Lsm proteins and RNA processing

J.D. Beggs

Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, King’s Buildings, Mayfield Road, Edinburgh EH9 3JR, U.K.

Abstract

Sm and Lsm proteins are ubiquitous in eukaryotes and form complexes that interact with RNAs involved in almost every cellular process. My laboratory has studied the Lsm proteins in the yeast Saccharomyces cerevisiae, identifying in the nucleus and cytoplasm distinct complexes that affect pre-mRNA splicing and degradation, small nucleolar RNA, tRNA processing, rRNA processing and mRNA degradation. These activities suggest RNA chaperone-like roles for Lsm proteins, affecting RNA–RNA and/or RNA–protein interactions. This article reviews the properties of the Sm and Lsm proteins and structurally and functionally related proteins in archaea and eubacteria.

Introduction

Lsm (like Sm) proteins are so named because of their structural similarity to the previously characterized Sm family of proteins. Sm proteins are a set of seven small polypeptides that are found associated with several snRNAs (small nuclear RNAs) as snRNP [small nuclear RNP (ribonucleoprotein)] particles. The Sm proteins were first identified through the precipitation of these RNA–protein particles from nuclear extracts by anti-Sm antibodies from patients with the autoimmune disorder systemic lupus erythematosus [1].

Further studies identified seven Sm proteins, named B/B′ (two products of the same gene, produced by alternative splicing), D1, D2, D3, E, F and G, which range in size from 24 kDa down to 8.5 kDa. These form a heteroheptameric complex on binding to a conserved Sm site [consensus AU(4-6)G] found in single-stranded regions of U1, U2, U4 and U5 snRNAs that are involved in nuclear pre-mRNA splicing. The Sm proteins play a role in the maturation of these RNAs. U1, U2, U4 and U5 snRNAs are produced in the nucleus by RNA polymerase II and exported to the cytoplasm, where the Sm proteins bind to them and promote the hypermethylation of the N7-monomethyl guanosine cap at their 5′-ends, to produce the 2,2,7-trimethyl guanosine cap structure that is a characteristic of these snRNAs. The core snRNP particle is imported into the nucleus, mediated by the trimethyl guanosine-specific nuclear import receptor, snurportin 1. Additional proteins, which are specific to individual snRNAs, associate with the core snRNPs and the mature snRNPs assemble along with the U6 snRNP in spliceosome complexes that catalyse the splicing of intron-containing pre-mRNAs (reviewed in [2–5]).

All the Sm proteins contain an Sm domain that includes two short stretches of conserved amino acids, referred to as Sm motifs 1 and 2, that are separated by a variable region [6]. X-ray crystallographic studies of two heterodimeric complexes, containing the Sm polypeptides B/D3 or D1/D2, revealed how the Sm domain determines the structure of the Sm polypeptides. The four Sm proteins whose structures were determined showed identical folding patterns, with an N-terminal α-helix followed by a five-stranded antiparallel β-sheet [4]. Based on the structures of the two heterodimers plus biochemical information about how the polypeptides interact, Kambach et al. [4] proposed a model for the Sm core complex in which the seven Sm proteins formed a seven-membered ring with a small central aperture (Figure 1A). Indeed, a doughnut-shaped structure for the Sm core particle was suggested by electron microscopy of purified human U1, U2 and U5 snRNPs [7,8].

Sm proteins are found associated with U1, U2, U4 and U5 snRNAs in all eukaryotes; however, the other spliceosomal snRNA, U6, is not directly bound by Sm proteins. Unlike the other U snRNAs, U6 snRNA is produced by RNA polymerase III, acquires a 5′ γ-monomethyl phosphate cap and is generally considered to be retained in the nucleus [9]. Nevertheless, in the budding yeast Saccharomyces cerevisiae, two proteins were identified that contained Sm domains but were shown to be associated with U6 snRNA [10,11]. Homology searches of the sequence of the budding yeast genome revealed the existence of eight Sm-like open reading frames in addition to the seven Sm genes, and yeast two-hybrid screens detected interactions amongst these
Sm-like proteins and also between them and pre-mRNA splicing factors [12,13].

**Lsm protein complexes**

The Sm-like proteins were named Lsm1 to Lsm8. Of these, Lsm2 to Lsm8 were demonstrated to associate with U6 snRNA and to be necessary for its stability [14–16]. Identification of orthologous human cDNAs indicated that the Lsm proteins are also highly conserved [17]. The human Lsm2–8 proteins interact to form a stable RNA-free complex that is similar in shape to the Sm core RNPs [17]. The Lsm2–8 complex binds to a uridine-rich sequence at the 5′-end of U6 snRNA. It remains associated with U6 when it becomes complexed with U4 snRNA in the U4/U6 di-snRNP as well as in the U4/U6·U5 tri-snRNP [15,17–20].

