Reconstitution of Two Recombinant LSm Protein Complexes Reveals Aspects of Their Architecture, Assembly, and Function

Bozidarka Zarič, Mohamed Chamiš, Hervé Rémy, Andreas Engel, Kurt Ballmer-Hofer, Fritz K. Winkler, and Christian Kambach

From the Paul Scherrer Institut, Biomolecular Research, CH5232 Villigen and the M. E. Müller Institute for Microscopy, Biozentrum, the University of Basel, Klingelbergstrasse 70, CH4056 Basel, Switzerland

© 2005 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in U.S.A.

Received for publication, December 22, 2004, and in revised form, February 7, 2005 Published, JBC Papers in Press, February 11, 2005, DOI 10.1074/jbc.M414481200

The Sm and Sm-like (LSm) proteins form complexes engaging in various RNA-processing events. Composition and architecture of the complexes determine their intracellular distribution, RNA targets, and function. We have reconstituted the human LSm1–7 and LSm2–8 complexes from their constituent components in vitro. Based on the assembly pathway of the canonical Sm core domain, we used heterodimeric and heterotrimeric sub-complexes to assemble LSm1–7 and LSm2–8. Isolated sub-complexes form ring-like higher order structures. LSm1–7 is assembled and stable in the absence of RNA. LSm1–7 forms ring-like structures very similar to LSm2–8 at the EM level. Our in vitro reconstitution results illustrate likely features of the LSm complex assembly pathway. We prove the complexes to be functional both in an RNA bandshift and an in vivo cellular transport assay.

The Sm and Sm-like (LSm) proteins are a widespread protein family with members in all kingdoms of life. Phylogenetic distribution suggests Sm proteins were already present in the last universal common ancestor of all present-day life forms and that the family underwent an explosive diversification with the advent of eukaryotes (1). Archaeabacteria harbor one or two SM/LSM genes each. Escherichia coli Hfq and its orthologues in other Gram-negative bacteria are so far the only known eubacterial LSm proteins (2, 3). In contrast, eukaryotic genomes appear to contain 24 or more Sm/LSm genes (1, 4). Thought to have originally arisen as chaperones mediating RNA-RNA interactions (5), Sm/LSm proteins have diversified through evolution and adopted new functionalities. LSm protein function in archaea is unknown. A structural and biochemical study on Archaeoglobus fulgidus LSm proteins showed that they bind to RNase P RNA in vivo and in vitro (6), a feature that has also been observed for several yeast LSm proteins (7). E. coli Hfq is a pleiotropic regulator of RNA metabolism (8).

The originally identified canonical Sm proteins engage in pre-mRNA splicing (9). Sm/LSm family members have been shown to participate in mRNA decapping and degradation (10, 11), histone pre-mRNA processing (12–14), telomere synthesis (15), rRNA maturation (16, 17), small nuclear ribonucleoprotein assembly (18), pre-tRNA processing (19), and trans-splicing (20, 21).

Sm/LSm proteins are characterized by a bipartite sequence motif of about 80 amino acids long situated in most members at the N terminus. Recently, divergent family members with additional domains have been identified (1, 4). The conserved motif translates into a fold common to all Sm/LSm proteins. This Sm fold mediates specific Sm-Sm interaction through a generic interface, which the various Sm protein family members use to build up homomeric (in prokaryotes) or heteromeric (in eukaryotes) ring-shaped complexes. These represent the functional form of all Sm/LSm proteins. The common fold, generic interface, and ring-like morphology of Sm/LSm complexes provide a rationale for the observed large variety of RNA targets bound, the diverse complex compositions and functions. LSm proteins appear as building blocks for complexes whose composition and architecture determines their intracellular distribution, interaction with RNA targets and non-Sm effector proteins, and function. The structural basis for the balance between interaction specificity and flexibility required for assembling different complexes with some subunits in common is unknown.

The canonical Sm core domain composed of the seven Sm proteins B, D1, D2, D3, E, F, and G was demonstrated to assemble in an ordered pathway onto a conserved, single-stranded stretch on their target RNAs, the Sm site of the spliceosomal snRNAs U1, U2, U4, and U5 (22, 23) (Fig. 1). The pathway is marked by the RNA-free sub-complexes D1D2, D3B, and EFG (22). The sub-complexes may constitute stages of the assembly pathway. Sm-snRNA assembly occurs in the cytoplasm. After hypermethylation of the snRNA moiety, the pre-snRNPs are transported to the nucleus, where they mature to functional particles (9). In vivo, Sm core domain assembly is a highly regulated process involving Sm protein modification by a methylase (24) and numerous assembly factors like the survival of motor neurons protein. In vitro, the Sm core domain can be assembled from Sm protein sub-complexes by the addition of target snRNA in the absence of any auxiliary factors (23).2 Spliceosomal U6 snRNA differs from the other snRNAs in many ways. It is thought to have an entirely nuclear life cycle (25, 26), does not bear an Sm site, and does not bind the canonical Sm proteins. However, a complex built up of seven Sm-like (LSm) proteins 2–8 was shown to interact with the 3′

2 C. Kambach, S. Waeye, and K. Nagai, unpublished observations.
end of U6 snRNA in the nucleus (27), stabilizing U6 snRNP and the U4/U6 snRNA interaction (Fig. 1).

