The EMBO Journal Vol.18 No.12 pp.3451–3462, 1999

Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin

Josefa Salgado-Garrido, Elisabeth Bragado-Nilsson, Stefanie Kandels-Lewis[1] and Bertrand Séraphin[2]

EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

[1]Present address: University of Edinburgh, King's Buildings, Michael Swann Building, ICMB, Mayfield Road, Edinburgh EH9 3JR, UK

[2]Corresponding author
e-mail: seraphin@embl-heidelberg.de

A group of seven Sm proteins forms a complex that binds to several RNAs in metazoans. All Sm proteins contain a sequence signature, the Sm domain, also found in two yeast Sm-like proteins associated with the U6 snRNA. We have performed database searches revealing the presence of 16 proteins carrying an Sm domain in the yeast genome. Analysis of this protein family confirmed that seven of its members, encoded by essential genes, are homologues of metazoan Sm proteins. Immunoprecipitation revealed that an evolutionarily related subgroup of seven Sm-like proteins is directly associated with the nuclear U6 and pre-RNase P RNAs. The corresponding genes are essential or required for normal vegetative growth. These proteins appear functionally important to stabilize U6 snRNA. The two last yeast Sm-like proteins were not found associated with RNA, and neither was essential for vegetative growth. To investigate whether U6-associated Sm-like protein function is widespread, we cloned several cDNAs encoding homologous human proteins. Two representative human proteins were shown to associate with U6 snRNA-containing complexes. We also identified archaeal proteins related to Sm and Sm-like proteins. Our results demonstrate that Sm and Sm-like proteins assemble in at least two functionally conserved complexes of deep evolutionary origin.

Keywords: human splicing factors/RNase P/ *Saccharomyces cerevisiae*/spliceosomal snRNP/U6 snRNA

Introduction

Small nuclear ribonucleoproteins (snRNPs) are particles that are found in the nucleus of eukaryotic cells. They consist of one RNA (snRNA) associated with one or more proteins. snRNPs are involved in a wide variety of functions including pre-mRNA splicing (e.g. U1, U2, U4–6, U11, U12, $U4_{ATAC}$, $U6_{ATAC}$ snRNPs and *trans*-spliced leader RNPs), histone mRNA 3′ end processing (U7 snRNP), rRNA processing (e.g. U3, U8, U13–72 snRNPs and RNase MRP), telomere replication (telomerase), tRNA maturation (RNase P) etc. (reviewed in Mattaj *et al.*, 1993). Most of these snRNPs are ubiquitously found in eukaryotes, even though some of them appear to be absent in some species (e.g. U7 in *Saccharomyces cerevisiae*). In addition, several snRNPs of unknown function have also been identified (Yu *et al.*, 1996; Smith and Steitz, 1997), some being encoded by viral rather than by cellular genomes (Lee *et al.*, 1988). The abundance of snRNPs is extremely variable depending on the snRNP and the species, ranging from a few copies per nucleus (e.g. yeast telomerase) to 10^6 particles per nucleus (e.g. human U1 snRNP; Lührmann *et al.*, 1990).

For unknown reasons, antibodies recognizing the protein or RNA moiety of snRNPs are frequently generated during autoimmune diseases. Among them, autoantibodies from patients suffering of systemic lupus erythemathosus were found to recognize an antigen named Sm (Tan, 1989). This antigen was found associated with the U7 snRNP, some snRNPs of unknown functions as well as all snRNPs involved in pre-mRNA splicing, with the notable exception of U6 (and probably $U6_{ATAC}$). This antigen was characterized as a group of eight small proteins named B, B′, D1, D2, D3, E, F and G, often referred to as Sm proteins (Lührmann *et al.*, 1990). An additional Sm protein, SmN, is expressed in the nervous system (McAllister *et al.*, 1989). The Sm antigen was found to be well conserved in metazoans, but is only weakly recognizable in more distant organisms (e.g. yeast and fungi; Tollervey and Mattaj, 1987). Antibodies of Sm specificity were used as tools to study the function of the Sm proteins. This revealed that Sm proteins are kept in the cytoplasm of metazoan cells until they assemble with the cognate snRNAs (Izaurralde and Mattaj, 1992). Sm protein binding involves the formation of specific interactions between Sm proteins (Raker *et al.*, 1996; Fury *et al.*, 1997; Camasses *et al.*, 1998; Kambach *et al.*, 1999) and the direct recognition of a short single-stranded sequence element whose generic structure is A(U)nG (Branlant *et al.*, 1982; Guthrie and Patterson, 1988). This RNA sequence is sufficient for Sm protein interaction (Mattaj, 1986) but does not entirely determine the efficiency of binding which is also modulated by snRNP specific proteins and/or neighboring RNA sequences and structures (Jarmolowski and Mattaj, 1993; Nelissen *et al.*, 1994). Once bound to the snRNA, Sm proteins trigger hypermethylation of the cap (Plessel *et al.*, 1994). The hypermethylated cap structure and the Sm proteins constitute two independent nuclear localization signals that will allow the transport of the newly assembled snRNP into the nucleus where it will be active (Izaurralde and Mattaj, 1992; Fischer *et al.*, 1994; Huber *et al.*, 1998). Along this pathway, Sm proteins interact with many cellular factors. Interestingly, these include the SMN protein encoded by the *SMA* gene, which is involved in human spinal muscular atrophy (Fischer *et al.*, 1997; Liu *et al.*, 1997). However,

it is not yet clear whether Sm proteins have any function other than promoting snRNA cap modification and targeting snRNPs to their appropriate cellular location.

Autoantibodies of the Sm specificity were also used to characterize the Sm antigen. These autoantibodies often recognize several Sm proteins in Western blot analyses. Most commonly they recognize SmB, SmB′, SmN and SmD1, but autoantibodies reacting with other Sm proteins have also been found (Rokeach *et al.*, 1992 and references therein). Immunoprecipitation analyses confirmed that some Sm autoantibodies independently recognize several of the Sm proteins (Brahms *et al.*, 1997). Most of these properties are also shared by a monoclonal antibody of the Sm specificity (Lerner *et al.*, 1981), suggesting that the various Sm proteins share a common epitope. Surprisingly, however, cloning and sequencing cDNAs coding for the SmB, SmB′, SmN, SmD1 and SmE proteins from various species did not originally reveal any obvious amino acid feature shared by all proteins. These studies nevertheless indicated that SmB and B′ are highly related, being generated by alternative splicing of a common primary transcript. Interestingly, this alternative splicing only occurs in human cells (Chu and Elkon, 1991). Consistent with previous immunological data, SmN was also found to be very similar to SmB and SmB′ (93% identity), even though it is encoded by a different gene (McAllister *et al.*, 1989). Upon careful re-examination of these Sm protein sequences, it became clear that a conserved domain is present in Sm proteins (Séraphin, 1995). This domain was independently identified following the cloning and sequencing of human cDNAs coding for additional Sm proteins (Hermann *et al.*, 1995). The Sm domain consists of two blocks of weak but significant sequence similarity interrupted by a spacer region of variable length that folds as a loop (Kambach *et al.*, 1999). The two blocks of sequence similarity are always found associated indicating that they are part of a single protein domain rather than independent protein motifs. Sensitive database searches revealed the presence of a large number of proteins containing an Sm domain in eukaryotes (Hermann *et al.*, 1995; Séraphin, 1995). Some of these proteins were highly similar to human Sm proteins, and therefore probably orthologues. However, other proteins had no clear counterpart in the Sm protein complex. These proteins were named Sm-like proteins (Séraphin, 1995). Phylogenetic analyses of this protein family demonstrated that Sm and Sm-like proteins could be grouped into various subtypes containing (putative) orthologues from various species. On one hand, the seven canonical Sm protein subtypes were named B, D1, D2, D3, E, F and G according to the corresponding human Sm protein member. On the other hand, Sm-like proteins could be divided in at least nine subtypes (Séraphin, 1995). Interestingly, some Sm-like protein subtypes showed close sequence relationship to specific canonical Sm protein subtypes. Characterization of two yeast Sm-like proteins revealed that they were specifically associated with the U6 snRNA (Cooper *et al.*, 1995; Séraphin, 1995).

With the completion of the yeast genome sequence, we can now address the function of the whole set of Sm proteins of an eukaryotic organism. We therefore identified all yeast Sm and Sm-like proteins. Characterization of these proteins reveals the existence of at least two distinct but phylogenetically related complexes. The first complex corresponds to the canonical Sm complex. The second complex contains Sm-like proteins and is associated with U6 snRNA and pre-RNase P RNA. Interestingly, this Sm-like protein complex appears to be required for the stability of the U6 snRNA. Identification and characterization of human Sm-like proteins indicate that this Sm-like protein complex is phylogenetically conserved. Furthermore, Sm-like proteins were identified in archaea, indicating the ancient origin of this family.

