We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,100
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

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

Various types of ncRNA, as well as RNPs containing tRNA, rRNA or snRNA, directly interact with mRNA at different stages of its life. Figure 1 presents an overview of the maturation of pre-mRNA and the fate of the mRNA generated. Pre-mRNA initially undergoes modification to enhance its stability: a 5’ methyl guanosine (m7G) cap added during transcription (Wen & Shatkin, 1999) and a poly(A)-tail placed in the 3’ region by the polyadenylation machinery (Proudfoot et al., 2002; Balbo & Bohm, 2007). Following initiation of spliceosomal assembly by recruitment of core particles in the cytoplasm, non-coding introns are spliced from the pre-mRNA sequence by the mature spliceosome in the nucleus (Crick, 1979; Pozzoli et al., 2002). This multi-megadalton complex itself contains 170 protein components and various types of snRNA, rivaling the ribosome in molecular complexity (Wahl et al., 2009).

Within the spliceosome, several distinct small nuclear RNP (snRNP) core complexes each contain snRNA organized around specific ring-structured protein assemblies. For those known as U1-, U2-, U4- and U5-snRNPs, these ring scaffolds are provided by members of the Sm protein family (Luhrmann et al., 1990), recruited to their specific snRNA partners in the cytoplasm at a distinct Sm-site of bases (Urlaub et al., 2001; Peng & Gallwitz 2004). The core snRNPs are reimported into the nucleus for further processing and spliceosome assembly (Will & Luhrmann, 2001; Patel & Bellini 2008). In contrast, U6 snRNA is first modified within the nucleoli and then engages with a related protein ring, in this case containing Lsm (“Sm-like”) proteins Lsm2-Lsm8. Together with the U1-U5 particles, the U6 snRNP is translocated to Cajal bodies for formation of the U4/U6*U5 tri-snRNP (Patel & Bellini, 2008). The mature snRNPs eventually assemble on pre-mRNA for intron removal steps (Will & Luhrmann, 2001; Patel & Bellini 2008).
Fig. 1. Lifecycle of mRNA from transcription to decay.
Following excision of introns, mRNA enters the cytoplasm via the nuclear pore complex to be either translated or degraded. In eukaryotes, two pathways are utilized for mRNA decay: i) 3'-to-5' degradation by the exosome or ii) 5'-decapping, followed by 5'-to-3' exonuclease degradation (Garneau et al., 2007). In either event, decay is initiated by shortening of the poly(A)-tail by deadenylases (Tucker et al., 2001; Garneau et al., 2007; Nissan et al., 2010). Protein machinery required for the 5'-decapping pathway is found enriched in cytoplasmic foci known as processing or P-bodies (Sheth & Parker 2003), which appear to control the sorting and storage of mRNA. Within P-bodies, a specific assembly of Lsm proteins (Lsm1-Lsm7) and ancillary protein factors expedites mRNA decapping and subsequent breakdown by ribonuclease (Nissan et al., 2010). While the extent to which mRNA decay is restricted to P-bodies is unclear, sequestered mRNA species are observed to leave P-bodies and may re-enter translation (Brengues et al., 2005).