Lsm1p, although not U6 associated, does interact with other Lsm proteins as shown indirectly, in two-hybrid screens [13]. In two-hybrid screens and co-fractionation experiments, Lsm proteins also interact with cytoplasmic proteins involved in mRNA decay, indicating the existence of an alternative, cytoplasmic Lsm complex with a role in mRNA turnover [13]. Indeed, there are at least two heptameric Lsm complexes: Lsm2–8, which is nuclear and a large fraction of which associates with U6 snRNA, and Lsm1–7, which is cytoplasmic and functions in mRNA degradation [21,22].

Sequence comparisons have aligned each of the Lsm proteins with one of the Sm proteins [12,16], and based on this a model can be proposed for the order of Lsm proteins in the heptameric ring complex, assuming the same arrangement as in the Sm complex (Figure 1B). This arrangement is partially supported by genetic studies in yeast [23] and fluorescence resonance energy transfer (FRET) analyses in human cells [24]; however, the precise organization of the Lsm complexes is yet to be determined.

**The U6-associated Lsm proteins**

In budding yeast, some of the Lsm genes are essential for viability, but lsm1Δ, lsm6Δ and lsm7Δ cells are viable, even though growth is heat-sensitive. This allows yeast cells that are deficient in Lsm1p, Lsm6p or Lsm7p to be propagated under permissive conditions (usually 23 °C) and the defect can be analysed. For the essential Lsm genes, conditional alleles can be generated by placing the gene under the control of a metabolically regulated promoter, such as the GAL1/10 gene promoter that is inducible by galactose and repressible by glucose. Depletion of any of the Lsm2 to 8 proteins but not of Lsm1p caused a reduced level of U6 snRNA and a mild pre-mRNA splicing defect. The residual U6 snRNA in lsm6Δ and lsm7Δ cells was found to be complexed with U4 snRNA in U4/U6·di-snRNPs, and the free U6 snRNA, normally present in excess, was absent or much reduced [15]. This prompted Verdone et al. [25] to investigate the nature of the defect in lsm6Δ and lsm7Δ cells. Extracts prepared from these mutant yeast cells were found to splice pre-mRNA as well as the wild-type extract, except that, after incubation with unlabelled pre-mRNA, mutant extracts were unable to continue to splice when challenged with additional radiolabelled pre-mRNA [25]. This suggested that an essential splicing factor became depleted during the first incubation and it was shown that this could be restored by incubation with the missing Lsm protein, produced as a recombinant polypeptide. As the limiting component was not ATP and the effect was dependent on incubation with a spliceable pre-mRNA, it was apparent that a splicing factor became limiting and that its regeneration depended on (at least) Lsm6 and Lsm7 proteins.

Following dissociation of spliceosomes, it is thought that, in order for U6 and U4 snRNPs to be recycled for further rounds of splicing, they must reassociate to form U4/U6·di-snRNPs and U4/U6·U5 tri-snRNPs; Prp24p, another U6-associated protein, has been proposed to promote recycling by facilitating association of the U4 and U6 snRNAs [26]. In an analysis of the fate of U6 snRNA in extracts from which U6 had been depleted, in vitro transcribed, radiolabelled U6 RNA was added back to reconstitute U6 snRNPs in the presence or absence of Lsm6p or Lsm7p. It was then shown that the Lsm6 and Lsm7 proteins were required for the efficient regeneration of annealed U4/U6 snRNAs in U4/U6·di-snRNPs and for the formation of functional U4/U6·U5 tri-snRNPs. These results indicate a role for nuclear Lsm2–8 proteins as RNA chaperones in modifying RNA–RNA and/or RNA–protein interactions, probably in association with Prp24p [25] (Figure 2).

**Lsm proteins interact with other stable nuclear RNAs**

The nuclear Lsm proteins have other functions in addition to supporting pre-mRNA splicing. Yeast lsm2, lsm3 and lsm8 mutants affect the level of pre-P RNA (the precursor of the RNA component of RNase P) [15], and Lsm2–7p, but not Lsm1 or Lsm8p, co-precipitated low levels of pre-P RNA but not of mature P RNA [16]. Lsm proteins may therefore play
Depletion of any of the essential Lsm proteins was also observed to result in a delay of pre-rRNA processing and aberrant processing intermediates accumulated as well as high levels of degradation products from both precursors and mature rRNAs. Many pre-rRNA species could be co-precipitated with tagged Lsm3p, indicating their direct interaction with an Lsm protein complex. These observations suggest a possible role for Lsm proteins to maintain the strict order of pre-rRNA processing events and possibly to facilitate RNA–protein interactions and structural changes required during ribosomal subunit assembly [28].