Results

Identification of yeast Sm and Sm-like proteins and phylogenetic relationship to canonical Sm proteins

We have used our previously described sequence profiles for Sm proteins (Séraphin, 1995) to search the complete yeast genome sequence (Goffeau *et al*., 1997). This scan, performed with the Searchwise software (Birney *et al.*, 1996) which allows for frameshifts and intron detection, revealed the existence of 16 yeast coding sequences potentially coding for Sm related proteins (Table I). Furthermore, no other coding sequences were detected by additional database searches using representative sequences of various Sm or Sm-like proteins (Séraphin, 1995). This strongly suggests that yeast does not code for more than 16 different Sm-related proteins (see also Fromont *et al.*, 1997). We then built a multiple alignment of the Sm domain for these 16 yeast proteins, seven human canonical Sm proteins, six human Sm-like proteins and five Sm-like proteins from archaebacteria. This multiple sequence alignment was used to build a neighbor-joining tree reflecting the relationship between the various sequences (Figure 1). Sm and Sm-like proteins group into distinct subtypes containing yeast and human members. (Figure 1: subtypes will be referred to by using the corresponding yeast protein name and human proteins by adding an h prefix.) Seven yeast proteins are highly similar to the seven canonical human Sm proteins. In addition, six subtypes of Sm-like proteins (named Lsm2–7) appear clearly related to, but nevertheless distinct from the canonical SmD1, D2, D3, E, F and G subtypes (Figure 1). Another Sm-like protein subtype (Lsm8) is weakly related to the SmB subtype. This relationship is further supported by the functional data presented below. Finally, two Sm-like protein subtypes (Lsm1 and Lsm9) are not clearly related to canonical Sm protein subtypes. The similarities between Sm-like and canonical Sm subtypes suggest a common origin for these two protein groups. The Sm-like protein names have been changed from the previously proposed general nomenclature (Séraphin, 1995) to facilitate the recognition of sequence (and functional, see results below) similarity between the Sm and Sm-like proteins. In this new system, Sm-like proteins are named Lsm (for like Sm) followed by a number. To be more easily remembered, numbers are selected such that the value reflects the alphabetical rank of the most similar canonical Sm protein (i.e. 2 for D1, 3 for D2, 4 for D3, 5 for E, 6 for F, 7 for G; see Table I and Figure 1). However, note that this relation does not hold true for SmB, which is functionally related to Lsm8, not Lsm1 (see results below).

Table I. Characteristics of Sm and Sm-like proteins: phenotypes and RNA directly associated

Gene	Chromosome	Intron	U1,2,4,5 snRNAs	U6 snRNA	pre-RNase P RNA	Phenotype	Disruption reference(s)
SMB	V	no	+	−	−	lethal	this work; Fromont *et al.* (1997)
SMD1	VII	no	+	−	−	lethal	Rymond (1993)
SMD2	XII	yes	+	−	−	lethal	this work
SMD3	XII	no	+	−	−	lethal	Roy *et al.* (1995)
SME	XV	no	+	−	−	lethal	Bordonné and Tarassov (1996)
SMF	XVI	no	+	−	−	lethal	this work
SMG	VI	no	+	−	−	lethal	Séraphin (1995)
LSM1	X	no	−	−	−	viable (Ts)	this work
LSM2	II	yes	−	+	+	lethal	this work
LSM3	XII	no	−	+	+	lethal	Séraphin (1995)
LSM4	V	no	−	+	+	lethal	Cooper *et al.* (1995)
LSM5	V	no	−	+	+	viable	this work
LSM6	IV	no	−	+	+	viable	this work
LSM7	XIV	yes	−	+	+	viable	this work
LSM8	X	no	−	+	−	lethal	this work
LSM9	III	no	−	−	−	viable	Séraphin (1995)

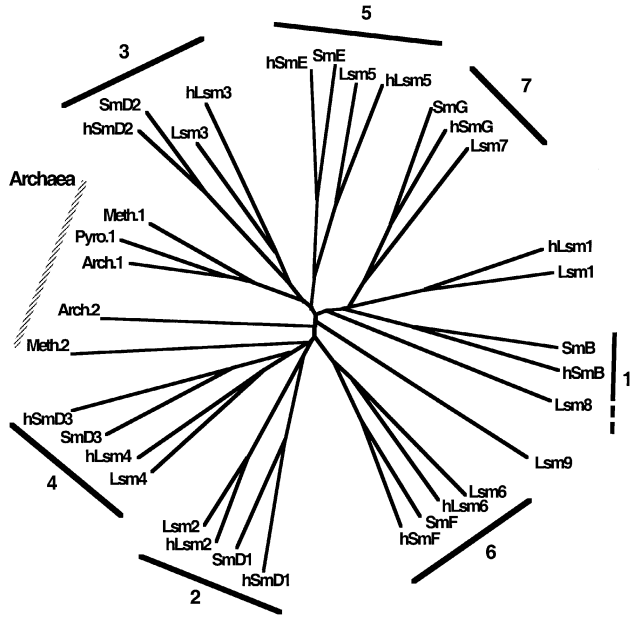

Fig. 1. Phylogenetic tree of yeast, human and archaeal Sm and Sm-like proteins. The tree was built from a multiple sequence alignment of the various Sm domains (Séraphin, 1995; see Materials and methods). Positions with gap were excluded. The seven subfamilies containing related subtypes of both the Sm and Sm-like (Lsm) proteins are indicated (1–7). Because sequence similarity in subfamily 1 is weak, it is indicated with a dashed lined. Human proteins are labeled with an 'h' prefix. Archaeal proteins are named with the first four letters of the species name (Arch, *Archaeoglobus fulgidus*; Meth, *Methanobacterium thermoautotrophicum*; Pyro, *Pyrococcus horikoshii* OT3) and a number. Reference for the sequences used for building the tree are indicated in Materials and methods.

Association of yeast Sm and Sm-like proteins with RNAs

To test whether yeast Sm and Sm-like proteins were associated with RNA, we followed the strategy that was previously used to characterize some of these proteins (Séraphin, 1995). Briefly, the putative promoter and coding region of each Sm or Sm-like protein was amplified by PCR and inserted in a yeast vector. In these plasmids, the Sm or Sm-like coding sequence is fused at its C-terminus to a cassette encoding two IgG-binding domains of *Staphyloccocus aureus* protein A (ProtA) (Materials and methods). These plasmids were introduced into a wild-type yeast strain and total cell extracts were prepared from individual transformants. Western blotting indicated that all fusion proteins were expressed and fractionated according to their calculated molecular weights (data not shown; see also Figure 4). These extracts were used in immunoprecipitation experiments. RNAs were recovered from immuno-pellets and analyzed by primer extension (Figure 2A). The yeast proteins highly similar to the human canonical Sm proteins precipitated efficiently the U1, U2, U4, U5 and U6 snRNAs (Figure 2A, lanes 3–9), as previously shown for some of them [SmF (formerly SmX3), SmG (formerly SmX2); Séraphin, 1995]. Except for SmB, these proteins were also found in the purified yeast U1 and U2 snRNPs (Neubauer *et al.*, 1997). Immunoprecipitation experiments revealed that it was associated with U1 snRNA (Gottschalk *et al.*, 1998) and we now demonstrate that it also associates with U2 snRNA. This indicates that the yeast SmB homologue has similar function as its metazoan counterparts and suggests that the original observation resulted from loss of this protein during biochemical purification. The yeast Sm proteins also interact with each other with the same specificity as their human counterparts (Raker *et al.*, 1996; Fury *et al.*, 1997; Camasses *et al.*, 1998). These results confirm that these seven yeast Sm proteins are homologous to the canonical human Sm proteins.

Seven Sm-like proteins (Lsm2–8) precipitated efficiently the U6 snRNA as well as a low level of U4 and U5 snRNAs (Figure 2A, lanes 10–16). This property was already known for Lsm3 (formerly SmX4; Séraphin, 1995) and Lsm4 (formerly Uss1; Cooper *et al.*, 1995). Finally, the two remaining Sm-like proteins, Lsm1 (formerly SmX8) and Lsm9 (formerly SmX1) and the Pop1 control, did not co-precipitate any spliceosomal snRNAs (Figure 2A, lanes 17–19) as expected in the two latter cases (Lygerou *et al.*, 1994; Séraphin, 1995). These results suggested that the Lsm2–8 proteins would be components of the U6 snRNP. To test this possibility, immunoprecipitated RNAs were fractionated on a native gel, blotted to a membrane and revealed with probes complementary to the U4, U5 and U6 snRNAs (Cooper *et al.*, 1995; Séraphin, 1995). The yeast SmB–G proteins