2. Phylogeny of Lsm protein sequences

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

A characteristic feature of the Lsm proteins is their natural tendency to form ring-shaped quaternary complexes, each of a precise composition related to cellular location and RNA target (Beggs, 2005; Spiller et al., 2007). In prokaryotes and archaea, homomeric complexes of six or seven Lsm protomers appear to be functional, whilst discrete heteromeric assemblies of seven distinct Lsm proteins are found in eukaryotes. The individual Lsm proteins vary in size from 8-25 kDa (78-240 amino acids); representative sequences are depicted in Figure 2. Within each, a bipartite consensus sequence (designated Sm1 and Sm2 motifs) can be identified. These motifs arise from strands $\beta_1-\beta_3$ and $\beta_4-\beta_5$ of the core $\beta$-sheet structure, respectively. A variable stretch of residues between these conserved segments is created by a surface-exposed interconnecting loop (Kambach et al., 1999; Collins et al., 2001). The N- and C-terminal tail regions of each Lsm sequence are often highly charged and differ markedly between members; these are considered to provide contact points for additional protein or RNA interactions (Reijns et al., 2008; Reijns et al., 2009; Weber et al., 2010). In the case of the eukaryotic Lsm1 and Lsm4 proteins, these tail segments are notably elongated. The most highly conserved sequence segments across the Lsm family include specific amino acid sidechains implicated in RNA-binding. These are localized to two specific loop features, as outlined in Figure 2. For archaeal and eukaryotic Lsm proteins, sequence motifs Asp-x-$\phi$-Asn ($\phi$ = hydrophobic) and Arg-Gly-(Asp) (Kambach et al., 1999; Collins et al., 2001; Toro et al., 2001) are characteristic of loops L3 and L5, respectively. In bacterial Hfq, these RNA-binding segments occur as Asp-x-$\phi$-$\phi$-$\phi$ (L3) and Tyr-Lys-His (L5) (Schumacher...
Fig. 2. Structure-based Lsm protein sequence alignment. Sequences displayed are for *S. cerevisiae* Lsm3 (yLsm3), *H. sapiens* Lsm3 (hLsm3), *S. cerevisiae* SmD2 (ySmD2), human SmD2 (hSmD2), *M. thermoautotrophicum* Lsmα (MtLsmα) and *E. coli* Hfq (EcHfq). Shaded residues represent areas with ≥ 80% sequence homology. Secondary structure assignment is based on the crystal structure of yLsm3 (Naidoo et al., 2008). Red bars indicate conserved residues implicated in RNA binding. # indicates additional truncated residues not displayed. Boxed insert shows organization of other Lsm multidomain proteins: AD, anticodon binding domain; MTD, methyl transferase domain; DFDF, DFDF-x(7)-F containing domain; FFD, Y-x-K-x(3)-FFD-x-(IL)-S containing motif; TFG: [RKHT]-x(2-5)-E-x(0-2)-[RK]-x(3-4)-[DE]-TFG containing domain. CTD, C-terminal domain2.
The Lsm Proteins: Ring Architectures for RNA Capture

et al., 2002). For this bacterial ortholog, a highly conserved Gln residue on the N-terminal α-helix is also implicated in RNA-binding (Schumacher et al., 2002). Overall, the bacterial protein Hfq shows little sequence conservation with its archaeal and eukaryotic orthologs, yet the archaeal and eukaryotic Lsm proteins share some limited sequence similarity (>20%). The following Lsm-Sm protein paralogs are identifiable: Lsm1-SmB, Lsm2-SmD1, Lsm3-SmD2, Lsm4-SmD3, Lsm5-SmE, Lsm6-SmF, Lsm7-SmG, Lsm8-SmB (Fromont-Racine et al., 2000). These specific sequence relationships suggest the eukaryotic Lsm proteins to have evolved from a common archaeal ancestor in two waves (Khusial et al., 2005; Veretnik et al., 2009). A first gene duplication event likely created eight distinct Lsm proteins, from which later evolved the Sm protein group. The diversity of biological activities of Lsm proteins compared to their more specialized Sm counterparts supports this two-step evolution model (Beggs, 2005; Khusial et al., 2005). The presence of up to three Lsm proteins in archaea, as well as an Hfq-like protein in archaeal M. jannaschii, further supports a common ancestor of eukaryotic and archaeal Lsm proteins (Fischer et al., 2011).

A few multidomain proteins incorporating Lsm components have been observed (summarized, Figure 2). Lsm12 includes t-RNA and methyltransferase domains (Albrecht & Lengauer, 2004), and Lsm13, Lsm14 and Lsm15 all contain a central DFDF-x(7)-F domain (Albrecht & Lengauer, 2004; Anantharaman & Aravind, 2004). Lsm16 features a remarkably disrupted Lsm variant (lacking both the N-terminal α-helix and a complete β4 strand) in addition to FDF and YjeF-N domains (Albrecht & Lengauer, 2004; Tritschler et al., 2007). This protein is suggested to be dimeric in solution (Ling et al., 2008). The archaeal protein Pa-Sm3 contains an Lsm-like domain in addition to a C-terminal domain of unknown function adopting an α/β-fold (Mura et al., 2003).