Many snoRNAs as well as snRNAs undergo 3′-end maturation, such as cleavage by the endonuclease Rnt1p and trimming by 3′ → 5′ exonucleases, including the exosome complex, which also participates in the 3′-maturation of the 5 S rRNA. Kufel et al. [29] analysed the maturation of the U3 snoRNA and found that normal processing of its 3′-end requires the essential Lsm proteins; depletion of any of Lsm2 or 5 or Lsm8p resulted in loss of the 3′-extended precursors and accumulation of truncated fragments of both mature and pre-U3 RNAs. In wild-type extracts, pre-U3 species could be co-precipitated with tagged Lsm3p, suggesting a likely transient association of Lsm proteins with the precursors. Lhp1p (yeast La protein) also binds to the 3′-extended U3 precursors. Results with Lsm-depleted and Lhp1-depleted extracts suggested that the interactions of Lsm proteins and of Lhp1p may be interdependent, consistent with an Lsm complex functioning as a chaperone in conjunction with Lhp1p to stabilize pre-U3 RNA species during 3′-processing [29]. Lhp1p binds to poly(U) tracts at the 3′-ends of many other newly synthesized RNA polymerase III transcripts, including precursors to tRNAs, 5 S rRNA and U6 snRNA. Lhp1p binding is thought to protect the 3′-end of many other newly transcribed U6 snRNA against degradation [14] and to stimulate the maturation of tRNA 3′-ends [31]. Genetic interactions have been found between LHP1 and LSM genes and it has been proposed that Lhp1p acts redundantly with the Lsm2–8 complex to stabilize nascent U6 snRNA [23].

With the exception of U6 snRNA, in all the above studies of the influences of Lsm proteins on stable nuclear RNAs, the effects of the Lsm6 and Lsm7 proteins (as well as of the Lsm1 protein) were minimal or absent. It may be that the non-essential Lsm6p and Lsm7p can be replaced by other Lsm proteins or by related Sm proteins in complexes that remain at least partially active. Therefore it is not clear, at present, whether the Lsm2–8 complex normally functions in these processes or whether alternative complexes exist in which Lsm6p and/or Lsm7p are replaced by other proteins that confer specificity for these processing pathways (Figure 3). Similarly, it has been reported that Lsm2–7 proteins associate, in the apparent absence of Lsm8p, with snR5, a yeast snoRNA that functions to guide site-specific pseudouridylation of rRNA [32], and it is not known whether another polypeptide replaces the missing Lsm protein to form a novel heteroheptameric complex. The Lsm proteins are not required for the pseudouridylation function of snR5 and, as for the interactions with pre-tRNAs, pre-rRNAs and snRNPs, indicating where defects occur as a result of depleting Lsm6p or Lsm7p.

Figure 2 | Spliceosome cycle of snRNAs, indicating where defects occur as a result of depleting Lsm6p or Lsm7p.

Schematic representation of the interactions of spliceosomal snRNPs while they form splicosomes. For simplicity, proteins are omitted. Thick lines and bars indicate the steps at which a lack of Lsm6p or Lsm7p results in defects in snRNP interactions in extracts from lsm6Δ or lsm7Δ. It is not certain whether Lsm proteins remain associated with the U6 snRNA in the spliceosome during the splicing reaction. A looser association of at least some of the Lsm proteins around the time of activation of the fully assembled spliceosome has been reported [49].
There are at least two separate Lsm protein complexes
Schematic representation of how Lsm proteins in S. cerevisiae interact with each other and with other factors to participate in different cellular processes. The compositions of the Lsm complexes that affect pre-snoRNA, pre-trNA, pre-rRNA and pre-P RNA processing are uncertain, as indicated by the dotted lines.

pre-snoRNAs, it is possible that the Lsm proteins may associate only transiently with snR5, rather than being components of the mature RNP particles. Also, in Xenopus, Lsm2–4p and Lsm6–8p bind to the U8 snoRNA, apparently without Lsm5p, and were postulated to facilitate the interaction between the 28S and 5.8S rRNAs [33].