The Sm core domain can only assemble onto its U snRNA target and is only stable in the presence of the RNA (23). In contrast, the native LSm2–8 complex has been shown to be stable in the absence of RNA (27). It is likely that the LSm2–8 complex is assembled in the cytoplasm, migrates as such to the nucleus, and there binds to U6 snRNA. The LSm2–8 assembly thus differs from the canonical core Sm domain pathway. In addition to LSm2–8, a cytoplasmic LSm1–7 complex exists that engages in mRNA decapping and degradation (10, 11). The two complexes have LSm proteins 2 to 7 in common, differing only in the seventh subunit (LSm8 and LSm1, respectively, Fig. 1). The LSm1–7 assembly pathway is even less well characterized than the LSm2–8 pathway. LSm1–7 has been shown to accumulate in cytoplasmic foci together with other components of the mRNA decapping and degradation machinery (28, 29). These foci are apparently active sites of mRNA turnover (30), but the available data do not indicate whether LSm1–7 assembles in these foci or elsewhere in the cytoplasm, nor whether it binds its mRNA targets as a preassembled complex. It is unknown whether LSm protein sub-complexes analogous to the Sm heterodimers and heterotrimers exist. Here we show that stable, soluble human LSm23, LSm48, and LSm567 sub-complexes corresponding to their paralogues SmD1,D2, SmD1,B, and SmEFG can be obtained by coexpression in E. coli. LSm1 and LSm4 are produced from monocistronic vectors. Isolated sub-complexes assemble into ring-like higher order structures, underscoring the preference of eukaryotic LSm proteins to associate with heterologous binding partners. The fact that both LSm1–7 and LSm2–8 complexes can be reconstituted from these components in the absence of RNA suggests that both species assemble in the cytoplasm and bind to their target RNAs as pre-assembled units. We show the recombinant LSm2–8 complex to be functional by in vitro band shift with U6 snRNA and by an in vivo cell microinjection/intracellular transport assay.

**MATERIALS AND METHODS**

**Cloning, Expression, and Purification of LSm1–8 Proteins and Sub-complexes—**LSM2, -3, and -5–8 were subcloned from a human lymphoma U937 cDNA library (Stratagene) into a modified pUC19 vector. LSM1 and LSM4 were subcloned from expressed sequence tag clones IMAG p989P2110673Q2 and IMAG p985A0418006Q2, respectively. Expressed sequence tag clones were obtained from the genetic resources center, Berlin (RZPD, available at www.rzpd.de). LSM2/3, LSM4/8, and LSM5/6/7 polyclustic T5 expression cassettes were constructed by successive compatible overhang cloning using engineered BamH1/BglII sites. The final cassettes were transferred to the pQE30 T5 expression vector (Qiagen, Basel). Expression constructs bear an MRGSH6 tag at the terminus of the first cistron, followed by a tobacco etch virus (TEV) cleavage site. LSM1 and LSM4 were subcloned into pQE30 as monocistrons. SG13009[pREP4] (for LSM1, LSm2/3, LSm4, and LSm5/6/7) or BLR[pREP4] (for LSm4/8) E. coli cells were transformed with plasmid DNA and plated out on selective media. LB starter cultures were grown at 30 °C overnight, and 2–12 liters of LB media were inoculated the next day. Cultures were grown to an A_600 of 0.8 at 37 °C and induced with 1 mM isopropyl 1-thio-D-galactopyranoside. Induction temperature was between 25 °C and 37 °C. Cells were harvested after 4–48 h of induction, depending on construct. Cell pellets were resuspended in lysis buffer (20 mM HEPES-Na, pH 7.50, 0.5–1.0 mM NaCl, 10 mM imidazole-Cl, pH 7.50, 5 mM β-mercaptoethanol), sonicated, and treated with DNase I. Insoluble material was removed by ultracentrifugation, and supernatants were purified by immobilized metal ion affinity chromatography (IMAC) on nickel-charged Hi-Trap chelating Sepharose columns (Amersham Biosciences). LSm proteins and sub-complexes were eluted with imidazole step gradients (60, 250, and 500 mM). If insufficiently pure, samples were subsequently dialyzed into 100 mM NaCl buffer without imidazole and subjected to ion exchange chromatography (100 mM to 1 mM NaCl). Samples were frozen in liquid nitrogen in ion exchange buffer. In some instances, the MRGSH6 tags were cleaved off by TEV protease (1:100 ratio, overnight at room temperature), and the sub-complexes purified by IMAC. Cloning and expression of Sm protein sub-complexes D1,D2 and D1,B has been described elsewhere (31). An SmEFG heterotrimer was produced from a pET15b vector (Novagen) and purified via consecutive IMAC and ion exchange chromatographies. 12% SDS-PAGE gels were run with equivalent amounts of total cell extracts at the time of induction (T_i) and time of harvest (T_h) for both soluble material and pellet (insoluble material). After staining with Coomassie Brilliant Blue R250 (MERK Biochemicals, Germany) and destaining, gels were scanned, and the bands corresponding to soluble and insoluble LSm proteins were integrated via densitometry using the program ImageJ 1.29x (Wayne Rasband, National Institutes of Health).