3. Structures of Lsm protein ring complexes

Crystal structures of Lsm and Sm proteins from diverse sources today provide many high-resolution views of the ring morphology of their assemblies. As shown in Figure 3, Lsm rings have been observed to range 58-75 Å in diameter and to contain a central pore of 6-15 Å. Some crystal structures solved to date (Table 1) have been obtained in the presence of specific RNA partners. The recent solving of the human U1-snRNP structures containing the Sm assembly bound together with U1 snRNA and proteins U1-70K and U1-A have been significant and exciting advances (Pomeranz Krummel et al., 2009; Weber et al., 2010). These provide the first molecular detail of L/Sm rings bound to the highly intertwined protein-RNA network within RNP complexes.

Within the various Lsm ring assemblies, each protomer occurs as a highly bent five-stranded antiparallel β-sheet overlaid in most cases by an N-terminal α-helix (Figure 4A). The pronounced twist of the β-sheet aligns strand β5 against β1, so forming an SH3-type barrel loosely related to the OB-fold (Kambach et al., 1999; Collins et al. 2001). Strands β4 and β5 each present on opposite ends of the module, so providing interaction sites for adjacent Lsm subunits via β4-β5’ pairing (Figure 4). Stacking of five to eight protomers in such a manner ultimately results in the formation of the toroid assembly characteristic of all Lsm assemblies (Figure 4).

Within this ring organisation, the N-terminal amphipathic α-helices of each Lsm component are gathered across one face of the toroid, from which also project the unstructured N- and
Fig. 3. Selected crystal structures solved for Lsm assemblies. A) Pentamer, cyanophage ECX21941 (PDB 3BY7) 60 Å ring, 9 Å pore. B) Hexamer, of C. parvum Lsm5 (PDB 3PGG) 60 Å ring, 10 Å pore. C) Hexamer, S. aureus Hfq (PDB 1KQ1) 65 Å ring, 11 Å pore. D) Hexamer, A. fulgidus Sm2 (PDB 1LJO) 58 Å ring, 6 Å pore. E) Hexamer, A. fulgidus Sm2 (PDB 1LJO) 58 Å ring, 6 Å pore. F) Heptamer, A. fulgidus Sm1 (PDB 1I4K) 65 Å ring, 13 Å pore. G) Heptamer, M. thermoautotrophicum Lsma (PDB 1I81) 65 Å ring, 10-15 Å pore. F) Heptamer, S. cerevisiae Sm-F (PDB 1N9R) 65 Å ring, 10-15 Å pore. H) Octamer, S. cerevisiae Lsm3 (PDB 3BW1) 75 Å ring, 15 Å pore.
### Table 1. Crystal structures solved for Lsm assemblies (to 2010)

