THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 278, No. 19, Issue of May 9, pp. 17291–17298, 2003

Printed in U.S.A.

Homomeric Ring Assemblies of Eukaryotic Sm Proteins Have Affinity for Both RNA and DNA

CRYSTAL STRUCTURE OF AN OLIGOMERIC COMPLEX OF YEAST SmF*

Received for publication, November 20, 2002, and in revised form, March 4, 2003
Published, JBC Papers in Press, March 4, 2003, DOI 10.1074/jbc.M211826200

Brett M. Collins‡, Liza Cubeddu§¶, Nishen Naidoo§, Stephen J. Harrop‖, Geoff D. Kornfeld‡‡, Ian W. Dawes‡‡, Paul M. G. Curmi‖**, and Bridget C. Mabbutt§ §§**

*From the ‡Cambridge Institute for Medical Research, University of Cambridge, Department of Clinical Biochemistry, Hills Road, Cambridge CB2 2XY, United Kingdom, and the §Department of Chemistry, Macquarie University, Sydney, New South Wales 2109, ‖School of Physics and ‡‡School of Biochemistry and Molecular Genetics, University of New South Wales, Sydney, New South Wales 2052, and **Centre for Immunology, St Vincent's Hospital, Darlinghurst, New South Wales 2010, Australia*

Sm and Sm-like proteins are key components of small ribonucleoproteins involved in many RNA and DNA processing pathways. In eukaryotes, these complexes contain seven unique Sm or Sm-like (Lsm) proteins assembled as hetero-heptameric rings, whereas in Archaea and bacteria six or seven-membered rings are made from only a single polypeptide chain. Here we show that single Sm and Lsm proteins from yeast also have the capacity to assemble into homo-oligomeric rings. Formation of homo-oligomers by the spliceosomal small nuclear ribonucleoprotein components SmE and SmF preclude hetero-interactions vital to formation of functional small nuclear RNP complexes *in vivo*. To better understand these unusual complexes, we have determined the crystal structure of the homomeric assembly of the spliceosomal protein SmF. Like its archaeal/bacterial homologs, the SmF complex forms a homomeric ring but in an entirely novel arrangement whereby two heptameric rings form a co-axially stacked dimer via interactions mediated by the variable loops of the individual SmF protein chains. Furthermore, we demonstrate that the homomeric assemblies of yeast Sm and Lsm proteins are capable of binding not only to oligo(U) RNA but, in the case of SmF, also to oligo(dT) single-stranded DNA.

Sm and Sm-like (Lsm)[1] proteins are core components of ribonucleoprotein (RNP) complexes involved in many nucleic acid processing events within the eukaryotic cell nucleus. The most highly characterized Sm/Lsm-containing RNPs are those involved in pre-mRNA splicing, the U1, U2, U4/U6, and U5 small nuclear RNPs (snRNPs) (1), whereas others are known to be important for telomere replication (2), *trans*-splicing (3), and mRNA degradation (4, 5).

All Sm/Lsm proteins contain two regions of conserved sequence, termed the "Sm motifs," separated by a segment of variable length and composition (6–8). They possess a common structure (the Sm domain) consisting of a five-stranded antiparallel β-sheet preceded by a short α-helix at the N terminus (9–12). The variable region of sequence comprises the flexible loop L4 connecting strands β3 and β4. The highly bent β-sheet forms a curved shape that encompasses a hydrophobic core extending to both edges of the molecule. This structure dictates that Sm/Lsm proteins have a preference for forming closed ring oligomers of seven subunits, whereby each subunit interacts with its neighbors through a combination of β-strand pairing and extensive hydrophobic contacts. In the cases of the bacterial Sm homolog Hfq and Sm2 from *Archaeoglobus fulgidus*, homomeric hexamer assemblies of the fold have been observed (13, 14).

In eukaryotes, it is a ring of seven different but specific Sm/Lsm proteins that binds to small nuclear RNA at a poly(U) sequence (the "Sm-binding motif") to form the core of each snRNP complex. For example, the spliceosomal snRNPs contain one copy each of the seven proteins SmB, SmD1, SmD2, SmD3, SmE, SmF, and SmG (15), assemblies of which are seen as ring structures when examined by electron microscopy (16, 17). The RNA binding interaction of the Sm/Lsm heptamers is governed by stacking of bases with conserved residues lining the inner surface on one face of the ring (12, 18, 19). The dominating features on the opposite ring face are the variable loops (loop L4) from individual Sm subunits.

Lsm genes have been identified in prokaryotic archaeal species, although only 1–3 Sm-like proteins are usually encoded per genome, compared with the 16 or more found for eukaryotic organisms (10, 20–22). In line with this observation, these single proteins form highly stable homo-oligomers with comparable RNA-binding affinities to the more complex mixed heptamers of the eukaryotic snRNPs (10–12, 23). Very recently it has been determined that the bacterial protein Hfq is a homolog of the Sm family (24, 25), mediating RNA-RNA interactions as a homo-hexameric assembly.

In this study we have focused on the functional organization of recombinant versions of Sm and Lsm proteins from the simple eukaryote *Saccharomyces cerevisiae*. We show that the three yeast proteins SmE, SmF, and Lsm3 are able to form stable homo-oligomeric ring structures in solution, in a similar

* This work was supported by grants from the Australian Research Council and Macquarie University. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1N9R and 1N9S) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ Present address: Centre for Biomolecular Sciences, University of St. Andrews, St. Andrews, KY16 9ST, UK.

§§ To whom correspondence should be addressed. Tel.: 61-29850-8282; Fax: 61-29850-8313; E-mail: bridget.mabbutt@mq.edu.au.

[1] The abbreviations used are: Lsm, Sm-like; GST, glutathione *S*-transferase; MtLsmα or β, *M. thermoautotrophicum* Lsmα or β protein; r.m.s.d., root mean square deviation; RNP, ribonucleoprotein; snRNP, small nuclear ribonucleoprotein; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; Ni-NTA, nickel-nitrilotriacetic acid.