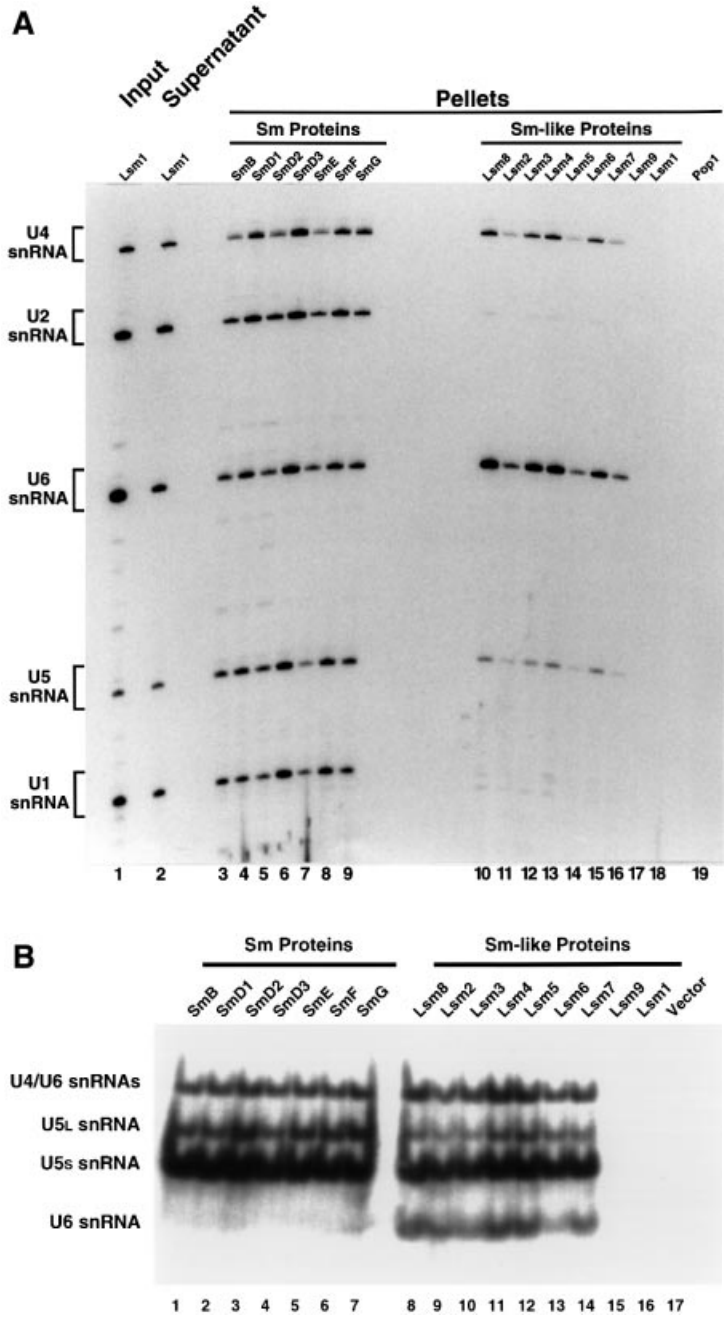

Fig. 2. Yeast contains two snRNA associated complexes made of Sm or Sm-like proteins. (**A**) Spliceosomal snRNA immunoprecipitated with the various yeast Sm (lanes 3–9) and Sm-like proteins (lanes 10–18) were detected by primer extension. RNA present in representative extract (lane 1) and supernatant (lane 2) are also presented. A ProtA–POP1 fusion (Lygerou *et al.*, 1994) was used to control for the specificity of the immunoprecipitation (lane 19). Primer extension products corresponding to the U1, U2, U4, U5 and U6 snRNA are labeled. Variations in immunoprecipitation efficiencies are likely to result from different expression of the tagged protein and competition with endogenous counterparts. (**B**) The nature of the association of Sm and Sm-like proteins with U6 snRNA was determined by fractionating immunoprecipitated snRNAs on a native gel. Sm proteins precipitate U6 molecules present in U4/U6 complexes (lanes 1–7). The Lsm2–8 proteins precipitate U6 snRNA present in free U6 snRNP in addition to U4/U6 (lanes 8–14). Lsm1 and Lsm9, or cells transformed with a vector, do not precipitate the U4–U6 snRNAs (lanes 15–17). The blot was probed with oligonucleotide complementary to the U4, U5 and U6 snRNAs. These species, including the short and long form of the yeast U5 snRNA (U5S and U5L, respectively), are indicated on the left. Because, the various oligonucleotide probes have different specific activities and hybridization efficiencies, the relative signals cannot be taken as an indication of the abundance of the various RNAs. The ratio of U6 snRNA to U4/U6 snRNA co-precipitated with the Lsm2–8 proteins was significantly higher than the same ratio for Sm proteins (data not shown).

precipitated U6 snRNA present in the U4/U6 snRNA as well as the two forms of yeast U5 snRNA as expected for canonical Sm proteins that associate directly with U4 and U5 snRNAs, indirectly with U6 snRNA present in the U4/U6 complex but not with free U6 snRNA (Lührmann *et al.*, 1990) (Figure 2B, lanes 1–7). In contrast, the Lsm2–8 proteins precipitated U6 snRNA present in both the U4/U6 and free U6 snRNPs (Figure 2B, lanes 8–14). Consistent with the primer extension analysis, U5 snRNA that is present in the U4/U6.U5 complex was also co-precipitated by the Lsm2–8 proteins (Figure 2B, lanes 8–14).

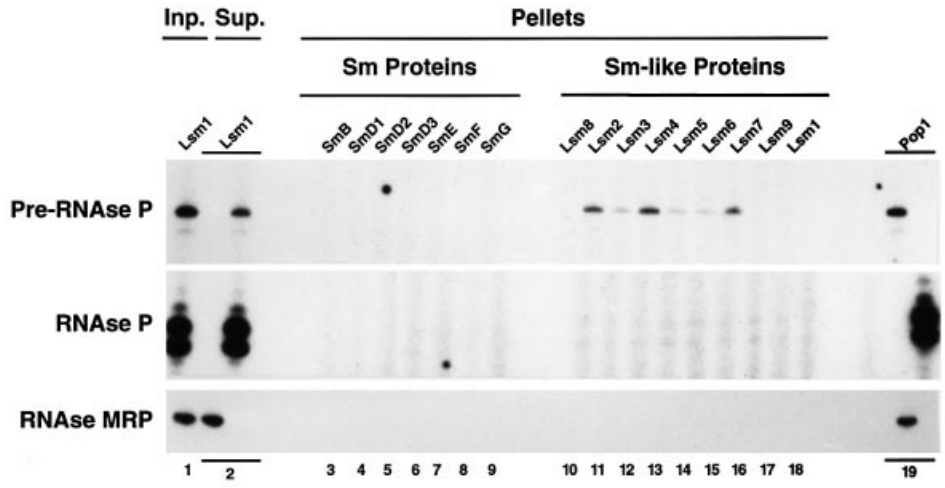

Fig. 3. Specific association of Lsm proteins with the pre-RNase P RNA. RNAs immunoprecipitated with Sm (lanes 3–9) and Sm-like (lanes 10–18) proteins were assayed by primer extension for the presence of the pre-RNase P (top row), RNase P (middle row) and RNase MRP (lower row) RNAs (see Séraphin, 1995). The pre-RNase P and RNase P signals result from extension of the same primer but have different lengths due to the presence of a 5′ end extension in the pre-RNase P species (Lee *et al.*, 1991). RNA presents in representative extract (lane 1) and supernatant (lane 2) are also presented. A ProtA–POP1 fusion that is associated with the RNase P/MRP RNAs (Lygerou *et al.*, 1994) was used as a positive control (lane 19).

We next investigated whether additional RNAs were associated with Sm or Sm-like proteins. For this purpose, we used two strategies. First, we labeled RNAs present in immuno-pellets with pCp. However, this did not reveal any new RNA species associated with the yeast Sm and Sm-like proteins (data not shown). It is well known that pCp labeling does not detect all RNA species (Uhlenbeck and Gumport, 1982; e.g. U6 snRNA was not detected with this assay consistent with previous findings). Therefore, we also analyzed some of the immuno-pellets for the presence of specific RNAs by primer extension. RNase P (Figure 3, middle row), RNase MRP (Figure 3, bottom row), scR1 and 5S (data not shown) RNAs were not found in the immuno-pellets. Interestingly, however, a low but significant level of pre-RNase P RNA (Lee *et al.*, 1991) was consistently co-immunoprecipitated with the Lsm2–7 proteins (Figure 3, top row) but not by Lsm1, Lsm8 or Lsm9. No RNA was found associated with the Lsm1 and Lsm9 proteins. This can not be attributed to a masking of the ProtA tag as these proteins were efficiently and nearly quantitatively recovered in immuno-pellets. In the case of Lsm1, the functionality of the tagged protein was further supported by its ability to complement the thermosensitive phenotype generated by the disruption of the corresponding gene. In an attempt to detect potential RNA species weakly associated with Lsm1 and Lsm9, we also repeated these immunoprecipitations at low salt concentration (50 mM NaCl) and analyzed the immuno-pellets for the presence of spliceosomal snRNAs by primer extension and the presence of other RNAs by pCp labeling (data not shown). No RNA associated with these two Sm-like proteins was detected. While these experiments do not rule out the association of RNA with the Lsm1 or Lsm9 proteins, they suggest that such RNA would be either present at low concentration and/or very heterogeneous in size.

Overall, our results indicate the presence of two Sm protein complexes bound to RNA in yeast (Table I). A first complex contains the canonical Sm proteins which bind to the U1, U2, U4 and U5 snRNAs. The second complex contains the Lsm2–Lsm8 proteins and is directly associated with the U6 snRNA. A slightly different complex (lacking Lsm8) is associated with the pre-RNase P RNA. At this stage, we cannot formally exclude that another RNA(s) is also associated with one or both of these complexes.