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID</th>
<th>Resolution (Å)</th>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>HsSmD3B</em></td>
<td>1D3B</td>
<td>2.00</td>
<td><em>H. sapiens</em></td>
<td>Kambach et al., 1999</td>
</tr>
<tr>
<td><em>Hfq</em></td>
<td>1KQ1</td>
<td>1.55</td>
<td><em>S. aureus</em></td>
<td>Schumacher et al., 2002</td>
</tr>
<tr>
<td><em>Hfq</em></td>
<td>1KQ2</td>
<td>2.71</td>
<td><em>S. aureus</em></td>
<td>Schumacher et al., 2002</td>
</tr>
<tr>
<td><em>AfSm2</em></td>
<td>1LJO</td>
<td>1.95</td>
<td><em>A. fulgidus</em></td>
<td>Toro et al., 2002</td>
</tr>
<tr>
<td><em>EhHfq</em></td>
<td>1HK9</td>
<td>2.15</td>
<td><em>E. coli</em></td>
<td>Sauter et al., 2003</td>
</tr>
<tr>
<td><em>PhHfq</em></td>
<td>1U1S</td>
<td>1.60</td>
<td><em>P. aeruginosa</em></td>
<td>Nikulin et al., 2003</td>
</tr>
<tr>
<td><em>PhHfq</em></td>
<td>1U1T</td>
<td>1.90</td>
<td><em>P. aeruginosa</em></td>
<td>Nikulin et al., 2003</td>
</tr>
<tr>
<td><em>MjSm</em></td>
<td>2QTX</td>
<td>2.50</td>
<td><em>M. jannaschii</em></td>
<td>Nielsen et al., 2007</td>
</tr>
<tr>
<td><em>CpLsm5</em></td>
<td>3PGG</td>
<td>2.14</td>
<td><em>C. parvum</em></td>
<td>Vedadi et al., 2007</td>
</tr>
<tr>
<td><em>AhHfq</em></td>
<td>3HFN</td>
<td>2.31</td>
<td>Anabaena sp.</td>
<td>Boggild et al., 2009</td>
</tr>
<tr>
<td><em>EhHfq</em></td>
<td>3GIB</td>
<td>2.40</td>
<td><em>E. coli</em></td>
<td>Link et al., 2009</td>
</tr>
<tr>
<td><em>SHfq</em></td>
<td>3HFO</td>
<td>1.30</td>
<td><em>Synchocystis sp.</em></td>
<td>Boggild et al., 2009</td>
</tr>
<tr>
<td><em>PaH57THfq</em></td>
<td>3INZ</td>
<td>1.70</td>
<td><em>P. aeruginosa</em></td>
<td>Moskaleva et al., 2010</td>
</tr>
<tr>
<td><em>PaH57AHfq</em></td>
<td>3M4G</td>
<td>2.05</td>
<td><em>P. aeruginosa</em></td>
<td>Moskaleva et al., 2010</td>
</tr>
<tr>
<td><em>BsHfq</em></td>
<td>3HSB</td>
<td>2.20</td>
<td><em>B. subtilis</em></td>
<td>Someya et al., 2010</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID</th>
<th>Resolution (Å)</th>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>MtLsmα</em></td>
<td>1I8L</td>
<td>2.00, 1.70</td>
<td><em>M. thermoautotrophicum</em></td>
<td>Collins et al., 2001</td>
</tr>
<tr>
<td><em>PaeSm1</em></td>
<td>1BF</td>
<td>1.75</td>
<td><em>P. aerophilum</em></td>
<td>Mura et al., 2001</td>
</tr>
<tr>
<td><em>AfSm1</em></td>
<td>1I4K</td>
<td>2.50</td>
<td><em>A. fulgidus</em></td>
<td>Toro et al., 2001</td>
</tr>
<tr>
<td><em>AfSm1</em></td>
<td>1I5L</td>
<td>2.75</td>
<td><em>A. fulgidus</em></td>
<td>Toro et al., 2001</td>
</tr>
<tr>
<td><em>MtLsmα</em></td>
<td>1JBM</td>
<td>1.85</td>
<td><em>M. thermoautotrophicum</em></td>
<td>Mura et al., 2003b</td>
</tr>
<tr>
<td><em>PaeSm1</em></td>
<td>1RI</td>
<td>1.75</td>
<td><em>P. aerophilum</em></td>
<td>Mura et al., 2003b</td>
</tr>
<tr>
<td><em>PaeSm1</em></td>
<td>1LNX</td>
<td>2.05</td>
<td><em>P. aerophilum</em></td>
<td>Mura et al., 2003b</td>
</tr>
<tr>
<td><em>PaHsm1</em></td>
<td>1H64</td>
<td>1.90</td>
<td><em>P. abyssi</em></td>
<td>Thore et al., 2003</td>
</tr>
<tr>
<td><em>PaHsm1</em></td>
<td>1M8V</td>
<td>2.60</td>
<td><em>P. abyssi</em></td>
<td>Thore et al., 2003</td>
</tr>
<tr>
<td><em>PaHsm3</em></td>
<td>1M5Q</td>
<td>2.00</td>
<td><em>P. aerophilum</em></td>
<td>Mura et al., 2003a</td>
</tr>
<tr>
<td><em>PaHsm1</em></td>
<td>1LOJ</td>
<td>1.90</td>
<td><em>M. thermoautotrophicum</em></td>
<td>Mura et al., 2003b</td>
</tr>
<tr>
<td><em>ScSmF</em></td>
<td>1N9R</td>
<td>2.80</td>
<td><em>S. cerevisiae</em></td>
<td>Collins et al., 2003</td>
</tr>
<tr>
<td><em>ScSmF</em></td>
<td>1N9S</td>
<td>3.50</td>
<td><em>S. cerevisiae</em></td>
<td>Collins et al., 2003</td>
</tr>
<tr>
<td><em>ScSm1</em></td>
<td>1TH7</td>
<td>1.68</td>
<td><em>S. solfataricus</em></td>
<td>Klic et al., 2005</td>
</tr>
<tr>
<td><em>U1-snRNP</em></td>
<td>3CW1</td>
<td>5.49</td>
<td><em>H. sapiens</em></td>
<td>Pomeranz Krummel et al., 2009</td>
</tr>
<tr>
<td><em>U1-snRNF</em></td>
<td>3PGW</td>
<td>4.40</td>
<td><em>H. sapiens</em></td>
<td>Weber et al., 2010</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID</th>
<th>Resolution (Å)</th>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>CphLsm</em></td>
<td>3BY7</td>
<td>2.60</td>
<td><em>Cyanophyceae</em></td>
<td>Das et al., 2009</td>
</tr>
<tr>
<td><em>ScLsm3</em></td>
<td>3BW1</td>
<td>2.50</td>
<td><em>S. cerevisiae</em></td>
<td>Naidoo et al., 2008</td>
</tr>
<tr>
<td><em>PfuQ8TZN2</em></td>
<td>1CYC</td>
<td>2.80</td>
<td><em>P. furiosus</em></td>
<td>Huang et al., 2004</td>
</tr>
</tbody>
</table>