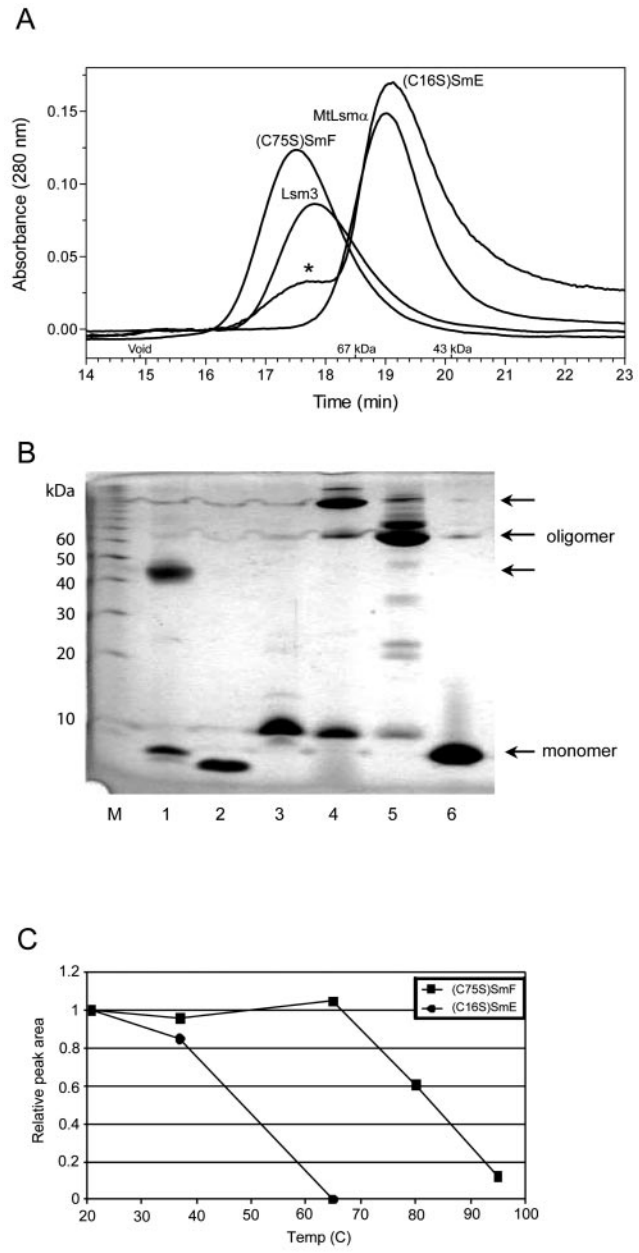

FIG. 1. **Gel filtration and electrophoresis of recombinant Sm/Lsm proteins.** *A*, gel filtration of Sm/Lsm proteins on a Superdex 75 column in 10 mM Tris (pH 8.0), 200 mM NaCl. (C75S)SmF and Lsm3 appear to form complexes twice the size of (C16S)SmE and the archaeal protein MtLsmα which forms a heptameric ring (10). Dimerization of the (C16S)SmE complex is sometimes observed in solution (*). *B*, silver-stained SDS-PAGE shows samples boiled for 5 min: *lane M*, 10-kDa marker; *lane 1*, MtLsmα; *lane 2*, MtLsmβ; *lane 3*, (C16S)SmE; *lane 4*, (C75S)SmF; *lane 5*, Lsm3; and *lane 6*, Lsm9. *C*, thermostability of (C16S)SmE and (C75S)SmF homo-oligomers measured from peak areas in gel-filtration traces following 15 min of incubation in a water bath. The SmF complex is resistant to denaturation to 65 °C.

manner to their archaeal and bacterial Lsm counterparts. SmE and SmF, thought to form a direct pairwise interaction within spliceosomal snRNPs *in vivo*, do not associate with each other when in these stable homo-oligomeric states. The ring-forming Sm/Lsm proteins are shown to bind poly(U) RNA. Surprisingly, in contrast to Sm/Lsm complexes described previously, the SmF complex actively binds single-stranded DNA as well as poly(U) RNA. The structure of SmF determined by x-ray crystallography reveals an Sm protein assembly with a novel higher order arrangement of 14 monomers.

EXPERIMENTAL PROCEDURES

Cloning of Yeast Sm and Lsm Genes—Genes encoding yeast proteins SmE, SmF, Lsm9, and Lsm3 were isolated from *S. cerevisiae* genomic DNA by PCR. Primers were designed to incorporate *Nde*I and *Eco*RI restriction sites at the 5′ and 3′ ends of the genes, respectively. Genes were inserted into the expression plasmid pETMCSIII, which relies on transcription by T7 RNA polymerase and results in a hexa-His fusion at the N terminus of expressed proteins (26). The plasmid was transformed into the BL21(DE3)/pLysS *Escherichia coli* strain for expression.

The single cysteine mutants (C16S)SmE and (C75S)SmF were made from pETMCSIII-derived plasmids using the Stratagene QuikChange™ Site-directed Mutagenesis kit according to the manufacturer's instructions. The (C75S)SmF gene was also inserted into the pGEX-4T-2 plasmid (Amersham Biosciences) for alternative expression as a glutathione *S*-transferase (GST) fusion protein. The gene was cloned from the pETMCSIII-derived plasmid by PCR using a 5′ primer that allowed insertion of the gene into pGEX-4T-2 using the *Bam*HI and *Eco*RI restriction enzymes. This plasmid was transformed into the BL21 *E. coli* strain for protein expression.

Protein Expression and Purification—Cells were grown at 37 °C in LB broth (containing appropriate antibiotics), and protein expression was induced by addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside to the culture medium at A_{595} of 0.5–0.6. Harvested cells were lysed by a French press in lysis buffer (20 mM Tris (pH 8.5), 50 mM NaH_2PO_4, 100 mM NaCl) containing 0.5% (w/v) Tween 20® (Calbiochem), 10 μg/ml RNase A, and protease inhibitor mixture (Sigma). The lysate was clarified by centrifugation (30 min at 15,000 × *g*, 4 °C), and expressed proteins were purified on Ni-NTA or glutathione-Sepharose affinity media as appropriate. The yeast protein Lsm9 was refolded from inclusion bodies by solubilizing in urea (8 M), glycerol (10%), imidazole (10 mM), binding to Ni-NTA-agarose and elution with gradual dilutions of urea. All proteins were further purified by gel-filtration chromatography using 10 mM Tris (pH 8.0), 200 mM NaCl as running buffer. Recombinant production of the *Methanobacterium thermoautotrophicum* Lsm protein (MtLsmα) has been described previously (10). For RNA binding assays, the site-specific mutants MtLsmα(R72L) and MtLsmα(R72L/N48A) were also prepared.

Molecular Weight Estimation—Proteins were analyzed by SDS-PAGE using the Tricine buffering system (27) or in solution by analytical gel filtration with Superose® 12 and Superdex 75 (Amersham Biosciences). Gel-filtration columns were run in 10 mM Tris (pH 8.0), 200 mM NaCl at a flow rate of 60 μl/min. Elution and void volumes were calibrated using standard globular proteins and blue dextran (28).

Electron Microscopy—Protein samples of (C75S)SmF, (C16S)SmE, Lsm3, and MtLsmα were diluted to 10–100 μg/ml in 10 mM Tris (pH 8.0), 100 mM NaCl. Copper grids (400 μm mesh) were coated with an ultra-thin layer of carbon (~10 nm). Preparations were stained for 10 s with 5.0% uranyl acetate, 2.0% acetic acid after binding the protein for 5 s. Grids were examined using a Philips CM10 transmission electron microscope, and electron micrographs were recorded at a magnification of ×73,000–145,000.

Protein Affinity Studies—Cultures of *E. coli* strains expressing bait GST-(C75S)SmF were mixed with cultures expressing prey His-tagged (C16S)SmE. The mixed cells were lysed with a French press in lysis buffer (see above). A 500-μl aliquot of the clarified lysate was incubated with 50 μl of glutathione-agarose and bound protein pelleted. The pellet was washed several times with buffer, and bound and unbound fractions were examined by SDS-PAGE.

