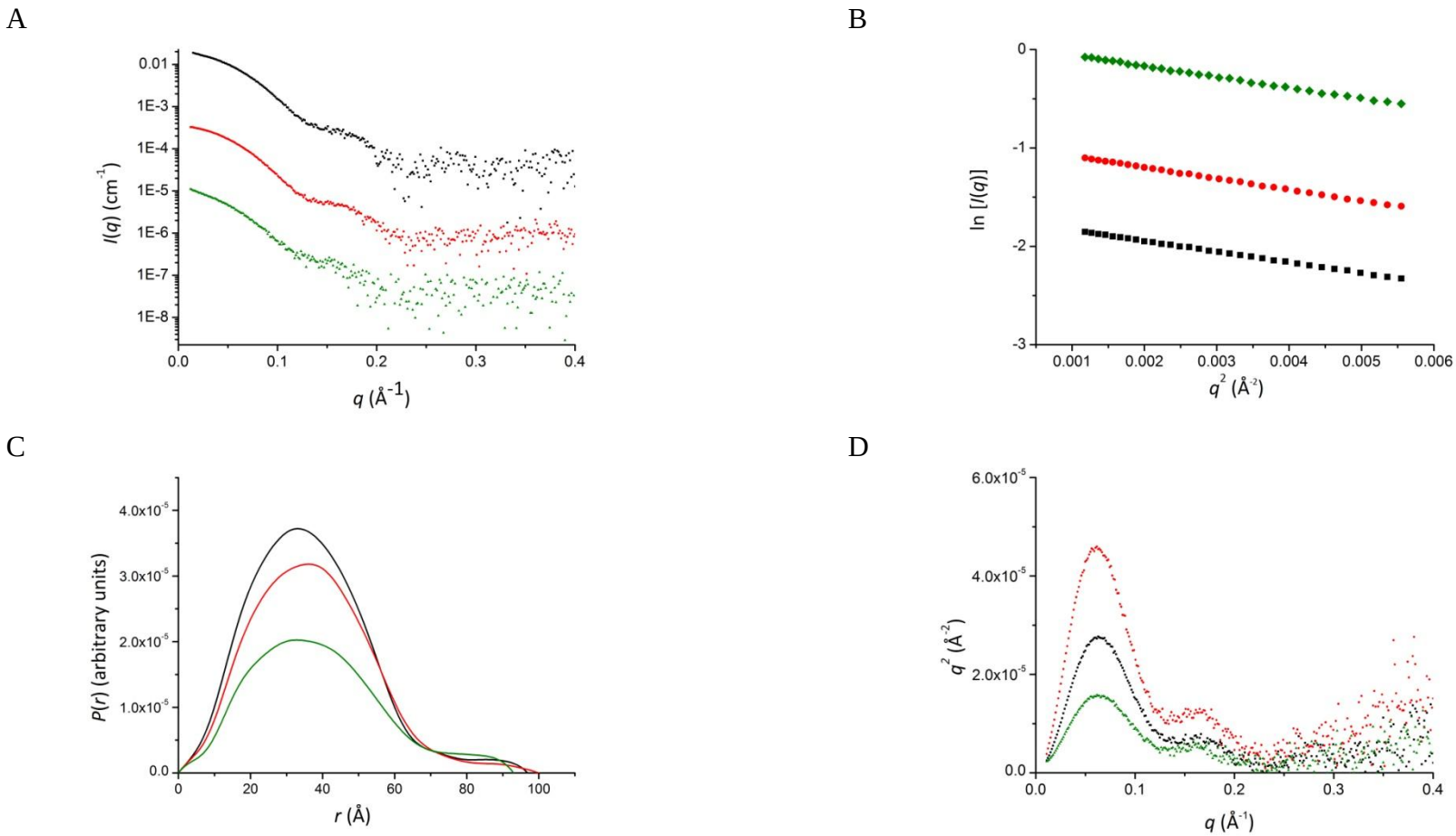


Figure 3.12| *(A) Scattering profiles, (B) Guinier plots, (C) Electron distribution function (P(r)), and (D) Kratky plots of Lsmα (black), N10A (red), and R65P (green). Curves have been arbitrarily displaced along the Y-axis for clarity in panel (A) and (B). Plots are normalised with respect to Y-axis in panel (C) and (D).*

Table 3.4| Geometric parameters derived from scattering data of Lsm proteins.

Sample	D_{max} [a] (nm)	R_g rec (nm) [a]	R_g real (nm) [b]	Volume [a] (nm^3)	M_w [a] (±0.1 kDa)	Subunits
Lsmα	9.6	2.7±0.1	2.8±0.2	99	62.0	6.7
N10A	9.9	2.8±0.2	2.8±0.2	99	63.0	6.9
R65P	9.2	2.9±0.6	3.0±0.3	95	55.0	6.0

[a] Calculated from Autoporod [32].

[b] Calculated from GNOM [33].

R_g^{rec}, reciprocal space radius of gyration derived from Guinier approximation; R_g^{real}, real space radius of gyration from distance distribution function; D_{max}, maximum particle diameter.

All data were obtained in PBS, pH 8.0.

For the Lsmα heptameric complex the R_g^{rec} is calculated as 2.7±0.1 nm and molecular weight of 62.0±0.1 kDa. A comparison of all parameters is listed in **Table 3.4**. Indirect Fourier transformation of collected data (GNOM [33]) produced a smooth bell-shaped electron pair distribution function ($P(r)$) for all complexes (**Figure 3.12C**). For all the complexes the results are consistent with heptameric ring formation and not with formation of any higher molecular species. However, R_g^{real} for R65P was 3.0±0.3 nm and the molecular weight calculated from the porod volume was 55.8±0.1 kDa, which corresponds to a hexameric composition.

In the case of R65P, the data reduction was performed over four sets of overlapping frames using ScatterBrain [30], with each set comprising of ten to fifteen frames of the scattering data, collected as the protein was eluting through the in-line size exclusion column. This allowed for molecular weight calculation covering the majority of the eluted fractions and concentrations, and provided with an accurate determination of the oligomeric state of R65P. The average weight obtained was 56 kDa, which is suggestive of a hexameric state of R65P.

Table 3.5| Geometric parameters derived for different elution fraction of R65P.

Number of frames	D_{max} [a] (nm)	R_g rec (nm) [a]	R_g real (nm) [b]	Volume [a] (nm^3)	M_w [a] (±0.1 kDa)	Subunits
7	10	2.9±0.6	2.9±0.3	91	55	6.1
10	9.8	2.9±0.6	2.9±0.3	96	56	6.1
10	9.9	2.9±0.6	2.9±0.3	96	56	6.1
7	9.6	2.9±0.6	2.9±0.3	96	56	6.1

[a] Calculated from Autoporod [32].

[b] Calculated from GNOM [33].

R_g^{rec}, reciprocal space radius of gyration derived from Guinier approximation; R_g^{real}, real space radius of gyration from distance distribution function; D_{max}, maximum particle diameter.

All data were obtained in PBS, pH 8.0.

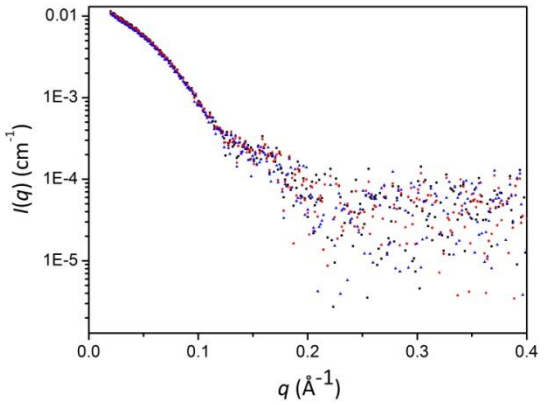


***Figure 3.13**| Overlay of scattering curves (non-scaled) of R65P taken from different positions of the elution peak, each representing a different concentration.*

The frames were chosen and analysed separately for the fractions. **Table 3.5** lists the parameters collected for different fractions of R65P. The molecular weight was calculated from the Porod volume using Autoporod [32]. The data suggest that different fractions of R65P have similar molecular weight and all exhibit hexameric composition. **Figure 3.13** shows the overlaid scattering curves of different fractions of R65P.

The real space radii of gyration are in good agreement with the reciprocal space data derived from the Guinier fits (**Section 2.6.7.1**). It is possible to deduce from these data that all

muteins exist in a folded globular state. All samples yield Kratky plots with a sharp peak at low *q* angles (**Figure 3.13D**). The peak at low angles was followed by a plateau, with no systemic signal increase indicating a folded state.

3.11 Generation of chimeric Lsm proteins

A second approach to control the pore size was tested by generating chimeric proteins, drawing from the sequence and structural information from different species. By studying the structural alignment of Lsm proteins, it was evident that the β4-, loop4 and β5-segments make up the protomeric interfaces of Lsm proteins [16; 34]. Out of these segments, the β4-strand and loop4 regions are the most variable (**Table 3.6**). If these segments are considered modular in nature, then a simple exchange between the Lsm proteins would change the bonding pattern at the interface but at the same time keep the monomer fold intact. The formation of the monomer should precede the oligomeric ring formation leading to altered assembly of monomers into different oligomeric states.

It was hypothesised that the difference in the molecular organisation is a result of altered angle of monomer packing and the differential extent of hydrogen bonding interactions along the interface (**Figure 3.14**). Exchange of sequences of these segments, β4 and β5, between the two protein species was trialled. Available crystal structures of Hfq (PDB ID 1HK9) and Lsmα (PDB ID 1I81) were studied, in particular the protomer interfaces. β4- and β5-segments of Hfq, namely VSQMV and ISTVVP respectively, were inserted into the Lsmα scaffold, to change the interface from that of a heptameric oligomer to a hexameric toroid. Such an arrangement was predicted to alter the bonding pattern at the protomeric interface and influence the angle between the monomeric modules and hence cause the assembly of a hexamer in place of a heptamer.

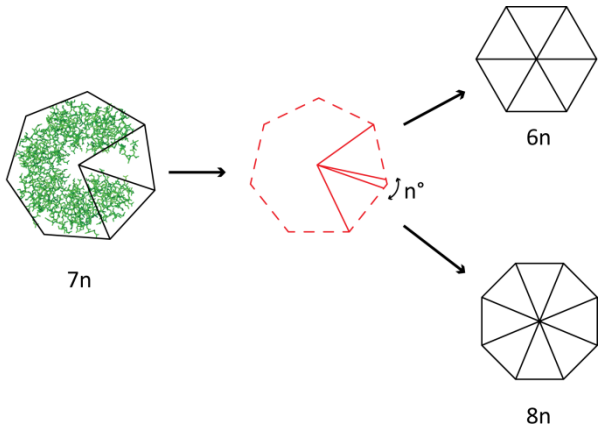


Figure 3.14| *Schematic showing the packing of monomers in a toroid. A change in the angle between the monomers was hypothesised to cause a change in the oligomeric state of the protein.*

Three such chimeric muteins were designed namely:

Chimera 1 – β4- and β5-strands were replaced with corresponding Hfq segments.

Chimera 2 – β4-strand was replaced with β4-segment of Hfq.

Chimera 3 – β5-strand was replaced with β5-segment of Hfq.

Table 3.6| Amino acid composition of Lsm monomer interface.

Protein	β4	loop5	β5
Ec Hfq	VSQMVY	KHA	ISTVVP
Mt Lsmα	EVTRRLGTVLI	RG	DNIVYISP
yeast Lsm3	ELSESERRCEMVFI	RG	DTVTLIST

The constructs were designed and appropriate plasmids ordered from Epoch Life Sciences and used to transform *E. coli* BL21 cells (**Section 2.4.1**). The small scale expression gave positive results with maximum expression at 6 h post protein expression by IPTG-induction. All three chimeric muteins were included for large scale production in 400 mL cultures. The

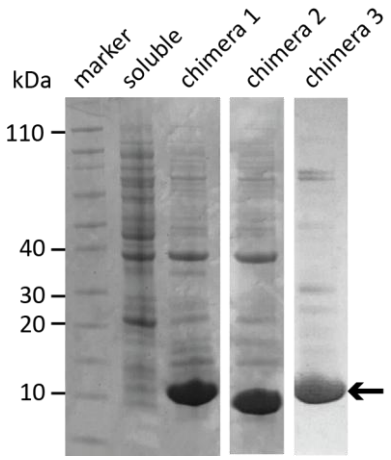


***Figure 3.15*|** *SDS-PAGE of chimeric muteins. All the protein aggregated and was present in the insoluble fraction (indicated by arrow).*

SDS-PAGE analyses, to check the solubility of the constructs, revealed that all the three chimeric muteins were insoluble, being obtained in the pellet and entirely absent from the supernatant. The changes introduced at the interface of the Lsmα scaffold caused that protein to form aggregates and not assemble into a soluble oligomer (**Figure 3.15**).

As explained in **Section 3.5**, the inter-subunit interactions are a complex network of hydrogen bonds and hydrophobic interactions. Although Lsm proteins show high levels of conservation at the amino acid sequence level, the structural assembly determines its organisation into a cyclic oligomer. As shown in the experiments above, a simple change of the interfacing residues prevents the ring assembly, indicating that the associations between each monomer is highly intricate and conserved between species. These results suggest that the natural assembly of Lsmα was perturbed when the protomeric interfaces were changed. In order to change its oligomeric state, it is necessary to refine the changes with additional site specific mutations of the relevant hydrophobic residues present at each protomeric interface.

3.12 Summary

Two approaches to change the oligomeric state and inner pore diameter of Lsmα were tested in this chapter. Single residue mutations were made in the subunit interface to probe the bonding between the monomers. In the second approach, chimeric proteins were generated by changing the β4- and β5-segments of Lsmα to mimic that of a hexameric protein. Generation of chimeric proteins did not yield soluble protein and hence the muteins with single residue change at the interface were tested for change in the oligomeric state by biophysical techniques.

The oligomeric distribution of the proteins changed with the introduction of the alanine residues as assessed by preparative SEC. It was observed that the prominent fraction of the muteins exhibited molecular weight similar to that of the wildtype heptameric Lsmα. A monomeric state was exhibited by the mutein L70A as observed by SEC and SDS-PAGE. It would be interesting to optimise the conditions to stabilise this monomeric form.

Muteins R65P and V77A exhibited molecular weights that are consistent with a hexameric state as determined by SEC-SLS and in the case of R65P, confirmed by SAXS. In order to unambiguously distinguish a heptamer and a hexamer, future detailed work should include X-ray crystallography and native mass spectrometry.

It is intriguing to look into the properties of the R65P mutein because the residue is part of the β4-strand, with its side-chain projecting outwards. The change to proline eliminates the positive charge and perhaps causes a bend in the strand, thereby inducing a change in the angle at the interface. This may lead to the monomers associating as a hexamer.

3.13 References

1. Phillips, A. J., Littlejohn, J., Yewdall, N. A., Zhu, T., Valery, C., Pearce, F. G., Mitra, A. K., Radjainia, M. & Gerrard, J. A. (2014). Peroxiredoxin is a versatile self-assembling tecton for protein nanotechnology. *Biomacromolecules* **15**, 1871-1881.
2. Ardini, M., Giansanti, F., Di Leandro, L., Pitari, G., Cimini, A., Ottaviano, L., Donarelli, M., Santucci, S., Angelucci, F. & Ippoliti, R. (2014). Metal-induced self-assembly of peroxiredoxin as a tool for sorting ultrasmall gold nanoparticles into one-dimensional clusters. *Nanoscale* **6**, 8052-8061.
3. Woolfson, D. N. & Mahmoud, Z. N. (2010). More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chemical Society Reviews* **39**, 3464-3479.
4. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R., Curmi, P. M. G. & Mabbutt, B. C. (2008). Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: implications for Lsm ring organisation and recruitment. *Journal of Molecular Biology* **377**, 1357-1371.
5. Audette, G. F. & Hazes, B. (2007). Development of protein nanotubes from a multi-purpose biological structure. *Journal of Nanoscience and Nanotechnology* **7**, 2222-2229.
6. Astier, Y., Bayley, H. & Howorka, S. (2005). Protein components for nanodevices. *Current Opinion in Chemical Biology* **9**, 576-584.
7. Grünberg, R. & Serrano, L. (2010). Strategies for protein synthetic biology. *Nucleic Acids Research* **38**, 2663-2675.
8. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M. G. & Mabbutt, B. C. (2001). Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs1. *Journal of Molecular Biology* **309**, 915-923.
9. Chen, C. S., Smits, C., Dodson, G. G., Shevtsov, M. B., Merlino, N., Gollnick, P. & Antson, A. A. (2011). How to change the oligomeric state of a circular protein assembly: switch from 11-subunit to 12-subunit TRAP suggests a general mechanism. *PloS One* **6**, e25296.
10. Wang, W. X., Dgany, O., Wolf, S. G., Levy, I., Algom, R., Pouny, Y., Wolf, A., Marton, I., Altman, A. & Shoseyov, O. (2006). Aspen SP1, an exceptional thermal,

protease and detergent resistant self assembled nano particle. *Biotechnology and Bioengineering* **95**, 161-168.

11. Braig, K., Adams, P. & Brünger, A. (1995). Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nature Structural Biology* **2**, 1083-1094.
12. Nishen, N. (2006). Molecular structures of the RNA-binding Lsm proteins (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
13. Zarić, B. L., Jovanović, V. B. & Stojanović, S. Đ. (2011). Non-covalent interactions across subunit interfaces in Sm proteins. *Journal of Theoretical Biology* **271**, 18-26.
14. Zaric, B., Chami, M., Rémigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F. K. & Kambach, C. (2005). Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *Journal of Biological Chemistry* **280**, 16066-16075.
15. Wu, D., Jiang, S., Bowler, M. & Song, H. (2012). Crystal Structures of Lsm3, Lsm4 and Lsm5/6/7 from *Schizosaccharomyces pombe*. *PloS One* **7**, e36768.
16. Mura, C., Randolph, P. S., Patterson, J. & Cozen, A. E. (2013). Archaeal and eukaryotic homologs of Hfq. *RNA Biology* **10**, 636-651.
17. Collins, B. M. (2001). Structural studies of archaeal and yeast Sm/Lsm proteins (PhD thesis). *Department of Chemistry, Macquarie University*.
18. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774-797.
19. Sobti, M., Cubeddu, L., Haynes, P. A. & Mabbutt, B. C. (2010). Engineered rings of mixed yeast Lsm proteins show differential interactions with translation factors and U-Rich RNA. *Biochemistry* **49**, 2335-2345.
20. Collins, B. M., Cubeddu, L., Naidoo, N., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M. G. & Mabbutt, B. C. (2003). Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA crystal structure of an oligomeric complex of yeast SmF. *Journal of Biological Chemistry* **278**, 17291-17298.
21. Morrison, K. & Weiss, G. (2001). Combinatorial alanine-scanning. *Current Opinion in Chemical Biology* **5**, 302-305.

22. Massova, I. & Kollman, P. (1999). Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *Journal of the American Chemical Society* **121**, 8133-8143.
23. Törö, I., Basquin, J., Teo-Dreher, H. & Suck, D. (2002). Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. *Journal of Molecular Biology* **320**, 129-142.
24. Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **60**, 523-533.
25. Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expression and Purification* **41**, 98-105.
26. Burgess, S. A., Walker, M. L., Thirumurugan, K., Trinick, J. & Knight, P. J. (2004). Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *Journal of Structural Biology* **147**, 247-258.
27. Mura, C., Phillips, M., Kozhukhovsky, A. & Eisenberg, D. (2003). Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4539-4544.
28. Graewert, M. A. & Svergun, D. I. (2013). Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Current Opinion in Structural Biology* **23**, 748-754.
29. Petoukhov, M. V. & Svergun, D. I. (2013). Applications of small-angle X-ray scattering to biomacromolecular solutions. *The International Journal of Biochemistry & Cell Biology* **45**, 429-434.
30. Kirby, N., Mudie, S. & Hawley, A. (2013). A low-background-intensity focusing small-angle X-ray scattering undulator beamline. *Journal of Applied Crystallography* **46.6**, 1670-1680.
31. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282.

32. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Applied Crystallography* **40**, 223-228.

33. Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495-503.

34. Murina, V. & Nikulin, A. (2011). RNA-binding Sm-like proteins of bacteria and archaea. Similarity and difference in structure and function. *Biochemistry (Moscow)* **76**, 1434-1449.

4 Chapter four

Engineering tubular structures of Lsmα

4.1 Introduction

Nanotubes have a number of potential uses in nanotechnology, such as biocontainers for the controlled release of drug formulations, or for applications in materials and electronics [1]. Examples of naturally occurring protein tubes include microtubules, pili, and flagella. However, these multi-protein structures are not ideal to work with as it is difficult to modify the structure and control their assembly *in vitro*. Therefore, nanometric protein structures are preferable, with a smaller diameter that can be easily assembled from protein units, expressed in large quantities, and have ease of surface modification to program the length of the nanotube.

This chapter discusses two approaches to promote stacking of Lsmα rings into nanotubes.

(i) Surface modification based on electrostatic-mediated stacking of the rings.

(ii) Surface modification based on covalent linking of the rings by disulfide bridges.

The Lsmα variants designed were characterised by SEC and TEM. If Lsmα is to be used as a molecular building block, there are number of issues that need to be addressed in order to make it more attractive for downstream applications. One of these properties is being able to control the supramolecular assembly by altering the ionic or redox conditions of the buffer to effectively provide a switch to control the quaternary structure arrangement. These studies are discussed in this chapter.

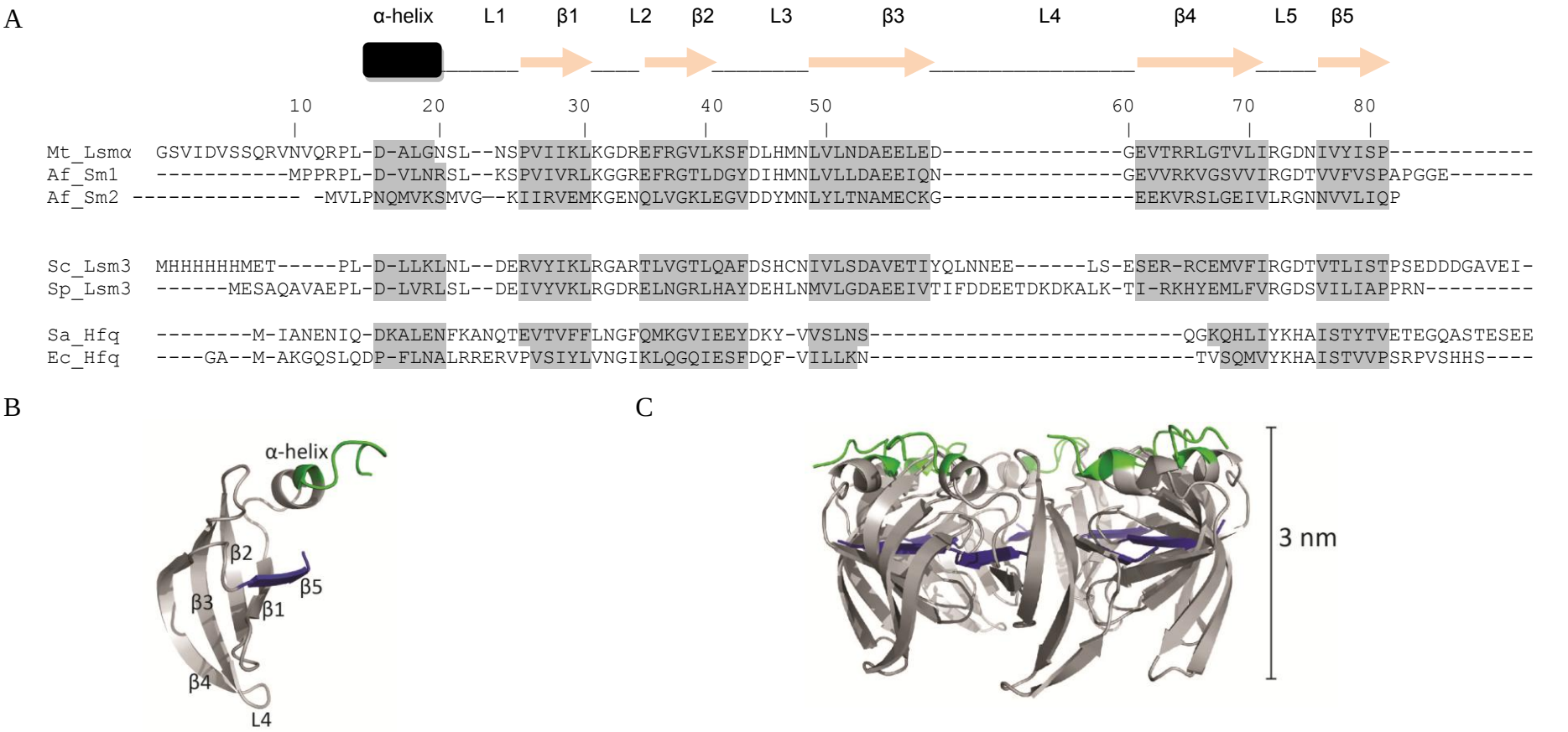


Figure 4.1| *(A) Structural alignment of Lsm proteins. The secondary structures are assigned with respect to the Lsmα protein. The species included in the alignment are: Mt Lsmα-M. thermoautotrophicum; Af Sm1 and Af Sm2-A. fulgidus; Sc Lsm3-S. cerevisiae; Sp Lsm3-S. pombe; Sa Hfq- S. aureus; Ec Hfq-E. coli. (B) Ribbon representation of Lsmα fold showing helix and β-strand elements. L4 is the loop between β4- and β5-strands. (C) Side view of Lsmα with a height of 3 nm.*

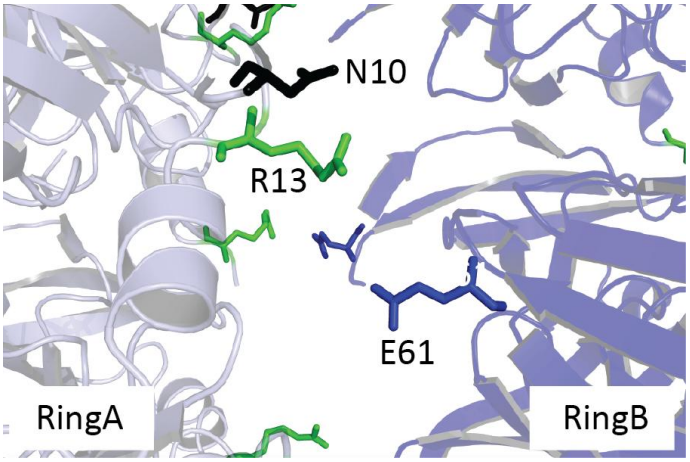
4.2 Polymerisation of Lsm proteins

Lsm proteins have the ability to assemble into fibres under certain conditions (**Section 1.9**). In the case of Lsm proteins of archaeal origin, also termed Sm-like archaeal proteins (SmAPs), from *M. thermoautotrophicum* and *P. aerophilum*, both form bundles of more than 200 nm in length and about 100 nm in width [2]. *Pae* SmAP1 (PDB ID 1LNX) has the tendency to form a disulfide-bonded dimer of rings via the loop-to-loop face. But under low ionic strength buffering and reducing conditions, the protein formed polymerised fibres [2] (**Figure 1.9E**). In the case of *Mt* SmAP1 (PDB ID 1JBM), the protein also assembles into striated fibres as studied by TEM and by examining the crystal structure packing (**Figure 1.9D**). The bacterial homologue of Lsm protein, Hfq, also exhibits the tendency to form large macromolecular structures [3]. It is interesting to point out the difference between the assemblies of the fibres in the cases mentioned above. For *Mt* SmAP1, it is head-to-tail stacking of the homoheptamers, whereas the Hfq fibres are formed in layered organisation, each layer formed by a hexamer of the hexameric rings [3]. These structures provide insight into the packing of Lsm rings, and inform the design of a controlled bottom-up fabrication of the rings in discrete steps.