Co-precipitations confirm the presence of a Sm-like protein complex in yeast cells

The results presented above suggest the existence of a complex composed of seven Sm-like proteins in yeast. To confirm the existence of this complex, we performed a co-precipitation experiment. For this purpose, a strain expressing a single copy of the *LSM8* gene tagged at its C-terminus with a CBP tag (Stratagene) was constructed (see Material and methods). This strain was transformed with the plasmids encoding the different Lsm proteins fused to the ProtA tag. Extracts were prepared from the various transformants. The Lsm8 protein, as well as associated factors, were precipitated using the CBP tag. The presence, or absence, of the ProtA tagged protein in the precipitated material was then assayed by Western blotting. To control for the specificity of the precipitation we performed the same assay with extracts of cells carrying the ProtA tagged proteins but expressing a wild-type (i.e. non-tagged) Lsm8. The results of this experiment are depicted in Figure 4. The upper panel displays a Western blot analysis of the Lsm–ProtA fusion present in the various extracts. This demonstrates that all fusion proteins were expressed and that the level of expression was not affected by the presence, or absence, of the CBP tag on the Lsm8 protein. In the lower panel, the precipitated material was analyzed by Western blotting to detect the Lsm–ProtA fusions that were co-precipitated with Lsm8. This demonstrated that Lsm2–Lsm7 are associated with Lsm8 (Figure 4, lanes 6, 8, 10, 12, 14 and 16). The co-precipitations are specific as no signal can be detected in pellets obtained from extracts lacking the Lsm8–CBP

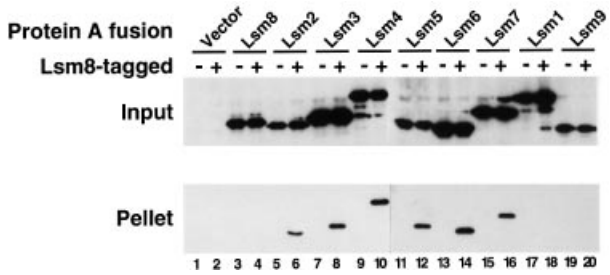

Fig. 4. The Lsm2–8 proteins are associated in a multi-protein complex. ProtA-tagged Lsm proteins were expressed in a strain carrying a Lsm8–CBP fusion (even lanes) or as negative control only the vector (odd lanes). The top row (Input) demonstrates that the signals detected are specific (compare with cells containing an empty vector, lanes 1 and 2) and that all fusions were expressed (lanes 3–20). The Lsm2–7 proteins were co-precipitated with the Lsm8 protein as revealed by Western blotting (lower row, lanes 6, 8, 10, 12, 14 and 16), while the Lsm8 (lane 4), Lsm1 (lane 18) and Lsm9 (lane 20) were not. The co-precipitations were indicative of specific association with Lsm8 as no signals were detected when the Lsm8–CBP fusion was absent (odd lanes).

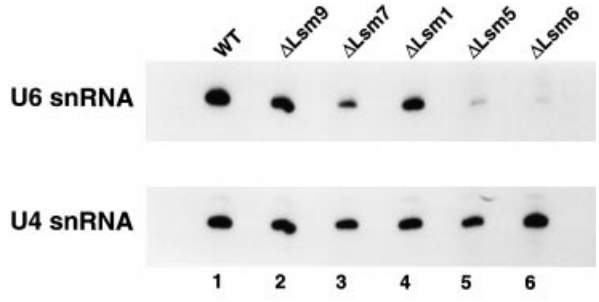

Fig. 5. The Lsm5–7 proteins are required for normal U6 snRNA accumulation. The level of U6 (top row) and, as a control, U4 (lower row) snRNAs present in wild-type (lane 1) and various Lsm mutants (lanes 2–6) were analyzed by primer extension.

fusion (Figure 4, odd lanes). Consistent with the results of our analysis of RNA associated with Lsm proteins (see Figure 2), ProtA-tagged Lsm1 and Lsm9 were not co-precipitated by the Lsm8–CBP fusion (Figure 4, lanes 18 and 20). Interestingly, ProtA-tagged Lsm8 was not co-precipitated by the Lsm8–CBP fusion (Figure 4, lane 4). This indicates that a single copy of Lsm8 is present in each complex. The canonical SmD1–ProtA fusion was also specifically co-precipitated with the Lsm8 protein in this assay, consistent with the presence of Lsm8 in the U4/U6 and U4/U6.U5 complexes (data not shown).

The results of this co-precipitation analysis confirm that the Lsm2–8 proteins associate in complex(es) from which Lsm1 and Lsm9 are excluded. It is noteworthy that this complex is evolutionarily related to the canonical Sm complex. Indeed, it contains the various Lsm proteins that show distant similarity to the canonical Sm proteins (Figure 1).

Construction and characterization of yeast Sm and Sm-like protein mutants

We next analyzed the phenotypes caused by the disruption of the Sm and Sm-like genes. Some of these mutants have been described previously and were therefore not re-constructed (Table I). The *SMB*, *SMD2*, *SMF*, *LSM1*, *LSM2*, *LSM5*, *LSM6*, *LSM7* and *LSM8* genes were disrupted through the replacement of the corresponding coding region by a selectable marker in a diploid strain (Baudin *et al.*, 1993; Puig *et al.*, 1998). Correct integration was verified by PCR and/or Southern blotting and tetrads were dissected. For the *SMB*, *SMD2*, *SMF*, *LSM2* and *LSM8* disruptions, two of the four spores grew in the majority (>85%) of the tetrads. (Two viable spores were recovered in all tetrads where all four spores germinated as determined by microscopic observation.) These spores lacked the disruption marker. The two other spores germinated and underwent a few (four or five) divisions before growth stopped (data not shown). This indicated that these genes are essential.

For the *LSM1*, *LSM5*, *LSM6* and *LSM7* strains, we recovered essentially tetrads with four viable spores. (All tetrads with full germination gave four viable spores.) This disruption indicated that the corresponding genes are not essential for vegetative growth. However, we noticed that the disrupted strains grew slower than the isogenic wild-type strains (especially for the *LSM5* disrupted strain) and that this phenotype was exacerbated at 37°C. These results are summarized in Table I.

Combined with previous analyses, our study reveals that the canonical yeast Sm proteins (SmB–G) are all essential for vegetative growth (Table I). Disruption of genes encoding Sm-like proteins directly associated with the U6 snRNA (Lsm2–8) generate, in contrast, variable phenotypes. The *LSM2–4* and *LSM8* genes are essential for vegetative growth while the *LSM5–7* are dispensable (Table I). However, these genes are required for efficient growth at 30°C and particularly 37°C. The two remaining Sm-like proteins, Lsm1 and Lsm9, are dispensable (Table I). However, Lsm1 is required for optimal vegetative growth at 30°C and is temperature sensitive (Table I).

We next analyzed the effect of the disruption of the *LSM5–7* gene on the levels of the associated U6 snRNA. *LSM1* and *LSM9* disrupted-strains were used, together with a wild-type strain, as control. Strains were grown at 37°C for 8 h, total RNAs were extracted and analyzed by primer extension. The levels of the U6 snRNA were strongly reduced in the strains harboring the *LSM5–7* disruptions (Figure 5, lanes 3, 5 and 6) while it was not affected by the *LSM1* or *LSM9* disruption (Figure 5, lanes 2 and 4). In contrast, the levels of the U4 snRNA (and of the remaining spliceosomal snRNAs) were not affected by any of the *LSM* gene disruptions (Figure 5, lanes 2–6; data not shown) demonstrating the specificity of this effect. We also analyzed the level of the pre-RNase P RNA in these strains but did not observe an effect of the *LSM1*, *LSM5–7* or *LSM9* gene disruption on this RNA species (data not shown).

Identification and preliminary characterization of metazoan Sm-like proteins

The results presented above indicate that a Sm-like protein complex is directly associated with the yeast U6 snRNA in the free U6 snRNP, and therefore also indirectly associated with the U4 and U5 snRNA present in the U4/U6 and U4/U6.U5 complexes. However, it was unclear whether this was a yeast specific feature or if a similar Sm-like complex was also present in other species. In this vein, it was intriguing that Sm-like proteins were not observed in the purified human U4/U6.U5 triple snRNP (Behrens and Lührmann, 1991) or in the *Trypanosoma* U6

(Palfi *et al.*, 1991) snRNP. On the other hand, database searches indicated that Sm-like proteins were present in a wide variety of species, including human, suggesting that they might have a widespread function (see Séraphin, 1995). To test this possibility we first recovered human cDNAs encoding Sm like proteins. This was done both by library screening or by obtaining the cDNAs corresponding to ESTs present in databases (Lennon *et al.*, 1996). Five of these cDNAs were sequenced and shown to encode full-length hLsm1, hLsm3, hLsm4, hLsm5 and hLsm6 proteins. In addition, two cDNAs encoding partial hLsm7 and hLsm8 proteins were also characterized (E.Bragado-Nilsson and B.Séraphin, unpublished). While this work was in progress, the sequence of the hLsm1 (Schweinfest *et al.*, 1997) and the hLsm2 (Fu, 1996) proteins were reported in the literature. These sequences were included in the phylogenetic analysis presented in Figure 1. This analysis indicated that the human proteins were highly similar to, and therefore probably orthologues of, the yeast proteins.

To test this hypothesis, we generated rabbit polyclonal antibodies against hLsm4 using a His-tagged recombinant protein overexpressed in *Escherichia coli* as immunogen. These antibodies were used to immunoprecipitate the hLsm4 protein and associated RNA(s) present in extracts of HeLa cells. RNAs were recovered from the immuno-pellets, fractionated on a denaturing gel and transferred to a membrane. The presence of U1, U2, U4, U5 and U6 snRNA in the pellets was then assayed by hybridization with specific probes. This analysis demonstrated that the U4–6 snRNAs were co-precipitated with the hLsm4 protein when the salt concentration was lower or equal to 400 mM (Figure 6A, lanes 3 and 5). Interestingly, only trace levels of these three snRNAs were recovered at higher salt concentration (500 mM; Figure 6A, lanes 7) suggesting that association of hLsm4 with RNA is quite labile. These signals are specific as they are not detected in precipitates obtained with pre-immune serum (Figure 6A, even lanes). Furthermore, the U1 snRNA was not co-precipitated under the same conditions also supporting the specificity of the assay. In contrast, we also observed co-precipitation of U2 snRNA at the lowest salt concentration tested (300 mM). This might reflect the presence of U2–U4/U6.U5 complexes and/or spliceosome in the extract. This interpretation is supported by the reduction of the U2 snRNA signal in immuno-pellets to background at higher salt concentration (Figure 6A, lanes 5 and 7). These results indicate that hLsm4 is associated with at least one of the U4–6 snRNAs.