Biosensor interaction studies were performed with a BIAcore2000 surface plasmon resonance instrument. His-tagged (C16S)SmE (75 μl, 200 nM) introduced to a Ni-NTA chip at a flow rate of 5 μl/min resulted in an increase of 3,300 response units above base line. Binding of (C75S)SmF (derived from the GST fusion protein by thrombin proteolysis) to His-tagged (C16S)SmE was tested by injecting 15 μl of 15 nM protein (*i.e.* 150 nM) at a flow rate of 5 μl/min. Similarly, binding of GST-(C75S)SmF fusion protein was tested by injecting 15 μl of protein at a concentration of 590 nM.

Crystallization of SmF—Following gel filtration, His-tagged versions of SmF and (C75S)SmF in Tris/NaCl (plus 10 mM dithiothreitol for SmF) were concentrated to 20 mg/ml. Crystals of both proteins were grown by sitting drop vapor diffusion in a reservoir containing 24% polyethylene glycol 3350, 0.1 M Tris (pH 8.5), and 0.1 M sodium acetate. Diffraction quality crystals measuring 2 × 0.5 × 0.5 mm formed within several weeks. From these conditions, two tetragonal crystal forms have been identified, $P4_122$ and $P4_32_12$.

Data Collection and Structure Determination—Data from the $P4_122$

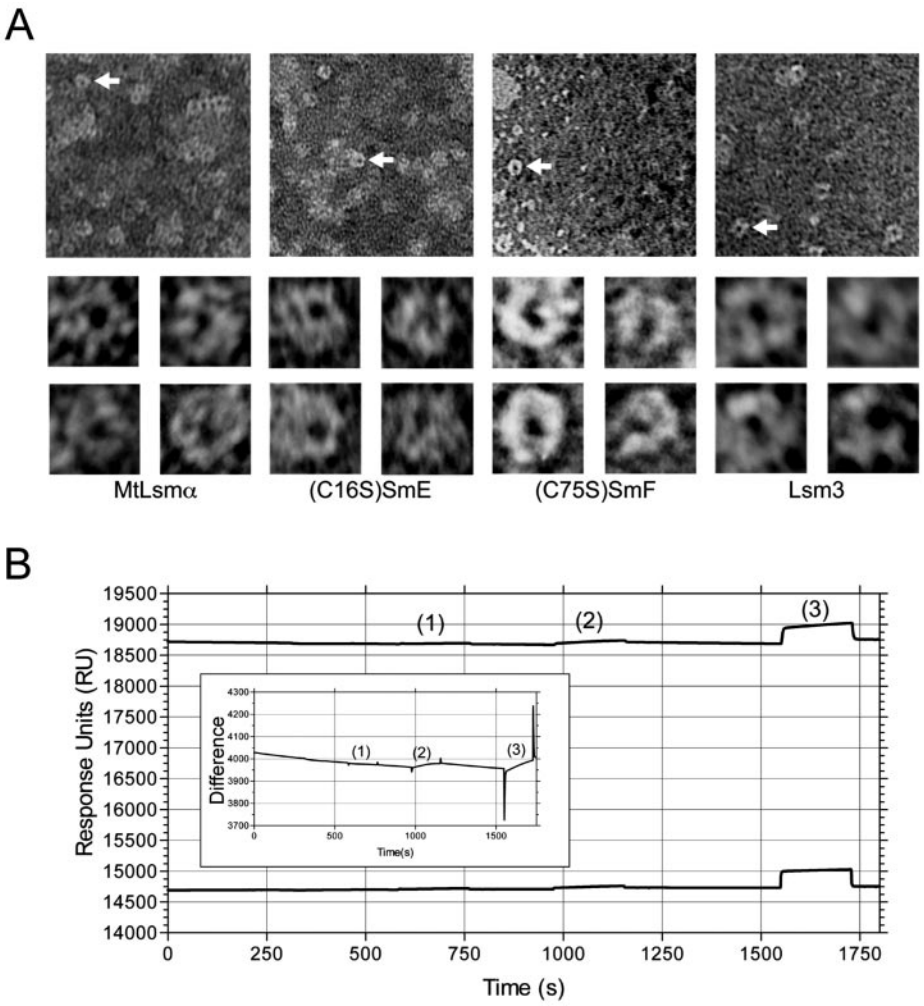

FIG. 2. *A*, electron micrographs of Sm/Lsm homo-oligomers. Purified proteins were bound to carbon-coated grids and negatively stained with uranyl acetate. The *upper panels* show selected fields 100 nm in width for samples of MtLsmα, (C16S)SmE, (C75S)SmF, and Lsm3. *White arrows* indicate a representative ring-shaped particle. The *lower panels* (10 nm in width) show galleries of four typical particles. *B*, biosensor response for Ni-NTA chip with bound His-tagged (C16S)SmE (upper sensorgram) probed with 15 μl of 15 nM (C75S)SmF (*1*), 150 nM (C75S)SmF (*2*), and 590 nM GST-(C75S)SmF (*3*). Lower sensorgram shows reference channel response with no chelated protein; *inset* shows the net difference.

crystal form of SmF was collected at beamline 14-4 at the European Synchrotron Radiation Facility. Data to 2.8 Å resolution was collected at 100 K after cryoprotecting the crystal in mother liquor plus 10% glycerol. Images were integrated using MOSFLM (29) and scaled using CCP4 programs (30). As a starting point for molecular replacement, the MtLsmα heptameric structure (Protein Data Bank code 1I81) was converted to polyserine and truncated by removal of the N-terminal helix of each protein chain. Molecular replacement was performed with AMORE (31) using data from 15 to 4 Å. Very clear rotationally related solutions constituted the top six results, with a seventh rotationally related solution appearing at the eighth position. The top solution (correlation coefficient 37.0, *R* factor 48.7) was subjected to rigid body and simulated annealing refinement in CNS (32), resulting in a model with R_{cryst} = 40% and R_{free} = 41% and a clearly interpretable electron density map. The model was rebuilt in the O program (33) and subjected to TLS refinement with REFMAC5 (34) to final R_{cryst} and R_{free} values of 25.4 and 26.7%, respectively (Table I). There is one heptameric ring in the asymmetric unit, but visual inspection revealed a striking co-axially stacked interaction with a symmetry-related heptamer. This involves contacts between all seven of the variable loops on one face of the structure with all seven of the loops from the symmetry-related molecule (see text for discussion).

Data from cryo-cooled crystals of the (C75S)SmF protein in space group $P4_32_12$ was collected on beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory using an ADSC Quantum4 CCD detector with x-rays at 1.0 Å wavelength. As calculations indicated two heptameric rings per asymmetric unit in this crystal form (35), a molecular replacement model was constructed by pairing the refined SmF heptameric structure (above) with its strongly interacting symmetry equivalent. Molecular replacement with AMORE showed that this dimeric arrangement of the heptameric rings constituted the asymmetric unit of the $P4_32_12$ crystal form. The structure of 14 subunits was rebuilt using the O program (33) and refined with REFMAC5 to yield final R_{cryst} and R_{free} values of 29.2 and 29.7%, respectively (Table I).