**Reconstitution and Purification of LSm1–7 and LSm2–8 Complexes—**Individual LSm protein or sub-complex preparations were incubated in 4 x urea, 1 x NaCl buffer for 2 h at 37 °C and then mixed in equimolar amounts for the assembly of the desired heptamer. The mix was incubated again for 2–5 h, and the sample was dialyzed against buffer with progressively less salt (1 M and 0.5 M NaCl) overnight at 4 °C. Reconstituted LSm1–7 and LSm2–8 were purified by consecutive gel filtration and anion exchange chromatographies.

**Electron Microscopy and Image Processing—**Samples were diluted at 10–20 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 (32). 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 were then aligned and classified iteratively. The class average with the best signal-to-noise ratio were selected and gathered in a gallery.

**Gel Permeation Chromatography/Static Light Scattering Analysis**—
LSm sub-complexes were run on a Superdex 200 HR10/30 gel permeation column using an ÄkTA Explorer FPLC (both Amersham Biosciences) coupled to a miniDAWN static light scattering analyzer and an OptiLab 95F refractometer (both Wyatt Technology Corp., CA). Data were analyzed using Wyatt’s ASTRA 4 software. Analytical gel filtrations were run on a Superdex 200 PC3/230 column using an Etan HPLC (Amersham Biosciences).

**Analytical Ultracentrifugation**—
LSm sub-complexes were subjected to an equilibrium sedimentation run in 20 mM HEPES-Na, pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol buffer on an Optima XL-A analytical ultracentrifuge (Beckman Coulter) at 12,000 rpm. Data analysis was performed using the program DISCREQ (50).

**Electromobility Shift Assays—**
Xenopus tropicalis U6 snRNA or X. laevis U1 snRNA (a kind gift from Iain Mattaj, EMBL Heidelberg) were in vitro transcribed and body-labeled with [32P]UTP. 20,000 cpm purified U snRNA was incubated with 5 pmol of an equimolar mix of the Sm D1D2, D3B, and EFG sub-complexes or EFG complexes and heptameric complexes in a buffer containing 20 mM HEPES-Na, pH 7.50, 300 mM NaCl, 5 mM MgCl₂, 10% glycerol, 0.5 μl ofRNasin (Promega) and 0.5 mg/ml yeast tRNA in a 10-μl assay at 30 °C for 1 h, then at 37 °C for 1 h. Samples were loaded on 6% native PAGE gels and run at 4 °C for 2.5 h, 160 V. Gels were autoradiographed for 14–16 h at -80 °C on x-ray film.

**Cell Microinjections—** REF52 rat fibroblasts were grown to 60–80% confluency on coverslips in Opti-MemTM medium with glutamine (In-Ref). Cells were microinjected at a concentration of 2.5 mg/ml.

**RESULTS**

Expression of canonical human Sm proteins in *E. coli* from single cistron vectors gives very low yields or insoluble protein. In contrast, high yields of soluble Sm proteins are obtained by coexpressing the SmD1D2, SmD3B, and SmEFG sub-complexes from polycistronic expression vectors (31). These correspond to the sub-complexes identified in HeLa cell nuclear extract (22).

Although some LSm proteins can be expressed more efficiently in a soluble form from monocistronic vectors than their canonical Sm protein paralogues, in general yield is very low and the obtained preparations tend to aggregate heavily. Based on our experiences with Sm protein coexpression, and to facilitate expression and purification of LSm proteins, we initially constructed polycistronic expression vectors encoding LSm2/3, LSm4/8, and LSm5/6/7 CDNA. These heterodimers and heterotrimers correspond to the canonical SmD1D2, SmD3B, and SmEFG sub-complexes, respectively. LSm1 and LSm4 were constructed and expressed as monocistrons for the reconstitution of LSm1–7. Expression yield and solubility of single-cistron LSm constructs could be greatly enhanced by fusing to two N-terminal Z tags (*Staphylococcus aureus* protein A IgG-binding domain), followed by a His₈ tag and a TEV cleavage site. This phenomenon is exemplified by the ZZ-His₈-TEV-LSm6 purification record (Fig. 2d). For crystallographic and other studies, we proceeded to express LSm5, LSm6, LSm8, and the complexes LSm5/6, LSm5/7, LSm6/7, and LSm5/3 (data not shown). In general, the solubility of a given LSm protein increased by up to 25-fold by coexpression, as measured by the supernatant:pellet ratio (Table I). Recombinant LSm protein sub-complexes were purified by Ni-IMAC followed by ion exchange chromatography where necessary. For each polycistron, only the first CDNA bears a His₈ tag; the other LSm proteins are isolated through sub-complex formation and co-purification. The complexes and single LSm proteins were purified to homogeneity, as shown by SDS-PAGE (Fig. 2, a–d). Sample integrity is further demonstrated by the successful crystallization of various LSm protein preparations. Weakly diffracting crystals could be obtained from LSm6 (Fig. 2e) and LSm5/6/7 (Fig. 2f).

The purified sub-complexes were characterized biophysically. Analytical ultracentrifugation (AUC) and static light scattering experiments combined with gel filtration chromatography yielded molecular weights that indicate formation of higher order structures (Fig. 3, a–c, and Table I): The LSm2/3 heterodimer has a nominal molecular mass of 25 kDa. In the analytical ultracentrifuge, the LSm2/3 oligomer distribution is bimodal at 10 μM concentration, containing a hexamer (10%) and an octamer (87%). The molecular mass of LSm5/6/7 is 33 kDa. Analytical ultracentrifugation yields a mixture of individual subunits (26%), trimer (25%), hexamer (40%), and nonamer (8%) species at 16 μM concentration. Analytical gel filtration combined with static light scattering measurements yields 85 kDa for LSm2/3 and 77 kDa for LSm5/6/7. These values reflect the heterogeneity in oligomer distribution found by AUC. LSm5/6/7 stays intact during gel filtration, and individual subunits are not observed. Upon incubation in up to 8 μM urea, the highest elution volume of LSm5/6/7 species corresponds to the trimer. This stands in contrast to LSm2/3, which at urea concentrations of 4 μM and higher, falls apart to some extent into its subunits (data not shown). LSm4/8 aggregates most strongly of the LSm sub-complexes and does not seem to form oligomeric higher order structures of defined stoichiometry (Table II).

Our concept on sub-complex higher order structure is confirmed by negative-stain electron microscopy. LSm2/3 (Fig. 4a, overview) shows up as ring-shaped structures with slightly smaller dimensions than the 8-nm outer diameter and 2 nm for the central hole that were measured for the native LSm2–8 complex from HeLa cell nuclear extract (27). LSm 5/6/7 shows up mainly as a ring-shaped structure as well but heterogeneities appear in the background of the electron micrographs (data not shown). After particles classification and subsequent class averaging, distinct ring particles can be observed in LSm2/3 galleries (Fig. 4a, bottom) having an outer diameter of ~7 nm and a cavity of ~1.5 nm. Considering the mass of the LSm2/3 heterodimer (25 kDa), we suggest that the resulting class averages correspond to octameric LSm2/3 ([LSm2/3]₈) in accordance with the AUC measurements. Although the LSm5/6/7 preparation did not show up as homogenous as LSm2/3, class averaging yielded ring-shaped particles having a size of ~7 nm and a cavity of 1.5 nm suggesting a hexameric arrangement ([LSm5/6/7]₆, 2 × 33 kDa) based on the AUC data. Nevertheless, from the electron microscopy analysis, a nonameric arrangement cannot be excluded. Smaller particles appear in the background (Fig. 4b, circles) and could represent LSm5/6/7 trimers. The small size of such particles was not suitable for classification.

The Sm core domain can be reconstituted in vitro from recombinant Sm sub-complexes and U snRNA with good efficiency under native buffer conditions (33). The LSm2–8 complex isolated from HeLa cell extract is, in contrast, stable and likely to assemble in the absence of RNA (27). For our LSm complex in vitro reconstitution protocol, we required the

---

3. C. Kambach, unpublished observations.
ruption of the higher order structures formed by the sub-complexes. The in vitro reconstitution process should then be guided by relative thermodynamic stability. Reconstitution was carried out by mixing equimolar amounts of LSm2/3, LSm4/8, and LSm5/6/7 (for LSm2–8) or LSm4, LSm1, LSm2/3, and LSm5/6/7 (for LSm1–7) under semi-denaturing conditions (for details, see “Materials and Methods”) followed by dialysis. The reconstituted complexes were then purified by ion exchange chromatography (Fig. 5, e and d) followed by gel filtration (Fig. 5, e and f). In both types of chromatography, they

Fig. 2. SDS-PAGE gels of LSm sub-complex purification protocols. a, LSm2/3; b, LSm4/8; c, LSm5/6/7; d, LSm6 (Z-tagged). a–d: lane 1: uninduced culture; lane 2: induced culture; lane 3: supernatant; lane 4: pellet. d: non-cleaved, Z-tagged LSm6 (lanes 1–6), TEV-cleaved LSm6 (lane 7). IMAC: immobilized metal ion affinity chromatography; IEX: ion exchange chromatography; FT: flow-through. e and f: crystals of LSm6 and LSm5/6/7, respectively.
LSm Complex Architecture

Solubilities are expressed as supernatant:pellet ratios (column: Ratio S:P), determined by band densitometry of SDS-PAGE gels as described under "Materials and Methods." Increase in solubility by heterologous coexpression is given for those LSm proteins that were also expressed as single cistrons, as judged by the change in the S:P ratio (column: Increase). Sub-complexes corresponding to assumed nearest neighbors in the LSm1–7 and LSm2–8 rings (see Fig. 1) are in bold, and non-nearest neighbor combinations are underlined.

| LSm protein | Construct | Supernatant | Pellet | Ratio S:P | Increase |
|-------------|-----------|-------------|--------|-----------|----------|
| LSm1        | LSm1      | 10090       | 10772  | 0.9       |          |
| LSm2        | LSm23     | 9349        | 2669   | 3.5       |          |
| LSm3        | LSm23     | 3060        | 2507   | 3.6       |          |
| LSm4        | LSm35     | 5797        | 2637   | 2.2       |          |
| LSm5        | LSm4      | 5292        | 12363  | 0.4       |          |
| LSm6        | LSm48     | 6734        | 6159   | 1.1       | 2.6      |
| LSm7        | LSm5      | 1543        | 8566   | 0.2       | 19.3     |
| LSm8        | LSm53     | 4546        | 1310   | 3.5       |          |
| LSm9        | LSm567    | 7561        | 3710   | 2.0       | 11.3     |
| LSm10       | LSm6      | 2154        | 20887  | 0.1       |          |
| LSm11       | LSm67     | 3614        | 1374   | 2.6       | 25.5     |
| LSm12       | LSm567    | 3751        | 1877   | 2.0       | 19.4     |
| LSm13       | LSm7      | 5751        | 3906   | 1.9       |          |
| LSm14       | LSm567    | 6514        | 4202   | 1.6       |          |
| LSm15       | LSm8      | 1717        | 8276   | 0.2       |          |
| LSm16       | LSm48     | 1722        | 1269   | 1.4       | 6.5      |

As research in genomics and RNA processing progresses, ever more proteins containing the Sm/LSm motif are discovered, and new functionalities of LSm protein complexes are identified. Still, very little is known about LSm complex assembly pathways, nor how the architecture of the often very similar complexes determines their specific function. Eukaryotic Sm/LSm proteins have a strong preference to form heterooligomers rather than homooligomers. Canoical Sm proteins form RNA-free heterodimers and heterotrimers that likely represent intermediates on the core snRNA interaction pathway. Specificity of Sm-LSm interaction impacts directly on the
assembly process, because lack of it must be overcome by the help of cellular assembly factors. Here we have presented results that show how LSm complex self-assembly can be successfully carried out \textit{in vitro} in the absence of such assembly factors and results in a correct architecture and functional heptameric and LSm2–8 and, presumably, the LSm1–7 complex.

LSm proteins tend to be more soluble than Sm proteins when produced singly. Nevertheless, providing another LSm protein as a heterologous binding partner in the same cell generally increases solubility by a factor of up to 25. In this way, we were able to produce soluble, stable LSm2/3, LSm4/8, and LSm5/6/7 sub-complexes, corresponding to SmD1D2, SmD3B, and SmEFG. However, the increase in solubility is independent of the combination and does not correlate with coexpression of assumed nearest neighbors in the LSm2–8 ring. We conclude there is a lower degree of LSm-LSm interaction specificity, as compared with the Sm-Sm interactions in the core snRNP domain. The results are in line with yeast two-hybrid data indicating a greater promiscuity for LSm than Sm proteins (39, 40). The findings impact on the cellular LSm2–8 assembly pathway: lower intrinsic interaction specificity puts a higher demand on assembly factors guiding productive ring assembly. Indeed, LSm2–8 assembly \textit{in vivo} could be promoted by snRNP assembly factors like survival of motor neuron, which has been demonstrated to interact with LSm4 \textit{in vitro} (41, 42). However, evidence that these interactions are also present \textit{in vivo} is as yet lacking.\textsuperscript{4}

We could show that the LSm2/3 and LSm5/6/7 sub-complexes assemble into higher order ring-shaped heterooligomers by negative-stain electron microscopy. From the AUC results that indicated predominantly octamers (tetramers of dimers) for LSm2/3, we assume the LSm2/3 rings represent octamers. However, all Sm or LSm rings reported so far have either six or seven subunits, and it remains to be proven that the generic Sm-Sm interface defined by the D3B and D1D2 heterodimers is capable of accommodating eight subunits in a ring. Alternatively, the LSm2/3 rings could be representing hexamers present in the LSm2/3 preparation at low concentration, in line with LSm5/6/7. Hexamer formation by an Sm sub-complex was previously demonstrated as a feature of the human EFG trimer (43). The physiological significance of this hexamer could not be demonstrated, and indeed the later establishment of the Sm core domain stoichiometry proved that the (EFG)\textsubscript{2} complex is not part of the final heptameric ring and most likely represents a storage form for the three proteins. Presence of the hexamer does not preclude heptamer formation \textit{in vitro}: recombinant EFG preparations also show the hexamer, but can efficiently be reconstituted into a functional Sm core domain by the addition

\textsuperscript{4} U. Fischer, personal communication.

**FIG. 3.** Gel filtration chromatograms (Superdex 200 HR 10/30 column) of LSm sub-complexes LSm2/3 (\textit{a}), LSm4/8 (\textit{b}), and LSm5/6/7 (\textit{c}). UV traces are in blue (280 nm) and red (260 nm).
Average molecular weights for LSm2/3 and LSm5/6/7 over the (single) gel-filtration peaks are given together with standard deviations and polydispersity figures as calculated by Wyatt's ASTRA software (see "Materials and Methods" for details). The terms "dimer," "trimer," etc. refer to multiples of the smallest heteromeric complex unit (2 for LSm2/3, 3 for LSm5/6/7, and 7 for LSm1–7). MMmean for the LSm5/6/7 run corresponds to the percentage monomers detected based on the averaged molecular weight for the three subunits LSm5, LSm6, and LSm7.

| Complex   | Average MW over peak (kDa) | Standard deviation % | Polydispersity % |
|-----------|----------------------------|----------------------|-----------------|
| LSm2/3    | 86                         | 2.0                  | 3.0             |
| LSm5/6/7  | 77                         | 2.0                  | 3.0             |
| LSm2–8    | 92                         | 1.2                  | 1.6             |

Losses of LSm sub-complexes

| Losses | MMmean |
|--------|--------|
| LSm2/3 binary complex | Trimer | Tetramer | Pentamer | Nonamer |
| %      | %      | %        | %        | %       |
| 40 µM  | 36     | 0        | 68       | 16      | 16      |
| 10 µM  | 7      | 10       | 87       | 0       | 3       |

Losses of LSm5/6/7 ternary complex

| Losses | MMmean |
|--------|--------|
| LSm5/6/7 | Monomer | Dimer | Trimer |
| %      | %      | %     | %      |
| 16.3 µM | 0      | 26    | 25     | 40      | 8       |

Losses of LSm1–7 heptameric complex

| Losses | Monomer | Dimer | Trimer |
|--------|---------|-------|--------|
| %      | %       | %     | %      |
| 50 µM  | 57      | 20    | 80     | 0      |
| 20 µM  | 59      | 45    | 55     | 0      |
| 10 µM  | 44      | 43    | 51     | 6      |

The reconstituted complexes elute as single peaks from ion exchange chromatography, demonstrating sample homogeneity in charge and in size. Because of the great variation in pI within the LSm subunits (from 4.3 for LSm8 to 10.0 for LSm4), it is thus very unlikely that the LSm1–7 and LSm2–8 preparations consist of several sub-populations, each lacking one particular LSm subunit and containing two of another instead. The SDS-PAGE gels of the purified heptamers clearly show the presence of all seven different subunits. Both LSm1–7 and LSm2–8 preparations are homogeneous in size as well: they elute as single, Gaussian peaks from gel filtration with elution volumes corresponding to the expected molecular weights of the heptamers. The accuracy of molecular weight determination for LSm2–8 by static light scattering is >6%. Because the smallest subunit, LSm6, has a molecular mass of 9.1 kDa, representing about 10% of the complex's mass, the value of 92 kDa obtained for LSm2–8 (nominal molecular mass = 86 kDa) is only compatible with a subunit number of seven. Similarly, the AUC analysis of LSm1–7 demonstrates the presence of a heptameric species in solution, which is at equilibrium with higher order oligomers, but not with smaller complexes like those observed in the sub-complex AUC runs. This result further illustrates the complex's stability and homogeneity. Taken together, sample homogeneity and composition together with the molecular weight determination results provide strong evidence for a "one of each subunit" stoichiometry of the recombinant LSm complexes, in line with the architecture of the canonical core Sm domain (44).

Negative stain electron micrographs show that recombinant LSm2–8 has a ring-like architecture with a diameter of ~8 nm. The shape and size are highly similar to the one previously observed for the native LSm2–8 complex isolated from HeLa cell nuclear extract (8 nm (27)) and core snRNP domain from the same source (45). The pore diameter we observed for the recombinant LSm2–8 complex is distinctly larger than in the native LSm2–8 complexes (3 versus 2 nm, respectively (27)). Because the recombinant complex shows the same RNA binding specificity, this difference must remain unexplained at the present time. The LSm1–7 rings appear to be slightly smaller,
measuring ~7 nm across and a pore diameter of <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, demonstrating that LSm1–7 architecture follows the generic Sm/LSm complex pattern.

A pore diameter of 15 Å agrees well with the range observed in archaeabacterial LSm protein complexes, free or bound to RNA. Distances vary from 8.8 Å for the narrowest point in Pyrobaculum aerophilum LSm1 (46) to about 13 Å in Archaeoglobus fulgidus LSm1 (6). The co-crystal structure of hexameric

---

**FIG. 5.** Reconstitution of LSm1–7 (a, c, and e) and LSm2–8 (b, d, and f) heptamers. SDS-PAGE gels show input sub-complexes or subunits (a and b), homogeneity in charge is demonstrated by the single peak elution profile from anion exchange chromatography (c and d), and in size by the elution profile from gel filtration chromatography (e and f, see text).
**LSm Complex Architecture**

**FIG. 6. RNA bandshifts.** [32P]UTP-labeled, *in vitro* transcribed U6 snRNA was incubated with different LSm sub-complexes or higher order structures reconstituted from them, run on native PAGE gels, and autoradiographed. Components present are indicated *underneath each lane*. Individual sub-complexes do not shift U6 snRNA (panel a, lanes 2–4), whereas LSm2–8 does (lane 5). LSm1–7 does not lead to complex formation with U6 snRNA (panel b, lane 3), nor do reconstituted particles lacking either LSm 2/3 (lane 4), LSm4/8 (lane 5), LSm5/6/7 (lane 6), or LSm6 (lane 7). Incubation of LSm2–8 with U6 snRNA in the presence of an LSm2/3-specific single chain Fv antibody leads to a complex shifted to higher molecular weight (panel c, lane 3). LSm2–8 does not shift U1 snRNA (panel d, lane 3), whereas a 1:1 mixture of the seven canonical Sm proteins does (lane 2). Under the same conditions, the Sm proteins lead to aggregation of U6 snRNA and shift to the well (panel d, lane 5). U6 snRNA complex formation with LSm2–8 is unaffected (lane 6).

*E. coli* Hfq with an AU6G RNA oligonucleotide shows that pore size increases from 12 Å for the RNA-free hexamer to 15 Å for the RNA complex (47). 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 (46). This stands in contrast to the original concept that in the core snRNP domain, the Sm site target RNA threads through the LSm2–8 complex (50). For LSm2–8-U6 snRNA interaction, the binding determinant has been shown to be the U5 stretch at the 3′ end of U6 snRNA (27). This target is freely accessible to a preassembled complex. Hence it is possible that the RNA 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′ untranslated regions 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′ untranslated regions are likely to prevent the RNA threading through the LSm1–7 hole. The established biochemical features of LSm1–7 and LSm2–8–RNA interaction fit very well with the concept that both LSm1–7 and LSm2–8 assemble in the absence of RNA, are transported to their site of action, and bind to their targets on site, possibly using different binding modes. Elucidation of the exact mode of LSm1–7 and LSm2–8–RNA interaction will have to await solution of the respective crystal structures.

Recombinant LSm2–8 binds to U6 snRNA *in vitro*, whereas LSm1–7 does not. The RNA binding characteristics of the two native complexes are thus reflected by their engineered counterparts. However, we do not as yet possess a suitably short RNA target to demonstrate specific interaction with LSm1–7. Indeed the precise nature of the binding determinants on target mRNA for the LSm1–7 complex is currently not known. The validity of using U6 snRNA interaction as a measure for LSm2–8 function is underscored by the fact that only the integral LSm2–8 complex specifically binds to U6. Leaving out a single LSm protein or one of the sub-complexes from the reconstitution procedure produces complexes incapable of binding U6 snRNA. This observation holds despite the likelihood that all these mixtures will form ring-shaped higher order structures, just as the sub-complexes themselves. Ring-shaped multimers are ubiquitous in nucleic acid binding complexes and other cellular processes (51–53). The ring architecture is thought in general to generate new biophysical properties on the resident protein subunits, and often to convey new functions (54). The failure of the LSm sub-complexes to bind U6 snRNA shows that the ring architecture and the presence of LSm family members in the complex are not sufficient for specific interaction. This goes in line with the need for strong RNA target discrimination based on the presence or absence of a single specific subunit.

Our cell microinjections of fluorescently labeled LSm complexes or proteins show that the intracellular distribution of the recombinant heptamers reflects the migration behavior of their native counterparts, implying that the *in vitro* reconsti-
tuted complexes are functional in vivo. LSm2–8 nuclear transport is active and not diffusive. Fluorescent labeling of the heptamers does not disrupt them. These observations provide some evidence that the transported species is the intact heptamer. In a transfection assay, LSm8 is found to accumulate in the nucleus (28). In contrast, in our cell microinjection assay, LSm8 (the subunit likely to bear the nuclear transport determinant of LSm2–8) fails to accumulate in the nucleus. The difference of our result could be due to the production mode of the protein and time course of the experiment: singly expressed, our recombinant LSm8 forms aggregates likely to mask a resident nuclear localization signal. The aggregates are probably also toxic to the cells, explaining the occurrence of pre-apoptotic granules. Conversely, within the ~36 h of the transfection experiment, it is conceivable that the YFP-labeled LSm8 (produced at levels only slightly higher than the endogenous protein) assembles into functional LSm2–8, which is then the transport substrate (28). On the basis of these experiments, nuclear migration of isolated LSm8 cannot be ruled out, however.

Our findings imply that we have to view the specific interactions and functions of individual LSm subunits in the context of the ring architecture: Exposure and probably juxtaposition of particular sequence elements in the subunits are likely to be instrumental in defining the interaction of the complex with its target RNA, with assembly factors (e.g., a presumptive nuclear import receptor for LSm2–8) and effector proteins (like the exonuclease Xrn1 and the decapping factor Dcp1/2 in the case of LSm1–7). Our recombinant LSm protein complexes represent an ideal test system to study these interactions in molecular detail. Our results should contribute to the understanding of the pathway of LSm complex assembly and its regulation of LSm-RNA and LSm-protein interaction and function.

Acknowledgments—We thank M. Steinmetz and U. Kutay for critical reading of the manuscript, Tewfik Soulimane for assistance in crystallography, H. J. Schönfeld and B. Pöschl, Roche Applied Science, Basel, for measurement of the static light scattering data, E. Kusznir and F. Müller, Roche Applied Science, for running and data analysis of the analytical ultracentrifugation, and the other members of the group for fruitful discussions and practical advice.

REFERENCES

1. Anantharaman, V., and Aravind, L. (2004) BMC Genomics 5, 45
2. Moller, T., Franch, T., Hjørup, P., Keene, D. R., Bachinger, H. P., Brennan, R. G., and Valentin-Hansen, P. (2002) Mol. Cell 9, 23–30
3. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C., and Storz, G. (2002) Mol. Cell 9, 11–22
4. Albrecht, M., and Lengauer, T. (2004) FEBS Lett. 569, 18–26
5. Donahue, W. F., and Jarrell, K. A. (2002) Mol. Cell 9, 7–8
6. Tollervey, D., and Luhrmann, R. (2001) J. Mol. Biol. 312, 1277–1294
7. Plessel, G., Luhrmann, R., and Kastner, B. (1997) Mol. Biol. 308, 49–58
8. van, D. E., Cougot, N., Meyer, S., Babiak, S., Wahle, E., and Seraphin, B. (2002) J. Biol. 20, 1239–1247
9. Yuan, H., Meheus, L., de, B., V, Fischer, U., and Luhrmann, R. (2001) FEBS Lett. 2256–2269