We also attempted to generate antisera specifically recognizing the hLsm3 protein. However, we only obtained antibodies of low affinity. We therefore introduced the coding sequence of the hLsm3 protein in an expression vector where it was fused to a cassette encoding two IgG-binding domains of *S.aureus* ProtA (Materials and methods). This plasmid and a vector control were transiently introduced into NIH 3T3 cells. Lysates were prepared 3 days after transfection and used in immunoprecipitation experiments with rabbit IgG–agarose beads. Immuno-pellets were then analyzed for the presence of the U1–6 snRNAs following the strategy described above. We observed the presence of U4 and U6 snRNAs in the pellet of cells transfected with the tagged hLsm3 protein

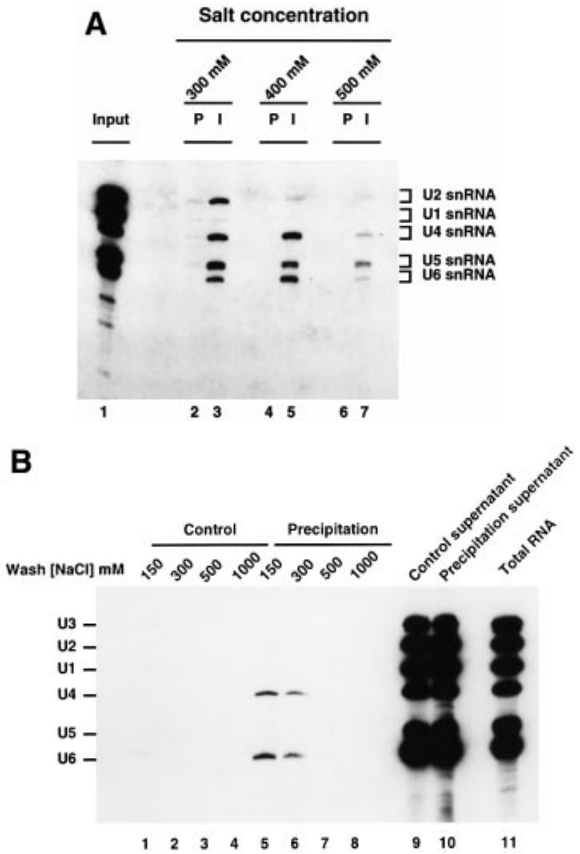

Fig. 6. Metazoan Lsm proteins are associated with U6 snRNA containing complexes. (**A**) Analysis of snRNAs co-precipitated by anti-hLsm4 antibodies. HeLa cell extracts were incubated with an anti-hLsm4 serum (I, lanes 3, 5 and 7) or the matching pre-immune serum (P, lanes 2, 4 and 6). snRNAs immunoprecipitated under various salt concentrations were detected by Northern blotting. snRNAs present in an aliquot of the starting extract are depicted in lane 1. Note the absence of significant U2 snRNA precipitation at 400 and 500 mM NaCl. The identity of the various snRNAs is indicated on the right. Signal intensities can not be used quantitatively because of difference in hybridization efficiency for the different probes. (**B**) snRNAs co-precipitated by the hLsm3–ProtA fusion transfected in NIH 3T3 cells. Cells transfected with vector alone (Control) or with a vector encoding the hLsm3–ProtA fusion (Precipitation) were lysed and incubated with IgG–agarose beads. Beads were pelleted and aliquots washed in buffers containing either 150 mM (lanes 1 and 5), 300 mM (lanes 2 and 6), 500 (lanes 3 and 7) or 1000 mM (lanes 4 and 8) NaCl. snRNAs present in the washed pellets (lanes 1–8), supernatants (lanes 9 and 10) and input lysate (lane 11) were fractionated on a denaturing gel and detected by Northern blotting. The identity of the various snRNAs is indicated on the left.

(Figure 6B, lanes 5 and 6) but not when cells were transfected with the vector alone (Figure 6B, lanes 1–4). The other U snRNAs (U1, U2, U3 and U5) were not co-precipitated, confirming that the association of hLsm3 with the U4 and/or U6 snRNAs was specific. The association of hLsm3 with the U4 and U6 snRNAs was lost when the salt concentration was raised above 300 mM salt (Figure 6B, lanes 7 and 8) suggesting that it is also quite labile. We also noted that only a low fraction of the input U4 and U6 snRNA was precipitated (Figure 6B, compare lanes 10 and 11). This might be due to low transfection and/or expression efficiency, poor assembly of the reporter protein into complexes due to the presence of the tag and/or loss of labile complex during extract preparation and immunoprecipitation.

Our results demonstrate that the hLsm3 and hLsm4 proteins are specifically associated with snRNP complexes containing the U6 snRNA. Our data do not allow us to conclude that these proteins are component of the U6 snRNP, however, this is highly likely in light of the results reported above for the homologous yeast proteins. Native gel fractionation of RNA co-immunoprecipitated with the hLsm4 protein demonstrated that it was associated with the U4/U6 complex (data not shown). Although some background level of free U6 snRNA were observed in control precipitation, a significant level of free U6 snRNA was also co-precipitated with the hLsm4 protein suggesting that this protein was also present in the free U6 snRNP (data not shown). Taken together with the identification of Sm-like protein sequences in plants as well as other fungi and animals, this strongly suggests that the Sm-like complex is widely distributed in eukaryotes.

Sm-like proteins in archaebacteria

Sm and Sm-like proteins have previously only been described in eukaryotes (Séraphin, 1995). Sensitive database searches revealed the presence of sequences related to Sm-like proteins in archaea (Figure 1). Among archaea, the genome of *Methanococcus jannaschii* was the first to be completely sequenced (Bult *et al.*, 1996). No Sm-like protein sequences were found in this organism. However, since then, three additional archaeal genomes have been completely sequenced. Sensitive database searches revealed that the genomes of *Methanobacterium thermoautotrophicum* (Smith *et al.*, 1997) and *Archaeoglobus fulgidus* (Klenk *et al.*, 1997) encode two Sm-like proteins each. In contrast, the genome of *Pyrococcus horikoshii* OT3 (Kawarabayasi *et al.*, 1998) encodes a single protein from this family. A phylogenetic analysis indicates that three of the archaeal Sm-like proteins are related while the two other ones are more divergent (Figure 1). Interestingly, the three related proteins belong to three different species suggesting the widespread distribution of an orthologue. In contrast, the two other, poorly related, archaeal proteins have probably diverged in different lineages through independent gene duplication events. Because archaea have been proposed to be related to the ancestor of the eukaryotic nuclear genome (Puhler *et al.*, 1989), the finding of Sm-like proteins in these organisms suggests that an Sm-like protein was already present in the latest common ancestor. Furthermore, the observation that archaea encode one or two Sm-like proteins support the proposal that Sm and Sm-like proteins originate from a single precursor.

Discussion

We have analyzed the yeast Sm protein family. This revealed the presence of a protein complex, the Sm-like complex, directly associated with the U6 and pre-RNase P RNAs that are present in the nucleus. Database searches identified Sm-like proteins in other eukaryotes and archaea. Two human proteins were characterized further and found to associate with U6 snRNA containing complexes. These results have important implications for our understanding of the evolution of the Sm protein family. Our analysis also provides a complete transversal overview of the functions of all proteins from this family in a model eukaryote and suggests that several of these functions are evolutionarily conserved.

Our analysis reveals that Sm and Sm-like proteins are present in archaea suggesting that they were present in a common ancestor shared by these organisms and eukaryotes. This is probably not unexpected given that several other eukaryotic nuclear proteins clearly evolved from archaea-related precursors (Puhler *et al.*, 1989). Sm-like proteins are only found encoded by three of the four complete archaeal genome sequences that are currently available (see Results) suggesting that it was recently deleted from the genome of *M.jannaschii*. At this stage, it appears that archaeal genomes code for only one or two proteins of this family. Eukaryotic proteins are more numerous with up to 16 members in yeast. Most of these yeast proteins have counterparts in metazoan and plants (Figure 1; Séraphin, 1995; B.Séraphin, unpublished results). Phylogenetic and functional analysis indicate that Sm and a subset of the Sm-like proteins form two related subgroups in eukaryotes. Each subgroup contains seven members: on the one hand, the seven canonical Sm proteins and on the other hand, the Lsm2–8 proteins. Proteins from these two subgroups associate in distinct subcomplexes. Interestingly, six proteins from the canonical Sm complex have a closely related counterpart in the Lsm2–8 complex. For the seventh protein, SmB, sequence similarity to Lsm8 is less apparent (dashes in Figure 1). Indeed, it has been suggested on the basis of pairwise alignments that yeast SmB is related to yeast Lsm1 rather than yeast Lsm8 (Fromont *et al.*, 1997) while human Lsm1 has been proposed to be more related to human SmG (Schweinfest *et al.*, 1997). These contradictory proposals are not supported by phylogenetic analysis (Figure 1). These original observations may have been biased by the limited number of sequences analyzed and the use of pairwise alignments of complete protein sequences rather than multiple alignment of the conserved Sm domain. Our analysis suggests that the weak similarity detected between Lsm8 and SmB is functionally relevant. This strongly suggests that the canonical Sm complex and the Lsm2–8 Sm-like protein complex have a common origin. According to this data, evolution of the Sm and Sm-like protein family would have proceeded in several steps. The ancestral Sm-like gene, related to the one currently present in archaea, would have first been duplicated several times to generate seven different Sm-like proteins associating in a complex. In a second step, the genes encoding these seven Sm-like proteins would have all been duplicated, possibly through polyploidization, leading to the appearance of two distinct, but nevertheless related, subgroups corresponding to the current Lsm2–8 and canonical Sm complexes of yeast, human cells and other eukaryotes. Other duplication events would have given rise to other Sm-like proteins like Lsm1 and Lsm9. Further studies will be required to establish whether the canonical Sm complex preceded the Lsm2–8 complex or vice versa.