The coordinates and structure factors for both the $P4_122$ and the $P4_32_12$ structures of yeast SmF have been deposited with the Protein Data Bank (accession codes 1N9R and 1N9S).

RNA/DNA Binding Assays—RNA/DNA oligonucleotides were ^{32}P-labeled at their 5′ ends, purified on a 10% polyacrylamide/7 M urea gel, and recovered by *n*-butanol precipitation (36). Labeled oligonucleotides were heat-denatured (2 min, 95 °C) and immediately cooled on ice before addition of reaction mixture. RNA binding mixture (15 μl) containing 20 fmol of RNA/DNA oligonucleotides, 5 μg of protein, and 1.5 μl of binding buffer (20 mM Tris/HCl (pH 8.0), 70 mM KCl, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 0.1 mM EDTA, 7% (w/v) glycerol, 4 mM dithiothreitol, and 20 units of ribonuclease inhibitor) were incubated at 22 °C for 30 min. Oligonucleotide species were fractionated on a 5% native polyacrylamide gel in Tris borate/EDTA buffer (with 3 mM β-mercaptoethanol) and detected by autoradiography.

RESULTS

Recombinant SmE, SmF, and Lsm3 Form Homo-oligomeric Rings—In order to characterize the yeast Sm and Lsm proteins, we have produced several as recombinant molecules in *E. coli*. SmE, SmF, and Lsm3 when expressed in bacteria were found to be soluble and easily purified in high yields by affinity chromatography. As there was some evidence of a proportion of covalently linked dimers within samples of His-tagged SmE and SmF on SDS-PAGE, non-native disulfide bond formation was avoided by mutating the single non-conserved cysteine residues of each sequence to serine, resulting in His-tagged (C16S)SmE and (C75S)SmF. These site-specific mutants yielded single bands by electrophoresis, yet otherwise appeared to possess an identical fold to proteins of native sequence as judged by gel-filtration, electron microscopy, and crystallogra-

TABLE I
Statistics of data collection and structure refinement

Values in parentheses apply to the high resolution shell.

Data collection		
Protein	SmF	(C75S)SmF
Space group	$P4_122$	$P4_32_12$
X-ray wavelength (Å)	0.9393	1.0
Unit cell dimensions	a = 79.9 Å; b = 79.9 Å; c = 253.2 Å; $\alpha = \beta = \gamma = 90°$	a = 105.6 Å; b = 105.6 Å; c = 236.6 Å; $\alpha = \beta = \gamma = 90°$
No. heptameric rings per asymmetric unit	1	2
Resolution (Å)	79.0–2.8 (2.95–2.80)	95.0–3.5 (3.69–3.50)
R_{merge} [a]	0.066 (0.390)	0.064 (0.353)
R_{meas} [b]	0.072 (0.415)	0.069 (0.381)
$\langle\langle I\rangle/\sigma\langle I\rangle\rangle$	19.6 (4.6)	23.5 (6.0)
Completeness (%)	99.9 (99.9)	99.3 (99.3)
Multiplicity	6.8 (7.1)	6.7 (6.8)
Wilson plot B (Å^2)	84.2	88.9
Refinement		
Resolution (Å)	20.0–2.8	20.0–3.5
R_{cryst}/R_{free} [c]	0.254/0.267	0.292/0.297
$\langle B\rangle$(Å^2)	50.3	44.8
No. reflections (no. in R_{free})	19,819/1075	16,292/876
No. atoms	3837	7887
r.m.s.d. bond lengths (Å)	0.023	0.023
r.m.s.d. bond angles (°)	2.1	2.3
Ramachandran violations [d]	0	0
Residues in final model	*A*, 19–86; *B*, 18–86; *C*, 19–86; *D*, 19–86; *E*, 19–86; *F*, 18–86; *G*, 19–86	*A*, 19–86; *B*, 19–86; *C*, 13–86; *D*, 16–86; *E*, 19–86; *F*, 17–86; *G*, 17–86; *H*, 16–86; *I*, 17–86; *J*, 15–86; *K*, 18–86; *L*, 13–86; *M*, 17–86; *N*, 18–86

[a] $R_{merge} = \Sigma\Sigma_i|I_h - I_{hi}|/\Sigma\Sigma_i I_h$ *where* I_h is the mean intensity of reflection h.
[b] $R_{meas} = \Sigma\sqrt{(n/n-1)}\Sigma|I_h - I_{hi}|/\Sigma\Sigma_i I_h$, the multiplicity weighted R_{merge} (48).
[c] $R = \Sigma(F_P - F_{calc})/\Sigma F_P$.
[d] From PROCHECK (49).

phy (see below). The molecular integrity of the recombinant products was verified by electro-spray mass spectrometry.

Gel-filtration chromatograms were used to ascertain the oligomeric composition of the various recombinant proteins in solution. By including in each run samples of archaeal MtLsmα characterized by us previously (10) as an oligomeric complex of seven subunits, we were able to compare the relative sizes of the homo-oligomers formed by (C16S)SmE, (C75S)SmF, and Lsm3 (Fig. 1). Of the group of proteins described here, only Lsm9 and MtLsmβ samples elute with an apparent molecular weight corresponding to a monomer (9 and 8 kDa, respectively). The majority of the (C16S)SmE sample elutes as a molecule with an apparent molecular mass of 55–60 kDa and at the same volume as the MtLsmα heptamer. The chromatograms of both (C16S)SmE and SmE often contained an additional shoulder corresponding to a size of ~110 kDa, indicating some dimerization of the major SmE species (Fig. 1*A*). The proteins (C75S)SmF and Lsm3, however, each elute as single species with apparent molecular masses of 100–110 and 95–105 kDa, respectively, *i.e.* they appear to be twice the molecular size of the (C16S)SmE and MtLsmα oligomers. The large (C75S)SmF and Lsm3 complexes are particularly stable, as shown by their electrophoretic properties (Fig. 1*B*); the large (C75S)SmF oligomer withstands 2% SDS and boiling for 5 min, and this treatment merely halves the size of the Lsm3 complex. In contrast, (C16S)SmE is more readily converted to a monomeric species. These results suggest that Lsm3 and (C75S)SmF associate in a similar dimeric arrangement of two ~55-kDa complexes in solution, with the Lsm3 dimerization being somewhat less stable. Note that the (C75S)SmF oligomer is significantly thermostable, resisting unfolding after heating to 65 °C, a point at which the (C16S)SmE complex is dissociated (Fig. 1*C*).

Transmission electron micrographs of the complexes of (C16S)SmE, (C75S)SmF, and Lsm3 stained with uranyl acetate were examined to determine the organization of the oligomeric assemblies. In all cases, ring structures are revealed (Fig. 2*A*). Each protein complex is the same dimension and is similar in size to the structure of the MtLsmα heptameric ring, showing a diameter of ~8 nm and a central area of stain ~2 nm wide. As the rings of (C75S)SmF and Lsm3 appear to be equal in dimension to those of MtLsmα and (C16S)SmE, despite having twice their effective molecular size in solution, this suggests that the former proteins are organized as co-axially stacked double rings. The ultra-structures of these yeast homo-oligomeric assemblies are highly similar to the hetero-heptameric Sm/Lsm protein complexes seen at the core of snRNPs from HeLa cells (16, 17, 20, 37).