1. Proteins are named by the first letters of the species, followed by the type of protein. Asterisked entries indicate structures solved in the presence of RNA.
2. Hypothetical protein adopting an Lsm fold.
3. Structure deposited without supporting publication.
C-terminal extensions. The opposite face of the ring, named the distal face, is predominantly composed of residues of the variable loop L4 segments. All the Lsm ring structures (across eukarya, archaea and bacteria) reveal clusters of positive residues lining the internal pore, as well as pronounced positive elements on the distal face (Toro et al., 2001; Brennan & Link, 2007; Naidoo et al., 2008).

The body of structural data adds to biochemical understanding concerning L/Sm-RNA interactions, and distinct RNA sites within the protein oligomer. These include i) a binding site within the lumen of the ring, ii) an external contact site on the helix face and iii) residues located on the distal face of the complex (Figure 4). The first of these sites engages residues from loops L3 and L5, contributed from all Lsm components to create a nucleotide-binding pocket running around the inner rim (Weber et al., 2010). The specific architecture and repeated circular location of these specific, highly conserved, sidechains enables one nucleotide base to be bound per L/Sm protomer. Crystal structures of archaeal and bacterial Lsm complexed with RNA clearly show the oligonucleotides to be threaded around this rim of the toroid (Toro et al., 2001; Schumacher et al., 2002). Each binding “slot” allows specific base stacking to a hydrophobic sidechain of loop L3, as well as contact with the signature Arg residues of loop L5 and H-bonding with Asn residues (strand β4). Further electrostatic contacts (involving conserved Asp (strand β2), Arg (loop L5) and Gly (loop L5) residues) enhance the stability of the Lsm-RNA complex (Toro et al., 2001). Figure 5 displays these relevant binding interactions for U5 within the lumen site of archaeal AfsSm1.

An external contact site for RNA at the helix face of the Lsm toroid (site ii) is suggested by the crystal structure of PaSm1 bound with U7 oligonucleotide (Thore et al. 2003). In this case, each of two sandwiched Lsm rings engage two nucleotides at the N-terminal α-helix (Arg, His) and strands β2 (Tyr) via base stacking and H-bonding.

A third distinct RNA-binding site (iii) is likely to be unique to the bacterial Hfq assembly, and its tripartite form has been detailed in the crystal structure of Hfq bound to poly(A) RNA (Link et al. 2009). The protein Hfq engages poly(A) sequences on its distal face via specific residues exposed from strands β2 and β4. There is, however, no evidence for poly(A) binding by eukaryotic Lsm proteins. In the structure of the Hfq/RNA complex, RNA contacts include electrostatic interactions from Lys (strand β2) and Gln (strand β4) sidechains, as well as stacking of bases between Tyr, Leu (strand β2) and Leu and Ile (strand β2’) of adjacent subunits. It is in this region of the toroid that sequence variability of the loop L4 across the Lsm family results in non-conservation of distal face chemistry, so explaining the unique binding properties of Hfq.