Interestingly, the human U7 snRNP is a clear example of a mature RNP complex that has an unusual core composition. The HeLa U7 snRNP contains the Sm proteins B, D3, E, F and G, plus two non-canonical Sm-like proteins, Lsm10p and Lsm11p, instead of the Sm proteins D1 and D2. The presence of the Lsm10 and Lsm11 proteins alters the binding specificity of the core complex to favour the U7 RNA’s Sm-binding site, which differs from the consensus Sm-site, and Lsm11p contributes to the function of the U7 snRNP in processing the 3′-ends of histone mRNAs [34,35]. Thus different combinations of Sm and Sm-like proteins can form distinct polypeptide complexes that have different RNA-binding specificities. These may interact transiently with their RNA targets, thus affecting their maturation, as appears to be the case with pre-trRNAs and pre-rRNAs, and/or they may become stable components of RNPs, contributing to their functions, as with the U7 snRNP.

Conditions have been found that permit the replication of BMV (brome mosaic virus) genomic RNA in yeast, allowing a genetic analysis of this process. Lsm1p was shown to be an essential host factor for BMV RNA replication in yeast [36]. Since BMV has a positive-stranded RNA genome that lacks a poly(A) tail, Lsm1p may facilitate the association of replication factors with the genomic RNA in a manner analogous to its recruitment of decapping activity to deadenylated mRNA [36].

Lsm proteins and mRNA turnover
The interaction of Lsm proteins in two-hybrid screens with cytoplasmic proteins involved in mRNA decapping (Dcp1p and Dcp2p) or mRNA turnover (Xrn1p and Pat1/Mrt1p) [12] suggested a role for Lsm proteins in cytoplasmic mRNA decay. Indeed, an allele of *LSM1*, called *spb8*, was isolated as a suppressor of a deletion of *PAB1*, which encodes the poly(A)-binding protein, and *spb8* resulted in a defect in mRNA decapping [37]. mRNA stability studies showed that depletion of Lsm1p to 7p or of Pat1p, but not that of Lsm8p, resulted in increased stability of mRNAs and accumulation of full-length, capped transcripts with shortened poly(A) tails. This led to the demonstration of a separate, cytoplasmic Lsm protein complex, containing Lsm1p instead of Lsm8p, that was required for the efficient removal of the 5′-cap from mRNA, thereby facilitating the 5′→3′ degradation of mRNA in the deadenylation-dependent pathway of mRNA turnover [21,22] (Figure 3). Purification of the yeast Lsm1–7 complex co-purified Xrn1p, the major 5′→3′ exonuclease involved in mRNA decay, as well as Pat1p/Mrt1p, a translation initiation factor also implicated in mRNA decapping [22]. In addition, a small amount of mRNA co-precipitated with Lsm1–7 proteins, and an RNA-dependent association of Lsm1–7p with Dcp1p was demonstrated, suggesting the possibility of an mRNA-mediated interaction between the Lsm1–Pat1 complex and the Dcp1/Dcp2 decapping enzyme [21]. A human homologue of Lsm1p has been identified and evidence was produced for a corresponding Lsm1–7 complex that co-localizes with Xrn1 and Dcp1/Dcp2 decapping enzyme, concentrated in foci in the cytoplasm of human cells [24]. Indeed, studies in yeast indicated that mRNA degradation intermediates co-localize with Lsm1p, Dcp1p and other decapping factors as well as with Xrn1p in discrete cytoplasmic foci. These and other results suggest that cytoplasmic mRNA decapping and 5′→3′ degradation occur in these foci, termed processing bodies or ‘P bodies’ [38].

In addition to the effects on mRNA decapping, mutations affecting *LSM1–7* or *PAT1* result in the accumulation of mRNAs that are shortened at their 3′-ends by 10–20 nt. This ‘trimming’ effect is independent of decapping and is accelerated at high temperatures. These results suggest that the Lsm1–Pat1 complex may have a distinct role in protecting mRNA against 3′-end trimming and may partly explain the temperature-sensitive growth of *lsm* and *pat1* mutant strains [39].

Lsm proteins are also involved in mRNA turnover in the nucleus. To distinguish between nuclear and cytoplasmic mRNA turnover, export of nuclear RNA to the cytoplasm was blocked by using a yeast strain that has a defect in the nucleoporin Nup145p. The stabilities of several nuclear-restricted mRNAs were shown to be significantly increased after the depletion of Lsm6p or Lsm8p, but not after the depletion of the cytoplasmic Lsm1p [30]. These nuclear-restricted, stabilized mRNAs remained polyadenylated, and Lsm8p could be cross-linked to the nuclear poly(A) RNA, indicating a direct interaction. Pre-mRNAs that contain
What is the origin of Sm/Lsm proteins?