Complexes of SmE and SmF Do Not Interact in Vitro—SmE and SmF are known to interact with each other, both within RNA-free hexameric complexes with SmG and in intact splicing snRNPs (38, 39). Two-hybrid studies and immunoprecipitations have also detected a direct pairwise interaction between these two core Sm molecules (39–41). The specific pairing of these proteins is a key event in the formation of a functional spliceosomal Sm complex (9, 39). We therefore examined whether the recombinant molecules are capable of a similar pairing *in vitro*. With (C16S)SmE attached to the sensor chip surface via its N-terminal hexa-His affinity tag, no binding event could be detected upon introduction of either (C75S)SmF or the heavier variant GST-(C75S)SmF (Fig. 2*B*). Similarly, no mass change was detected when the biosensor experiment was configured for binding of GST-(C16S)SmE to His-tagged (C75S)SmF on the Ni-NTA chip.

We were also unable to detect any co-complex of GST-(C75S)SmF and (C16S)SmE in solution phase following the mixing of expression host cell lysate and pull-down with glutathione-agarose (not shown). The lack of observable interaction between the recombinant forms of SmE and SmF does not exclude very weak interactions between the proteins but certainly rules out the possibility *in vitro* of the 1:1 complex

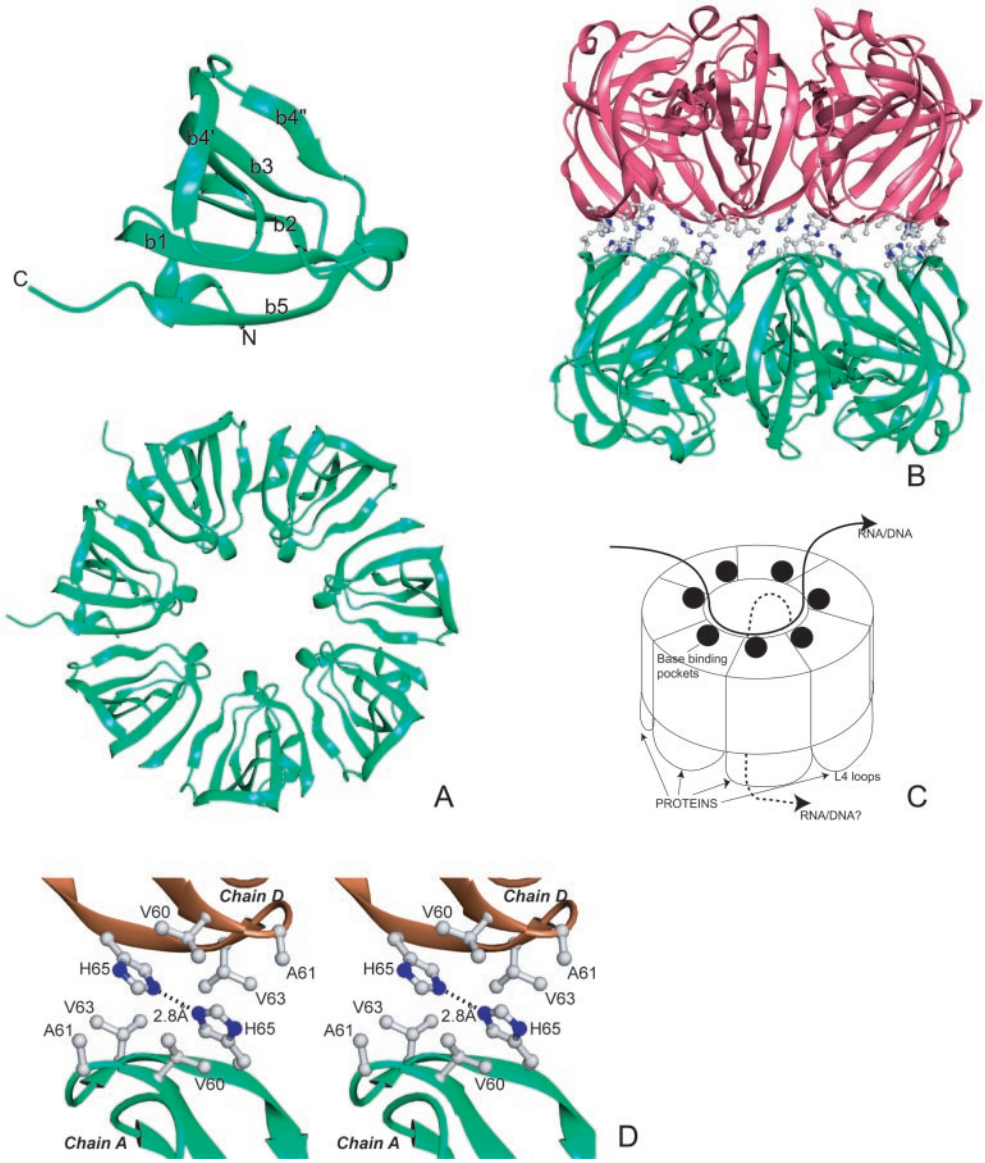

FIG. 3. **Ribbon structures of the SmF homo-oligomeric complex.** *A*, structure of the SmF assembly in the $P4_122$ crystal form. A single SmF subunit is shown at the *top*. The heptameric ring forms extensive contacts with a symmetry-related heptamer in an identical arrangement to that seen in the $P4_32_12$ crystal form. *B*, two heptameric rings are shown in *magenta* and *green*, and residues in the L4 loop that form the interface between the two rings are shown as *ball* and *stick models*. *C*, predicted interactions of Sm/Lsm complexes with nucleic acids and proteins. RNA may pass across one face of the ring (*heavy line*) or through the central hole (*dashed line*). We propose that one face of a single ring docks with conserved nucleic acid Sm-binding sequences, whereas the variable loops on the opposite loop L4 face are ideally located for interacting with associated proteins. *D*, stereo-diagram showing a close-up of the interface between two SmF heptamers (shown in *tan* and *green*). The interaction is governed by well ordered contacts between the variable L4 loops of individual SmF chains.

formation thought to occur in native snRNPs. It appears that the formation of such a mixed complex is inhibited by the highly organized and stable oligomeric states formed separately by each protein component.

Crystal Structure of SmF—We have determined the crystallographic structures of the homo-oligomeric complexes of both SmF and the mutant form (C75S)SmF. Both crystallize under identical conditions, although we have observed that (C75S)SmF crystallizes in two different forms ($P4_32_12$ and $P4_122$) while so far observing only the $P4_122$ form for native SmF. We report here the structure of SmF in the $P4_122$ crystal form and (C75S)SmF in the $P4_32_12$ form. The statistics of data collection and structure refinement are given in Table I. The two structures were determined sequentially by molecular replacement using the previously solved MtLsmα heptamer as a starting model to first solve the structure of the SmF $P4_122$ crystal form.