Within the crystal structures of the human U1-snRNP complex, multiple RNA interactions made by the ring of Sm proteins include binding sites i) and ii) outlined above (Weber et al., 2010). However, the U1-snRNP structure also clearly demonstrates the role of the Sm sequence extensions and loop regions as additional interaction sites, particularly the C-terminal extensions of SmD3 and SmB. In the lumen of the toroid (i.e. site i), snRNA threads to stack single nucleotides of the Sm site against the key loop L3 and L5 residues, notably the aromatic sidechains. From the helix face of the ring are projected residues of the N-terminal α-helix and loop L3 of SmD2, forming an external contact site (reminiscent of site ii) that guides the snRNA into the ring pore. Residues from the loop L2 regions of SmD1 and SmD2 appear to guide RNA out from the Sm ring. Protruding beyond the distal face, residues of the elongated L4 loops of SmD2 and SmB provide another important interaction point to clamp and secure a stem-loop of the snRNA.
Fig. 4. Lsm fold and quaternary structure. Ribbon diagrams of MtLsmα (A, B; PDB 1I81) are displayed. A) Dimer interface of MtLsmα. Chain A is represented in green, chain B in blue. Residues involved in hydrophobic packing at the dimer interface (Chain A: Ile27, Val77, Tyr78 of chain A; Chain B: Leu30, Phe36, Leu66, Val69, Ile71) are shown in stick representation. B-C) Top and side view of heptameric MtLsmα. D) Homo-heptameric MtLsmα (PDB 1I81). E) Homo-octameric yeast Lsm3 (PDB 3BW1). Space-filled models highlight in red conserved residues implicated in RNA binding: Asp in β2, Asn in L3, Arg and Gly in L5.
Fig. 5. Three general sites for RNA binding within specific examples of Lsm complexes. Site i) AfsSm1 (PDB 1I5L) bound to U1 RNA viewed from helix face. Site ii) Two PaSm1 (PDB 1M8V) heptamers are bridged by uridine heptamer. Site iii) EcHfq (PDB 3GIB) bound to poly(A) viewed from distal face. U1-snRNP) Figure includes side view of the Sm-core of the human U1 snRNP structure (PDB 3PGW).
The majority of crystal structures of Lsm obtained to date portray the hexa- and heptameric protein assemblies that correspond to fully functional homomeric or heteromeric protein groupings. It is, for instance, assumed that complexes of SmD1-SmD2, SmD3-SmB and SmE-SmF-SmG can exist independently in the cytoplasm, yet rearrange into mixed heptamers in the presence of RNA during snRNP formation (Peng & Gallwitz, 2004). However, a few crystal structures suggest that other compositions, e.g. pentamers and octamers, may be stable for eukaryotic Lsm (Naidoo et al., 2008; Das et al., 2009). While it is currently not clear if these organizations are peculiar to recombinant preparations of the Lsm family, they suggest possibilities for a variety of multimeric assemblies in vivo. Our own interaction studies indicate that Lsm assemblies may be relatively dynamic in solution, providing capacity to engage in alternative protein partnerships and stable groupings (Sobti et al., 2010).

4. Functional roles for Lsm proteins

Sm and Lsm proteins are known to interact with a diversity of RNA partner species. Specific RNA sequences recognized by various Lsm complexes include the Sm-site (A$_2$U$_5$GA) (Raker et al., 1999) and U-rich stretches at the 3’ end of oligoadenylated mRNA (Chowdhury et al., 2007) and RNA polymerase III transcripts, including snRNA (Achsel et al., 1999). Other binding partners include snoRNA (Kufel et al., 2003a), P RNA (Kufel et al., 2002), tRNA (Kufel et al., 2002) and rRNA (Kufel et al., 2003b). Depletion of Lsm proteins 2-5 and 8 in yeast results in defects in post-transcriptional processing of tRNA, P RNA, rRNA, snoRNA and snRNA precursors (Kufel et al., 2002; Kufel et al., 2003b; Kufel et al., 2003a). Yet only minor (or no) effects are observed on depletion of Lsm6 and Lsm7. A summary of some specific Lsm-ncRNA interactions is presented in Table 2.