4.3 Design rationale to engineer protein nanotubes

As mentioned in **Section 3.1**, Lsmα exists as a homomeric ring containing seven repeating subunits, associated through a network of hydrophobic interactions and hydrogen bonds between β-strands lining adjacent protomers. Due to the radial symmetry of the associated monomers, a toroidal face is created containing seven N-terminal helical elements, termed the helix face, and the obverse surface contains the loop4, of each monomer, and is termed as loop face. Thus, the ring scaffold provides four surfaces that

A



B

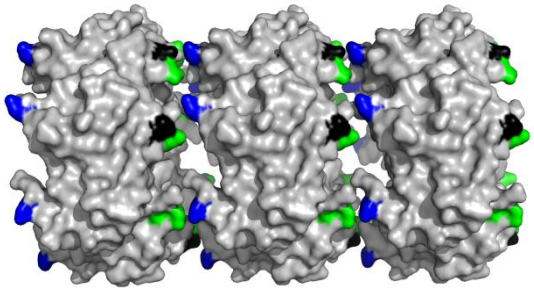


Figure 4.2| *(A) Inter-ring interface of Lsmα showing N10 (black), R13 (green), and E61 (blue) as stick representation. (B) Surface representation of stacks formed in the crystal lattice of Lsmα (PDB ID 1I81) showing the location of the residues mutated.*

can be modified for manipulation of the quaternary organisation of the rings. The helix and loop faces are easily accessible and provide a large surface area for ring-to-ring binding.

From the structural alignment of the Lsm sequences (**Figure 4.1A**), it is known that the N-terminal and loop4 regions are the most variable segments. The oligomerisation of the ring itself is largely dependent on the interactions between the β4- and β5-strands as discussed in Chapter three. As highly exposed structural segments, the N-terminal and loop4 regions were chosen as mutation sites and thereby targeting the opposite faces of the Lsmα tecton. Selecting the variable structural segments for the mutations was not predicted to cause any changes to the heptameric toroidal scaffold. Even though the N-terminal residues are part of the hydrophobic pockets stabilising the subunit interface, care was taken to select residues

that are solvent exposed and facing outward to allow probable interaction with another Lsmα ring module.

A study of the interface between paired ring modules from the crystal packing in X-ray crystallographic structures (PDB 1I81) suggested that the position of side chains of Asn-10 and Arg-13, located at N-terminal helix and Glu-61 located within loop4 were well-suited for Ala and Cys substitutions (**Figure 4.2A**). Hence, post assembly of the ring, there would be seven substitutions at the helix face and the loop face of the scaffold, thereby conferring the ring with seven potential attachment points for conjugation. In order to have additional control on the stacking of these rings, each site was separately modified with double and single substitutions.

(i) Electrostatically-mediated ring stacking

Electrically charged and acidic residues on the N-terminal helix and loop face were chosen which were exposed towards the outside solvent. Single (R13A) and double (N10A/E61A) muteins were generated by changing specific residues to Ala. R13 and N10 are part of the helix face, whereas, E61 is located on the loop face. The archaeal Lsmα structure shows a distinct polarity in its electrostatic potential, with loop4 side being more acidic than the helix face (**Figure 4.3**). It was hypothesised that altering the surface charge would cause multimerisation of the rings, as previously observed in the case of Lsm polyproteins [4], thereby allowing control over the stacking of rings.

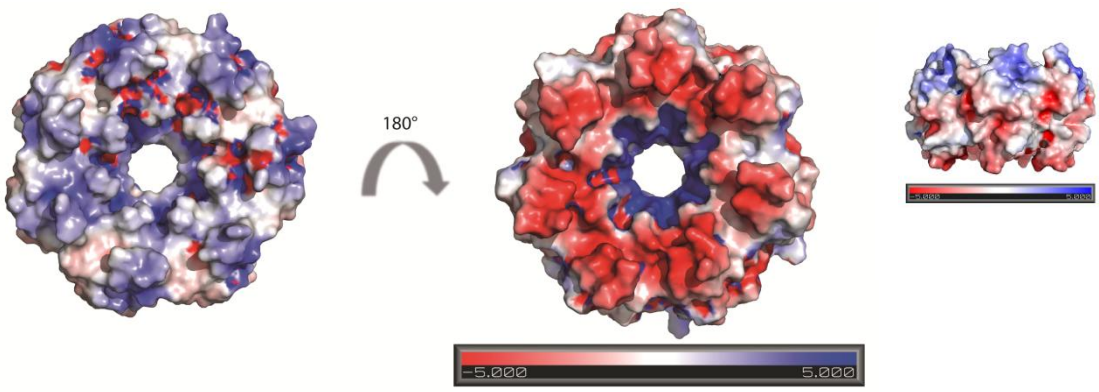


Figure 4.3| *Surface electrostatics of the Lsmα ring complex showing the helix face and loop faces. Inset shows the side view of the ring. Structures were generated in Pymol* [5] *and electrostatic surfaces were computed with Adaptive Poisson-Boltzmann Solver (APBS) algorithm* [6] *and contoured so that red and blue represent -5kT/e and 5kT/e respectively.*

(ii) Disulfide-mediated ring stacking

Doughnut-shaped proteins have been engineered to polymerise into tubes by introducing Cys residues on the surface of the ring scaffold. There are two literature precedents for this, each using rings of larger dimensions than the Lsmα tecton presented here. The first case was TRAP protein from *B. stearothermophilus*, which polymerised in the presence of the reducing agent, DTT, acting as a cross-linker bridging the rings together [7]. The second example is the HCP1 protein, a secretary protein from *P. aeruginosa*, which was shown to assemble as stacks by disulfide conjugation. Furthermore, the authors were able to plug the ends of the tubes, thereby creating a hollow protein capsule [8]. This work served as an inspiration for creating Lsmα tubes. A similar strategy was employed for polymerisation of Lsmα rings.

A scheme of the possible combinations of stacks that can be formed is depicted in **Figure 4.4**. Since the Lsmα ring exhibits two chemically distinct helix and loop faces, three possible stacking events are hypothesised; helix-to-loop, loop-to-loop, and helix-to-helix. This combination can either be coaxial or skewed in nature. The polymerisation of these basic stacked units can give rise to an extended tubular structure of varying length.

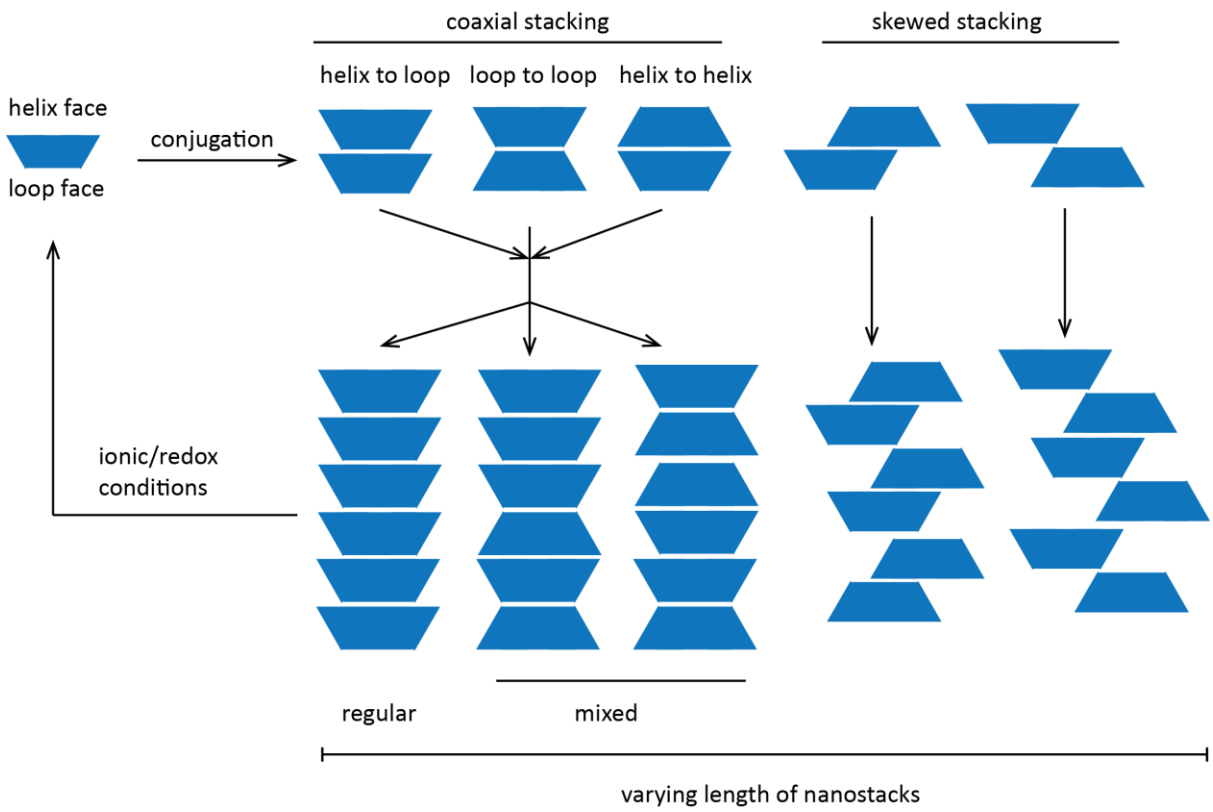


Figure 4.4| *Schematic showing the different possible combinations of the modified Lsmα rings for polymerisation into tubes.*

4.4 Preparation and isolation of recombinant protein complexes

The muteins were generated as GST-fusion products. It was shown in Chapter three, that this strategy produces sufficient yield of the wildtype and the muteins with easy on-column cleaving of the GST-tag. Although the GST-tag was attached to each of the monomers, it did not interfere with the assembly of the oligomeric ring. Therefore, as a successful purification methodology, the designed Lsmα muteins were generated as GST-fusion products.

Five GST-fusion constructs namely R13A, N10A/E61A, N10C, E61C and N10C/E61C were designed and the genes cloned in pGEX plasmids coding for an N-terminal GST-tag (**Section 2.5.3**). These constructs were ordered from Epoch Life Sciences (**Section 2.2.3**) and used to transform a conventional strain of *E. coli* cells. The protein products were expressed in *E. coli* BL21 cells following induction by IPTG (0.2 mM). The yield was monitored by SDS-PAGE analysis and was found to be optimal after growth at 26 °C for 8 h post-induction. Typically 5-10 mg of protein was produced per 400 mL of growth media. All three constructs showed good levels of expression appropriate for further large scale preparation.

All the muteins were successfully recovered using the glutathione Sepharose affinity matrix (**Section 2.5.3**). **Figure 4.5** shows the SDS-PAGE gel for the various steps of purification for E61C. Similar yields were obtained for R13A, N10A/E61A, N10C and N10C/E61C. The fusion proteins were cleaved with thrombin whilst still attached to the column and then subjected to preparative SEC to ascertain their size distribution and to separate them from contaminating thrombin and small amounts of GST.

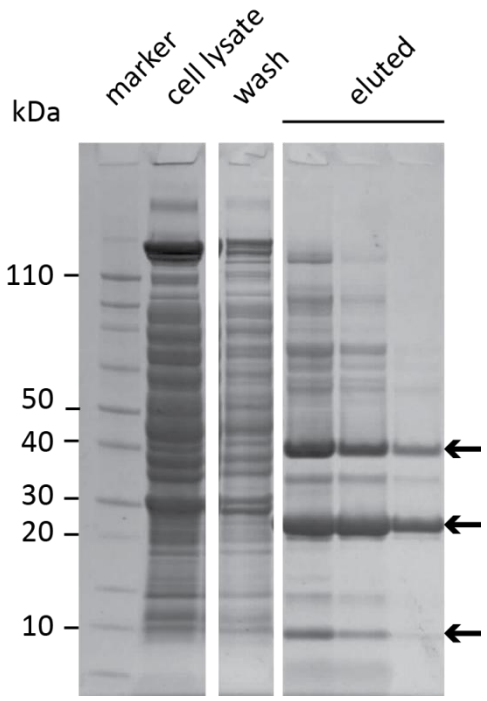


Figure 4.5| *SDS-PAGE gel showing the affinity purification steps of recombinant E61C from E. coli. The cell lysate shows the complete protein content in the cell after induction of protein expression by IPTG. The wash is glutathione affinity matrix wash. The eluted fraction represents the protein released following thrombin cleavage of bound fusion protein. Bands are obtained at ~9 kDa (monomer), ~25 kDa and 40 kDa (non-dissociated oligomers).*

4.5 Characterisation of R13A and N10A/E61A

R13A and N10A/E61A were subjected to analytical SEC in different salt concentrations. **Figure 4.6** shows the SEC traces obtained for both the muteins. A phosphate buffer was used at pH 8.0, at three salt concentrations i.e. 20 mM, 200 mM and 800 mM. The samples were collected after preparative SEC and dialysed in the appropriate buffer for 12 h at 4 °C (**Section 2.6.2**). For all the conditions tested, the elution volume of the samples remained at the K_{av} corresponding to a single ring (~ 63 kDa). Therefore, the stacking of Lsmα muteins was not induced. As a control, the wildtype Lsmα was also dialysed into the above mentioned buffers. It did not exhibit any change in the supramolecular assembly with change in the salt concentration. R13A and N10A/E61A were also dialysed against buffers with pH 7, 8, and 9.

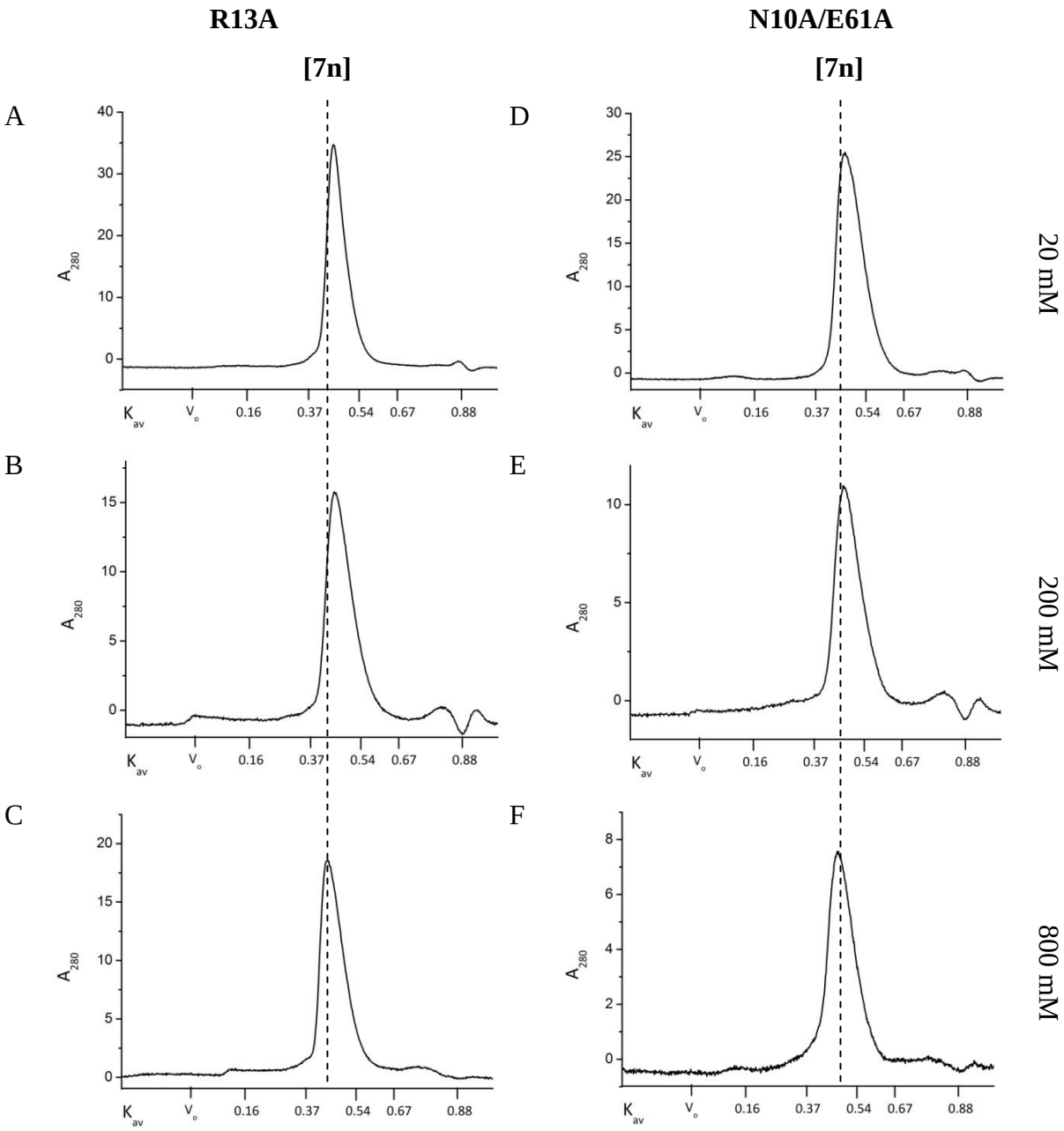
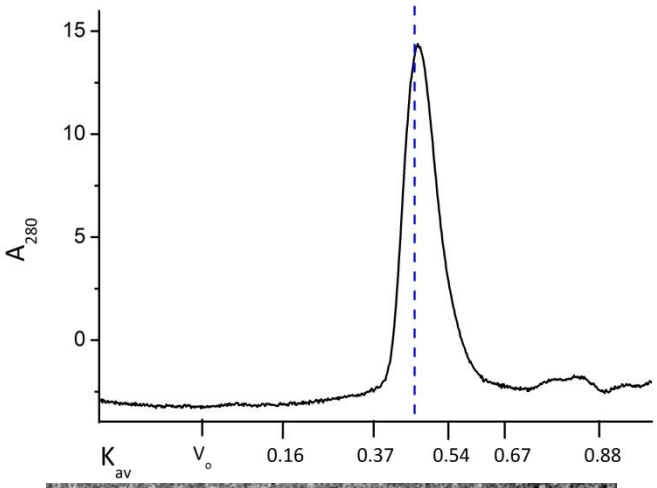


Figure 4.6| *Analytical SEC of R13A and N10A/E61A in different salt concentrations. (A), (B), and (C) show the mutein R13A in 20 mM, 200 mM, and 800 mM NaCl, respectively. (D), (E), and (F) show the mutein N10A/E61A in 20 mM, 200 mM, and 800 mM NaCl, respectively. The SEC column was equilibrated with PBS, pH 8.0 and run at a flow rate of 0.4 mL/min. The dashed line represents the single ring [7n] elution position.*

The samples were analysed by analytical SEC and did not show any change in the molecular weight (data not shown).

As mentioned in **Section 4.2**, fibril structures of Lsmα were obtained when dialysed into a low ionic strength buffer [2]. Similar conditions were tested on Lsmα by dialysing the protein against 10 mM Tris, pH 7.5, 60 mM NaCl for 12 h at room temperature and tested by analytical SEC and TEM. In the low ionic strength buffer, the Lsmα behaved as a single ring and did not undergo any change into higher order structures (**Figure 4.7A, B**).

A



B

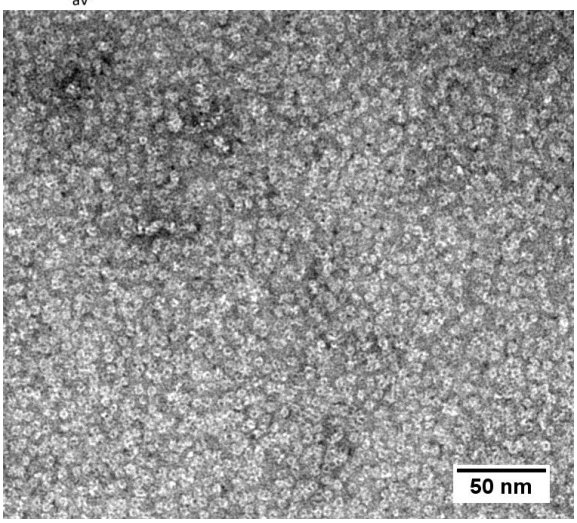


Figure 4.7| *(A) Analytical SEC chromatograph of Lsmα. Dashed line represents elution volume of single ring. (B) TEM of Lsmα dialysed in 10 mM Tris, pH 7.5, 60 mM NaCl.*

4.6 Disulfide-mediated ring stacking

In order to induce the formation of supramolecular assemblies in which intact rings are covalently associated, faces of the protein ring were modified with Cys residues to allow disulfide conjugation. The native Lsmα protein sequence does not contain any Cys residues, ensuring there will be no interference with these newly engineered thiol groups. The three Cys muteins, namely N10C, E61C, and N10C/E61C were characterised by SEC and TEM.

4.6.1 Characterisation of E61C

E61C existed as a mixture of species as judged by preparative SEC under non-reducing conditions in a phosphate buffer at pH 8.0. The trace obtained (**Figure 4.8A**) shows a distribution of three species. Fraction I eluted in the void volume, and is most likely indicative of aggregated product formed during the affinity purification step.

Fraction II eluted at K_{av} 0.3 (~110 kDa) corresponding to 2 rings that are perhaps associated via an axial or random stacking mechanism. However, the SDS-PAGE gel of the sample showed little protein content for fraction II. The majority of the sample eluted at K_{av} 0.4 corresponding to ~ 60 kDa which is the single ring form. Fraction III appeared as three bands at 9 kDa (monomeric form), and 40 kDa and 60 kDa (thermostable oligomers) (**Figure 4.8B**). Since there are no bands observed at higher molecular weights, it can be concluded that E61C (fraction III) exists predominantly as a toroidal ring with no sulfhydryl associations in non-reducing conditions.

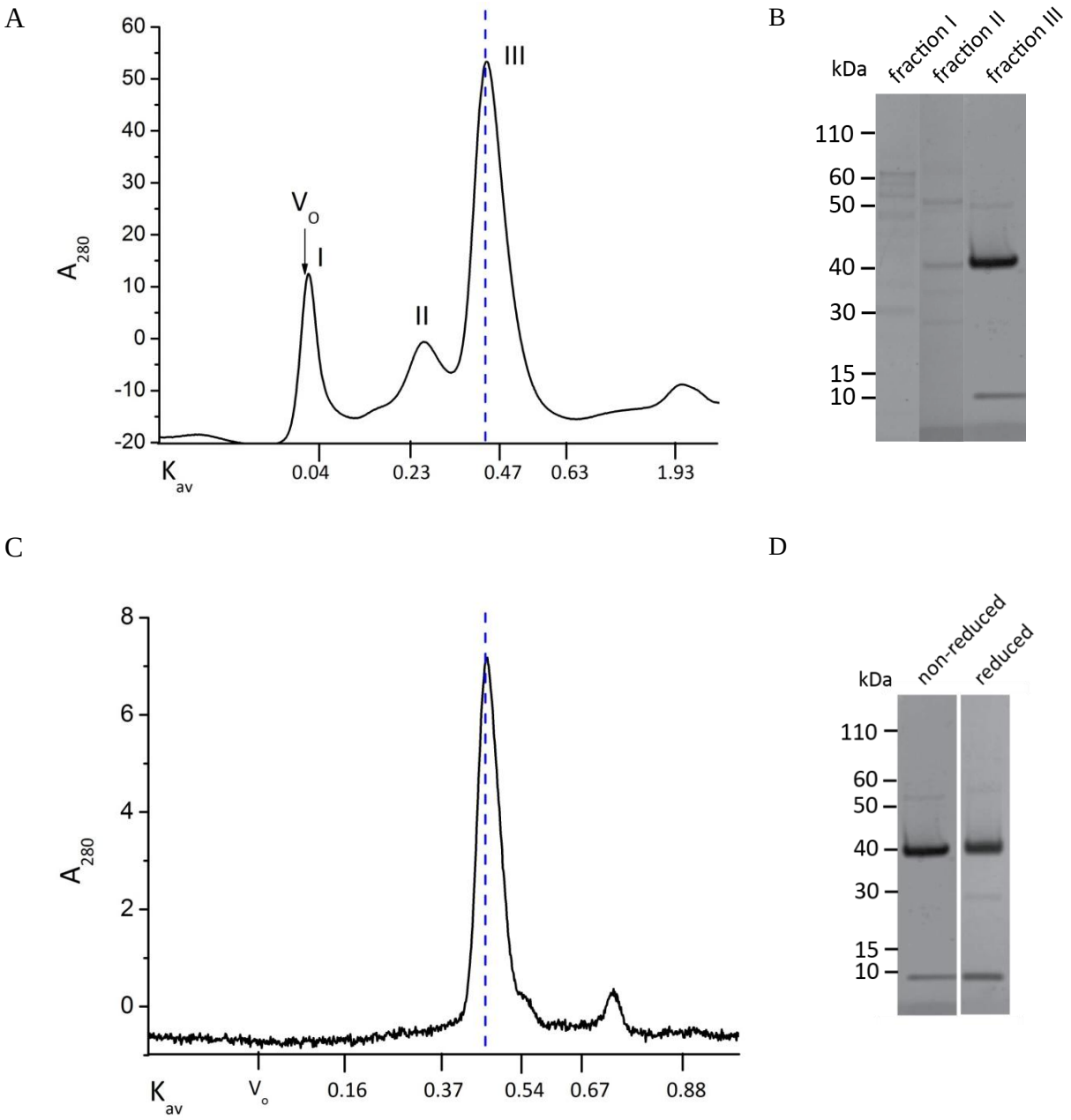


Figure 4.8| *SEC chromatograph of E61C in non-reducing conditions. (A) Preparative grade SEC (Superdex 200 16/600) pre-equilibrated with 10 mM PBS pH 8.0 and run at a flow rate of 1 mL/min. (B) SDS-PAGE of fractions obtained on preparative SEC. (C) Analytical SEC run with same buffer conditions at a flow rate 0.4 mL/min. The dashed line indicates the fraction with K_{av} corresponding to a single ring. (D) SDS-PAGE of E61C in non-reducing and reducing conditions.*

To check if E61C interconverts into other species, fraction III was collected and injected in an analytical SEC under reducing and non-reducing conditions.

The samples were dialysed into PBS, pH 8.0 with and without reducing agent for 2 h at room temperature prior to loading onto the SEC column. **Figure 4.8C** and **Figure 4.9** shows the SEC traces obtained for E61C. Under both conditions the sample eluted as a single species with a K_{av} 0.4 (~ 63 kDa), confirming that the species is present in the single ring state. The SDS-PAGE showed the characteristic three bands. From the size distribution of the sample, it is evident that modification of the loop face did not result in disulfide conjugation. Therefore, in the case of E61C, no higher order assembly was observed.

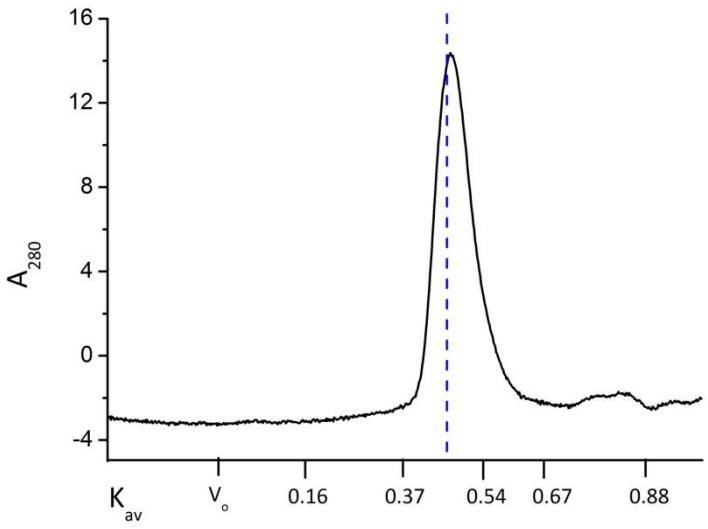


Figure 4.9| *Analytical SEC of E61C in reducing conditions (10 mM TCEP) at a flow rate of 0.4 mL/min. The dashed line indicates the position of a single ring.*

4.6.2 Characterisation of N10C

The preparative SEC trace in non-reducing conditions obtained for N10C exhibited a mixture of species similar to E61C (**Figure 4.10A**). Fraction I eluted in the void volume, whereas Fraction II shifted to K_{av} 0.14, indicating stacking of ~4 rings which is different from E61C. The majority of N10C, eluted at K_{av} 0.30, with the molecular weight calculated corresponding to two rings. This is suggestive of the two rings linked via disulfide bridges. Fraction III was collected and subjected to analytical SEC both in reducing and non-reducing conditions. The samples were dialysed into PBS, pH 8.0 with and without reducing agent for 2 h at room temperature prior to loading onto the SEC column.

(i) Non-reducing conditions

The elution profile from **Figure 4.10C** shows that N10C eluted as a single species but with an altered K_{av} of 0.30, indicating a molecular weight of ~ 110 kDa, indicative of two rings associated together in non-reducing conditions.