SmF from *S. cerevisiae* forms a homo-heptameric ring (Fig. 3), where each subunit binds to its neighbor via β-strand pairing and hydrophobic interactions. Each SmF polypeptide chain essentially adopts an identical fold to that defined for its human and archaeal Sm and Lsm counterparts (9–12), a curved β-sheet composed of five anti-parallel strands. These strands comprise residues His^{25}–Leu^{31} (β1), Thr^{36}–Asp^{46} (β2), Asn^{50}–Val^{60} (β3), Val^{63}–Arg^{74} (β4), and Tyr^{80}–Leu^{84} (β5) of the SmF sequence, respectively. In our data for SmF we do not observe any significant density for the N-terminal α-helix common to the other Sm/Lsm structures. Most subunits in the two SmF crystals structures show a single-turn distorted helix near the N terminus, which in some subunits is preceded by an extended chain, particularly in the $P4_32_12$ structure.

With stacking of β-strands of seven neighboring subunits into a ring, two circular faces are formed, which we term the loop L4 face and the RNA-binding face according to their most

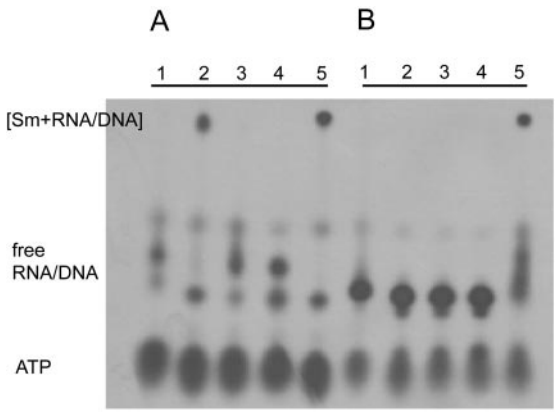

FIG. 4. **Autoradiograph demonstrating *in vitro* interaction of ^{32}P-labeled oligo(U) RNA and oligo(T) DNA with Sm/Lsm complexes.** *A*, specific binding of radiolabeled A_2U_5 RNA assayed by gel shift and visualized on a native PAGE gel. RNA (20 fmol) was incubated alone (*lane 1*) or with 5 μg of purified protein as follows: MtLsmα (*lane 2*), (R72L)MtLsmα mutant (*lane 3*), (R72L/N48A)MtLsmα mutant (*lane 4*), and (C75S)SmF (*lane 5*). A shift in RNA mobility, indicating RNA-protein complex formation, is noted for MtLsmα and SmF. Mutants of MtLsmα altered the strictly conserved RNA binding pocket. *B*, gel shift assays of radiolabeled A_2T_5 DNA (*lane 1*) and following incubation with proteins (*lanes 2–5*) as described for *A*. A_2T_5 produces a band shift for (C75S)SmF at the same intensity as that seen for the A_2U_5/(C75S)SmF interaction.

dominant structural features (10). In the $P4_122$ crystal form of SmF, there is one heptameric ring per asymmetric unit (see "Experimental Procedures"), but this heptamer makes extensive contacts with a symmetry-related complex along one face, where the two heptamers are stacked co-axially. This interaction is governed by binding between residues from the loop L4 faces of each heptamer, and we believe that this dimeric arrangement corresponds to the complex that we observe in solution and in electron micrographs. In strong support of this, our crystal structure of (C75S)SmF in the $P4_32_12$ form shows two heptamers per asymmetric unit and reveals an identical arrangement to the SmF symmetry-related dimer (Fig. 3*B*). The observation of an identical 14 subunit complex in a different crystal form strongly implies that the dimeric interaction is not simply the result of crystal packing.

Examination of the dimeric interface reveals the unique interactions responsible for the dimerization of SmF heptamers into an oligomeric complex of 14 subunits, which results in the burial of 1,030 Å^2 of accessible surface per heptamer. The stacking of two rings is a result of hydrophobic interactions between Val^{60} and Val^{63} residues on one L4 loop and His^{65} on the opposite L4 face, as well as a specific hydrogen bond between the two histidine residues (His^{65}) close to opposing L4-loops (Fig. 3*D*). It is conjectured that this interaction would be destabilized at low pH due to the protonation of each imidazole group.

RNA and DNA Binding by Oligomeric Sm/Lsm Protein Complexes—To investigate the binding properties of the prepared Sm/Lsm protein complexes, we carried out electrophoretic mobility shift assays with ^{32}P-labeled synthetic oligonucleotides containing the specific U-rich sequences to which these proteins bind *in vivo*. The human U snRNP core assembly has been shown to have the same high affinity for AAUUUUU as it does for the complete Sm site sequence AAUUUUUGA, and results in a stable complex (16). For our study, the RNA sequences U_5 and A_2U_5 were screened, as well as the corresponding DNA oligonucleotide, A_2T_5.

Fig. 4 shows that a strong band shift occurs for A_2U_5 in the presence of either (C75S)SmF or MtLsmα. The position of the gel band indicates that the RNA-protein binding interaction involves an intact Sm oligomer in both cases. Gel shifts of similar size are obtained for the binding assay in the presence of U_5. The specificity of the interaction is seen when key residues of the RNA-binding site in MtLsmα, Arg^{72} and Asn^{48} (10), are altered and result in complete abolition of binding of A_2U_5. Oligomeric complexes of (C16S)SmE and Lsm3 both showed some binding to U_5, but to a lesser degree than heptameric MtLsmα and (C75S)SmF (gels not shown). The suggestion by Toro *et al.* (14) that an aromatic residue at the midpoint of the L3 loop (*e.g.* Tyr^{48} in yeast SmF, Phe^{46} in SmE, and His^{46} in MtLsmα) promotes strong RNA binding is not well supported by these results, as we observed a range of RNA interactions for Sm proteins containing this sequence feature. Rather, our data (summarized in Table II) indicate that it is the ability to preform ring assemblies that promotes specific interactions of Sm proteins with RNA. Lsm9 and MtLsmβ are not able to form assemblies in solution, and these recombinant versions give no response to RNA. DNA gel shifts were performed with our complete set of Sm proteins and yielded the surprising result that (C75S)SmF is capable of interacting with DNA.

DISCUSSION

Several previous reports show that homotypic interactions between eukaryotic Sm/Lsm proteins are able to occur, at least *in vitro*. Using two-hybrid methods, others have identified weak interactions between yeast SmE and itself (40, 41). GST interaction studies have shown that the trypanosomal SmB protein is able to interact with itself (42). However, to the best of our knowledge, this is the first report of Sm/Lsm proteins from a eukaryotic organism forming stable homo-oligomers with a ring-like morphology. This overall structure is similar to that of the hetero-heptamers at the heart of known Sm/Lsm-containing snRNPs. Furthermore, these complexes not only bind specifically to identical poly(U) RNA sequences as native Sm/Lsm assemblies, they also bind to single-stranded poly(dT) DNA. This binding has not been observed previously.