Sm-like proteins are also present in Archaeabacteria [15,16]. These form doughnut-shaped complexes that have remarkable structural similarity to Sm and Lsm complexes, indicating an ancient origin for this family of proteins [40,41] (Figure 4A). Different species of *Archaeoglobus* contain either one or two Sm-like proteins, suggesting that a gene duplication event gave rise to these related proteins [40]. Similar to Lsm proteins, the *A. fulgidus* proteins associate with RNase P RNA, probably indicating a role in tRNA processing [41]; however, it is not clear whether AF-Sm proteins interact with P RNA in a stable complex or in a transient interaction during pre-P RNA maturation. Recombinant AF-Sm1 protein of *A. fulgidus* forms a homoheptameric complex *in vitro* in the absence of RNA (similar to Lsm proteins) [41]. Recombinant AF-Sm2 protein forms complexes only in the presence of RNA (similar to Sm proteins) [40] or at low pH [42] and has been reported to form a homoheptameric complex as visualized by electron microscopy [40] or a homohexameric complex in crystallographic studies [42] (Figure 4A). It is not known whether these alternative forms of the AF-Sm2 complex exist *in vivo*. It is possible that archaeal Sm-like proteins may represent a primitive form of the Sm proteins of eukaryotic snRNPs, consistent with archaeal organisms being closely related to the precursor of eukaryotic nuclei. The Lsm and Sm groups of proteins may have arisen through further gene duplication events in eukaryotes followed by divergence to acquire their distinct functions and cellular locations.

Eubacteria also contain proteins with an Sm-like fold. Crystallographic analyses of Hfq show a striking similarity to the Sm and Lsm proteins and it forms a homohexameric ring-shaped complex [43,44] (Figure 4B). Hfq (also known as HF-1) was originally identified in *Escherichia coli* as a host factor required for bacteriophage QB replication, but was later found to affect multiple cellular processes, including the regulation of translation of certain target mRNAs and the degradation of others (reviewed in [45,46]). Hfq binds to the 3′-ends of some mRNAs, stimulating their polyadenylation and thereby targeting them for degradation. Hfq also binds a variety of small non-coding RNAs (sRNAs or ncRNAs) that function as post-transcriptional regulators, or riboregulators, controlling the stability of some mRNAs and affecting the translation of others. Some of these sRNAs are protected by Hfq against endonucleolytic cleavage by RNase E, which is involved in the processing and also the degradation of RNAs. In addition, Hfq facilitates the interaction of sRNAs with their target mRNAs; for example, the interaction of Spot 42 RNA with its target galK mRNA. Spot 42 RNA inhibits the translation of galK mRNA by binding to its translation initiation region. Hfq interacts with both Spot 42 RNA and galK mRNA, facilitating their interaction (reviewed in [46]). Mutagenesis and RNA binding studies have indicated that Hfq has independent RNA interaction surfaces for binding to sRNA and its target mRNA [47]. Possible mechanisms of action previously proposed to explain how Hfq promotes intermolecular base-pairing include: the functioning of Hfq as a chaperone, unfolding or refolding one or both of the RNAs, and/or by binding of Hfq to two different RNA molecules simultaneously, thus increasing their chance of interaction through proximity [45,48].

Conclusions

Sm and Sm-like proteins represent a large family of structurally and functionally related proteins of ancient origin that have apparently arisen through gene duplication and divergence, giving rise to complexes of proteins having different...
substrate specificities and functions. A common property of the various protein complexes is that they interact with RNAs, protecting them against inappropriate nuclease activity and/or modifying their structures, in many cases affecting their interactions with other RNAs or with proteins. It seems quite probable that more members of this protein family are yet to be discovered, especially in higher eukaryotes, where the results of studies with U7 snRNP indicate that the particular combination of Sm and Sm-like proteins present in a complex determines the specificity of RNA binding by the complex. This suggests the potential for enormous versatility in these complexes and we can look forward to many interesting discoveries in the future.

I thank my past and present colleagues for their invaluable contributions to the work described here, and M. Reijns for comments on this manuscript. I thank the Biochemical Society and Novartis for contributions to the work described here, and M. Reijns for comments. I thank the Biochemical Society and Novartis for contributions to the work described here, and M. Reijns for comments. I thank the Biochemical Society and Novartis for contributions to the work described here, and M. Reijns for comments.

References

10 Cooper, M., Parkes, V., Johnston, L.H. and Beggs, J.D. (1999) EMBO J. 18, 2066-2075
20 Vidal, V.P., Verdone, L., Mayes, A.E. and Beggs, J.D. (1999) RNA 5, 1470-1481
34 Pillai, R.S., Williams, C.L., Schumperli, D. and Muller, B. (2001) EMBO J. 20, 5470-5479

Received 28 January 2005