(ii) Reducing conditions

With addition of a reducing agent (10 mM TCEP), the elution peak shifted to K_{av} 0.4, corresponding to ~ 60 kDa (**Figure 4.11**), and hence a single ring state. Thus, under reducing conditions, the dimers are dissociated.

SDS-PAGE in reducing and non-reducing conditions reveals results consistent with SEC. **Figure 4.10C** shows N10C in higher molecular weight states in the non-reduced condition. The lowest band at ~ 18 kDa corresponds to two monomers associated via disulfide bonds. There is formation of a ~ 50 kDa species, followed by higher order forms above 60 kDa, with a complete absence of the tetrameric species. Whereas, in the reduced state, the bands are typical of Lsmα protein, with species at ~ 9 kDa, ~ 40 kDa and ~ 50 kDa.

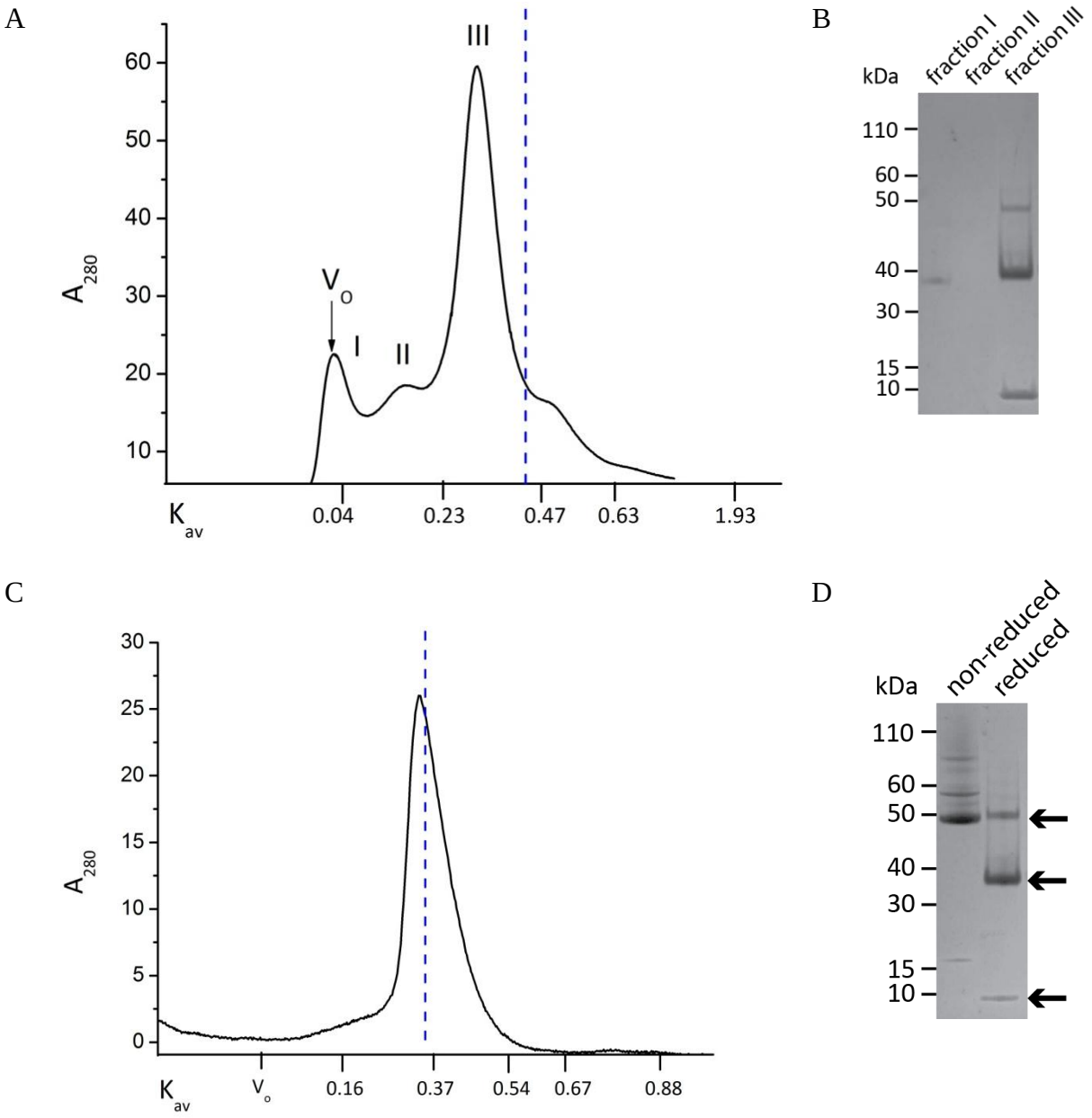


Figure 4.10| *SEC chromatograph of N10C. (A) Preparative grade SEC (Superdex 200 16/600) pre-equilibrated with 10 mM PBS pH 8.0 and run at a flow rate of 1 mL/min. (B) SDS-PAGE of fractions from preparative SEC. (C) Analytical SEC run with the same buffer conditions at a flow rate 0.4 mL/min. The dashed line indicates the fraction with K_{av} corresponding to a single ring. (D) SDS-PAGE of N10C in non-reducing and reducing conditions.*

Hence, the data are consistent with ring formation being followed by disulfide linking of two rings, resulting in helix-to-helix stacking. This could be attributed to the fact that the N-terminal structural elements, including the N10 residues, are more flexible and exposed than the loop elements, thereby available for thiol-thiol associations, in contrast to the data presented for E61C. This results in a locked dimer of rings.

Hence, the helix face is more conducive towards linking but results in a dimer of rings, not capable of further polymerisation and extension to nanotubes. The loop face does not link together due to structural constraints, and therefore has a functionalised face available for further polymerisation reactions with the double Cys mutein, N10C/E61C. The surface onto which the Cys residues are engineered affects the conjugative properties of the toroidal scaffold.

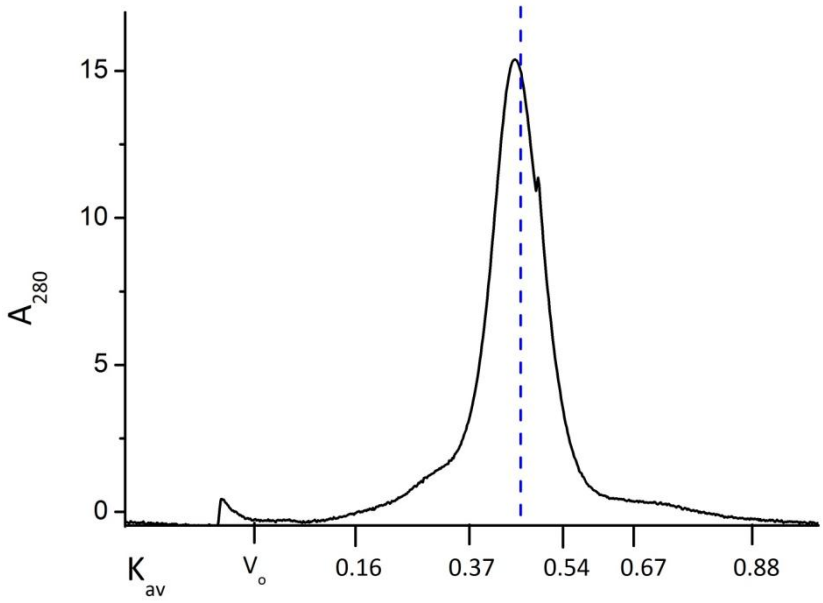


Figure 4.11| *Analytical SEC of N10C in reducing conditions (10 mM TCEP) at a flow rate of 0.4 mL/min. The dashed line indicates the elution volume of a single ring.*

4.6.3 Characterisation of N10C/E61C

The double mutein N10C/E61C was analysed by preparative SEC. Since there are engineered Cys residues on the two faces of the ring scaffold, the solution state of the sample was assessed in reducing and non-reducing conditions. The samples were dialysed into PBS, pH 8.0 with and without reducing agent for 2 h at room temperature prior to loading onto the SEC column.

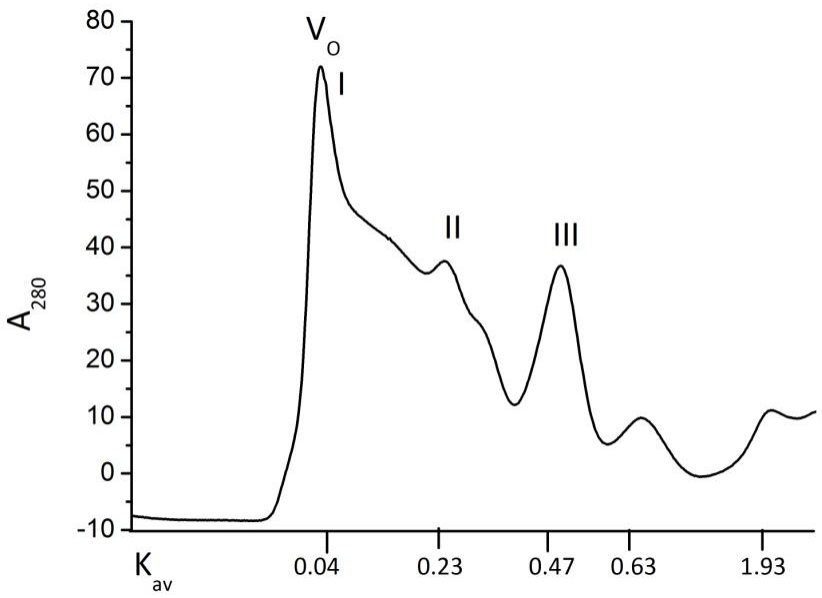
(i) Non-reducing conditions

Preparative SEC revealed a mixture of species in phosphate-based buffer. As shown in **Figure 4.12A**, the majority of the sample eluted in the void volume, indicating a species > 1300 kDa (fraction I). Fraction II elutes at K_{av} 0.20 corresponding to ~ 340 kDa, indicating a probable stacking of ~5 rings. Fraction III corresponds to a K_{av} 0.5. Since the formation of disulfide bridges was hypothesised between the rings, leading to the generation of larger molecular weight species, fraction I was collected and analysed further by analytical SEC and TEM. Since, the higher-order structures were formed during the purification of the sample, the disulfide association of N10C/E61C rings is a spontaneous event most likely being oxidised when exposed to air. Furthermore, to confirm the nature of conjugation, N10C/E61C was also subjected to reducing conditions in preparative SEC.

(ii) Reducing conditions

In reducing conditions, N10C/E61C exhibited a drastic change in the oligomeric distribution as assessed by SEC (**Figure 4.12B**). Contrary to the mixture of species obtained in a non-reducing environment, in the presence of 10 mM TCEP, the equilibrium shifted towards a predominantly single-ring species. Fraction I was obtained in the void volume, which could indicate the non-reduced higher molecular species being >1300 kDa. Fraction II had a K_{av}

A



B

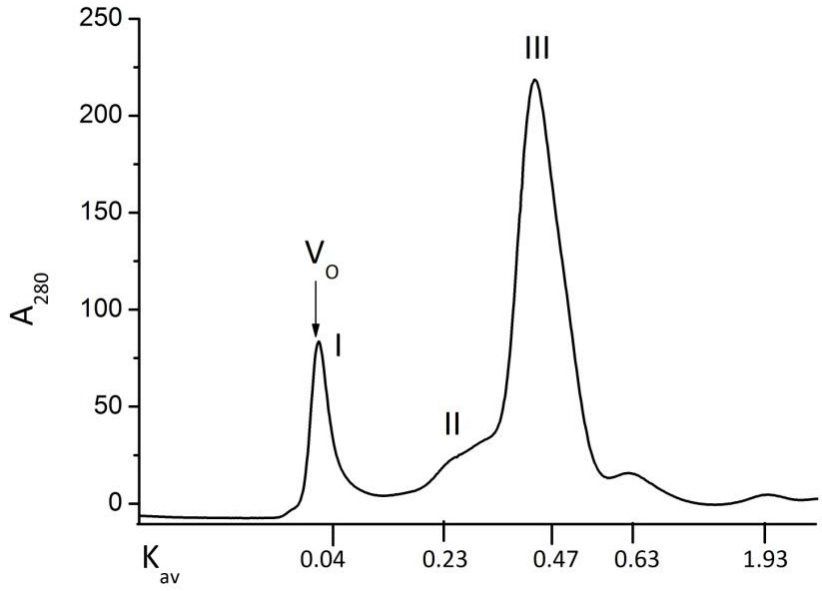


Figure 4.12| *Preparative grade SEC of N10C/E61C. Superdex 200 16/600 column pre-equilibrated with 10 mM PBS pH 8.0 and run at a flow rate of 1 mL/min. (A) Non-reducing conditions and (B) Reducing conditions (10 mM TCEP).*

0.2, indicating stacks of 2-3 rings. The majority of the sample was obtained as Fraction III with K_{av} corresponding to the single ring form. This is consistent with the nature of the linkage between the rings occuring via disulfide bridges. Fraction III was collected and analysed by analytical SEC.

N10C/E61C eluted as a single species on analytical SEC. Inspection of the chromatograph showed the presence of shoulders which may indicate stacking events even in the presence of reducing agent. A large portion of the sample was completely reduced to single rings in the presence of molar excess of reducing agent (**Figure 4.13A**). Consistent with the solution studies, the SDS-PAGE analysis showed higher-order oligomeric forms in non-reduced conditions (**Figure 4.13B**).

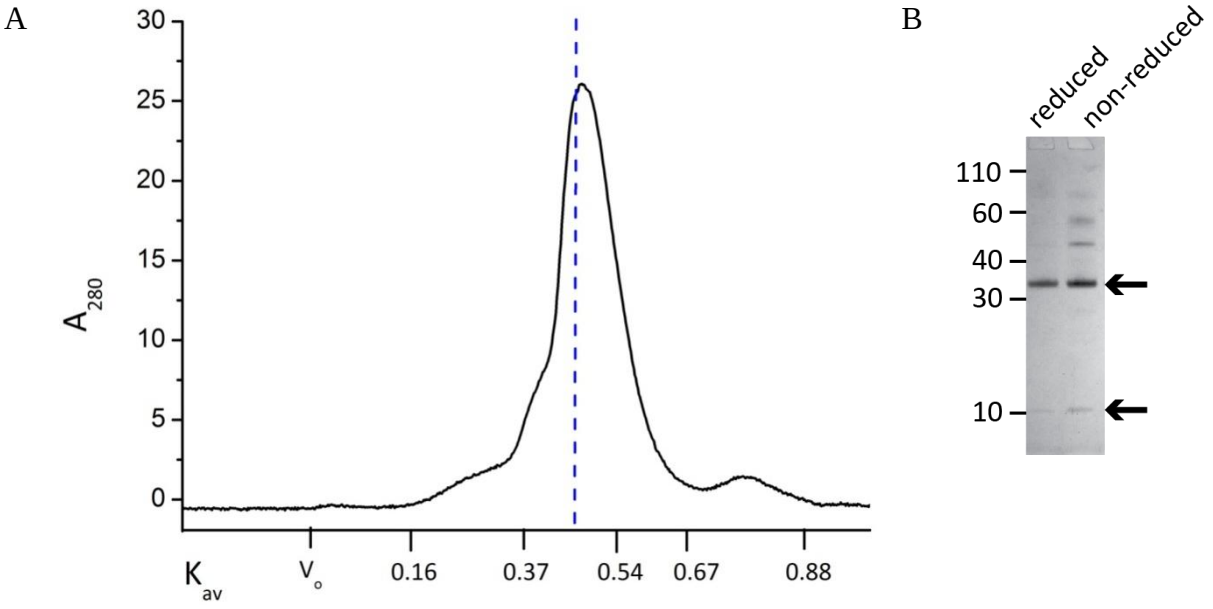


Figure 4.13*| *(A) *Analytical SEC of N10C/E61C in 10 mM PBS pH 8.0, 10 mM TCEP at a flow rate 0.4 mL/min. The dashed line indicates the position of a single Lsmα ring on the analytical column. (B) SDS-PAGE showing reduced and non-reduced sample of N10C/E61C.*

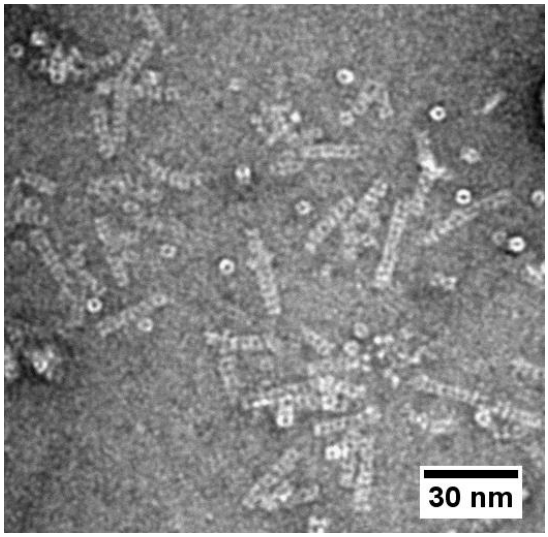
4.7 TEM of protein stacks

N10C/E61C spontaneously assembled into nanostacks as observed by negative stain TEM. As discussed in **Section 4.3**, the ring-to-ring stacking can be axial i.e. they share a common axis, or skewed i.e. the tube axis is staggered and misaligned. **Figure 4.14A** shows the organisation of rings into tubes of length ranging from 10 nm to 50 nm, suggesting about three to sixteen rings (the height of a single ring being 3 nm) associated in axial stacking. From the size distribution histogram, the majority of the tubes were ~20 nm in length (**Figure 4.14B**). Although most of them were linear in geometry, curved tubes were also observed, indicating the presence of skewed stacking. The same sample, in the presence of reducing agent, disassembles into single rings as shown in **Figure 4.14C**. Thus, the gross organisation of Lsmα can be controlled by changing the redox environment.

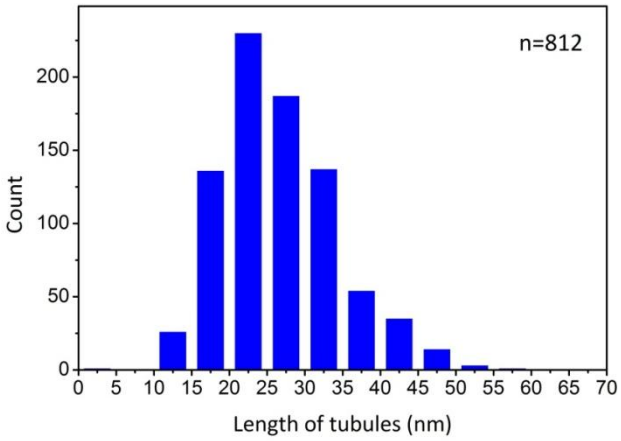
4.8 Polymerisation of nanotubes

As discussed in **Section 4.3**, other ring-shaped proteins have been engineered to assemble as stacks. In order to control the length of N10C/E61C, the two ends of the tube were blocked using the single Cys version of Lsmα. As observed from the solution behaviour of the two single Cys constructs, N10C is most likely constrained into a head-to-head dimer, thereby removing the ability to form thiol-thiol associations with other modules. N10C and E61C have one functionalised face capable of forming disulfide linkages through the exposed sulfhydryl (–SH) groups, with E61C showing the cleanest behaviour in solution (**Section 4.6.1**).

A



B



C

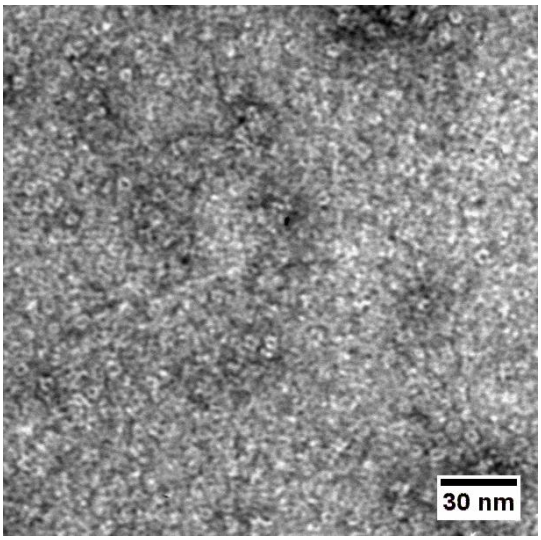
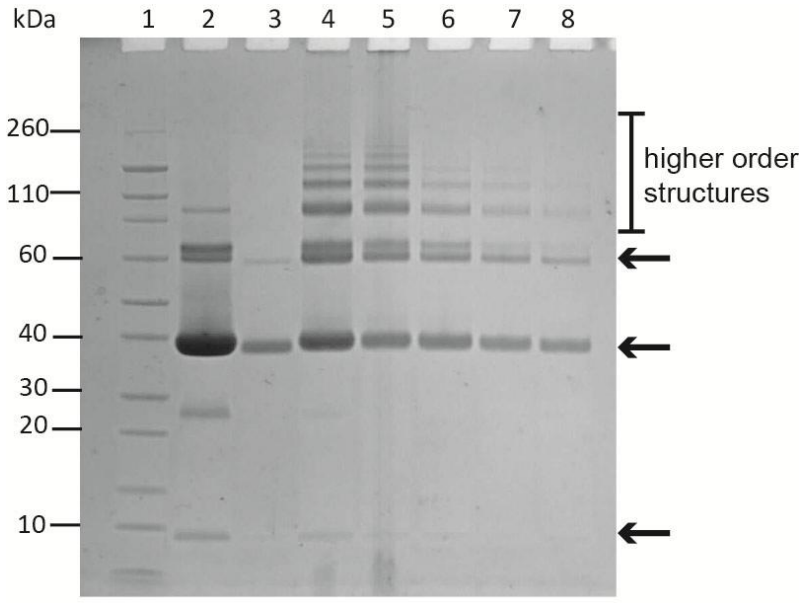


Figure 4.14| *Negative stain transmission microscopy of N10C/E61C. (A) Micrograph showing formation of tubes prepared in a phosphate buffer. (B) Histogram showing the distribution of length of tubes. (C)TEM of N10C/E61C in presence of the reducing agent.*

The samples were prepared by dialysing into PBS pH 8.0 containing 10 mM TCEP for 6 h at room temperature, to ensure that the thiol groups are reduced. Different stoichiometries of N10C/E61C and E61C were mixed together and dialysed into PBS, pH 8.0 buffer to remove the reducing agent alongside providing sufficient interaction time. The samples were analysed by non-reducing SDS-PAGE to assess the formation of higher order structures.

As seen from **Figure 4.15A**, higher molecular weight bands are obtained on the SDS-PAGE when the two constructs are mixed together. It was observed that there is a shift in the molecular weights by an increment of ~ 9 kDa. This suggests that there are disulfide linkages between the monomers of N10C/E61C and E61C which decreases as the ratio decreases. The size distribution of the tubules also changes and the length distribution is tighter when compared with spontaneous tube formation (**Figure 4.16C**). Therefore, a higher concentration of N10C/E61C leads to longer polymerised chains and hence the appearance of more number of discrete bands on the gel.

A



B

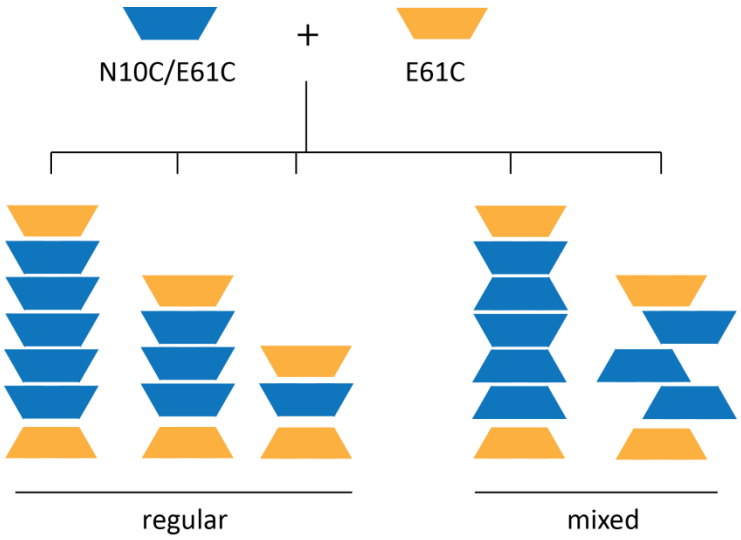
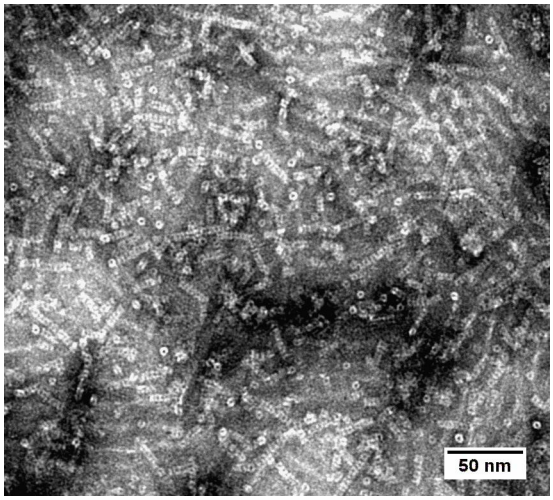
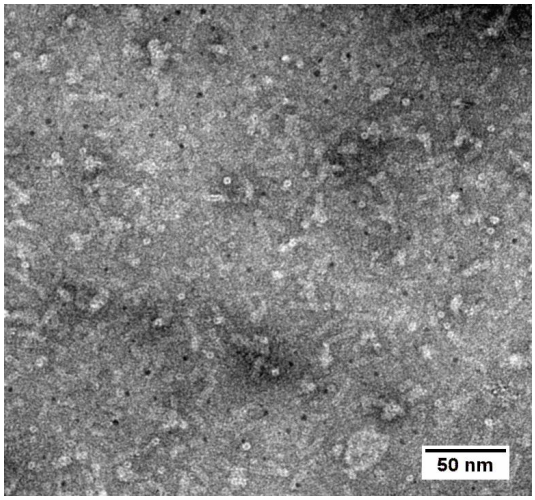


Figure 4.15| *(A) Non-reducing SDS-PAGE showing formation of higher order structures with capping of N10C/E61C with E61C module. 1-marker; 2-N10C/E61C (reduced); 3-E61C (reduced). Lanes 4 to 8 show ratios of N10C/E61C : E61C. 4 -3:1; 5 - 2:1; 6 - 1:1; 7 - 1:2; 8 - 1:3. (B) Schematic showing the possible combinations of stacking obtained by mixing N10C/E61C with E61C.*

A



B



C

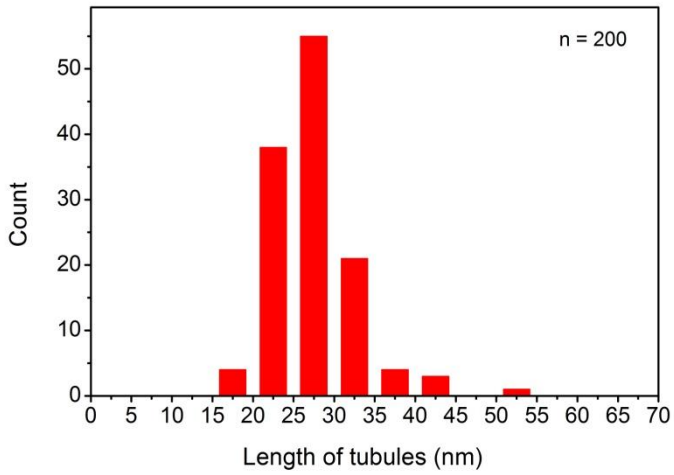


Figure 4.16| *TEM micrographs showing N10C/E61C and E61C in (A) 3:1 and (B) 1:3 ratios corresponding to the SDS-PAGE in Figure 4.13.(C) Size distribution histogram of N10C/E61C mixed with E61C in 3:1 ratio.*

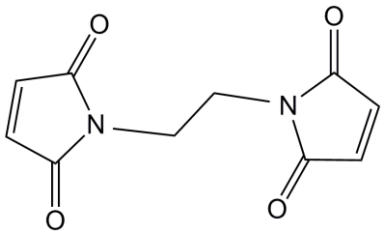
4.8.1 Cross-linking of N10C/E61C

As demonstrated in **Sections 4.6.2** and **4.6.3**, the artificially inserted Cys residues can induce the disulfide linkage and formation of tubes, however, as observed from TEM (**Figure 4.14A**), there are single rings present along with the nanotubes, indicating dissociation of rings from the tubular structure or not being incorporated into the tube. This could be caused by the local surface effects created as the sample was being deposited on the grid surface. To ensure a more rigid and stable structure, other chemical linkages were tested to enhance the formation of higher molecular weight species.