The heptamer unit of the SmF 14-mer is very similar to previously observed archaeal Lsm heptamers (Table III), generally within an R.M.S.D. of 1.4–1.6 Å for all structures (*M. thermoautotrophicum* 1I81 (10) and 1JRI; *A. fulgidus* 1I5L and 1I4K (12)). The exception is the Lsm from *Pyrobaculum aerophilum* (11), with an R.M.S.D. of 2.0–2.4 Å from all other heptamers. This difference appears to be due to the packing of two aromatic residues (Phe/Tyr) on the interface of the archaeal Lsm proteins and SmF (Tyr^{38} and Tyr^{80} in SmF). In *P. aerophilum* Tyr^{38} is replaced by an isoleucine residue, which alters the monomer packing geometry.

Distortion of the N-terminal helix in SmF is also seen in human SmB (9), although the latter structure lacks an Sm-Sm interface as it is not part of a heptameric complex in the crystal. Two factors appear to contribute to the disruption of the helix in SmF: the presence of a proline at residue 17, and a steric clash between Phe^{18} and a residue on strand β4, Phe^{72}. The Pro^{17} residue is unique to yeast (and human) SmF and SmG and is preceded by Pro^{15}, which is conserved in numerous Sm protein sequences. The steric clash is likely to be absent in the heterogeneous spliceosomal core complex, since SmF would directly interface with SmE which instead contains a leucine residue at the position corresponding to Phe^{72}. We note that the human SmD1-SmD2 interface (9) also has the packing problem seen for SmF between residues 18 and 72, with SmD1 Phe^{6} (equivalent to Phe^{18}) packing against SmD2 Phe^{100} (equivalent to Phe^{72}), yet the structure manages to maintain the full N-terminal helix.

Our structure of the SmF protein complex reveals an unexpected higher order arrangement of Sm/Lsm proteins. Although the homo-heptameric ring formation is identical to that

TABLE II
Purification and characterization of Sm/Lsm proteins

Protein	Predicted mass	ES-MS	SDS-PAGE[a]	Gel filtration[b]	Possible no. subunits	RNA binding[c]	Loop L3 sequence
			kDa				
Lsm9	10.679	10.679	9	14	1	–	DAQMN
Lsm3	10.984	10.984	(10) 60 (70)(100)	95–105	10–14	+	DSHCN
MtLsmα	9.173	9.174	(9)45	60	7[d]	++	DLHMN
(C16S)SmE	11.313	11.313	11	55–60 (110)	5–7 (10–14)	+	DEFMN
(C76S)SmF	10.597	10.598	(10)(60) 100	100–110	14[d]	++	DNYFN
MtLsmβ	8.073	8.038	8	8–9	1	–	DNYLN

[a] Numbers in parentheses indicate less abundant species; sample was boiled (5 min) prior to gel loading.
[b] Size range by gel filtration in 10 mM Tris (pH 8.0), 200 mM NaCl.
[c] Strength of binding according to relative intensity of autoradiograph band.
[d] Number of subunits defined by crystallography.

TABLE III
Least squares alignment of Sm/Lsm heptamer structures

R.m.s.d. calculated in Å by LSQMAN in the program O (33). Numbers in parentheses indicate the number of C_α atoms used in the superposition. *Mth*, *M. thermoautotrophicum; Af*, *A. fulgidus; Pa*, *P. aerophilum.*

PDB code	1I81 *Mth*	1JRI *Mth*	1I4K *Af*	1I5L *Af*	1I8F *Pa*
1I81					
1JRI	0.78 (494)				
1I4K	1.23 (486)	1.20 (487)			
1I5L	1.05 (482)	0.93 (495)	0.99 (496)		
1I8F	2.05 (423)	2.07 (427)	2.22 (455)	2.21 (428)	
SmF	1.59 (446)	1.36 (460)	1.62 (449)	1.49 (467)	2.40 (336)

observed for archaeal Lsm proteins, the dimerization of two rings is novel. Our data also suggest that Lsm3, and to some extent SmE, are both able to organize similarly. The dimerization of the rings is mediated by the variable L4 loop segments that comprise one face of each heptameric complex, opposite to the ring face containing the determinants for RNA binding (10, 12). Sequence alignments of the Sm/Lsm protein family show loop L4 to be a region of variable length and amino acid composition between more structurally conserved protein segments (6–8). Thus this conformationally flexible portion of the Sm fold appears to be a key determinant for the protein-protein interactions we observe here. The Sm-like bacterial homolog Hfq, which is also found organized as a co-axial dimeric ring assembly in the crystalline state (each of six subunits), lacks this loop L4 sequence feature. The Hfq dimer interface instead utilizes contacts from a different portion of the Sm fold to that seen for SmF, instead involving the edge of strand β2 via Phe^{60} (13).

In vivo there are two major types of interaction governed by Sm/Lsm protein complexes; a core RNA-binding interaction that is common to all of the multitude of Sm/Lsm complexes so far identified, and interactions with many different protein factors associated with the diverse snRNP complexes. Coupled with previous structures of Sm/Lsm assemblies (10–12, 14, 18), the structure of SmF demonstrates an inherent dichotomy between the two faces of the heptameric ring; one side governs nucleic acid binding, and the other side is capable of protein-protein interactions. This suggests to us that the various Sm/Lsm complexes may have a polarity in their functional interactions within the cell, whereby RNA binds in conserved grooves on one side of the ring and proteins bind to the variable loops on the other (Fig. 3*C*).

In this study we have identified at least three yeast Sm and Lsm proteins that can form homo-oligomeric rings. The fact that Sm and Lsm proteins with different functions are able to self-associate in a common manner shows that this is a quite general phenomenon. Of course, one particularly important question arises, *i.e.* what is the role of these eukaryotic homo-assemblies *in vivo*? It has been rigorously shown (15) that yeast spliceosomal snRNPs contain only one copy each of SmB, SmD1, SmD2, SmD3, SmE, SmF, and SmG. Thus should stable SmE and SmF oligomeric complexes exist *in vivo*, they are definitely not components of mature snRNP particles. Future functional studies will need to pay particular regard to discriminating between heterotypic and homotypic protein interactions.

It is known that spliceosomal snRNP assembly, at least in higher eukaryotes, requires interactions with the SMN protein, which appears to play a role in guiding the correct associations of different Sm hetero-complexes (43–45). The results of this study clearly show that Sm/Lsm protein-protein interactions are not limited to highly specific hetero-associations. Furthermore the homo-assemblies we observe here (specifically SmE and SmF) preclude the formation of hetero-complexes required for functional snRNP formation. This demonstrates why there is a need for proteins such as SMN or the complex pIC1n (46) in assisting correct formation of eukaryotic snRNPs containing several different Sm/Lsm proteins. Without such chaperones, the necessary specificity of Sm-Sm interactions may be severely compromised.