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

A specific role for Lsm1-Lsm7 concerns activation of mRNA decay in P-bodies; depletion of individual yeast Lsm proteins results in the accumulation of capped, oligoadenylated mRNA transcripts (Boeck et al., 1998; Bonnerot et al., 2000; Bouveret et al., 2000; Tharun et al., 2000). This specific Lsm complex is recruited alongside other decay factors to U-rich tracts by the protein Pat1, after its displacement of cap-binding translation factors (Parker & Sheth, 2007). It is likely that Pat1 and Lsm1-Lsm7 are then involved in subsequent activation of the Dcp1-Dcp2 enzyme (Nissan et al., 2010). A variety of studies have demonstrated the interaction of Lsm1-Lsm7 with decapping factors and exoribonuclease Xrn1 (Bonnerot et al., 2000; Bouveret et al., 2000; Tharun et al., 2000; Coller et al., 2001).
<table>
<thead>
<tr>
<th>RNA species</th>
<th>Lsm function</th>
<th>Selected experimental evidence</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>snRNA</td>
<td>assembly, processing</td>
<td>Lsm2-8 binds 3’ end of U6 snRNA</td>
<td>Achsel et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lsm2-8 initiates structural rearrangements of U6 snRNA</td>
<td>Karaduman et al., 2006; 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depletion of Lsm2-Lsm8 results in splicing defects</td>
<td>Mayes et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Splicing activity recovered through recombinant Lsm proteins</td>
<td>Verdone et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lsm2-8 localizes U6 snRNA to the nucleus</td>
<td>Spiller et al., 2007</td>
</tr>
</tbody>
</table>

| tRNA | splicing, 3’ and 5’ end-processing | Accumulation of unprocessed pre-tRNA and reduced La/Lhp1 binding upon Lsm2-Lsm5 and Lsm8 depletion | Kufel et al., 2002 |
|      |               | Direct interaction of Lsm3 with tRNA and its splicing factors | Fromont-Racine et al., 1997 |

| P RNA | chaperone | Depletion of Lsm2-Lsm5 and Lsm8 reduces pre-PRNA levels | Mayes et al., 1999 |
|       |           | Reduced La/Lhp1 binding upon Lsm2-Lsm5 and Lsm8 depletion | Kufel et al., 2002 |
|       |           | Lsm2-Lsm7 proteins coprecipitate with pre-PRNA | Salgado-Garrido et al., 1999 |

| rRNA | 3’ and 5’ end-processing | Depletion of Lsm2-Lsm5 and Lsm8 delays pre-rRNA processing and increases rRNA decay rate | Kufel et al., 2003b |
|      |                         | Pre-rRNA coprecipitates with Lsm3 but not Lsm1 | Kufel et al., 2003b |
|      |                         | Deletion of Lsm6 and Lsm7 genes impairs 20S pre-rRNA processing | Li et al., 2009 |

| snoRNA | 3’ end-processing | Lsm2-Lsm5 and Lsm8 depletion results in U3-snoRNA degradation and loss of its 3’ extended precursor | Kufel et al., 2003a |
|        |                  | Reduced La/Lhp1 binding upon Lsm3 or Lsm5 depletion | Kufel et al., 2003a |
|        |                  | Lsm2-Lsm7 but not Lsm1 or Lsm8 coprecipitate with snR5 snoRNA | Fernandez et al., 2004 |
|        |                  | Lsm2-4 and 6-8 but not Lsm5 coprecipitate with U8 snoRNA | Tomasevic & Peculis, 2002 |

Table 2. Lsm binding interactions with ncRNA
In contrast to its enhancement of mRNA decay, however, the Lsm1-Lsm7 complex can also protect mRNA against 3' end trimming (He & Parker, 2001). This may involve steric hindrance of nuclease attack at mRNA locations on which Lsm1-Lsm7 and Pat1 proteins are bound.

5. Specific functions of bacterial Hfq

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

As for the eukaryotic Lsm proteins, Hfq is required for deadenylation-dependent mRNA decay. An RNase E-Hfq-sRNA complex is thought to function in translational repression and subsequent mRNA destabilization and degradation (Morita et al., 2005; Morita et al., 2006). Additional functions of Hfq include ATPase activity (Sukhodolets & Garges, 2003), cellular stress response and modulation of virulence in some bacterial strains (Tsui et al., 1994; Fantappie et al., 2009; Liu et al., 2010). Interestingly, the virulence of the multi-drug resistant human pathogen S. aureus was decreased in Hfq-deletion strains. (Liu et al., 2010).