Cross-linking using reactive reagents was tested to control the polymerisation of N10C/E61C. Homo-bifunctional crosslinkers have been used to determine the arrangement of subunits with homo-oligomeric proteins [9], and to study the conformational changes in proteins. Two types of cross-linkers were trialled to observe the assembly of N10C/E61C in their presence.

Bismaleimides are a class of compounds with two maleimide groups connected by the nitrogen atom via a carbon linker (**Figure 4.17A**) and are commonly used as cross-linking reagents in polymer chemistry and to characterise protein structures and interactions [10]. In this study, bis(maleimide)ethane (BMOE) is used as a cross-linker, containing a two carbon linker, to specifically react with the engineered sulfhydryl groups on the surface of

A



B

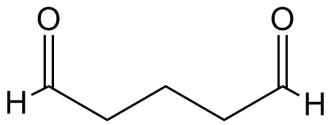


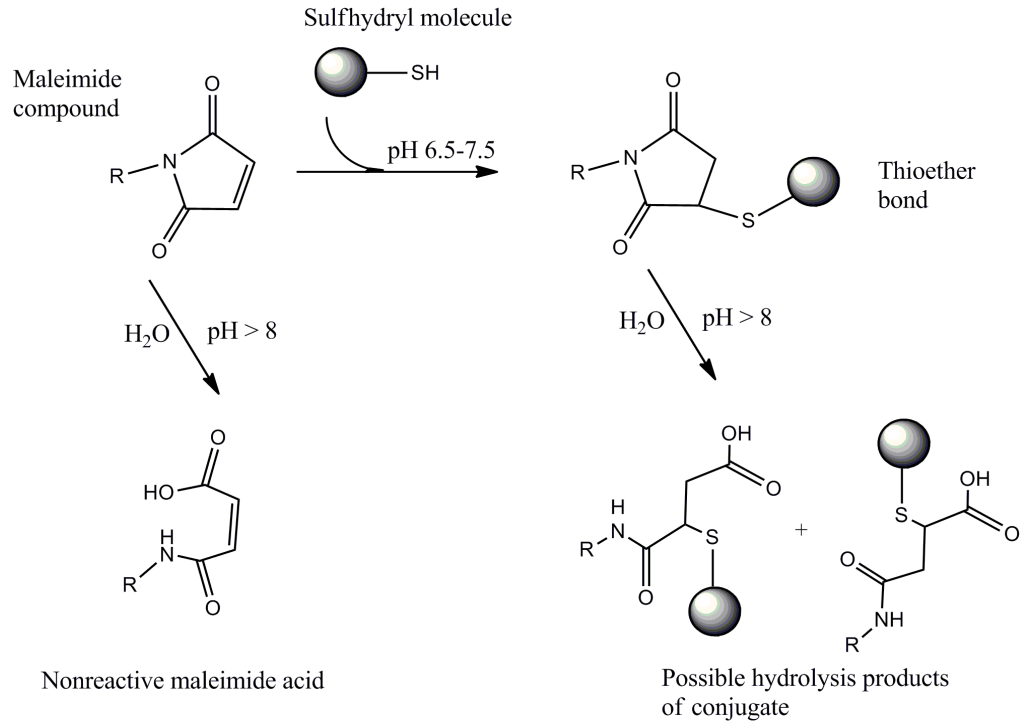
Figure 4.17| *Structures of (A) bis(maleimido)ethane and (B) glutaraldehyde.*

N10C/E61C. This is an alternative method to form stable tubes via the thiol-maleimide chemistry.

Reaction of a sulfhydryl to the maleimide group results in the formation of a stable thioether linkage, which cannot be cleaved by reducing agents or physiological buffer conditions [11]. **Figure 4.18A** shows the thiol-maleimide reaction. The reaction of maleimides is very specific to sulfhydryls at pH 6.5-7.5 and they do not react with Tyr, His or Met residues [12]. The maleimide moiety is temporarily stable in aqueous solutions devoid of reactive sulfhydryl targets, but hydrolysis to a nonreactive maleimide acid can occur during storage at pH > 8. Hydrolysis of the ring structure can also occur following conjugation, resulting in an open-ring linkage [13] (**Figure 4.18A**).

Glutaraldehyde (GA) is a bi-functional cross-linker containing two aldehyde groups on a five carbon chain (**Figure 4.17B**). GA is most commonly used in enzyme technology, leather tanning, and as a fixative for microscopy [14]. In this study, GA was used to link the N10C/E61C rings into an irreversible tube form and further serve as a comparison to the maleimide cross-linking agent.

A



B

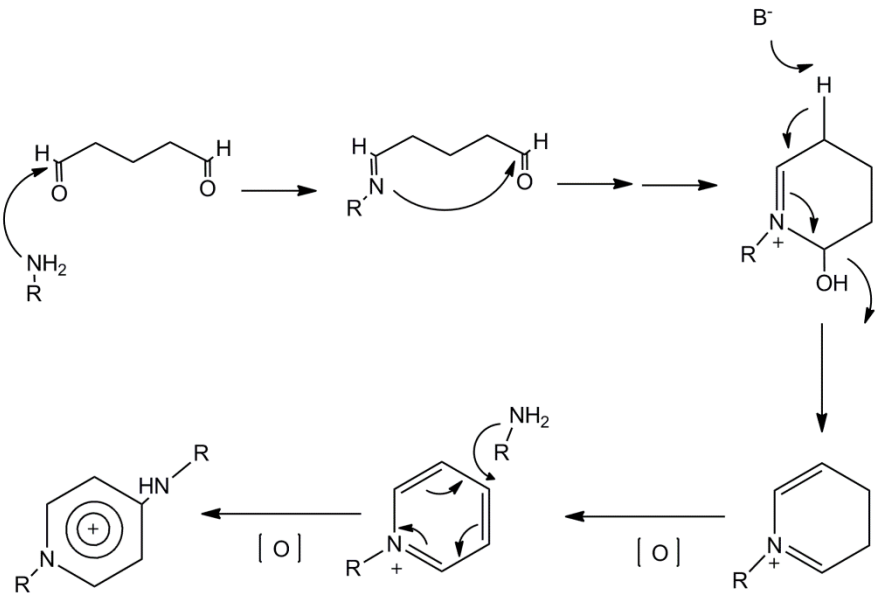


Figure 4.18| *(A) Reaction of maleimide-activated compounds to sulfhydryls. Adapted from Heredia et al.* [16]*. (B) Proposed mechanism of pyridium cross-links. Adapted from Meade et al.* [17]*.*

GA is known to react with several functional groups of proteins, including, amine, thiol, phenol and imidazole moieties [15]. The most reactive species towards GA are the ε-amino groups, followed by the α-amino groups, which correspond to the primary amino groups present in the Lys residues, and the N-terminal amino group present in proteins, respectively [14]. Since, N10C/E61C also contains engineered thiol groups, the reaction of GA via thiol groups also has to be taken into account. The mechanism by which GA reacts with the amino groups of proteins is still under debate. One possible mechanism is via the formation of a pyridinium cross-link, as shown in **Figure 4.18B**.

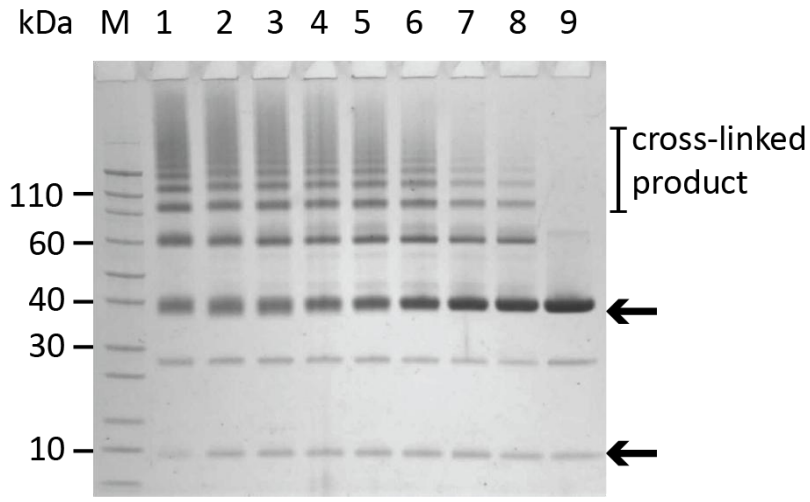
(i) Sulfhydryl cross-linking

BMOE was used to link the N10C/E61C ring modules through their exposed Cys residues. Since N10C/E61C instantaneously forms the elongated stacks, it was reduced by dialysing in a phosphate buffer, pH 7.5, containing 10 mM TCEP for 2 h prior. TCEP was chosen as it does not interfere with the maleimide chemistry. The reduced N10C/E61C (5 mg/mL) was filtered using a 0.2 µm syringe filter before adding BMOE. Care was taken to make fresh solutions of BMOE to avoid hydrolysis. The mixture was incubated for 2 h and the reaction stopped by adding 50 mM DTT to the reaction mix. 15 µL aliquots were taken and loaded on the SDS-PAGE. **Figure 4.19A** shows the SDS-PAGE of N10C/E61C incubated with BMOE, at increasing concentration. Reduced N10C/E61C (**Figure 4.19A**, lane 1) exhibits three bands whereas on controlled addition of the cross-linker, there were higher molecular weight species present. The samples investigated with SDS-PAGE showed that N10C/E61C incubated with a very low concentration of BMOE (> 0.062 mM, lane 7 and lane 8) exhibited formation of discrete bands indicating cross-linking of the subunits. Whereas, higher concentration (>0.125 mM) shows smearing on the gel which suggests non-specific cross-linking.

(ii) Glutaraldehyde cross-linking

GA is a homo-bifunctional reagent that reacts with the primary amine groups i.e. ε-amino groups of Lys residues, capable of forming inter- and intra-subunit covalent bonds. N10C/E61C was cross-linked with GA to compare the pattern obtained to that obtained with the use of BMOE. The sample was prepared as described above with omission of the stopping of the reaction with DTT. The SDS-PAGE (**Figure 4.19B**) showed smears for all concentrations of GA tested, indicating that there were non-specific cross-linking events. GA has the capacity to create large and often poorly defined protein aggregates and is difficult to control the intra- and intermolecular cross-linking [18]. Maleimide chemistry proves to be more effective in cross-linking N10C/E61C than using glutaraldehyde as observed from the extent and nature of bands formed by SDS-PAGE. The degree of linkage is more controlled in the case of N10C/E61C being incubated with E61C.

A



B

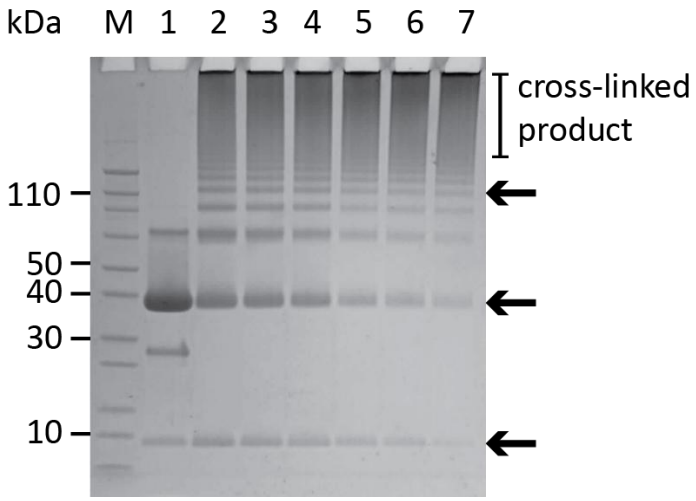


Figure 4.19| *Non-reducing SDS-PAGE. (A) N10C/E61C with BMOE for 30 min at room temperature; M-ladder; 1-20 mM; 2-10 mM; 3-1 mM; 4-0.5 mM; 5-0.25 mM; 6-0.125 mM; 7-0.062 mM; 8-0.031 mM; 9-reduced N10C/E61C. (B) N10C/E61C with GA incubated for 5 min at room temperature; M-ladder; 1- reduced N10C/E61C; 2-0.025 mM; 3-0.5 mM; 4-1 mM; 5-2 mM; 6-2.5 mM; 7-5 mM. Arrows indicate the monomeric form (~ 9 kDa), tetrameric form (~ 35 kDa) and the higher order forms.*

4.9 Summary

Lsmα consists of two chemically distinct faces i.e. the helix face and the loop face. Two approaches were tested to manipulate the quaternary structure of the rings and let them self-assemble into stacks of rings. The muteins that were engineered to include Cys residues polymerised into tubes by disulfide conjugation. In the case of N10C/E61C, the associations can take place through either of the two faces i.e. helix-to-loop, loop-to-loop and helix-to-helix. As observed from the size distribution histogram, the average length of the spontaneously formed tubes is ~20 nm, suggesting that polymerisation includes up to ~7 rings conjugated in an extended tubular arrangement. From the micrographs, the majority of the stacks are linear, suggesting axial stacking of the ring scaffold. A minority of the stacks are skewed, suggesting random association of rings through thiol-thiol interactions.

Incubation of N10C/E61C with E61C provided some insight into the orientation of stacking of the modified Lsmα rings. Since only the loop face is available for conjugation for E61C and it was found that N10C/E61C conjugates with E61C (**Figure 4.15**), these observations lead to the conclusion that the majority of the linkages were in a head-to-tail manner. E61C presumably acts as a chain terminating module at one end. In order to comment on the exact nature of the Lsmα tube structure, high resolution structural data is required.

The results are consistent with the tubes being formed by direct disulfide links between the engineered Cys side chains. From the solution studies performed on N10C/E61C, the tubes completely dissociate on addition of a reducing agent, thereby confirming that the associations are mediated through disulfide bridge formation. Considering the fact that there are seven Cys residues on both faces of the Lsmα ring, a skewed non-axial linking of the rings cannot be eliminated. As seen from the TEM micrograph (**Figure 4.14A**), some of the tubes are non-axially stacked. This may arise due to incomplete alignment of the rings on top

of each other but resulting in disulfide conjugation nonetheless. The schematic in **Figure 4.4** sums up the various arrangements available to the Cys modified Lsmα rings.

By generating a synthetic interface between the ring modules, high aspect-ratio tubes were created, otherwise not observed in the native state of Lsmα. The nature of these associations is shown to be through disulfide bridge formation. This feature provides an additional control over the assembly and disassembly of the proteinaceous tubes.

The various interactions present and observed in Lsm proteins which result in the formation of higher oligomeric structures of stacks and fibres include salt bridge formation [19], His interactions [19; 20], disulfide linking [2], and electrostatic interactions [20; 21]. These interactions result in either stacking of the ring modules to form 14-mer species or result in assembly of fibril ultrastructures. As mentioned in **Section 4.2**, higher order oligomerisation properties have been observed for other archaeal Lsm proteins. *Mt* SmAP1 (PDB ID1LOJ) also exhibits oligomerisation properties assembling into fibres stretching to several nanometres in length. These were obtained by treating purified protein in very low salt conditions which resulted in head-to-tail stacking of the heptamers in low ionic strength buffer [2]. The same conditions were used to induce fibrogenesis in the case of Lsmα (**Section 4.2**). It is interesting to point out the difference in the preparation of the proteins. *Mt* SmAP1 was constructed as a C-terminally His_6-tagged protein, which was followed by trypsin cleaving to remove the tag. The N-terminus of the protein did not have any extra residues post-cleavage. On the other hand, the strategy used in this study involved using GST as the affinity tag, which results in the incorporation of two extra residues, Gly and Ser, at the N-terminus after protease cleavage. The fact that protein preparations had the extra residues could contribute to it failing to produce fibres by head-to-tail stacking when treated in low ionic strength buffers.

Although various proteins exist that naturally assemble into filaments and tubular structures, these are not optimal for nanotechnological applications [22; 23]. The features required for downstream applications include ease of modification, *in vitro* control over assembly, and easily accessible structural features. Doughnut-shaped proteins allow *in vitro* assembly post-expression of the self-assembling ring module. As discussed in **Section 4.3**, other ring-shaped proteins have been engineered to generate tubular structures, with pore dimension ranging from 13 nm to 3 nm [24; 25]. Lsmα tubes offers a pore size of ~1.5 nm which is a new dimension reached compared to other the protein tubes generated by genetic engineering methods.

4.10 References

1. Son, S. J., Bai, X., Nan, A., Ghandehari, H. & Lee, S. B. (2006). Template synthesis of multifunctional nanotubes for controlled release. *Journal of Controlled Release* **114**, 143-152.
2. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M. & Eisenberg, D. (2003). The oligomerization and ligand binding properties of Sm like archaeal proteins (SmAPs). *Protein Science* **12**, 832-847.
3. Arluison, V., Mura, C., Guzman, M. R., Liquier, J., Pellegrini, O., Gingery, M., Regnier, P. & Marco, S. (2006). Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. *Journal of Molecular Biology* **356**, 86-96.
4. Molls, J. (2011). Structural characterisation of *yeast* Lsm protein complexes (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
5. DeLano, W. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*.
6. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10037-10041.
7. Miranda, F. F., Iwasaki, K., Akashi, S., Sumitomo, K., Kobayashi, M., Yamashita, I., Tame, J. R. H. & Heddle, J. G. (2009). A self-assembled protein nanotube with high aspect ratio. *Small* **5**, 2077-2084.
8. Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y. & Mougous, J. D. (2008). *In vitro* self-assembly of tailorable nanotubes from a simple protein building block. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3733-3738.
9. Mischke, R., Kleemann, R., Brunner, H. & Bernhagen, J. (1998). Cross-linking and mutational analysis of the oligomerization state of the cytokine macrophage migration inhibitory factor (MIF). *FEBS Letters* **427**, 85-90.
10. Brinkley, M. (1992). A brief survey of methods for preparing protein conjugates with dyes, haptens and crosslinking reagents. *Bioconjugate Chemistry* **3**, 2-13.
11. Wong, S. S. (1991). Chemistry of protein conjugation and cross-linking, *CRC press*.

12. Smyth, D. G., Blumenfeld, O. & Konigsberg, W. (1964). Reactions of N-ethylmaleimide with peptides and amino acids. *Biochemical Journal* **91**, 589-595.
13. Brewer, C. F. & Riehm, J. P. (1967). Evidence for possible nonspecific reactions between *N*-ethylmaleimide and proteins. *Analytical Biochemistry* **18**, 248-255.
14. Migneault, I., Dartiguenave, C., Bertrand, M. J. & Waldron, K. C. (2004). Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme cross-linking. *Biotechniques* **37**, 790-806.
15. Walt, D. R. & Agayn, V. I. (1994). The chemistry of enzyme and protein immobilization with glutaraldehyde. *TrAC Trends in Analytical Chemistry* **13**, 425-430.
16. Heredia, K. L. & Maynard, H. D. (2006). Synthesis of protein–polymer conjugates. *Organic & Biomolecular Chemistry* **5**, 45-53.
17. Meade, S. J., Miller, A. G. & Gerrard, J. A. (2003). The role of dicarbonyl compounds in non-enzymatic crosslinking: a structure–activity study. *Bioorganic & Medicinal Chemistry* **11**, 853-862.
18. Hermanson, G. T. (2013). Bioconjugate techniques, *Academic Press*.
19. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R., Curmi, P. M. G. & Mabbutt, B. C. (2008). Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: implications for Lsm ring organisation and recruitment. *Journal of Molecular Biology* **377**, 1357-1371.
20. Mura, C., Phillips, M., Kozhukhovsky, A. & Eisenberg, D. (2003). Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4539-4544.
21. Mura, C., Randolph, P. S., Patterson, J. & Cozen, A. E. (2013). Archaeal and eukaryotic homologs of Hfq. *RNA Biology* **10**, 636-651.
22. Hou, S., Wang, J. & Martin, C. R. (2005). Template-synthesized protein nanotubes. *Nano Letters* **5**, 231-234.
23. Nuraje, N., Banerjee, I. A., MacCuspie, R. I., Yu, L. & Matsui, H. (2004). Biological bottom-up assembly of antibody nanotubes on patterned antigen arrays. *Journal of the American Chemical Society* **126**, 8088-8089.
24. Wang, W. X., Dgany, O., Wolf, S. G., Levy, I., Algom, R., Pouny, Y., Wolf, A., Marton, I., Altman, A. & Shoseyov, O. (2006). Aspen SP1, an exceptional thermal,

protease and detergent resistant self assembled nano particle. *Biotechnology and Bioengineering* **95**, 161-168.

25. Braig, K., Adams, P. & Brünger, A. (1995). Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nature Structural Biology* **2**, 1083-1094.

5 Chapter five

Preparation and biophysical characterisation of complexes of His_6-tagged Lsmα: evidence for higher order organisation

5.1 Introduction

It is useful to trigger a change in the quaternary structure with metals or to use a biological structure to template metal assembly. Previous studies on using biomolecules for metallisation have yielded components for electronic circuits [1; 2], templates to create NPs [3; 4] and nanowires [5]. Much more recently, another doughnut-shaped protein system, peroxiredoxin, has been used to assemble Au NPs into one dimensional space [6].

As a purification strategy, a poly-His tag was employed as an affinity tag. This approach served a dual advantage. Along with purifying the protein, the tag conferred a multivalent display of functional groups for each of the Lsmα monomers. The ready access to imidazole moieties with their affinity for metal ions, was used to generate higher order architectures for the Lsmα system. This chapter describes the effects of attachment of a His_6-tag to the termini of Lsmα polypeptides and their resulting assembly into higher molecular weight species.

5.2 Recombinant design and production of His_6-tagged Lsmα

The β-propeller of the Lsmα assembly forms a robust core, leaving both the N- and C-termini of each monomer exposed and available for modification (as discussed in **Section 1.1**). A consideration of the Lsmα heptamer structure in **Figure 5.1B**, shows the N-terminal residues as seven flexible segments preceding the α-helices circling one face of the ring. While the Lsmα C-terminal segments are not specifically defined in the crystal structures, it is likely that these extend out from the equatorial locations around the ring, i.e. remote from the helical face (**Figure 5.1B**). Therefore, the ring scaffold provides two positions for the placement of His-rich segments in each of the seven monomeric units of Lsmα heptamer (**Figure 5.1A**). In order to assess the properties conferred to the Lsmα scaffold by addition of the His_6-tag at each position, constructs were generated by cloning into plasmids that incorporated His_6-tags at the N-terminus and C-terminus of the protein.

Two plasmids designed to express the fusion constructs, $[H_6Lsm\alpha]_7$ and $[Lsm\alpha H_6]_7$ were generated by cloning the Lsmα gene into plasmid pET24a, ordered from Epoch Biolabs (**Section 2.2.3**). These were used to transform the BL21(DE3) strain of *E. coli* (**Section 2.4.1**). Small-scale production was used to assess the relative levels of protein expression across several growth protocols with variation in media, expression strain and induction protocols (**Section 2.4.2**). The best yields were obtained by expressing the protein at an OD_{600} of 0.6 by IPTG (0.2 mM) and grown further for 6 h at 26 °C. SDS-PAGE analysis was used to verify the optimum growth conditions for the two fusion proteins (**Figure 5.2**). MRE5 cells did not show any expression of the His_6-tagged Lsmα proteins under the conditions tested in the small scale expression trials (**Section 2.4.2**), and were not pursued as an expression strain. **Figure 5.2C** shows the total protein content in the MRE5 cells after protein

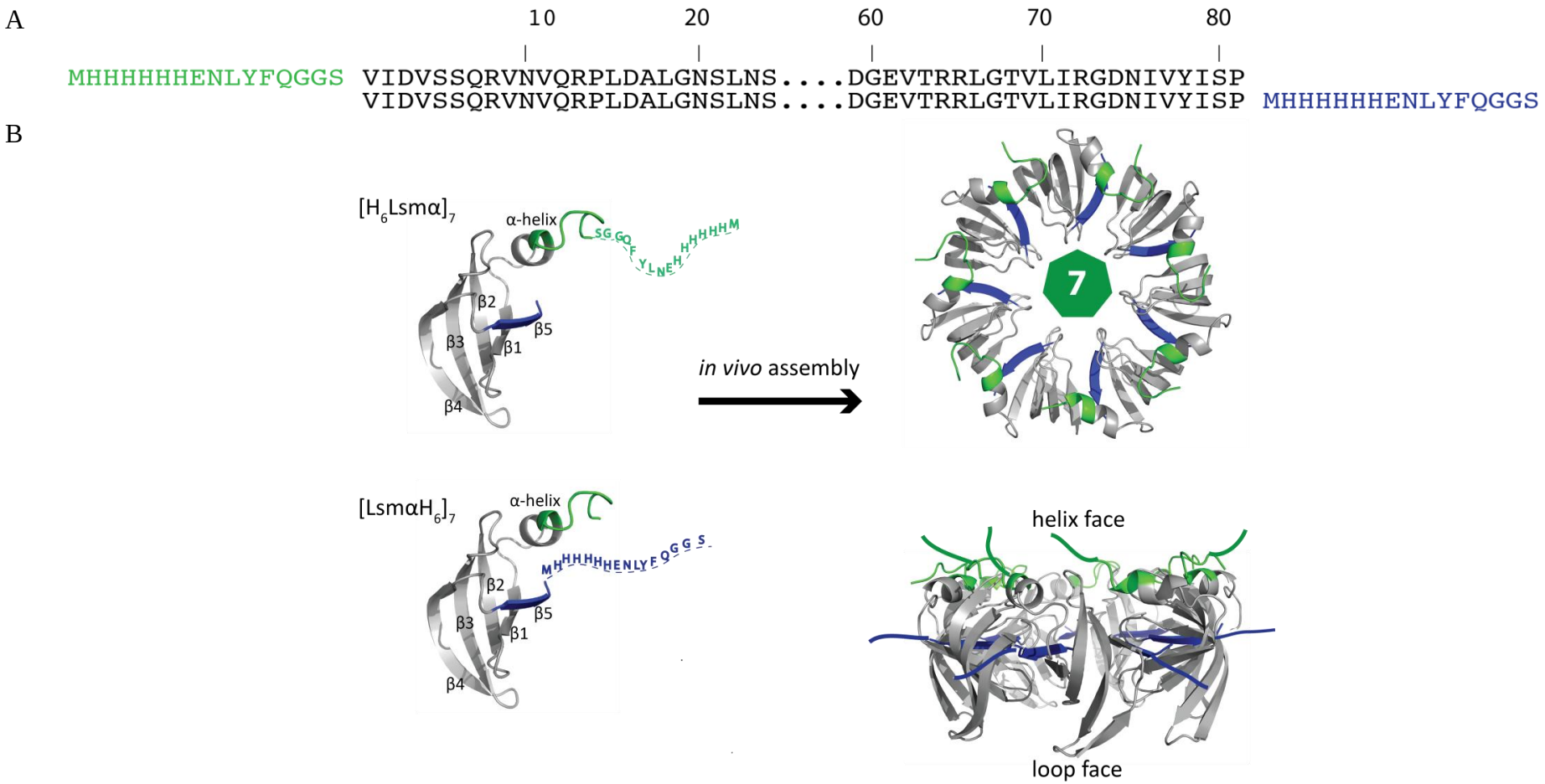


Figure 5.1| *Sequence and ribbon diagram of Lsmα (PDB ID 1I81) depicting the placement of His_6 segments. (A) Lsmα sequence showing His_6 segment at the N- and C-termini. Dots between the sequences represent omitted residues. (B) Ribbon representation of the Lsmα fold with β-strands labelled. The overall placement of the His_6 segment is also shown after self-assembly into a ring scaffold. Images were prepared in Pymol* [7].

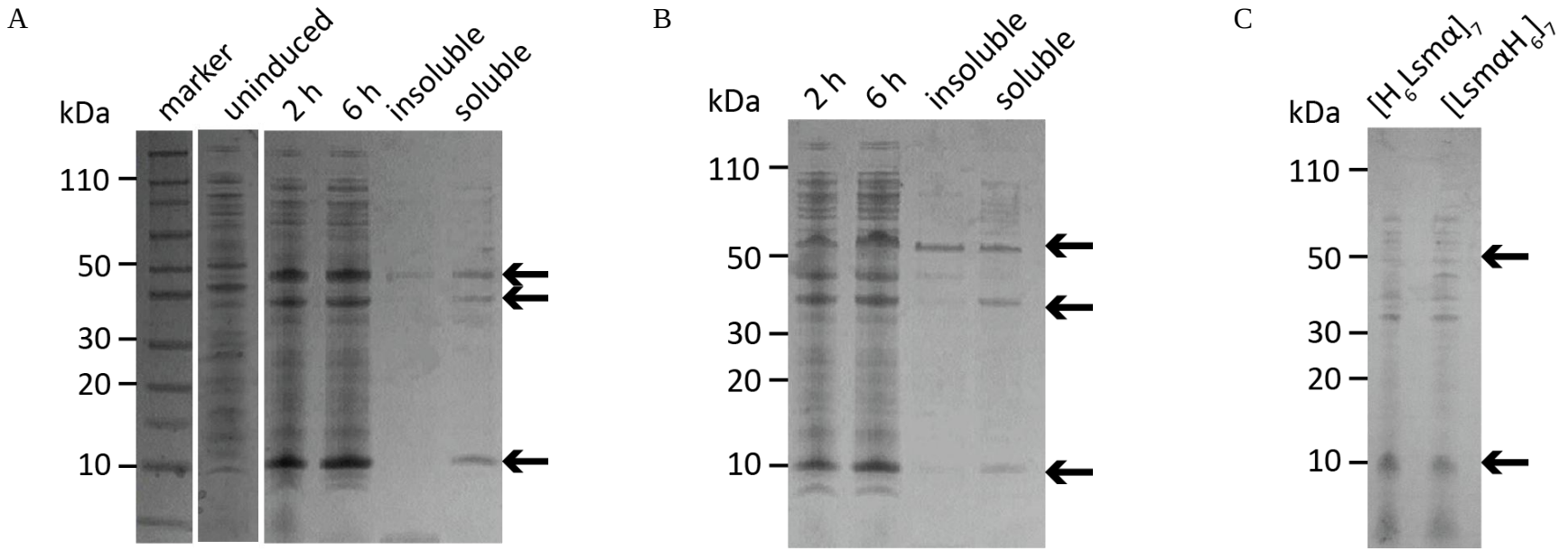


Figure 5.2| *SDS-PAGE showing protein composition for cells containing Lsmα constructs. Protein content analysed following growth of hosts for (A) $[H_6Lsm\alpha]_7$ in BL21(DE3) cells, (B) $[Lsm\alpha H_6]_7$ in BL21(DE3) cells and (C) expression in MRE5 strain showing the absence of His_6-tagged Lsmα with IPTG induction in the crude, grown at 26 °C. The gel was run at 200 V for 35 mins at room temperature before staining with Simply blue stain.*

expression was induced by IPTG. There is no noticeable expression of the protein. Due to reasonable levels of expression in the soluble fraction in the BL21(DE3) strain, both the constructs, $[H_6Lsm\alpha]_7$ and $[Lsm\alpha H_6]_7$, were selected for further large-scale preparation.

As was found for the untagged protein (**Section 3.7**), $[H_6Lsm\alpha]_7$ ran as multiple bands on a denaturing SDS-PAGE gel, at 10, 40 and 50 kDa (**Figure 5.2A**), consistent with the wildtype Lsmα and other muteins isolated. A separate band for $[Lsm\alpha H_6]_7$ (**Figure 5.2B**), at ~60 kDa, is present both in the insoluble and soluble fractions, and is suggestive of a different stable oligomer, corresponding to a hexamer, formed during the denaturation step compared to $[H_6Lsm\alpha]_7$. This was attributed to the placement of the His_6-tag at the C-terminal. A higher yield was observed for $[H_6Lsm\alpha]_7$ recovered in the soluble cell fraction, compared to $[Lsm\alpha H_6]_7$ as observed from the SDS-PAGE analysis.

5.3 Isolation of tagged recombinant products

The matrix most commonly used for immobilised metal affinity chromatography (IMAC) of His-tagged protein incorporates nickel (Ni^{2+}) [8; 9]. However, other divalent transition metal ions have also been tested successfully for IMAC procedures, taking advantage of the affinity of metal ions to the aromatic imidazole ring of the His residues [10]. The metals most widely used are cobalt (Co^{2+}), zinc (Zn^{2+}), and copper (Cu^{2+}) [8; 11]. Following the application of a protein mixture to the adsorption bed at pH 8.0, high concentrations of imidazole facilitated the desorption of a His-rich target from the matrix, after washing off unbound proteins.

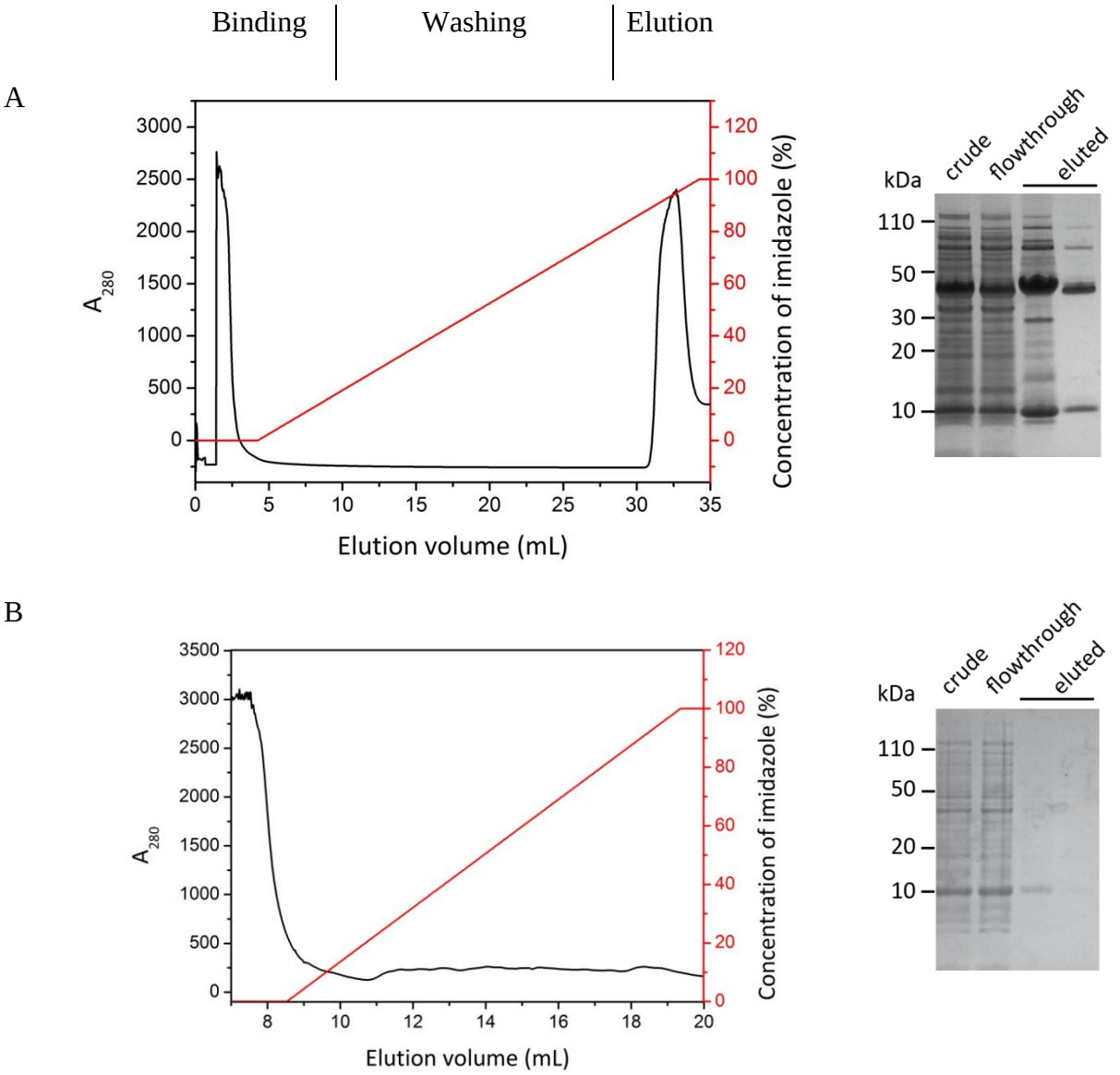


Figure 5.3| *Elution profile for IMAC purification of $[H_6Lsm\alpha]_7$. (A) $[H_6Lsm\alpha]_7$ was successfully purified using a chelating matrix pre-charged with $CuSO_4$. Similar results were obtained for $[Lsm\alpha H_6]_7$. (B) HiTrap Chelating (G E Healthcare) columns were pre-charged with different metal ions (**Section 5.3**) and proteins eluted at a flow rate of 1 mL/min. The trace indicates matrix charged with Ni^{2+}. Traces are A_{280} (black) and conductance (red). Binding, washing and elution steps are indicated at the top of the IMAC trace. The SDS-PAGE gel shows molecular weight standards with crude, flowthrough and eluted fractions.*

Samples of $[H_6Lsm\alpha]_7$ and $[Lsm\alpha H_6]_7$ were successfully recovered by selectively using a Cu-based matrix. **Figure 5.3A**, shows the elution profile of $[H_6Lsm\alpha]_7$ at 450 mM of imidazole, yielding ~20 mg/mL of protein. Similar results were obtained for $[Lsm\alpha H_6]_7$. Purification of the constructs failed when metals such as $NiCl_2$, $ZnCl_2$ and $CoSO_4$ were trialled (**Figure 5.3B**). Once purified, the samples were subjected to preparative scale SEC in order to assess the species distribution in the samples.

5.3.1 $[H_6Lsm\alpha]_7$ is selective for Cu^{2+}

To understand the reason for selective affinity of $[H_6Lsm\alpha]_7$ and $[Lsm\alpha H_6]_7$ to the Cu^{2+}-matrix, it is imperative to know the chemistry of the metal-ligand binding to the matrix used and its interaction with the accessible His residues from the His_6-tag. Loading different metal ions onto a resin, results in variable affinity and specificity for His_6-tagged proteins. Generally, cobalt exhibits the highest binding specificity of the commonly used IMAC metal ions. Copper lies at the other end of the spectrum and has a high affinity leading to higher yields but nonspecific binding [10]. Nickel lies in the middle of the spectrum and provides specific binding and reasonable yields. Although IMAC procedures have been widely established, self-assembling proteins can pose a unique challenge due to the number of polyHis-tags on one assembled protein structure [12] and their interaction with the cation-charged matrix.

The coordination potential of a metal ion is largely dependent on the localisation and topography of the His residues exposed on the protein surface [8]. Moreover, the spatial distribution of the His residues plays a critical role on the retention capacity of a metal-ion matrix to adsorb the protein of interest [13]. The partial burial of His residues due to the quaternary structure of Lsmα and possible oligomerisation in the matrix itself, may limit the strength of coordination. Under the given conditions of pH and salt, the His_6-tagged Lsmα

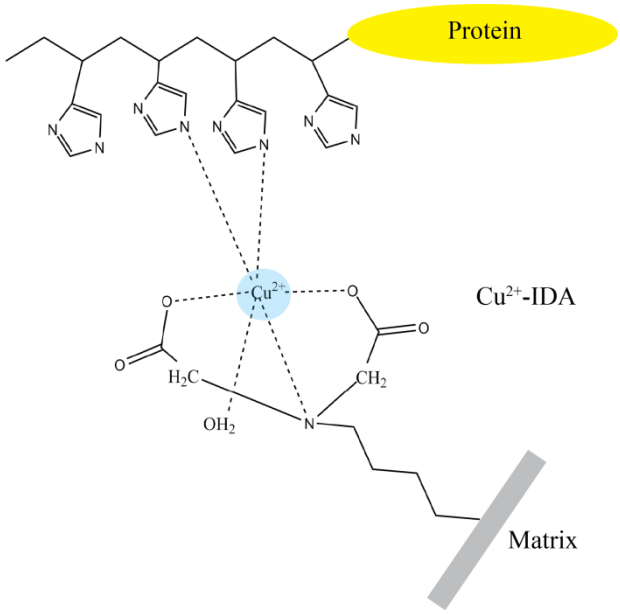
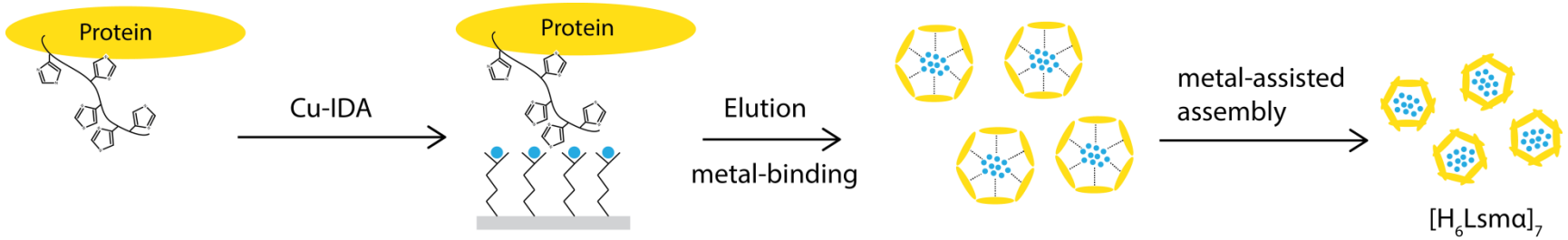


Figure 5.4| *Schematic diagram of chromatographic media used for affinity purification of recombinant proteins. Adapted from Block el al [14]. Cu^{2+}-IDA is iminodiacetic acid forming a complex with copper ions.*

constructs showed optimum binding to the Cu^{2+}-matrix. Failure of binding and retention in the case of Ni^{2+}, Co^{2+}, and Zn^{2+} can be attributed to the formation of a relatively weak ternary complex formed between the chelate, metal ion and the protein. **Figure 5.4** shows the complex formed by the His_6-tagged protein, the ligand and the matrix.

Such behaviour also provides insights into the accessibility of the His_6-tags. Failure to capture this recombinant protein by other metal ions suggests that the His_6-tags are not entirely exposed to the surface and hence not accessible for binding to the cation. This observation is also indicative of possible metal-assisted oligomerisation of the quarternary structure of the His_6-tagged Lsmα proteins during the absorption step of the procedure (**Figure 5.5**). The differential interactions of M^{2+}-IDA with the His_6-tagged Lsmα constructs in the IMAC procedure demonstrates the metal-binding capacity of this ring scaffold and provides tantalizing insights for further work in the field of metal-assisted assembly of the Lsmα complex and other protein complexes [15; 16].

A



B

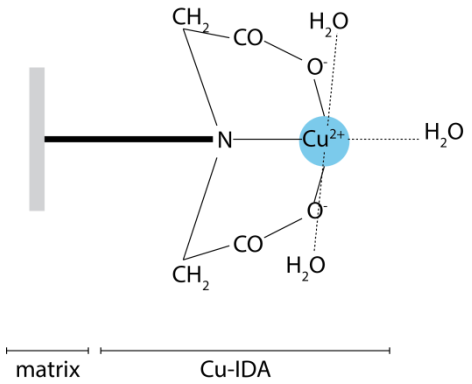


Figure 5.5| *Schematic of putative mechanism of formation of metal-bound [H_6Lsmα]$_7$. (A) Possible assembly steps and interaction with copper ions during affinity chromatography. (B) Intermolecular metal-to-ligand coordination: metal complexing material is a polymer support derivatised with Cu^{2+}-iminodiacetate (G E Healthcare). Given the multiplicity of the His_6-tags oriented towards one surface of the recombinant Lsmα, a microenvironment was created at the attachment point of the cation and protein, leading to the formation of a stronger coordination bond and hence a strong chelate complex.*

5.4 Solution properties of His_6-tagged Lsmα

SEC was used to assess the distribution of oligomeric species for the His_6-tagged constructs using methods described in the previous chapter. Lsmα prepared as a GST-fusion protein was characterised using a Superdex 75 matrix which was consistent with heptameric assembly [17], in agreement with SEC-SLS results in **Section 3.9.1**. A similar strategy was employed to determine the octameric organisation of Lsm3 [18]. Therefore, after affinity chromatography, both $[H_6Lsm\alpha]_7$ and $[Lsm\alpha H_6]_7$ were subjected to preparative and analytical grade SEC.

Preparative SEC for $[H_6Lsm\alpha]_7$ revealed a bimodal distribution of species (**Figure 5.6A**). Peak 1 eluted at the void volume (V_o), indicating a species of >1300 kDa, possibly an aggregate formed during the affinity purification. Peak 2 eluted within the sizing range of the column and corresponded to a size of 400-500 kDa. Fraction II was collected and re-chromatographed on an analytical SEC column. $[H_6Lsm\alpha]_7$ eluted as a single symmetrical peak spanning K_{av} 0.2-0.3 i.e. corresponding to a native mass of ~440 kDa. The estimated molecular weight was consistent with a species containing six heptameric rings which have a predicted mass of 460 kDa. The $[H_6Lsm\alpha]_7$ sample was stable for several weeks stored at 4 °C as assessed by SEC. By SDS-PAGE, a pure preparation of $[H_6Lsm\alpha]_7$ ran as three bands at 10, 50, 80 kDa (**Figure 5.6B**). As seen from the gel, the relatively minor species ran at an expected size of 11 kDa. The other two pronounced bands running at ~50 kDa and ~80 kDa, can be attributed to formation of stable oligomers during the denaturation step. This triplet of bands is diagnostic for His_6-tagged Lsmα assembly and was also noted above during purification (**Section 5.3**). Thus, $[H_6Lsm\alpha]_7$ exhibited similar characteristics consistent with the robust and stable properties of the Lsmα heptameric ring scaffold formed, at the same time, exhibiting higher order species formation, not previously observed for Lsmα.

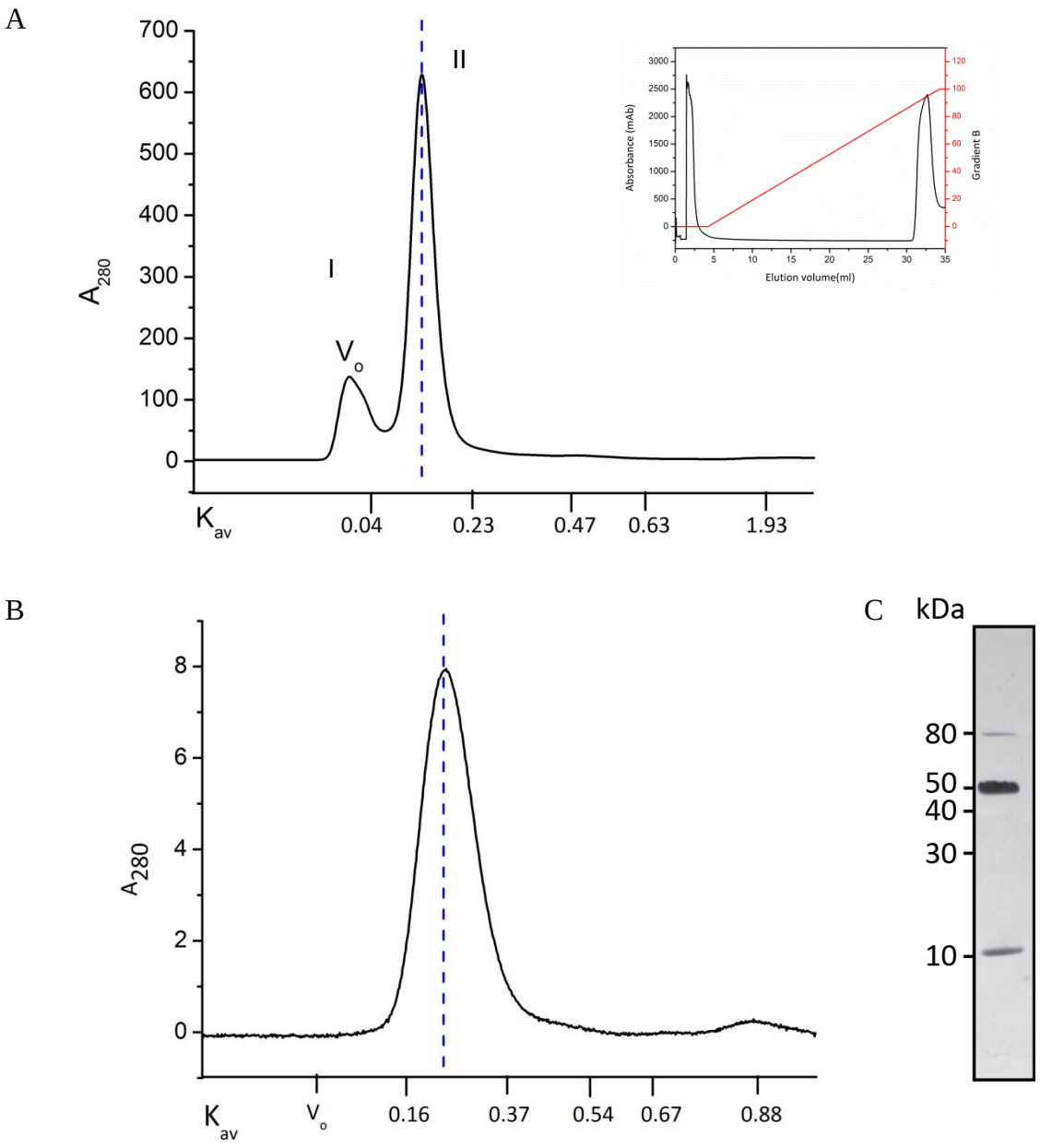


Figure 5.6| *SEC chromatograph of [H_6Lsmα]$_7$. (A) Preparative grade SEC (Superdex 200 16/600) column pre-equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, 2% glycerol with a flow rate of 1 mL/min. IMAC elution profile provided in inset. (B) Analytical SEC (Superdex 200 10/300) column, in the same buffer as above with a flow rate of 0.4 mL/min. (C) SDS-PAGE of purified [H_6Lsmα]$_7$.*

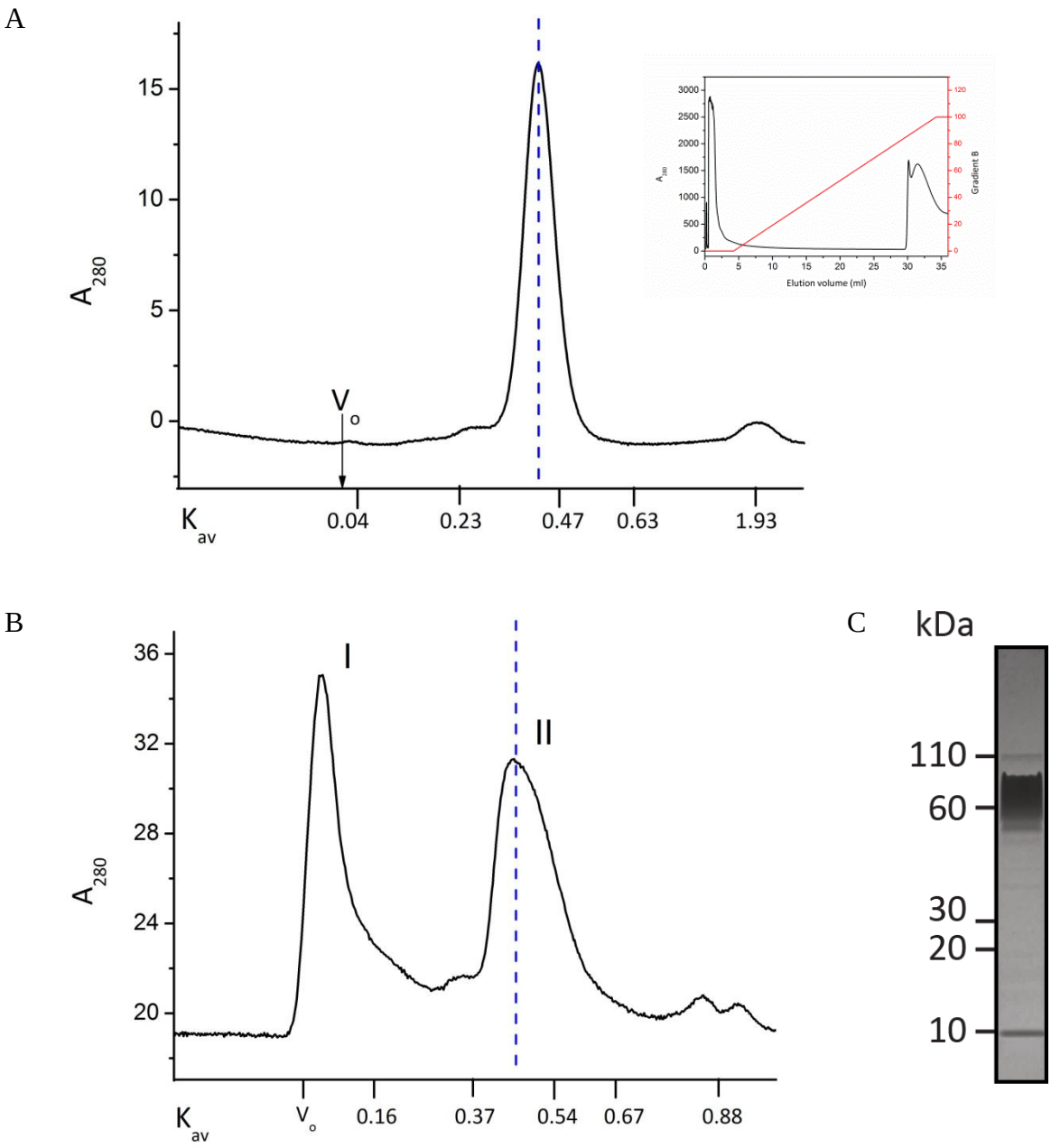


Figure 5.7| *SEC chromatograph of $[LsmαH_6]_7$. (A) Preparative grade SEC (Superdex 200 16/600) column pre-equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl, 2% glycerol with a flow rate of 1 mL/min. IMAC elution profile provided in inset. (B) Analytical SEC (Superdex 200 10/300), in the same buffer as above with a flow rate of 0.4 mL/min. (C) SDS-PAGE of $[LsmαH_6]_7$.*

$[LsmαH_6]_7$ exhibited a different SEC profile compared to $[H_6Lsmα]_7$. On the preparative SEC, it eluted as a single species with negligible material eluting in the void volume (**Figure 5.7A**), in contrast to the bimodal distribution obtained for $[H_6Lsmα]_7$. The calculated K_{av} of 0.43 corresponds to ~ 50 kDa. This fraction separated as two species when injected onto an analytical sizing column in Tris/NaCl pH 8.0 buffer (**Figure 5.7B**). The majority formed aggregates eluting at the void volume, whereas the second fraction had a K_{av} of 0.45 corresponding to ~66 kDa, indicative of a single heptameric ring. Characterisation of this fraction proved to be difficult as it precipitated within a few hours at 4 °C. Therefore, due to the unstable behaviour of $[LsmαH_6]$ compared to $[H_6Lsmα]_7$, the latter was chosen for further study.

From the solution studies of the two His_6-tagged constructs, the position of the tag plays a critical role in the multimerisation of the Lsmα rings to higher order species. For $[H_6Lsmα]_7$, the His_6-moiety is predicted to be oriented towards one face of the ring (**Section 5.1**), whereas, for $[LsmαH_6]_7$, since the tags are directed at the equatorial plane of the ring, the orientation of the His residues may not be conducive for interaction with metal ions. It is difficult to predict whether $[LsmαH_6]_7$ assemblies are metal directed due to the unstable nature of the sample and lack of subsequent characterisation.

Six to seven $[H_6Lsmα]_7$ rings form a discrete assembled structure. This species is in contrast to the single-ring species obtained by engineering Lsmα as a GST-fusion. Thus, the inclusion of a His_6-tag in the native sequence of Lsmα provides a new level of control for the hierarchical organisation of its quaternary structure.

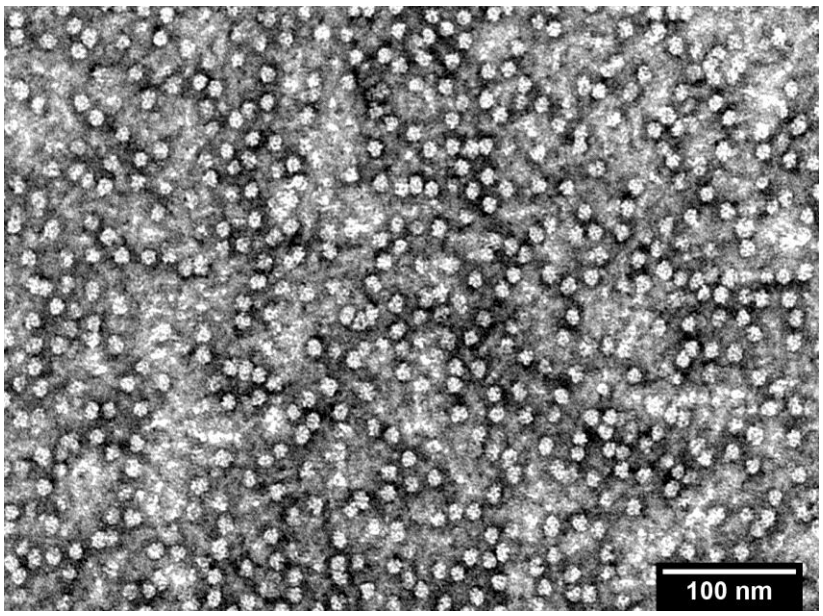
5.5 Visualisation of $[H_6Lsm\alpha]_7$ and $[Lsm\alpha H_6]_7$ by TEM

$[H_6Lsm\alpha]_7$ oligomerised into a defined cluster of individual rings as ascertained by negative stain transmission electron microscopy. The cluster had a diameter of about ~9 nm (**Figure 5.8**). In most of the particles, single rings with pores were visible. It is evident from the micrographs, that individual $[H_6Lsm\alpha]_7$ rings are arranged in a cage-like architecture. Such an arrangement is suggestive of a hollow cavity, formed by an angular arrangement of the rings. These dimensions are much larger than the single His_6-Lsmα rings (6.5 nm wide with central pore of 1.5 nm) as well for other Lsm proteins such as Lsm3 octamer [18] (7.5 nm wide and 2 nm pore). Therefore, the insertion of the His_6-tag at the N-terminal of the protein sequence triggered the association of Lsmα rings into a higher ordered organisation of a defined shape and dimension.

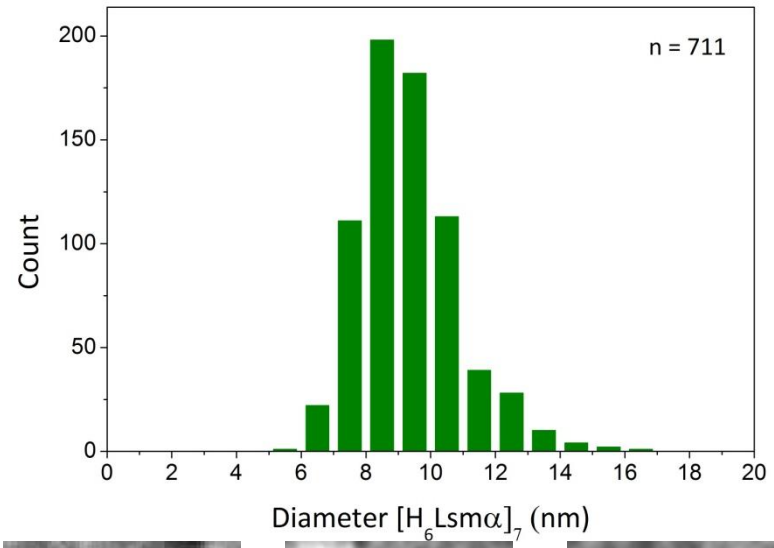
$[Lsm\alpha H_6]_7$ on the other hand, did not exhibit any definite assemblies as observed by TEM (**Figure 5.9**). Although, single rings were visible, the overall sample was heterogenous and nature of the sample indicated aggregate formation. Due to lack of any discrete and uniform organisation of $[Lsm\alpha H_6]_7$ rings, this construct was not characterised further.

Single particle averaging was attempted on $[H_6Lsm\alpha]_7$ using the EMAN suite. Due to the heterogeneous nature of the sample and high contrast due to negative staining, the obtained averaged image was not useful.

A



B



C

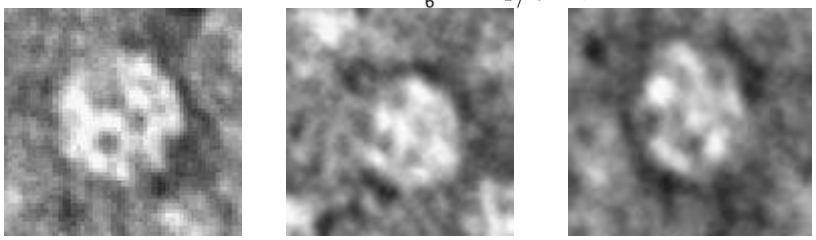


Figure 5.8| *Negative stain TEM of $[H_6Lsm\alpha]_7$. (A) Micrographs showing well-distributed distinct clusters of $[H_6Lsm\alpha]_7$ (0.05 mg/mL) prepared in Tris/NaCl buffer. (B) Size distribution histogram of $[H_6Lsm\alpha]_7$. (C) Magnified particles from the original micrographs.*

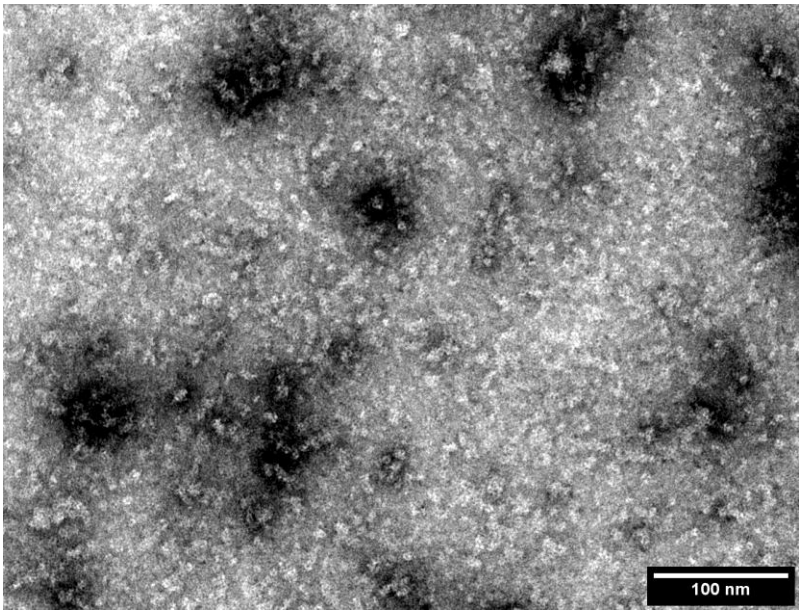


Figure 5.9| *TEM of $[LsmαH_6]_7$. Single rings are visible but the sample is heterogeneous in nature.*

Thus, TEM and analytical SEC analysis revealed that the $[H_6Lsmα]_7$ complex obtained is comprised of about six individual rings associated through contacts mediated by the His residues. Although His-tags do not cause structural changes in most proteins [19], there have been examples where inclusion of a His-tag in the native sequence has caused changes in disulfide bonding patterns [20] and conformational changes at a binding site [21]. Metal-assisted oligomerisation of His-tagged proteins has been seen in Pbx protein fused with a His_{10}-tag but the exact nature of the oligomer is not known [15]. Only recently there have been reports of rationally engineered single His and Cys residues in a protein sequence to artificially induce metal-mediated assembly of proteins. New crystal forms were produced by engineering metal contact points between protein interfaces for T4 lyzozyme [16]. This approach is also being used to develop novel protein nanostructures and assemblies.[22]

5.6 Polydispersity of $[H_6Lsm\alpha]_7$

Although $[H_6Lsm\alpha]_7$ eluted as a single species on the analytical SEC, the failure to produce an averaged structure of the TEM images suggested that the structures formed were polydisperse. In order to assess the polydispersity of the complex, SEC was coupled with static light scattering (SEC-SLS). This technique provides a direct measure of the molecular weight from the change in the light scattering based on the Debye-Zimm equation [23] (**Section 2.6.5**). $[H_6Lsm\alpha]_7$ samples eluted from the SEC column, were directly passed through a right angle light scattering detector to ascertain the molecular weight of each point. This measure of molecular weight was not constant over the peak as it trails down over the elution profile (**Figure 5.10**). The molecular weight range measured at each point is 400 to 500 kDa, indicating that $[H_6Lsm\alpha]_7$ consists of five to seven rings forming a complex. Attempts to generate a more monodisperse preparation of $[H_6Lsm\alpha]_7$ were trialled by employing different gel filtration matrices (Superose 12 and Sephacryl 400) to separate the five-membered complex from the six- and seven-membered complexes, but efficient separation could not be achieved. Since Lsmα is thermostable, it was hypothesised that heating it would encourage the formation of a stable homogenous product by rearrangement of the complex, but again such a strategy did not succeed in generating a more homogenous state. The purified $[H_6Lsm\alpha]_7$ sample was heated upto 60 °C, 70 °C and 80 °C, and analysed by analytical SEC. The results were similar to that obtained for the non-heated sample, hence failed to resolve the $[H_6Lsm\alpha]_7$ mixture (data not shown).

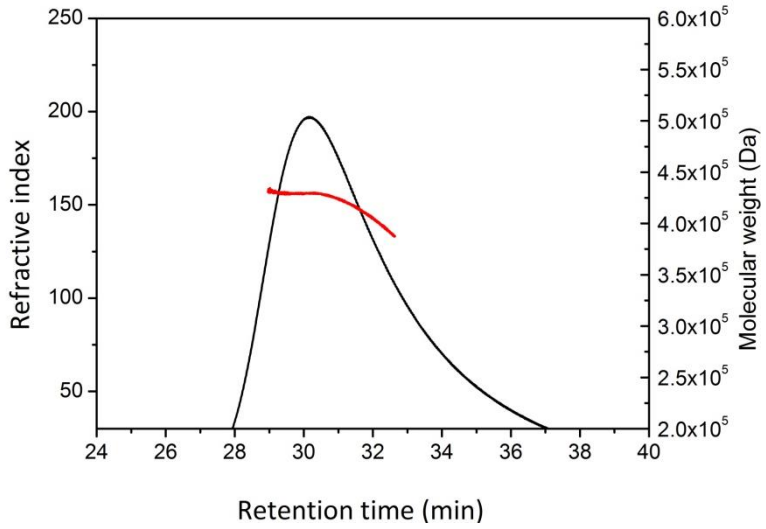


Figure 5.10| *Static light scattering showing the molecular weight distribution of* $[H_6Lsm\alpha]_7$. *The run was operated in 20 mM Tris, pH 8.0, 150 mM NaCl, 2% glycerol at a flow rate of 0.4 mL/min.*

5.7 Factors affecting the assembly of $[H_6Lsm\alpha]_7$

In order to probe the factors that control assembly of $[H_6Lsm\alpha]_7$ rings, several sets of solution conditions were tested. Previous studies on Lsm proteins have indicated that the ionic strength of the buffer influences the oligomerisation state of Lsm[4+1] polyprotein [24]. pH plays a role in stacking events of Lsm3, with low pH encouraging formation of higher molecular weight species and assembly of multiple octameric rings [25]. As noted in **Section 1.9**, head-to-tail stacking events have been observed in the case of untagged Lsmα (PDB ID ILOJ) thought to be facilitated by the electrostatic interactions of the loop4 and the helical face. Therefore, in order to understand the nature of $[H_6Lsm\alpha]_7$ assembly, solution characterisation with varying pH and salt concentrations was carried out.

The oligomeric assembly of $[H_6Lsm\alpha]_7$ was found to be influenced by pH. This behaviour was seen by aliquoting the affinity-purified samples into phosphate buffer at pH 7.0 and Tris buffer at pH 8.0 and pH 9.0. The salt concentration was kept constant at 150 mM NaCl. At pH 8.0, $[H_6Lsm\alpha]_7$ was observed to elute solely as a single species (~ 440 kDa). However, when the same affinity-purified protein was applied to the sizing column at pH 7.0, the majority eluted near the void volume, indicating aggregation. This could be attributed to the pI of His_6-Lsmα being 6.5 and nearing this value could trigger aggregation. At pH 9.0, the assembly seems to be reorganised into three species with the dominant fraction eluting at K_{av} 0.36 corresponding to ~300 kDa. At pH 9.0 and above, the imidazole ring is deprotonated which, as shown later in the chapter, is predicted to weaken the interactions between the metal ion and the His residues. The behaviour of $[H_6Lsm\alpha]_7$ at high pH could also be explained by ionisation of the pyrrole hydrogen of the His side-chain [26 27] which may cause perturbation in the already formed assembly and cause it to rearrange to a more stable complex at high pH [28].

Affinity-purified samples of $[H_6Lsm\alpha]_7$ were also dialysed into buffers with different ionic strength (100-800 mM NaCl). These conditions did not alter the $[H_6Lsm\alpha]_7$ organisation, thus, ruling out electrostatically-mediated interactions between the rings.

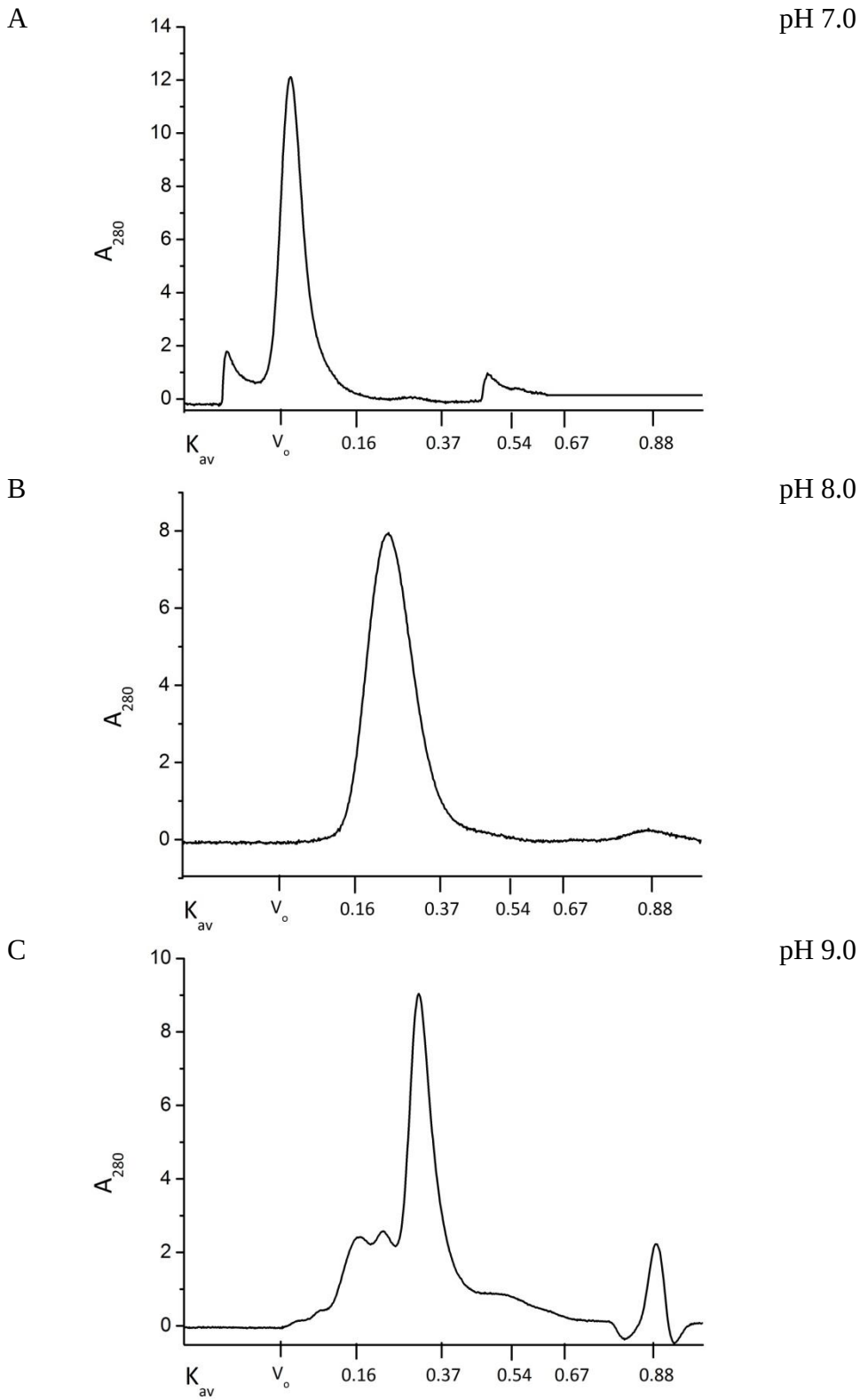


Figure 5.11| *SEC traces of $[H_6Lsm\alpha]_7$ at different pH carried out using a Superdex 200 10/300 column at a flow rate of 0.4 mL/min.*

5.8 $[H_6Lsm\alpha]_7$ is Cu^{2+} bound

As described in **Section 5.3**, $[H_6Lsm\alpha]_7$ was prepared using a Cu^{2+}-matrix. During the purification procedure, the soluble lysate was exposed to this matrix for 1 h, allowing for maximum product adsorption. This may have resulted in metal leaching from the IMAC adsorbant, forming a chelation product with the recombinant protein. To test this hypothesis, $[H_6Lsm\alpha]_7$ samples were washed with competing chelating agent, EDTA, to ensure an apo (metal-free) form of the complex. **Figure 5.12**, shows a time course of SEC chromatograms of $[H_6Lsm\alpha]_7$ material treated over 48 hours. This experiment opens up the possibility of switchable conversion from a large complex to a smaller species by removal of bound metal.

(i) Size-distribution at time 4 h

$[H_6Lsm\alpha]_7$ was treated with a molar excess of EDTA (10 mM). **Figure 5.12A** shows the elution profile of this sample taken after 4 h of exposure to the chelating agent. It showed a bimodal distribution with fraction I at K_{av} 0.2 (~ 400 kDa) and fraction II at 0.4 (~ 70 kDa). These two fractions corresponded to the six-ring membered $[H_6Lsm\alpha]_7$ assembly and single ring form, respectively. The isolated sample from fraction II, when re-chromatographed in the sizing column, did not rearrange to form other oligomeric forms, consistent with the assembly process being metal dependent. In order to drive the equilibrium towards smaller species, a more extended treatment of EDTA was employed.

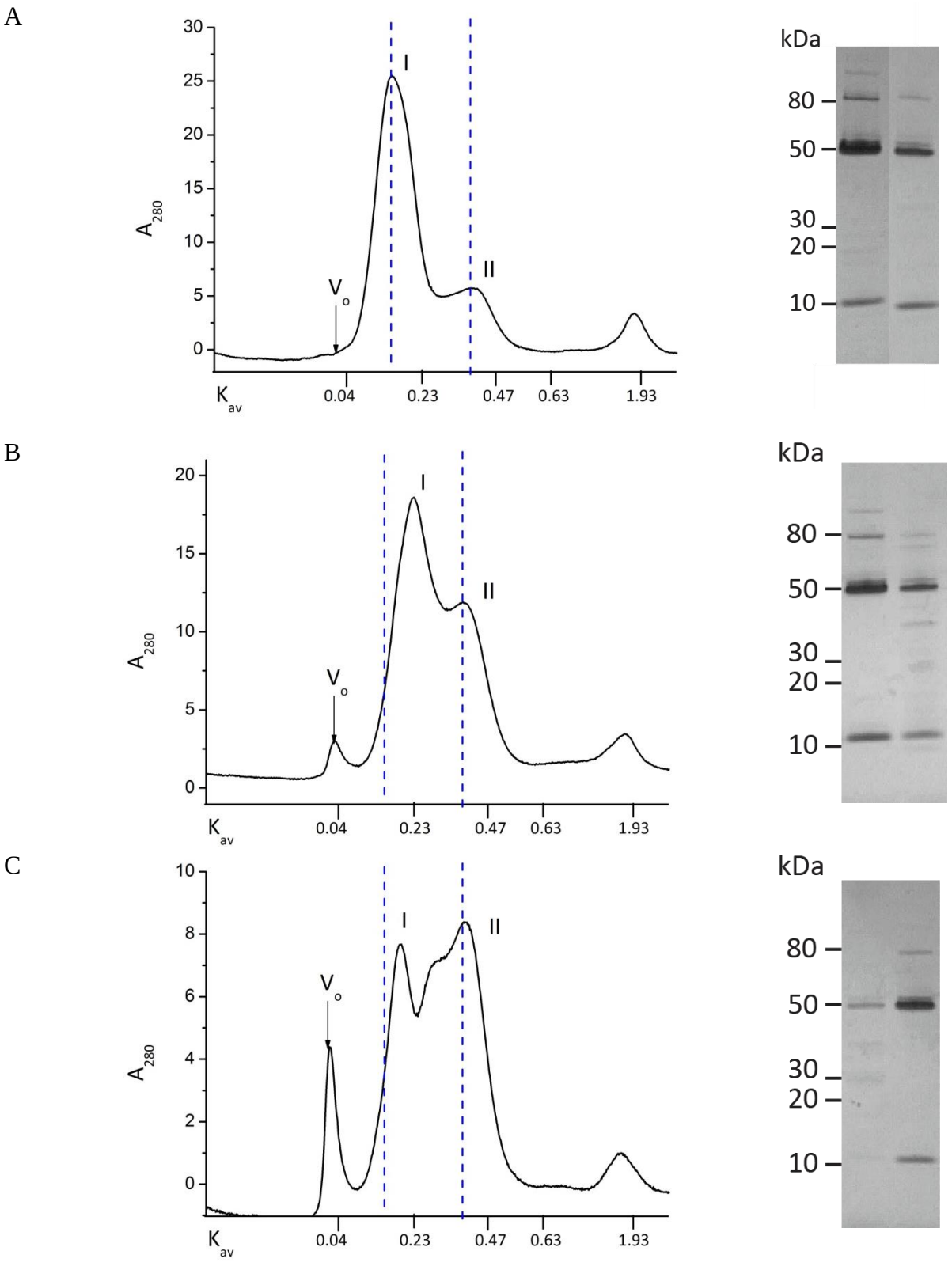


Figure 5.12| *SEC profiles of $[H_6Lsm\alpha]_7$ in presence of EDTA carried out in Superdex 200 16/600 with a flow rate of 1 mL/min. (A) 10 mM EDTA, 4 h. (B) 15 mM EDTA, 24 h. (C) 15 mM EDTA, 48 h. Dashed lines indicate position of the $[H_6Lsm\alpha]_7$ complex and single ring.*

(ii) Size-distribution at time 24 h

The percentage of metal-free single ring form of $[H_6Lsm\alpha]_7$ increased with an increase in concentration of EDTA and extending the contact time to 24 h. From **Figure 5.12B**, fraction I shifted its elution volume to K_{av} 0.23, indicating further dissociation into smaller species comprising of roughly four rings. There was also formation of aggregates as seen in the void volume.

(iii) Size-distribution at time 48 h

The SEC analysis revealed a highly mixed population after 48 h of incubation with EDTA (**Figure 5.12C**). There was a marked increase in the formation of the single-ring apo (metal-free) form of $[H_6Lsm\alpha]_7$. Due to the long duration of treatment there was more aggregate formation as seen from the fraction eluting at the void volume.

The next step was to explore the conditions which provide a more efficient switch to transition between the six-ring membered metal-bound assembly and the single-ring metal-free form.

(iv) Complete dissociation of $[H_6Lsm\alpha]_7$ metal-bound assembly

As seen from the pH studies carried out on $[H_6Lsm\alpha]_7$ (**Section 5.7**), the assembly dissociated into smaller species at pH 9 and above. Therefore, this condition was applied to the affinity-purified $[H_6Lsm\alpha]_7$ in the presence of 20 mM EDTA for 26 h at 4 °C. A complete dissociation of the cage-like assembly took place, as seen from the SEC trace (**Figure 5.13A**). This fraction was analysed by TEM, which also showed the presence of single rings (**Figure 5.13B**). Thus, this condition is a suitable means to switch from a Cu^{2+}-bound oligomeric assembly to a metal-free single-ring form.

Table 5.1| Table showing percentage fraction of the species formed during EDTA treatment of $[H_6Lsm\alpha]_7$.

Condition[a]	Percentage fraction[b]		
	Single ring	Intermediate species	Metal-bound $[H_6Lsm\alpha]_7$
20 mM EDTA, 26 h, pH 9.0	100	-	-
15 mM EDTA, 48 h, pH 8.0	54	46	-
15 mM EDTA, 24 h, pH 8.0	42	58	-
10 mM EDTA, 4 h, pH 8.0	28	-	72

[a] 20 mM Tris pH 8.0, 150 mM NaCl

[b] Area under curve calculated in OriginPro software suite.

(v) Formation of metal-bound $[H_6Lsm\alpha]_7$

In metal affinity chromatography, the metal ion is not directly attached to the polymer support but via a chelating group, which in turn is covalently attached to the polymer (**Figure 5.5B**). Thus, the protein engineered with metal-coordinating groups (affinity-groups) adsorb onto the functionalised polymer for affinity separation. The molecular mechanism of protein adsorption to the derivatised functional polymer surface involves protein-to-metal interactions. Since the binding energy of metal-to-ligand is larger than that of protein-to-metal binding [29; 30; 31], the IMAC matrix is a robust and efficient system for recombinant protein isolation. A special case arises for variants with multiple surface metal-coordinating groups which have larger binding affinity. This causes the protein to simultaneously interact with multiple metal ions on the derivatised surface. The binding selectivity in affinity chromatography methodology is largely influenced by the multiplicity and microenvironment created by the metal-binding sites (His residues and deprotonated amines) exposed on the protein surface [32; 33].

In $[H_6Lsm\alpha]_7$, seven affinity-segments are oriented towards one side of the ring scaffold, with six His side-chains on each segment, serving as potential sites to interact with metal ions. Strong binding of $[H_6Lsm\alpha]_7$ is achieved by multi-point attachment of these residues. This complex interaction leads to the formation of a very strong protein-Cu^{2+}-chelate complex, greater than the Cu^{2+}-IDA interaction. Therefore, it is plausible that $[H_6Lsm\alpha]_7$ is formed in the microenvironment created by the ligand-metal-protein complex and the assembly thus formed leaches Cu^{2+} from the matrix. Subsequently, with reasonable exposure time of adsorption to the Cu^{2+}-matrix, $[H_6Lsm\alpha]_7$ eluted out as a metal-bound form in high ionic strength buffer. Taking into account the three-dimensional scaffolding by the folded protein, the rings conform to a closed-structure yielding to $[H_6Lsm\alpha]_7$ rings surrounding a cluster of Cu^{2+} ions.

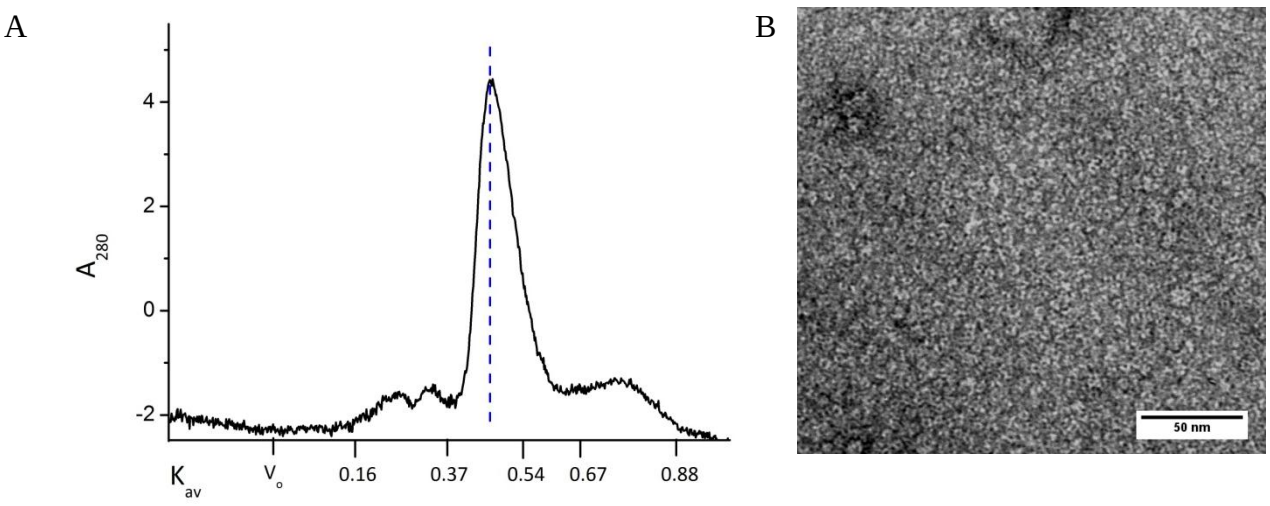


Figure 5.13| *Single-ring form of $[H_6Lsm\alpha]_7$ shown by (A) Analytical SEC; the dashed line indicates the elution volume of a single ring state and (B) TEM of $[H_6Lsm\alpha]_7$ as single rings post-EDTA treatment.*

5.9 Metal conjugation of $[H_6Lsm\alpha]_7$

Selective metals were chosen depending on their affinity to His residues and incubated with $[H_6Lsm\alpha]_7$. Analytical SEC was used to analyse any change in oligomeric state of the complex formed. As discussed in the previous **Section 5.3**, transition metals such as Ni^{2+}, Zn^{2+}, and Co^{2+} have high specific affinity to His groups exposed on the protein surface.. Addition of a molar excess of nickel ions to the $[H_6Lsm\alpha]_7$ sample shifted the elution peak towards a higher-order molecular weight species (**Figure 5.14B**). This fraction when isolated and treated with a competing chelating agent, shifts back to a ~400 kDa form (**Figure 5.14C**). This indicates a reversible metal-binding capacity of affinity-purified $[H_6Lsm\alpha]_7$ complex. Thus, there is formation of a bi-metallic $[H_6Lsm\alpha]_7$ complex, that binds to Cu^{2+} ions from the affinity-chromatography and Ni^{2+} ions when added externally. The most likely explanation of this occurrence is the high specific affinity of Ni^{2+} to the imidazole groups exposed on the surface of the protein, which leads to multimerisation of the rings onto the affinity-purified $[H_6Lsm\alpha]_7$. Similar experiments carried out with Zn^{2+} and Co^{2+} ions, did not cause any change in multimerisation and higher-order oligomerisation of $[H_6Lsm\alpha]_7$.

Metal ions were added to un-tagged Lsmα to assess its metal tolerance. Equimolar ratio of metal to protein invariably resulted in white precipitate indicating that the His_6-tag moiety is required for metal binding and increases the robustness of the Lsmα complex formed.

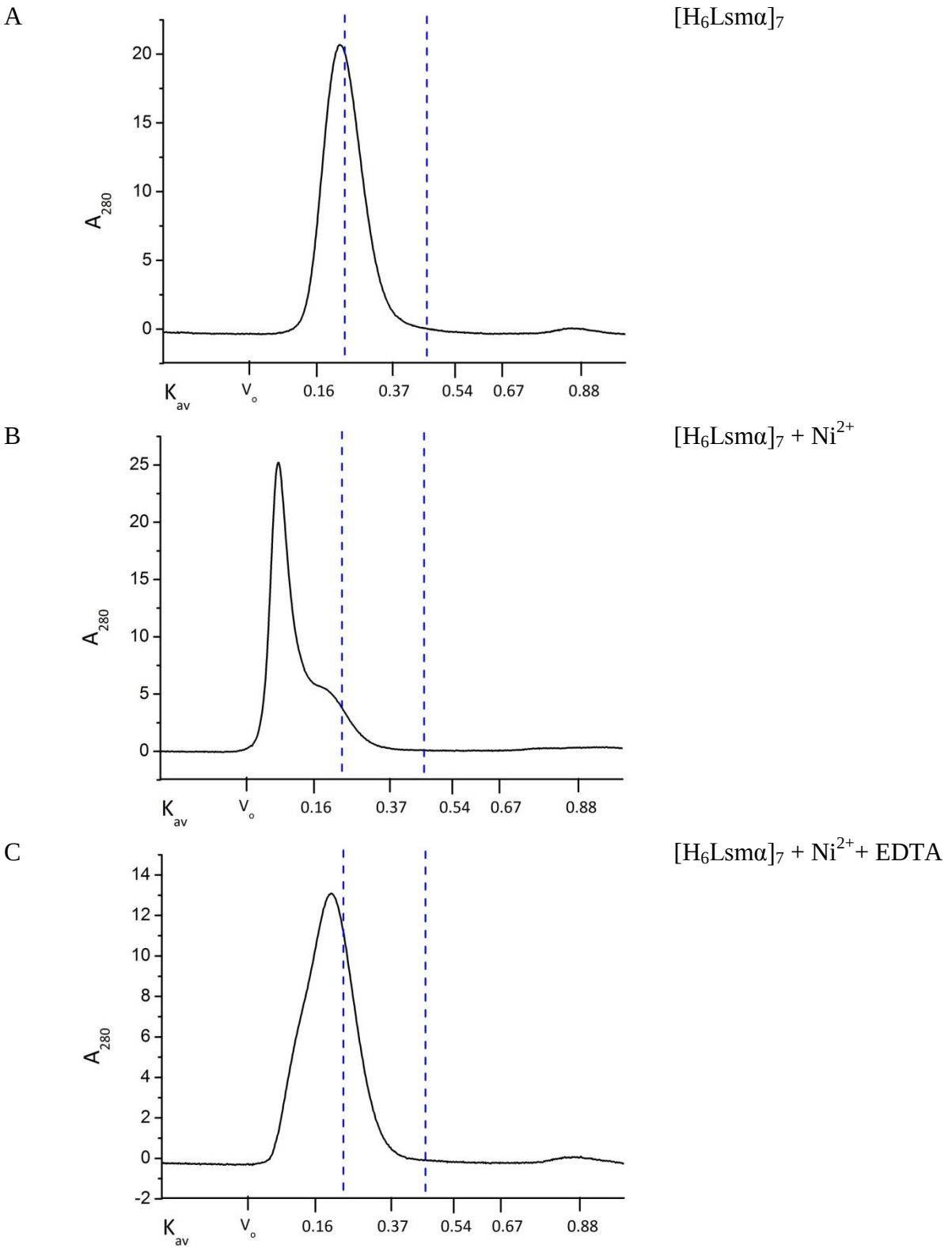


Figure 5.14| *SEC traces showing solution behaviour of $[H_6Lsm\alpha]_7$ on addition of Ni^{2+}(0.2 mM). The dotted lines represent elution position of $[H_6Lsm\alpha]_7$ complex and single-ring form. The running buffer was 20 mM Tris, pH 8.0, 150 mM NaCl, 2% glycerol. Sizing column used was a Superdex 200 10/300 column run at a flow rate of 0.4 mL/min.*

5.10 Small-angle X-ray scattering (SAXS)

In previous solution state characterisation of Lsm proteins at the Macquarie Protein Structural Group, none of the prepared Lsm proteins have shown discrete higher order assemblies. Although stacking has been inferred by SEC and seen in crystal packing, both axial and staggered, in different cases of Lsm proteins, including Lsmα, clusters have not been previously observed. Therefore, SAXS was used to pursue structural characterisation of the $[H_6Lsm\alpha]_7$ construct.

Samples of $[H_6Lsm\alpha]_7$, were analysed using the SAXS beamline at the Australian Synchrotron [34]. The SAXS beamline has a protein specific arrangement with an in-line SEC system to fractionate protein preparations into their individual components. A Superdex 200 5/150 GL gel filtration column was placed in-line with the SAXS beamline. This arrangement ensured monodispersity and buffer matching which are crucial for obtaining quality SAXS data for proteins [35]. Data were continuously collected as the protein eluted from the column. Scattering patterns for $[H_6Lsm\alpha]_7$ were radially averaged to derive one-dimensional scattering curves and scattering contributions of the buffer were subtracted from the presented curves. This data reduction was carried out using ScatterBrain software developed at the Australian Synchrotron [34].

To assess data quality and provide a more precise definition of molecular dimensions, Guinier plots were calculated. The fact that the plots were linear and the scattering data did not show any upward curvature at low angles, allows us to eliminate the possibility of aggregate formation (**Figure 5.15B**). The Guinier plots provide data quality assurance for further processing and shape determination. The slope of Guinier plots [36] yield radius of gyration, R_g values in reciprocal space. For $[H_6Lsm\alpha]_7$ the R_g^{rec} was calculated as 5.1±0.05 nm.

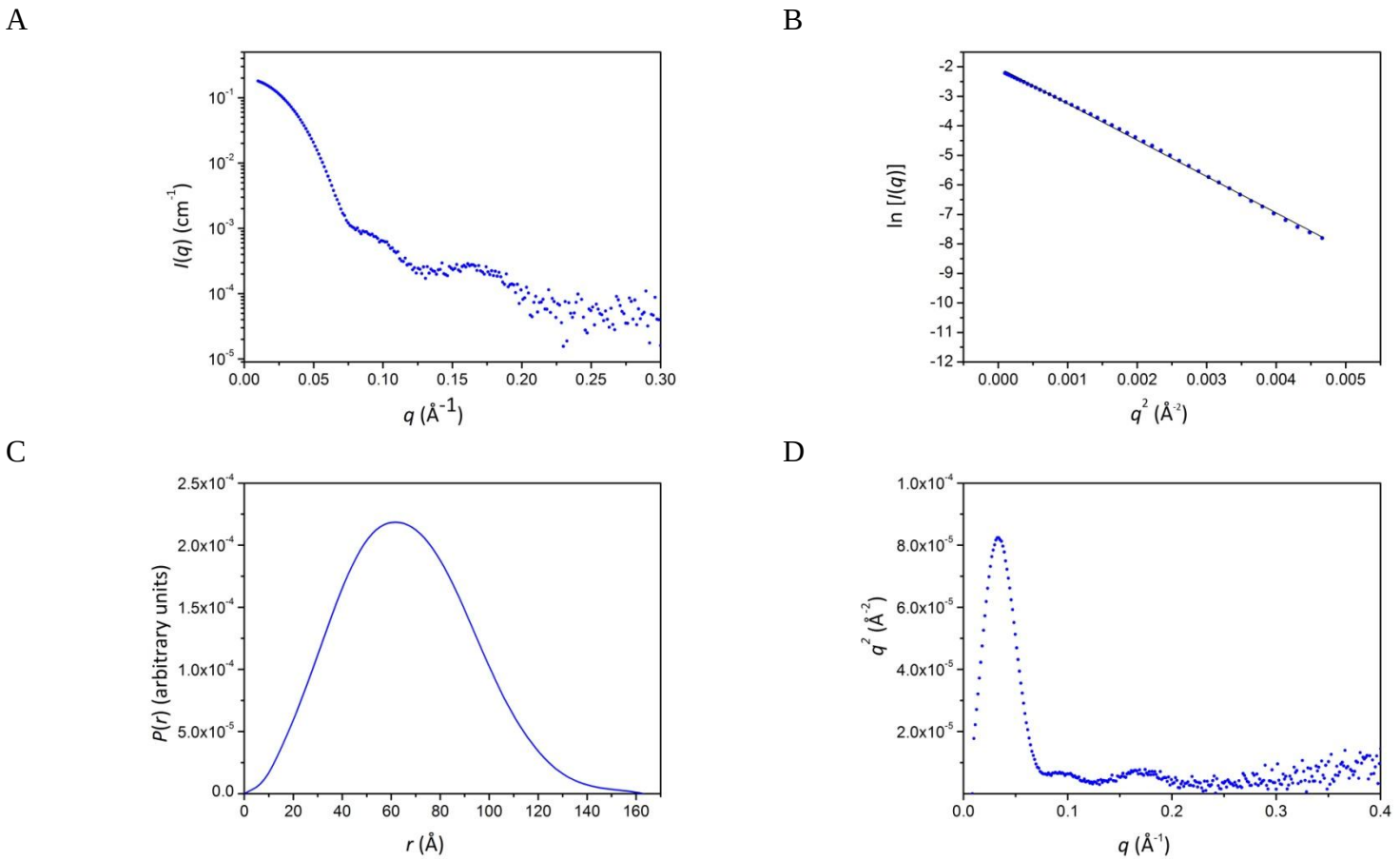


Figure 5.15| *SAXS data recorded at SAXS/WAXS beamline at Australian Synchrotron. The following parameters were calculated for* $[H_6Lsm\alpha]_7$ *loaded onto the SEC column at 2 mg/mL (A) Scattering curve. (B) Guinier approximation. (C) Distance distribution function. (D) Kratky representation of scattering data.*

Fourier transformation of collected data produced smooth bell-shaped electron-pair distribution function (*P(r)*) for this complex (**Figure 5.15C**). The real space radius of gyration (R_g [real]) was 5.16±0.1 nm and the maximum particle dimension, D_{max} was 16.3 nm calculated from Autoporod [37]. The real space radius of gyration is in good agreement with reciprocal space data derived from the Guinier fit. In order to determine that $[H_6Lsm\alpha]_7$ exists in a folded globular state, a Kratky plot was derived (**Figure 5.15D**). The sample yielded a sharp clear peak at low values of *q*, with a plateau at higher values. This trend is indicative of a folded globular protein.

SAXS data can be used to derive three-dimensional molecular envelopes [38]. This provides a detailed picture regarding the structure of the protein being studied. *Ab initio* reconstruction of $[H_6Lsm\alpha]_7$ using GASBOR [35] yielded an obtuse ellipsoid geometry. To validate the solution structure of $[H_6Lsm\alpha]_7$ obtained, the coordinate file, PDB 1I81, was used to calculate a theoretical X-ray scattering curve. The resulting curve was compared to the experimental scattering curve. Both visual inspection and the Chi^2 value of 5.1 suggested that it is a very poor fit (**Figure 5.16**) confirming that the $[H_6Lsm\alpha]_7$ complex exists as a higher-order organisation of individual Lsmα rings.

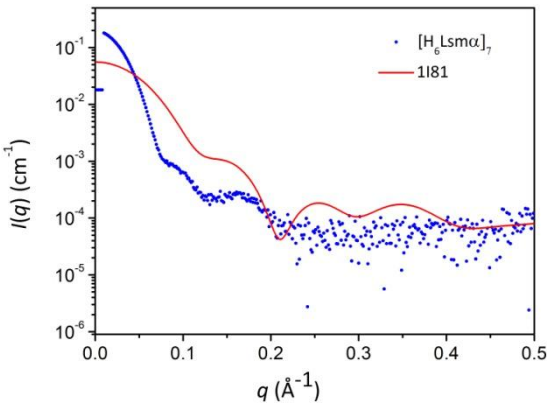


Figure 5.16| *Overlay of experimental scattering by $[H_6Lsm\alpha]_7$ (blue) and theoretical scattering of Lsmα (PDB ID 1I81) (red) using CRYSOL* [39].

In order to check whether the oligomerisation of $[H_6Lsm\alpha]_7$ was concentration dependent, different sections of the eluted peak as the sample eluted from the SEC column were analysed. The frames were chosen and analysed separately for the fractions. **Table 5.2** lists the parameters collected for different fractions of $[H_6Lsm\alpha]_7$. The molecular weight was calculated from the Porod volume using Autoporod [37]. The data suggest that the oligomerisation of $[H_6Lsm\alpha]_7$ is similar at different sections of the eluted peak, thus indicating concentration independent oligomerisation. **Figure 5.17** shows the overlaid scattering curves of the frames collected.

$[H_6Lsm\alpha]_7$ has a molecular weight distribution of 400-500 kDa as determined by SEC-SLS (**Section 5.6**). This indicates a modular combination of five to seven Lsmα rings. Pymol was used to generate combinations of rings using symmetry functions and PDB files were generated. Theoretical curves from these PDB files were then compared to the experimental scattering curve of $[H_6Lsm\alpha]_7$ to assess the arrangement of rings in the complex (**Figure 5.18**).

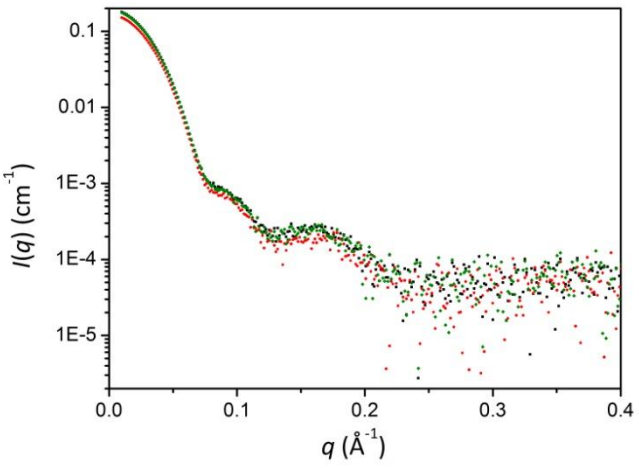


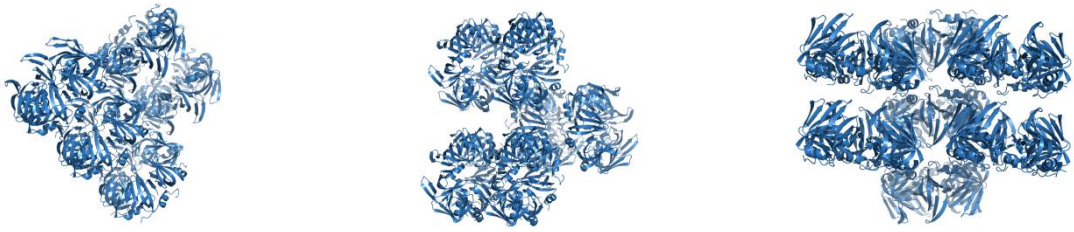
***Figure 5.17**| Overlaid non-scaled scattering curves of $[H_6Lsm\alpha]_7$ calculated from different regions of the eluted proteins at different concentrations.*

Table 5.2| Geometric parameters collected for different fractions of $[H_6Lsm\alpha]_7$.

Number of frames	D_{max} [a] (nm)	R_g [rec] (nm) [a]	R_g [real] (nm) [b]	Volume [a] (nm^3)	M_w [a] (±0.1 kDa)	Number of rings
7	16.1	5.1±0.05	5.1±0.1	920	567	7.3
10	16.3	5.1±0.05	5.1±0.1	908	556	7.2
10	16.0	5.1±0.05	5.1±0.1	900	554	7.2

[a] Calculated from Autoporod [32].
[b] Calculated from GNOM [33].
R_g^{rec}, reciprocal space radius of gyration derived from Guinier approximation; R_g^{real}, real space radius of gyration from distance distribution function; D_{max}, maximum particle diameter.
All data were obtained in PBS, pH 8.0.

Visual inspection confirms poor agreement (**Figure 5.19B(ii)**) although it is better than the comparison between PDB 1I81 and the experimental data. It can therefore, be concluded that $[H_6Lsm\alpha]_7$ does not contain a stacked arrangement but maybe an angular formation of rings. This is also indicative of a hollow cavity re-enforcing the notion that this complex has a cage-like architecture as seen by TEM.



5-mer 6-mer 7-mer

Figure 5.18| *Different combinations of rings obtained from the crystal lattice of Lsmα (PDB ID 1I81) used to fit into ab initio model of* $[H_6Lsm\alpha]_7$. *Structures were generated in Pymol* [40].

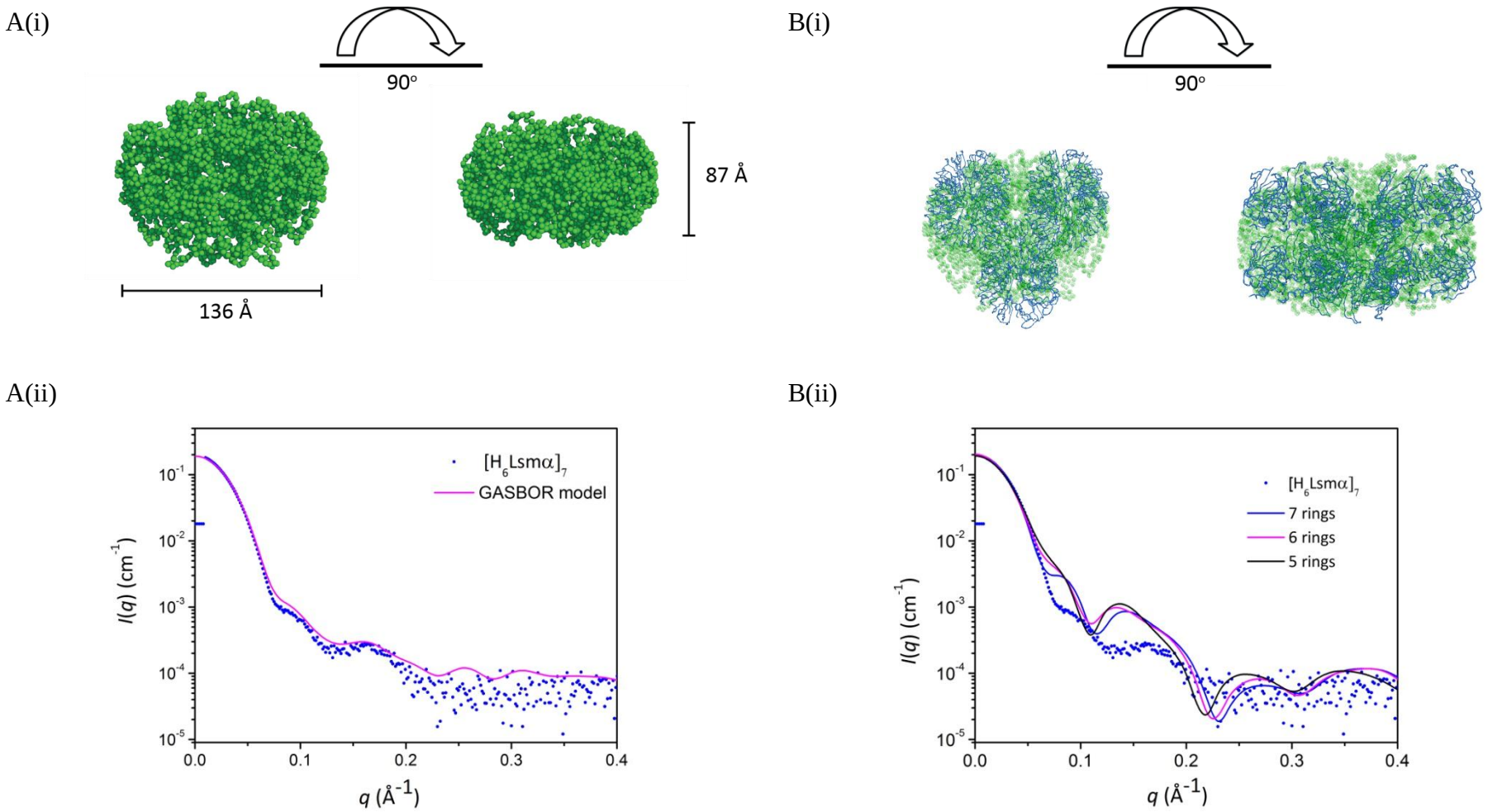


Figure 5.19| *(A(i)) Ab initio models of $[H_6Lsm\alpha]_7$ generated using GASBOR [35]. (A(ii)) Overlay of the ab initio model obtained (pink) and the experimental data (blue). (B(i)) Superimposition of six rings on the GASBOR model of $[H_6Lsm\alpha]_7$ using SUPCOMB [41]. (B(ii)) Fits of different combination of rings on the scattering curve of the ab initio model.*

5.11 Summary

The manipulation of the quaternary structure of His_6-tagged Lsmα into higher-order hierarchical organisations was achieved by placement of His_6-segments at one end of the sequence of the self-assembling system. The solution behaviour of Lsmα is influenced by the position of the tag at either the N- or C-terminus of the protein sequence, with the N-terminal $[H_6Lsm\alpha]_7$ complex exhibiting a more robust and stable solution behaviour and assembly into a discrete higher order organisation.

The $[H_6Lsm\alpha]_7$ assembly was shown to be Cu^{2+}-bound. Such behaviour has been observed in the case of other proteins isolated by affinity chromatography but $[H_6Lsm\alpha]_7$, being ring-shaped, organises into regular cage-like structures. Such a desirable architecture for generation of biological nanostructures is often lacking in globular proteins. The $[H_6Lsm\alpha]_7$ metal-bound assembly is consistent with:

i. Preferential multi-point interactions with high binding strength to copper ions via coordination bonds between the His moiety [27].
ii. Selective binding of the metal ion to the exposed and available deprotonated amines on the protein surface [9].

Both mechanisms provide an explanation for the formation of a thermodynamically stable metal-bound $[H_6Lsm\alpha]_7$ assembly resulting in a tightly bound closed-architecture. Such an organisation has not been observed for other Lsm protein systems. The types of interactions present in other Lsm proteins are:

i. Face-to-face stacking present in the case of Lsm3 of two octameric rings in the crystal structure mediated via the N-termini, involving salt bridge formation and His interactions [18].

ii. Head-to-tail stacking observed in Lsmα by electrostatic contributions of acidic residues present on the loop face and exposed basic and polar residues on the α-helix face [42] [43; 44].

In contrast, the interactions involved in the $[H_6Lsm\alpha]_7$ complex are mainly due to metal-coordination bonds and have been shown not to be electrostatic in nature. Although, pH plays a role in the stabilisation of the complex, conditions were optimised to switch it to a single-ring form. With the tag on the surface on the ring scaffold, it was possible to obtain ordered multimers of the modified ring through metals. The tags form reversible yet stable metal-chelate complexes and confer upon the Lsmα ring scaffold the ability to form supramolecular architectures.

5.12 References

1. Medalsy, I., Klein, M., Heyman, A., Shoseyov, O., RemacleF, Levine, R. D. & Porath, D. (2010). Logic implementations using a single nanoparticle-protein hybrid. *Nature Nanotechnology* **5**, 451-457.
2. Butts, C. A., Swift, J., Kang, S.-g., Di Costanzo, L., Christianson, D. W., Saven, J. G. & Dmochowski, I. J. (2008). Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* **47**, 12729-12739.
3. van Rijn, P. & Boker, A. (2011). Bionanoparticles and hybrid materials: tailored structural properties, self-assembly, materials and developments in the field. *Journal of Materials Chemistry* **21**, 16735-16747.
4. McMillan, R. A., Paavola, C. D., Howard, J., Chan, S. L., Zaluzec, N. J. & Trent, J. D. (2002). Ordered nanoparticle arrays formed on engineered chaperonin protein templates. *Nature Materials* **1**, 247-252.
5. Tsukamoto, R., Muraoka, M., Seki, M., Tabata, H. & Yamashita, I. (2007). Synthesis of CoPt and $FePt_3$ nanowires using the central channel of tobacco mosaic virus as a biotemplate. *Chemistry of Materials* **19**, 2389-2391.
6. Ardini, M., Giansanti, F., Di Leandro, L., Pitari, G., Cimini, A., Ottaviano, L., Donarelli, M., Santucci, S., Angelucci, F. & Ippoliti, R. (2014). Metal-induced self-assembly of peroxiredoxin as a tool for sorting ultrasmall gold nanoparticles into one-dimensional clusters. *Nanoscale* **6**, 8052-8061.
7. DeLano, W. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*.
8. Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599.
9. Gaberc-Porekar, V. & Menart, V. (2001). Perspectives of immobilized-metal affinity chromatography. *Journal of Biochemical and Biophysical Methods* **49**, 335-360.
10. Sundberg, R. J. & Martin, R. B. (1974). Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. *Chemical Reviews* **74**, 471-517.
11. Ueda, E., Gout, P. & Morganti, L. (2003). Current and prospective applications of metal ion-protein binding. *Journal of Chromatography A* **988**, 1-23.

