Reconstitution of two human LSm protein complexes reveals aspects of their architecture, assembly and function

Author(s):
Zaric, Bozidarka L.

Publication Date:
2005

Permanent Link:
https://doi.org/10.3929/ethz-a-004946401

Rights / License:
In Copyright - Non-Commercial Use Permitted
RECONSTITUTION OF TWO HUMAN LSM PROTEIN COMPLEXES
REVEALS ASPECTS OF THEIR ARCHITECTURE, ASSEMBLY AND
FUNCTION

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH
for the degree of
Doctor of Natural Sciences

Presented by

Bozidarka L. Zaric

Master of Biochemistry, University of Belgrade

Born January 6th, 1972
citizen of Serbia and Montenegro

accepted on the recommendation of

Prof. Dr. Fritz K. Winkler, examiner
Prof. Dr. Ulrike Kutay, co-examiner

Zürich, 2005
ACKNOWLEDGEMENT

First of all I want to thank my husband for unending support and his understanding that helped me succeed in my work. Even though we were far away from each other his love and encouragement inspired me and gave me the strength to succeed.

Two special people, Prof Dr. F.K. Winkler and Dr. Christian Kambach guided me throughout my work.

Prof. Dr. F.K. Winkler gave me the solid background, support and freedom to advance in my work. I appreciate the valuable discussions and advices from an extremely knowledgeable person such as Prof.Dr. F.K. Winkler. I want to thank him for always being accessible to his colleagues and open for discussion.

I thank Dr. Christian Kambach, my supervisor for his support and for successfully guiding my thesis from the beginning till the very end. I truly admire his enthusiasm and energy. I enjoyed our fruitful discussions that helped me throughout my work.

I also thank Prof. Ulrike Kutay, member of my thesis committee, for her lively participation in my thesis evaluation, for her valuable comments on my work and for always being open for questions.

I am gratefully thankful to Prof. Kurt Ballmer-Hoefefer for introducing me to valuable cell biology techniques, which I used during my work.

Special thanks to Prof. A. Engel’s group, for excellent collaboration in electron microscopy characterization of my proteins.

I am thankful to PD. Dr. T.Soulimane for helping me in crystallization trials.

Last but not the least I want to thank my parents for their unselfish support and understanding throughout difficult times.
CONTENTS

ABSTRACT .................................................................................................................. 1

ZUSAMMENFASSUNG ............................................................................................... 3

1. INTRODUCTION .................................................................................................... 5

1.1. DNA, RNA and the central dogma of molecular biology ................................. 5

1.2. RNA as a primordial molecule .......................................................................... 5

1.3. The RNAissance: Multiple roles for small RNAs ............................................. 6

1.3.1. A multitude of snoRNAs and rRNAs .......................................................... 6

1.3.2. tRNA ............................................................................................................. 7

1.3.3. Ribozymes .................................................................................................. 7

1.3.4. Telomerase .................................................................................................. 7

1.3.5. Small nuclear RNAs ..................................................................................... 8

1.3.6. Micro RNAs and Small interfering RNAs ................................................... 8

1.4. Structure of RNA and association with proteins ............................................. 8

1.5. The unifying frame: All RNA processing steps are coupled ........................... 9

1.6. RNA capping, and the role of the cap structure ............................................. 9

1.7. Polyadenylation of RNA .................................................................................. 10

1.8. Removal of introns and the splicing reaction .................................................. 11

1.9. Types of introns ............................................................................................... 12

1.10. Splicing .......................................................................................................... 13

1.10.1. Small Nuclear Ribonucleoproteins, snRNPs .............................................. 13

1.10.2. U6snRNP biogenesis ............................................................................... 14

1.10.3. Spliceosome assembly .............................................................................. 15

1.10.4. Role of U4/U6 specific proteins in formation of the catalytic site of the spliceosome .......................................................... 17

1.11. Export of RNAs ............................................................................................ 18

1.12. mRNA stabilization, degradation ................................................................... 20

1.13. Family of Sm-LSm proteins .......................................................................... 22

1.13.1. Sm proteins, assembly of U1, U2, U4, U5 snRNPs ..................................... 22

1.13.2. LSm proteins ............................................................................................ 27

1.13.3. Role of LSm proteins in U6snRNP assembly .......................................... 29

1.13.4. Role of LSm proteins in protecting mRNA 3' end termini from degradation .......................................................... 30

1.13.5. Role of LSm proteins in U8 snoRNP organization .................................... 30

1.13.6. Sm-like proteins as part of U7snRNP ...................................................... 30

1.13.7. LSm proteins in mRNA degradation ....................................................... 31

1.13.8. LSm proteins promote regeneration of pre-mRNA splicing activity ....... 32

1.13.9. LSm proteins in the processing of pre-tRNAs ........................................... 32

1.13.10. LSm 2-7 complex associated with snR5 ............................................... 32
1.13.11. Prp24 associates with LSm 2-8 promoting annealing of U4/U6
1.13.12. The family of LSm proteins is growing
1.13.13. The Bacterial LSm protein: Hfq

2. MATERIALS AND METHODS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Chemicals and enzymes</td>
<td>36</td>
</tr>
<tr>
<td>2.2. Plasmids</td>
<td>36</td>
</tr>
<tr>
<td>2.3. E.Coli Strain Genotypes</td>
<td>37</td>
</tr>
<tr>
<td>2.4 Media</td>
<td>38</td>
</tr>
<tr>
<td>2.5. Buffers and Solutions</td>
<td>38</td>
</tr>
<tr>
<td>2.6. cDNA sequences for human LSm proteins</td>
<td>39</td>
</tr>
<tr>
<td>2.7. PCR amplification</td>
<td>45</td>
</tr>
<tr>
<td>2.8. Restriction enzyme digestion</td>
<td>45</td>
</tr>
<tr>
<td>2.9. Dephosphorylation of the digested vector</td>
<td>46</td>
</tr>
<tr>
<td>2.10. Ligation</td>
<td>46</td>
</tr>
<tr>
<td>2.11. Preparation of competent cells</td>
<td>46</td>
</tr>
<tr>
<td>2.12. Transformation of competent cells</td>
<td>47</td>
</tr>
<tr>
<td>2.13. Quik-Change™ Site -Directed Mutagenesis</td>
<td>47</td>
</tr>
<tr>
<td>2.14. Dam_ cells</td>
<td>47</td>
</tr>
<tr>
<td>2.15. Plasmid preparation</td>
<td>48</td>
</tr>
<tr>
<td>2.16. Cloning of LSm2/3 bicistron into pQE30 expression vector</td>
<td>48</td>
</tr>
<tr>
<td>2.17. Cloning of LSm4/8 bicistron into pQE30 expression vector</td>
<td>49</td>
</tr>
<tr>
<td>2.18. Cloning of LSm5/6/7 tricistron into pQE30 expression vector</td>
<td>49</td>
</tr>
<tr>
<td>2.19. Cloning of LSm1 into pQE30 expression vector</td>
<td>50</td>
</tr>
<tr>
<td>2.20. Cloning of LSm4 into pQE30 expression vector</td>
<td>50</td>
</tr>
<tr>
<td>2.21. Cloning of LSm8 into pQE30 expression vector</td>
<td>50</td>
</tr>
<tr>
<td>2.22. Cloning of LSm6 into pQE30 expression vector</td>
<td>51</td>
</tr>
<tr>
<td>2.23. Cloning of LSm5:pQE30</td>
<td>51</td>
</tr>
<tr>
<td>2.24. Cloning of LSm5/7: pQE30</td>
<td>51</td>
</tr>
</tbody>
</table>
2.25. Cloning of LSm5/6:pQE30 .............................................................. 52
2.26. Cloning of LSm5/3:pQE30 .............................................................. 52
2.27. Cloning of LSm3/6:pQE30 .............................................................. 52
2.28. Cloning of LSm4/2:pQE30 .............................................................. 53
2.29. Cloning of LSm4/7:pQE30 .............................................................. 53
2.30. Cloning of ZZLSm4:pQE70 .............................................................. 54
2.31. Cloning of ZZLSm6:pQE70 .............................................................. 54
2.32. Cloning of ZZLSm8:pQE70 .............................................................. 54
2.33. Kinasing and annealing oligonucleotide linkers ..................... 55
2.34. Determining optimal expression parameters in E. coli .......... 55
2.35. Growing expression cultures of E. coli ......................................... 56
2.36. Harvesting and disruption of the cells ........................................ 56
2.37. Immobilized Metal Affinity Chromatography (IMAC) ................. 57
2.38. Anion exchange chromatography ................................................ 57
2.39. Cation exchange chromatography ................................................ 58
2.40. Reversed Phase Chromatography ................................................ 58
2.41. Gel filtration chromatography ....................................................... 59
2.42. Determination of protein concentration ...................................... 59
2.43. Tobacco Etch Virus (TEV) protease cleavage .............................. 60
2.44. Sodium Dodecylsulfate- Polyacrylamide Gel Electrophoresis (SDSPAGE) .................................................. 61
2.45. Coomassie- staining ................................................................. 61
2.46. In Vitro transcription of U1snRNA and U6 snRNA ....................... 61
2.47. Preparation of radioactively labeled U1 and U6 snRNA ............ 62
2.48. Microinjection in to the REFS2 cells .......................................... 66
2.49. Coupling of proteins to the fluorescent dyes ............................... 67
2.50. Electron microscopy ................................................................. 67
2.51 Crystallization trials ................................................................. 67
### 3. RESULTS AND DISCUSSION

3.1. Cloning strategy
- 3.1.1. Construction of pUC19.1
- 3.1.2. Subcloning of the LSm2/3 and LSm4/8 bicistrons into the pQE30 expression vector
- 3.1.3. Cloning of the LSm5/6/7 tricistron into the pQE30 expression vector

3.2. Protein expression and purification
- 3.2.1. Expression and purification of His6-TEV-LSm2/3
- 3.2.2. Expression and purification of His6-TEV-LSm4/8
- 3.2.3. Expression and purification of His6-TEV-LSm5/6/7
- 3.2.4. Expression and purification of His6-TEV-LSm1
- 3.2.5. Expression and purification of His6-TEV-LSm4
- 3.2.6. Expression and purification of His6-TEV-LSm5
- 3.2.7. Expression and purification of His6-TEV-LSm6
- 3.2.8. Expression and purification of His6-TEV-LSm5
- 3.2.9. Expression and purification of ZZ-His6-TEV-LSm5
- 3.2.10. Expression and purification of His6-TEV-LSm4/6
- 3.2.11. Expression and purification of His6-TEV-LSm5/3
- 3.2.12. Expression and purification of His6-TEV-LSm4/7
- 3.2.13. Expression and purification of His6-TEV-LSm4/2 and His6-TEV-LSm6/7
- 3.2.14. Expression and purification of His6-TEV-LSm5/7
- 3.2.15. Biophysical characterization of the His6-TEV-LSm2/3, His6-TEV-LSm4/8 and His6-TEV-LSm5/6/7
- 3.2.15.1. His6-TEV-LSm2/3 characterization
- 3.2.15.2. His6TEV LSm4/8 characterization
- 3.2.15.3. His6TEV LSm5/6/7 characterization
- 3.2.16. Summary of Protein Expression and Characterization Results

3.3. Sub-complex characterization under semi-denaturing conditions

3.4. In vitro reconstitution of LSm2-8 and LSm1-7 complexes
- 3.4.1. In vitro reconstitution of LSm2-8
- 3.4.2. In vitro reconstitution of LSm1-7
- 3.4.3. Electron micrographs of reconstituted LSm2-8 and LSm1-7 complexes

3.5. In vitro reconstitution of LSm2-8 in the presence of U6 snRNA

3.6. Electrophoretic mobility shift assays (EMSA)

3.7. In vivo functional assays

3.8. Crystallization trials

### 4. SUMMARY AND OUTLOOK

### 5. REFERENCES

### 6. CURRICULUM VITAE
ABSTRACT

The Sm and LSm proteins are an ancient, widespread protein family with members in all kingdoms of life. *Archaeabacteria* harbour between one or two Sm/LSm protein genes each. The Escherichia coli Hfq protein and its orthologues represent the family in several bacterial lineages. Eukaryotes genomes contain a minimum of 24 Sm/LSm genes. Sm/LSm proteins are generally mediating RNA-protein, or RNA-RNA interactions. They always act in the form of homo (prokaryotes) or hetero (eukaryotes) -multimeric complexes. Complex composition and architecture determine their intracellular distribution, RNA target, and specific function. Often, only a subset of subunits (or one) is exchanged from one complex to another, bringing all the observed changes about. Sm/LSm proteins thus can be regarded as exchangeable building blocks, with complex formation as the single most important factor for determining Sm/LSm function. As research in RNA processing progresses, ever more proteins containing the Sm/LSm motif are discovered, and new LSm complex functionalities identified. Still, very little is known about LSm complex assembly pathways, nor how the architecture of the often very similar complexes determines their specific functions.

The canonical Sm core domain binds to a conserved RNA binding motif on several small nuclear RNAs (snRNAs), and functions in the removal of non-coding sequences (introns) from eukaryotic pre-mRNA. Seven Sm-like proteins (LSm2-8) bind to the 3' end of U6 snRNA to form U6 snRNP. The LSm1-7 complex differs from LSm2-8 by only one subunit (LSm1 instead of LSm8). It has a role in cytoplasmic mRNA degradation. We chose to study the structure-function relationship in two human LSm complexes with seven subunits each, LSm1-7 and LSm2-8. The aim was to learn more about how exchange of one subunit (LSm1 with LSm8) can alter function, RNA target, and intracellular distribution of the two heptamers. A longer term goal of the project which extends beyond the time frame of this thesis is to solve the X-ray structure of the heptamers.

In order to study LSm1-7 and LSm2-8 assembly, as well as address their structure and function at the molecular level, I have reconstituted the heptameric complexes from recombinant LSm proteins or, in analogy to the canonical Sm core domain assembly pathway, from heterodi- and trimeric sub-complexes (LSm1, LSm4, LSm2/3, LSm5/6/7, and LSm4/8), characterized the sub-
complexes and heptamers biophysically, and demonstrated that the latter are functional both in vitro (RNA bandshift) and in vivo (intracellular migration). As shown by electron microscopy, isolated subcomplexes assemble into ring-like higher order structures, as do the heptamers. Crystallization trials yielded weakly diffracting crystals for LSm6 and LSm5/6/7, but not for the heptamers. In summary, the recombinant LSm complexes represent an ideal test system for the study of LSm protein biochemistry, cell biology, and structure.
Zusammenfassung

Die Sm und LSm Proteine sind eine alte, weitverbreitete Proteinfamilie mit Mitgliedern in allen Bereichen des Lebens. Archaeabakterielle Genome enthalten alle ein bis zwei Sm/LSm Gene. Das Escherichia coli Hfq Protein und seine Orthologen vertreten die Familie in einigen bakteriellen Linien. Eukaryontische Genome enthalten jeweils mindestens 24 Sm/LSm Gene.

Sm/LSm Proteine vermitteln im allgemeinen RNS-Protein oder RNS-RNS Wechselwirkungen. Sie fungieren immer in Form von homo (in Prokaryonten) oder hetero (in Eukaryonten) - multimeren Komplexen. Aufbau und Architektur der Komplexe legen deren intrazelluläre Verteilung, RNS-Substrat und spezifische Funktion fest. Von einem Komplex zu einem anderen werden häufig nur wenige Untereinheiten (oder auch nur eine) ausgetauscht. Dieser Austausch bringt dann alle beobachteten funktionellen Änderungen hervor. Sm/LSm Proteine können daher als auswechselbare Bausteine betrachtet werden, deren Komplexbildung als der im Einzelnen wichtigste Faktor für die Bestimmung der Funktion jedes Sm/LSm Proteins gelten kann. Im Zuge des Fortschritts der Forschung über RNS-Prozessierung werden immer weitere Proteine, die das Sm/LSm Sequenz-Motiv enthalten, entdeckt und neue Funktionalitäten für LSm-Komplexe definiert. Im Gegensatz dazu ist immer noch sehr wenig über die Biogenese von LSm-Komplexen bekannt, noch wie die Architektur der häufig sehr ähnlichen Komplexe ihre spezifischen Funktionen bestimmt.

Die kanonische, klassische Sm Kern-Domäne bindet an ein konserviertes Uridin-reiches Motiv auf einigen kleinen nukleären RNAs (small nuclear RNAs, snRNAs) und agiert in der Entfernung nicht kodierender Sequenzen (Introns) aus eukaryontischer Vorläufer-mRNS. Sieben Sm ähnliche Proteine (LSm2-8) binden an das 3'-Ende der U6 snRNA, um das U6 snRNP zu bilden. Der LSm1-7 Komplex unterscheidet sich von LSm2-8 durch nur eine Untereinheit (LSm1 anstelle von LSm8). Er spielt eine Rolle in der cytoplasmatischen mRNA-Degradierung.

Wir beschlossen, die Struktur-Funktions-Wechselbeziehung von humanem LSm1-7 und LSm2-8 zu studieren. Das Ziel war, mehr darüber zu lernen, wie der Austausch einer einzigen
Untereinheit (LSml durch LSm8) Funktion, RNS-Bindung und intrazelluläre Verteilung der zwei heptameren Komplexe ändern kann. Das langfristige Ziel des Projektes, das über dem Zeitrahmen dieser Doktorarbeit hinausgeht, ist es, die Röntgenstruktur der Komplexe zu lösen.

Mit dem Ziel des Studiums der LSml-7 und LSm2-8 Biogenese, sowie um deren Funktion und Architektur auf molekularer Ebene besser zu verstehen, habe ich die beiden heptameren Komplexe aus rekombinanten LSm Proteinen, oder, in Anlehnung an die Biogenese der kanonischen Sm Kern-Domäne, ausgehend von heterodi- und trimeren Unterkomplexen (LSml, LSm4, LSm2/3, LSm5/6/7, und LSm4/8) in vitro rekonstituiert, sowohl die Unterkomplexe als auch die Heptamere biophysikalisch charakterisiert, und gezeigt, dass die Letzteren sowohl in vitro (per RNA bandshift) als auch in vivo (intrazelluläre Verteilung) funktionell sind. Durch Elektronenmikroskopie konnte gezeigt werden, dass isolierte Unterkomplexe sich zu ringförmigen, höher geordneten Strukturen assemblieren, genauso wie die Heptamere. Kristallisationsversuche führten zu schwach beugenden Kristallen von LSm6 und LSm5/6/7, jedoch nicht der Heptamere. Zusammenfassend stellen die rekombinanten LSm Protein-Komplexe ein ideales Testsystem zum Studium von LSm Protein Biochemie, Zellbiologie, und Struktur dar.
1. INTRODUCTION

1.1. DNA, RNA and the central dogma of molecular biology

In 1957, Crick postulated the central dogma of molecular biology, which stated that genetic information only flows in one direction, from DNA to proteins via an intermediate called messenger ribonucleic acid (mRNA) [1,2,3]. Originally RNA, in the form of mRNA, tRNA and rRNA was thought to have relatively passive roles (in information transfer and structure maintenance). Today we know that RNA performs a remarkable range of functions in the living cell. In addition to its role in information transfer, it has a role in the control of gene expression, chromosome –end maintenance, housekeeping activities, sorting of proteins in the cell, metazoan development [3].

1.2. RNA as a primordial molecule

For a long time, only proteins were considered to have enzymatic activities. In the early 1980s Tom Cech and Sidney Altman showed that RNA molecules can catalyse a chemical reaction, namely the cleavage of a bond between the sugar ribose and a phosphate group. RNAs with catalytic activity are called ribozymes. The discovery of ribozymes led to the hypothesis that RNA could have been the original molecule of life on earth about four billion years ago; a biopolymer with the ability to self-replicate and that could both store information and catalyse chemical reactions. Until ribozymes were discovered, it had not been possible to fully explain the origin of life and evolution on the basis of DNA or proteins. DNA can only be formed and have an effect with the aid of enzymes, which are proteins. Proteins are the product of the genetic information that begins with DNA; how, then, could either have originated? DNA cannot exist without proteins, and proteins cannot exist without DNA. RNA would have been self-sufficient as the original molecule of life [4].
1.3. The RNAissance: Multiple roles for small RNAs

Discovery of the unexpectedly wide variety of functions carried out by RNA was accompanied by the identification of a multitude of further types of small, non-coding RNAs (small nuclear RNA, small nucleolar RNAs, small interfering RNAs, micro RNAs) highlighting the versatility of RNA as a biochemical tool for the cell [2].

![Figure 1.1. The RNA family](Reprinted from reference 2)

1.3.1. A multitude of snoRNAs and rRNAs

rRNAs represent essential structural, and, as more recently established, catalytic elements of the ribosome. In the nucleolus of eukaryotic cells, more than 100 tandemly repeated units of rRNA genes are transcribed by RNA polymerase I into long precursor transcripts [5]. Following transcription, numerous residues of ribosomal RNA are modified by either sugar methylation or pseudouridylation [6,7]. The resulting pre-rRNA is subsequently cleaved to form mature 18S, 5.8S, 28S rRNAs and assembled with 5S rRNA and approximately 80 proteins to form the large and small ribosomal subunits prior to their export to the cytoplasm. SnoRNAs participate in both the modification and cleavage events that occur during ribosome biogenesis. The snoRNA can be categorized according to certain sequence motifs [5]. A unique feature of these snoRNAs is that they possess extensive regions of complementarities to the rRNA. Most are encoded within the introns of protein coding genes.
1.3.2. tRNA

A second group of small RNAs are the tRNAs, which are essential in translation. For each codon in the DNA there is a tRNA which becomes loaded with the corresponding amino acid. Interestingly, tRNA genes contain non-coding sequences, introns, in organisms from all kingdoms of life [8].

1.3.3. Ribozymes

Small RNAs with catalytic activity are called ribozymes. Ribozymes are remarkably active for their chemical simplicity, with rate enhancements ranging from $10^4$-10$^7$-fold. Three main mechanisms of catalysis have emerged: general acid-base catalysis, electrophilic catalysis and conformational control/transition-state stabilization [4].

RNaseP is a ribonucleoprotein enzyme which is required for the processing of the 5' termini of tRNA in the pathway of biosynthesis of tRNA from precursor tRNAs. RNase P catalyzes endonucleolytic 5' maturation of pre-tRNA, in all three kingdoms of life (Archaea, Bacteria and Eukarya) as well as chloroplasts and mitochondria. This enzyme consists of a number of protein subunits (in humans ten subunits) and one RNA subunit, which has a catalytic role [9, 158].

1.3.4. Telomerase

Telomeres are the natural ends of eukaryotic chromosomes. Their length is maintained using a ribonucleoprotein called telomerase. Telomerase was recognized to be a reverse transcriptase when it became clear that its mechanism is to copy an RNA template into DNA. Telomerase requires both protein and RNA components for its activity [10,160].
1.3.5. Small nuclear RNAs

Small nuclear RNAs are components of the macromolecular machinery (spliceosome) that has a role in the removal of introns. They are termed U snRNAs (stands for uridyl rich small RNA). U1, U2, U4, U5, U7, U11, and U12 are synthesized in the nucleus by RNA polymerase II. After that they are transported to the cytoplasm where association with the U snRNP proteins occurs, followed by re-import into the nucleus [1, 67].

1.3.6. Micro RNAs and Small interfering RNAs

Micro RNAs are non-coding RNAs of 22-24 nucleotides in length. They downregulate gene expression by attaching themselves to mRNA, thereby preventing them from being translated into protein. Another type of non-coding RNA is the small interfering RNA. These small RNAs mediate RNA interference by attaching to mRNA and labelling it for destruction [2, 4].

1.4. Structure of RNA and association with proteins

DNA and RNA have very similar covalent structures, the only difference being the change from a 2'-deoxyribose sugar to a ribose sugar and from a methyl group in thymine to a hydrogen in uracil. RNA clearly has much wider biological activities and adopts a wider range of structures. DNA double helices preferentially assume the B-form structure in solution and RNA double helices are found in the A-form. The RNA A-form double helix has a narrow and deep major groove, which prevents proteins to recognize RNA in a manner analogous to the way they recognize DNA. An RNA molecule can locally adopt several types of secondary structure. Bulges occur when one or more nucleotides on one strand are not base paired, while all nucleotides on the opposite strand are. The exposed unpaired nucleotides represent targets for site-specific recognition of RNA by proteins and may form tertiary contacts between distant regions of the RNA secondary structure. Internal loops form when two double helices are separated on each strand by several non-Watson-Crick-paired nucleotides. Hairpins are structures where the phosphodiester backbone folds back to form a double-helical tract leaving
unpaired nucleotides. Tertiary structures form when singly-stranded loops form Watson-Crick base pairs with remote sequences and some of these structures are called pseudoknots [12, 21, 159].

Eukaryotic mRNAs are almost always associated with RNA-binding proteins that control each and every aspect of RNA metabolism. RNA-binding proteins generally have a modular structure and contain RNA-binding domains of 70–150 amino acids that mediate RNA recognition. Three major classes of eukaryotic RNA-binding protein domains are known: the RNA-recognition motif (RRM), the double stranded RNA binding domain (dsRBD) and the K-homology (KH) domain. Proteins containing Puf domains interact with protein cofactors to form complexes that bind mRNAs and control diverse developmental events. They have been implicated in regulating deadenylation, translation, and mRNA stability [13, 14, 15].

1.5. The unifying frame: All RNA processing steps are coupled

Eukaryotic gene expression is a complex, stepwise process that begins with transcription [16, 17]. Mature mRNAs are produced in the cell nucleus from primary transcripts of coding genes (pre-mRNAs) by a series of processing events which include capping, splicing, and 3’ end polyadenylation. Mature mRNAs are transported to the cytoplasm. Transport of mRNA, its stability and intracellular localization are all subject to regulation. All processes mentioned above are mediated by numerous RNA-binding proteins. All modification steps are coupled and influence each other. RNA polymerase II is a key molecular coordinator of these processing events, and phosphorylation of it has regulatory role [18, 19, 20, 22, 25, 156, 162, 163].

1.6. RNA capping, and the role of the cap structure

The first pre-mRNA processing event occurs co-transcriptionally after about 20-30 nucleotides have been synthesized. A three-step reaction adds a cap structure to the 5’ end of all mRNAs and produces the 7'-methyl guanosine cap structure [22].
This initial cap structure is recognized in the nucleus by a cap binding complex (CBC), composed of two proteins [23]. The cap structure plays an important role in activating mRNA processing and enhances 3' end formation [24]. Upon export through the nuclear pore complex, the nuclear cap binding proteins are replaced by the cytoplasmic translation initiation factor, eIF-4E [25]. The cap has a role in control of cell growth [23] and protection of mRNA from exonucleolytic degradation [26]. The cap structure can promote degradation of the mRNA in the presence of some stimuli [27]. There are reports that indicate association of poly (A) binding protein \textit{in vitro} with the cap structure, preventing the access of the decapping machinery [28].

\textbf{1.7. Polyadenylation of RNA}

All protein-encoding mRNAs (with the exception of histone mRNAs) contain a uniform 3' end consisting of around 200 adenosine residues. The formation of this poly (A) tail is directed by sequences present on the pre-mRNA and is carried out by the polyadenylation machinery. The 160 kDa protein complex has a role in cleavage and polyadenylation. Poly (A) polymerase (PAP) is required for poly (A) addition [18].
1.8. Removal of introns and the splicing reaction

In 1977, a number of research groups discovered that the genes of higher organisms are often made up of a sequence of coding and non-coding base sequences. Analysis of the gene coding for the β-chain of the blood pigment haemoglobin in the mouse revealed that it contained three coding sequences (called exons) separated by two non-coding sequences (called introns). During transcription, all parts of the gene are copied to form a strand of pre-mRNA. Introns are removed and the exons stitched together so that the now continuous exons can be translated to produce a protein. This splicing of the pre-mRNA is a multistage process, carried out by a complex macromolecular machinery known as the spliceosome. The spliceosome is among the most complex macromolecular machineries in the cell [29, 30, 147, 161].
Splicing of precursors to mRNAs occurs in two steps, both involving a single transesterification reaction. The first step generates a 2'-5' bond at the branch site upstream of the 3' splice site and a free 3' hydroxyl group on the 5' exon. In the second step, an attack of the 3' hydroxyl on the phosphodiester bond at the 3' splice site displaces the lariat intron resulting in the joining of the two exons [30]. Assembly and function of the spliceosome requires approximately 300 polypeptides and five snRNAs, not considering gene-specific RNA-binding factors [30, 31, 33]. There are two distinct types of spliceosome in most cells. The major class U2-type spliceosome is universal in eukaryotes, whereas the minor class or U12-type spliceosome is not present in some organisms. The evolutionary relation between these two spliceosomes is uncertain.

1.9. Types of introns

The pre-mRNA contains conserved elements at its intron/exon boundaries that determine the proper sites for the splicing reaction (Figure 5). The 5' splice site contains a conserved consensus sequence, which is AG/GURAGU (R = purine, / denotes the exon/intron boundary). The branch site lies between 100 and 18 bases upstream of the 3' splice site and has the consensus: CURAY for vertebrates (A branching nucleotide, Y is pyrimidine). In higher eukaryotes, a polypyrinic tract variable in length is often located between the branch site and the 3' splice site. The 3' splice site has the consensus: YAG/R for mammals. This class of introns is spliced by U2
spliceosome. The U12 type introns have different consensus sequences and are spliced by the U12 spliceosome [31, 148, 149, 151, 153]. Members of the U12 group of introns have strongly conserved 5' and 3' splice site sequences, either /AUAUCCUUU and YAC/ or /GUAUC-CUUU and YAG/, respectively (where "/" represents the splice junction). These introns also have a conserved branch site with the consensus TCCTTAAC (A branching nucleotide), located between 10 and 20 nucleotides upstream of the 3' splice site.

The number of known U12 introns is still very small. U12 type introns are present in many vertebrates, nematodes, insect, and plant species.

Figure 1.5. Splice site consensus sequences
Comparison of splice site consensus sequences for human U2 dependent and U12 dependent introns. The most conserved regions, 5' splice site (5' SS), branch point (BP), 3' splice site (3' SS), are shown with their consensus sequences (R = purine, Y = pyrimidine). The polypyrimidine tract often present in metazoan U2 dependent introns is indicated as (Py). (Reprinted from reference 33)

1.10. Splicing

1.10.1. Small Nuclear Ribonucleoproteins, snRNPs

Small ribonucleoproteins (RNPs) are tight complexes of one or more proteins with a short RNA molecule (usually 60-300 nucleotides). RNPs inhabit nuclear and cytoplasmatic compartments of the eukaryotic cell [11, 33]. Those that reside in the nucleus, the small nuclear RNPs (snRNPs) can themselves be divided into two families. There are snRNPs of the nucleoplasm, whose function lies in preparing messenger RNA for export into the cytoplasm. A different set of snRNPs, called snoRNPs, reside in the nucleolus [34]. There are about 200 distinct kinds of snoRNPs with abundances between $10^4$ (for snoRNPs) to over $10^6$ copies per cell (for snRNPs in major spliceosomes). They generally play a role in gene expression. One exception is the
telomerase snRNP, essential for genome maintenance, which is present only in a few copies per cell. Spliceosomes are formed around the premRNA substrate by the successive assembly of five small nuclear ribonucleoprotein-particles (snRNPs): U1, U2, U4, U5, and U6. These particles are composed each of a small nuclear RNA (snRNA), seven Sm core proteins common to all snRNPs (except for the U6 snRNP, which contains a related set of seven proteins, the Sm-like proteins) and several snRNP-specific proteins. The snRNPs play a central role in the process of splicing. They are responsible for the recognition of splice sites and definition of exon/intron boundaries. In addition, the snRNAs build the framework of the spliceosome by interacting with each other, and with the pre-mRNA. These interactions are partially mediated through base pairing and are dynamic so that the spliceosome complex changes during the process of splicing.

1.10.2. U snRNP biogenesis

Subsequent to transcription by RNA polymerase II and capping, pre-U1 snRNA assembles with several factors including cap-binding proteins (CBP), a phosphorylated adaptor for RNA export (PHAX), Crm1, and Ran-GTP, which all together mediate export of U1 snRNA to the cytoplasm. After export, dephosphorylation of PHAX and hydrolysis of Ran-GTP leads to dissociation of this complex. In the next step, Sm proteins interact with the U snRNAs to form the snRNP Sm core. This step is facilitated by the SMN complex. Nuclear re-import is mediated by snurportin-1 (SPN1), which binds to the snRNAs m3G cap structure. After import, these factors dissociate. The U1 specific proteins are imported independently into the nucleus, where assembly into mature U1 snRNP occurs [11]. This is pathway shared between U1, U2, U4 and U5 snRNPs.
1.10.3. Spliceosome assembly

Assembly of a spliceosome for excision of an intron requires recognition of sequences at the 5' splice site as well as the branch site and nearby 3' splice site. Recognition of the 5' consensus sequence is based on RNA sequence complementarities: U1 snRNA binds to the 5' end of the intron. For a typical intron, U1 snRNP association with the 5' sequence is critical for splicing in vitro, but there are introns that are very efficiently processed in the absence of U1 snRNP. In this case, the 5' sequence is recognized by U6 snRNA. There are reports which show that the U1 snRNA recognizes the 5' splice site in a preassembled penta-snRNP complex [35]. The primary function of U1 snRNP is to promote the association of a U2 snRNP complex with the branch

Figure 1.6: The U snRNP biogenesis pathway (Reprinted from reference 11)
region. Early snRNP/pre-mRNA complexes are preferentially committed to splicing as compared to free RNA and thus are called commitment complexes (CCs). The process of U2 snRNP association is ATP-dependent. Four proteins, Prp5, Prp9 (SF3a60), Prp11 (SFa66), and Prp21 (SFa120) are critical for recognition [36,37,157]. Subsequent to the binding of U2 snRNP complex, a tri-snRNP complex containing U4/U6 snRNP and U5 snRNP associates in an ATP-dependent manner to form complex A2-1. It is likely that the U1 snRNA/pre-mRNA duplex dissociates at this stage. The 5' splice site sequence is probably paired on the intron side to U6 snRNA and on the exon side to U5 snRNA [152,154].

The transition between complex A2-1 and A1 requires destabilization of at least U4/U6 di-snRNA. This transition again requires ATP and possibly the putative RNA helicase Brr2. As only three snRNAs U2, U6 and U5 are associated with the spliceosome at the moment of catalysis, and as U5 snRNA pairing with exon sequences is not essential, the catalytic site is either created by U6 snRNA by U2 snRNA or both. The action of Prp16, Prp17, Prp18 and Slu7 is required for the transition to the second step in splicing. The catalytic site for the second step is created by either U6 snRNA, U2 snRNA, or associated proteins. The spliced exon RNA is released from the spliceosome through the action of Prp22 and perhaps other factors. The reannealing of released U4 and U6 snRNP and association with U5 forms the U4/U6 U5 tri-snRNP complex which is then ready to reassemble on another commitment complex.

A 61 kDa protein (Prp31 in yeast) is associated with U4/U6 snRNP and promotes via the 102 kDa, U5 snRNP specific protein formation of the tri-snRNP particle. The 61 kDa protein is encoded by a gene (PRPF31) that has been shown to be linked to autosomal dominant retinitis pigmentosa [38]. In the absence of Prp31, U4/U6disnRNP accumulates in Cajal bodies. U5 snRNP remains in nucleocytoplasmic speckles [39].

The classical view of spliceosome assembly has been challenged by Stevens et al [40]. This group isolated from yeast a penta-snRNP complex which when supplied with soluble components, does splice pre-mRNA.
1.10.4. Role of U4/U6 specific proteins in formation of the catalytic site of the spliceosome

During activation of the spliceosome, the U4/U6 snRNA duplex is dissociated, releasing U6 for base-pairing with U2 snRNA. Proteins that directly bind U4/U6 snRNP are involved in this change. In humans, there are five U4/U6 snRNP specific proteins in addition to the Sm core on U4 and LSm2-8 on U6 snRNA.

Upon spliceosome activation, stem II of U6 (paired with U4 snRNA) folds back on itself to form a new intramolecular stem–loop, and the stem I region of U6 snRNA basepairs with U2 snRNA to form part of the catalytic centre (Figure 8). Part of U6 snRNA interacts with the 5' splice site of the intron. In contrast, U4 snRNA is dissociated from the spliceosome or remains only loosely
attached to it. The establishment of the U6/5' splice site interaction is thought to be a prerequisite for U4/U6 dissociation and U5 snRNA is likely to be involved in this process by formation of the U5/5' splice site early in assembly. Prior to the first step in the reaction, U5 snRNA interacts with the 3' extremity of the upstream exon. After this first step it interacts with the two exon extremities which have to be ligated [150, 103, 41, 155].

![Diagram of spliceosome activation](image)

**Figure 1.8.** Proposed mechanism for disruption of U4/U6 snRNA during activation of spliceosome. (Reprinted from reference 41)

### 1.11. Export of RNAs

All macromolecular transport between nucleus and cytoplasm must occur via nuclear pore complexes (NPC) [42]. Most of nucleocytoplasmatic transport is mediated by factors that belong to a family of nuclear transport receptors termed karyopherins or importins/exportins. Different members of this protein family bind to the distinct cargo molecules or to the adapter proteins which bind cargo, and interact with specific nucleoporins. Nucleoporins are components of nuclear pore complexes.
Crm1-dependent export of U snRNA

In higher eukaryotes, the U snRNA components of the snRNPs are synthesized in the nucleus but, with the exception of U6 snRNA, they are assembled in the cytoplasm and re-imported into the nucleus. Transport of U snRNAs is mediated by Crm1, RanGTP and PHAX [43].

Nuclear export of rRNAs

In a complex and highly ordered pathway, 28S, 5.8S, and 5S rRNA are assembled and together with 50 ribosomal proteins form the 60S preribosomal subunit. The 18S rRNA is assembled with 33 proteins into 40S preribosomal subunit [44]. These subunits are then separately transported into the cytoplasm. A protein called Nmd3p associates with the 60S ribosomal subunit in the nucleus through direct interaction of Nmd3p with proteins of the 60S ribosomal subunit [45]. The export of the 60S preribosomal subunit involves Crm1.

Nuclear export of tRNAs

Exportin-t binds in the nucleus to mature tRNAs and exports them into the cytoplasm [46].

Nuclear export of mRNAs

The export of most mRNAs is independent of the RanGTP systems and karyopherins [47, 48]. A key factor for mRNA export is a heterodimer formed by Tap and a small cofactor termed Nxt or p15. In humans there are 6 nuclear export factors, of which Tap is the best characterized [49]. In the past years, evidence of a coupling between mRNA splicing and export has emerged and appears mediated by the exon junction complex [50, 51, 52, 53].
1.12. mRNA stabilization, degradation

Regulation of mRNA decay rates is an important control point in determining the abundance of cellular transcripts. For successful gene expression one of the important steps is mRNA degradation. Some mRNA has half-lives that are 100 times shorter than cellular generation times and some mRNA have half-lives spanning several cell cycles [27]. The poly (A) tail is important in stabilization of mRNA. It interacts with the poly (A) binding protein (PABP), which makes direct contact with a specific region of the translation-initiation factor (eIF4E). Translation initiation factor (eIF4) interacts with the cap binding proteins. In this way, a ternary (PABP-translation initiation, -cap binding protein, poly (A) tail) complex is formed which circularizes mRNA in vitro, promoting translation and stabilization of mRNA, by preventing access of deadenylation and decapping enzymes to their targets [27].
Several sequence elements can regulate the mRNA turnover rate, either by its promotion (destabilizer elements) or by its inhibition (stabilizer elements). Important elements are A+U-rich elements (ARE), located in the 3' untranslated regions (UTR) of mRNAs [54, 55].

At least four different ways of mRNA degradation have recently been reported in eukaryotic cells [56].

In most cases, degradation of the transcript begins with the shortening of the poly (A) tail at the 3' end of mRNA. After shortening of the poly (A) tail follows the removal of the 5' cap structure (decapping), thereby exposing the transcript to digestion by a 5' to 3' exonuclease. Transcripts can be degraded in the 3'-5' direction after deadenylation. This process is catalyzed by the exosome [57].

One mRNA degradation pathway is the nonsense mediated decay (NMD), which provides strongest evidence for a link between translation and turnover. Substrates for NMD are transcripts from mutated genes containing either non-sense or frame shift mutations, or transcripts from normal genes that contain errors as result of splicing or transcription. In both cases, the result is an mRNA that contains premature stop codons [58]. Normal stop codons are located at the last exon. mRNA containing termination codon located 50-55 nucleotides upstream of the last exon-exon junction are subject to rapid mRNA decay [59]. To explain the mechanism of NMD, a nuclear-scanning model was proposed in which there is a nuclear ribosome-like apparatus [60]. The exact cellular localization of NMD is controversial, as both nuclear and cytoplasmic NMD have been reported [61]. After mRNA splicing on exon –exon junctions the exon junction complex (EJC, composed from five proteins) remains as a marker [59]. This marker recruits NMD factors during the first round of translation, and triggers degradation of this mRNA. Three interacting proteins: Upf1, Upf2, Upf3 are essential for NMD [62, 63]. Nonsense mediated mRNA decay degrades mRNA both in the 5'-3' and 3'-5' directions.
1.13. Family of Sm-LSm proteins

1.13.1. Sm proteins, assembly of U1, U2, U4, U5 snRNPs

The common proteins for U1, U2, U4 and U5 snRNPs are named Sm proteins due to their recognition by anti-Sm autoantibodies (isolated from the serum of patients with autoimmune diseases) [64, 65, 66, 67]. To date eight proteins: B', B, D1-D3, E, F and G have been identified in human cells. All of the Sm core proteins are encoded by separate genes [68], with the exception of B and B'. B' is the product of alternative splicing of B protein and differs only in 11 amino acids at the C-terminus [69]. In neural tissues, SmN replaces SmB and SmB' [70]. Two sequence motifs, named Sm1 and Sm2, are found in all known Sm proteins, what is reason that they are named Sm proteins [71]. The N terminal Sm1 motif is composed from 32 amino acids. The Sm2 motif, located closer to the C terminus, is shorter spanning only 14 amino acids [68]. Sm motif 1 and Sm motif 2 are separated by a linker of variable length. The alignment of the sequences of human Sm proteins reveals a striking conservation of the two motifs. Majority of the Sm proteins have amino or carboxy-terminal extensions.

Figure 1.10. Primary and secondary structure of Sm proteins (Reprinted from reference 72)

Amino acid sequence alignment of the human Sm (D1, D2, D3, B'/B', E, F, and G) proteins with secondary structure elements. Wavy line, helix; arrows, β strands. The β strands within the Sm1 and Sm2 motifs are colored blue and yellow, respectively. The β strands and interconnecting loops are numbered consecutively from the N terminus. The conserved Sm1 and Sm2 motifs are indicated and the conserved residues within these motifs are highlighted in blue (hydrophobic), grey (hydrophobic, less well conserved), orange (neutral polar), red (basic) and green (acidic). (Reprinted from reference 72)
Recently, a couple of structures of this protein family, including those from *Archaeabacteria* and *E.coli* have been solved, showing that the fold is highly conserved. It is defined by an N-terminal helix, followed by a five-stranded anti-parallel β sheet. Strands β1, β2, and β3 are part of the Sm1 motif, whereas the Sm2 motif forms strands β4 and β5. The five-stranded β sheet is strongly bent in the middle and the conserved hydrophobic residues form a hydrophobic core [72].

![Figure 1.11. Proposed Higher-Order Assembly of the Human Core snRNP Proteins.](image)

The Sm proteins bind to the Sm site of U snRNAs [73]. The Sm site consensus sequence (PuAU4-6GPu) has a central, uridine rich tract and flanking purines. *In vitro* studies reveal that the single-stranded U rich region and the 5' adenosine of the Sm site play critical roles in Sm protein assembly. The uridine bases and the 2' hydroxyl groups collectively provide binding determinants [74]. In the absence of U snRNA, the seven Sm proteins form three stable subcomplexes (D3B, or D3B', D1D2, and EFG). These sub complexes then form a heptameric ring around the snRNA Sm site, and as such the complex is termed the Sm core. SnRNP core assembly is an ordered pathway that involves formation of a sub-core particle followed by formation of the full Sm core, which promotes cap hypermethylation and pre-snRNP import [74, 76, 75]. The Sm fold is necessary and sufficient for the formation of specific inter-subunit interactions.

Biochemical results indicate that there is one copy of each Sm protein in the snRNP core domain and therefore support the heptameric ring model of the snRNP core domain [77].
None of the single Sm proteins has a known RNA recognition motif, so another type of interactions with RNA must be involved [68]. Very little is known about the manner in which Sm proteins are interacting with the Sm site RNA. Crosslinking studies indicate that Sm motif 1 is responsible for interactions with RNA, and Sm motif 2 for protein–protein interactions [74, 78]. Within the yeast commitment complex, it has been demonstrated that SmB, SmD1 and SmD3 make direct contact with pre-mRNA, close to the 5' splice site [79]. Basic residues of human and yeast SmB, SmD1 and SmD3 are reported to be responsible for import of the Sm core particle [80, 81].

In vitro, the snRNP core domain can be assembled from purified components [166]. Assembly of the spliceosomal class of snRNPs in vivo it is an active process that is mediated by several factors, including the product of the SMN gene that is involved in SMA (spinal muscular atrophy disease). Spinal muscular atrophy is an autosomal recessive disorder correlated with loss of motor neurons, as a result of a mutation on the SMN gene [82, 83]. The SMN protein is ubiquitously expressed in all tissues of metazoan organisms reflecting the fact that it provides a fundamental activity required by all cells. The SMN protein is predominantly cytosolic but it is also found in the nucleus, namely in a few spherical nuclear domains that overlap with the so-called Cajal bodies (where snRNPs and snoRNPs are localized). These spherical domains have been called Gemini of Cajal bodies (Gems). Proteins associated with the SMN protein are called Gemins. The SMN complex interacts in vitro with Sm and LSm proteins which contain symmetrically methylated RG repeats [86, 87, 168]. Symmetrically methylated RG repeats of SmD1, SmD3 and LSm4 are generated by action of the so-called methylosome [88]. The SMN complex also interacts with p53 and profilin, but the functional significance of these interactions is not known [84, 85].

![Figure 1.12. The SMN complex.](image)

Schematic representation of the SMN complex. Gemins 2, 3, 5, and 7 bind directly to SMN, while Gemins4 and 6 are associated through direct interaction with Gemins3 and 7, respectively. (Reprinted from reference 90)
For generation of import competent Sm core particle, cap hypermethylation and assembly of the Sm proteins around the Sm site are necessary.

Figure 1.13. Model of TGS1 and SMN interactions in cytoplasmic snRNP biogenesis. (Reprinted from reference 89)

The SMN complex binds to the human hypermethylase which suggests that SMN may have a role in formation of the snRNA m^3G cap structure. SMN and TGS1 (human hypermethylase) recognize a similar interaction determinant on the heptameric Sm-core domain. It has been proposed that after binding of SMN to the Sm core proteins, SMN promotes engagement of TGS1 to the m^7G-capped snRNP particle. According to that model, SMN dissociates from the C terminal part of the B/B' Sm proteins followed by association of TGS1 and Sm core. This step allows formation of the m^3G-cap. The association of Snurportin 1 with the m^3G-cap can promote release of TGS1 and generate import-competent snRNP [89, 90]. According to these data, the SMN complex interacts with protein components of U snRNPs, but there are reports [91] on sequence-specific interactions between U1 snRNA and the SMN complex. The nuclear localization signal of U snRNP is composed of the U snRNAs 2,2,7 tri-methylguanosine cap and Sm core domain. Snurportin 1 binds to the m^3G cap but not to the Sm core. Snurportin1 has an N-terminal importin beta binding domain and a carboxy terminal m^3G-cap binding domain [94, 95]. The Importin beta binding domain allows for snRNPs cargo to be imported in a Ran
independent fashion [96]. After import of snRNPs into the nucleus, Snurportin1 dissociates from its cargo and is exported back into the cytoplasm using Crm1, a receptor for leucine-rich nuclear export signals [97].

It is conceivable that the SMN complex plays a direct role in snRNP import because one factor that interacts with the Sm core in snRNP import remains unidentified. The SMN complex not only mediates snRNP core assembly but is an integral complex component during the entire snRNP core biogenesis in the cytoplasm [92, 93]. It is not excluded that SMN is actually the long-sought Sm core nuclear localization signal receptor [94].

Figure 1.14. Schematic model of the role of SMN in snRNP core biogenesis in the cytoplasm. (Reprinted from reference 92)
1.13.2. LSm proteins

Sm and Sm-like proteins are found in all kingdoms of life. Sensitive database searches have revealed the presence of sequences related to the Sm proteins even in Archaea. Because Archaea have been proposed to be related to the ancestor of the eukaryotic nuclear genome, this fact suggests that an LSm protein gene was present in the last common ancestor. Archaeabacteria harbour between one and two Sm/LSm genes each. Eukaryotic genomes have more than 20 Sm/LSm genes each, corresponding to the LSm and Sm proteins which are components of Sm and LSm complexes. Database searches in the yeast genome, revealed 16 Sm motif containing proteins. Multiple sequence alignment of the 16 Sm motif containing yeast proteins, the seven human LSm proteins, the seven human Sm canonical proteins and the five Sm-like proteins from archaeabacteria, illustrates the relationship between the various sequences [71]. Some Sm-like proteins were found to interact weakly with some Sm proteins, most probably via non-specific Sm domain interactions [98], but some of the LSm proteins interact with Sm proteins as part of U7 snRNP. In yeast there are nine LSm proteins, in humans more than eight. Each of the human LSm proteins has one orthologue in yeast. Yeast LSm2p-LSm7p share sequence identity with human LSm2-LSm7 ranging form 41-62%. LSm9p appears to be present only in yeast. Yeast LSm8p aligns best with human LSm8 (26% identity).

Figure 1.15. Phylogenetic tree of yeast, human and archaeal Sm and Sm-like proteins. The tree was built from a multiple sequence alignment of the various Sm domains (reference 168). The seven subfamilies containing related subtypes of both the Sm and Sm-like (LSm) proteins are indicated (7). Because sequence similarity in subfamily 1 is weak, it is indicated with a dashed lined. Human proteins are labeled with an 'h' prefix. Archaeal proteins are named with the first four letters of the species name (Arch, Archaeoglobus Fulgidus, Meth, Methanobacterium thermoautotrophicum, Pyro, Pyrococcus hirikoshii) Reprinted from reference 71.
In addition, LSm proteins are highly conserved throughout all eukaryotic kingdoms, as the homologues in insect, nematode and plant database share between 50 and 75% identity with their human counterparts. Each of the LSm proteins in humans can clearly be best aligned with one of the canonical Sm proteins. However, their sequence identities are not high enough to allow the conclusion that LSm proteins undergo the same protein-protein interactions like Sm proteins [99].

![Structural Alignment of Human Sm/LSm Proteins](image)

**Figure 1.16.** Structural Alignment of Human Sm/LSm proteins

Similar to canonical Sm proteins, the LSm proteins are recognized by antibodies from patients suffering from systemic lupus erythematosus (SLE) [100].

All canonical Sm proteins are essential for vegetative growth of yeast. LSm proteins have variable effects after depletion in yeast: LSm2, LSm3, LSm4 and LSm8 genes are essential for vegetative growth, while LSm5, LSm6 and LSm7 are dispensable, like LSm1 and LSm9. They are required for efficient growth at 30 degrees and higher temperatures [71]. In mice embryos, LSm4-null zygotes survived to the blastocyst stage, but died shortly after [101].
1.13.3. Role of LSm proteins in U6snRNP assembly

The LSm2-8 complex was isolated from Hela nuclear extract in an RNA free form. Electron micrographs revealed a doughnut-shaped heterooligomer, similar to the Sm core snRNPs [99, 102,100]. LSm proteins have a high affinity for single-stranded oligo-U, but they do not recognize the canonical Sm binding site. In yeast and humans, LSm2-8 forms a heteroheptameric ring around the 3' end of U6 snRNA, consisting of a U rich tract. The Sm core RNP is extremely salt stable; however, LSm-U6 snRNA dissociates at salt concentrations higher than 0.5M, or in the presence of competitor RNA, suggesting that the LSm-U6 complex is less stable [99].

U6 snRNA has no conserved Sm site and does not associate with Sm proteins. Its biogenesis pathway differs in many respects from the U1, U2, U4 and U5snRNP pathways; it is transcribed by RNA polymerase III and capped by γ-monomethyltriphosphate. The 3' end of pre-U6 snRNA is elongated during maturation and subsequently trimmed leaving in most organisms a 2'-3' cyclic phosphate. The enzymes involved in this process are specific for U6 snRNA, and U6 snRNA does not leave the nucleus [103]. Mature U6 snRNA shows nucleoplasmic localization. U6 snRNA has eight sites of 2'-O-ribose methylation and three pseudouridylation sites. These modifications are guided by a snoRNA, and they occur in the nucleolus with the consequence that U6 snRNA transiently assumes nucleolar localization [104]. The first protein that binds to nascent, newly transcribed U6 snRNA, is the La protein (Lhp1p in yeast), by way of the UUUoh sequence at the 3' terminus. The La protein stabilizes U6 snRNA. It serves as a molecular chaperone, not only for U6 snRNA but for other RNA polymerase III transcripts as well [106, 107]. Experimental evidence suggests that U6 snRNA is present in the cytoplasmic compartment of mouse fibroblast cells [108]. This result suggests that the LSm2-8 complex may act as a nuclear localization signal, but the cytoplasmic localization of the U6 snRNP is highly questionable.

The actual function of the LSm 2-8 complex associated with U6 snRNA appears to be connected to U6 snRNP assembly and function. Mutants with decreased levels of LSm2-8 show splicing defects correlating with a reduced level of U6 snRNA. Moreover, overexpression of U6 snRNA can partially compensate for the loss of LSm proteins. How the LSm2-8 complex affects U6 snRNP remains unclear. One possibility is that LSm proteins facilitate conformational rearrangements during the splicing cycle, U4/U6 annealing and formation of U4/U6/U5 tri-snRNP [109].
Some LSm mutants show abnormalities in U5 snRNA, indicating that LSm proteins have a role in U5 processing, although it can be an indirect effect due to alterations in the U6 snRNA level [110].

1.13.4. Role of LSm proteins in protecting mRNA 3' end termini from degradation

LSm proteins have additional roles apart from splicing. Yeast strains which lack LSm1-7p fail to grow at higher temperatures, and accumulate mRNA shortened at the 3' end by 20-30 nucleotides. These 3' shortened species arise by degradation of the 3' end of the mRNA, following deadenylation. The mechanism by which LSm1-7 protects mRNA 3' termini is unclear. The simplest model proposes that LSm1-7 complex binds to the mRNA and sterically inhibits endo and exo-nucleases. Nuclear LSm2-8 binds to the U6 snRNA 3' end, suggesting, that LSm2-8 protects the 3' end of U6 snRNA from degradation [110].

1.13.5. Role of LSm proteins in U8 snoRNP organization

U8 snoRNP is required for processing of 5.8S and 28S rRNAs, which together with the 5S rRNA build up the large ribosomal subunit. In Xenopus extract, LSm2, 3, 4, 6, 7, and 8 are bound to U8 snoRNA on the conserved third stem-loop sequences [111,112,113].

1.13.6. Sm-like proteins as part of U7 snRNP

Maturation of the non-polyadenylated histone mRNAs 3' ends occurs by endonucleolytic cleavage mediated by U7 snRNP [114, 115, 116, 117]. U7 snRNA contains a non-canonical Sm site. Purified U7 snRNP lacks D1 and D2 proteins but has LSm10 (14kDa) and LSm 11 (50kDa) instead [118,119].

The long N-terminal part of LSm11 has a direct role in histone RNA processing. The C terminal part of LSm11 contains the bi-partite Sm motif. The linker between Sm motif 1 and Sm motif 2 can depending on species be up to > 100aa long. The Sm motif of LSm11 is required for interaction with LSm10 and interacts with the U7 snRNA-specific Sm site. LSm10 has a more compact Sm motif. U7 snRNP core assembly is an SMN- and ATP-mediated process [118]. In contrast to the spliceosomal snRNPs, whose base-pairing interactions with the substrate include
the nucleotides involved in the transesterification reaction of splicing, base pairing by the U7 snRNP acts at a distance to the processing site.

1.13.7. LSm proteins in mRNA degradation

Yeast two hybrid assays reveal multiple interactions between the eight LSm proteins, suggesting the existence of more than one LSm protein complex. Each human LSm protein is capable of interacting with multiple other LSm proteins and splicing factors, like prp24, prp4, and SmD1 [120]. Coprecipitation experiments demonstrated that LSm1p (LSmXp, is the nomenclature for yeast LSm proteins) together with LSm2p-LSm7p forms a new seven-subunit complex [121, 122].

The LSm complex LSm1-7 plays a role in mRNA degradation [121, 123], and LSm2-8 has a role in the stabilization of U6 snRNP. These two protein complexes thus have very different functions. LSm1p mutants accumulate full length capped transcripts, but mutations on LSm1p do not stabilize mRNA containing premature stop codons, suggesting that the LSm1-7 complex is not involved in NMD [122]. In yeast two-hybrid assays, LSm proteins were found to interact with proteins involved in NMD [124].

The function of the LSm1-7 complex is most likely to interact with the mRNA substrate and accelerate decapping, perhaps by promoting rearrangements in mRNP organization that lead to decapping. Decapping is mediated by a decapping enzyme that is consisting of Dcp1a, Dcp1b, and the catalytic subunit Dcp2. Dcp1 enhances the activity of Dcp2. The crystal structure of yeast Dcp1 has been determined and reveals that the enzyme has conserved hydrophobic residues located on the surface that are critical for cap recognition [125].

The LSm1-7 proteins are localized in discrete cytoplasmic foci. The foci contain key decapping factors required for 5'-3' mRNA degradation. Coexpression of LSm proteins increases the number of foci. The cytoplasmic foci contain LSm1-7 proteins [126].

LSm1 and LSm8 are closely related to each other, and to the SmB protein. The 33 C terminal amino acids of LSm1 are necessary but not sufficient for proper cellular localization of hLSm1 [126].

A protein termed GW182 has multiple glycine-tryptophane repeats and an RNA recognition motif (RRM). The function of the GW182 protein is not known, but this protein has been found
to co-localize with hLSm4 and hDep1, enriched in the foci described above [127]. Finally it has been demonstrated [128] that the foci are actual degradation centres, where mRNA degradation occurs. This suggests that the cytoplasm of cells is more organized than previously thought.

1.13.8. LSm proteins promote regeneration of pre-mRNA splicing activity

U4 and U6 snRNA contain extensive sequence complementarities and interact to form the U4/U6di-snRNP, which associates with U5 to yield the U4/U6/U5tri-snRNP. LSm proteins dissociate from U6 snRNA within the spliceosome at the time of activation [129]. Following the splicing reaction, snRNPs must be recycled. Little is known about this process. In *Saccharomyces cerevisiae*, LSm6 and LSm7 are nonessential genes. Their deletion causes a splicing defect, namely they fail to support splicing during sequential incubation. This is probably due to inability to reassemble U4/U6 and U4/U6U5 snRNP complexes. Probably U6 snRNA that is released from dissociating spliceosomes may be degraded in the absence of the LSm complex. The LSm 2-8 complex seems to function as a chaperone complex to support multiple rearrangements of U6-containing complexes, including U4/U6 annealing/stabilization.

1.13.9. LSm proteins in the processing of pre-tRNAs

It has been reported that depletion of LSm proteins in yeast leads to strong accumulation of unspliced tRNA species. The absence of LSm proteins most probably alters the pattern of processing intermediate [130,131]. LSm proteins are similarly required for normal processing and stability of ribosomal RNAs.

1.13.10. LSm 2-7 complex associated with snR5

An LSm2-7 hexameric complex is found to be associated with snR5 in *Saccharomyces cerevisiae*. This RNA is a member of the box H/ACA class of snoRNAs that function in pseudouridylation of rRNA. Previously, it was also reported that LSm2-7 complex associates with pre RNase P RNA [71]. The SnR5 associated LSm complex may be a hexamer, but it is not
excluded that one LSm protein in this complex is present in more than one copy, or an as yet unidentified yeast protein associates with LSm2-7, thereby closing the heptameric ring [132]. LSm2-8 interacts with external 3' sequences on U6 snRNA. The LSm1-7 complex interacts with 3' UTR mRNA but there could well be secondary structure elements between the LSm1-7 binding site and the mRNA 3' end. U8snoRNA and snR5 bind LSm proteins via internal RNA sequences, suggesting that LSm rings can assemble onto the RNA. LSm proteins have a role in the biogenesis and function of at least a subset of nucleolar RNAs.

One possibility is that LSm proteins assist base pairing of snoRNAs with their rRNA targets, to conduct pseudouridylation and ribose modifications.

1.13.11. Prp24 associates with LSm 2-8 promoting annealing of U4/U6

In order for U4 and U6 to be reutilized for subsequent rounds of splicing, a mechanism must exist for their association. In yeast, the Prp24 protein has an important role in this process. It has been shown that recombinant Prp24 accelerates the formation of base-paired U4/U6 of in vitro transcribed U4 and U6 snRNAs. The stimulatory effect of Prp24 is greater using protein-containing U4 and U6 snRNPs as a substrate. Recent work from several laboratories have provided evidence that the U6 LSm proteins contribute to this difference. Prp24 and the LSm2-8 complex act cooperatively to promote association of U4 and U6 snRNA. The human orthologue of Prp24 is SART3 (p110nrp) [133, 134, 135]. Prp24 has 4 RNA recognition motifs (RRMs), and one NLS. The C-terminal domain is required for Prp24-LSm protein interaction, and for efficient U4/U6 annealing. A cooperative model for LSm2-8 and Prp24-mediated U4/U6 annealing does not exclude the possibility that each of these factors can promote U4/U6 association without the other. Prp24 initially binds to free U6 snRNP and subsequently recruits U4 snRNP.
1.13.12. The family of LSm proteins is growing

Based on a comprehensive bioinformatics approach, divergent Sm/LSm family members have been found as domains of larger proteins which are involved in neurodegenerative diseases, in the regulation of the cell cycle, and ones that have methyl transferase activities. However, the function of Sm domains in these proteins is not known. [136].

1.13.13. The Bacterial LSm protein: Hfq

Hfq was identified as an *E. coli* host factor required for the replication of the Qβ phage RNA [137].

After disruption of the gene for the Hfq protein, many pleiotropic effects were observed [138,139]. The ability of Hfq to associate with various RNAs, binding U-rich tracts on its targets, and the identification of an Sm1-like motif in the N-terminal part of the protein suggested that Hfq is a bacterial homologue of the Sm/LSm proteins. Hfq forms a hexameric ring, in which each subunit has an N-terminal helix and a five-stranded antiparallel beta sheet. Bacterial Hfq and human Sm/LSm proteins display nearly the same structure. Hfq binds single-stranded RNA, similar to archaeabacterial Sm proteins. Nucleotides are bound in the centre of the pore [140].

The eubacterial Sm-like Hfq protein promotes bimolecular RNA-RNA interactions in general. The discovery of the Hfq protein gives further credibility to the assumption that the Sm protein
family has evolved from a common ancestor. Hfq protein offered to the cell a simple device for establishing bimolecular RNA interactions, which may have had a role in versatile RNA-based gene regulatory systems. It is likely that the presence of specialized Sm proteins that carry extra subdomains was a prerequisite for the establishment of primitive eukaryotic cell [141, 142].


Archaeoglobus fulgidus has two Sm-like proteins (AFSm1 and AFSm2). They associate with RNase P-RNA, suggesting a role in tRNA processing. The crystal structure of AF-Sm1 reveals a seven-membered ring, with the RNA interacting inside the central cavity [143]. The A. fulgidus Sm2 protein forms homoheptamers as well, with sevenfold symmetry when it is bound to oligo-U RNA [144]. The Sm protein from Pyrobaculum aerophilum shows metal-dependent assembly properties and the formation of 14-mers [145]. The crystal structure of the LSm protein from Methanobacterium thermoautotrophicum has revealed the formation of a homoheptamers whose function presently is not known [146].
2. MATERIALS AND METHODS

2.1 Chemicals and enzymes

All chemicals used in this work were obtained from Biorad, Roche, Fluka, Sigma and Merck if not specified otherwise. Enzymes used in this work were obtained by MBI Fermentas, and New England Biolabs. Dyes for fluorescent labelling were from Molecular Probes. Crystallization solutions were obtained from Hampton Research, Sigma, and Emerald Biostructures.

2.2. Plasmids

pUC19 is a commonly used plasmid cloning vector in E. coli. The molecule is a small double-stranded circle, 2686 base pairs in length, and has a high copy number. pUC19 carries a 54 base-pair multiple cloning site polylinker that contains unique sites for 13 different hexanucleotide-specific restriction endonucleases. pUC19 is isolated from Escherichia coli ER2272 (dam⁺ dcm⁺ EcoK M⁻) by a standard plasmid purification procedure and generally is used for cloning, has Col El origin of replication, high copy number (100+) and carries an Ampicillin resistance gene.

pQE30 Carries the coding sequence for an N-terminal His6- tag, T5 promoter, bla, β-Lactamase, Ampicillin resistance gene.

This vector has been modified to carry a recognition sequence for the tobacco etch virus (TEV) protease following the His6- tag.

pQE702Z.3 Represents a modified pQE702Z vector (obtained from Dr. U. Kutay, Biochemistry ETH Zürich). This vector has been modified to carry two N-terminal Z-tags (major epitope of S. aureus protein A), an N-terminal His6- tag, and a recognition sequence for the tobacco etch virus (TEV) protease following the His6- tag, T5 promoter; bla, β-Lactamase, Ampicillin resistance gene.
2.3. E. Coli Strain Genotypes

BLR(pREP4) F-ompT [lon] hsdSb (rb-mb-; an E. coli B strain) recABL21(DE3) F-ompT [lon] hsdSb (rb-mb-; an E. coli B strain) with DE3, a prophage carrying the T7 RNA polymerase gene

M15(pREP4) NalS, StrS, RifS, Thi-, Lac-, Ara+, Gal+, Mtl-, F-, RecA+, Uvr+ , Lon+ (an E. coli K12 strain)

SG13009(pREP4) NalS, StrS, RifS, Thi-, Lac-, Ara+, Gal+, Mtl-, F-, RecA+, Uvr+, Lon+ (an E. Coli K12 strain)
XL1 Blue  
*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F'proAB lacIq ZDM15 Tn10 (TetO)]*

### 2.4 Media

SOC medium: Tryptone 20 g, Yeast extract 5 g, NaCl 0.5 g, 1 M Glucose solution 400 μl
Millipore Water 930 ml

LB- Medium: Tryptone 10 g, Yeast extract 5 g, Millipore Water 950 ml

### 2.5. Buffers and Solutions

**IMACA:** 20 mM Hepes- Na (pH 7.5), 1 M NaCl, 5 mM β- Mercaptoethanol (ME)
10 mM Imidazole- HCl (pH 7.5)

**IMACB:** 20 mM Hepes- Na (pH 7.5), 1 M NaCl, 5 mM β- ME 0.5 M Imidazole- HCl (pH 7.5 or 6.5)

Lysis buffer: IMAC A supplemented with protease inhibitor cocktail tablets (Roche, Basel). 50 ml of IMAC A buffer was supplemented with 1 tablet.

**MonoQ- A1000:** 20 mM HEPES- NaOH (pH 7.5) 5 mM DTT, 150 mM NaCl

**MonoQ- B1000:** 20 mM HEPES- NaOH (pH 7.5) 5 mM DTT, 1 M NaCl

**SDS- dye 6x (for 10 ml):** 4x Tris- HCl (pH 6.8) 7 ml Glycerol, SDS 1 mg DTT 0.93 mg

4 % High-TEMED stacking gel: [169]
Acrylamide 30 % 10 ml, 1.5 M Tris- HCl (pH 6.8) 15 ml, Ammonium Persulfate 10 % 300 μl
TEMED 150 μl, Millipore Water 35 ml

12 % High-TEMED running gel: [169]
Acrylamide 30 % 40 ml, 1.5 M Tris- HCl (pH 8.8) 25 ml, Ammonium Persulfate 10 % 333 μl
TEMED 333 μl, Millipore Water 35 ml
Laemmli Buffer 10x:
Sodium Dodecyl sulfate 20g, Glycine 288g, Tris base 60g, Millipore Water to 2 liters

Fairbanks A stain solution: [170]
Coomassie stain 0.05% (w/v), Isopropanol 25% (v/v), Acetic acid 10% (v/v)

Fairbanks B stain solution: [170]
Coomassie 0.005% (w/v) Isopropanol 10% (v/v) Acetic acid 10% (v/v)

Fairbanks D destain solution: [170]
Acetic acid 10% (v/v)

2.6. cDNA sequences for human LSm proteins

LSm 2 nucleotide sequence

1 atgcttttcattctttttaaagtctccttagtgccaaaggtgatgtgttcgatgagaactaaag
61 aatgacctgagcatctgtggacctccatctgtggatcatatatctcaactcagaacta
121 actgacatcagtcgtcagatacctgagttacctgagtcagctgccagcagatgaggtcgcacacag
181 ttcattcggcctcagtggtccgatacgtgcagctgccagcagatgaggtcgcacacag
241 tgctacaggatcgggcaaggaagaggacccagcagccacagaacag

LSm2 protein sequence

MLFYSFFKSLVGDKVDVVVELKNDLSRICGTLSQYDDSDISVTDPKYPKMLSVKNC
FIRGSSVRYVQLPLADEVDTQLQQDAARKEALQQKQ

LSm3 nucleotide sequence

1 atgcggcagaagctgagaccaacacaactaccaacactgttagaggagcccctgctcattt
61 atcaggctcaagcattggttagaaattgaaatgacgcatctgacgaagatcgagctcag
121 ggcagattacatgcttatgataacatttaaatatgtcctgtgaggagtgggaagacgtaaacta
181 gtgactactataaattgattagatatatatatcgaacgaaacgg
241 aatattcctagtgaagagagatcggggagatcgctgttctggggtccaccactgagaacgacag
301 gctggtctga
LSm3 protein sequence

MADDVDQQQTNTVVEEPLDLIRSLDERIYVKMRNDRELRLGRLHAYDQHLNMILGDVEETVTTIEIDEETYEEIYKSTKRNIPLFMLVGRGDGVVLVAPPLRVG

LSm4 nucleotide sequence

1  atgcctccct tgctcactgct gaagacggct cagaatcacc ccatgtttgt ggagctgaaa
61  aatggggaga cgtacaatgg acacctggtg agctgcgaca actggatgaa cattaacctg
121  cgagaagtcg tetgcaegtc cagggaacgg gacaagttct gcggagtgac cgagtgctac
181  atccgccgca gcacaccataa gccacctgaa gccgctgttgc ggcggatgcc cgagtgctac
241  gaggagttgg tcggcagaggg ccgcggcgc gcggagttgg gcggagttgg gcggagttgg gcggagttgg
301  gcgcggccga ggcgcggcgc gcgcggccga ggcgcggccga ggcgcggccga ggcgcggccga ggcgcggccga
361  gcgcggccga ggcgcggccga ggcgcggccga ggcgcggccga ggcgcggccga ggcgcggccga ggcgcggccga

LSm4 protein sequence:

MLPLSLLKTAQNHPPMLVELKNGETYNGHVLVSCDNWMINLREVICTSRDGDKFWRMPECYIRGSTIKYLRRIDMVKEEVVKMGPRGRGLQQKQPQKGRGMGGAGVFGGRGRGGIPGTPGRGQPEKKPGPQAGKQ

LSm5 nucleotide sequence

1  atgcggcgcga acgcctactac caacccgtcg cagctgctga cgttagagct tgtggacaaa
61  tgtataggat caagaattca catcgtgatg aagagtgata aggaaattgt tggtactctt
121  cttaggattg atgcattttg caatatggta ctggaagatg tcactgagtt tgaaatcaca
181  ccagaagcga gaaggattac taaattagat cagattttgc taaatggaaa taatataaca
241  atgcctggc tcggcagaggg ccgcggcgc gcggagttgg gcggagttgg gcggagttgg gcggagttgg

LSm5 protein sequence:

MAANATTNPSQLPLELVDOKCIPSHVMKSDEIVGTLGDFDFFMNVDVTEEHTEIPERGRITKDQIILGNNITMLVPGGEGPEV
LSm6 nucleotide sequence:
1 atgagttc ggaagcaaac ccctagtgac ttcttaaagc aaatcatcgg acgaccagtt
61 gtggtaaat taattctgg agtggattat cgaggggtcc tggcttgcct ggatggctac
121 atgaatatag ccctggagca gcacagagaa tatgtaaatg gacaactgaa gaataagtat
181 gggagtctat ttatccgagg aaacaaatgtg ttgtatctca gtacacagaa gagacgggatg
241 tga

LSm 6 protein sequence:
MSLRKQTPSDFLKQIIGRPVVKLNSGVDYRGVLACLDGYMNIALEQTEEYVNGQLKN
KYGDAFIRGNVLYISTQKRRM

LSm7 nucleotide sequence:
1 atggcggata aggagaagaa gaaaaaggag agcatcttgg acttgctcaaa gtacatcgtac
61 aagagctatcc gggtaaggt ccagggagac ggcgaagccga gtggatcct gaagggcttc
121 gacccacttc tcaaccttg gctgaaeggc acctaggagc acctggagacgac
181 cagttacaagc ttaacgggagg cttgggacttg tgggtgagcc gggcaagcttc
241 tgggtatgaa ttcttgcatc gaacggcagag ccacccctcc catcccagcag
301 caggaegctt ag

LSm7 protein sequence
MADKEKKKESILDLSKYIDKTIRVKFQGGREASGILKGFDPILLVLVDGTIEYMRDPPD
QYKLTIEDTRQLGLVRCRGTSDVVLICPQDGMEAIPNPFIQQQA

LSm8 nucleotide sequence
1 atgaegctcg cttgggagaa ctacatcatac ggaactggtga ccgttattac atcagatggg
61 agaatgattg tgggaacact gaaaggtttt gaccagacca ttaattttcg tttggatg
121 agccatgaac gagtattcag ctcttcacag gggtagaac aagtggtact aggattatac
181 atgtaagag gtgacaaactg tcagtgctgc ggaagaatcg atgaagaac agattctgcg
241 cttggattgg ggaatattcg ctaacccct ccaccccttc tattgcttg tagacaactg a
LSm8 protein sequence

MTSALENYINRTVAVITSDGRMLKVTLGFDQTIINLILDESHERVFSSSQQGEQVVLGLY1
VRGDNVAVIGEIDEETDSALDLGNIRAEPNLNSV4H

LSm1 nucleotide sequence

1   atgaactata tgcctggcac gcgccagctc atcgaggacattgacaaaaa gcacttggtt
61  ctgcctegag atggaaggac acttatagc ttttaagaa gcattgatca atttgcaac
121 ttagagctac atcagactgt ggaggtgatt catgtgggca aaaaatacgg tgaattcct
181 cgagggattt tgtggtcag aggagaaat gtgtctctac taggagaat agacttggaa
241 aaggagagtg acacacccct ccagcaagta tccattgaag aaattctaga agaacaaggg
301 gtggaacage agacagcttggaagagcagaa ggtgaagtc tccagccct gaaggacag
361 gtctctgcaga gatacatttt cgtgagttt aa

LSm1 protein sequence

MNYMPGTASLIEDIDKKHLVLLRDGRTLIGFLRSIDQFANLVLHQTVERIHVGGKKGDIP
RGIFVVRGENVVLGEIDLE

Sequence of primers used for amplification of LSm proteins

LSm2 5' primer
GGGGAGCTCTGATCAATGCTCTTCTATTCTTTTTTCAAGTCCC

LSm2 3' primer
CCCAAGCTTAGATCTATCGATCATTACTGTTTCTGCTGCAGGGCTTCCTTCC

LSm3 5' primer
GTTGCTGGTCTACGTCGTCGCCATATGTGATCAGAGCTCCCC

LSm3 3' primer
CCCAAGCTTAGATCTATCGATCATTAGCCAACTCTCAGTGGAGGGGCAACC

LSm4 5' primer
GGGGAGCTCTGATCAATGCTCTTCTATTCTTCTAGACGGGCTTCCTTCC
LSm4 3' primer
CCCAAGCTTAGATCTATCGATCATTACTTGGCCTGTCTGCCAGGC

LSm5 5' primer
GGGGAGCTCTGATCAATGGCGGCTAACGCTACTACCAACCCG

LSm5 3' primer
CCCAAGCTTAGATCTATCGATCATTACTTGGCCTGTCTGCCAGGC

LSm6 5' primer
GGGGAGCTCTGATCACATATGAGTCTTCGGAAGCAACCCC

LSm6 3' primer
CCCAAGCTTAGATCTATCGATCATTACACTTCAGGTCCTTCTCCAGG

LSm7 5' primer
GGGGAGCTCTGATCACATATGCGGCTAACGCTACTACCAACCCG

 SacI  BclI  Ndel

LSm7 3' primer
CCCAAGCTTAGATCTATCGATCATTACACTTCAGGTCCTTCTCCAGG

 HindII  BglII  ClaI

LSm8 5' primer
GGGGAGCTCTGATCACATATGACGTCCGCTTTGGAGAACTAC

 SacI  BclI  Ndel

LSm8 3' primer
CCCAAGCTTAGATCTATCGATCATTAGTAGTGTGCTACAGAAATTTAAAGGTGTTCTGC

 BglII  ClaI

LSm1 5' primer
GGGGAGCTCTGATCACATATGACTATGACGTCCGCTTTGGACCCGCC

 SacI  BclI  Ndel
LSm1 3’primer
CCCAAGCTTAGTCTATCGATCATTAAGTAGTATCTGCTGAGG

Bgl2  Cla1

Intergenic linker sequence (for vectors which carry more than one gene):

5’ GGATCC TCTAGA aataattttgtaaaggggatat CATATG GTCGAC ATCGAT AAGCTT 3’
(BamH1)  (Xba1)  RBS (SD)  (Nde1)  (SalI)  (Cla1)  (HindIII)

Oligonucleotide for removal internal Nde1 site on position 185 of pUC19 vector

CTGAGAGTGCACCATCTGCGGTGTGAAATA  (upper oligonucleotide for removal Nde1 185 position)
TATTTTCACACCACAGTGGTGCACTCTCAG  (lower oligonucleotide for removal Nde1 on 185 position)

Silent mutagenesis oligos

GATTTATATATCTCTCTCTGATGTTTCTTCATC  (LSm3 NdeI removal at position 240 lower)
GATGAAGAAACATACGAAGAGATATATAAATC  (LSm3 NdeI removal at position 240 upper)

GCCTGTGCCGGATTGATCCACCTCCGCCCAG (LSm4 BamH1 removal at position 401 lwr.)
CGGGGCGAGGTGCGATCCCGGGGCACAGGC (LSm4 BamHI removal at position 401 upr)

CTTCATCAGTGATTCTCTGATCCTCATAAC (LSm5 EcoRI removal at position 82 lwr)
GTATAGGATCAAGAATCCACATCGTGATGAAG (LSm5 EcoRI removal at position 82upr)
2.7. PCR amplification

LSm2, 3 and 5, 6, 7, and 8 were subcloned from a human lymphoma U937 cDNA library (Stratagene) into a modified pUC19 vector. LSm1 and LSm4 were subcloned from EST clones IMAGp998P2110673Q2 and IMAGp958A041800Q2, respectively. EST clones were obtained from the genetic resources centre, Berlin (RZPD, http://www.rzpd.de)

In a reaction volume of 100µl, 2ng of cDNA containing target sequence to be amplified were incubated on ice containing 1x Vent buffer (20mM Tris-Cl pH8.8, 10mM KCl, 10mM (NH₄)₂ SO₄, 2mM MgSO₄, 0.1% triton X-100) with 0.2mM dNTP and 0.2µM primers). Vent DNA polymerase (2 units, New England Biolabs) was added after 1 min at 95°C. Samples were heated in thermocycler (Perkin Elmer). The samples were incubated 95°C min and then amplification was carried out automatically according to the following scheme: 30 cycles of 30seconds at 95°C (denaturing), 90 seconds at the temperature 5-10 °C below the lowest melting temperature of the primer (annealing), and 2 minutes at 72°C for extension. The final elongation step was achieved by an incubation of 5 minutes at 72°C, followed by incubation at 4°C. Products were analyzed by 2% agarose gel. The PCR products were extracted using QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

The DNA fragment was excised from the gel, and 3 volumes of the buffer QG was added, and incubated with gel slice for 10 minutes at 50°C. After that one gel volume of the isopropanol was added to the sample and mixed. Subsequently DNA was bind to the QIA spin column, and washed by centrifugation with buffer PE. After washing step DNA was eluted from the column by adding 50µl water and centrifugation in the table top centrifuge for 1 min.

2.8. Restriction enzyme digestion

Restriction digestions were carried out in volumes of 100µl in appropriate restriction buffer for the enzymes. The digestion mix contained 10-20 units of restriction enzyme, and was incubated at 37°C for 2-4 hours. For analytical purposes, 10µl of reaction mixture was added to 2µl of 6 x DNA loading buffer, and samples were loaded onto 2% Agarose gels which contain ethidium bromide (0.5%) final concentration, and results were observed under UV transiluminator. For
preparative purposes, digestion mixtures were cut from agarose gel and purified using PCR purification kit (Qiagen).

2.9. Dephosphorylation of the digested vector

The digested vector was treated with 2 miliunits /μl of calf intestinal phosphatase (alkaline) for 2 hours at 37°C and purified using Qiagen Enzyme clean up kit according to the manufacturer protocol.

2.10. Ligation

For cohesive ends ligation, approximately 200ng of vector and 1 to 3-fold insert were mixed in the volume of 10μl ligation buffer (50mM Tris-Cl pH7.8, 10mM MgCl₂, 10mM DTT, 25μg/μl BSA) containing 40units of T4 DNA ligase (New England Biolabs) and incubated overnight at 16°C. Reaction mixture was centrifuged briefly and then used for transformation of competent cells.

2.11. Preparation of competent cells

Competent cells were prepared according to the protocol described in reference [171]. Single colony was transferred into 25 ml of LB medium in 250 ml flasks, and culture was incubated for 6-8 hours at 37°C. This starter culture was used for inoculation of three 1-liter flasks, which were grown at 18°C-22°C over night. When OD of one of the cultures reaches 0.6, it was transferred to an ice-water bath. Cells were harvested by centrifugation at 3000rpm for 15 minutes. Supernatant was discarded and cells were resuspended in 80ml of ice-cold Inoue buffer (55mM MnCl₂,15mM CaCl₂, 250mM KCl, 10mM PIPES pH 6.7) followed by centrifugation at 3525rpm in Eppendorf Centrifuge 5810. Supernatant was discarded and cells were re suspended in 20 ml of ice-cold Inoue buffer in which 1.5 ml of DMSO was added. Aliquots of competent cells were distributed into sterile eppendorf tubes. Tubes were stored at -70°C till they were used.
2.12. Transformation of competent cells

A 100µl aliquot of competent cells was thawed whilst standing on ice and the plasmid DNA (0.1 ng, or 10µl ligation reaction product was added to it. The cell suspension was gently mixed and incubated on ice for 40 minutes. The cells were heat-shocked at 42°C for 60 seconds and 0.5 ml of the SOC medium immediately was added. Cells were transferred into 37°C heating block. 300µl of the transformed cells were then plated onto LB agar plates containing appropriate antibiotics.

2.13. Quik-Change™ Site-Directed Mutagenesis

A Quik-Change™ site-directed mutagenesis kit (Stratagene) was used for introducing point mutations, exchanging or introducing single amino acids. PfuTurbo DNA polymerase replicates both plasmid strands with high fidelity and without displacing the mutant oligonucleotide primers. The basic procedure utilizes a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by PfuTurbo DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with Dpn I. The Dpn I endonuclease (target sequence: 5'-Gm6ATC-3') is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. DNA isolated from almost all E. coli strains is dam methylated and therefore susceptible to Dpn I digestion. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue supercompetent cells which repair nick.

2.14. Dam_cells

Methylation of DNA by the host strain can have effect on DNA cleavage by some restriction enzymes. Most strains of E. coli contain 3 DNA methylases that methylate distinct nucleotides in specific DNA sequences. When DNA is isolated from strains expressing one of these methylases, especially Dam or Dcm methylase because their sites are much more common than EcoK1 sites, some or all sites for a restriction enzyme may be resistant to cleavage. This occurs because DNA
is protected from cleavage when a particular base in the recognition site of the restriction enzyme is methylated. For example, plasmid DNA isolated from Dam\(^+\) *E. coli* is completely resistant to cleavage by Mbo I, which cleaves GATC sites. Not all restriction enzymes are sensitive to methylation. For example, BamH I still cleaves GGATCC sites of plasmid DNA from Dam\(^+\) *E. coli*. In some cases it is necessary to use a Dam- and/or Dcm-negative *E. coli* strain, for instance GM48, JM110 or SCS110 cells. SCS and JM110 competent cells are deficient in Dam and Dcm methylases, and DNA propagated in these cells can be digested by methylation-sensitive enzymes, like BclI. These SCS110 cells were used during the sub-cloning of His\(_6\)LSm2/3 for restriction digestion with BclI.

### 2.15. Plasmid preparation

Overnight cultures of *E. coli* (100 ml of 2xTY media) were harvested for 10 minutes at 4000 rpm in 50 ml Falcon tubes. The pelleted cells were resuspended in 5 ml of resuspension buffer. After resuspension cells were lysed in Qiagen buffer P2 (25 mM Tris-Cl pH8.0, 50 mM glucose, 10 mM EDTA) and after adding Qiagen buffer P3 (0.2 M NaOH, 1% SDS) were incubated on ice for 5 minutes. Sample was centrifuged for 30 minutes at 18 krpm in Sorvall superspeed centrifuge at 4°C. After centrifugation of the samples the supernatant was filtered, and DNA was bound to the Qiagen Midi prep 100 columns by applying soluble fraction. DNA bound to the column was eluted by adding Qiagen buffer QF, and DNA was precipitated using Isopropanol followed by centrifugation at 15 krpm for 30 minutes at 4°C. The pellet was washed with 1 ml of 70% ethanol, air dried, and resuspended by addition of 44\(\mu\)l of a 3M Sodium acetate solution (pH5.2) and 1ml of cold (-20°C) Ethanol. After incubation on ice for 10 minutes, the DNA was centrifuged 30 minutes at 4°C. Pelleted plasmid was washed with 70% ethanol and air dried for 10-15 minutes. The pellet was dissolved in 101\(\mu\)l Millipore water, and concentration of DNA was determined.

### 2.16. Cloning of LSm2/3 bicistron into pQE30 expression vector

For the cloning of several LSm proteins, we used a human U937 lymphoma cDNA library from Stratagene. LSm3 and LSm2 cDNAs were amplified from the cDNA library using oligonucleotides which include a Sacl site on the 5’ end, and a HindIII site on the 3’end. They were subcloned into
pUC19. The correctness of the cloned cDNAs was verified by sequencing. LSm 3 contains an internal Ndel site, which was removed by site directed mutagenesis. LSm3 cDNA was inserted into pUC19.1 using Ndel and HindIII restriction sites. SCS 110 cells were transformed with LSm2 (Ndel site removed):pUC19. After preparation of miniprep DNA which was not methylated on the GATC sequence, Bcl1 and Bgl2 restriction enzymes were used for subcloning into LSm3:pUC19.1, and afterwards cut with BamH1. The LSm2 open reading frame was joined to this construct by making a joint with BamHI and Bgl2. Correct orientation of inserted cDNAs was verified by restriction analysis. After verification of the clone, EcoRI/HindIII digestion was performed to transfer LSm2-LSm3:pUC19.1 as a cassette into pQE30, the T5 expression vector. Expression constructs bear a His6-tag at the N-terminus of the first cistron, followed by a tobacco etch virus (TEV) cleavage site.

2.17. Cloning of LSm4/8 bicistron into pQE30 expression vector

We were not able to obtain LSm4 cDNA from the same library; the source for LSm4 cDNA was the EST Clone IMAGE: 4902795, ACCESSION BG826318, VERSION BG826318.1 GI: 14173905. From the EST clone, LSm 4 was amplified and subcloned into the pUC19 vector. The LSm4 internal BamH1 site was removed by a site directed mutagenesis procedure. LSm8 was subcloned into pUC19.1 using Ndel and Hind III enzymes. Cutting this construct with BamH1 enzyme and ligation of LSm4 (BamH1, Xba1 fragment) yielded a bicistronic construct in which LSm4 is the first cistron. EcoRI/ HindIII cutting was used to transfer His6-tag-TEV site segment followed by the LSm4 and LSm8 expression cassettes into the expression vector pQE30.

2.18. Cloning of LSm5/6/7 tricistron into pQE30 expression vector

LSm5, LSm6 and LSm7 cDNAs are amplified from cDNA library, using oligonucleotides which include on 5' Sac I site and 3' Hind III site. After PCR amplification, these three cDNAs are separately cloned into pUC19 cloning vector. From LSm 5 internal EcoRI site was removed by site directed mutagenesis, because this site would interfere with later cloning steps. LSm6 was introduced as Ndel/Hind III fragment and transferred to pUC19.1 vector. Incorporation of LSm6 ORF was checked by different restriction enzymes analysis. LSm5 cDNA was inserted as BamHI and Xba1 fragment into LSm6:pUC19.1. LSm7 was subcloned into pUC19.1 vector using Ndel and Hind III sites. For introducing LSm7 as a third cistron, BamH1/HindIII fragment was ligated.
with LSm5-6 dicistron was cut with Bgl2 and HindIII enzymes. BamH1 and Bgl2 are making joint, and result of all these cloning steps is His tag TEV LSm5-6-7-pUC19 cloning vector. As a cassette (Eco1, Hind III sites used) it is transferred into pQE30 expression vector.

2.19. Cloning of LSm1 into pQE30 expression vector

Similarly, we were not able to utilize the cDNA library for obtaining human LSm1 protein so, cDNA for LSm1 was obtained from an EST clone, Homo sapiens cDNA clone IMAGE: 4793732 5', ACCESSION BG714535VERSION BG714535.1 GI: 13993466
LSm1 was amplified using NdeI 5' site and HindIII 3' site oligos, and subcloned as a BamH1/HindIII fragment into a modified pUC19 vector (pUC19.3). As a cassette EcoR1/HindIII fragment which contains His6-tag-TEV site cDNA was transferred into the pQE30 expression vector.

2.20. Cloning of LSm4 into pQE30 expression vector

For obtaining single cistron LSm4 expression construct, LSm4 was subcloned as 5'BamHI /3' HindIII fragment into pUC19.3. After sequencing, transferred as a cassette into pQE30.

![Figure 2.3. H6-TEVLSm4:pQE30 vector](image)

2.21. Cloning of LSm8 into pQE30 expression vector

BamH1/ Hind III fragment, which contains LSm8 cDNA was introduced into pUC19.3, and after sequencing transferred into the pQE30 expression vector.

![Figure 2.4. H6-TEV LSm8:pQE30 vector](image)
2.22. Cloning of LSm6 into pQE30 expression vector

BamH1/HindIII fragment LSm6 was ligated with pUC19.3 (NdeI/HindIII cut), and after sequencing transferred as a cassette into the pQE30 expression vector.

![Figure 2.5. H6-TEV LSm6:pQE30 vector](image)

2.23. Cloning of LSm5:pQE30

The LSm5 BamH1/HindIII fragment was introduced into pUC19.3, and transferred as a cassette into the pQE30 T5 expression vector.

![Figure 2.6. H6-TEV LSm5:pQE30 vector](image)

2.24. Cloning of LSm5/7: pQE30

NdeI restriction digestion removes LSm6 gene and leaves bicistronic LSm5/7, which was transferred using EcoR1 and Hind III sites into T5 pQE30 expression vector.

![Figure 2.7. H6-TEV LSm5/7:pQE30 vector](image)
2.25. Cloning of LSm5/6:pQE30

First steps in cloning of tricistronic LSm5/6/7 was introducing LSm6 into pUC19.1 using NdeI/HindIII, and after BamH1/XbaI fragment of LSm 5 into generated monocistron. This was called LSm5/6: pUC19.1 intermediate in cloning of LSm5/6/7 tricistron. EcoRI/ HindIII fragment was cut and ligated into pQE30 (EcoR1/HindIII cut vector).

![Diagram](image)

**Figure 2.8. H6-TEV LSm5/6:pQE30 vector**


LSm3 NdeI/HindIII fragment (from LSm2/3 construction) was inserted into LSm5:pUC19.1. Transfer EcoR1/HindIII fragment to pQE30.

![Diagram](image)

**Figure 2.9. H6-TEV LSm5/3:pQE30 vector**

2.27. Cloning of LSm3/6:pQE30

NdeI/HindIII fragment of LSm6 was inserted into pUC19.1. LSm3 was inserted as a BamH1/XbaI fragment, using BamH1 site on pUC19.1. EcoR1/HindIII fragment was transferred to pQE30.
2.28. Cloning of LSm4/2:pQE30

LSm2 NdeI/HindIII fragment (5' and 3' sites respectively) was inserted into LSm4:pUC19.1 after sequencing transfer EcoR1/HindIII fragment to pQE30.

2.29. Cloning of LSm4/7:pQE30

Insert LSm7 NdeI/HindIII fragment (from LSm567 construction) into LSm4:pUC19.1 Transfer EcoR1/HindIII fragment to pQE30.
2.30. Cloning of ZZLSm4:pQE70

First step in obtaining this clone was PCR using EST clone and 5' LSm4 BamHI oligos and 3' LSm4 XbaI oligos. Then LSm4 cDNA was subcloned into the pUC19.6 vector, and transferred into pQE702ZZ.3 vector.

Figure 2.13. ZZ H6-TEV LSm4:pQE70 vector

2.31. Cloning of ZZLSm6:pQE70

The first step was a PCR using 5' BamHI LSm6 and 3' XbaI oligos, followed by subcloning into pUC19.6 as a BamHI/HindIII fragment and transfer to the pQE702Z.3 vector.

Figure 2.14. ZZH6-TEV LSm6:pQE70 vector

2.32. Cloning of ZZLSm8:pQE70

LSm8 was subcloned using BamHI and XbaI sites into pUC19.6 and after into pQE702Z.3.

Figure 2.15. ZZH6-TEV LSm8:pQE70 vector
2.33. Kinasing and annealing oligonucleotide linkers

This reaction was used in several modification steps of pU19, in introducing six histidine tag, the tobacco etch protease cleavage site, and introducing intergenic region linker for vectors which carry more than one gene. Oligolinker (upper and lower strand were kinased) using ATP and polynucleotide kinase in 20µl reaction. The reaction mixture was made to the following recipe:

- 5µl of 10µM oligo (upper and lower)
- 2µl of 10X polynucleotide kinase buffer (PNK buffer)
- 2µl of 10mM ATP
- 1µl T4 Polynucleotide kinase (PNK) 10 000 U/ml
- 10 µl of Millipore water.

The reaction of kinasing the oligolinkers was performed for 40 minutes at 37 °C. The enzyme was inactivated by incubation for 10 minutes at 70°C. Upper and lower strand oligonucleotide were mixed and allowed to anneal by cooling slowly down to room temperature. The kinased linker was then ligated with the vector in the following assay:

- 2µl of cut, phosphatased vector DNA (0.1mg/ml)
- 1µl 10X T4 ligase buffer
- 0.5µl T4 DNA ligase
- 6.5µl oligolinker

Reaction was incubated for more than 4 hours at 16°C, and after this reaction of transformation of cells was performed according to the standard procedure.

2.34. Determining optimal expression parameters in E. coli

To conduct a test expression in E. coli the expression plasmids were transformed into competent cells as described in 2.12. After incubation over night (~12-14 hours) in 37 °C on LB- agar plates with the appropriate antibiotics, one colony of cells was used to inoculate cultures of 100 ml, supplemented with the appropriate antibiotic, and shaken at 200 rpm at 37 °C. Here either LB or 2xTY were tried. Cells were then grown at 37 °C until they reached an OD600 ~0.8 and were induced by 1 mM IPTG. Prior to induction 1 ml of the culture was centrifuged at 13’000 rpm for one minute and frozen. Here conditions for optimal protein expression were screened by varying temperature (18°C, 25 °C, 30 °C, 37 °C). After 1, 3, 5, 20, and 48 hours of induction again samples of 1 ml were taken. For each different temperature and different time point the OD600 of
the cells was determined. The related samples showing higher OD600 were diluted to the lowest value, for comparison reasons. These samples were centrifuged as described above and frozen. In order to determine expression levels, the pellets were re-suspended in 150 μl MPW, and lysed by the addition of 30 μl of 6x SDS dye and boiling for 10 minutes at 95 °C. Samples were then analyzed by SDS-PAGE on a 12 % SDS-High TEMED gel.

2.35. Growing expression cultures of E. coli

For protein purification, usually 1 liter culture, or multiples there of were grown. A single colony from an LB agar plate was inoculated into cultures of 100 ml, supplemented with the appropriate antibiotic. The medium used here was usually LB medium. Cultures were then grown at 37 °C and shaken at 200 rpm over night. The following morning, 20 ml of the resulting dense culture were diluted into 1 liter of the optimal medium, supplemented with the appropriate antibiotic. The OD600 resulting was always found to be ~0.1. Cells were then grown to an OD600 of ~0.8. At OD600 ~0.8 a 1 ml sample was centrifuged as described above. Cells were cooled to the appropriate temperature (ten minutes at 4°C), and induced with 1 mM IPTG. Cells were then grown for the optimal amount of time as determined by the procedure described in 2.30. For most of the proteins optimal expression temperature was 25°C and expression time 16-18 hours (over night). LSm4/8 dicistron construct was expressed for 4 hours at 25°C. LSm5/6/7 tricistronic construct was expressed at 37°C for 48 hours.

2.36. Harvesting and disruption of the cells

Cells were harvested by centrifugation for 30 minutes at 5000 rpm in as Sorvall RC3B+ centrifuge at 4°C. After harvesting, cells were either frozen for further usage at -20 °C or directly processed. Cell pellet from one liter culture was re-suspended in 20 ml Lysis buffer (IMAC A with EDTA free protease inhibitor cocktail tablets, one tablet was dissolved into 50 ml of IMAC A buffer). If the large scale preparations were carried out (12 l cultures) then, cells were disrupted after harvesting using Constant Systems high pressure disruptor. If small scale preparations (till 6 l cultures) were carried out than for disruption of the cells sonication was sufficient to disrupt cells, using a 50% duty cycle (1” on, 1” off) at 40% output with a microtip (horn diameter 6 mm, Sonics Vibra Cell sonicator), at 4°C for 20 minutes. To reduce viscosity of the suspension, 10 μg/ml DNaseI was added (in the ratio 1:50) and the mixture was incubated for 1 hour on ice.
If the suspension was still viscous after this treatment, the suspension was further subjected to sonication. The preparation of a cleared lysate was achieved by centrifugation for 40 minutes to remove cell debris and still non disrupted cells in Optima XL-100 Ultracentrifuge (Beckman), using a 45 Ti rotor at 4 °C, at 35000rpm.

2.37. Immobilized Metal Affinity Chromatography (IMAC)

As described previously all expression constructs were designed to carry on the first protein six Histidine Tag and another proteins were co purified via complex formation. First and the main purification step was IMAC. To purify His6- tagged LSms proteins, the supernatant prepared as described in chapter 2.36. was applied onto a Ni-NTA column at a flow rate of 1ml/min. If the larger yields of protein are expected 5 ml Ni Hi Trap Chelating column was used, and for purification test purposes 1ml Ni Hi Trap Chelating column. In both cases supplier was Amersham Pharmacia. Prior to application a sample of 50 µl was taken for subsequent analysis by SDS- PAGE (input). The pellet was resuspended in 20 ml lysis buffer, and 50 µl was taken for subsequent analysis by SDS- PAGE (pellet). To remove unbound sample from the column, the column was washed with 10 column volumes (CV) of IMAC A at 1 ml/min. Samples were eluted by a step gradient of 10 CV of 10 % IMAC B (60 mM Imidazole elution), 10 CV 50 % IMAC B (250mM Imidazole elution), and 10 CV 100 % IMAC B (500mM Imidazole elution) at 1ml/min. For each elution step, from the fractions samples of the 50 µl were taken and analyzed later by SDS- PAGE.

2.38. Anion exchange chromatography

If the LSms proteins were not sufficiently pure after first affinity chromatography purification step, they were prepared for next purification step, anion exchange chromatography, by dialysis into MonoQ-A150, or by dilution with MPW until conductivity of the sample correspond to that of MonoQ-A150, or using PD10 desalting column (Amersham Biosciences). Samples were applied at Resource Q 6ml column if larger preps were worked out, or Resource Q 1ml if small scale cultures were purified) at flow rate of 1 ml/min. Non-bound samples were washed with 10 CV of MonoQ- A150. Samples were eluted with a linear gradient from 0- 100 % MonoQ- B1000 over a length of 30 CV.
2.39. Cation exchange chromatography

During course of this work cation exchange chromatography was used for purification of ZZH6TEVLSm6 after affinity chromatography step, and for separation of cleaved LSm6 from ZZH6tags.

After purification of ZZH6TEV LSm6 using affinity chromatography, LSm6 was released from ZZH6 using TEV protease. Purification of LSm6 from the fusion tags was achieved using Resource S 6 ml (or for the small scale preps Resource S 1 ml column). Sample after digestion with TEV protease was prepared for chromatography using PD10 desalting columns (Amercham Biosciences), by diluting with MPW until conductivity of the sample corresponds to the conductivity of low salt buffer, or by dialysis against: 150mM NaCl, 20mM BisTris 6.5, 5mM betamercaptoethanol. Sample was applied on Resource S 6 ml column at flow rate 1 ml/min. Non bound fraction was washed with 10 column volumes of 150mM NaCl, 20mM BisTris 6.5, 5mM betamercaptoethanol, and bound sample was eluted using salt gradient from 150 mM NaCl-1M NaCl over 30 column volumes.

2.40. Reversed Phase Chromatography

Reversed phase chromatography has been used for analytical applications, namely for judging stoichiometry of protein sub-complexes, of heteroheptamers. In these experiments C18 column was used, at 50 °C. Linear gradient was used.

Mobile phase A was: 0.1% Trifluoroacetic acid in MPW. Mobile phase B was: 0.1% Trifluoroacetic acid in Acetonitrile. Linear gradient from 0-100% mobile phase B was used over 30 column volumes at speed 5μl/minute. While separation was in progress peaks corresponding to different LSm5s were collected manually and subjected to the MALDI TOF, for the identification, and SDS PAGE analysis. Since collected protein is acidic Tris base was added in order that SDS PAGE can be conducted. Area under identified peak was divided with number of tyrosines and triptophans which are present in protein. If in the protein subcomplex (LSm5/6/7) all proteins are present in equimolar ratios divided numbers should be close to each other.
2.41. Gel filtration chromatography

For semi-preparative gel filtration chromatography pre-packed columns Superdex 75 HR10/30, and Superdex 200 HR 10/30 (Amercham Biosciences) were used. Columns were equilibrated with 2 CV of buffer prior to the actual chromatography run. Sample was concentrated to 500-1000 µl and injected onto the column via a 1 ml sample loop. The loop was flushed with 1 ml of buffer, thereby allowing injection of the sample on the column. Sample was eluted by isocratic elution at a flow-rate of 0.5 ml/min for 1.5 CV.

Figure 2.16. Calibration curve for Superdex 200 HR10/30 column

2.42. Determination of protein concentration

Concentration of protein samples was determined by measuring absorbance spectra from 220-340 nm. Concentration was determined by the equation: \( c = \frac{A_{280}}{c_{280}}d \) to determine the molar concentration. To determine the concentration in mg/ml, the obtained molar concentration value was multiplied with the calculated molecular weight of the protein.
### Table 2.1: Physical parameters of LSm protein constructs

<table>
<thead>
<tr>
<th>Protein or complex</th>
<th>MW</th>
<th>Mabs</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSm1</td>
<td>15289.61</td>
<td>3840</td>
<td>5.11</td>
</tr>
<tr>
<td>His6-TEV-LSm1</td>
<td>17776.12</td>
<td>5120</td>
<td>6.05</td>
</tr>
<tr>
<td>LSm1 post TEV</td>
<td>15483.8</td>
<td>3840</td>
<td>5.23</td>
</tr>
<tr>
<td>LSm2</td>
<td>10833.75</td>
<td>5360</td>
<td>6.09</td>
</tr>
<tr>
<td>His6-TEV-LSm2</td>
<td>13171.06</td>
<td>6640</td>
<td>6.76</td>
</tr>
<tr>
<td>LSm2 post TEV</td>
<td>10977.87</td>
<td>5360</td>
<td>6.09</td>
</tr>
<tr>
<td>LSm3</td>
<td>11844.59</td>
<td>5120</td>
<td>4.58</td>
</tr>
<tr>
<td>His6-TEV-LSm23</td>
<td>25015.65</td>
<td>11760</td>
<td></td>
</tr>
<tr>
<td>LSm23 post TEV</td>
<td>22822.46</td>
<td>10480</td>
<td></td>
</tr>
<tr>
<td>LSm4</td>
<td>15348.91</td>
<td>15580</td>
<td>10.07</td>
</tr>
<tr>
<td>His6-TEV-LSm4</td>
<td>17698.28</td>
<td>16860</td>
<td>10.07</td>
</tr>
<tr>
<td>LSm4 post TEV</td>
<td>15493.03</td>
<td>15580</td>
<td>10.02</td>
</tr>
<tr>
<td>LSm5</td>
<td>9937.16</td>
<td>120</td>
<td>4.42</td>
</tr>
<tr>
<td>His6-TEV-LSm5</td>
<td>12274.47</td>
<td>1400</td>
<td>5.34</td>
</tr>
<tr>
<td>LSm5 post TEV</td>
<td>10081.28</td>
<td>120</td>
<td>4.42</td>
</tr>
<tr>
<td>LSm6</td>
<td>9126.77</td>
<td>6520</td>
<td>9.61</td>
</tr>
<tr>
<td>ZZ-His6-TEV-LSm6</td>
<td>24921.37</td>
<td>10360</td>
<td>7.99</td>
</tr>
<tr>
<td>LSm6 post TEV</td>
<td>9126.77</td>
<td>6520</td>
<td>9.61</td>
</tr>
<tr>
<td>LSm7</td>
<td>11601.67</td>
<td>4080</td>
<td>5.11</td>
</tr>
<tr>
<td>His6-TEV-LSm567</td>
<td>33002.91</td>
<td>12000</td>
<td></td>
</tr>
<tr>
<td>LSm567 post TEV</td>
<td>30809.72</td>
<td>10720</td>
<td></td>
</tr>
<tr>
<td>LSm8</td>
<td>10402.06</td>
<td>2560</td>
<td>4.34</td>
</tr>
<tr>
<td>LSm8 post TEV</td>
<td>10459.11</td>
<td>2560</td>
<td>4.34</td>
</tr>
<tr>
<td>His6-TEV-LSm48</td>
<td>28100.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSm48 post TEV</td>
<td>25895.09</td>
<td>18140</td>
<td></td>
</tr>
</tbody>
</table>

**2.43. Tobacco Etch Virus (TEV) protease cleavage**

The concentration of the protein sample to be cleaved was determined, as described in 2.42. TEV protease was added in a 1:50 (protease: protein) ratio. The sample containing protease was subjected to cleavage reaction 40-48 hours at room temperature (~25 °C). Cleaved sample, the cleaved tag, the His-tagged TEV protease, and uncleaved sample were then purified by IMAC. The IMAC procedure was further modified insofar as elution was carried out with a linear gradient from 0-100% IMAC B over 20 CV.
2.44. Sodium Dodecylsulfate- Polyacrylamide Gel Electrophoresis (SDSPAGE)

To analyze purity of protein samples, SDS-PAGE was used, as described by [172].
For SDS-PAGE the Hoefer Mini-VE system (Amercham Biosciences) was utilized. The stacking gel contained 5 %, and the resolving gel contained 12 % SDS. A high TEMED gel system was used, because resolution was found to be better. To protein samples that had to be analyzed, 6x SDS dye was added in a 1:6 (dye: sample) ratio. Samples were then boiled at 95 °C for 5 minutes. The gel was run for 15 minutes at 150 V, then for 50 minutes at 200 V. After this procedure the dye front was found to have reached the bottom of the gel.

2.45. Coomassie- staining

After SDS-PAGE the gel was transferred into Fairbanks A solution, heated in a microwave, and incubated for ~20 minutes. Next the gel was transferred into Fairbanks B solution, again heated in a microwave, incubated for ~45 minutes. Gels were distained in Fairbanks D [170].
Gels were documented by scanning.

2.46. In Vitro transcription of U1 snRNA and U6 snRNA

U6 snRNA was obtained by transcription in vitro of cut, linearized template. Template was Xenopus tropicalis T7 U6 snRNA (obtained as gift from Prof. I. Mattai, EMBL, Heidelberg). Template was linearized with Dra1 enzyme, and purified after. U6 snRNA template was digested over night with Dra1 enzyme in appropriate buffer (MBI buffer B+) in reaction volume of 5 ml at 37°C. Usually 500 µg template was cut in 5ml reaction volume. Linearized template was supplemented with 1ml of 3M Na-Acetate pH5.2 and 5 ml Phenol. Mixture was vortexed and phases were separated in the Centrifuge 5810R for five minutes at 4000 rpm. Upper, water phase was transferred into falcon tube with addition of 12 ml ice cold ethanol after addition of linear polyacrilamide at 0.25% as carrier. Mixture was vortexed and centrifuged in the SW28 swinging bucket rotor in 40000 rpm for 30 minutes, in ultracentrifuge. Pellet with linearized template is on the bottom of the centrifugal tube. Yield from 500 µg template was 390 µg of the cut, linearized purified template for the transcription in vitro.

Transcription in vitro was carried out in the 5 ml reaction, with 100 µg of linearized template, in transcription buffer with addition of 5mM rNTPs, and T7 RNA polymerase. After 5 hours incubation at 37 degrees, transcribed U6 snRNA was purified on 8% 8M urea denaturing gel.
Band was excised and cut into the slices, and eluted from polyacrilamide by the soaking over night in milipore water. Sample was after dialyzed against milipore water at 4°C, for 4 hours, and concentrated using speedvac till 100 μM or higher. Yield of in vitro transcribed U6 snRNA was 590 μg form such reaction where 500 μg template was used. Using same protocol was obtained U1 snRNA, and template was Xenopus laevis T7 U1snRNA:pUC19 BamH1 cut template.