One of the most intriguing results presented here is the affinity of the SmF homo-oligomeric complex for both single-stranded poly(U) RNA and poly(dT) DNA. Further studies are needed to assess the functional significance of this affinity for single-stranded DNA. It has been proposed that the diverse Sm/Lsm species found in eukaryotic organisms have evolved from one single Sm protein, presumably with the ability to form ring structures with RNA-binding affinity (9). Studies of the bacterial Sm-like Hfq protein have given rise to the postulate that Sm/Lsm proteins have evolved specifically to modulate RNA-RNA interactions (24, 25). A recent review (47) has highlighted the potential role of Sm/Lsm proteins (along with many other RNA-interacting domains) as a simple scaffold for mediating RNA interactions in an ancestral organism relying primarily on RNA-based metabolism. Our work suggests an even broader role in nucleic acid stabilization that includes single-stranded DNA as well as RNA. The ability of archaeal and bacterial Sm/Lsm proteins to form RNA-binding homo-oligomers has led to the general consensus that these may represent the ancestors of the Sm/Lsm protein family. However, the identification of eukaryotic proteins with the same self-associative and nucleic acid-binding properties shows that these po-

tentially ancestral traits are common to proteins from all domains of life.

Acknowledgments—We thank Dr. Nick Dixon (Australian National University) for providing us with the pETMCSIII plasmid, Debra Birch (Macquarie University) for technical assistance with electron microscopy, and Dr. Matthew Seaman for the gift of yeast genomic DNA. We also thank Dr. David Owen (University of Cambridge) for the contribution of some materials and equipment and Dr. Jan Löwe and Suzanne Cordell for their assistance with crystallographic data collection.

REFERENCES

1. Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds) 2nd Ed., pp. 525–560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L., and Cech, T. R. (1999) *Science* **401,** 177–180
3. Blumenthal, T. (1995) *Trends Genet.* **11,** 132–136
4. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Séraphin, B. (2000) *EMBO J.* **19,** 1661–1671
5. Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000) *Nature* **404,** 515–518
6. Séraphin, B. (1995) *EMBO J.* **14,** 2089–2098
7. Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., and Lührmann, R. (1995) *EMBO J.* **14,** 2076–2088
8. Cooper, M., Johnston, L. H., and Beggs, J. D. (1995) *EMBO J.* **14,** 2066–2075
9. Kambach, C., Walke, S., Young, R., Avis, J. M., delaFortelle, E., Raker, V. A., Luhrmann, R., Li, J., and Nagai, K. (1999) *Cell* **96,** 375–387
10. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M. G., and Mabbutt, B. C. (2001) *J. Mol. Biol.* **309,** 915–923
11. Mura, C., Cascio, D., Sawaya, M. R., and Eisenberg, D. S. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 5532–5537
12. Törö, I., Thore, S. P., Mayer, C., Basquin, J. R. M., Séraphin, B., and Suck, D. (2001) *EMBO J.* **20,** 2293–2303
13. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P., and Brennan, R. G. (2002) *EMBO J.* **21,** 3546–3556
14. Toro, I., Basquin, J., Teo-Dreher, H., and Suck, D. (2002) *J. Mol. Biol.* **320,** 129–142
15. Walke, S., Bragado-Nilsson, E., Séraphin, B., and Nagai, K. (2001) *J. Mol. Biol.* **308,** 49–58
16. Raker, V. A., Hartmuth, K., Kastner, B., and Lührmann, R. (1999) *Mol. Cell. Biol.* **19,** 6554–6565
17. Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001) *Nature* **409,** 539–542
18. Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003) *J. Biol. Chem.* **278,** 1239–1247
19. Urlaub, H., Raker, V. A., Kostka, S., and Lührmann, R. (2001) *EMBO J.* **20,** 187–196
20. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999) *EMBO J.* **18,** 5789–5802
21. Mayes, A. E., Verdone, L., Legrain, P., and Beggs, J. D. (1999) *EMBO J.* **18,** 4321–4331
22. Salgado-Garrido, J., Bragodo-Nilsson, E., Kandels-Lewis, S., and Séraphin, B. (1999) *EMBO J.* **18,** 3451–3462
23. Achsel, T., Stark, H., and Lührmann, R. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3685–3689
24. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C., and Storz, G. (2002) *Mol. Cell* **9,** 11–22
25. Moller, T. M., Franch, T., Jrup, P. H., Keene, D. R., Bächinger, H. P., Brennan, R. G., and Valentin-Hansen, P. (2002) *Mol. Cell* **9,** 23–30
26. Neylon, C., Brown, S. E., Kralicek, A. V., Miles, C. S., Love, C. A., and Dixon, N. E. (2000) *Biochemistry* **39,** 11989–11999
27. Schagger, H., and von Jagow, G. (1987) *Anal. Biochem.* **166,** 368–379
28. Laurent, T. C., and Killander, J. (1964) *J. Chromatogr.* **14,** 317–330
29. Leslie, A. G. W. (1992) *Joint CCP + ESF-EAMCB Newsletter on Protein Crystallogr.* **26**
30. Dodson, E. J., Winn, M., and Ralph, A. (1997) *Methods Enzymol.* **277,** 620–634
31. Navaza, J. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 157–163
32. Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kuntsleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
33. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
34. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 122–133
35. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497
36. Cathala, G., and Brunel, C. (1990) *Nucleic Acids Res.* **18,** 201
37. Kastner, B., Bach, M., and Lührmann, R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1710–1714
38. Plessel, G., Lührmann, R., and Kastner, B. (1997) *J. Mol. Biol.* **265,** 87–94
39. Raker, V. A., Plessel, G., and Lührmann, R. (1996) *EMBO J.* **15,** 2256–2269
40. Camasses, A., Bragado-Nilsson, E., Martin, R., Séraphin, B., and Bordonne, R. M. (1998) *Mol. Cell. Biol.* **18,** 1956–1966
41. Fury, M. G., Zhang, W., Christodouloupolos, I., and Zieve, G. W. (1997) *Exp. Cell Res.* **257,** 63–69
42. Palfi, Z., Lücke, S., Lahm, H.-W., Lane, W. S., Kruft, V., Bragado-Nilsson, E., Séraphin, B., and Bindereif, A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8967–8972
43. Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997) *Cell* **90,** 1013–1021
44. Mackenzie, A. E., and Gendron, N. H. (2001) *Nat. Struct. Biol.* **8,** 13–15
45. Selenko, P., Sprangers, R., Stier, G., Bühler, D., Fischer, U., and Sattler, M. (2001) *Nat. Struct. Biol.* **8,** 27–31
46. Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C., and Fischer, U. (2001) *Curr. Biol.* **11,** 1990–1994
47. Anantharaman, V., Koonin, E. V., and Aravind, L. (2002) *Nucleic Acids Res.* **30,** 1427–1464
48. Diederichs, K., and Karplus, P. A. (1997) *Nat. Struct. Biol.* **4,** 269–275
49. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291

Homomeric Ring Assemblies of Eukaryotic Sm Proteins Have Affinity for Both RNA and DNA: CRYSTAL STRUCTURE OF AN OLIGOMERIC COMPLEX OF YEAST SmF

Brett M. Collins, Liza Cubeddu, Nishen Naidoo, Stephen J. Harrop, Geoff D. Kornfeld, Ian W. Dawes, Paul M. G. Curmi and Bridget C. Mabbutt

J. Biol. Chem. 2003, 278:17291-17298.
doi: 10.1074/jbc.M211826200 originally published online March 4, 2003