12. Tsai, S.-Y., Lin, S.-C., Suen, S.-Y. & Hsu, W.-H. (2006). Effect of number of poly (His) tags on the adsorption of engineered proteins on immobilized metal affinity chromatography adsorbents. *Process Biochemistry* **41**, 2058-2067.
13. Hemdan, E. S., Zhao, Y.-J., Sulkowski, E. & Porath, J. (1989). Surface topography of histidine residues: a facile probe by immobilized metal ion affinity chromatography. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1811-1815.
14. Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J. & Schäfer, F. (2009). Immobilized-metal affinity chromatography (IMAC): a review. *Methods in Enzymology* **463**, 439-473.
15. Hom, L. G. & Volkman, L. (1998). Nickel-induced oligomerization of proteins containing 10-histidine tags. *Biotechniques* **25**, 20-22.
16. Laganowsky, A., Zhao, M., Soriaga, A. B., Sawaya, M. R., Cascio, D. & Yeates, T. O. (2011). An approach to crystallizing proteins by metal-mediated synthetic symmetrization. *Protein Science* **20**, 1876-1890.
17. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M. G. & Mabbutt, B. C. (2001). Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs1. *Journal of Molecular Biology* **309**, 915-923.
18. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R., Curmi, P. M. G. & Mabbutt, B. C. (2008). Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: implications for Lsm ring organisation and recruitment. *Journal of Molecular Biology* **377**, 1357-1371.
19. Carson, M., Johnson, D. H., McDonald, H., Brouillette, C. & DeLucas, L. J. (2007). His-tag impact on structure. *Acta Crystallographica Section D: Biological Crystallography* **63**, 295-301.
20. Klose, J., Wendt, N., Kubald, S., Krause, E., Fechner, K., Beyermann, M., Bienert, M., Rudolph, R. & Rothemund, S. (2004). Hexa-histidine tag position influences disulfide structure but not binding behavior of in vitro folded N-terminal domain of rat corticotropin-releasing factor receptor type 2a. *Protein Science* **13**, 2470-2475.
21. Chant, A., Kraemer-Pecore, C. M., Watkin, R. & Kneale, G. G. (2005). Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene

regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expression and Purification* **39**, 152-159.

22. Huard, D. J., Kane, K. M. & Tezcan, F. A. (2013). Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nature Chemical Biology* **9**, 169-176.
23. Zimm, B. H. (2004). The scattering of light and the radial distribution function of high polymer solutions. *The Journal of Chemical Physics* **16**, 1093-1099.
24. Molls, J. (2011). Structural characterisation of *yeast* Lsm protein complexes (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
25. Nishen, N. (2006). Molecular structures of the RNA-binding Lsm proteins (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
26. Martin, R. B. (1974). Pyrrole hydrogen ionization of imidazole derivatives in metal ion complexes and carbonic anhydrase. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 4346-4347.
27. Deschamps, P., Kulkarni, P., Gautam-Basak, M. & Sarkar, B. (2005). The saga of copper (II)–l-histidine. *Coordination Chemistry Reviews* **249**, 895-909.
28. Sigel, H. & Martin, R. B. (1982). Coordinating properties of the amide bond. Stability and structure of metal ion complexes of peptides and related ligands. *Chemical Reviews* **82**, 385-426.
29. Johnson, R. & Arnold, F. (1995). Review: Multipoint binding and heterogeneity in immobilized metal affinity chromatography. *Biotechnology and Bioengineering* **48**, 437-443.
30. Martell, A. E. & Smith, R. M. (1974). *Critical Stability Constants*, **1**, Springer.
31. Mohan, M. S., Bancroft, D. & Abbott, E. (1979). Mixed-ligand complexes of copper (II) with imidazole and selected ligands. *Inorganic Chemistry* **18**, 1527-1531.
32. Chaga, G. S. (2001). Twenty-five years of immobilized metal ion affinity chromatography: past, present and future. *Journal of Biochemical and Biophysical Methods* **49**, 313-334.
33. Zachariou, M. (2004). Immobilized metal ion affinity chromatography of proteins. In *HPLC of Peptides and Proteins*, pp. 89-102. Springer.

34. Kirby, N., Mudie, S. & Hawley, A. (2013). A low-background-intensity focusing small-angle X-ray scattering undulator beamline. *Journal of Applied Crystallography* **46.6**, 1670-1680.
35. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* **80**, 2946-2953.
36. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**, 1277-1282.
37. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Applied Crystallography* **40**, 223-228.
38. Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495-503.
39. Svergun, D., Barberato, C. & Koch, M. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773.
40. DeLano, W. L. (2002). The PyMOL molecular graphics system.
41. Kozin, M. B. & Svergun, D. I. (2001). Automated matching of high-and low-resolution structural models. *Journal of Applied Crystallography* **34**, 33-41.
42. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M. & Eisenberg, D. (2003). The oligomerization and ligand binding properties of Sm like archaeal proteins (SmAPs). *Protein Science* **12**, 832-847.
43. Mura, C., Phillips, M., Kozhukhovsky, A. & Eisenberg, D. (2003). Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4539-4544.
44. Mura, C. (2005). The Structures, Functions, and Evolution of Sm-like Archaeal Proteins (SmAPs) (PhD thesis), *University of California Los Angeles*.

6 Chapter six

Discussion and suggestions for future work

6.1 Introduction

Self-assembling proteins have been recognised in the literature as being among a number of biomolecules and biomolecular assemblies that have the potential to be used as nanomaterials [1; 2; 3]. The aims of this project were to explore the utility of the self-assembling Lsmα system and to design and characterise the various nanostructures that could be adopted by the ring-shaped protein. A number of other protein systems have been identified and explored for incorporation into nanodevices [4; 5] by modifying the protein interface to generate suitable structures. This approach provides an opportunity to design the proteins to assemble in a precise manner in three dimensional space [6; 7]. In order to exploit a bottom-up approach for the fabrication of molecular biomaterials, it has become important to better understand and control the process of self-assembly of protein units into their supramolecular organisations. Therefore, the study of more protein tectons with suitable properties is warranted.

Molecular building blocks should possess properties which make them suitable for downstream applications and provide better control over their assembly. First and foremost, the system should exhibit an inherent ability to self-assemble into clearly organised forms and combinations. This feature provides a basis from which to develop general principles for assembly of higher order structures from simple protein components. A naturally self-assembling system provides opportunities to engineer scaleable structures from the protein componentary.

Secondly, the tectons should possess adequate functionality that can be explored for potential applications. The chemistry of protein modules allows metals to be readily introduced or deposited on the surface or cavities of proteins depending on the purpose. Such methods of using proteins as metal-depositing templates have been explored with a variety of metals [8; 9; 10]. Certain protein parts have affinity towards other biomolecules such as nucleotides, and hence can act as capturing-units when used for fabrication of biosensors or drug delivery vehicles. Proteins also provide attachment points for embellishment or molecular modification, thereby providing unique functionality to the supramolecular structures.

The third requirement is the stability of the protein assembled structure, which increases the viability of downstream applications. Availability of thermostable versions of protein tectons is an added advantage for fabrication purposes. Additional properties, such as easy expression and purification of the proteins and its variants, are required to explore a molecular building block for its applications in bionanotechnology.

Lsm proteins and their variants were identified as meeting these requirements. They constitute a large indispensible protein family of RNA chaperones in all domains of life [11]. Lsm proteins from diverse organisms have been found in ring morphologies with diameters ranging from 5.5 – 7.5 nm, with a central cavity of 0.5 – 1.5 nm [12]. Regardless of the amino acid sequence, Lsm proteins show the propensity to assemble into ring-shaped quaternary structures, with five to eight monomers forming the toroid complex. This feature made Lsmα an attractive target to explore as a molecular tecton.

All Lsm proteins possess RNA binding motifs making them suitable as capturing devices. Thermostable versions of Lsm proteins are also available which exhibit protein stability up to 90 °C. Although Lsm proteins have been studied extensively for their role in RNA processing, investigating the assembling properties for its role in bionanotechnology is addressed here for the first time.

This thesis aimed to investigate the properties and adaptability of Lsm proteins suitable for applications in bionanotechnology. A bottom-up approach was used to fabricate molecular structures from the self-assembling thermostable Lsmα system. The native Lsmα assemblies were modified by site-directed mutagenesis to generate oligomers of different pore sizes, introduce novel functionality and generate supramolecular assemblies of interest. The structures were determined through biophysical techniques of SEC-SLS, SAXS and finally TEM. The scheme of various approaches utilised to change the assembly and supramolecular organisation of Lsm proteins is shown in **Figure 6.1**.

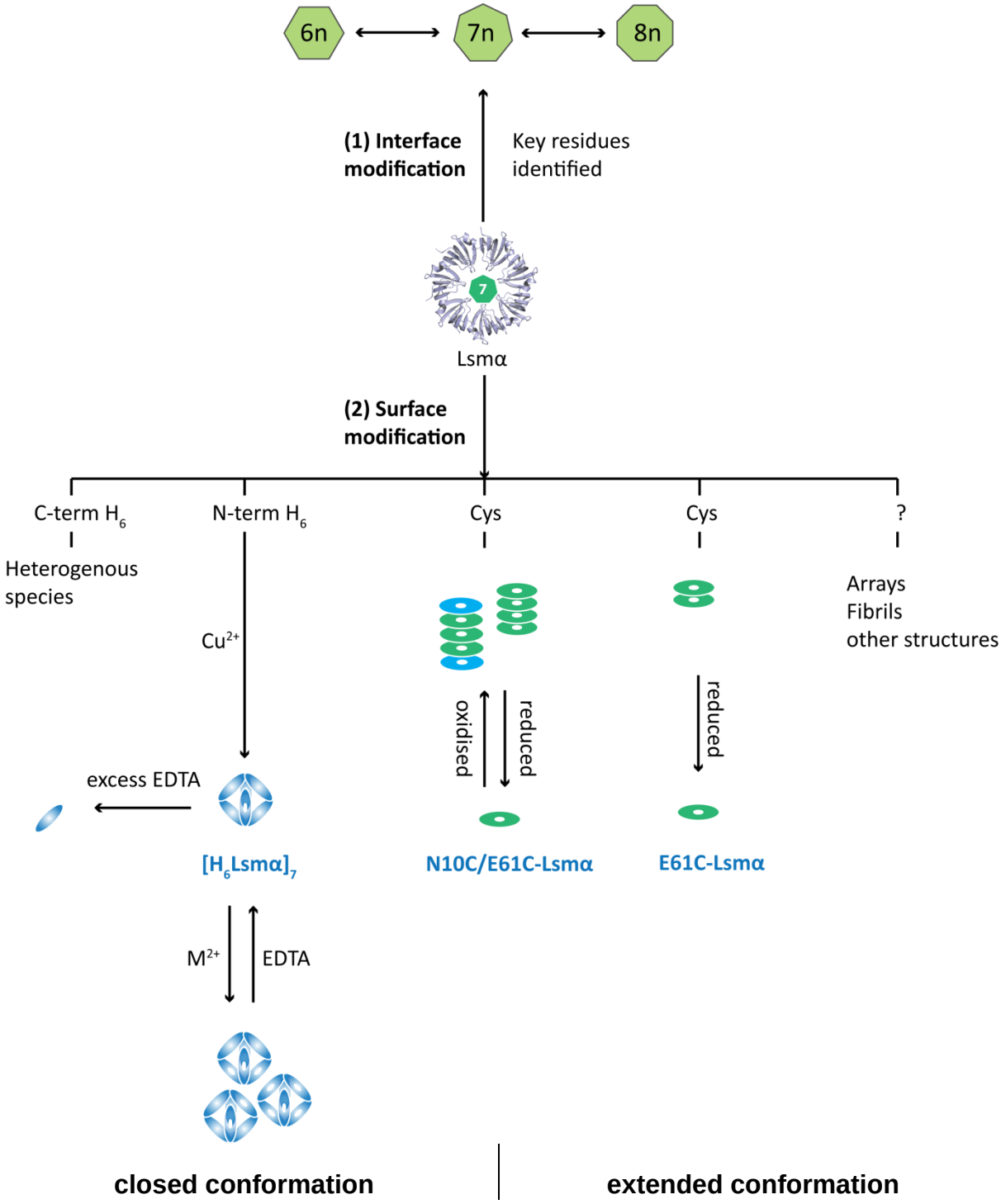


Figure 6.1| *Schematic diagram of multimer formation of Lsmα proteins generated by targeting two regions of the ring scaffold. (i) Interface modification (ii) Surface modification. Six possible outcomes can be achieved using Lsmα as a building block, out of which five have been explored in this thesis.*

6.2 Control of the pore size

Initially, two sources of Lsm proteins were chosen, namely the eukaryotic Lsm3 (*S. cerevisiae*) [13] polypeptide that assembled as an octamer, and the archaeal Lsmα (*M. thermoautotrophicum*) which formed a heptameric ring. After initial expression studies, Lsmα easily expressed as a soluble, highly stable protein, and was therefore chosen as the model protein for this thesis.

The Lsm protein complexes assemble into cyclic oligomers and exhibit a high degree of oligomeric plasticity assembling into different states ranging from a trimeric to an octameric form in different species when generated *in vitro*. Despite their conservation at the amino acid sequence level and similar 3D folds, Lsm proteins are commonly found in hexameric and heptameric states and less frequently as a pentamer and octamer. It was hypothesised that the quaternary structure could be altered by judicious changes at the inter-subunit contacts and hence, cyclic oligomers of different pore sizes obtained.

Probing the inter-subunit interactions is important to the understanding of self-assembly of circular protein oligomers. A strategy employed by Chen et al. to change the oligomeric state of TRAP protein from *B. stearothermophilus* and *B. halodurans* was to (i) delete the C-terminal region and (ii) insert a bulky hydrophobic residue. In both cases, it caused rotation of the subunits leading to the formation of a 12-mer from an 11-mer [14]. This strategy was successful because key hydrogen bonds were disrupted, allowing for the formation of a 12-mer circular oligomer.

Lsm protein interface presents a complex network of hydrogen bonds and hydrophobic interactions, hence a different approach was employed to investigate the subunit interface. The protomeric interface of Lsmα was studied by alanine screening. As observed by SAXS and SEC-SLS, there was no apparent change in the oligomeric state of most of the muteins.

The error inherent in these techniques made it difficult to distinguish between a heptamer and a hexamer. For all the samples tested, none of them showed molecular weights greater than that of a heptamer, therefore, octameric and higher order oligomeric states can be ruled out. The mutein, R65P, consistently showed results corresponding to the molecular weight of a hexamer and it would be interesting to verify the quaternary structure of this mutein by X-ray crystallography or native mass spectrometry.

Two muteins, L70A and I71A, exhibited complete breakdown into the monomeric form as assessed by SDS-PAGE, indicating that residues L70 and I71 are important in the stability of the cyclic oligomer. For L70A, a monomeric form was observed by preparative SEC. It would be interesting to optimise the conditions to isolate this monomeric form in solution and characterise it further. The I71A mutein did not show presence of any monomeric form in solution but dissociated completely on the SDS-PAGE gel.

In addition to changing single residues at the interface, chimeric muteins were also generated to alter the oligomeric assembly. However, this change in the entire interfacing segments, caused the protein to dissociate into monomers as assessed by SDS-PAGE. It most likely disrupted the protomeric interface, which resulted in failure of ring formation. This indicates that the associations are conserved within species and also between each of the protomers.

6.3 Generation of nanotubes

Previous studies on synthetically constructed Lsm proteins have shown electrostatically-mediated multimerisation events which were thought to occur via stacking of the rings:

(i) Helix-to-helix face stacking of two Lsm protein rings has been observed in the crystal structure of Lsm3 [13]. Packing interactions include a salt bridge (Glu-2 and Lys-9), and contacts between the imidazole rings of His-5 and His-6 of the N-terminal His_6-tag.

(ii) Helix-to-loop face interactions have been previously inferred in archaeal Lsm protein complexes. Lsmα has been found to produce fibril structures in low ionic strength conditions. The interactions are hypothesised to take place via electrostatic contributions of acidic residues in loop4 (EDGE) and exposed charged and polar residues in the flexible N-terminus α-helix.

(iii) Loop-to-loop face intermolecular stacking has been observed in the crystal structure of yeast SmF [15] utilising loop4 residues.

The aim of this research work was to produce a new level of control over the stacking of Lsmα rings. The interactions that engage the same binding interfaces of two rings are "symmetrical" in nature and self-limiting. Therefore these interactions cannot provide for regular multi-ring polymer formation i.e. extension via stacking into tubes. However, an "asymmetrical" interaction i.e. helix-to-loop face utilises two distinct binding interfaces. This type of interaction could lead to polymerisation and subsequent formation of high-aspect ratio structures. Furthermore, should such a polymer engage in both "symmetrical" and "asymmetrical" interactions it is possible to obtain large multimers through repeated combinations.

An important outcome of solution studies of polyproteins conducted by Jens Molls demonstrated that synthetic constructs of Lsm proteins exhibited multimersation events with possible stacking [16]. The existing propensity of Lsmα to form fibril-like polymers and the stacking observed in the crystal lattice (PDB ID 1I81), led to the design of engineered disulfide bridges to covalently stabilise the stacking events in solution in this work. The results confirmed this hypothesis, demonstrating that engineered Lsmα heptameric complexes successfully assembled into supramolecular structures thus creating a reversible nanotube of relatively narrow length distribution. It provided a new level of hierarchical control over the assembly of Lsmα rings which had not been observed before.

Organisation into higher-order structures of N10C/E61C was confirmed by TEM which showed the arrangement of rings into tubes. Solution based analysis of N10C by SEC was consistent with association into a dimer of rings. The mutein N10C showed a "symmetrical" interaction via helix-to-helix face covalent association. This could be attributed to the highly exposed engineered Cys residues on the flexible N-terminal region. E61C did not associate at the loop-to-loop face, perhaps due to the very short loop4 of Lsmα and the potential constraint provided by the other structural elements in the vicinity of loop4. Further polymerisation with the double Cys mutein N10C/E61C was achieved by incubation of N10C/E61C with E61C which lead to the polymerisation of the rings in a discrete step-wise manner. Thus, the Cys muteins have provided three separate types of molecular tectons – (i) a single ring with reactive Cys groups (E61C), (ii) a dimer of rings (N10C) and (iii) an extended tube form (N10C/E61C) as shown in **Figure 6.1**.

6.4 Effect of the His_6-tag on the supramolecular assembly

Lsmα was developed as a construct with an His_6-tag. Its oligomerisation was found to be dependent on the placement of the tag at either the N- or C-terminus. Placement of the His-tag at the N-terminus caused a marked change in the supramolecular organisation of Lsmα due to chelation of copper ions. The oligomerisation was found to be metal-dependent and large quantities, 40 mg per litre of culture, of highly pure and stable multimeric species of Lsmα were isolated. Such nanostructures were formed only when exposed to certain metal ions indicating high metal specificity.

The position of the His_6-tag played a role in the supramolecular assembly of the Lsmα. The β5-strand extends into the C-terminus which then protrudes laterally away from the heptamer core at the equatorial plane of the ring. Both solution studies and TEM revealed that $[LsmαH_6]_7$ differed from the properties of $[H_6Lsmα]_7$. $[LsmαH_6]_7$ exhibited a heterogeneous mixture, making it unsuitable for further studies. It could be possible to obtain metal-assisted two dimensional arrays as the His_6-tag is exposed laterally from the ring structure. It would be interesting to optimise conditions for $[LsmαH_6]_7$ to obtain a stable homogenous sample.

Placing the tag at the N-terminus, completely changed the assembly of Lsmα. $[H_6Lsmα]_7$ assembled as cage-like structures at alkaline pH. This was assessed by SEC and confirmed by TEM. The 3D model from the SAXS data corresponded to the dimensions presented by TEM. It was also shown that the assembly is bound to Cu^{2+} and further can bind to Ni^{2+} in a reversible manner. This is an important property of a molecular building block, conferring it with a metal-binding functionality.

From the standpoint of using metal to induce assembly of proteins, Huard et al. engineered ferritin protein to oligomerise in presence of the Cu^{2+}, generating a cage-like structure [17]. In terms of changing the assembling properties, TMV exhibits remarkable changes and assembling properties with the attachment of a His-tag, with change in pH and the position of the tag [18]. The modified TMV coated protein (CP) was also used as a one dimensional template for Au NPs synthesis [19]. Recently, DNA modified with His_6-tag and NTA, was assembled into arrays with addition of Ni^{2+} through formation of metal-chelate complex [20]. The incorporation of the His_6-tag into the Lsmα scaffold drives a metal-mediated assembly of the protein which exhibited different levels of supramolecular assembly with variation in the metal concentration. Thus, Lsmα was engineered to show metal-binding properties, in addition to assembling as stacks in a controlled manner.

6.5 Future perspectives and outlook

Solution and electron microscopy studies demonstrated that Lsmα rings can form stable supramolecular structures with both "closed" and "extended" conformations (**Figure 6.1**). The Lsmα proteins in this research retain their RNA binding capacity and therefore, the next step would be to study the RNA binding ability of the structures generated. This feature is advantageous when such a tecton is used as a component for a biosensor to capture nucleotides. This motif could further be modified with metal-binding sites or serve as attachment points to an inorganic surface [21]. Furthermore, within the Lsm structure, each monomer is folded such that the unstructured N- and C-terminal tails are naturally exposed. These segments can be derivitised with antimicrobial peptides or even conjugated to industrially useful enzymes [5], making Lsm a useful nanoscaffold. Lsmα derivitised with the His_6-tag could be assembled on a NTA-modified surface to obtain an array of protein.

It would be of interest to apply the metal-binding capability to the nanostructures generated and to fabricate a biosensor with RNA-binding capacity. The nanotubes can be engineered to bind metal which could lead to the formation of metal wires through a bottom-up approach, with its dimensions confined to the nanometer range. The strategies employed to engineer supramolecular structures for Lsmα can be extended to other Lsm systems including Lsm polyproteins. Peroxiredoxin protein has been recently used to capture Au NPs into one dimensional space [22]. Similar, approaches have been employed for TRAP protein [23] and SP1 protein [24], wherein the doughnut-shaped structures act as NP assembling templates. Lsm proteins offer smaller pore dimensions (0.5 – 1.5 nm) compared to other protein systems, which is likely to be advantageous for applications where small and uniform NPss are required [25; 26].

Completely synthetic versions of Lsm proteins, namely Lsm polyproteins, are being routinely generated by Protein Structure Group, Macquarie University. Lsm polyproteins feature individual Lsm proteins covalently linked via unstructured sequence extensions and allow for simultaneous expression of two Lsm proteins as heteroheptamers. The knowledge gained from the Lsmα system can be further extended to Lsm polyproteins for tailor-made properties.

As depicted in the schematic (**Figure 6.1**), four possible strategies were explored in the given time frame, out of which three gave fruitful results. It would be of interest to generate arrays and fibrils in a controlled manner, for applications in bionanotechnology. Such protein tectons can be employed to control the assembly of nanomaterials and help generate metallic, inorganic, or organic materials with precise control over their dimensions. For metallic materials, the spherical, rod-like, and planar architectures of proteins have found success in

generating predesigned NPs and nanowires [27; 28]. Other avenues where such building blocks can be potentially applied are drug delivery, biosensing and polymer-composites.

In this study, insights were gained into the controlled hierarchical assembly of Lsmα proteins and their potential use as a molecular tecton. Hence, the results obtained from Lsmα complexes in this study aid our understanding of self-assembly and contribute to the growing field of molecular building blocks.

6.6 References

1. Bittner, A. M. (2005). Biomolecular rods and tubes in nanotechnology. *Naturwissenschaften* **92**, 51-64.
2. Channon, K., Bromley, E. H. C. & Woolfson, D. N. (2008). Synthetic biology through biomolecular design and engineering. *Current Opinion in Structural Biology* **18**, 491-498.
3. Sinclair, J. C. (2012). Self-assembly: Proteins on parade. *Nature Chemistry* **4**, 346-347.
4. Medalsy, I., Klein, M., Heyman, A., Shoseyov, O., RemacleF, Levine, R. D. & Porath, D. (2010). Logic implementations using a single nanoparticle-protein hybrid. *Nature Nanotechnology* **5**, 451-457.
5. Heyman, A., Levy, I., Altman, A. & Shoseyov, O. (2007). SP1 as a novel scaffold building block for self-assembly nanofabrication of submicron enzymatic structures. *Nano Letters* **7**, 1575-1579.
6. King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E. & Gonen, S. (2014). Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108.
7. Song, W. J., Sontz, P. A., Ambroggio, X. I. & Tezcan, F. A. (2014). Metals in protein–protein interfaces. *Biophysics* **43**, 409-431.
8. Jutz, G. & Boker, A. (2010). Bionanoparticles as functional macromolecular building blocks-A new class of nanomaterials. *Polymer* **52**, 211-232.
9. Lee, S.-Y., Royston, E., Culver, J. & Harris, M. (2005). Improved metal cluster deposition on a genetically engineered tobacco mosaic virus template. *Nanotechnology* **16**, 435-441.
10. Kadri, A., Maiss, E., Amsharov, N., Bittner, A. M., Balci, S., Kern, K., Jeske, H. & Wege, C. (2011). Engineered Tobacco mosaic virus mutants with distinct physical characteristics *in planta* and enhanced metallization properties. *Virus Research* **157**, 35-46.
11. Zaric, B. L. (2014). Oligomerization of Biomacromolecules–Example of RNA Binding Sm/LSm Proteins, Oligomerization of Chemical and Biological Compounds, *Dr. Claire Lesieur (Ed.), ISBN: 978-953-51-1617-2, InTech.*
12. Wilusz, C. & Wilusz, J. (2013). Lsm proteins and Hfq: Life at the 3' end. *RNA Biology* **10**, 592-601.

13. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R., Curmi, P. M. G. & Mabbutt, B. C. (2008). Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: implications for Lsm ring organisation and recruitment. *Journal of Molecular Biology* **377**, 1357-1371.
14. Chen, C. S., Smits, C., Dodson, G. G., Shevtsov, M. B., Merlino, N., Gollnick, P. & Antson, A. A. (2011). How to change the oligomeric state of a circular protein assembly: switch from 11-subunit to 12-subunit TRAP suggests a general mechanism. *PloS One* **6**, e25296.
15. Collins, B. M. (2001). Structural studies of archaeal and yeast Sm/Lsm proteins (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
16. Molls, J. (2011). Structural characterisation of *yeast* Lsm protein complexes (PhD thesis). *Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW*.
17. Huard, D. J., Kane, K. M. & Tezcan, F. A. (2013). Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nature Chemical Biology* **9**, 169-176.
18. Bruckman, M., Soto, C., McDowell, H., Liu, J., Ratna, B., Korpany, K., Zahr, O. & Blum, A. (2011). Role of hexahistidine in directed nanoassemblies of tobacco mosaic virus coat protein. *ACS Nano* **5**, 1606-1622.
19. Liu, N., Wang, C., Zhang, W., Luo, Z., Tian, D., Zhai, N., Zhang, H., Li, Z., Jiang, X., Tang, G. & Hu, Q. (2012). Au nanocrystals grown on a better-defined one-dimensional tobacco mosaic virus coated protein template genetically modified by a hexaHis tag. *Nanotechnology* **23**, 335602-335607.
20. Mitchell, N., Ebner, A., Hinterdorfer, P., Tampé, R. & Howorka, S. (2010). Chemical tags mediate the orthogonal self-assembly of DNA duplexes into supramolecular structures. *Small* **6**, 1732-1735.
21. Heyman, A., Medalsy, I., Bet Or, O., Dgany, O., Gottlieb, M., Porath, D. & Shoseyov, O. (2009). Protein scaffold engineering towards tunable surface attachment. *Angewandte Chemie International Edition* **48**, 9290-9294.
22. Phillips, A. J., Littlejohn, J., Yewdall, N. A., Zhu, T., Valery, C., Pearce, F. G., Mitra, A. K., Radjainia, M. & Gerrard, J. A. (2014). Peroxiredoxin is a versatile self-assembling tecton for protein nanotechnology. *Biomacromolecules* **15**, 1871-1881.

23. Heddle, J. G., Fujiwara, I., Yamadaki, H., Yoshii, S., Nishio, K., Addy, C., Yamashita, I. & Tame, J. R. H. (2007). Using the ring shaped protein TRAP to capture and confine gold nanodots on a surface. *Small* **3**, 1950-1956.
24. Behrens, S., Heyman, A., Maul, R., Essig, S., Steigerwald, S., Quintilla, A., Wenzel, W., Bürck, J., Dgany, O. & Shoseyov, O. (2009). Constrained synthesis and organization of catalytically active metal nanoparticles by self assembled protein templates. *Advanced Materials* **21**, 3515-3519.
25. Peng, S., McMahon, J. M., Schatz, G. C., Gray, S. K. & Sun, Y. (2010). Reversing the size-dependence of surface plasmon resonances. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14530-14534.
26. Wang, C., Daimon, H., Onodera, T., Koda, T. & Sun, S. (2008). A general approach to the size-and shape-controlled synthesis of platinum nanoparticles and their catalytic reduction of oxygen. *Angewandte Chemie International Edition* **47**, 3588-3591.
27. Howorka, S. (2007). Creating regular arrays of nanoparticles with self-assembling protein building blocks. *Journal of Materials Chemistry* **17**, 2049-2053.
28. Gazit, E. (2007). Use of biomolecular templates for the fabrication of metal nanowires. *FEBS Journal* **274**, 317-322.