**Figure 2.17. In Vitro transcribed U6 and U1 snRNA**

### 2.47. Preparation of radioactively labeled U1 and U6 snRNA

U6 Xenopus tropicalis snRNA, and U1 Xenopus laevis snRNA subcloned into pUC19 vector were obtained from laboratory of Ian Mattai (EMBL, Heidelberg). U1 snRNA template was linearized using BamH1 site and U6 snRNA using Dra1 site. Linearized templates were purified as described in chapter 2.46. For obtaining radioactively labeled (^32P U snRNA), hot transcription was performed. Reaction mixture was composed from following compounds:
Table 2.2. Hot transcription reaction mixture for obtaining $^{32}$P labeled U1 and U6 snRNA

Transcription was performed one till two hours at 37°C, in lid container. Transcribed RNA was applied on 8% denaturing RNA gel (8%, 8M urea 0.5xTBE) dimensions 20x40x0.15cm. Electrophoresis was carried out at 4 °C, 200 V, 20mA for 5 hours, till bromphenol blue is at the bottom of the gel. After this gel was wrapped with cellophane and transferred to cassette for autoradiography. Kodak X-ray film was exposed for 5-10 minutes, to evaluate purity of obtained transcripts.

Figure 2.18. Transcribed U1 and U6 snRNA on exposed X ray film
The obtained transcripts shown in figure 2.18. were not sufficiently pure. Bands were excised and eluted in 3-4 ml of milipore water rotating head-over-tail at 4°C. Excised bands were filtrated 0.45μm filter and dialyzed against milipore water 2x2 hours. After releasing from urea and contaminants from gel, RNAs were concentrated using Speedvac till 100 μl. Small 12x12x0.15 cm denaturing gel was run to check purity of snRNA after procedure.

**Figure 2.19.** Transcribed and purified U1 and U6 snRNA on exposed X ray film

After ensuring that U1 snRNA and U6 snRNA are sufficiently pure, for conducting gel retardation assay, scintillation was used to count radioactivity of samples (using Cerenkov detection method) Per each reaction was used 10000cpm (counts per minute). 1μl of purified, 32P SnRNA (U1 and U6 snRNA) was measured in plastic vial in scintillating machine, and from this values volume necessary to reach 10000 cpm was calculated.

**Table 2.2.** Counts of UTP labeled U1snRNA and U6snRNA

<table>
<thead>
<tr>
<th></th>
<th>BG</th>
<th>U1-#1</th>
<th>U6-#1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts</td>
<td>7</td>
<td>4936</td>
<td>26401</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>4461</td>
<td>29067</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>5161</td>
<td>28409</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>3887</td>
<td>28719</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td>31122</td>
</tr>
<tr>
<td>Mean:</td>
<td>11</td>
<td>4611</td>
<td>28744</td>
</tr>
<tr>
<td>SD:</td>
<td>3</td>
<td>564</td>
<td>1685</td>
</tr>
<tr>
<td>Mean overall:</td>
<td></td>
<td>4611</td>
<td>28744</td>
</tr>
<tr>
<td>μl for 10,000cpm</td>
<td></td>
<td>2.17</td>
<td>0.35</td>
</tr>
</tbody>
</table>

In general reaction mixture for EMSA in these experiments was:
10000 cpm of appropriate RNA per reaction

1 µl of tRNA (1mg/ml)

5 pmol of proteins

0.5 µl of RNAsin

1x Reconstitution buffer (300mM NaCl, 12mM NaHepes7.5, 1.5mM MgCl₂, 0.1mg/ml of tRNA final concentration and 0.1% triton X100, 10% Glycerol)

<table>
<thead>
<tr>
<th>Component:</th>
<th>Ul snRNA</th>
<th>U6snRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ul snRNA</td>
<td>free</td>
<td>Sm</td>
</tr>
<tr>
<td></td>
<td>1.78</td>
<td>1.78</td>
</tr>
<tr>
<td>U6 snRNA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>tRNA 1 mg/ml</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RNAsin</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Sm proteins 5 µM</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LSm2-8 new 5 µM</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Buffer</td>
<td>6.72</td>
<td>5.72</td>
</tr>
</tbody>
</table>

Table 2.3. Pipeting scheme for EMSA (Electrophoretic mobility shift assay)

First radioactive RNA, tRNA, and reconstitution buffer were mixed together and incubated at 80°C for 5 minutes, and immediately transferred on ice. Probably during incubation at 80°C re-annealing of RNA occurs and conversion of RNA into monomeric form. On ice 5 pmol of appropriate proteins were added and 0.5 µl of RNAsin. The final volume of the reaction mixture was 10 µl. After all components were present in the reconstitution mixture, it was incubated at 30°C for one hour. After reconstitution samples were loaded onto a native 6% RNA gel (0.075% Bisacrylamide 80:1 in 500 ml of 0.5 xTBE buffer). Native 6 % RNA gel was made from 30 g acrylamide, 0.375g Bisacrylamide, 25 ml of 10xTBE to 500 ml of Millipore water), filtrate
0.22\textmu m. Per one gel 40 ml of this mixture as well as 400\mu l of 10\% ammonium persulfate, and 40\mu l of TEMED were used. The gel was poured immediately and pre-run 30 minutes at 160V. Upon reconstitution with RNA samples were loaded on slots and gel was run, 160 V for 4 hours, and after this wrapped into saran wrap, and transferred into cassette. In dark room Kodak film was overlaid on gel. Exposure was performed for 12 hours at -80\degree C, followed by developing of film.

2.48. Microinjection in to the REF52 cells

In the line of all other experiments which are done during my dissertation, and for deeper characterization of my system I used microinjection technique. Before material is microinjected it has to fulfill some requirements:

Protein solutions were filtered with 0.2mm filter to remove particulates which may clog the injection needle. Rinse Eppendorf tube and tips with injection buffer (water, PBS 1X) to remove any residual contaminants. During these experiments proteins were microinjected in 150 mM NaCl, 20 mM HEPES-NaOH pH 7.5, 5 mM \( \beta \)-Mercaptoethanol (i.e. in a buffer in which they were purified and stable).

Before microinjection, the solution was centrifuged 15 min. at 13.000 rpm RT. Cells used during course of experiments were REF52 cells, and they were grown in Optimem medium (composition is same like Dulbeco Modified Medium, except it does not contain phenol red). After microinjection into the cytoplasm of REF52 cells, they were immediately visualized using an Olympus confocal microscope. For inhibition of active transport WGA was used according to the procedure described in reference 105.

![Microinjection into the cells](image)

Figure 2.20. Microinjection into the cells
2.49. Coupling of proteins to the fluorescent dyes

After purification of proteins of interest or reconstitution of heteroheptameric rings, they were coupled to fluorescent dyes. We used Alexa 488 (green) or Alexa 555 (red), succinimidyl esters. They were prepared according to the recommendation of the supplier (Invitrogen). These dyes are reacting with amino terminal part of the protein or with exposed Lysine residues. Coupling was performed for 2 hours at room temperature with stirring. Three times molar access of dye over protein was used. Non-coupled dye was removed from labeled protein using dialysis over night at 4 °C against the same buffer in which is protein.

2.50. Electron microscopy

All electron microscopy analysis of prepared specimens was done in the laboratory of Prof. A. Engel (Biozentrum, Basel), by Dr. M. Chami. Samples were diluted at 10 to 20 µg/ml. Aliquots of 5 µl were stained with 1 % (w/v) uranyl acetate after sample adsorption onto glow discharged 400 mesh carbon-coated grids. The micrographs were recorded at an accelerating voltage of 100 kV and a magnification of 50,000×, using a Hitachi 7000 electron microscope. All micrographs were recorded on Kodak SO-163 film. Reference-free alignment was performed on manually selected particles from digitized electron micrographs using EMAN image processing package. After multivariate statistical analysis of a set of rotational and translational invariants previously generated, a reference free k-means classification was performed on the resulting footprint file. The resulting classified images are then aligned and classified iteratively. The class average with the best signal to noise ratio were selected and gathered in a gallery.

2.51 Crystallization trials

Crystallization of macromolecular samples is an empirical process that relies upon many variables, such as pH, buffers, concentration of the sample, concentration and type of the precipitant, temperature, ionic strength, homogeneity of the sample, detergents, additives, and conformational stability of the sample (McPherson 1990). Crystallization trials were performed, using the sittingdrop vapour diffusion method. Determination of primary crystallization conditions was achieved by using “Sparse- matrix screens/ incomplete factorials” (Carter and
Carter 1979; Jancarik and Kim 1991). When an initial hit was found, 2-dimensional grid-screens were used to optimize crystallization conditions.
3. RESULTS and DISCUSSION

The composition of the ribonucleoprotein particles (RNPs) which are forming on mRNA determines its fate, that is, where, whether and how efficiently it will be translated, and how stable it is. In consequence, the functional role of a given protein has to be viewed in the context of the RNP of which is part. This fact highlights the importance of addressing the structure-function relationship of RNPs: Only by studying the structure of RNPs it is possible to dissect the functional role of a given protein subunit in its different complexes.

A particularly interesting example for this phenomenon is the Sm/LSm protein family, whose members are engaged in a very wide variety of RNA processing events, forming complexes which differ sometimes only by one out of seven subunits (see introduction). Another important aspect of the Sm/LSm protein family is that these proteins never occur in isolation; for proper functioning they require complex formation. Hence, the way to better understand Sm/LSm protein function is to study Sm/LSm complexes. It is difficult to determine the connection between the oligomeric state of a given protein and its function in vivo. In consequence, the interdependence of Sm/LSm protein complex formation and function suggests that it is advisable to work with complexes that have been reconstituted in vitro from recombinant proteins.

We chose to study the structure-function relationship in two human LSm complexes with seven subunits each, LSm1-7 and LSm2-8.

When trying to obtain the large amounts of proteins necessary for X-ray crystallographic studies, it is often convenient to use the *Escherichia coli* (*E. coli*) overexpression system. This system has the advantage of easy handling of cultures and recovery of large amounts of the protein product and wide availability and versatility of cloning and expression vectors. The only source for sufficient quantities of LSm proteins pure enough for biophysical studies, crystallization and reconstitution of the heptamers was overexpression of recombinant proteins in a suitable host.

We chose the T5 expression system, and as expression host for the LSm proteins *E. coli*.

This choice was based on successful coexpression of the homologous canonical Sm proteins carried out previously, using the same expression host and system [72].

Each of the eight human LSm (LSm1-LSm8) proteins present in the LSm1-7 and LSm2-8 complexes has a closest parologue among the proteins which build up the canonical Sm core domain (D1, D2, D3, E, F, G, B), as defined by sequence identity levels.
LSm2 has the highest sequence identity to SmD1, LSm3 is most similar to SmD2, LSm4 to SmD3, LSm5 to SmE, LSm7 to SmG, and the highest pairwise identity (46%) occurs between LSm6 and SmF (Figure 3.2).

It is generally assumed that these pairwise highest identity levels translate into biochemically similar behavior, in particular binding specificity with other LSm proteins. Hence, the current hypothesis on the architecture of the LSm rings has it that closest paralogues take up equivalent positions to those in the canonical core snRNP domain. This concept is partially corroborated by yeast two-hybrid data [120, 124].

![Figure 3.1](image)

**Figure 3.1.** Three different heptameric complexes contain Sm or LSm proteins, reprinted from [123]

The Sm proteins bind to the single stranded Uridine-rich sequence (the Sm site) that is found between two stem-loops in spliceosomal U snRNAs. The LSm2-8 proteins bind to the Uridine-rich tract at the 3' end of U6 snRNA. The binding site of the LSm1-7 complex on mRNA has not been determined. Apart from the yeast two hybrid data there is no experimental evidence for order of the subunits in the LSm rings. The order of the subunits in the Sm ring is based on biochemical and genetic data on the pairwise nearest-neighbour relationships. In essence, the structural model of the Sm ring was designed such as both to satisfy the biological data and to be compatible with the topology of the Sm-Sm interfaces, i.e. the two crystal structures plus the concept of a repetitive interface [72, 73, 77].
Obtaining Sm proteins from monocistronic expression vectors was tried previously, but singly expressed Sm proteins were insoluble (C. Kambach, unpublished result). SmD1D2, SmD3B, and SmEFG are found as stable heterodimers and heterotrimers during Sm core assembly in vivo, [73] (Described in more detail in the introduction). Moreover, they can be expressed in E. coli in the same format.

Coexpression of Sm/LSm proteins as a tool for obtaining large amounts of recombinant proteins has several advantages, like increasing yields, solubility and stability of the coexpressed Sm/LSm proteins. Furthermore, usage of heterodimers and -trimers should facilitate in vitro reconstitution of the two heteroheptamers and, by pre-defining several nearest neighbour relationships, possibly provide some information on their architecture.

Figure 3.2: Sm/LSm proteins correspondence according to sequence similarities
3.1. Cloning strategy

The choice of given LSm combinations (LSm2/3, LSm4/8, LSm5/6/7, LSm4/1) was based on pair wise sequence identity levels compared to Sm proteins, and the established composition of the canonical Sm protein sub-complexes [73].

![Subunit interaction for LSm2-8 and LSm1-7 complexes](image)

**Figure 3.1.1.:** Subunit interaction for LSm2-8 and LSm1-7 complexes, reprinted from [124]

Our chosen T5 expression system relies on low copy number vectors (pQE30, pQE70) that are difficult to propagate and manipulate in sufficient yields for complex subcloning work. For this reason, we used a cloning intermediate (pUC19 vector).

The building of the dicistronic vectors for coexpression of LSm2/3, LSm4/8 and other dicistronic constructs in general was achieved by subcloning the second cistron as a cassette in form of an NdeI/HindIII fragment first, followed by introducing the first cistron as a BamHI/XbaI or, in case of LSm2/3 construction, as a BclI/BglII fragment. In case of constructing the LSm5/6/7 tricistronic vector, the last cistron was subcloned into the pUC19.1 intermediate and subcloned into dicistronic LSm5/6, pUC19.1 intermediate using compatible overhangs (BamHI/HindIII for the donor LSm7 fragment, and BglII/HindIII for the recipient, the LSm5/6 dicistronic vector).
The general features of all our expression constructs are a His\textsubscript{6}-tag followed by a Tobacco Etch Virus (TEV) cleavage site, a ribosomal binding site, the respective cDNA, then an intergenic region linker, followed by the second cDNA, and, in the case of the LSm5/6/7 tricistron, again an intergenic region linker and the third cDNA. Because of the reasons mentioned above (T5 expression vectors are low copy number vectors, hence difficult to use for cloning purposes), all cloning was performed in the pUC19.1 vector (see section 3.1.1.) and the cDNAs of interest were transferred in form of a cassette (using EcoRI/HindIII sites) into the pQE30 T5 expression vector as the last subcloning step. The latter was used for transforming the expression host strains.

3.1.1. Construction of pUC19.1

The original pUC19 vector was modified by replacing the BamH1/Hind III fragment of the multiple cloning site with an intergenic region linker for the construction of expression vectors that carry more than one open reading frame. The general features of the modified pUC19 vector used for cloning (pUC19.1) are a His\textsubscript{6}-tag, a TEV protease cleavage site in front of the first citron followed by the intergenic region linker which contains a second ribosomal binding site, and restriction sites at the 3' end making it possible to introduce further expression cassettes by compatible overhang cloning.

![Figure 3.1.2. Map of pUC19.1](image)

3.1.2. Subcloning of the LSm2/3 and LSm4/8 bicistrons into the pQE30 expression vector

The cloning strategy described above was used to obtain the LSm2/3 and LSm4/8 expression constructs. The detailed procedure is described in the Materials and Methods part of this thesis.

73
The second cistron (LSm3 or LSm8, respectively) was subcloned into the pUC19.1 vector (Figure 3.1.2.), using the NdeI and Hind III sites. The first cistron was placed downstream of the His6-tag and TEV site using compatible BamH1/XbaI sites in the case of LSm4, and BclI/BglII sites with BamH1, in the case of LSm2.

![Figure 3.1.3]: His6-TEV-LSm2/3:pQE30 and His6-TEV-LSm4/8 expression constructs

3.1.3. Cloning of the LSm5/6/7 tricistron into the pQE30 expression vector

The strategy described above was used to obtain the LSm5/6/7 tricistronic vector. The detailed procedure is described in the Materials and Methods part.

![Figure 3.1.4]: His6-TEV-LSm5/6/7:pQE30 expression constructs

Our original concept was that the sequence – binding specificity relationship works backwards for coexpression: Only those pairs of LSm proteins representing true physiological sub-complexes or at least reflecting nearest-neighbour relationships in the final heptamers were supposed to yield well expressed, soluble heterodimers and -trimers. In order to test this
hypothesis we constructed several negative controls (LSm6/7, LSm4/2, LSm5/3), representing non-nearest neighbours in the assumed order of subunits in the rings. The general cloning strategy was the same as in the case of constructing the nearest-neighbour pairs. The cloning maps can be found in the Materials and Methods part of the thesis.

A construct required for the reconstitution of the LSm1-7 ring was the LSm 4/1 heterodimer, but we never obtained this construct, most likely due to toxic effects on cells. Similar difficulties were encountered during construction of other LSm protein expression vectors. We subcloned LSm4 and LSm1 as a monocistrons. We cloned LSm6, LSm5, LSm8 in the pQE30 vector as monocistrons for comparison (yield, solubility) with polycistronic constructs.

Monocistronic constructs were in general subcloned into pUC19.3 vector. In this cloning vector, the intergenic region linker was omitted. Its main features are a His6-tag and a TEV cleavage site at the 5' end of the construct. The genes of interest were subcloned between the single NdeI and HindIII sites of the vector (Figure 3.1.5).

Figure 3.1.5. represents the pUC19.3 cloning intermediate in construction of single cistrons and Figure 3.1.6. the His6-TEV-LSm1 single cistron expression construct. Vector maps for the other single expression vectors are shown in the Materials and Methods part and the cloning strategy was the same in all described cases.

Monocistronic LSm6, subcloned in to the pQE30 vector, was insoluble when expressed under conditions which we used to express most of our constructs. In consequence, we chose another approach for single cistrons, subcloning them into the pQE70.2Z.3 T5 vector, which contains two
Z tags (corresponding to the IgG binding domain of *Staphylococcus aureus* protein A). The Z-tags have the capacity to greatly increase solubility of a downstream fused ORF. The cloning intermediate was pUC19.6 (Figure 3.1.7.) pQE70.ZZ vector, obtained from Prof. U. Kutay (ETH Zurich), was modified by introducing a His6-TEV cleavage site segment after the two Z tags. Single cistrons were introduced into this cloning intermediate using BamHI/HindIII restriction sites and transferred as an EcoRI/HindIII cassette into the pQE70.ZZ.3 vector. Indeed, single cistrons were produced in higher amounts, but with tri- and di-cistronic constructs we were faced with the problem of non-stoichiometrical production. Generally, the pQE70.ZZ.3 vector was used during this work for large amount production of single-cistrons LSm proteins. The produced proteins could be purified by IMAC chromatography. The obtained fusion proteins can potentially be used for coimmunoprecipitation experiments. For single cistrons cloned into pQE70.ZZ.3, the same strategy was applied. The cloning maps are in the Materials and Methods part of the thesis.

**Figure 3.1.7.** Map of pUC19.6

**Figure 3.1.8.:** ZZ-His6-TEV-LSm4:pQE70ZZ.3
<table>
<thead>
<tr>
<th>Number</th>
<th>LSm protein</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LSm1</td>
<td>His6-TEV-LSm1:pQE30</td>
</tr>
<tr>
<td>2</td>
<td>LSm2,3</td>
<td>His6-TEV-LSm2/3:pQE30</td>
</tr>
<tr>
<td>3</td>
<td>LSm3,2</td>
<td>His6-TEV-LSm3/2:pQE30</td>
</tr>
<tr>
<td>4</td>
<td>LSm3,5</td>
<td>His6-TEV-LSm5/3:pQE30</td>
</tr>
<tr>
<td>5</td>
<td>LSm4,8</td>
<td>His6-TEV-LSm4/8:pQE30</td>
</tr>
<tr>
<td>6</td>
<td>LSm4,7</td>
<td>His6-TEV-LSm4/7:pQE30</td>
</tr>
<tr>
<td>7</td>
<td>LSm4,2</td>
<td>His6-TEV-LSm4/2:pQE30</td>
</tr>
<tr>
<td>8</td>
<td>LSm5,6</td>
<td>His6-TEV-LSm5/6:pQE30</td>
</tr>
<tr>
<td>9</td>
<td>LSm5,7</td>
<td>His6-TEV-LSm5/7:pQE30</td>
</tr>
<tr>
<td>10</td>
<td>LSm6,7</td>
<td>His6-TEV-LSm6/7:pQE30</td>
</tr>
<tr>
<td>11</td>
<td>LSm5,6,7</td>
<td>His6-TEV-LSm5/6/7:pQE30</td>
</tr>
<tr>
<td>12</td>
<td>LSm4</td>
<td>His6-TEV-LSm4:pQE30</td>
</tr>
<tr>
<td>13</td>
<td>LSm6</td>
<td>His6-TEV-LSm6:pQE30</td>
</tr>
<tr>
<td>14</td>
<td>LSm8</td>
<td>His6-TEV-LSm8:pQE30</td>
</tr>
<tr>
<td>15</td>
<td>LSm5</td>
<td>His6-TEV-LSm5:pQE30</td>
</tr>
<tr>
<td>16</td>
<td>LSm1</td>
<td>His6-TEVLSm1:pQE70.ZZ.3</td>
</tr>
<tr>
<td>17</td>
<td>LSm4</td>
<td>His6-TEVLSm4:pQE70.ZZ.3</td>
</tr>
<tr>
<td>18</td>
<td>LSm6</td>
<td>His6-TEVLSm6:pQE70.ZZ.3</td>
</tr>
<tr>
<td>19</td>
<td>LSm8</td>
<td>His6-TEVLSm8:pQE70.ZZ.3</td>
</tr>
<tr>
<td>20</td>
<td>LSm5/6/7</td>
<td>His6-TEVLSm5/6/7:pQE70.ZZ.3</td>
</tr>
<tr>
<td>21</td>
<td>LSm3/6</td>
<td>His6-TEV-LSm3/6:pQE30</td>
</tr>
</tbody>
</table>

Table 3.1: List of LSm protein expression constructs which were made and used for expression trials.

3.2. Protein expression and purification

The previous chapter describes the cloning strategy which was used during the course of this work. Most of the LSm proteins were subcloned as heterodimers and heterotrimers, with only the first cistron bearing a His6-tag. The other overexpressed proteins of interest were co-purified via complex formation with the protein bearing the His6-tag. This suggests that the sub-complexes produced are in general very stable. In all cases, the purified LSm protein subcomplexes are high salt tolerant (the first affinity chromatography was performed in 0.5 M, or even 1 M NaCl). From the experiences we have had with LSm protein expression, we can recommend that purification of those LSm proteins which show the highest tendency to aggregate (LSm4/8 and singly
expressed LSm proteins) should be conducted under semi denaturing conditions (4 M urea), although during the course of this work this was not yet the case.

In general, the first purification step was metal ion affinity chromatography, followed where needed by anion or cation exchange chromatography. The identity of the produced proteins was confirmed by MALDI TOF (Protein service laboratory, Institute for Molecular Biology and Biophysics, ETH Zürich), Western blot using anti-penta-His antibodies, and N-terminal sequencing. The oligomeric state of purified species was determined using calibrated gel filtration chromatography columns, and, in some instances, by combination with static light scattering measurements, or, independently, by analytical ultracentrifugation (AUC).

Our overexpressed LSm proteins have His6-tags, followed by a TEV cleavage site. After purification, some of the sub-complexes were subjected to TEV protease cleavage in order to remove the His6-tag. The TEV protease bears a His6-tag as well, so separation of protein (sub-complexes) released from the His6-tag, the TEV protease and the non-cleaved fraction of proteins can in general be achieved by IMAC.

In the same vein as the LSm protein tendency to aggregate, cleavage of the N-terminal His6-tags proved difficult. His6-tags from most of the constructs turned out to be non-cleavable, or the separation of the cleaved from the non-cleaved material was incomplete even when chaotropes were included in the purification. In rare cases (LSm5/6/7 and LSm6) cleavage was complete.

The following sections describe the expression, purification and characterization of the produced LSm proteins.

3.2.1. Expression and purification of His6-TEV-LSm2/3

After verifying the expression vector by sequencing for each respective construct, expression parameters were optimized. The generic expression screen tested two different cell strains, M15 [pREP4] and SG13009 [pREP4], one temperature (25 °C) and an induction time of ~20 hours. Lauria Broth (LB) medium was used in these test trials. We did not observe any production differences between the two tested cell strains on SDS PAGE gels (data not shown). These first chosen expression parameter settings turned out to be beneficial for protein production, hence we did not change them any more. For the His6-TEV-LSm 2/3 heterodimer, we chose 20 hours of induction at 25 °C, LB medium supplemented with the appropriate antibiotics and SG13009
[pREP4] as the expressor strain for optimal expression parameter values. According to the parameter settings described above, 6 x 11 of LB cultures were grown. Cells were harvested and disrupted as described in the Materials and Methods part of the thesis. The cleared lysate was subjected to the Affinity Chromatography step. Most of the His6-TEV-LSm2/3 heterodimer eluted in the 250 mM Imidazole fraction. His6-TEV-LSm2/3 heterodimer was eluted in 60 mM Imidazole elution fractions, but the sub-complex was not sufficiently pure. A second purification step was anion exchange chromatography conducted as described in the Materials and Methods part. The heterodimer was subjected to a salt gradient, and eluted at \( \sim 300 \) mM NaCl concentration.

![Figure 3.2.1. Affinity chromatography of His6-TEV-LSm2/3](image)

![Figure 3.2.2.: Anion exchange chromatography of His6-TEV-LSm2/3](image)

79
Figure 3.2.3. SDS PAGE of all purification steps for His$_6$-TEV-LSm2/3:

M: Marker, Lane1: Uninduced sample, Lane2: total cell extract after 20 hours of induction, Lane3: Supernatant after ultracentrifugation, Lane4: Pellet after ultracentrifugation, Lane5: LSm2/3 heterodimer after Affinity chromatography, Lane6: LSm2/3 heterodimer after anion exchange chromatography

After each purification step, samples were analyzed on SDS PAGE. Fig 3.2.3 shows a gel which summarizes all expression and purification steps.

After the final anion exchange chromatography, a Western blot was performed with the purified protein to confirm the identity of His$_6$-tagged and non-tagged LSm protein. It is known that many non-related bacterial proteins are capable to attach to the IMAC columns.

Figure 3.2.4. SDS PAGE (left) and corresponding Western blot (right) with anti-penta His.

The lower band corresponds to His$_6$-tagged LSm2, although His$_6$-TEV-LSm2 has a higher molecular weight than LSm3. For the SDS PAGE molecular weight marker (10-150 kDa) was used, and for the Western Blot, a molecular weight marker in the same range (10-150kDa) with His$_6$ tagged proteins.
Figure 3.2.5. MALDI TOF of the His6-TEV-LSm2/3 heterodimer.

The nominal molecular weight of His6-TEV-LSm2 is 13171Da. The LSm3 nominal molecular weight is 11844Da. The difference in calculated molecular weight and the one obtained can be due to processing of the leading Methionine residue. The second amino acid after methionine in the LSm3 sequence is alanine. This would fit with the concept of the N-end rule, where the presence of a smaller amino acid after the first Methionine would favor Methionine processing [164].

After the last purification step, the yield of pure His6-TEV-LSm2/3 was 10 mg/l of bacterial culture. The purified sub-complex was characterized biophysically, as described in section 3.2.15.
3.2.2. Expression and purification of His6-TEV-LSm4/8

We established expression conditions for expressing large amounts of His6-TEV-LSm4/8. Three cell strains (SG13009 [pREP4], M15 [pREP4], and BLR [pREP4] were tested, and different expression times (2 hours after induction, 4 hours and overnight induction). As judged from the SDS PAGE gels, the strain BLR [pREP4] produces the highest amounts of the desired proteins. We chose to induce the cells for 4 hours because at that time point the background, i.e. contaminating bacterial proteins were not yet very prominent.

Cell harvesting, disruption, and preparation of the cleared lysate for affinity chromatography were identical to those employed for purification of the LSm2/3 heterodimer and are described in more details in the Materials and Methods part of the thesis. The LSm4/8 heterodimer was purified to near homogeneity by affinity chromatography, as judged by SDS PAGE (Figure 3.2.7). However, the same SDS PAGE gel, Lane 6, figure 3.2.7, shows that the second protein (LSm8) is to some extent present in the non-bound fraction. This indicates that the LSm4/8 heterodimer is not as stable as LSm2/3. In conclusion, we conducted the affinity chromatography step at slower flow rates (0.5 ml/min) than in the usual procedure (see Materials and Methods). Interestingly, when we tried to further purify the fractions eluting at 60 mM imidazole using anion exchange chromatography, the sub-complex precipitated, because it is low salt intolerant. Hence IMAC remained the only purification procedure for this heterodimer.

Figure 3.2.6: Affinity purification of His6-TEV-LSm4/8
Figure 3.2.7. SDS PAGE gel of all purification steps for His6-TEV-LSm4/8:

M: Marker, Lane 2: Non-induced sample, Lane 3: total cell extract after 4 hours of induction, Lane 4: Supernatant after ultracentrifugation, Lane 5: pellet after ultracentrifugation, Lane 6: Wash, Lane 7: 60 mM Imidazole elution, Lane 8: 250 mM Imidazole elution, Lane 9: 500 mM Imidazole elution.

The yield of purified His6-TEV-LSm4/8 heterodimer was 8 mg/l of bacterial culture.

Figure 3.2.8: SDS PAGE (left) and Western blot (right) of purified His6-TEV-LSm4/8

For this experiment, two different SDS PAGE gels were used. This is the reason why on the Western, the His6-tagged LSm4 appears to run in roughly the same position as the upper band in the Coomassie-stained gel. However, this correspondence is not kept in the case of LSm2/3 and LSm5/6/7 (Figures 3.2.4 and 3.2.18).
The molecular weights of proteins detected correspond to those expected for His$_6$-TEV-LSm4 and LSm8. His$_6$-TEV-LSm4/8 was expressed in high amounts and purified to homogeneity. The His$_6$-tag on LSm4 can be cleaved which was not the case for the His$_6$-TEV-LSm2/3 heterodimer, but separation of the cleaved from the non-cleaved part of the protein was not possible, probably due to aggregation problems. For complete cleavage, 8 days at room temperature were required, which increases the possibility for protease contaminations.

3.2.3. Expression and purification of His$_6$-TEV-LSm5/6/7

Several expression parameters were tried for expression of His$_6$-TEV-LSm5/6/7. Cell strains tested were SG13009 [pREP4], BLR [pREP4], and BLR [pREP4] RIL. Four different temperatures were tested, 18°C, 25°C, 28°C, and 37°C. Three different induction times were tested, 4, 24, and 52 hours. In each combination all three proteins were produced, but using
standard expression conditions, i.e. 20 hours at 25°C in SG13009 [pREP4], LSM6 was underrepresented. When LSM5/6/7 was overexpressed using the described conditions, cells were disrupted as listed in the Materials and Methods part of thesis, and cleared lysate was applied to an IMAC column. Possible explanations could be that LSM6 does not very strongly bind to the other two LSM proteins LSM5 and LSM7 and that it elutes in the unbound fraction of the IMAC. Another possibility is that LSM6 is mainly insoluble under such conditions and remains in the cell pellet after ultracentrifugation, but LSM6 was not found in the pellet in larger amount than the other two proteins.

![Figure 3.2.10. Affinity chromatography of the His6-TEV-LSM5/6/7 expressed under non-optimized conditions (SG13009 [pREP4] cell strain, 25 °C, 20 hours of induction).](image)

![Figure 3.2.11. SDS PAGE gel which corresponds to the elution from IMAC (Figure 3.2.10)](image)

From Figures 3.2.10 and 3.2.11, as judged by eye the fractions E4 and E5 that correspond to the 250 mM Imidazole elution and are concentrated in the sharp peak (see chromatogram, Figure

85
3.2.10), contain all three proteins present in almost stoichiometrical amounts. These fractions were pooled and dialyzed against low salt buffer (150 mM NaCl). His6-TEV-LSm5 and LSm7 have isoelectrical points of 5.2 and 5.11, respectively, and LSm6 is basic (pI = 9.6). The conclusion from these values is that a His6-TEV-LSm5/6/7 sub-complex that contains all three proteins in equal amounts should bind less strongly to the anion exchange column, while the fraction of His6-TEV-LSm5/6/7 containing substoichiometrical amounts of LSm6 should be found in the gradient fraction of the anion exchange chromatography.

![Anion exchange chromatography of fractions from IMAC containing stoichiometrical His6-TEV-LSm5/6/7](image)

**Figure 3.2.12.** Anion exchange chromatography of fractions from IMAC containing stoichiometrical His6-TEV-LSm5/6/7

![SDS PAGE with fractions from anion exchange chromatography](image)

**Figure 3.2.13.** SDS PAGE with fractions from anion exchange chromatography (Figure 3.2.12)

Most of the stoichiometrical His6-TEV-LSm5/6/7 is in the unbound fraction from the anion exchange chromatography. The excess of overproduced components (LSm5 and LSm7) is

86
binding stronger to the anion exchange column. Unbound fractions from the anion exchange chromatography were applied to a gel filtration column, for determination of oligomeric state.

![Graph](image)

Figure 3.2.14. Gel filtration chromatography of unbound fraction from the Anion exchange

The elution volume of 9.8 ml would correspond to a MW of 55.3 kDa. The calculated MW of the heterotrimer is 33 kDa. This was the first indication that His\textsubscript{6}-TEV-LSm5/6/7 forms higher order structures.

This was one approach to overcome difficulties of non stoichiometrical expression of His\textsubscript{6}-TEV-LSm5/6/7 under standard expression conditions which were beneficial for most of the other LSm proteins.

After extensive parallel screening for optimal expression conditions for all three proteins, we established that BLR [pREP4], at 37°C, for 48 hours of induction can overproduce all three proteins in equal amounts.

The heterotrimer was purified to homogeneity using a combination of affinity chromatography and anion exchange chromatography. The final yield is ~14 mg per liter of bacterial culture. Harvesting of the cells, disruption, and the two chromatographies were performed as described in the Materials and Methods part of the thesis.
Figure 3.2.15: Affinity chromatography of His$_6$-TEV-LSm5/6/7 under optimized conditions

Figure 3.2.16: Anion exchange chromatography of His$_6$-TEV-LSm5/6/7 after affinity chromatography
Figure 3.2.17: SDS PAGE of all purification step His$_6$-TEV-LSm5/6/7 under optimized conditions:

M: Marker, Lane1: Non-induced sample, Lane2: total cell extract after 48 hours of induction, Lane3: supernatant after ultracentrifugation, Lane4: pellet after ultracentrifugation, Lane5: Wash, Lane6: 60 mM Imidazole elution, Lane7: 250 mM Imidazole elution. Lane9: Anion exchange step.

Figure 3.2.18 SDS PAGE (right) and Western blot (left) using anti-penta-His antibodies on purified His$_6$-TEV-LSm5/6/7.

Only the first cistron (LSm5) bears a His$_6$-tag. LSm6 and LSm7 are co-purified by formation of a stable trimeric complex. His$_6$-TEV-LSm5 has a higher molecular weight (12289Da) than LSm7.
but migrates lower than LSm7 on the SDS Page gel. For this experiment, two different SDS PAGE gels were used. This is the reason why on the Western, the His6-tagged LSm5 appears to run in roughly the same position as the upper band in the Coomassie-stained gel.

Figure 3.2.19: Ni-affinity chromatography of cleaved His6-TEV-LSm5/6/7 sub complex.

Purified His6-TEV-LSm5/6/7 heterotrimer was subjected to Tobacco etch virus protease (TEV) protease, which cleaves off the His6-tag. Cleavage was performed for 24 hours at room temperature, and the preparation subjected to a second affinity chromatography. Cleaved LSm5/6/7 lacking the His6-tag eluted in the unbound fraction. The non-cleaved fraction of His6-TEV-LSm5 and the (His6-tagged) TEV protease eluted in the imidazole gradient.
Figure 3.2.20. SDS PAGE of all purification step His<sub>6</sub>-TEV-LSm5/6/7.

Marker, Lane 1, Uninduced sample, Lane 2: total cell extract after 48 hours of induction, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: IMAC purified His<sub>6</sub>-TEV-LSm5/6/7, Lane 6: ALEX purified His<sub>6</sub>-TEV-LSm5/6/7, Lane 7: His<sub>6</sub> tag cleaved IMAC purified LSm5/6/7.

3.2.4. Expression and purification of His<sub>6</sub>-TEV-LSm<sub>1</sub>

His<sub>6</sub>-TEV-LSm<sub>1</sub> was produced from a monocistronic expression vector and overexpressed in the strain SG13009 [pREP4]. After induction with IPTG, cells were grown 20 hours at 25°C. After harvesting, the cells were disrupted as described in Materials and Methods. The first step was an affinity chromatography followed by anion exchange chromatography according to the procedure explained in the Materials and Methods part of this thesis. Size exclusion chromatography was used to determine the oligomeric state of protein. The elution volume from calibrated size exclusion chromatography indicates formation of aggregates of non-defined oligomeric state. An elution volume 8.68 ml would correspond to an apparent molecular weight of more than 400 kDa. The nominal molecular weight of His<sub>6</sub>-TEV-LSm<sub>1</sub> is 17776.12 Da.
Figure 3.2.21: Affinity purification of His\textsubscript{6}-TEV-LSm1

Figure 3.2.22: Anion exchange chromatography of His\textsubscript{6}-TEV-LSm1

Figure 3.2.23: SDS PAGE of His\textsubscript{6}-TEV-LSm1 purification:
M: Marker, Lane 1: Uninduced sample, Lane 2: total cell extract after 20 hours, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: His\textsubscript{6}-TEV-LSm1 after IMAC, Lane 7: Anion exchange chromatography
3.2.5. Expression and purification of His6-TEV-LSm4

His6-TEV-LSm4 was obtained using a monocistronic expression vector. The cell strain used for expression was SG13009 [pREP4] at 25°C for 20 hours induction. Centrifugation of the cells and disruption was like described for the other constructs. Protein of sufficient purity was obtained.
after one affinity chromatography step. His$_6$-TEV-LSm4 is only stable in the presence of 1 M or higher NaCl, and forms higher order aggregates as judged by size exclusion chromatography.

Figure 3.2.26: SDS Page (left) and Western blot (right) of purified His$_6$-TEV-LSm4.

For this experiment, two different SDS PAGE gels were used. This is the reason why on the Western, the His$_6$-tagged LSm4 appears to run upper than in Coomassie-stained gel.

Figure 3.2.27: SDS PAGE of expression and purification of His$_6$-TEV-LSm4.

M: Marker, Lane 1: Uninduced sample, Lane 2: total cell extract after 20 hours, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: His$_6$-TEV-LSm4 after affinity chromatography

Figure 3.2.28: His$_6$-TEV-LSm4 on size exclusion chromatography
3.2.6. Expression and purification of His\textsubscript{6}-TEV-LSm8

His\textsubscript{6}-TEV-LSm8 was obtained using a monocistronic expression vector, whose map is depicted in the Materials and Methods part of the thesis. The cell strain used for expression was SG13009 [pREP4] at 25°C for 20 hours. Protein of sufficient purity was obtained after one affinity chromatography step. His\textsubscript{6}-TEV-LSm8 forms higher order aggregates of non-defined oligomeric state as judged by size exclusion chromatography.

![Figure 3.2.29: SDS PAGE (right panel) and Western of His\textsubscript{6}-TEV-LSm8 (left panel)](image)

**Figure 3.2.29:** SDS PAGE (right panel) and Western of His\textsubscript{6}-TEV-LSm8 (left panel)

![Figure 3.2.30: SDS PAGE of expression and purification of His\textsubscript{6}-TEV-LSm8. M: Marker, Lane 1: Uninduced sample, Lane 2: total cell extract after 20 hours, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: 60 mM Imidazole elution, Lane 7: 250 mM Imidazole elution](image)

**Figure 3.2.30:** SDS PAGE of expression and purification of His\textsubscript{6}-TEV-LSm8. M: Marker, Lane 1: Uninduced sample, Lane 2: total cell extract after 20 hours, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: 60 mM Imidazole elution, Lane 7: 250 mM Imidazole elution
3.2.7. Expression and purification of His$_6$-TEV-LSm5

His$_6$-TEV-LSm5 was expressed from a monocistronic vector in the SGI3009 [pREP4] cell strain as soluble protein. Cells were harvested 20 hours after induction with IPTG at 25°C. Disruption and purification was the same as described in the previous sections.

Figure 3.2.32: SDS PAGE expression and purification of His$_6$-TEV-LSm5.

Marker, Lane 1: Uninduced sample, Lane 2: total cell extracts after 20 hours, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: 60 mM Imidazole elution, Lane 7: 250 mM Imidazole elution
3.2.8. Expression and purification of His₆-TEV-LSm6

A His₆-TEV-LSm6 expression trial was conducted using the same procedure as for LSm5 and other single cistrons. The result (as shown on the SDS PAGE gel, Figure 3.2.33.) was insoluble protein, present in the pellet after ultracentrifugation. In the 250 mM Imidazole elution, a small amount of soluble LSm6 was observed.

![Figure 3.2.33: SDS PAGE expression and purification of His₆-TEV-LSm6](image)

M: Marker, Lane 1: Uninduced sample, Lane 2: total cell extract after 20 hours, Lane 3: Supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: 60 mM Imidazole elution, Lane 7: 250 mM Imidazole elution

3.2.9. Expression and purification of ZZ-His₆-TEV-LSm6

LSm6 was obtained by overexpression from a monocistronic pQE70ZZ.3 vector (see chapter 3.1.). The cell strain used for expression was SG13009 [pREP4] at 25°C for 20 hours. Protein of sufficient purity was obtained after one affinity chromatography step followed by cation exchange chromatography. Cell harvesting, disruption, and two chromatographical steps proceeded as described in the Materials and Methods part. TEV protease was used to remove the ZZ and His₆ tags. Separation of the LSm6 from the ZZ-His₆-tag was carried out by cation exchange chromatography; ZZ-His₆ remains in an early part of the gradient (see Figure 3.2.36). Pure LSm6 elutes as the second peak. When we tried to determine its oligomeric state, the protein precipitated on the gel filtration column. The yield of cleaved LSm6 from one liter of bacterial culture was ~10 mg.
Figure 3.2.34: Affinity chromatography for ZZ-His<sub>6</sub>-TEV-LSm6

Figure 3.2.35: Cation exchange purification step for ZZ-His<sub>6</sub>-TEV-LSm6

Figure 3.2.36: Cation exchange chromatography step after TEV protease cleavage.

Uncleaved protein, the ZZ-His<sub>6</sub>-tag and TEV protease are found in the unbound fraction or in the broad peak in the early gradient (36.91 min). The peak at 39.75 min with corresponding fractions J6 and J5 contains purified, cleaved LSm6.
3.2.10. Expression and purification of His$_6$-TEV-LSm5/6

The LSm5/6 heterodimer was expressed in SG13009 [pREP4] cell strain for 20 hours at 25 °C after induction using 1 mM IPTG. Cells harvesting, disruption, and first affinity purification were the same as for the other constructs described in the previous chapters. After an affinity purification step, the sub-complex is sufficiently pure. Like the other LSm sub-complexes, it forms defined higher order structures, namely hexamers, as estimated by size exclusion chromatography. Interestingly, the LSm5/6 interaction is observed in a yeast two hybrid screen [124]. However, an interaction between SmE and SmF, the paralogues of LSm5 and LSm6 was observed in coimmunoprecipitation assays [73].
Figure 3.2.38. (a): SDS PAGE expression and purification of His6-TEV-LSm5/6

M: Marker, Lane 1: Uninduced sample, Lane 2: Supernatant after ultracentrifugation, Lane 3: pellet after ultracentrifugation, Lane 4: Unbound fraction from IMAC, Lanes 5 and 6: 250 mM Imidazole elution

Figure 3.2.38 (b) Size exclusion chromatography profile on a Superdex 200 HR 10/30 column.
Purified LSm5/6 heterodimer does not run as monomer. Similarly to other heterodimers of LSm proteins, it forms higher order structures, in this case of a molecular weight of 54 kDa (the nominal molecular weight of His6-TEV-LSm5/6 is 20kDa).

3.2.11. Expression and purification of His6-TEV-LSm5/3

His6-TEV-LSm5/3 which based on the assumed order of subunits in the LSm rings is not a nearest neighbour pair (Figure 3.1.1.) was expressed and purified using strain SG13009 [pREP4], for 16 hours at 25° C. After one affinity purification step, the protein is sufficiently pure. Judged by SDS PAGE gel both gene products are expressed in a stoichiometrical and soluble format, c.f. lanes 3 and 4 (Figure 4.39). LSm5/3 is a new interaction between LSm proteins, as observed in this work, but not detected in the yeast two hybrid assay [124]. As well the paralogues, the canonical D2 and E proteins were never coexpressed in E. coli, or in any other expression system in a soluble format.
Figure 3.2.39. (a): SDS PAGE expression and purification of His<sub>6</sub>-TEV-LSm5/3

M: Marker, Lane 1: Uninduced sample, Lane 2: total cells before disruption Lane 3: supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lane 6: 60 mM Imidazole elution from affinity chromatography, Lane 7: 250 mM Imidazole elution

Figure 3.2.39. (b): Size exclusion chromatography profile on Superdex 200 HR 10/30.

The purified LSm5/3 heterodimer does not run as a monomer. Similarly to other LSm protein heterodimers, it forms higher order structures, whose apparent molecular weight in this case is 62 kDa. The nominal molecular weight of the His<sub>6</sub>-TEV-LSm5/3 heterodimer is 25kDa.

3.2.12. Expression and purification of His<sub>6</sub>-TEV-LSm4/7

His<sub>6</sub>-TEV-LSm4/7 (representing a pair of assumed nearest neighbours in the heteroheptameric LSm2-8 ring, Figure 3.1.1.) was expressed and purified like described in previous sections. After one affinity purification step, the complex is pure. An LSm4/7 interaction was originally identified in a yeast two hybrid assay [124] and confirmed with this experiment. There is no experimental evidence that their paralogues D3 and G are capable to form stable heterodimers. In analogy to other LSm heterodimers, it forms higher order structures as judged by size exclusion chromatography. His<sub>6</sub>-TEV-LSm4/7 on calibrated gel filtration chromatography runs at an
elution volume which would suggest formation hexamers. The nominal molecular weight of the heterodimer is 28 kDa.

Figure 3.2.40 (a): SDS PAGE expression and purification of His<sub>6</sub>-TEV-LSm4/7
M: Marker, Lane 1: supernatant after ultracentrifugation Lane 2: pellet after ultracentrifugation, Lane 3: Unbound fraction from IMAC, Lane 4: 60 mM Imidazole elution from Affinity, Lane5: 250 mM Imidazole elution, Lane6: 500 mM Imidazole elution

Figure 3.2.40 (b) Size exclusion chromatography profiles on Superdex 200 HR 10/30.

3.2.13. Expression and purification of His<sub>6</sub>-TEV-LSm4/2 and His<sub>6</sub>-TEV-LSm6/7

His<sub>6</sub>-TEV-LSm 6/7 was an assumed pair of non nearest neighbours in the heteroheptameric LSm2-8 complex (Figure 3.1.1.). It was expressed and purified like the other constructs described in the previous sections. An LSm6/7 interaction was not identified in the yeast two hybrid assay [124]. There is no evidence that their paralogues F and G are capable of making stable heterodimers [73]. The LSm6/7 heterodimer forms higher order structures. The main peak in gel filtration on a Superdex 200 HR10/30 column (calibration curve is in Materials and Methods part) at 13.73 ml would correspond to a molecular weight of 92 kDa (The nominal molecular weight of the heterodimer is 22 kDa). The smaller peak at an elution volume of 18.08 ml would correspond to a molecular weight of 10 kDa, which roughly fits with the molecular weight of the
individual subunits. Fractions from gel filtration chromatography were applied on the SDS PAGE gel but none of them contain individual proteins. Judged by eye, the ratio between two proteins is the same in all fractions.

Figure 3.2.41. (a) SDS PAGE expression and purification of His6-TEV-LSm6/7
M: Marker, Lane 1: Uninduced sample, Lane 2: total cells before disruption, Lane 3: supernatant after ultracentrifugation, Lane 4: pellet after ultracentrifugation, Lane 5: Unbound fraction from IMAC, Lanes 6: 60 mM Imidazole elution from affinity chromatography, Lane 7: 250 mM Imidazole elution

Figure 3.2.41. (b) Size exclusion chromatography profile on Superdex 200 HR 10/30.

His6-TEV-LSm4/2 was expressed and purified in cell stain SG13009 [pREP4] for 20 hours at 25°C. An LSm4/2 interaction was not identified in the yeast two hybrid assays. In this work, His6-TEV-LSm4/2 was used as one of the negative controls (i.e. as assumed non-nearest neighbour in the LSm2-8 ring, figure 3.1.1.). There is no evidence that their paralogues D3 and D2 are capable to make stable heterodimer. The oligomeric state of this heterodimer was not assessed during the course of this work.
3.2.14. Expression and purification of His6-TEV-LSm5/7

His6-TEV-LSm5/7 was expressed and purified using standard conditions, i.e. expression in SG13009 [pREP4], for 20 hours at 25°C. After one affinity purification step, the protein was sufficiently pure. There is no evidence from the coimmunoprecipitation assays that their paralogues: SmE and SmG are interacting [73].
3.2.15. Biophysical characterization of the His6-TEV-LSm2/3, His6-TEV-LSm4/8 and His6-TEV-LSm5/6/7

The oligomeric state of the three sub-complexes was addressed using gel filtration chromatography. Determination of the exact molecular weight and polydispersity was carried out using static light scattering and analytical ultracentrifugation. The shape and size of sub-complexes was addressed by electron microscopy, in collaboration with the group of Prof. A. Engel, Biozentrum Basel. Reversed phase chromatography was used for determining the stoichiometry of the sub complexes.

3.2.15.1. His6-TEV-LSm2/3 characterization

As judged by eye, both LSm2 and LSm3 proteins were produced and purified in equimolar amounts (Figure 3.2.3). This was confirmed by reversed phase chromatography. The fractions corresponding to the two observed main peaks were transferred to MALDI TOF for identification. The peak eluting at 55.20% of mobile phase B is LSm2, and the peak eluting at 57.20% of mobile phase B is LSm3. After integration of the peak areas (signal at 280nm) and dividing by the number of Tyrosine and Tryptophane residues, we came to the conclusion that both proteins were produced and purified in equal amounts. This indicates that they make a stable heterodimeric complex which can withstand stringent purification steps.
Figure 3.2.44.: Reversed phase chromatography of the His6-TEV-LSm2/3 heterodimer.

The purified heterodimer was concentrated up to 7 mg/ml, and subjected to gel filtration chromatography according to the procedure described in the Materials and Methods part. The heterodimer elutes as a single, Gaussian peak at 14.13 ml. The elution profile indicates sample homogeneity and formation of higher order structures. According to calibration with proteins of known molecular weight, this elution volume corresponds to a molecular weight of 66 kDa. The molecular weight of the heterodimer is 25 kDa. This GPC result was the first indication that LSm2/3 forms higher order structures. Within the error of estimating the MW of a macromolecular species by calibrated GPC (about 15%), the deduced value for LSm2/3 would thus correspond to a hexamer (i.e. a trimer of dimers).
Next, in collaboration with the group of Prof. A. Engel at the Biozentrum, University of Basel we took negative stain electron micrographs of our heterodimer. These reveal a homogeneous population of ring-shaped particles. Class averages were created from 641 particles, from an initial data set containing 1000 particles of LSm2/3. LSm2/3 (Figure 3.2.46.) shows up as ring-shaped structures with the central hole, as published for the native LSm 2-8 heteroheptamer isolated from HeLa cells [99].

After particle classification and subsequent class averaging, distinct ring particles can be observed in the LSm2/3 galleries (Figure 4.46.a, bottom) having an outer diameter of ~7 nm and a cavity of less than 1.5 nm. Already from the negative stain electron micrographs, we have an indication that on first approximation we have homogeneous population with small deviations. (Electron Microscopy work was carried out in Biozentrum Basel, Group. Prof. A. Engel, by Dr.M. Chami)
Figure 3.2.46: Negative-stain electron micrographs and picture gallery of the His6-TEV-LSm2/3 heterodimer

The His6-TEV-LSm2/3 sub complex was subjected to an equilibrium sedimentation run in an OPTIMA XL-A analytical ultracentrifuge at 12000 rpm. (Roche, Basel). In the analytical ultracentrifuge, the LSm2/3 heterodimer distribution is bimodal at 10μM concentration, containing a hexamer (10%) and octamer (87%). The LSm2/3 sub complex was run on a gel filtration column coupled with mini DAWN static light scattering analyzer and a refractometer (Wyatt, USA). Data were analyzed using Wyatt’s ASTRA 4 software.

<table>
<thead>
<tr>
<th>c</th>
<th>Losses</th>
<th>trimer</th>
<th>tetramer</th>
<th>pentamer</th>
<th>nonamer</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 μM</td>
<td>36%</td>
<td>0%</td>
<td>68%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>10 μM</td>
<td>7%</td>
<td>10%</td>
<td>87%</td>
<td>0%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 3.2.1. Analytical ultracentrifugation distribution for LSm2/3

<table>
<thead>
<tr>
<th>complex</th>
<th>Average MW over peak</th>
<th>Standard deviation</th>
<th>Polydispersity</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSm2/3</td>
<td>86 kDa</td>
<td>2.0%</td>
<td>3.0%</td>
</tr>
</tbody>
</table>

Table 3.2.2. His6-TEV-LSm2/3 sub-complex static light scattering analysis results

Considering the mass of the LSm2/3 heterodimer (25 kDa), from the analytical ultracentrifugation results and static light scattering result it can be suggested that the
predominant oligomeric state of the His6-TEV-LSm2/3 heterodimer is octameric LSm2/3 ([LSm2/3]₄), which cannot be inferred from the negative stain electron micrographs alone.

3.2.15.2. His6TEV LSm4/8 characterization

The purified His₆-TEV-LSm₄/₈ heterodimer was concentrated to 8 mg/ml, and subjected to gel filtration chromatography for determining the oligomeric state. The column used was the same as in section 3.2.15.1, but the running buffer contained 1 M NaCl, because of the low salt intolerance of His₆-TEV-LSm₄/₈ as mentioned above. On calibrated gel filtration chromatography His₆-TEV-LSm₄/₈ elutes at 9.06 ml, which would correspond to a molecular weight of 400 kDa. The nominal molecular weight of the heterodimer is 28 kDa (Figure 3.2.47). The formation of aggregates was confirmed by negative stain electron micrographs (Figure 3.2.48).