6. Lsm proteins in human disease and viral replication

Aberrations in functions of Lsm proteins have been associated with a number of human diseases. Sm proteins are known to be targeted by auto-antibodies in systemic lupus erythematosis (Lerner & Steitz, 1979). In fact, the proteins were first identified in nuclear extracts of a patient suffering from this disease. A mutation of the SMN gene resulting in diminished assembly of snRNPs is the cause of spinal muscular atrophy (Lefebvre et al., 1995; Wan et al., 2005). Three Lsm proteins (Lsm1, Lsm3 and Lsm7) have now been directly connected to different cancer types. Lsm1 (also named cancer associated Sm-like protein, CaSm) was upregulated in pancreatic, prostate and breast cancer, as well as in several cancer-derived cell lines (Schwein fest et al., 1997; Fraser et al., 2005; Streicher et al., 2007). Remarkably, overexpression of antisense Lsm1 has been demonstrated to promote tumor reduction (Kelley et al., 2000; Kelley et al., 2001; Yan et al., 2006). Elevated levels of Lsm7 have been identified in malignant thyroid tumors, and a reduction in Lsm7 expression was observed in breast cancers (Conte et al., 2002; Rosen et al., 2005). The copy number and expression for the Lsm3 gene was found to be elevated in cervical cancer (Lyng et al., 2006). Observations concerning Lsm proteins in viral replication underlines some interesting functional diversity. Bacterial Hfq was initially described as a host factor required for phage Q8 replication (Franze de Fernandez et al., 1968). A role for Lsm1 as an effector of HIV replication has been reported (Chable-Bessia et al., 2009). It has also been suggested more recently that positive-strand RNA viruses may directly bind to the host Lsm1-7 protein complex via tRNA-like structures and A-rich stretches, so diverting normal mRNA
regulation (Galao et al., 2010). The requirement of host Lsm proteins for the replication of this class of virus has additionally been demonstrated in plant brome mosaic virus (Diez et al., 2000; Noueiry et al., 2003; Mas et al., 2006) and human hepatitis C virus (Scheller et al., 2009).

7. References


Bonnerot, C., Boeck, R., and Lapeyre, B. (2000), 'The two proteins Pat1p (Mrt1p) and Spb8p interact in vivo, are required for mRNA decay, and are functionally linked to Pab1p', Mol Cell Biol, 20.


Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007), 'The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs', RNA, 13.


Crick, F. (1979), 'Split genes and RNA splicing', Science, 204 (4390), 264-71.


He, W. and Parker, R. (2001), 'The yeast cytoplasmic Lsm1/ Pat1p complex protects mRNA 3' termini from partial degradation', Genetics, 158 (4), 1445-55.


Khusial, P., Plaag, R., and Zieve, GW. (2005), 'LSm proteins form heptameric rings that bind to RNA via repeating motifs', Trends Biochem Sci 30 (9).


The Lsm Proteins: Ring Architectures for RNA Capture

Licht, K., Medenbach, J., Luhrmann, R., Kambach, C., and Bindereif, A. (2008), '3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins', RNA, 14 (8), 1-7.


Sukhodolets, M. V. and Garges, S. (2003), 'Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq', *Biochemistry*, 42 (26), 8022-34.


RNA functions broadly as informational molecule, genome, enzyme and machinery for RNA processing. While these functions reflect ancient activities, they also remain vital components of contemporary biochemical pathways. In eukaryotic cells RNA processing impacts the biogenesis of RNA molecules of essentially every shape and function. The collection of articles in this volume describes the current state of understanding of the broad array of RNA processing events in animal and plant cells, key unanswered questions, and cutting edge approaches available to address these questions. Some questions discussed in this volume include, how viruses subvert the RNA processing machinery of the host cell, how the coordination of co-transcriptional RNA processing is regulated at the level of chromatin, the status of RNA processing in plant organelles, and how micro RNA machinery is biosynthesized and regulated.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:
