A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*

Tilmann Achsel[1], Hero Brahms[1], Berthold Kastner[1], Angela Bachi[2], Matthias Wilm[2] and Reinhard Lührmann[1,3,4]

[1]Institut für Molekularbiologie und Tumorforschung, Universität Marburg, Emil Mannkopff-Strasse 2, 35037 Marburg, [2]EMBL, Protein and Peptide Group, Meyerhofstrasse 1, 69117 Heidelberg and [3]Max-Planck-Institute of Biophysical Chemistry, Am Faβberg 11, 37070 Göttingen, Germany

[4]Corresponding author
e-mail: luehrmann@imt.uni-marburg.de

We describe the isolation and molecular characterization of seven distinct proteins present in human [U4/U6·U5] tri-snRNPs. These proteins exhibit clear homology to the Sm proteins and are thus denoted LSm (like Sm) proteins. Purified LSm proteins form a heteromer that is stable even in the absence of RNA and exhibits a doughnut shape under the electron microscope, with striking similarity to the Sm core RNP structure. The purified LSm heteromer binds specifically to U6 snRNA, requiring the 3′-terminal U-tract for complex formation. The 3′-end of U6 snRNA was also co-precipitated with LSm proteins after digestion of isolated tri-snRNPs with RNaseT[1]. Importantly, the LSm proteins did not bind to the U-rich Sm sites of intact U1, U2, U4 or U5 snRNAs, indicating that they can only interact with a 3′-terminal U-tract. Finally, we show that the LSm proteins facilitate the formation of U4/U6 RNA duplices *in vitro*, suggesting that the LSm proteins may play a role in U4/U6 snRNP formation.
Keywords: human splicing factors/Sm proteins/snRNP/ U6 snRNA

Introduction

Pre-mRNA splicing is catalysed by the spliceosome, which is formed by the ordered interaction of the U1 and U2 snRNPs, the [U4/U6·U5] tri-snRNP particle and an undefined number of non-snRNP splicing factors with the pre-mRNA (Krämer, 1996; Will and Lührmann, 1997; Burge *et al*., 1999). The tri-snRNP particle is especially important for the splicing reaction, as it constitutes a major part of the spliceosome, consisting of three snRNAs and, in the mammalian system, at least 25 proteins, many of which have been demonstrated to be essential splicing factors (Will and Lührmann, 1997). This particle has to be assembled from U4, U5 and U6 snRNPs prior to association with the spliceosome. First, the U4 and U6 snRNPs interact with each other through extensive base-pairing of their respective snRNAs to form the U4/U6 snRNP. Under splicing conditions, the U4/U6 particle then interacts with U5 snRNP and additional proteins to form

the tri-snRNP. Immediately after integration of the tri-snRNP into the spliceosome, the U4/U6 duplex dissociates and U4 snRNP leaves the spliceosome. After completion of the splicing reactions, the spliceosome is dismantled and it is generally believed that the tri-snRNP is re-formed and takes part in new rounds of splicing (Moore *et al*., 1993; Staley and Guthrie, 1998).

An important feature of the U1, U2, U4 and U5 snRNAs is their Sm site. This consists of a single-stranded uridine-rich region that is usually flanked by two stem-loop structures. The seven Sm proteins B/B′, D1, D2, D3, E, F and G bind to the Sm site, thus forming the Sm core particle (Branlant *et al*., 1982; Liautard *et al*., 1982). In the nervous system, SmN can replace SmB (McAllister *et al*., 1989). The Sm core is essential for the biogenesis and function of the snRNPs. Not only is it required for cap hypermethylation (Mattaj, 1986; Plessel *et al*., 1994), but it also forms part of the snRNP nuclear localization signal (Fischer and Lührmann, 1990; Hamm and Mattaj, 1990) and influences the integration of at least some snRNP-specific proteins into the snRNPs (Nelissen *et al*., 1994). In electron micrographs, all Sm core RNPs possess a similar doughnut-shaped structure (Kastner *et al*., 1990). The Sm proteins interact with each other strongly and specifically (Lehmeier *et al*., 1994; Hermann *et al*., 1995; Raker *et al*., 1996; Fury *et al*., 1997; Camasses *et al*., 1998), but none of the individual Sm proteins binds stably to RNA. Instead, the RNA-binding site is generated only after the formation of Sm protein heteromers: E·F·G and D1·D2 protein complexes are minimally required to form a stable intermediate RNP, the so-called sub-core RNP. Sm core assembly is completed by the subsequent interaction of a B/B′·D3 heteromer (Raker *et al*., 1996).

In contrast to the other spliceosomal snRNAs, U6 has no Sm site and consequently does not associate with the Sm proteins. Moreover, its biogenesis pathway differs in many respects, as it is transcribed by RNA polymerase III (Dahlberg and Lund, 1991, and references therein) and capped with γ-monomethyl triphosphate (Singh and Reddy, 1989). The 3′-oligo(U) end of pre-U6 RNA is elongated during maturation and is subsequently trimmed, leaving, in most organisms, five uridines and a terminal 2′3′-cyclophosphate (Lund and Dahlberg, 1992; Terns *et al*., 1992). Both enzymes involved in this processing exhibit a specificity for U6 snRNA (Booth and Pugh, 1997; Trippe *et al*., 1998). Finally, U6 does not leave the nucleus during snRNP biogenesis (Vankan *et al*., 1990; Terns *et al*., 1993; Boelens *et al*., 1995).

All Sm proteins share a conserved Sm sequence motif consisting of two segments, Sm1 and Sm2, interrupted by a spacer region of variable length (Cooper *et al*., 1995; Hermann *et al*., 1995; Séraphin, 1995). In addition, SmB has a proline-rich C-terminal domain, and SmD1 and SmD3 of most organisms, but not budding yeast, have a

C-terminus rich in arginine and glycine. Recently, the two Sm protein complexes B·D3 and D1·D2 have been crystallized and their three-dimensional structure solved (Kambach *et al.*, 1999). Significantly, the Sm motifs of all four proteins show identical folds. Furthermore, the arrangement of the proteins in the two dimers is virtually identical. Based on these findings, a model has been proposed, in which the Sm proteins oligomerize to form a seven-member ring (Kambach *et al.*, 1999). This model is in good agreement with the observed morphology and size of the Sm core under the electron microscope (Kastner *et al.*, 1990).

Based on sequence homology, Sm-like proteins were identified in a variety of organisms (Cooper *et al.*, 1995; Hermann *et al.*, 1995; Séraphin, 1995). In addition to the seven Sm protein genes, the genome of budding yeast contains open reading frames (ORFs) encoding nine Sm-like proteins (Fromont-Racine *et al.*, 1997), termed 'like Sm', i.e. Lsm1–Lsm9. Seven of these proteins have been shown to associate with U6 as well as with U4/U6 snRNPs and tri-snRNP particles (Cooper *et al.*, 1995; Séraphin, 1995; Pannone *et al.*, 1998; Mayes *et al.*, 1999; Salgado-Garrido *et al.*, 1999). Furthermore, these Lsm proteins are also present in purified yeast tri-snRNPs (Gottschalk *et al.*, 1999; Stevens and Abelson, 1999). At present, however, it is not clear whether the U6-specific Sm proteins recognize the U6 snRNA directly. Because the protein composition of the tri-snRNP particle is well conserved between yeast and man, it is likely that similar proteins are also present in the human tri-snRNP. In search of human Sm-like proteins, we fractionated proteins present in purified [U4/U6·U5] tri-snRNPs using glycerol-gradient centrifugation at high salt concentrations. This method, in combination with an ion-exchange chromatography step, allowed us to isolate an RNA-free protein complex that contained seven Sm-like proteins. These proteins have been termed LSm2–LSm8, adopting the nomenclature used for the yeast Sm-like proteins. We show that the LSm proteins form a doughnut-shaped structure similar to the Sm core, however, unlike the Sm core, they do so in the absence of RNA. We have identified the binding site of LSm proteins on U6 snRNA by demonstrating that the LSm heteromer binds directly to U6 snRNA at its 3′-terminal oligo(U) tract. Finally, we show that the LSm proteins may play an important role in the annealing of the U4 and U6 snRNAs.

Results

Isolation of an RNA-free complex of low molecular mass proteins from purified human [U4/U6·U5] tri-snRNPs

Recently, we used high concentrations of monovalent salt to dissociate specific proteins from U5 snRNPs (Laggerbauer *et al.*, 1998). In search of human Sm-like proteins, we have adopted this method. Thus, 25S [U4/U6·U5] tri-snRNPs, which were purified using immunoaffinity chromatography and glycerol-gradient centrifugation (Laggerbauer *et al.*, 1998), were subjected to a second glycerol-gradient centrifugation in the presence of 0.7 M NaCl. The fractionation of proteins and RNAs on such a glycerol gradient is shown in Figure 1A. As described previously, most of the U5-specific proteins

remain associated with the U5 snRNA under these conditions and thus form a rapidly sedimenting complex with a Svedberg constant of ~20S (Laggerbauer *et al.*, 1998; fractions 22–27). Most of the U4 and U6 snRNAs sediment as a 10S U4/U6 core snRNP complex containing one set of the Sm proteins B/B′, D3, D2, D1, E, F and G (fractions 10–14), whereas other tri-snRNP proteins dissociate from the RNPs and are mostly found in the top third of the gradient (fractions 1–9). Examples of proteins sedimenting as monomers are the 15.5kD or 27kD proteins, which sediment at the top of the gradient in fractions 1–4. The previously described RNA-free heteromer of the U4/U6-specific proteins SnuCyp20, 60kD and 90kD (Teigelkamp *et al.*, 1998) sediments with a higher S value and peaks in fractions 6–8. Interestingly, several proteins with molecular masses ranging from 8 to 16 kDa (Figure 1A, labelled by asterisks) co-sediment with a slightly lower S value than the SnuCyp20/60kD/90kD heteromer and peak in fractions 5–7. Close inspection of the SDS–polyacrylamide gels shows that these proteins do not exactly co-migrate with the Sm proteins D1, D2, D3, F and G, suggesting that they may represent hitherto uncharacterized human tri-snRNP proteins. Their migration behaviour on the glycerol gradient and the absence of snRNA in these fractions suggest that these low molecular mass proteins are part of a salt-stable, RNA-free protein complex. Moreover, the proteins peak in exactly the same fractions and do not appear to co-sediment with any of the other proteins present in these fractions, suggesting that the low molecular mass proteins may form a complex with each other.

To purify these low molecular mass proteins further, fractions 4–8 of the glycerol gradient were subjected to ion-exchange chromatography on a MonoS column, and the proteins were fractionated by a 0.1–0.4 M NaCl gradient. Figure 1B (top), shows the UV-absorption profile, and the Figure 1B (bottom) shows the protein composition of each fraction as analysed by SDS–PAGE. The low molecular mass proteins eluted as a single, homogeneous peak at 250 mM NaCl (fractions 10–12). At least six protein bands could be distinguished by SDS–PAGE. All other proteins loaded onto the MonoS column eluted at higher salt concentrations (fractions 18–25), demonstrating that they do not interact stably with the low molecular mass proteins. Moreover, the fractions containing the low molecular mass proteins were devoid of RNA, as shown either by silver staining or by labelling with [^{32}P]pCp and [γ^{32}P]ATP (not shown). The absence of RNA from fractions 10–12 is also strongly indicated by the high ratio of the UV absorption profiles measured at 280 and 260 nm, respectively (Figure 1B, top).

Molecular characterization of seven distinct Sm-like proteins present in [U4/U6·U5] tri-snRNPs

In order to characterize them in more detail, the proteins present in the MonoS-purified complex were separated by preparative SDS–PAGE and subjected to partial peptide sequencing. In preparative gels, the protein bands 1a and 1b, as well as 2a and 2b (Figure 1B), were not resolved and therefore were excised together as one band. Edman sequencing yielded a total of 15 peptide sequences from the four bands (Table I). These peptides were found in the ORFs of several expressed sequence tags (ESTs) in
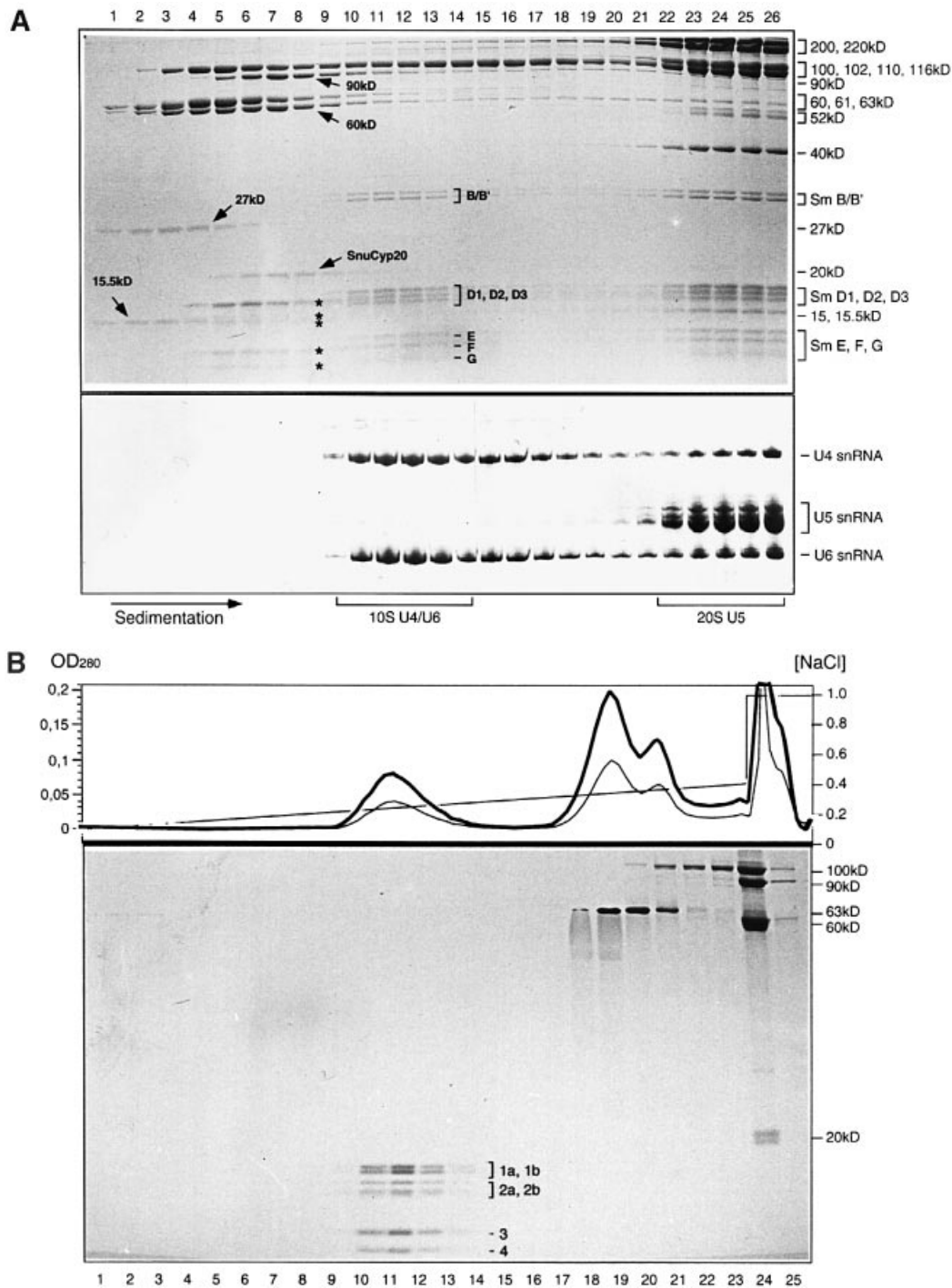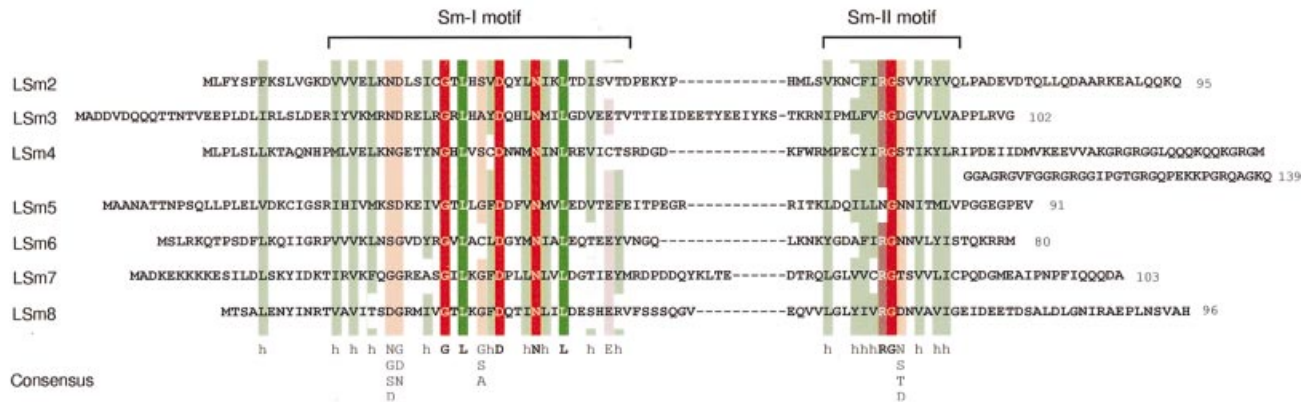
Fig. 1. A set of low molecular mass proteins present in human tri-snRNP particles. (**A**) Fractionation of tri-snRNPs by glycerol-gradient centrifugation at high salt. Approximately, 0.5 mg of purified tri-snRNP particles were concentrated by ultracentrifugation, resuspended in buffer containing 0.7 M NaCl and then separated on a 5–20% glycerol gradient prepared with the same buffer. The gradient was harvested manually from the top, and the protein composition of each fraction was analysed by SDS–PAGE and staining with Coomassie Blue (upper panel). The positions of the tri-snRNP proteins are indicated on the right and the proteins mentioned in the text are labelled by arrows. Previously unidentified low molecular mass proteins are marked by asterisks. The RNA was analysed by 8 M urea–PAGE and staining with ethidium bromide (lower panel, the negative image is shown). The positions of the snRNAs are indicated on the right. (**B**) Purification of the novel proteins by cation-exchange chromatography. Glycerol-gradient-purified proteins were fractionated by MonoS as described in Materials and methods. At the top, the UV absorption at 280 nm (thick profile) and at 260 nm (thin profile), as well as the shape of the salt gradient is shown. The protein composition of each fraction was analysed by SDS–PAGE and staining with Coomassie Blue (lower panel). The positions of the known tri-snRNP proteins are indicated on the right, whereas the additional proteins are numbered (Table I).
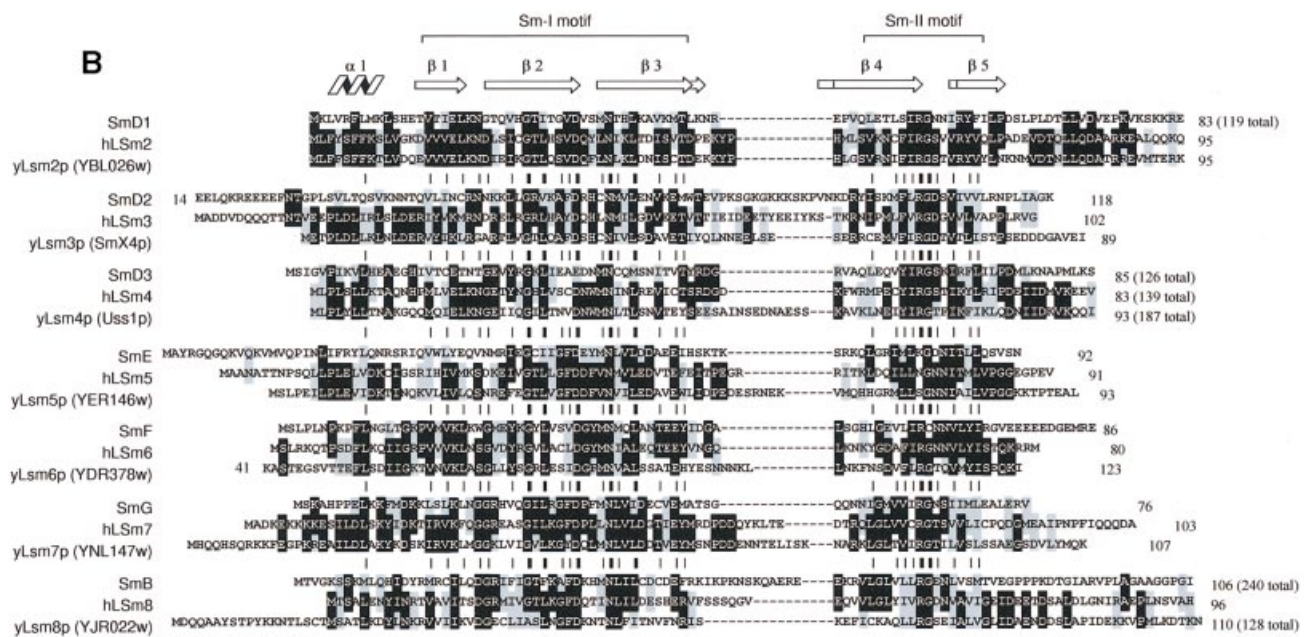
the database that encode four distinct proteins (see below). To ensure the most comprehensive peptide sequence analysis, the proteins recovered from the four bands were analysed independently using tandem mass spectroscopy, which is a much more sensitive protein-sequencing method (Wilm and Mann, 1996; Wilm *et al*., 1996). Utilizing this method, we obtained several peptides which confirmed the identity of the proteins identified by Edman sequencing.
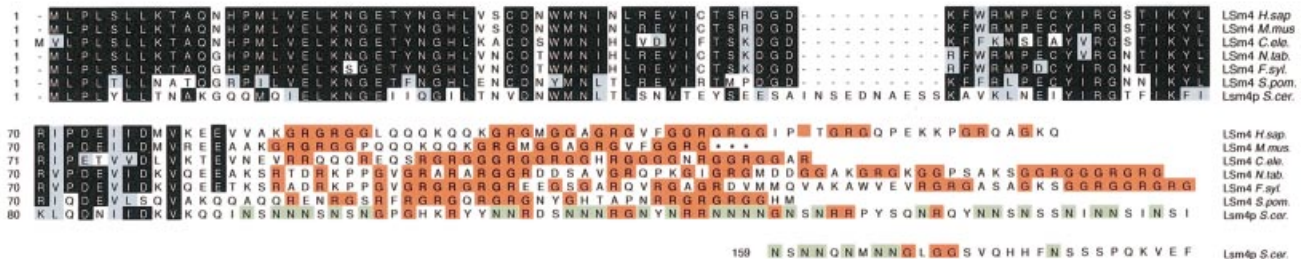
A

Sm-I motif   Sm-II motif

	Sm-I motif	Sm-II motif	
LSm2	MLFYSFFKSLVGKDVVVELKNDLSICC H HSV QYLN IK TDISVTDPEKYP------------	HMLSVKNCFIR G SVVRYVQLPADEVDTQLLQDAARKEALQQKQ	95
LSm3	MADDVDQQQTTNTVEEPLDLIRLSLDERIYVKMRNDRELR GR HAYD QHLN MI GDVEETVTTIEIDEETYEEIYKS-TKRNIPMLFV RG DGVVLVAPPLRVG	102
LSm4	MLPLSLLKTAQNHPMLVELKNGETYNGH VSC NWMN INU REVICTSRDGD------------ KFWRMPECYIR GS TIKYLRIPDEIIDMVKEEVVAKGRGRGGLQQQKQQKGRGM	
	GGAGRGVFGGRGRGGIPGTGRGQPEKKPGRQAGKQ	139
LSm5	MAANATTNPSQLLPLELVDKCIGSRIHIVMKSDKEIV TL LGF DDFVN MVL EDVTEFEITPEGR------------ RITKLDQILLN GNN ITMLVPGGEGPEV	91
LSm6	MSLRKQTPSDFLKQIIGRPVVVKLNSGVDYR GL ACLD GYMN IAL EQTEEYVNGQ------------ LKNKYGDAFIR GN NVLYISTQKRRM	80
LSm7	MADKEKKKKESILDLSKYIDKTIRVKFQGGREAS GIL KGF DPLLN LVD GTIEYMRDPDDQYKLTE------------ DTRQLGLVVC RGT SVVLICPQDGMEAIPNPFIQQQDA	103
LSm8	MTSALENYINRTVAVITSDGRMIV TL KGF DQTIN LIL DESHERVFSSSGV------------ EQVVLGLYIV RG DNVAVIGEIDEETDSALDLGNIRAEPLNSVAH	96
Consensus	h h h h NG h G L GhD hNh L h Eh	h hhhRGN h hh	
	GD S	S	
	SN A	T	
	D	D	

B

Sm-I motif   Sm-II motif

α1   β1   β2   β3   β4   β5

SmD1		83 (119 total)
hLSm2		95
yLsm2p (YBL026w)		95
SmD2	14	118
hLSm3		102
yLsm3p (SmX4p)		89
SmD3		85 (126 total)
hLSm4		83 (139 total)
yLsm4p (Uss1p)		93 (187 total)
SmE		92
hLSm5		91
yLsm5p (YER146w)		93
SmF		86
hLSm6		80
yLsm6p (YDR378w)	41	123
SmG		76
hLSm7		103
yLsm7p (YNL147w)		107
SmB		106 (240 total)
hLSm8		96
yLsm8p (YJR022w)		110 (128 total)

C

CaSm		89 (133 total)
Lsm1p (YJL124c)	34	127 (172 total)
LSmα A.fulgidus		77
LSmα M.thermauto.		81
LSmα P.horikoshii		75
LSmβ M.thermauto.		67
LSmγ A.fulgidus		75

D

LSm4 H.sap	
LSm4 M.mus	
LSm4 C.ele.	
LSm4 N.tab.	
LSm4 F.syl.	
LSm4 S.pom.	
Lsm4p S.cer.	

Lsm4p S.cer. (159)

Moreover, nine peptides were obtained which belong to ESTs encoding three additional proteins (Table I). In summary, seven distinct proteins were identified in the MonoS-purified sample. Note that no peptides were identified in the protein mixture that corresponded to the Sm proteins, confirming the purity of the MonoS-derived sample.

To characterize the entire ORF of each protein, the ESTs in the database containing the longest ORFs (Table I) were sequenced completely. Five of these ESTs encoded the respective full-length protein (see below). In contrast, ESTs AA534490 and N42439 did not contain a proper initiator methionine codon. Therefore, the 5′-ends of the corresponding mRNAs were amplified by PCR using a HeLa cDNA library (see Materials and methods). To exclude errors due to mutations introduced during the PCR, the 5′-ends of five independent clones were sequenced. From these sequences and the corresponding ESTs, a contiguous sequence was assembled that encoded the respective full-length protein. The amino acid sequences of the seven proteins, as deduced from the ORFs, are shown in Figure 2A. The identity of the cDNAs was confirmed by the findings that: (i) all peptides were found in the deduced amino acid sequences (Table I; Figure 2A), and (ii) proteins synthesized by translation *in vitro* co-migrated on SDS-gels with the band from which the respective peptides were obtained (data not shown).

A database search revealed that all seven cDNAs encode novel proteins. Sequence comparison showed that all seven proteins share significant homology with the known Sm core proteins but are nonetheless distinct proteins. Moreover, they are highly conserved evolutionarily (see below), and we therefore named them LSm proteins adopting the nomenclature used in yeast. Like the Sm core proteins, all Sm-like proteins contain the Sm sequence motif, which consists of two regions separated by a linker of variable length; at the majority of conserved positions, only a certain physico-chemical property is conserved (mostly hydrophobicity), but there are also seven residues that are highly conserved (Cooper *et al.*, 1995; Hermann *et al.*, 1995; Séraphin, 1995).

To determine whether the seven LSm proteins identified in the human tri-snRNP are evolutionarily conserved, we carried out an extensive database search for putative homologues. Of particular interest was a comparison of the human proteins with those in the yeast database, as sequencing of the yeast genome revealed nine genes, *LSM1* to *LSM9*, that encode Sm-related proteins. Sequence alignments demonstrated that Lsm2p, Lsm3p (SmX4p), Lsm4p (Uss1p), Lsm5p, Lsm6p, and Lsm7p are orthologues of the human proteins LSm2 to LSm7, with sequence identities ranging from 41 to 62%. We used this relationship as a criterion to name the human proteins (Figure 2B; the score for Lsm4p excludes the C-terminal domain, see below). Of the additional three Lsm proteins found in yeast, Lsm1p exhibits clear homology with the human 'cancer-associated Sm' protein CaSm, (Schweinfest *et al.*, 1997), sharing 33% identity (Figure 2C). Lsm9p/SmX1p appears to be present only in budding yeast, as no clearly homologous protein could be identified in the EST databases of other organisms (Séraphin, 1995; our database searches). The third protein, Lsm8p, cannot be aligned as well with the Sm and Sm-like proteins, because its Sm domain deviates slightly from the consensus. Among the human LSm proteins, it aligns best with LSm8 (26% identity) and these proteins are therefore probably functional counterparts. Consistent with this idea, Lsm8p and LSm8 have been identified in purified [U4/U6·U5] tri-snRNP particles of yeast and human, respectively (this study; Gottschalk *et al.*, 1999; Stevens and Abelson, 1999). In conclusion, each of the human Sm-like proteins has an orthologue in yeast.

In addition, the human LSm proteins are highly conserved throughout all eukaryotic kingdoms, as homologues in the insect, nematode and plant databases have been identified which generally share between 50 and 75% identity with their human counterparts. Moreover, the observed conservation is not restricted to the Sm sequence motif, but as a rule also includes the N- and C-terminal extensions. The strong phylogenetic conservation of the Sm motif and surrounding sequences is illustrated by the alignment of the LSm4 proteins from various organisms (Figure 2D). Interestingly, all LSm4 proteins contain an additional C-terminal domain that varies in length (52–108 amino acids). Although the sequences are not highly conserved, the biochemical character of the C-terminal domains is very similar, as they are enriched in glycine and arginine. Only Lsm4p from budding yeast has a divergent C-terminus that is particularly rich in asparagine. Interestingly, similar C-termini containing RG dipeptides

Fig. 2. The low molecular mass proteins are novel, Sm-like proteins. (**A**) All seven protein sequences contain a highly conserved Sm motif. The amino acid sequences of the proteins were aligned with respect to the Sm motif. Conserved amino acids are marked according to their properties: bulky hydrophobic residues (L, I, M, V, F, Y, W) are shown in green, small polar residues (G, S, D, N) in red and bulky polar residues (E, K, R) in purple. Positions that are identical in most Sm motifs are indicated by dark colours, whereas less conserved positions are marked with light colours. At the bottom, the Sm consensus based on 80 Sm and Sm-related proteins is shown (note that this consensus is somewhat more stringent than reported previously; Hermann *et al.*, 1995). 'h' signifies a bulky hydrophobic residue as defined above. The cDNA sequences were deposited in the DDBJ/EMBL/GenBank (accession numbers are given in Table I). (**B**) Comparison of the human LSm proteins with the human Sm and yeast Lsm proteins. Each of the LSm proteins was aligned to the human Sm and *Saccharomyces cerevisiae* Lsm protein that exhibits the highest sequence similarity. Residues that are identical to the respective LSm protein are boxed in black, conserved residues are shaded grey. Conserved residues are grouped as follows: (L, I, M, V), (F, Y, W), (K, R, H), (D, E), (N, Q), (S, T) and (G, A). The fold of the Sm domain (Kambach *et al.*, 1999) is drawn schematically at the top; residues conserved in the Sm motifs are connected by vertical lines. Numbers at the end indicate the position of the last shown amino acid; in those cases where only a portion of the sequence is shown, the total length of the protein is additionally indicated. (**C**) Additional Sm-related proteins identified in the database. The sequences of the human CaSm (AF000177) and the yeast Lsm1p protein, as well as five Sm-related proteins from the archaea *Archaeoglobus fulgidus* (Klenk *et al.*, 1997), *Methanobacterium thermautotrophicum* (Smith *et al.*, 1997) and *Pyrococcus horikoshii* (Kawarabayasi *et al.*, 1998) are aligned. Identical and conserved residues are highlighted as above. (**D**) Evolutionary conservation of the LSm4 protein. LSm4 is aligned with its orthologues from mouse (assembled from twelve largely overlapping ESTs; note that this contiguous sequence does not contain the 3′-end of the ORF), *Caenorhabditis elegans* (U20864), *Nicotiana tabacum* (S54169), *Fagus sylvatica (*the European beech; AJ130887), *Schizosaccharomyces pombe* (Z97992) and *S.cerevisiae* (Uss1p; YER112w). Residues identical in at least five sequences are boxed in black, conserved residues (grouped as in B) are shaded grey. In the C-terminal domain, arginine and glycine residues are shaded in red. In addition, asparagines present in the C-terminus of the *S.cerevisiae* sequence are shaded in green.

Table I. Peptide sequences derived from human LSm proteins

Protein	Band	Peptides	EST accession No.	Protein accession No.
LSm2	3	[2]LFYsFFK[8] [14]DVVVELK[20] (MS) [22]DlsI-GT—VDQYLN[37] [40]LTDISVTDPE[49] [40]LTDISVTDPEKYPH[53] (MS) [41]TDISVTDPEKY[51] [51]YPHMLSVK[58] [59]NCFIR[63] (MS) [69]YVQLPADEVDTQLLQDAAR[87] (MS)	AA636093	AF182288
LSm3	1	[2](Ac)ADDVDQQQTTNTVEEPLDLIR[22] (MS) [23]LSLDER[28] (MS) [81]NIPMLFVR[88] (MS) [89]GDGVVLVAPPLR[100] (MS)	N42583	AF182289
LSm4	1	[9]tAQNHPMLVe[18] [11]QNHPMLVELK[20] [49]DGDKFW[54] (MS) [53]FWRMpE-Yirgst [65] [56]MPECYIR[62] (MS) [71]IPDEIIDMVK[80] (MS) [101]gXGMGGAGXGVFGGXGXGGIPG [122]	AA112507	AF182290
LSm5	3	[2](Ac)AANATTNPSQLLPLELVDK[20] (MS)	W80883	AF182291
LSm6	4	[5]KQTPSDFLK[13] (MS) [6]qTPSDFLK[13] [14]qIIGRPVVVK[23] (also MS) [24]LNSGVDYR[31] (MS) [58]NKYGDAFIR[66] (MS) [60]YGDAFIRGNNVLYIS[74] [67]GNNVLYISTQKR[78] (MS)	AA054771	AF182292
LSm7	2	[10]ESILDLSK[17](also MS) [27]FQGGREASGILK[38] [70]QLgLVVcR[77](also MS)	AA534490	AF182293
LSm8	2	[2](Ac)TSALENYINR[11] (MS) [12]TVAVITSDGR[21](MS) [29]GFDQTINLILDESHER[44](MS) [88]AEPLNSVAH[96](MS)	N42739	AF182294

The seven LSm proteins are listed along with the SDS–PAGE band in which they were identified (Figure 1B), all peptide sequences, the accession No. of the most complete human EST and the accession No. under which the entire cDNA sequence was deposited in the DDBJ/EMBL/GenBank. The numbering of the peptide sequences indicates their position in the complete amino-acid sequence (Figure 2A). Amino acids that could not be identified unambiguously are set in small type and positions that could not be determined are marked '–'. X designates a modified amino acid. Peptide sequences obtained by mass spectroscopy are labelled (MS); all other sequences were obtained by Edman degradation.

are also found in the canonical SmD1 and SmD3 proteins of all organisms, but not in those of budding yeast. From these sequence comparisons, we conclude that the tri-snRNP particles of all the species mentioned above in all likelihood contain a set of seven LSm proteins.

During the database searches, five additional ORFs encoding Sm-related proteins were found in the genomes of three archaea (Figure 2C). Three of these proteins are homologous to each other and are thus probable orthologues (Lsmα, Figure 2C). Thus, the Lsmα protein may well represent the progenitor of the Sm protein family. The other two, LSmβ and LSmγ, diverge somewhat from the Sm consensus and it remains to be seen whether they are bona fide Sm-like proteins.

Finally, we compared the sequences of the human LSm and Sm proteins in order to determine whether a given LSm protein is structurally more closely related to a particular Sm protein. Because of the considerable variation in the length of the proteins, these comparisons were restricted to the Sm domain as defined crystallographically (i.e. the residues equivalent to positions 1–57 and 64–86 in LSm8; Kambach *et al.*, 1999). Significantly, each of the LSm proteins clearly aligned better with one of the

canonical Sm proteins (Figure 2B). The similarities are significant, ranging from 26 to 38% identity, with an exceptionally high identity of 48% between LSm6 and SmF. These values are, however, not high enough in themselves to allow the conclusion that the respective proteins undergo analogous protein–protein interactions.

The RNA-free LSm protein complex exhibits a doughnut-shaped structure under the electron microscope with striking similarity to the canonical Sm core RNP structure

To learn whether the Sm-like proteins were organized in a specific higher order assembly, electron-microscopy was performed to visualize the MonoS-purified LSm complex. As in previous studies with the Sm core RNP (Kastner *et al.*, 1990), the proteins were negatively stained with uranyl formate using the double-carbon-film method. Figure 3A shows a typical overview of the LSm protein preparation. Interestingly, most of the LSm proteins appear as a round projection with a diameter of ~8 nm and 80% of the projections exhibit an accumulation of stain in the middle. This central stain can best be interpreted as a hole in the middle of the structure. Therefore, the structure of
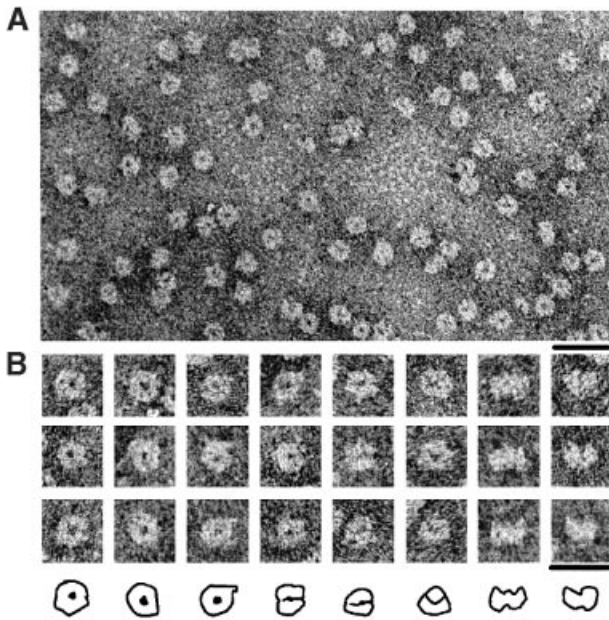
Fig. 3. Electron microscopy reveals the ring-shaped structure the LSm complex. MonoS-purified LSm proteins were absorbed on carbon foils, stained with uranyl formate and examined by electron microscopy. (**A**) General view of the preparation. The bar represents 25 nm. (**B**) Selected specimens representing typical projections. The characteristic shape of the projection is drawn schematically at the bottom of each row. Bar, 20 nm.

Fig. 4. [U4/U6·U5] tri-snRNPs and U4/U6 snRNPs co-precipitate with the LSm4 protein. Immunoprecipitations from HeLa nuclear extract using α-LSm4 serum (lanes 4–6), the corresponding pre-immune serum (lanes 1–3) or α-Sm antibodies (lane 7) were performed at the salt concentrations indicated above each lane. Co-precipitating RNA was recovered, labelled with [^{32}P]pCp, fractionated by urea–PAGE and visualized by autoradiography. The position of the snRNAs is indicated on the right. Note that U6 snRNA is very poorly labelled by pCp and therefore does not appear in the autoradiograph.

the LSm proteins resembles a doughnut and bears a striking similarity to the structure of the U1, U2, U4 and U5 Sm core RNPs. Moreover, the size of the doughnut can only be explained by assuming that an LSm multimer is formed and the electron micrographs thus strongly support the biochemical evidence that the LSm proteins form a heteromer. The similarity between the structure of LSm proteins and Sm core RNPs is further supported by several structural details evident in the electron micrograph. Typical images showing these details are enlarged in Figure 3B. In the centre of many projections (about one in three), a 2-nm-wide accumulation of stain can be seen (rows 1–3). Somewhat less frequently, projections are observed that seem to be bisected by a line of stain (rows 4 and 5); most rare are projections with a wedge-shaped inner structure (row 6). All of these details were also found in Sm core RNPs (Kastner *et al.*, 1990). In addition, ~20% of the images have an elongated form ~9 nm long and 3–5 nm thick (rows 7 and 8). These images may represent side views of the ring, which would provide a measure for the thickness of the ring structure. In conclusion, the LSm images display great similarity, in both size and shape, with the structure of the Sm core RNPs. Only the central cavity, as evidenced by the accumulation of stain, is ~0.5 nm narrower in the Sm core RNP than in the LSm complex. Thus, the Sm core RNP appears to have extra mass at its centre.

In the [U4/U6·U5] tri-snRNP, the LSm proteins are associated primarily with the U4/U6 snRNP

As the LSm proteins were isolated from purified tri-snRNP complexes, it was not possible to conclude whether they associate primarily with U4/U6 snRNPs, as described fo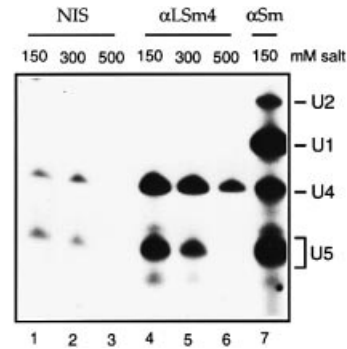r the yeast Lsm proteins, or whether they require the presence of U5 in order to associate with the tri-snRNP. As a prerequisite for immunoprecipitation studies, we raised antibodies against a peptide derived from the C-terminus of LSm4 (see Materials and methods). On Western blots of purified tri-snRNP proteins, these antibodies recognized only the LSm4 protein (not shown). Furthermore, the α-LSm4 antibodies precipitated only the U4, U5 and U6 snRNAs, and not U1 or U2 snRNA from HeLa nuclear extracts, confirming their specificity for LSm as opposed to Sm proteins (Figure 4, lane 4. Note that U6 snRNA is poorly labelled with pCp. Its presence was confirmed by Northern blotting, not shown). As purified 20S U5 snRNPs were not precipitated by α-LSm4 antibodies (not shown), we conclude that only [U4/U6·U5] tri-snRNP particles are precipitated from nuclear extract at 150 mM NaCl (note that HeLa cell nuclear extract contains minimal amounts of free U4/U6 snRNPs). To determine whether the LSm proteins associate with U4/U6 snRNPs, we raised the salt concentration during the immunoprecipitation washes in order to disrupt the tri-snRNP particle. Significantly, only U4 and U6 snRNAs are precipitated at 500 mM NaCl, albeit at a reduced level, whereas the U5 snRNA is washed away completely (Figure 4, lane 6; data not shown). This finding is consistent with the idea that the LSm4 protein, and most likely the other LSm proteins as well, associates primarily with U4/U6 in human tri-snRNP particles. Similarly, the association of LSm3 and LSm4 proteins with human U4/U6 and [U4/U6·U5] tri-snRNPs was reported by others while this work was in progress (Salgado-Garrido *et al.*, 1999).

The LSm protein heteromer binds directly to U6 snRNA and requires the 3'-terminal U-tract for complex formation

Whereas the LSm4 protein interacts primarily with the U4/U6 snRNP, it is not clear whether it and the other LSm proteins bind directly to RNA or associate indirectly via other tri-snRNP proteins. As we had a highly purified LSm protein complex available, it was possible to analyse its RNA-binding properties. Thus, we incubated the LSm proteins with ^{32}P-labelled U1, U2, U4, U5 and U6 snRNAs
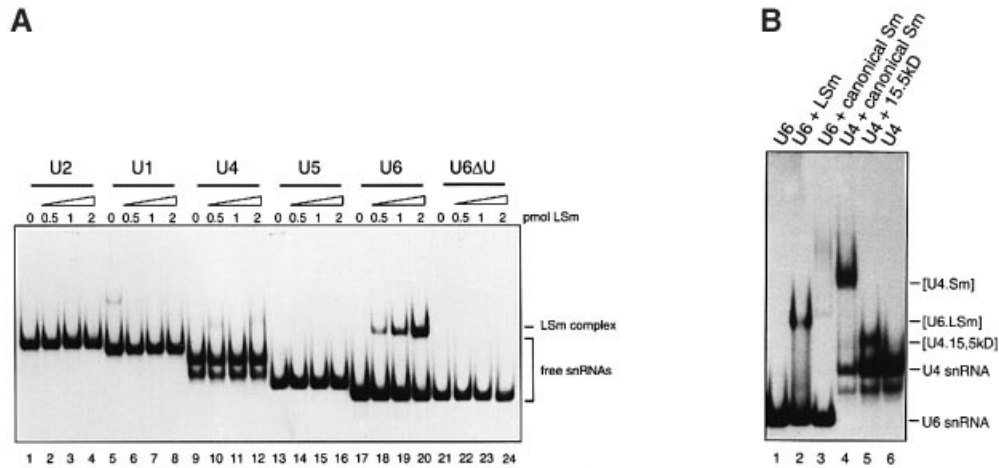
Fig. 5. The LSm proteins bind directly and specifically to U6 snRNA. (**A**) The snRNAs U2 (lanes 1–4), U1 (lanes 5–8), U4 (lanes 9–12), U5 (lanes 13–16), U6 (lanes 17–20) or U6ΔU (lacking the five 3′-terminal Us, lanes 21–24) were prepared by *in vitro* transcription, incubated with increasing amounts (as indicated above each lane) of MonoS-purified LSm proteins for 15 min at 30°C, chilled on ice, fractionated by native 6% PAGE and visualized by autoradiography. The position of the free snRNAs and the LSm complexes is indicated on the right. (**B**) The snRNAs U6 (lanes 1–3) and U4 (lanes 4–6) were incubated with buffer (lanes 1 and 6), with 2 pmoles of LSm proteins (lane 2), ~2 pmoles of canonical Sm proteins derived from purified U1 and U2 snRNPs (lanes 3 and 4) or with 2 pmoles of the U4-15.5kDa protein (lane 5) as in (A).

synthesized by transcription *in vitro* and fractionated the mixture by PAGE under non-denaturing conditions. Significantly, only U6 snRNA formed a complex with the LSm proteins, as indicated by a U6 band migrating significantly more slowly than free U6 snRNA (Figure 5A, lanes 17–20). This interaction was highly specific, as U1, U2 and U5 snRNAs did not form a complex (lanes 1–8 and 13–16). Only a very small amount of U4 snRNA was shifted in a manner independent of the LSm protein concentration (lanes 9–12; the double band observed with free U4 snRNA is most likely to represent different conformers, as the same RNA migrates as a single band after denaturing PAGE; not shown). The specificity of LSm protein binding with U6 snRNA was further supported by the finding that canonical Sm proteins (isolated from a mixture of U1 and U2 snRNPs, see Materials and methods) did not result in a mobility shift of U6 under the same conditions (Figure 5B, lane 3). In contrast, the Sm proteins interacted with the U1, U2, U4 and U5 snRNAs, as demonstrated here by the bandshift of the U4 snRNA (Figure 5B, lane 4). It is not clear from these results whether all or a subset of the LSm proteins bind to the U6 snRNA. However, the gel mobility shift of U6 snRNA caused by the LSm proteins is very similar in magnitude to the shift observed with U4 snRNA in the presence of the Sm proteins (compare Figure 5B, lanes 4 with lane 5), whereas the 1:1 complex of U4 RNA and the 15.5kD protein (Nottrott *et al.*, 1999) migrates significantly faster than the U4 Sm core RNP. This suggests that the pre-formed LSm protein complex binds to U6 snRNA as such.

Next we investigated the structural requirements of U6 snRNA for LSm protein binding, initially testing various deletion mutants. Whereas the 5′-end up to the AUAUA sequence (i.e. up to position 24) can be deleted with no effect on LSm protein binding (not shown; Figure 6), deletion of the five uridines at the 3′-end abolished complex formation completely (Figure 5A, lanes 21–24), indicating that the oligo(U) terminus is essential for LSm protein binding. To determine whether additional elements of the U6 snRNA are also needed for efficient binding of

the Sm-like complex, we deleted the 5′-terminal domain up to position 24 and replaced the entire central domain (nucleotides 37–93) with an artificial tetraloop sequence (UUCG). Three guanosine residues were added to the new 5′-end to ensure efficient transcription *in vitro*. Note that this RNA still forms the lower part of the central stem of U6 snRNA (Figure 6A). In bandshift assays, the RNA is efficiently recognized by the Sm-like proteins only if the 3′-terminal oligo(U) stretch is present (compare Figure 6B, lanes 4–6 with lanes 1–3). In U6 snRNA, as well as in the deletion mutant, the 3′-terminal uridines can potentially base-pair with a stretch of adenosines (Figure 6A). We thus tested whether the LSm proteins recognize the 3′-terminal uridines as part of a helical stem or as a single strand. To this end, we deleted the stretch of four adenosines, thus preventing base-pairing of the terminal uridines. This mutant was also bound efficiently by the LSm proteins (Figure 6B, lanes 7–10). We conclude that the LSm proteins recognize the 3′-terminal uridines of U6 snRNA, and that they have a preference for single-stranded oligouridylic acid. This idea is corroborated by our finding that the LSm proteins do not bind at all to the U6 deletion mutant lacking the 3′-terminal uridines (lanes 1–3), nor to a similar RNA having an oligo(C) stretch instead of oligo(U) (lanes 11–14).

The 3′-terminus of U6 snRNA is specifically co-precipitated with α-LSm4 antibodies after RNase T₁ digestion of purified tri-snRNP complexes

The results presented above demonstrate that the single-stranded oligo(U) terminus of U6 snRNA is a major determinant of LSm protein binding *in vitro*. To investigate whether the 3′-terminus of U6 snRNA is also associated with the LSm proteins in native tri-snRNPs, we performed experiments involving RNase T₁ digestion of purified tri-snRNPs followed by immunoprecipitation. Co-precipitating RNA fragments were labelled with [γ ³²P]ATP and analysed by denaturing PAGE. Only one RNA fragment specifically and efficiently co-precipitated with LSm4
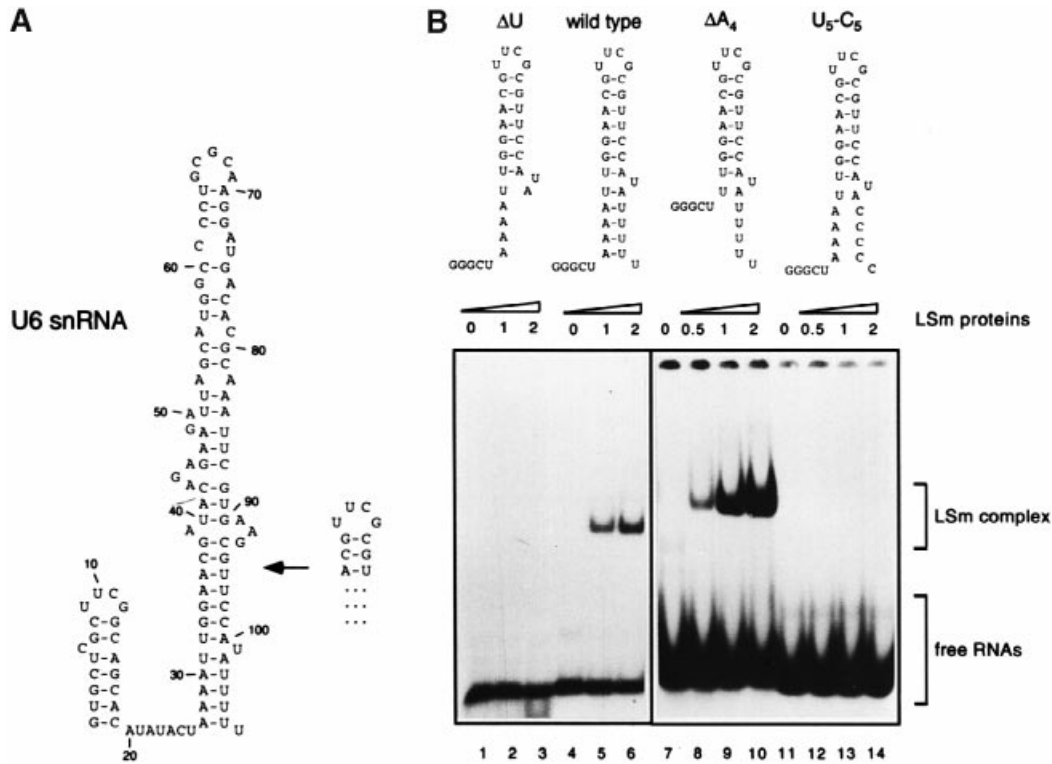
Fig. 6. The 3′-terminal uridines are the major determinant for LSm protein binding. (**A**) Construction of a deletion mutant. The sequence and secondary structure of human U6 snRNA are shown; the position at which the artificial tetraloop sequence (UUCG) was introduced is indicated on the right. (**B**) LSm protein binding of the U6 mutants. RNAs with the sequence indicated on the top were prepared by *in vitro* transcription and assayed for gel mobility shifts as in Figure 5. The position of the free RNAs and the LSm complexes is indicated on the right.

(Figure 7, right-hand panel, lanes 1 and 2). Antibodies directed against another U4/U6-specific protein, namely the 60kD protein (Lauber *et al.*, 1997), did not precipitate the same RNA fragment (lanes 3 and 4), thereby ruling out the possibility that its co-precipitation with LSm4 is due to an association with fragments of U4/U6 snRNPs. Moreover, antibodies directed against the U5-116kD protein (Fabrizio *et al.*, 1997) also failed to precipitate this RNA (lanes 5 and 6). To identify the RNA fragment, we eluted it from a preparative polyacrylamide gel and subjected it to enzymatic sequencing using RNase PhyM, RNase U$_2$, and RNaseT$_1$. As shown in Figure 7 (Figure 7, left), the following partial sequence could be deduced from the sequencing reactions: 5′-CNAUAUUUU-3′. This sequence can be assigned unambiguously to the 3′-terminal T$_1$ fragment of the U6 snRNA (5′-uuCCAUAUUU-Uu-3′). The only other RNA fragment that specifically co-precipitates with LSm4 and migrates as a minor band slightly above the main fragment (Figure 7, right, lane 1) yielded exactly the same sequence, but with one additional uridine at the 3′-end (not shown); it is therefore derived from the minor population of U6 snRNAs containing six terminal uridines. The LSm complex appears to dissociate from this specific RNA fragment at salt concentrations lower than that at which it dissociates from the tri-snRNP (compare Figure 7, lanes 1 and 2 with Figure 4), suggesting that additional contacts might stabilize LSm protein binding in the intact particle. However, no other part of the U4, U5 or U6 snRNAs associates with the LSm proteins stably enough to be co-precipitated.

Binding of the LSm complex to U6 snRNA facilitates annealing with U4 snRNA in vitro

An initial step in the assembly of tri-snRNP is the annealing of the U4 and U6 snRNAs to form the U4/U6 snRNP. In the mammalian system, this step does not take place in the absence of proteins, as the U4/U6 duplex is severely destabilized by an intramolecular structure adopted by naked U6 snRNA (Brow and Vidaver, 1995). Disrupting the base of the central stem of U6 snRNA alleviates this destabilization, and we therefore hypothesized that binding of the LSm proteins to this region might have a similar effect. To test whether the LSm proteins affected U4/U6 duplex formation, we compared the annealing of naked U6 snRNA and the U6·LSm complex with U4 snRNAs prepared by transcription *in vitro*. As shown in Figure 8, incubation of U6 snRNA with an excess of non-radioactive U4 snRNA at 30°C leads to the formation of a slower-migrating U4/U6 complex, albeit slowly and inefficiently (lanes 7–10). In contrast, when the U6 snRNA is associated with LSm proteins, the U4/U6·LSm complex forms more rapidly, and more efficiently than the naked U4/U6 duplex (lanes 3–6). Formation of these complexes involves a specific interaction between U4 and U6, as addition of U1 snRNA as a control RNA does not induce the formation of slower-migrating complexes (lanes 11 and 12). This demonstrates that the Sm-like complex enhances the formation of U4/U6 snRNA duplices *in vitro*. The question whether the LSm proteins bind first to U6 snRNA and facilitate base-pairing with U4 snRNA or whether a pre-formed U4/U6 duplex is
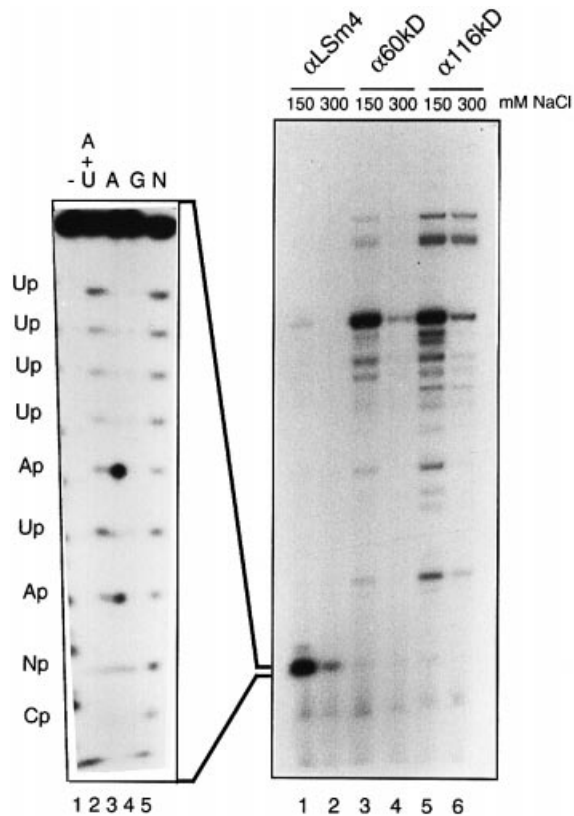
Fig. 7. LSm4 associates with the 3′-terminal fragment of U6 snRNA. (Right) For each lane, 10 µg of purified tri-snRNPs were digested with RNaseT$_1$ and resulting RNA fragments were immunoprecipitated with antibodies directed against LSm4 (lanes 1 and 2), U4/U6-60kD (lanes 3 and 4) or U5-116kD (lanes 5 and 6). Immunoprecipitations were performed at 150 mM (lanes 1, 3 and 5) or 300 mM NaCl (lanes 2, 4 and 6). Co-precipitating RNA fragments were eluted from the beads, labelled at their 5′-ends with ^{32}P, fractionated by urea–PAGE and visualized by autoradiography. (Left) The major band precipitated in lane 1 of the right panel was recovered from a preparative gel and partially digested with RNasePhyM (lane 2), RNaseU$_2$ (lane 3), RNaseT$_1$ (lane 4) or by alkaline treatment (N; lane 5). The interpretation of the sequencing reactions is shown on the left.

stabilized by subsequent binding of the LSm proteins to the 3′-end of U6 snRNA, cannot be answered at present.

Discussion

In this study, we isolated seven new proteins from human [U4/U6·U5] tri-snRNPs which are closely related to the seven Sm core proteins and are therefore termed LSm proteins, i.e. LSm2–LSm8. All seven LSm proteins possess highly conserved Sm1 and Sm2 sequence motifs that are also present in the Sm proteins (Figure 2). Owing to the high degree of conservation of the Sm motifs, it is likely that the tertiary fold of the LSm proteins is similar to that of the Sm proteins and that the LSm proteins are also involved in protein–protein interactions. Indeed, we find that the LSm proteins interact with each other and form a stable protein complex. We cannot rigorously exclude the possibility that the LSm proteins are part of several distinct complexes. This is, however, not very likely for several reasons. Most importantly, all proteins elute from the cation-exchange column in a single homogeneous peak (Figure 1B). Fortuitous co-elution of individual LSm proteins is highly unlikely, because the charge of the
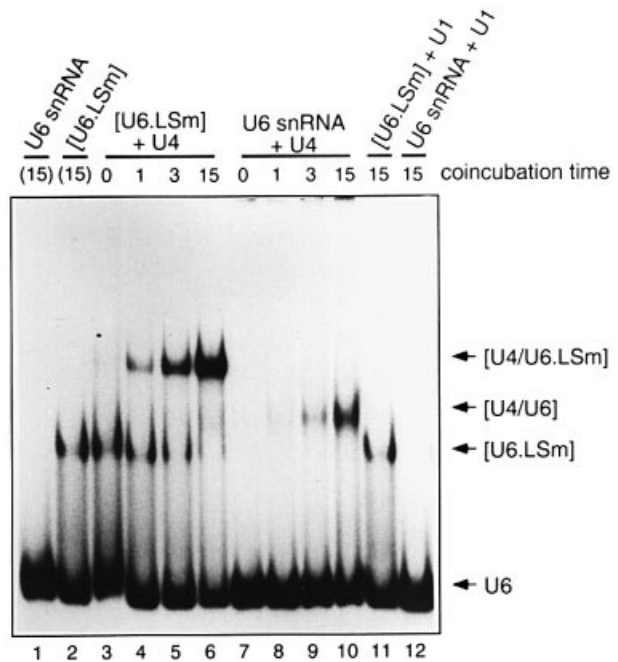

Fig. 8. The LSm proteins facilitate U4/U6 snRNA annealing. U6 snRNA prepared by transcription *in vitro* was pre-incubated for 15 min at 30°C with (lanes 2–6, 11) or without (1, 7–10, 12) 1 pmol of LSm proteins. Subsequently, a 10-fold excess of unlabelled U4 (lanes 3–10) or U1 snRNA (lanes 11 and 12) were added and incubation was continued for the time indicated above each lane. In lanes 1 and 2, incubation was continued for 15 min without addition of unlabelled RNA. The samples were chilled on ice and fractionated by native PAGE as described above. An autoradiograph of the gel is shown; the positions of the free RNA and of the complexes are indicated on the right.

various proteins is rather different, with theoretical pI values ranging from 4.3 to 10.0. Secondly, a high proportion of images in the electron micrographs exhibit a doughnut-like structure (Figure 3). Finally, the LSm proteins were isolated from purified tri-snRNP particles, and their monospecific RNA binding activity (Figure 5A) provides strong evidence against the possibility that several different LSm complexes bind to distinct sites on the U4, U5 or U6 snRNA. Therefore, the tri-snRNP-associated LSm proteins are most probably all part of the same complex. Moreover, the LSm complex is stable even in the presence of chaotropic salt (0.4 M sodium thiocyanate; not shown). The biochemical stability of the LSm complex is all the more remarkable as this heteromer is free of RNA, whereas the seven Sm proteins form a stable complex only in the presence of RNA. The Sm motifs of the Sm and LSm proteins show no obvious divergence that could explain the differential stability of these complexes. It is possible that the long, proline-rich domain of SmB, which has no counterpart in the LSm proteins, inhibits stable complex formation of all seven Sm proteins in the absence of RNA.

The doughnut-shaped structure of the RNA-free complex observed under the electron microscope deserves special attention because it shows a high degree of similarity, both in size and shape, to the structure of the Sm core RNPs (Figure 3). Even several structural details, such as fine lines of stain in the doughnut structure are found in both the LSm complex (Figure 3) and the Sm core RNP (Kastner *et al.*, 1990). Based on the crystal

structures of two Sm dimers, D1·D2 and B·D3, a structural model has been proposed for the Sm core RNP (Kambach *et al.*, 1999). This model predicts that all Sm proteins fold in a similar manner and that seven Sm proteins then oligomerize to form a ring which is consistent with the dimensions and morphology of the Sm core RNP structure seen under the electron microscope. Moreover, the accumulation of stain in the middle is consistent with the model which predicts that there is a hole in the centre. Interestingly, the central cavity appears to be ~0.5 nm wider in the RNA-free LSm complex than in the Sm core RNP, indicating that the Sm core RNP has some extra mass in the middle. Based on an accumulation of positively charged residues, the model of Kambach *et al.* (1999) predicts that the RNA may bind to the centre of the Sm ring structure, which may be consistent with the different appearances of the central cavities. Similar to the situation in the Sm core RNP, the purified LSm complex contains seven different LSm proteins and it is tempting to speculate that the LSm protein complex may also be a heptamer containing one copy of each protein. To substantiate the model further, it will be important to determine the stoichiometry of the protein composition in both the Sm core RNP and the LSm protein complex. Taken together, these data suggest that proteins with a conserved Sm sequence motif may generally form heteromers of this particular shape and size.

The LSm proteins are highly conserved throughout all eukaryotic organisms (Figure 2D; Salgado-Garrido *et al.*, 1999). In budding yeast, the orthologues of the human LSm proteins, Lsm2p–Lsm8p, associate with U6 as well as U4/U6 snRNPs and tri-snRNP particles, and are also present in purified tri-snRNPs (see Introduction). It is highly likely that these proteins also form a complex. Indeed, a tagged version of Lsm8p co-precipitates the Lsm2p–Lsm7p proteins, but not another copy of Lsm8p, supporting the idea that the LSm complex contains only one copy of each protein (Salgado-Garrido *et al.*, 1999). In addition to the tri-snRNP-associated LSm proteins, both yeast and human cells contain an additional Sm-like protein, Lsm1p/CaSm, which is not present in either tri-snRNP. This indicates that the specificity of the LSm protein complex assembly has been conserved in evolution. Genetic interactions of the *LSM1* gene in yeast suggest that its function is related to mRNA decapping (Boeck *et al.*, 1998). It will be interesting to see whether its human homologue, CaSm, has a similar function, and whether the Lsm1p/CaSm protein functions individually or is also part of an Sm-like protein complex.

The Sm domain is an ancient protein sequence motif, as it is also found in archaea. Archaeal genomes, however, encode only one or two Sm-like proteins. These proteins may thus be the progenitor of the Sm and Sm-like proteins, and the eukaryotic Sm and LSm proteins have probably evolved from this root by a combination of gene duplication events and diversification. It remains to be seen whether these proteins also form doughnut-shaped higher order assemblies or whether this property has been acquired by the Sm and LSm proteins later in evolution.

The yeast Lsm proteins have previously been shown to associate with U6 snRNPs, but whether they interact directly with the U6 snRNA was unknown (see Introduction). Here, we have shown that the LSm proteins bind directly to RNA. Specifically, the LSm proteins bind to U6, but not U1, U2, U4 or U5 snRNAs, and the 3′-terminal oligo(U) tract of U6 snRNA is essential for LSm protein binding (Figure 5A). In U6 snRNA, the oligo(U) tract potentially basepairs with adenosines to form part of the central stem (Figure 6A). Mutational analysis, however, has shown that the binding efficiency is not reduced when the stem is disrupted (Figure 6B). We conclude that the LSm complex is most likely to unwind the helix at the bottom of the central stem of the U6 snRNA and binds to single-stranded oligouridylic acid. This idea is supported by the finding that the addition of four single-stranded uridines to the 3′-end of the U4 snRNA induces efficient binding of the LSm proteins (not shown). Thus, a single-stranded oligo(U) terminus is a major determinant for LSm protein binding. Nevertheless, additional contacts to other RNA elements and/or other tri-snRNP proteins may well stabilize the association of the LSm complex with native particles, as the LSm proteins appear to bind more stably to native tri-snRNPs than to a 3′-terminal RNase T_1 fragment (compare Figure 4 with Figure 7).

Interestingly, a single-stranded uridine-rich tract is also a major determinant for RNA binding of the Sm proteins. Replacement of the U-tract in the Sm site for oligo(C) completely abolishes binding of the Sm proteins (Raker *et al.*, 1999). Similarly, substitution of the uridines with cytidines at the 3′-terminus of U6 abolishes LSm protein binding (Figure 6). Thus, proteins containing Sm motifs appear to exhibit more generally oligo(U)-specific RNA binding activity. Another feature that appears to be common to proteins with Sm sequence motifs is that only complexes of several Sm or Sm-like proteins exhibit RNA-binding activity. For example, co-operative interaction of at least five Sm proteins is necessary for stable RNA binding (see Introduction). Similarly, the LSm proteins appear to bind as a complex to RNA. This is suggested by the migration behaviour of the U6 snRNA·LSm complex, compared with the U4 snRNA·Sm complex, on native gels (Figure 5B). This idea is further supported by electron micrographs of 14S U4/U6 particles which contain both Sm and LSm proteins (O.V.Makarova, E.M.Makarov, B.Kastner and R.Lührmann, in preparation); many of the images possess two doughnut-shaped structures, indicating that the structure of the LSm complex does not change significantly upon binding to the U6 snRNA.

Despite these similarities, there are significant differences in the properties of the RNP complexes formed with the Sm and LSm proteins. The Sm core RNP is extremely salt stable (Liautard *et al.*, 1982), and does not dissociate upon addition of competitor RNA (Raker *et al.*, 1996). In contrast, the LSm complex dissociates from the U6 snRNA at salt concentrations >0.5 M (Figures 1A and 4) or in the presence of competitor RNA (not shown), indicating that the LSm protein–RNA complex is unstable kinetically. It has been shown that the Sm proteins require only a short ribo-oligonucleotide, AAUUUUUGA, for stable binding (Raker *et al.*, 1999). This motif is similar to the 3′-terminus of U6, UAUUUUU, and thus the differences in binding stability are probably due to differences in the proteins rather than the RNA-binding sites. Despite their similar binding sites, the LSm proteins do

not recognize the canonical Sm sites within the snRNAs (Figure 5A). One possible explanation for this difference might be that they interact only as a pre-formed complex. For example, the interaction of the LSm proteins could involve threading of the U-tract through the centre of the ring, which may be too narrow for the stem-loop structures that flank canonical Sm sites (Kambach *et al.*, 1999). In this case, the pre-formed LSm complex could bind only to terminal sequences, whereas the Sm protein complex, which is assembled on the RNA in a stepwise manner and would not require a threading process, could bind to an internal site. Conversely, the Sm proteins do not bind to U6 snRNA (Figure 5B). Possibly, the Sm proteins need a AA dinucletide before the U-tract for maximal binding efficiency. Furthermore, it can be envisioned that the internal duplex (see above) prevents initiation of Sm core assembly, but not binding of the pre-assembled LSm complex.

In α-LSm4 precipitation and pCp-labelling experiments, we detected only the tri-snRNP-associated snRNAs (Figure 4). There are, however, many more RNAs containing 3′-terminal U-tracts that might potentially associate with the LSm proteins. Owing to limitations of the pCp-labelling procedure, we could not detect minor RNAs or RNAs with a modified 3′-terminus, such as U6. For example, the U4atac/U6atac snRNP has been shown by immunoprecipitation and Northern blotting with specific probes to associate with the LSm proteins (C.Schneider and R.Lührmann, unpublished), but owing to its low abundance it escaped detection by our methods. Furthermore, the yeast LSm proteins associate with pre-RNaseP RNA (Salgado-Garrido *et al.*, 1999). This, and perhaps additional RNAs, may also bind to the human LSm proteins. Nonetheless, we can conclude from our pCp-labelling experiments that the bulk of RNA polymerase III transcripts do not associate with the LSm proteins. Most likely, the potential LSm-protein-binding sites on these RNAs are sequestered by the highly abundant La protein. The latter has been shown to bind to the oligo(U) terminus of polymerase III transcripts (Stefano, 1984), including pre-U6 snRNA (Rinke and Steitz, 1985). Mature U6 snRNA, in contrast, contains a 2′,3′-terminal cyclic phosphate (Lund and Dahlberg, 1992) which inhibits the binding of La (Stefano, 1984; Terns *et al.*, 1992). The Sm-like proteins bind efficiently both to RNAs that have a 3′-hydroxyl end (Figures 5 and 6) and to mature U6 snRNA containing a 2′3′-cyclic phosphate, as they were isolated from native tri-snRNP particles and associate with the 3′ terminus of native U6 snRNA (Figure 7). Therefore, mature U6 snRNA is a target for LSm protein binding, whereas the La protein binds only to primary RNA polymerase III transcripts. The idea that the La protein and the LSm complex compete for the same binding site on U6 snRNA is also consistent with the finding that mutation of both the La and Lsm8p proteins in yeast dramatically destabilizes U6 snRNA leading to a lethal phenotype (Pannone *et al.*, 1998). In wild-type cells, the La protein and the Sm-like protein complex probably bind consecutively to U6, as La binds only to pre-U6 snRNAs and the Sm-like proteins associate with the mature U6 snRNA present in tri-snRNPs. Although it is currently not known at what stage the Sm-like proteins bind, it will be very interesting to determine whether they are already functional during the maturation of U6 snRNA.

Our finding that the LSm proteins bind to the 3′-end of U6 snRNA is also interesting with respect to U4/U6 RNA duplex formation. In the mammalian system, U4/U6 duplex formation with naked snRNAs is destabilized by the secondary structure formed by the 3′-end of U6 snRNA (Brow and Vidaver, 1995), and U4 and U6 snRNAs anneal only in the presence of nuclear proteins (Wolff and Bindereif, 1993). Therefore, any protein that binds to the 3′-end of U6 and alters the conformation of this region could affect the stability of the U4/U6 duplex. This is exactly what the LSm proteins appear to do; specifically, we were able to demonstrate with purified components that the U6 snRNA·LSm complex anneals more efficiently with U4 snRNA than does isolated U6 snRNA (Figure 8). This raises the possibility that the LSm complex may also facilitate the annealing of the U4/U6 snRNAs *in vivo*. In yeast, Prp24p has been demonstrated to facilitate U4/U6 duplex formation *in vitro* (Raghunathan and Guthrie, 1998). In the human system, no functional homologue of Prp24p has yet been identified. Nevertheless, we cannot exclude that there is also a chaperone activity which, in addition to the LSm proteins, facilitates U4/U6 snRNP assembly, e.g. by modulating RNA structure. Furthermore, the LSm proteins may interact with other snRNP-specific proteins and thus have an additional role in snRNP assembly. Since the LSm proteins readily dissociate from U6 snRNA (see above), it is quite possible that their interaction is dynamic in nature, and it will be interesting to see whether they indeed dissociate from and re-associate with U6 at certain stages of the spliceosomal cycle.

Finally, there is strong evidence that U6 snRNA is retained in the nucleus during biogenesis (Vankan *et al.*, 1990; Terns *et al.*, 1993; Boelens *et al.*, 1995). It is noteworthy that the oligo(U) 3′-terminus contributes to U6 nuclear retention (Boelens *et al.*, 1995). This effect has been ascribed to the interaction of the La protein with the oligo(U) tract, but the LSm proteins might also mediate this retention. As U6 remains in the nucleus, the LSm proteins apparently enter the nucleus without RNA and future studies will have to show whether the LSm complex assembles already in the cytoplasm and is transported to the nucleus as such. Interestingly, the survival of motor neurons (SMN) protein and SMN-interacting proteins (SIPs) associate with several Sm proteins in the cytoplasm and are essential for U snRNP core assembly *in vivo* (Fischer *et al.*, 1997; Liu *et al.*, 1997). In light of the extensive similarities between the Sm and LSm proteins and the close resemblance of their higher order structures, SMN and the SIPs, or perhaps related proteins, might well play a similar role in the assembly of the LSm complex.

Materials and methods

Purification and characterization of the Sm-like proteins
Dissociation of purified snRNP particles in the presence of 0.7 M NaCl and separation of the resulting snRNP components on glycerol gradients has been described previously (Laggerbauer *et al.*, 1998). For further purification, fractions containing the Sm-like proteins were taken from a glycerol gradient containing 0.7 M NaCl and dialysed against MonoS buffer (20 mM HEPES, 1.5 mM MgCl$_2$, final pH 7.6) containing 125 mM

NaCl. The sample was loaded onto a 0.1 ml MonoS column (Pharmacia) and eluted with a 2-ml, linear gradient of MonoS buffer containing 100–400 mM NaCl, followed by 0.5 ml of MonoS buffer containing 1 M NaCl. The Sm-like proteins elute at 250 mM salt.

For peptide sequencing, proteins were excised from preparative SDS–polyacrylamide gels, digested with LysC and subjected to Edman sequence analysis (Toplab, Munich, Germany). Alternatively, the proteins were digested with trypsin and analysed by nanoelectrospray mass spectrometry as described previously (Neubauer *et al.*, 1997). Database searches using the TBLASTN and BLASTP programs were performed on the NIH server. The ESTs were obtained from RZPD (Berlin, Germany) and sequenced using an automated DNA sequencer (ABI). The 5′-termini of the LSm7 and LSm8 cDNAs were amplified by PCR from a human cDNA bank (Marathon DNA; Clonetech) as described by the manufacturer using gene-specific primers that anneal to nucleotide positions 227–256 and 259–285, respectively, of the ORFs.

For electron-microscopy analysis of the MonoS-purified proteins, negative staining with 2.5% (w/v) uranyl formate was performed using the double-carbon-film method as described previously (Kastner *et al.*, 1990). Preparations were examined with a Philips CM120 Biotwin electron microscope operating with 120 kV. Electron micrographs were taken at a magnification of 105 000.

Antibody production and immunoprecipitations

A peptide containing the C-terminal 20 amino acids of LSm4 and an additional cysteine at its N-terminal end was synthesized by automated peptide synthesis and coupled to bovine serum albumin (BSA) via a sulfo-SMCC cross-linker (Pierce). The conjugate was used for immunization of rabbits as described previously (Lauber *et al.*, 1997).

RNPs were precipitated from 100 μl of HeLa nuclear extract using 50 μl of serum bound to 10 μl of pre-blocked protein A–Sepharose essentially as described by Raker *et al.* (1996). Briefly, the extract was diluted with phosphate-buffered saline (PBS), added to the antibody beads and incubated for 1 h at 4°C with constant rotation. Subsequently, the beads were washed five times with 1 ml precipitation buffer (20 mM Tris–HCl, pH 8.0, 150–500 mM NaCl, 0.05% Nonidet P40), changing the reaction container once. RNA was released by treatment with 1% SDS and phenol extraction and labelled with [^{32}P]pCp as described previously (Fabrizio *et al.*, 1997).

In the RNase protection/immunoprecipitation experiments, 10 μg of glycerol gradient-purified tri-snRNP particles (Laggerbauer *et al.*, 1998) were digested with 10 μg RNase T$_1$ (Boehringer) for 30 min at 25°C, diluted with 300 μl precipitation buffer and used for precipitations as described above. The recovered RNA was labelled with 10 μCi (2 pmol) [γ^{32}P]ATP by using T4 polynucleotide kinase and fractionated on 8.5 M urea, 20% polyacrylamide gels. Partial digestions of individual RNA fragments were performed under denaturing conditions using sequencing grade RNases as described by the manufacturer (Pharmacia).

Electrophoretic mobility shift assays

snRNAs were generated by run-off transcription from linearized plasmid DNA encoding U1, U2 and U5 snRNAs (Fischer *et al.*, 1991) in the presence of [α^{32}P] UTP using SP6 or T7 RNA polymerase as described by the manufacturer (Promega). In the case of the U4 and U6 snRNAs, the proper end of the template was generated by PCR amplification of the coding region of the plasmid using the M13 sequencing primer and primers that pair to the desired 3′-terminal sequences. Templates encoding artificial U6 constructs and including the T7 promoter sequence were synthesized by automated DNA synthesis and made double-stranded using the T7 primer and *Taq* DNA polymerase.

RNAs prepared by transcription *in vitro* (10 000 c.p.m.; 0.4 pmol) were incubated with up to 2 pmol of LSm proteins (assuming a combined molecular mass of 100 kDa) for 15 min at 30°C in 5 μl of buffer containing 12 mM HEPES pH 7.6, 1.5 mM MgCl$_2$, 300 mM NaCl, 10% glycerol, 0.1% Triton X100 and 0.1 μg/μl competitor tRNA. The samples were then chilled on ice and loaded onto a 6% AA, 0.075% bis-AA gel prepared with 0.5× TBE buffer and separated by electrophoresis at 10 V/cm for 4 h in the cold room. RNA-free Sm proteins were prepared from glycerol gradient-purified U1 and U2 snRNPs as described previously (Sumpter *et al.*, 1992). In addition to the Sm proteins, this preparation also contained U1- and U2-specific proteins, but no tri-snRNP-specific proteins. In U4/U6 annealing experiments, the U6 snRNA and the LSm proteins were first incubated as described above. Subsequently, 4 pmol of unlabelled U4 snRNA was added and incubation was continued for up to 15 min, The reaction mixture was then analysed on native PAGE as described above.

References

Boeck,R., Lapeyre,B., Brown,C.E. and Sachs,A.B. (1998) Capped mRNA degradation intermediates accumulate in the yeast spb8-2 mutant. *Mol. Cell. Biol.*, **18**, 5062–5072.

Boelens,W.C., Palacios,I. and Mattaj,I.W. (1995) Nuclear retention of RNA as a mechanism for localization. *RNA*, **1**, 273–283.

Booth,B.L. Jr and Pugh,B.F. (1997) Identification and characterization of a nuclease specific for the 3′ end of the U6 small nuclear RNA. *J. Biol. Chem.*, **272**, 984–991.

Branlant,C., Krol,A., Ebel,J.P., Lazar,E., Haendler,B. and Jacob,M. (1982) U2 RNA shares a structural domain with U1, U4 and U5 RNAs. *EMBO J.*, **1**, 1259–1265.

Brow,D.A. and Vidaver,R.M. (1995) An element in human U6 RNA destabilizes the U4/U6 spliceosomal RNA complex. *RNA*, **1**, 122–131.

Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) Splicing of precursors to mRNAs by the spliceosomes. In Gesteland,R.F., Chech,T.R. and Atkins,J.E. (eds), *The RNA World*, 2nd edn. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Camasses,A., Bragado-Nilsson,E., Martin,R., Séraphin,B. and Bordonné,R. (1998) Interactions within the yeast Sm core complex: from proteins to amino acids. *Mol. Cell. Biol.*, **18**, 1956–1966.

Cooper,M., Johnston,L.H. and Beggs,J.D. (1995) Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.*, **14**, 2066–2075.

Dahlberg,J.E. and Lund,E. (1991) How does III × II make U6? *Science*, **254**, 1462–1463.

Fabrizio,P., Laggerbauer,B., Lauber,J., Lane,W.S. and Lührmann,R. (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.*, **16**, 4092–4106.

Fischer,U. and Lührmann,R. (1990) An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science*, **249**, 786–790.

Fischer,U., Darzynkiewicz,E., Tahara,S.M., Dathan,N.A., Lührmann,R. and Mattaj,I.W. (1991) Diversity in the signals required for nuclear accumulation of U snRNPs and variety in the pathways of nuclear transport. *J. Cell Biol.*, **113**, 705–714.

Fischer,U., Liu,Q. and Dreyfuss,G. (1997) The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell*, **90**, 1023–1029.

Fromont-Racine,M., Rain,J.C. and Legrain,P. (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.*, **16**, 277–282.

Fury,M.G., Zhang,W., Christodoulopoulos,I. and Zieve,G.W. (1997) Multiple protein:protein interactions between the snRNP common core proteins. *Exp. Cell Res.*, **237**, 63–69.

Gottschalk,A., Neubauer,G., Banroques,J., Mann,M., Lührmann,R. and Fabrizio,P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6·U5] tri-snRNP. *EMBO J.*, **18**, 4535–4548.

Hamm,J. and Mattaj,I.W. (1990) Monomethylated cap structures facilitate RNA export from the nucleus. *Cell*, **63**, 109–118.

Hermann,H., Fabrizio,P., Raker,V.A., Foulaki,K., Hornig,H., Brahms,H. and Lührmann,R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. *EMBO J.*, **14**, 2076–2088.

Kambach,C., Walke,S., Young,R., Avis,J.M., de la Fortelle,E., Raker, V.A., Lührmann,R., Li,J. and Nagai,K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, **96**, 375–387.

Kastner,B., Bach,M. and Lührmann,R. (1990) Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5:

evidence for a common structure-determining principle in the major U snRNP family. *Proc. Natl Acad. Sci. USA*, **87**, 1710–1714.

Kawarabayasi,Y. *et al.* (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.*, **5**, 55–76.

Klenk,H.P. *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364–370.

Krämer,A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.*, **65**, 367–409.

Laggerbauer,B., Achsel,T. and Lührmann,R. (1998) The human U5–200kD DEXH-box protein unwinds U4/U6 RNA duplices *in vitro*. *Proc. Natl Acad. Sci. USA*, **95**, 4188–4192.

Lauber,J., Plessel,G., Prehn,S., Will,C.L., Fabrizio,P., Groning,K., Lane,W.S. and Lührmann,R. (1997) The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA*, **3**, 926–941.

Lehmeier,T., Raker,V., Hermann,H. and Lührmann,R. (1994) cDNA cloning of the Sm proteins D2 and D3 from human small nuclear ribonucleoproteins: evidence for a direct D1–D2 interaction. *Proc. Natl Acad. Sci. USA*, **91**, 12317–12321.

Liautard,J.P., Sri-Widada,J., Brunel,C. and Jeanteur,P. (1982) Structural organization of ribonucleoproteins containing small nuclear RNAs from HeLa cells. Proteins interact closely with a similar structural domain of U1, U2, U4 and U5 small nuclear RNAs. *J. Mol. Biol.*, **162**, 623–643.

Liu,Q., Fischer,U., Wang,F. and Dreyfuss,G. (1997) The spinal muscular atrophy disease gene product, SMN and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell*, **90**, 1013–1021.

Lund,E. and Dahlberg,J.E. (1992) Cyclic 2′,3′-phosphates and nontemplated nucleotides at the 3′ end of spliceosomal U6 small nuclear RNAs. *Science*, **255**, 327–330.

Mattaj,I.W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, **46**, 905–911.

Mayes,A.E., Verdone,L., Legrain,P. and Beggs,J.D. (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J.*, **18**, 4321–4331.

McAllister,G., Roby-Shemkovitz,A., Amara,S.G. and Lerner,M.R. (1989) cDNA sequence of the rat U snRNP-associated protein N: description of a potential Sm epitope. *EMBO J.*, **8**, 1177–1181.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursors to messenger RNAs by the spliceosome. In Gesteland,R.F. and Atkins,J.E. (eds), *The RNA World*. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Nelissen,R.L., Will,C.L., van Venrooij,W.J. and Lührmann,R. (1994) The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.*, **13**, 4113–4125.

Neubauer,G., Gottschalk,A., Fabrizio,P., Séraphin,B., Lührmann,R. and Mann,M. (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl Acad. Sci. USA*, **94**, 385–390.

Nottrott,S., Hartmuth,K., Fabrizio,P., Urlaub,H., Vidovic,I., Ficner,R. and Lührmann,R. (1999) Functional interaction of a novel 15.5kD [U4/U6·U5] tri-snRNP protein with the 5′ stem–loop of U4 snRNA. *EMBO J.*, in press.

Pannone,B.K., Xue,D. and Wolin,S.L. (1998) A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J.*, **17**, 7442–7453.

Plessel,G., Fischer,U. and Lührmann,R. (1994) m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) *in vitro*: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol. Cell. Biol.*, **14**, 4160–4172.

Raghunathan,P.L. and Guthrie,C. (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science*, **279**, 857–860.

Raker,V.A., Plessel,G. and Lührmann,R. (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. *EMBO J.*, **15**, 2256–2269.

Raker,V.A., Hartmuth,K., Kastner,B. and Lührmann,R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol.*, **19**, 6554–6565.

Rinke,J. and Steitz,J.A. (1985) Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res.*, **13**, 2617–2629.

Salgado-Garrido,J., Bragado-Nilsson,E., Kandels-Lewis,S. and Séraphin,B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.*, **18**, 3451–3462.

Schweinfest,C.W., Graber,M.W., Chapman,J.M., Papas,T.S., Baron,P.L. and Watson,D.K. (1997) CaSm: an Sm-like protein that contributes to the transformed state in cancer cells. *Cancer Res.*, **57**, 2961–2965.

Séraphin,B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.*, **14**, 2089–2098.

Singh,R. and Reddy,R. (1989) Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc. Natl Acad. Sci. USA*, **86**, 8280–8283.

Smith,D.R. *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.*, **179**, 7135–7155.

Staley,J.P. and Guthrie,C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell*, **92**, 315–326.

Stefano,J.E. (1984) Purified lupus antigen La recognizes an oligouridylate stretch common to the 3′ termini of RNA polymerase III transcripts. *Cell*, **36**, 145–154.

Stevens,S.W. and Abelson,J. (1999) Purification of the yeast U4/U6·U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl Acad. Sci. USA*, **96**, 7226–7231.

Sumpter,V., Kahrs,A., Fischer,U., Kornstädt,U. and Lührmann,R. (1992) *In vitro* reconstitution of U1 and U2 snRNPs from isolated proteins and snRNA. *Mol. Biol. Rep.*, **16**, 229–240.

Teigelkamp,S., Achsel,T., Mundt,C., Göthel,S.F., Cronshagen,U., Lane,W.S., Marahiel,M. and Lührmann,R. (1998) The 20kD protein of human [U4/U6·U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA*, **4**, 127–141.

Terns,M.P., Dahlberg,J.E. and Lund,E. (1993) Multiple *cis*-acting signals for export of pre-U1 snRNA from the nucleus. *Genes Dev.*, **7**, 1898–1908.

Terns,M.P., Lund,E. and Dahlberg,J.E. (1992) 3′-End-dependent formation of U6 small nuclear ribonucleoprotein particles in *Xenopus laevis* oocyte nuclei. *Mol. Cell. Biol.*, **12**, 3032–3040.

Trippe,R., Sandrock,B. and Benecke,B.J. (1998) A highly specific terminal uridylyl transferase modifies the 3′-end of U6 small nuclear RNA. *Nucleic Acids Res.*, **26**, 3119–3126.

Vankan,P., McGuigan,C. and Mattaj,I.W. (1990) Domains of U4 and U6 snRNAs required for snRNP assembly and splicing complementation in *Xenopus* oocytes. *EMBO J.*, **9**, 3397–3404.

Will,C.L. and Lührmann,R. (1997) Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.*, **9**, 320–328.

Wilm,M. and Mann,M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal. Chem.*, **68**, 1–8.

Wilm,M., Shevchenko,A., Houthaeve,T., Breit,S., Schweigerer,L., Fotsis,T. and Mann,M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature*, **379**, 466–469.

Wolff,T. and Bindereif,A. (1993) Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4–U6 interaction and for the first step of slicing. *Genes Dev.*, **7**, 1377–1389.