Our study represents the analysis of all members from a protein family from a model eukaryote. Such comprehensive studies are now possible with the availability of complete genome sequences. Our results demonstrate that the Sm and Sm-like protein family can be subdivided into smaller groups containing functionally related proteins.

First, seven yeast proteins from this family are true Sm proteins. Interestingly, these seven proteins appear to be essential for vegetative growth in yeast (Table I; this work; Rymond, 1993; Roy *et al.*, 1995; Séraphin, 1995; Fromont *et al.*, 1997), indicating their importance. These proteins are associated with the U1, U2, U4 and U5 snRNAs and indirectly to U6 snRNA present in the U4/U6 snRNP (Figure 2B). Immunoprecipitation followed by pCp labeling of associated RNA did not reveal other RNAs associated with these proteins. Even though some RNAs may not be detected by this technique, this suggests that yeast has a lower number of Sm-associated RNAs than higher eukaryotes (Yu *et al.*, 1996). The yeast Sm proteins have significant similarities to their human counterparts even though they are not fully interchangeable with them (Rymond *et al.*, 1993). Furthermore, the yeast Sm proteins appear to interact with each other with the same specificity as their mammalian counterparts (Raker *et al.*, 1996; Fury *et al.*, 1997; Camasses *et al.*, 1998). Given these strong similarities, it is likely that the yeast Sm complex shares additional functional properties with its metazoan counterpart.

The second subset of yeast Sm and Sm-like proteins contains the Lsm2–8 proteins. These proteins co-precipitate the U4, U5 and U6 snRNAs as previously shown for two of them (SmX4/Lsm3 and USS1/Lsm4). They directly associate with the U6 snRNA present in the free U6 snRNP. Given the relative ratio of the different RNAs recovered, it is likely that the co-precipitation of U4 and U5 snRNAs is indirect and results from their presence in the U4/U6 and U4/U6.U5 particles. Similar results for Lsm8 were reported by others while this work was in progress (Pannone *et al.*, 1998). In addition, the Lsm2–7 proteins were found to be associated with the pre-RNase P RNA but not the mature RNase P RNA. We could not detect association with the RNase MRP RNA, the 5S rRNA and the abundant cytoplasmic scR1 RNA (Felici *et al.*, 1989). In addition, pCp labeling also failed to reveal additional RNAs associated with the Lsm2–8 complex. However, it remains possible that RNAs poorly labeled by this method or RNAs of low abundance also associate with these proteins. It is unclear why both the pre-RNase P RNA and the U6 snRNA are associated with these proteins since these two RNAs do not share obvious specific primary or secondary structural elements. While both of these RNAs are transcribed by pol III, this seems insufficient to explain their association with these proteins as this is not the case for other pol III transcripts (5S rRNA, scR1 RNA). In any case, this result demonstrates that proteins of the Sm and Sm-like family are not restricted to spliceosomal factors. Additional studies will be required to determine whether Sm and Sm-like proteins interact directly with RNA and, if so, the sequence and structure requirements for specific interaction.

Our data suggest that these proteins associate with RNA as a complex, rather than individually. Indeed, co-precipitation experiments demonstrate that the Lsm2–7 proteins can associate with the Lsm8 proteins. In addition, nearly all the U6 snRNA is associated with the Lsm3 protein in a cell expressing a Lsm3–ProtA fusion as the only source of Lsm3 activity (data not shown), indicating that the various Lsm proteins can not be associated with different subpools of U6 snRNA. It is interesting that both the U6 associated Sm-like complex and the canonical Sm complex contain seven subunits. While the exact stocheiometry of the protein in these complexes remains to be determined, our co-precipitation experiment indicates that there is a single copy of the Lsm8 protein per complex. The analysis of protein–protein interactions inside the canonical Sm complex (Raker *et al.*, 1996; Fury *et al.*, 1997; Camasses *et al.*, 1998) and structural data (Kambach *et al.*, 1999) are also compatible with the presence of a single copy of each monomer per complex. Some Sm-like proteins were found to interact weakly with canonical Sm proteins in quantitative two-hybrid analysis (Camasses *et al.*, 1998). This was interpreted as resulting from weak non-specific interaction occurring between Sm domains. Our immunoprecipitation data strongly suggest the existence of two distinct complexes containing either Sm or Sm-like proteins demonstrating the need for additional data to validate protein–protein interactions suggested on the basis of two-hybrid assays. Altogether, our data and the conservation of a structural domain in Sm and Sm-like proteins suggest that these proteins form a pseudo-regular structure containing seven repeats of the basic unit.

In contrast to the results obtained for canonical Sm proteins, disruptions of the *LSM2–LSM8* genes produced variable phenotypes. Previous studies had demonstrated that the *LSM3* (*SMX4*; Séraphin, 1995) and *LSM4* (*USS1*; Cooper *et al.*, 1995) genes were essential for viability. The same is true for the *LSM2* and *LSM8* proteins. On the other hand, the *LSM5–7* genes are not essential. However, they are required for normal vegetative growth as their disruptions lead to slow growth especially at elevated temperatures. The level of U6 snRNA was significantly reduced in *LSM5–7* disrupted cells. This suggests that the Lsm2–8 complex is functionally important to stabilize this RNA by interacting with it, although we cannot rule out at this stage that these proteins are also involved in U6 snRNA transcription. A function of Sm-like proteins in U6 snRNA stabilization is consistent with the observations that canonical Sm proteins are required for normal U1, U2, U4 and U5 snRNA accumulation (Rymond, 1993; Roy *et al.*, 1995; Bordonné and Tarassov, 1996; Camasses *et al.*, 1998). In contrast, the Lsm2–8 proteins were not required for stability of the RNase P RNA precursor (data not shown). This might result from the transient nature of this species and/or its association with additional proteins as shown for Pop1 (Lygerou *et al.*, 1994).

The two remaining yeast Sm-like proteins, Lsm1 and Lsm9, were not essential and were not found associated with the Lsm8 protein or with RNA under a variety of conditions. At least for Lsm1, the lack of co-precipitation cannot result from the non-functionality of the tagged protein as this construct suppressed the growth phenotype associated with a *LSM1* disruption (see Results). Consistent with a lack of association of these two proteins with spliceosomal RNAs, disruption of the corresponding genes did not affect U snRNA and pre-RNase P levels (Figure 5; data not shown). Disruption of the *LSM1* gene, however, impaired cell growth and produced a thermosensitive phenotype, suggesting that it is involved in some non-essential cellular process. While this work was in progress, an lsm1 mutation was identified. This mutation suppresses a *PAB1* gene disruption and affects

mRNA degradation through inhibition of the decapping step (Boeck *et al.*, 1998). This suggests that *LSM1* is involved in an RNA degradation pathway. The function of Lsm9 is still unclear. However, *lsm9* mutants were reported to be defective in the maintenance of yeast killer virus-like particles (Toh-E and Sahashi, 1985).

The results that we obtained in yeast are likely to be of general significance. Indeed, we identified and sequenced several cDNAs encoding Sm-like proteins from human cells. The conservation of the Lsm1–6 proteins in human and other species suggests the widespread existence of an Sm-like complex in eukaryotes. Immunoprecipitation experiments demonstrated that the hLsm3 and hLsm4 proteins associate with the U4 and U6 snRNAs. hLsm4 aso co-precipitated the U5 snRNA. It is not clear why hLsm3 did not co-precipitate the human U5 snRNA. However, it is possible that the protein expressed from the transfected DNA was poorly expressed and therefore inefficiently incorporated in the U4/U6.U5 complexes (compare signals in supernatant and pellet in Figure 6). Surprisingly, the Lsm proteins were also not identified in the purified human U4/U6.U5 triple snRNP (Behrens and Lührmann, 1991). As for the yeast SmB, it is possible that these proteins were lost during biochemical fractionation. Indeed, our results demonstrate that association of Lsm proteins with U snRNAs is salt sensitive. In any case, our results strongly suggest that, like their yeast counterparts, human Lsm proteins form a complex associated with the U6 snRNA. We conclude that the Lsm2–8 function is conserved. A Lsm1 homologue is also found in humans (Figure 1). It is likely that the function of this protein is also conserved between yeast and humans. In contrast, no Lsm9 related proteins have been found in human cells yet. It is therefore possible that the Lsm9 function only evolved recently in yeast cells. Further studies will be required to determine whether other RNAs, beside U4, U5 and U6, are associated with hLsm proteins. In particular, it will be interesting to find out whether these proteins also associate with the U6-ATAC snRNA involved in the splicing of U12-dependent introns.

The function of archaeal Sm-like proteins is currently unknown. No RNA related to spliceosomal snRNAs have been reported in these organisms. However they are likely to contain small RNAs involved in various processes (e.g. RNase P RNA). Further studies will be required to establish whether archaeal Sm-like proteins are associated with any of them. Such observation could provide some clues about the origin of spliceosomal snRNAs.