Figure 3.2.47: Size exclusion chromatography of His₆-TEV-LSm₄/₈ heterodimer.
Figure 3.2.48. Negative stain electron micrographs of His6-TEV-LSm4/8 heterodimer

3.2.15.3. His6TEV LSm5/6/7 characterization

Figure3.2.49: RPC of His6-TEV-LSm5/6/7
Reversed phase chromatography was used as a technique that can indicate whether all three proteins are present in a stoichiometrical ratio. Three separated peaks were collected and individually subjected to MALDI TOF for identification. The peak that elutes at 22.292 minutes (54.20% of mobile phase B) is LSm6; the molecular weight corresponds to the expected value. The peak that elutes at 22.453 minutes (54.40% of mobile phase B) is LSm7. The molecular weight determined for this fraction is correct for this species. The last peak eluting at 27.879 minutes (72.60% mobile phase B) was manually collected and corresponds to His6-TEV-LSm5. The areas under each peak were integrated and divided with the number of Trp and Tyr residues for each of the proteins. The resulting numbers are 1, 0.8, and 1.1 which within the error of the method indicate that His6-TEV-LSm5/6/7 heterotrimer after purification using optimized conditions contains all three proteins expressed in equimolar amount.

**Figure 3.2.50.:** Sample1 from RPC is LSm6
Figure 3.2.51: Sample 2 from RPC is LSm7

Figure 3.2.52: Sample 3 from RPC is His6 LSm5
Figure 3.2.53: Size exclusion chromatography of His6-TEV-LSm5/6/7

On calibrated size exclusion column, purified His6-TEV-LSm5/6/7 heterotrimer elutes at an elution volume corresponding to ~60 kDa. The nominal molecular weight of the heterotrimer is 33 kDa. This result presents an indication that His6-TEV-LSm5/6/7 forms defined oligomeric structures, namely a dimer of trimers.

Size exclusion chromatography was first indication that the His6-TEV-LSm5/6/7 heterotrimer forms higher order structures. Negative stained electron micrographs of LSm5/6/7 preparation shows up as homogenous ring shaped particles. Class averaging yielded ring shaped particles having a size of ~7 nm and a cavity of 1.5 nm.
To characterize better oligomeric state of His6-TEV-LSm5/6/7, we performed static light scattering analysis, which suggests that the His6-TEV-LSm5/6/7 heterotrimer may form hexamers.

<table>
<thead>
<tr>
<th>complex</th>
<th>Average MW over peak</th>
<th>Standard deviation</th>
<th>Polydispersity</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSm5/6/7</td>
<td>77 kDa</td>
<td>2.0%</td>
<td>3.0%</td>
</tr>
</tbody>
</table>

Table 3.2.3. His6-TEV-LSm5/6/7 static light scattering analysis

Analytical ultracentrifugation yields a mixture of individual subunits (26%), heterotrimer (25%), hexamers (40%) and nonamers (8%), at 16μM concentration.

<table>
<thead>
<tr>
<th>c</th>
<th>Losses</th>
<th>MMmean</th>
<th>monomer</th>
<th>dimer</th>
<th>trimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.3 μM</td>
<td>0%</td>
<td>26%</td>
<td>25%</td>
<td>40%</td>
<td>8%</td>
</tr>
</tbody>
</table>

Table 3.2.4. Analytical ultracentrifugation distribution for His6-TEV-LSm5/6/7
Analytical ultracentrifugation results, taken together with static light scattering results indicate that LSm5/6/7 assumes hexameric arrangement ([LSm5/6/7]₂, 2 x 33kDa). Negative stained electron micrographs reveal ring shaped particles but from the EM analysis, a nonameric arrangement cannot be excluded. Smaller particles appear in the background and could represent LSm5/6/7 trimers. The small size of such particles was not suitable for classification. (Dr. M. Chami, Biozentrum Basel, making and interpretation of all negative stain electron micrographs presented in this work).

3.2.16. Summary of Protein Expression and Characterization Results

In this chapter I described expression and purification of 14 different LSm protein constructs. The previous working hypothesis was that LSm complex architecture and assembly are guided by homology to the known canonical Sm proteins. Initial constructs made as guided by the sequence homology with the canonical D1D2 (paralogue of LSm2/3), D3B (paralogue of LSm4/8) and EFG (paralogue of LSm5/6/7) support this concept, but there are some discrepancies in the proteins’ behavior and characteristics. All the paralogues of canonical Sm proteins mentioned above are forming higher order structures of defined oligomeric state, except for LSm4/8, which aggregates (as shown by electron microscopy and gel filtration). This does not hold for the D3B paralogue of LSm4/8. D3B runs as a heterodimer on gel filtration chromatography. LSm2/3 forms defined oligomers (hexamers predominantly, according to size exclusion chromatography). The analytical ultracentrifugation result indicates that there is a mixture between octamers and hexamers, with octamers dominating. The D1D2 paralogue forms heterodimers. LSm5/6/7 forms a dimer of trimers like its paralogue EFG. The LSm5/6/7 size exclusion chromatography elution profile and AUC indicate the formation of a mixture between trimers, hexamers, and nonamers. Canonical Sm proteins are very difficult to express singly in a soluble way (C. Kambach, unpublished result). In contrast, during my thesis I purified LSm1, LSm4, LSm8, and LSm5 from monocistronic expression vectors. LSm6 was obtained from a monocistronic T5 expression vector containing two Z tags and one His₆-tag followed by a TEV protease cleavage site. Singly expressed LSm proteins are forming higher order structures as well and they are prompt to aggregate. The solubility and stability of recombinant human LSm proteins increases as well when they are coexpressed. In order to test our hypothesis about the
role of nearest neighbour relationships in governing subcomplex stability, we also expressed LSm5/6, LSm5/3, LSm4/7, LSm5/7, LSm4/2, LSm6/7, heterodimers. All mentioned subcomplexes are expressed in high amounts, and in soluble form. They all form higher order structures, predominantly hexamers (as concluded from size exclusion chromatography). Correlating the ratio between soluble protein (in the supernatant) and non-soluble (in the pellet), we came to the conclusion that the solubility of those LSm proteins which were also expressed as monocistrons increases when they are co-expressed with another LSm protein by between 2.5- and 25-fold. Which coexpression partner was chosen apparently has no great influence on solubility. In conclusion from this extensive study, our coexpression results match well with the yeast two hybrid data showing that LSm proteins interact promiscuously with other LSm proteins, and that one given protein can interact with more than 2 other LSm proteins, i.e. more than two adjacent neighbours in the ring (Figure 3.1.). In the coexpression trials, we found some interactions that were not previously detected by yeast two-hybrid screens (Figure 3.1.1.). These are LSm5/3, LSm4/2, and LSm6/7.

<table>
<thead>
<tr>
<th>LSm protein</th>
<th>Construct</th>
<th>Supernatant</th>
<th>Pellet</th>
<th>Ratio S:P</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSm1</td>
<td>LSm1</td>
<td>10090</td>
<td>10772</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>LSm2</td>
<td>LSm23</td>
<td>9349</td>
<td>2669</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>LSm3</td>
<td>LSm23</td>
<td>9060</td>
<td>2507</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>LSm3</td>
<td>LSm53</td>
<td>5797</td>
<td>2637</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LSm4</td>
<td>LSm4</td>
<td>5292</td>
<td>12363</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>LSm4</td>
<td>LSm48</td>
<td>6734</td>
<td>6159</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>LSm5</td>
<td>LSm53</td>
<td>4546</td>
<td>1310</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>LSm5</td>
<td>LSm5/67</td>
<td>75/61</td>
<td>3710</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LSm6</td>
<td>LSm6</td>
<td>2154</td>
<td>20887</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>LSm6</td>
<td>LSm67</td>
<td>3614</td>
<td>1374</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>LSm6</td>
<td>LSm5/67</td>
<td>3751</td>
<td>1877</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>LSm7</td>
<td>LSm67</td>
<td>5751</td>
<td>3006</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LSm7</td>
<td>LSm5/67</td>
<td>6514</td>
<td>4202</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LSm8</td>
<td>LSm8</td>
<td>1717</td>
<td>8276</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>LSm8</td>
<td>LSm48</td>
<td>1722</td>
<td>1269</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 3.2.5. Increasing solubility of LSm proteins by coexpression

Supernatant-pellet ratios increase from 0.1-0.9 for singly expressed LSm proteins to 1.1-3.5 for polycistrons, and the observed solubility increases for those LSm proteins which were expressed both as mono and polycistronic between 2.5 and 25 times. Assumed nearest-neighbour relationships have no influence on the described effect.
What could be the reason that all LSm proteins are forming higher order structures? One could be that LSm proteins perform many more functions than canonical Sm proteins and that these higher order structures are some kind of storage form in the living cell which dissociate upon arrival of an appropriate signal, governing the formation of the specific heptamer from the hexameric storage complexes. The Sm proteins need only to accommodate the interactions in the canonical core snRNP domain (plus LSm10/11 in the U7 snRNP). Nobody knows how many different LSm complexes of differing composition and architecture there are, most of them probably sharing several subunits. The implication is that LSm proteins have to be much more flexible (hence less specific) in their pairwise neighbour interactions to adapt to all these differing architectures.

3.3. Sub-complex characterization under semi-denaturing conditions

As shown in the previous chapters all LSm proteins which were purified have the tendency to form higher order structures. In some cases (His6-TEV-LSm2/3, His6-TEV-LSm5/6/7, and all other heterodimers) these are defined oligomers, as judged by size exclusion chromatography, electron microscopy and analytical ultracentrifugation. The His6-TEV-LSm4/8 heterodimer has the strongest tendency to aggregate, and it is stable only in buffer which contains 1 M NaCl. All singly expressed LSm proteins have the tendency to aggregate as well, as judged by gel filtration chromatography. We chose to use urea for the disruption of aggregates. Solid urea was added to the protein solution in the respective buffer in which it was maintained. The amount of added urea was calculated to reach a final concentration of 4 M. No corrections for the volume increase were made, so the actual final concentration is somewhat lower than 4 M. The proteins were incubated in 4M urea at 37 °C for four hours. For comparison, gel filtration chromatography was performed with the same batch of protein under native conditions and with protein incubated in 4 M urea. In the latter case, the running buffer was supplemented with 4 M urea. Under native conditions, the His6-TEV-LSm2/3 heterodimer elutes at an elution volume (14.13 ml) corresponding to an apparent molecular weight of 66 kDa on size exclusion chromatography (Figure 3.3.1). The nominal molecular weight of His6-TEV-LSm2/3 is 25kDa.
Figure 3.3.1. Size exclusion chromatography profile of His6-TEV-LSm2/3 under native conditions.

The elution volume of 14.13 ml corresponds to an apparent MW of 66 kDa. This corresponds to a hexamer within the error margin of the gel filtration method (~15%). The calculated molecular weight of the hexamer is 75 kDa.

After incubation in urea and running a gel filtration in a buffer containing 4 M urea, several different oligomeric states are visible. The main peak runs at a slightly higher apparent molecular weight (13.71 ml, 76 kDa) than under native conditions. This could indicate that oligomer adopts a more relaxed, open conformation. The peak at 15.55 ml would correspond to a molecular weight of 40 kDa which can not be fitted with the molecular weight of a tetramer or any other defined multiple of the heterodimer. The peak at 18.58 ml corresponds to a molecular weight of 14 kDa, possibly indicating disruption to a small extent into the individual protein subunits. However, the predominant species remains a higher order structure of His6-TEV-LSm2/3 of defined oligomeric state, namely the hexamer, probably adopting a slightly more open conformation than under native conditions.
Figure 3.3.2. Size exclusion chromatography profile of His6-TEV-LSm2/3, after incubation in 4 M urea at 37°C.

A similar trend was observed for the His6-TEV-LSm5/6/7 heterotrimer. Under native conditions His6-TEV-LSm5/6/7 does not elute from gel filtration chromatography as a trimer. (Figure 3.3.3)

The elution volume from the GPC column fits to a molecular weight (determined by external calibration) that best corresponds to a hexameric oligomer – but that is assuming that there are only trimers of defined composition and stoichiometry that can assemble into multiples thereof.

Figure 3.3.3. Size exclusion chromatography profile of His6 LSm5/6/7.

After incubation in urea and running gel filtration chromatography under semi-denaturing conditions, His6-TEV-LSm5/6/7 still elutes as a single homogeneous peak from gel filtration but at an elution volume of 14.50 ml (in the native case 14.95 ml). An elution volume of 14.50 ml
would correspond to molecular weight of 65 kDa, which could indicate that this sub-complex assumes as well a more open conformation (Figure 3.3.4).

Figure 3.3.4. Size exclusion chromatography profile of His6 LSm5/6/7 after incubation in 4 M urea at 37°C.

This result indicates that the His6-TEV-LSm5/6/7 hexamer (Figure 3.3.4.) does not fall apart in 4M urea. The slight increase in apparent molecular weight indicates that this sub-complex adopts a more open conformation.

His6-TEV-LSm4/8, as mentioned previously, has the strongest tendency to form aggregates, as judged by size exclusion and negative stain electron micrographs. Under native conditions on gel filtration chromatography it runs at an elution volume which would correspond to a molecular weight higher than 394 kDa (Figure 3.3.5). The nominal molecular weight of this heterodimer is 28kDa.
Figure 3.3.5. Size exclusion chromatography profile of His6-TEV-LSm4/8.

Together with negative stain electron micrographs this indicates that this heterodimer makes aggregates of a non-defined oligomeric state.

After incubation in urea, the aggregates were disrupted as judged by the size exclusion chromatography profile. A small peak at 9.14 ml is still present suggesting that longer incubation times may be required. The predominant peak elutes at 15.38 ml, corresponding to a molecular weight of ~42 kDa. The 42 kDa in the semi-denaturing and native conditions lies between a tetramer (2x2, nominally 56kDa) and the dimer (28kDa).

Figure 3.3.6. Size exclusion chromatography profile of His6 LSm4/8, after incubation in 4 M urea at 37°C.
In figures 3.3.1.-3.3.6, it is demonstrated that LSm2/3 and LSm 5/6/7 sub-complexes upon incubation in 4 M urea are adopting more open conformation and that aggregates formed by the LSm4/8 are disrupted. Pronounced dissociation of these sub-complexes into the individual proteins was not observed on gel filtration, except for the LSm2/3 subcomplex. Apparently the higher aggregates but not the smaller subcomplexes are falling apart under these semi denaturing conditions and the subcomplexes appear to assume more open structures. This suggests that they are real near-neighbours in the heptameric LSm2-8 and LSm1-7 rings.

3.4. In vitro reconstitution of LSm2-8 and LSm 1-7 complexes

One of the main goals of my thesis was the reconstitution (assembly) of the LSm2-8 and LSm1-7 heteroheptameric complexes in vitro. This task was achieved using protein components whose cloning, expression and purification is described in chapters 3.1 and 3.2. The following section describes reconstitution of the two LSm complexes in the absence of RNA and reconstitution trials in the presence of U6 snRNA which is the natural RNA target, of LSm-8.

3.4.1. In vitro reconstitution of LSm2-8

In vitro reconstitution was carried out using the two heterodimers His6-TEV-LSm2/3, His6-TEV-LSm4/8 and the heterotrimer His6-TEV-LSm5/6/7. Initially, the reaction was carried out in the buffer in which the sub-complexes were stored. For His6-TEV-LSm2/3 and His6-TEV-LSm5/6/7 this was 150 mM NaCl, 20 mM HEPES-Na pH 7.5, 5 mM β-mercaptoethanol. For His6-TEV-LSm4/8, the storage buffer composition was similar, replacing 150 mM with 1 M NaCl (all other components identical).

First reconstitution trials were carried out by mixing the purified sub-complexes (proteins) in equimolar ratios and incubating at 37°C. In the case of the canonical Sm proteins, a modification of this protocol involving the addition of UsnRNA was used successfully [165]. The LSm2-8 complex was isolated from HeLa cell nuclear extract in an RNA-free form [99]. This prompted us to attempt assembly of the LSm2-8 complex in the absence of U6 snRNA. During the biophysical characterization of our LSm sub-complexes, we found that semi-denaturing
concentrations of urea can disrupt higher order aggregates (see chapter 3.3.) and we made use of this fact. Reconstitution was carried out by mixing equimolar amounts of LSm2/3, LSm4/8 and LSm5/6/7 (previously incubated in urea separately and afterwards together), followed by dialysis against 1 M NaCl, 20 mM HEPES-Na pH 7.5, 5 mM β-mercaptoethanol. After six hours the NaCl concentration was reduced to 0.5 M and dialysis was continued over night. After concentration of dialysate to a final volume of 1 ml, it was applied to size exclusion chromatography, using Superdex 200 HR10/30, at 4°C, in a buffer containing 150 mM NaCl, 20 mM, HEPES-Na pH 7.5, 5 mM DTT.

Figure 3.4.1. First size exclusion chromatography profile

Figure 3.4.2. SDS PAGE gel with fractions from the first size exclusion chromatography
From Figures 3.4.1 and 3.4.2, it is clear that the peak at 11.62 ml and the following fractions contain proteins with molecular weights between 9 and 17 kDa. The smallest member of the LSm2-8 ring (LSm6) has a calculated molecular weight of 9.1 kDa, and the largest LSm4, of 17.7 kDa. Fractions G9-G13 contains all seven proteins although the resolution of the gel is not sufficient to make this clear. The peak at 19.48 ml (fraction H13) does not contain LSm6. According to the isoelectric points of the individual sub-complexes, the estimated isoelectric point of a complex containing one copy of each of the seven proteins should be around 6.4. We thus decided to purify the complex by anion exchange chromatography at pH 7.5. Fractions G9-G13 from the gel filtration step were applied to a Resource Q 1 ml column and elution is carried out using a linear salt gradient. The composition of the buffers and the running conditions are described in Materials and Methods.

Figure 3.4.3. First Anion exchange size chromatography profile

Figure 3.4.4. SDS PAGE gel with fractions from first anion exchange chromatography step
From Figures 3.4.3 and 3.4.4., it is clear that at least four different peaks with different protein distribution are present. Fractions D2 and D3 elute at about 250 mM NaCl and contain all seven proteins, although the resolution of the gel is not sufficient to resolve them all. The fractions eluting at lower and higher NaCl, contain LS proteins of variable composition which probably were not incorporated into heptameric rings. For the next anion exchange step the pooled D2 and D3 fractions were reapplied onto the Resource Q 1 ml column, and the protein eluted as a single sharp peak indicating that it contains a stable complex (Figures 3.4.5., and 3.4.6).

![Second anion exchange chromatography profile](image)

**Figure 3.4.5.** Second anion exchange chromatography profile

![SDS PAGE gel with fractions from the second anion exchange chromatography step](image)

**Figure 3.4.6.** SDS PAGE gel with fractions from the second anion exchange chromatography step
The final step for determining the molecular weight of the obtained complex was size exclusion chromatography on a Superdex 200 HR10/30 column in a buffer containing 150 mM NaCl, 20 mM Hepes-Na 7.5.

Figure 3.4.7. Second size exclusion chromatography step

Figure 3.4.8. SDS PAGE gel with fractions from second size exclusion chromatography step
Fractions D14, D15, E1, E2, and E3 (Figure 3.4.6.) were concentrated and subjected to this second gel filtration chromatography as shown in Figures 3.4.7 and 3.4.8. The proteins elute in a single peak at 12.60 ml which corresponds to an apparent molecular weight of 95 kDa and all seven proteins appear to be present in equal amounts. The calculated molecular weight of the LSm2-8 heteroheptamer is 87 kDa.

Although these data do not unambiguously prove that we have isolated homogeneous LSm2-8 heteroheptameric complex this appears very likely.

To further enhance the stringency of the reconstitution process, we tried a modified approach. The idea was to use previously described reconstitution conditions, with three different sub-complexes but only one carrying a H6-tag. After reconstitution, the first purification step would be affinity chromatography. While the His6-tag from the LSm5/6/7 trimer is cleavable, that from LSm2/3 is not and that from LSm4/8 is only completely cleaved after seven days at room temperature with additional purification using gel filtration column. The reconstitution procedure was the same; sub-complexes were incubated in 4 M urea and dialyzed against native buffer. The reconstitution mixture, after over night dialysis, was applied onto a Ni-affinity column, and the bound proteins were eluted using a linear gradient of 0-500 mM imidazole pH7.5, over 10 column volumes (Figures 3.4.9. and 3.4.10.).
Figure 3.4.9. Ni-Affinity chromatography of reconstituted LSm2-8

Figure 3.4.10. SDS PAGE gels with fractions that correspond to the first (48.56 min) peak (left SDS PAGE gel). Fractions which correspond to the second (56.15 min) peak are on the right SDS PAGE gel.
Fractions G15-H15 contain all seven protein in stoichiometrical ratio (although this is not visible from the upper gel, Figure 3.4.10), and they were subjected to a second chromatography step, being anion exchange.

Figure 3.4.11. Anion exchange chromatography after affinity purification step.

Figure 3.4.12. SDS PAGE of first anion exchange chromatography after affinity
Fractions M9 and M10 (Figure 3.4.12.) contains all seven proteins and they were pooled for further purification by a second anion exchange and a size exclusion chromatography step. The numbers indicate seven resolved protein bands (Figure 3.4.12) but do not correspond to the protein names.

Figure 3.4.13. Second anion exchange chromatography after affinity purification step.

Figure 3.4.14. SDS PAGE of second anion exchange chromatography.
Figures 3.4.15 and 3.4.16 illustrate the final size exclusion chromatography step and the results indicate the presence of a homogeneous complex preparation. This was confirmed by a static light scattering experiment (Figure 3.4.17).

Figure 3.4.15. Size exclusion chromatography step after second anion exchange

Figure 3.4.16. SDS PAGE of size exclusion chromatography step after second anion exchange
Figure 3.4.17. Static light scattering chromatogram and report for LSm 2-8 complex. Complex has molecular weight of 92 kDa and a polydispersity of 1.6%.
3.4.2. *In vitro* reconstitution of LSm 1-7

The reconstitution was carried out using the following components: heterodimer His6-TEV-LSm2/3 and heterotrimer His6-TEV-LSm5/6/7 and separately produced His6-TEV-LSm1 and His6-TEV-LSm4. The components were mixed in equimolar ratios. The reconstitution procedure and the purification of the LSm1-7 complex was done in the same way as for LSm2-8 complex (Chapter 3.4.1).

![Size exclusion chromatography of LSm1-7](image)

**Figure 3.4.18.** Size exclusion chromatography of LSm1-7

![SDS PAGE of size exclusion chromatography of LSm1-7](image)

**Figure 3.4.19.** SDS PAGE of size exclusion chromatography of LSm1-7
Fractions A12 to A15 were used for the subsequent anion exchange chromatography. The sample was loaded onto a resource Q 1 ml column and elution carried using a linear gradient over 30 column volumes, from 150 mM NaCl to 1 M NaCl. The buffer composition was the same as in the case of the LSm2-8 reconstitution.

Figure 3.4.20. Size exclusion chromatography of LSm1-7

Figure 3.4.21. Anion exchange chromatography of LSm1-7
After three steps of chromatography (size exclusion, anion exchange chromatography and final size exclusion chromatography step) homogeneous LSml-7 complexes were obtained that appeared suitable for electron microscopy analysis, for functional assays and for crystallization trials.
3.4.3. Electron micrographs of reconstituted LSm2-8 and LSm1-7 complexes

Negative stain electron micrographs show that reconstituted LSm2-8 has a ring-like architecture with a diameter of about 8 nm. The overall dimensions are similar to those previously observed for the native LSm2-8 complex isolated from HeLa cell nuclear extract (8 nm) and core snRNP domain [99,102]. The central cavity observed for the recombinant LSm2-8 complex is larger than in the native LSm2-8 complexes (3 vs. 2 nm, respectively). Since the recombinant complex shows the same RNA binding specificity, this difference must remain unexplained at present. The LSm1-7 rings appear to be slightly smaller, measuring ~7 nm across with a pore diameter of less than 1.5 nm. Thus, recombinant LSm1-7 and LSm2-8 complexes are similar to one another and to the native Sm/LSm complexes at this level. In all LSm co-crystal structures solved with RNA oligonucleotides, the RNA molecules mainly wrap around the rim of the pore, although in one case, additional binding sites on the ring surface have been observed. This is in contrast to the original concept that in the core snRNP domain, the Sm site target RNA threads through the heptamer's pore. The concept was based on the electrostatics of the core domain model and the position of conserved residues assumed (and later shown) to bind RNA. Structural evidence to corroborate this idea has so far only been obtained at the ultrastructural level, by cryo-electron microscopy of the U1 snRNP [166]. For LSm2-8 – U6 snRNA interaction, the binding determinant has been shown to be the U5 stretch at the 3' end of U6 snRNA. Indeed all Sm/LSm complex–RNA interaction studies have come to the conclusion that this protein family binds with strong preference to single stranded, uridyl-rich targets, including poly-U. LSm2-8 assembles and is stable in the absence of RNA. Since its target, U6 snRNA 3' end, is freely accessible to a preassembled complex, it is possible that it threads through the LSm2-8 central cavity. However, the smaller pore diameter of the recombinant LSm1-7 complex could indicate differences to LSm2-8 in RNA binding. LSm1-7 binds to the 3' UTRs of deadenylated mRNAs. Although the RNA binding determinants for the LSm1-7 complex have not been characterized in detail, LSm1-7 presumably does not bind to the extreme 3' end of its target mRNAs. At least in some cases, secondary structure elements found in many of its target 3' UTRs are likely to prevent the RNA threading through the LSm1-7 hole, as is possibly the case for the LSm2-8 – U6 snRNA interaction.
Figure 3.4.24. Electron micrographs of reconstituted complex LSm1-7 (c) and LSm2-8 (d).

Scale bars are 30 nm. Class averages were created from 993 and 988 particles, respectively, from initial data set containing 1000 particles of, LSm1-7 and LSm2-8.
3.5. *In vitro* reconstitution of LSm2-8 in the presence of U6 snRNA

An alternative way to obtain LSm protein complex is carrying out the assembly reaction in the presence of the U6 snRNA. The LSm2-8 complex is reported to be extremely stable, up to 2 M NaCl, or 0.4 M isothiocyanate [99]. The LSm2-8/U6 snRNA complex is apparently not very stable, unlike the canonical Sm core domain [99].

U6 snRNA was obtained as described in Materials and Methods. Purified U6 snRNA was added to the proteins in a ratio of 2:1. In this and other trials 3 nmoles of RNA and 1.5 nmoles of protein complex was used. The mixture was incubated for one hour at 30 °C and another hour at 37 °C. After dilution into 20 mM Hepes-Na7.5 the mixture was applied onto an anion exchange chromatography column and eluted using a linear gradient of the NaCl from 100 mM to 1 M NaCl. This procedure relies on the experiences with canonical Sm core domain assembly *in vitro*, which was successfully carried out as described here (Christian Kambach, unpublished result).

In our case this approach failed probably because the LSm2-8/U6 snRNA complex does not withstand anion exchange chromatography. The anion exchange chromatography profile and the corresponding SDS PAGE gel (Figures 3.5.1 and 3.5.2) suggest that there is little or no complex formation or that the complex falls apart on the anion exchange column resulting in very heterogeneous elution profile.
Figure 3.5.1 Anion exchange chromatography of LSm2-8 reconstitution trial in the presence of U6snRNA

Figure 3.5.2 SDS Page gel which corresponds to the upper chromatogram.
Lane 1: C9+C10, Lane 2: D15+E1, Lane 3: E4+E5, Lane 4: E9, Lane 5: E10, Lane 6: E11, Lane 7: E12, Lane 8: E13+E14, Lane 9: E15+F1
3.6. Electrophoretic mobility shift assays (EMSA)

One of the methods which can be used for the identification and characterization of the RNA binding proteins is the electrophoretic mobility shift assay (EMSA). The basis of this method is the change in the electrophoretic mobility of a nucleic acid molecule upon binding to a protein or another molecule. Initially a labeled RNA, which contains the binding sequence, is incubated with a sample containing the RNA binding proteins and the mixture is then analyzed on a non-denaturing gel. The unbound RNA will have a characteristic electrophoretic mobility. If the RNA is bound by one or more proteins, then the movement of RNA through the gel is retarded. This gives rise to a characteristic shift in the position of the RNA band on the gel. The mobility shift method is a common technique for identifying DNA binding proteins. This method is less developed for identifying RNA binding proteins but it has been applied successfully in many instances.