Materials and methods

Biocomputing

Database searches were performed using the SearchWise (Birney *et al.*, 1996), WUBLAST (W.Gish, unpublished) and GCG (Devereux *et al.*, 1984) softwares. CLUSTAL_X (Thompson *et al.*, 1997) was used for multiple sequence alignments that were further edited by hand. Treetool (Maidak *et al.*, 1997) was used for tree building. The names of the open reading frames (Goffeau *et al.*, 1997) encoding yeast Sm and Sm-like proteins are as follows (with former names in the general SmX nomenclature in parenthesis; Séraphin, 1995): SmB/YER029C, SmD1/YGR074W, SmD2/YLR275W, SmD3/YLR147C, SmE/YOR159C, SmF/YPR182W (SmX3), SmG/YFL018WA (SmX2), Lsm1/YJL124C (SmX8), Lsm2/YBL026W (SmX5), Lsm3/YLR438C-A (SmX4), Lsm4/YER112W (SmX6), Lsm5/YER146W (SmX12), Lsm6/YDR378C (SmX14), Lsm7/YNL147W (SmX7), Lsm8/YJR022W (SmX9) and Lsm9/YCR020CA (SmX1). References for the human Sm protein sequences can be found in Séraphin (1995) and Hermann *et al.* (1995). Human Sm-like protein sequences are from Schweinfest *et al.* (1997), Fu (1996) and from this publication (see below). Identifiers for the archaeal proteins are as follows: Arch1/AF0875, Arch2/AF0362, Meth1/MTH649, Meth2/MTH1440 and Pyro1/PHS042.

Yeast plasmids and strains

The putative promoter and coding sequence regions of each Sm and Sm-like protein from yeast were amplified by PCR and inserted together with a ProtA cassette in vector pRS425 (Christianson *et al.*, 1992). The SmG/SmX2, SmF/SmX3, Lsm3/SmX4, Lsm9/SmX1 constructs built using this strategy has been described previously (Séraphin, 1995). Plasmids carrying the SmB(pBS1361), SmD1(pBS747), SmD2(pBS824), SmD3(pBS1297), SmE(pBS1360), Lsm1(pBS1298), Lsm2(pBS826), Lsm4(pBS1303), Lsm5(pBS959), Lsm6(pBS1296), Lsm7(pBS957) and Lsm8(pBS1302) fusions were built during this study. The absence of PCR-generated mutation in the Sm or Sm-like coding sequence of these constructs was verified by sequencing. These constructs were introduced into yeast strain MGD343-13D as described previously (Séraphin, 1995).

To analyze protein co-precipitation, the Lsm8 coding sequence from plasmid pBS1302 was inserted together with a C-terminal CBP tag (Stratagene) in a vector pRS424 (Christianson *et al.*, 1992), generating plasmid pBS1611. This construct was introduced in the diploid BSY699 [*MAT***a**/*MAT*α, *ade2/ade2, arg4/arg4, leu2-3,112/leu2-3,112, trp1-289/trp1-289, ura3-52/ura3-52, LSM8/lsm8::URA3* (*Kluveromyces lactis*)] strain carrying a disrupted *LSM8* gene (see below). After sporulation and dissection, a strain, BSY754, carrying the disrupted genomic *lsm8* allele complemented by the plasmid-borne CBP tagged Lsm8 protein was recovered. This strain was transformed with the plasmids encoding the various yeast Lsm–ProtA fusions described above.

Gene disruption was performed using PCR fragments harboring the sequence flanking the target gene and the marker (Baudin *et al.*, 1993; Puig *et al.*, 1998). The *SMB, SMD2, SMF, LSM5, LSM6, LSM7* or *LSM8* gene were disrupted in strain in BSY320 (*MAT***a**/*MAT*α, *ade2/ade2, arg4/arg4, leu2-3,112/leu2-3,112, trp1-289/trp1-29, ura3-52/ura3-52*) with the *K.lactis URA3* marker (Langle and Jacobs, 1995; Puig *et al.*, 1998) giving diploid strains BSY722, BSY729, BSY748, BSY667, BSY716, BSY719, BSY731 and BSY699, respectively. The second diploid strain carrying a disrupted *SMF* gene, BSY795, was obtained by crossing a spore carrying a disrupted smf::*URA3* (*K.lactis*) allele complemented by a plasmidic copy with a strain of the opposite mating type. *LSM2* was disrupted in strain BSY320 with the *TRP1* marker (Puig *et al.*, 1998) giving the diploid strain BSY541. LSM3 was disrupted in the diploid strain BMA38 (Baudin *et al.*, 1993) with the *HIS3* marker giving strain BSY486.

Correct integration was verified by Southern blotting and/or PCR. At least nine complete tetrads were dissected for each construct. Spore germination was assessed by microscopic observation. For *LSM8*, we ruled out the possibility that the lethal phenotype resulted from the disruption of the open reading frame overlapping with the Lsm8 coding sequence on the opposite strand (Goffeau *et al.*, 1997) by showing that a plasmid expressing Lsm8 but not the complementary open reading frame rescued the lethality associated with the disruption (data not shown). For *SMF* disruption with *URA3*, dissection of several independent transformants gave mostly tetrads with a single viable spore even though we also obtained a few tetrads with two or no viable spores. All viable spores were Ura$^-$ suggesting that the SMF disruption was lethal but that another recessive lethal mutation was also segregating. To ascertain whether the SMF disruption truly causes a lethal phenotype, we dissected tetrads from the same diploid strains transformed with a plasmid encoding the SmF–ProtA fusion. The plasmid rescued the lethality of associated with the *URA3* marker but we obtained a maximum of two viable spores consistent with the segregation of another recessive lethal marker. One of these spores was then crossed with a wild-type strain of the opposite mating type. The resulting diploid was cured from the plasmid and sporulated. After dissection, two of the four spores grew in the majority (>85%) of the tetrads and all viable spores turned out to be Ura$^-$, confirming that SMF is essential. Viable haploid strains carrying the disrupted *LSM1* [BSY673 (*MAT***a**) and BSY674 (*MAT*α)], *LSM5* [BSY717 (*MAT***a**) and BSY718 (*MAT*α)], *LSM6* [BSY72 (*MAT***a**) and BSY721 (*MAT*α)] and *LSM7* [BSY732 (*MAT***a**) and BSY733 (*MAT*α)] genes were recovered. Transformation of the thermosensitive BSY673 strain with the pBS1302 plasmid restored growth at 37°C.

Human Lsm proteins

cDNA clones were obtained from the IMAGE EST project (Lennon *et al.*, 1996) or by screening cDNA libraries. Sequences obtained on

both strands for cDNA encoding hLsm1 (accession No. AJ238094), hLsm3 (accession No. AJ238095), hLsm4 (accession No. A3238096), hLsm5 (accession No. A3238097) and hLsm6 (accession No. AJ238098) are deposited in the EMBL/DDBJ/GenBank databases (the corresponding entry numbers are indicated in parentheses). Partial cDNA clones encoding the putative hLsm7 and hLsm8 protein were also sequenced (data not shown).

The hLsm4 coding sequence was inserted in vector pRSET B (Invitrogen). The resulting plasmid, pBS1169, was used to generate recombinant hLsm4 protein carrying a His tag at its N-terminus. The protein purified from *E.coli* lysates was used to immunize rabbits to generate antibodies.

The hLsm3-coding sequence was inserted in vector pSFFV6 (Chen *et al.*, 1993) together with a ProtA cassette (Séraphin, 1995). The ProtA tag was fused at the C-terminus of the hLsm3 protein. This construct, and the pSffvport vector control, were transfected in mouse NIH 3T3 cells using the calcium phosphate method (Bachetti and Graham, 1977).

Extracts, precipitation and RNA analyses

Yeast extracts, immunoprecipitation of tagged yeast protein, RNA extraction, primer extension and native gel analysis were performed as described previously (Lygerou *et al.*, 1994; Séraphin, 1995). Co-imunoprecipitation of yeast proteins 50 µl of extract was diluted 1:1 in dialysis buffer (Séraphin and Rosbash, 1989). This material was bound to 15 µl calmodulin–agarose beads (Stratagene) in 300 µl of IPP150Ca buffer (10 mM Tris–Cl pH 8.0, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 3 mM $CaCl_2$, 0.1% NP-40, 10 mM β-mercapto-ethanol) for 2 h at 4°C. Beads were pelleted and washed four times for 20 min with IPP150-Ca buffer and once for 10 min with IPP150-O buffer (10 mM Tris–Cl pH 8.0, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 0.1% NP-40, 10 mM β-mercaptoethanol). Complexes were eluted in 50 µl of IPP150-EGTA buffer (10 mM Tris–Cl pH 8.0, 150 mM NaCl, 1 mM magnesium acetate, 1 mM imidazole, 3 mM EGTA, 0.1% NP-40, 10 mM β-mercaptoethanol) and transferred to a new tube. Half a volume of 2× loading buffer was added and samples were loaded on a 10% denaturing polyacrylamide gel (Sambrook *et al.*, 1989). Following blotting to nitrocellulose, ProtA fusions were detected using PAP (Sigma) and Renaissance (NEN).

Transfected mouse cells were collected and stored frozen at –80°C. Cell pellets were lysed by thawing in ice-cold IPP150 buffer (10 mM Tris–Cl pH 8.0, 150 mM NaCl, 0.1% NP-40, 0.1% NaN_3). Immunoprecipitations of ProtA-tagged hLsm3 protein were essentially performed as described previously for tagged yeast proteins (Séraphin, 1995) except that wash buffers adjusted at different salt concentration were used (Figure 6B). For hLsm4, immune or pre-immune sera were incubated with ProtA–agarose beads. After removal of unbound sera components, these antibody-covered beads were then incubated with HeLa extracts (Dignam *et al.*, 1983; kind gift of J.Valcárcel) and immunoprecipitation were performed as described above except that buffers adjusted to various salt concentrations were used. RNA was extracted from the pellets, input extracts and/or supernatant fractions. RNA was fractionated by electrophoresis of denaturing acrylamide gels and detected by Northern blotting (Sambrook *et al.*, 1989).

Acknowledgements

We thank E.Bouveret, F.Caspary, P.Fortes, S.Kuersten, P.Lopez, M.Luukkonen, O.Puig, R.Ramirez-Morales, G.Rigaut, B.Rutz, D.Suck, J.Valcárcel and I.Mattaj for discussion and comments on the manuscript. We are grateful to J.C.Jauniaux, M.Jacquet and C.Herbert for providing yeast sequences before publication, MIPS for sequence searches, the IMAGE consortium and RZPD for cDNA clones, F.Mauxion for gift of the vector pSFFV6 and J.Valcárcel for HeLa extracts, respectively. The support of the EMBL services was highly appreciated. B.S. is on leave from the CNRS.

References

Bachetti,S. and Graham,F.L. (1977) Transfer of the gene for thymidine kinase to thymidine kinase-deficient human cells by purified herpex simplex viral DNA. *Proc. Natl Acad. Sci. USA*, **74**, 1590–1594.

Baudin,A., Ozier,K.O., Denouel,A., Lacroute,F. and Cullin,C. (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **21**, 3329–3330.

Behrens,S.E. and Lührmann,R. (1991) Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev.*, **5**, 1439–1452.

Birney,E., Thompson,J.D. and Gibson,T.J. (1996) PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res.*, **24**, 2730–2739.

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.

Bordonné,R. and Tarassov,I. (1996) The yeast *SME1* gene encodes the homologue of the human E core protein. *Gene*, **176**, 111–117.

Brahms,H., Raker,V.A., van,V.W. and Lührmann,R. (1997) A major, novel systemic lupus erythematosus autoantibody class recognizes the E, F and G Sm snRNP proteins as an E–F–G complex but not in their denatured states. *Arthritis Rheum.*, **40**, 672–682.

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.

Bult,C.J. *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1073.

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

Chen,H.J., Remmler,J., Delaney,J.C., Messner,D.J. and Lobel,P. (1993) Mutational analysis of the cation-independent mannose 6-phosphate/insulin-like growth factor II receptor. A consensus casein kinase II site followed by 2 leucines near the carboxyl terminus is important for intracellular targeting of lysosomal enzymes. *J. Biol. Chem.*, **268**, 22338–22346.

Christianson,T.W., Sikorski,R.S., Dante,M., Shero,J.H. and Hieter,P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.

Chu,J.-L. and Elkon,K.B. (1991) The small nuclear ribonucleoproteins, SmB and B′, are products of a single gene. *Gene*, **97**, 311–312.

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

Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.

Felici,F., Cesareni,G. and Hughes,J.M. (1989) The most abundant small cytoplasmic RNA of *Saccharomyces cerevisiae* has an important function required for normal cell growth. *Mol. Cell. Biol.*, **9**, 3260–3268.

Fischer,U., Heinrich,J., van Zee,K., Fanning,E. and Lührmann,R. (1994) Nuclear transport of U1 snRNP in somatic cells: differences in signal requirement compared with *Xenopus laevis* oocytes. *J. Cell Biol.*, **125**, 971–980.

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,R.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.

Fu,Y.H. (1996) Identification of a novel protein, DMAP, which interacts with the myotonic dystrophy protein kinase and shows strong homology to D1 snRNP. *Genetica*, **97**, 117–125.

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.

Goffeau,A. *et al.* (1997) The yeast genome directory. *Nature Suppl.*, **387**, 1–105.

Gottschalk,A. *et al.* (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA*, **4**, 374–393.

Guthrie,C. and Patterson,B. (1988) Spliceosomal snRNAs. *Annu. Rev. Genet.*, **22**, 387–419.

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

Huber,J., Cronshagen,U., Kadokura,M., Marshallsay,C., Wada,T., Sekine,M. and Lührmann,R. (1998) Snurportin1, an m^3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.*, **17**, 4114–4126.

Izaurralde,E. and Mattaj,W.I. (1992) Transport of RNA between nucleus and cytoplasm. *Semin. Cell Biol.*, **3**, 279–288.

Jarmolowski,A. and Mattaj,I.W. (1993) The determinants for Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and non-identical. *EMBO J.*, **12**, 223–232.

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.

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.

Langle,R.F. and Jacobs,E. (1995) A method for performing precise alterations in the yeast genome using a recycable selectable marker. *Nucleic Acids Res.*, **23**, 3079–3081.

Lee,J.Y., Evans,C.F. and Engelke,D.R. (1991) Expression of RNase P RNA in *Saccharomyces cerevisiae* is controlled by an unusual RNA polymerase III promoter. *Proc. Natl Acad. Sci. USA*, **88**, 6986–6990.

Lee,S.I., Murthy,S.C.S., Trimble,J.J., Desrosiers,R.C. and Steitz,J.A. (1988) Four novel U RNAs are encoded by a herpesvirus. *Cell*, **54**, 599–607.

Lennon,G., Auffray,C., Polymeropoulos,M. and Soares,M.B. (1996) The IMAGE consortium: an integrated molecular analysis of genomes and their expression. *Genomics*, **33**, 151–152.

Lerner,E.A., Lerner,M.R., Janeway,C.A.,Jr and Steitz,J.A. (1981) Monoclonal antibodies to nucleic acid-containing cellular constituents: probes for molecular biology and autoimmune disease. *Proc. Natl Acad. Sci. USA*, **78**, 2737–2741.

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.

Lührmann,R., Kastner,B. and Bach,M. (1990) Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochem. Biophys. Acta*, **1087**, 265–292.

Lygerou,Z., Mitchell,P., Petfalski,E., Séraphin,B. and Tollervey,D. (1994) The *POP1* gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes Dev.*, **8**, 1423–1433.

Maidak,B.L., Olsen,G.L., Larsen,N., Overbeek,R., McCaughey,M.J. and Woese,C.R. (1997) The RDP (ribosomal database project). *Nucleic Acids Res.*, **25**, 109–110.

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

Mattaj,I.W., Tollervey,D. and Séraphin,B. (1993) Small nuclear RNAs in messenger RNA and ribosomal RNA processing. *FASEB J.*, **7**, 47–53.

McAllister,G., Roby,S.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.

Nelissen,R.L.H., 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.

Palfi,Z., Günzl,A., Cross,M. and Bindereif,A. (1991) Affinity purification of *Trypanosoma brucei* small nuclear ribonucleoproteins reveals common and specific protein components. *Proc. Natl Acad. Sci. USA*, **88**, 9097–9101.

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.

Puhler,G., Leffers,H., Gropp,F., Palm,P., Klenk,H.P., Lottspeich,F., Garrett,R.A. and Zillig,W. (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl Acad. Sci. USA*, **86**, 4569–4573.

Puig,O., Rutz,B., Luukkonen,B.G., Kandels,L.S., Bragado,N.E. and Séraphin,B. (1998) New constructs and strategies for efficient PCR-based gene manipulations in yeast. *Yeast*, **14**, 1139–1146.

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.

Rokeach,L.A., Jannatipour,M., Haselby,J.A. and Hoch,S.O. (1992) Mapping of the immunoreactive domains of a small nuclear ribonucleoprotein-associated Sm-D autoantigen. *Clin. Immunol. Immunopathol.*, **65**, 315–324.

Roy,J., Zheng,B., Rymond,B. and Woolford,J. (1995) Structurally related but functionally distinct yeast Sm D core small nuclear ribonucleoprotein particle proteins. *Mol. Cell. Biol.*, **15**, 445–455.

Rymond,B.C. (1993) Convergent transcripts of the yeast PRP38-SMD1 locus encode two essential splicing factors, including the D1 core polypeptide of small nuclear ribonucleoprotein particles. *Proc. Natl Acad. Sci. USA*, **90**, 848–852.

Rymond,B.C., Rokeach,L.A. and Hoch,S.O. (1993) Human snRNP polypeptide D1 promotes pre-mRNA splicing in yeast and defines nonessential yeast Smd1p sequences. *Nucleic Acids Res.*, **21**, 3501–3505.

Sambrook,J., Fritsch,E.J. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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.

Séraphin,B. and Rosbash,M. (1989) Identification of functional U1 snRNA–pre-mRNA complexes committed to spliceosomal assembly and splicing. *Cell*, **59**, 349–358.

Smith,C.M. and Steitz,J.A. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669–672.

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

Tan,E.M. (1989) Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv. Immunol.*, **44**, 93–151.

Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.

Toh-E,A. and Sahashi,Y. (1985) The PET18 locus of *Saccharomyces cerevisiae*: A complex locus containing multiple genes. *Yeast*, **1**, 159–171.

Tollervey,D. and Mattaj,I.W. (1987) Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.*, **6**, 469–476.

Uhlenbeck,O.C. and Gumport,R.I. (1982) In Boyer,P.D. (ed.), *The Enzymes*. Vol. 15B. Academic Press, Inc., New York, NY, pp. 31–60.

Yu,Y.T., Tarn,W.Y., Yario,T.A. and Steitz,J.A. (1996) More Sm snRNAs from vertebrate cells. *Exp. Cell Res.*, **229**, 276–281.

Received December 2, 1998; revised April 13, 1999; accepted April 29, 1999