In this work EMSA was used to evaluate the functionality of reconstituted LSml-7 complex whose natural substrate is U6 snRNA. LSml-7 binds to the 3' end of the untranslated region of mRNA (3' UTR) but in this work we did not establish binding of LSml-7 complex to such RNA.

![Figure 3.6.1. Electrophoretic mobility shift assay (EMSA) with U6 snRNA as a probe.](image)

Individual subcomplexes LSml-2/3, LSml-4/8, LSml-5/6/7 are not capable to assemble with RNA and to slow electrophoretic mobility. Only the reconstituted LSml-2-8 complex shifts U6 snRNA.
Figure 3.6.2. Electrophoretic mobility shift assay (EMSA) with U6 snRNA as a probe. LSmi-7 complex does not shift RNA, nor do the combinations of sub complexes, missing the full complement of LSmi-8 components. Leaving out LSmi2/3 (lane 4, Figure 3.6.2) leads to aggregation of the sample and retardation into the wells. Only LSmi2-8 (Lane 2, Figure 3.6.2) shifts U6snRNA.

Figure 3.6.3. Electrophoretic mobility shift assay (EMSA) with U6 snRNA as a probe. LSmi1-7 complex does not shift RNA (Figure 3.6.3, Lane 4). LSmi2-8 complex shifts U6 snRNA (Figure 3.6.4, lane 2). Incubation of LSmi2-8 with U6 snRNA in the presence of an LSmi2/3-
specific single chain Fv antibody leads to a complex shifted to higher molecular weight (Figure 3.6.3. Lane 3 supershift)

Figure 3.6.4. Electrophoretic mobility shift assay (EMSA) with U1 snRNA as a probe. LSm2-8 does not shift U1 snRNA (Lane 3, Figure 3.6.4.), whereas a 1:1 mixture of the 7 canonical Sm proteins does (Lane 2, Figure 3.6.4.)

The native LSm2-8 complex was initially isolated from U4/U6 di-snRNP particle as an RNA free complex and shown to bind to the U6 snRNA in vitro [99]. In order to test our reconstituted LSm complexes for binding specificity we performed electrophoretic mobility shift assays (bandshift) with U6 snRNA. Pre-assembled, purified LSm2-8 does shift U6 snRNA. Individual subcomplexes (LSm2/3, LSm4/8, and LSm5/6/7) do not shift U6 snRNA. We also used reconstituted samples in which either one of the sub-complexes or a single LSm protein was left out. Such samples do not shift U6 snRNA. LSm1-7 also does not shift U6 snRNA. The specificity of complex formation is further demonstrated by adding an LSm2/3-specific antibody to the reaction mixture, which leads to a supershift. Using same assay conditions LSm 2-8 does not bind to the U1 snRNA. In contrast, a mixture of the canonical Sm proteins (D1D2, D3B, EFG), binds U1 snRNA as
expected. In summary, the recombinant LSm complexes show the same RNA binding properties like their native counterparts, and we can conclude that our recombinant LSm2-8 complex is functional in this respect. Apart from establishment that LSm1-7 complex does not bind to U6 snRNA, its substrate RNA binding was not further investigated because appropriate RNA substrate was difficult to obtain.

3.7. In vivo functional assays

We decided to test the functionality of the LSm complexes in living cells using microinjection of reconstituted LSm1-7 and LSm2-8, labeled with fluorescent dyes. The complexes were assembled and purified to homogeneity, as judged by the methods mentioned above. The stability of the LSm2-8 complex was evaluated after the labeling procedure using analytical gel filtration.

Figure 3.7.1. Analytical size exclusion chromatography of reconstituted LSm2-8 complex
Figure 3.7.2. Analytical size exclusion chromatography of reconstituted LSm2-8 labeled with Alexa 488 succinimidyl ester. The fluorescent dye absorbs at 495nm (violet). One peak at 1.43 ml is dominates.

Figure 3.7.3 SDS page gel of non-labeled and labeled LSm2-8 complex. Lane 1 represents non labeled LSm2-8 complex which contains all seven subunits present, without any apparent contaminations or degradations. Lane 2 represents Alexa 488 labeled reconstituted LSm2-8 complex. Similarly, all seven bands are present on the SDS PAGE gel, and no contamination or degradation is visible.
Figure 3.7.4. SDS page gel of non-labeled and labeled LSml-7 complex

LSm2-8 and LSml-7 stability (integrity) after coupling with fluorescent dye was evaluated by size exclusion chromatography and SDS PAGE. It is concluded that both complexes are stable after coupling and that the dye does not interfere with the stability or integrity of the complexes. Reconstituted complexes were labeled and used for microinjection into the cytoplasm of Embryo Fibroblast Ref52 cells (in further text REF52), according to the procedure described in Materials and Methods.
Immediately after microinjection, the cells were transferred to a 37°C incubator for 120 minutes, fixed, and observed under a fluorescence microscope. LSm1-7 complex is observed in the cytoplasmic compartment, although it is not localized in cytoplasmic foci [99], but evenly distributed through cytoplasm. The LSm2-8 complex is clearly localized in the nuclear compartment. This distribution is in accordance with their natural occurrence. To test the kinetics of LSm2-8 nuclear import, we observed microinjected cells immediately after injection, because during a long incubation, complex inside the cell might degrade and individual subunits might diffuse into the nucleus. Proteins which are up to 40 kDa can passively diffuse through nuclear pore complexes, and the largest LSm protein is 17 kDa (LSm4).
Figure 3.7.6. LSm2-8 complex microinjected into the cytoplasm of REF52 cells, at different time points. 20 minutes after injection, the complex becomes localized in the nucleus.

Figure 3.7.7. LSm2-8 complex microinjected into the cytoplasm of REF52 cells at later time points. Up to at least 120 minutes the labeled complex is in the nucleus.

To increase the stringency of the conclusions that LSm2-8 complex is preassembled in the cytoplasmic compartment and then transported into nucleus, we performed an experiment in
which we coinjected wheat germ agglutinin (WGA) together with the LSm2-8 complex. Wheat
germ agglutinin (WGA) blocks nuclear transport [105]. Its inhibitory effect persists for about one
hour after injection into the cytoplasm of cells. WGA does not block export of some RNAs from
the nucleus, and does not inhibit passive diffusion of FITC dextran (17.9kDa).

Figure 3.7.8. LSm2-8 complex microinjected into the cytoplasm of REF52 cells together with
wheat germ agglutinin, at a concentration of 2.5mg/ml. After 20 minutes and 60 minutes post
injection the injected material resides in the cytoplasm.

Figure 3.7.9. LSm2-8 complex co-injected with WGA, reversibility of nuclear import is
observed at, after 120 minutes.

Whether the complex is transported as an intact entity or whether it is degraded in the cell
cytoplasm cannot be answered with these experiments, although the WGA control experiment
suggests the former. Assembly and nuclear import of the endogenous LSm2-8 complex remains
unanswered to this date. There are indications from the literature that the complex assembles in a one-step pathway, but it is not yet known whether the SMN complex is involved in the assembly or not [99]. There are only some data suggesting that SMN interacts with some LSm proteins (LSm4) in vitro [86,87].

I tried to identify the LSm protein nuclear import receptor in the laboratory of Prof. U. Kutay (ETH, Zürich) using permeabilized HeLa cells and recombinant known transport factors, but the experiments did not lead to a clear conclusion. The experiment was repeated using a buffer containing 0.048 M K₂HPO₄, 0.0045 M KH₂PO₄, 0.014 M NaH₂PO₄, pH: 7.12, rather than: 150 mM NaCl, 20 mM HEPES-Na 7.5, 5 mM DTT. Assessment of the stability of labeled complex in this buffer on analytic gel filtration indicates that it is not stable in this buffer. I also tried to exchange 150 mM NaCl with 150 mM KCl as well, but without success. It appears unlikely that exchanging Na⁺ for K⁺ can cause such stability problems. Possibly this preparation was contaminated with proteases.

Figure 3.7.10. Control experiment with nucleoplasmin.

As shown in the upper panel of Figure 3.7.10., injected nucleoplasmin was found as expected in the nucleus after 30 minutes. When coinjected with WGA, about two hours after microinjection reversibility of nuclear import inhibition is observed.

Microinjection experiments were also done with reconstituted LSm1-7 complex (Figure 3.7.11) and as further controls with some individual subunits and subcomplexes (Figures 3.7.12, 3.7.13 and 3.7.14)
Reconstituted LSm1-7 complex was labeled with Alexa 488 (upper panel of Figure 3.7.11) and with Alexa 555 (lower panel of Figure 3.7.11). As expected it remains in the cytoplasmic compartment, for the time of observation (180 minutes).

Figure 3.7.11. LSm1-7 complex injected into REF52 cells

LSm sub-complexes injected into REF52

LSm2/3 was labeled with Alexa 488, LSm4/8 with Alexa 350 and LSm5/6/7 with Alexa 555. The labeled subcomplexes were coinjected immediately after mixing into the cytoplasm of REF52 cells. On the overlay it is visible that the sub complexes are colocalizing but are not transported into the nucleus.

Figure 3.7.12. LSm sub-complexes injected into REF52
At all time points after microinjection (Figure 3.7.13), LSm8 was found to be localized in the cytoplasmic compartment. Other authors, after transfection of HeLa cells with YFP-LSm8, found this protein exclusively in the nuclear compartment. Indeed it is known that LSm 8 is part of several complexes (snR5, snoU8, LSm2-8) and they are all nucleolar or nuclear. In our experimental system LSm 8 shows a different localization probably due to aggregation.

Figure 3.7.14. LSm1 injected into cytoplasm of REF52 cells
We employed the microinjection technique for the evaluation of the functionality of our reconstituted complexes. The complexes were preassembled like described in the previous chapters, and coupled to a fluorescent dye. After this, we confirmed by analytical size exclusion chromatography and SDS GEL electrophoresis, that the complexes were not degraded, or disassembled during labeling. In REF52 cells, the LSm2-8 complex is found predominantly in the cell nucleus 20 minutes after microinjection. In contrast the LSm1-7 complex is localized in the cytoplasmic compartment.

This strongly suggests that our complexes are functional in vivo as they localize in their appropriate subcellular locations. Moreover, it seems that the semi denaturing step with urea for assembly of the complexes is necessary as well as all subsequent purification steps, because mixed LSm2/3, LSm4/8 and LSm5/6/7 are not capable to bind radioactively labeled $^{32}$P U6 snRNA and, inside the cell, they do not assemble into functional complexes. It is generally assumed that the LSm2-8 complex undergoes a one step assembly pathway in the cytoplasm and is subsequently imported into the nucleus. We tried by coinjection with wheat germ agglutinin to answer the question whether the LSm2-8 complex is transported intact or whether after disruption the individual subunits diffuse passively into the nucleus because of their low molecular weights (9kDa to 17kDa). Individually, in theory they could enter the nucleus by passive diffusion. As wheat germ agglutinin binds to the nuclear envelope and blocks active transport, only passive diffusion is possible [105]. From the experiments performed it seems that WGA blocks active transport of LSm2-8 complex. After 120 minutes WGA becomes degraded and active transport through nuclear pore complexes resumes. This is clear evidence that preassembled LSm2-8 complex moves into the nucleus.

We tried to address the question of which nuclear import receptor is involved in collaboration with Prof. U. Kutay (ETH Zürich). In this experiment, I used digitonin permeabilized HeLa cells, where the whole cytoplasmic content is washed away from the cell. Recombinant, labeled LSm2-8 labeled complex was incubated with different import factors like importinβ, importinβ/α, reticulocite cell lysate, buffer, RanGTP, and an energy mix. In these experiments the LSm2-8 complex appears to move into the nucleus in the absence of any import factors, which does not make biological sense. (Figures 3.7.15. and 3.7.16.) Presently it is not clear whether this is an experimental artifact and further experiments will be needed to clarify this point.
3.8. Crystallization trials

One of the goals at the start of the entire project was to crystallize and solve the X-ray structure of the two heteroheptameric complexes whose preparation and functional characterization has been described in the previous chapters. The X-ray structure of complexes would give insight into the selectivity of the subunit interactions and provide a basis for understanding the different functions. We also wanted to crystallize LSm2-8 with its target U6 snRNA. None of the individual LSm proteins has a known RNA recognition motif. A high resolution X-ray structure would show the mode of interaction with U6 snRNA. We would have identified the part played by individual proteins in the interaction with U6 snRNA. The identified residues could then have been mutated and their capacity assessed to assemble via their Sm motifs into a ring-shaped, functional structure. An X-ray structure would provide evidence how a group of proteins that do not contain a known RNA interaction motif collectively can provide an RNA interaction interface. We tried to crystallize not only the full heteroheptameric complexes, but heterodimers
and trimers as well. Some of these could be crystallized but the diffraction quality was not sufficient to solve any of these crystal structures.

<table>
<thead>
<tr>
<th>LSm</th>
<th>Expression vector</th>
<th>Expression host</th>
<th>yield / liter L.B</th>
<th>Crystallization</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSM2/3</td>
<td>His6-TEV-LSm2/3:pQE30</td>
<td>SG13009[pREP4]</td>
<td>10 mg</td>
<td>No crystals obtained</td>
</tr>
<tr>
<td>LSm5/6/7</td>
<td>His6-TEV-LSm5/6/7:pQE30</td>
<td>BLR[pREP4]</td>
<td>15 mg</td>
<td>Crystals diffracting to 8Å</td>
</tr>
<tr>
<td>LSm5/7</td>
<td>His6-TEV-LSm5/7:pQE30</td>
<td>SG13009[pREP4]</td>
<td>10 mg</td>
<td>No crystals obtained</td>
</tr>
<tr>
<td>LSm5/3</td>
<td>His6-TEV-LSm5/3:pQE30</td>
<td>SG13009[pREP4]</td>
<td>8 mg</td>
<td>No crystals obtained</td>
</tr>
<tr>
<td>LSm6</td>
<td>ZZ-His6-TEV-LSm6:pQE70</td>
<td>SG13009[pREP4]</td>
<td>10 mg</td>
<td>Crystals diffracting to 4.3Å</td>
</tr>
<tr>
<td>LSm2-8</td>
<td>Reconstitution in vitro</td>
<td></td>
<td>5% efficiency</td>
<td>Obtained non diffracting crystals</td>
</tr>
<tr>
<td>LSm1-7</td>
<td>Reconstitution in vitro</td>
<td></td>
<td>5% efficiency</td>
<td>No crystals obtained</td>
</tr>
</tbody>
</table>

Table 3.8.1.: Crystallization trials with complexes, subcomplexes and individual subunits
Figure 3.8.1. LSm5/6/7 crystals

Initial crystals were obtained using 20% PEG 8000, 100 mMTris8.5, 0.2 M MgCl2. Optimization trials did not yield another form of crystals. At the Swiss Light Source (PSI Villigen) they were diffracting to 8Å.

Figure 3.8.2. LSm6 crystals

Initial crystals of LSm6 were obtained under 20%PEG 1000 0.1 M Na-Citrate 0.1 M Amoniumacetate pH 6. After optimization they diffracted to 4.3 Å, on a rotating anode, X-ray generator. Further attempts to improve the existing crystals were not successful. The other targets did not yield any crystals.
4. Summary and Outlook

The introduction part of this thesis briefly describes all processing steps during the life of mRNA: transcription, capping, splicing, 3' end formation, export, and degradation. It is important to realize that all steps influence each other. During the entire life cycle, the mRNA is coated with proteins. At each step of the life cycle, the protein composition of the ribonucleoprotein particle changes and in such way decides the further fate of mRNA.

We concentrated our work on the splicing step of the mRNA life cycle. SnRNPs play a central catalytic role in intron removal. Probably the transesterification reactions occur via the action of the snRNAs, giving credibility to the theory that the spliceosome as one of the most complex macromolecular machineries has evolved from type II self splicing introns. U6 snRNP has a central catalytic role in splicing. U6 snRNA is the most conserved snRNA and its life cycle is completely different from the one of other U snRNAs.

In the eukaryotic genome there are more than 24 Sm/LSm genes. We chose to study the structure-function relationship in two human LSm complexes containing seven subunits each, LSm1-7 and LSm2-8. The aim was to learn more about how the exchange of one subunit (LSm1 with LSm8) can alter function, RNA target, and intracellular distribution of the two heptamers.

The function of the LSm/Sm proteins has to be viewed in the context of the RNP of which it is part. LSm proteins were first identified as part of U6 snRNP, and probably they have role in splicing, and promote annealing with U4 snRNP, making the di/snRNP particle. 7 LSm proteins (LSm2-8) form a heptameric ring around the 3' end of U6 snRNA. It is likely that the LSm2-8 complex undergoes a one-step assembly pathway but this is presently not sure. The LSm2-8 complex has a role in U6snRNP assembly, and concomitantly in recycling splicing activity.

Although belonging to the same protein family, the Sm core domain assembly pathway and the LSm complex assembly pathway differ. The nature of their respective protein-protein interactions is different: Interactions of Sm proteins are highly specific, and LSm proteins interact more promiscuously with each other in vitro. The other heptameric, complex LSm1-7, has a role in mRNA degradation. Each of the LSm proteins has a closest paralogue among the canonical Sm proteins. We decided to use a coexpression strategy for obtaining LSm proteins in good yield. Usage of heterodimers and -trimers should also facilitate in vitro reconstitution of the two heteroheptamers and, by pre-defining several nearest neighbour relationships, possibly provide
some information on their architecture. According to experiences with canonical Sm proteins, coexpression can increase yield, solubility and stability. Sm proteins were never expressed in a soluble form as single cistrons in *E. coli* in a functional properly folded form.

Our original concept was that the sequence – binding specificity relationship works backwards for coexpression: Only those pairs of LSm proteins representing true physiological subcomplexes or at least reflecting nearest-neighbour relationships in the final heptamers were supposed to yield well expressed, soluble heterodimers and -trimers. In order to test this hypothesis, we constructed several negative controls (LSm6/7, LSm4/2, LSm5/3), representing non-nearest neighbours in the assumed order of subunits in the rings.

During this work, 21 different LSm protein expression clones were constructed all based on the T5 expression system. I describe the determination of the optimized expression parameters in *E. coli*, and the purification, and characterization of the sub-complexes or of the singly expressed proteins. Most of the LSm proteins were subcloned as heterodimers and heterotrimers, with only the first cistron bearing a His6-tag. The other overexpressed proteins of interest were co-purified via complex formation with the protein bearing the His6-tag. This suggests that the sub-complexes produced are in general very stable.

An LSm5/6 interaction was observed in a yeast two hybrid screen [124]. Interaction between the paralogues of LSm5 and LSm6 (SmF and SmE) was observed in coimmunoprecipitation assays [73]. LSm5/3 is a new interaction between LSm proteins, as observed in this work, but not detected in the yeast two hybrid assay [124]. Their paralogues, the canonical D2 and E proteins were never coexpressed in *E. coli* or in any other expression system in a soluble form. An LSm4/7 interaction was also identified in the yeast two hybrid assay [124] and confirmed with our experiment. There is no experimental evidence that their paralogues D3 and G, are capable to form stable heterodimers.

His6-TEV-LSm 6/7 was assumed to be a pair of non nearest neighbours in the heteroheptameric LSm2-8 complex (Figures 3.1.1.). An LSm6/7 interaction was not identified in the yeast two hybrid assay [124] and there is no evidence that their paralogues F and G are capable of making stable heterodimers.

LSm2/3, Lsm4/8 and LSm5/6/7 correspond to subcomplex paralogues for the canonical Sm proteins, and they were all expressed in soluble forms. The oligomeric state of these subcomplexes was analysed using gel filtration chromatography. Determination of the exact
molecular weight and polydispersity was carried out using static light scattering and analytical ultracentrifugation. The shape and size of the sub-complexes was addressed by electron microscopy, in collaboration with the group of Prof. A. Engel, Biozentrum Basel. Reversed phase chromatography was used for determining the stoichiometry of the sub complexes.

The results from analytical ultracentrifugation, negative stain electron micrographs, static light scattering and calibrated size exclusion chromatography all suggest that the predominant form of LSm2/3 in solution is an octamer. This can not be concluded from the electron micrographs and gel filtration alone.

H6-TEV-LSm4/8 forms aggregates of a non-defined oligomeric state. The formation of aggregates was confirmed by negative stain electron micrographs.

The H6-TEV-LSm5/6/7 heterotrimer runs as a hexamer on gel filtration chromatography. Negative stain electron micrographs reveal a homogeneous population of ring-shaped particles. Analytical ultracentrifugation shows formation of different oligomeric states in solution: trimers, hexamers, nonamers, and individual subunits.

Altogether, this work describes the expression, purification and characterization of 14 different LSm protein constructs. Some of them were expressed from monocistronic expression vectors and some from bicistronic vectors. Coexpression of LSm proteins enhances solubility, yield, and stability, irrespective of which partners were chosen.

The task of reconstituting LSm2-8 and LSm1-7 heteroheptameric rings was achieved using protein components overexpressed in E. coli. Complexes were reconstituted in the absence of RNA. Aggregates of individual subunits were disrupted using semi-denaturing conditions, and assembled complexes were purified using a combination of anion exchange and gel filtration chromatography. Sample homogeneity is illustrated on the anion exchange chromatography profile where complexes are eluting as single homogeneous peaks.

Negative stain electron micrographs (in collaboration with the group of Prof. A. Engel, Biozentrum Basel), are showing ring shaped particles indistinguishable from each other at the electron microscopy level. The overall dimensions of LSm2-8 particles are similar to those which have been purified from HeLa cell nuclear extract [124]. Negative stain electron micrographs of the LSm1-7 complex provide first evidence that this complex assembles into a structure that is very similar to the LSm2-8 complex.
The many crystallization attempts did not yield usable crystals. Crystals of the LSm5/6/7 diffracted to 8 Å. The best LSm6 crystals diffracted to 4.3 Å still not sufficient for deriving an atomic model. Crystals of the LSm2-8 complex were not diffracting at all and were observed once.

In order to test the functionality of the complexes, we performed electrophoretic mobility shift assays (EMSA) with U1 and U6 snRNA. Reconstituted LSm2-8 complex shifts U6 snRNA, whereas U1snRNA migration is not affected. Individual sub-complexes (LSm2/3, LSm4/8 and LSm5/6/7) are not able to slow migration of the U6 snRNA, neither does a reconstituted LSm preparation in which either of the sub-complexes or a single LSm protein has been left out. The LSm1-7 complex is not binding to U6 snRNA. Thus, the LSm protein complexes are showing their expected characteristics in this in vitro functional assays.

The functionality of the preassembled complexes was addressed using microinjection into somatic cells. The LSm1-7 and LSm2-8 complexes were reconstituted and coupled to fluorescent dyes. The LSm2-8 complex is found in the nuclear compartment 20 minutes after microinjection. In contrast, the LSm1-7 complex remains in the cytoplasmic compartment, although evenly distributed and not located in cytoplasmic foci, [126] possibly due to a higher level of the injected material compared to endogenous complex.

In summary, in our functional assays in vivo and in vitro, the recombinant LSm1-7 and LSm2-8 complexes show the expected characteristics of the native complexes.

It is of high importance to determine which parts of the LSm1 and LSm8 proteins have a decisive role in appropriate cellular localization, RNA target recognition and function. This could be investigated by constructing swap mutants, with an LSm1 carboxy terminal part joined to an N-terminal LSm8 section or vice versa.

Including such chimeras in LSm1-7/2-8 complexes, followed by microinjection into somatic cells could provide an answer to this question. Moreover, LSm4 has symmetrically methylated arginine-glycine repeats [86, 87]. LSm4 is part of both complexes, cytoplasmic and nuclear. Since they are similar to each other, it might be that the cytoplasmic LSm1-7 complex does not contain methylated arginine-glycine repeats, whereas the nuclear LSm 2-8 complex has methylated LSm4. This could represent a major factor in the functional differences between these
two complexes. It would be possible to investigate this issue by using antibodies recognizing symmetrically modified arginines.

Of great importance for understanding this model system will be further experiments in the direction of cell biology. This requires improvement and better control of experimental procedures, especially the stabilization of preformed, labelled complexes.

The structure-function relationships within the Sm/LSm protein family reflect three major interconnected features which illustrate why it is so important to solve the structures of Sm/LSm hetero-oligomeric complexes: First, Sm/LSm protein function is in general strictly dependent on complex formation. This holds for RNA binding, Sm/LSm-protein containing RNP biogenesis, interaction with non-Sm protein effector proteins, and RNA processing activity. The required interaction interfaces are apparently always structural sites/epitopes built up from several Sm/LSm subunits. High resolution structural information is clearly required to explain the molecular basis for this phenomenon. Second, exchange of only one or two subunits from one to another heterooligomeric (mostly heptameric) Sm/LSm complex changes its whole biology (see above). How such subtle structural changes can have these very large functional effects can only be addressed by solving the crystal structures of the respective complexes. Lastly, the ability of individual Sm/LSm proteins to assemble with different homologous binding partners to form architecturally very similar, yet functionally diverse complexes argues for a very fine balance between flexibility and specificity for the respective Sm-Sm interactions. Some Sm-Sm interfaces seem to be invariable; others tolerate the exchange of one partner. To this day, the only heterotypic Sm-Sm interfaces that have been characterised structurally remain the D1D2 and D3B interfaces from two human canonical Sm protein heterodimers. Clearly, in order to understand the "molecular recognition code" governing the specificity balance mentioned above, more structural information on such interactions is indispensable. The crystal structures of the LSm1-7 and LSm2-8 heptamers would advance this quest enormously.
5. REFERENCES


165


\[\text{arb}\] associates with the 5' splice site within large transcription/splicing complexes, The EMBO Journal, (23): 1782-1791


160. Evans S.K (2004) Telomeres, Current Biology, 10 (11), R418


168. Seraphin B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs, *The EMBO Journal*, 14(9):2089-98.


6. CURRICULUM VITAE

Born: 06.01.1972, Belgrade, Serbia and Montenegro

Nationality: Serbian

Education:

- 1996 B.Sci Department of Biochemistry, University of Belgrade, with average mark 9.17 (maximum 10)

- 1997 recipient of the Government fellowship for the young talents

- 1997-2000 M.Sci., Department of Biochemistry, University of Belgrade, with average mark 10 (10 is maximum)

- 06/2000 Recipient of Deutscher Akademischer Austauschdienst (DAAD) fellowship

- 07 and 08/2000 Language school, Goethe Institute Mannheim

- 10/2000-05/2001 Fellowship training at the University of Giessen, Strahlenbiologie

- 06/2001-03/2005 PhD work at Paul Scherrer Institute, in group of Prof. Dr. Fritz K. Winkler, under guidance of Dr. Christian Kambach
Publications and Presentations


Posters and talks


Zaric B. and C.Kambach, Reconstitution of two Recombinant LSm protein complexes: Architecture, Assembly and Function, (June 2004), Warsaw, FEBS meeting, poster presentation

Zaric B. and C.Kambach, Reconstitution of two Recombinant LSm protein complexes (May 2004), ETH Zürich, Electron Microscopy user meeting, poster presentation

Zaric B. and C.Kambach, Reconstitution of two Recombinant LSm protein complexes (May 2004), D-Biol. Symposium, DAVOS, Switzerland