
The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs)

CAMERON MURA,¹ ANNA KOZHUKHOVSKY,¹ MARI GINGERY,¹ MARTIN PHILLIPS,²
AND DAVID EISENBERG^{1,2}

¹Howard Hughes Medical Institute, Molecular Biology Institute, and UCLA-DOE Center for Genomics and Proteomics, Los Angeles, California 90095-1570, USA

²Department of Chemistry and Biochemistry and Department of Biological Chemistry, University of California Los Angeles, Los Angeles, California 90095, USA

(RECEIVED July 19, 2002; FINAL REVISION December 13, 2002; ACCEPTED December 19, 2002)

Abstract

Intron splicing is a prime example of the many types of RNA processing catalyzed by small nuclear ribonucleoprotein (snRNP) complexes. Sm proteins form the cores of most snRNPs, and thus to learn principles of snRNP assembly we characterized the oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs) from *Pyrobaculum aerophilum* (*Pae*) and *Methanobacterium thermoautotrophicum* (*Mth*). Ultracentrifugation shows that *Mth* SmAP1 is exclusively heptameric in solution, whereas *Pae* SmAP1 forms either disulfide-bonded 14-mers or sub-heptameric states (depending on the redox potential). By electron microscopy, we show that *Pae* and *Mth* SmAP1 polymerize into bundles of well ordered fibers that probably form by head-to-tail stacking of heptamers. The crystallographic results reported here corroborate these findings by showing heptamers and 14-mers of both *Mth* and *Pae* SmAP1 in four new crystal forms. The 1.9 Å-resolution structure of *Mth* SmAP1 bound to uridine-5'-monophosphate (UMP) reveals conserved ligand-binding sites. The likely RNA binding site in *Mth* agrees with that determined for *Archaeoglobus fulgidus* (*Afu*) SmAP1. Finally, we found that both *Pae* and *Mth* SmAP1 gel-shift negatively supercoiled DNA. These results distinguish SmAPs from eukaryotic Sm proteins and suggest that SmAPs have a generic single-stranded nucleic acid-binding activity.

Keywords: Ribonucleoprotein; Sm protein; protein polymerization; uridine binding; OB fold

Excision of noncoding regions (introns) is a vital step in the maturation of precursor mRNAs. Most eukaryotic protein-coding genes contain multiple introns (Long et al. 1995), and thus high-fidelity pre-mRNA processing is essential to

ensure production of mature mRNAs with correctly registered exons. The simultaneous excision of introns and splicing of exons in eukaryotic pre-mRNA is catalyzed by a transiently stable assembly of five small nuclear ribonucleoproteins (snRNPs). This large assembly of uridine-rich snRNPs (U snRNPs) is known as the spliceosome, and at various stages in its catalytic cycle it consists of the U1, U2, U4/U6, and U5 snRNPs (Yu et al. 1999). Five small nuclear RNAs (snRNAs) and at least 80 proteins are contained within the spliceosome (Burge et al. 1999), making it roughly the same size as the ribosome (sedimentation coefficient of ~60S; Muller et al. 1998); furthermore, assembly of U snRNPs into spliceosomes is likely to be independent of pre-mRNA binding, as suggested by recent isolation of a novel U1●U2●U4/U6●U5 penta-snRNP devoid of mRNA (Stevens et al. 2002).

Reprint requests to: David Eisenberg, Department of Chemistry and Biochemistry and Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA 90095, USA; e-mail: david@mbi.ucla.edu; fax: (310) 206-3914.

Abbreviations: *Afu*, *Archaeoglobus fulgidus*; DTT, dithiothreitol; EM, electron microscopy; MPD, 2-methyl-2,4-pentanediol; *Mth*, *Methanobacterium thermoautotrophicum*; NCS, noncrystallographic symmetry; nt, nucleotide; OB-fold, oligosaccharide/oligonucleotide-binding fold; *Pae*, *Pyrobaculum aerophilum*; ss(D/R)NA, single-stranded (D/R)NA; SmAP, Sm-like archaeal protein; snRNP, small nuclear ribonucleoprotein; UMP, uridine-5'-monophosphate; *wt*, wild type.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0224703>.

Extensive biochemical and genetic data have shown that a key step in snRNP assembly is stepwise binding of seven cytoplasmic Sm proteins to exported snRNAs (Will and Luhrmann 2001). Each U snRNP is a complex formed from an ~110–180-nucleotide (nt) snRNA and two classes of proteins: (1) snRNP-specific proteins that confer snRNP-specific functions (e.g., U1A protein of U1 snRNPs) and (2) the Sm or Sm-like (Lsm) proteins that are common to each snRNP core (Will and Luhrmann 1997). The snRNAs contain a single Sm or Lsm binding site with the uridine-rich consensus sequence PuAU₋₄₋₆GPu (Pu = purine). However, specificity for this sequence is not stringent and there can be redundancy in Sm-snRNA binding (Jones and Guthrie 1990). The Sm sites are predicted to be single-stranded RNA regions flanked by stem-loop structures (Burge et al. 1999; Yu et al. 1999). Sm binding is highly sensitive to modifications of the flanking stem-loops and the Sm site of a given snRNA, and varies from one snRNA to another (Jarmolowski and Mattaj 1993). Sm-snRNA binding also may be modulated by interactions between certain Sm proteins and the survival of motor neurons (SMN) protein complex (Selenko et al. 2001), and by symmetric dimethylation of arginine residues in some of the RG dipeptide repeats of Sm (Brahms et al. 2000; Friesen et al. 2001; Meister et al. 2001) and Lsm (Brahms et al. 2001) proteins by a putative “methylosome” (Friesen et al. 2002). In eukaryotes, Sm D1•D2 and E•F•G heteromers simultaneously bind to snRNA to yield a “subcore” snRNP complex (Raker et al. 1996, 1999; Will and Luhrmann 2001). The final component to join the Sm complex is the B/B'•D3 heterodimer, and this triggers hypermethylation of the 5' m⁷G cap of snRNA to a trimethylated guanosine cap (m₃G). The m₃G cap and the snRNA•Sm core complex form a bipartite nuclear localization signal that results in transit of the snRNP core to the nucleus, where association of various snRNP-specific proteins completes the assembly process.

The importance of Sm proteins in RNP assemblies is underscored by their phylogenetic distribution: In addition to the canonical Sm and Lsm proteins found in eukaryotes ranging from yeast to humans, an Sm-like archaeal protein (“SmAP”) family has been discovered (Salgado-Garrido et al. 1999; Mura et al. 2001). The recent demonstration that the *E. coli* bacteriophage host factor *Hfq* is an Sm-like protein provides the first example of a eubacterial Sm protein (Moller et al. 2002; Zhang et al. 2002). These results imply fundamental roles for Sm proteins in the early evolution of RNA metabolism. Sm proteins probably mediate critical RNA-RNA, RNA-protein, and protein-protein interactions in snRNP cores. The vast network of protein-protein interactions in which Sm proteins participate was recently suggested by genome-wide two-hybrid screens of yeast Lsm proteins (Fromont-Racine et al. 2000).

Sm proteins have a tendency to associate into cyclic oligomers. Prompted by biochemical and genetic data, elec-

tron microscopic (EM) investigations of U snRNP particles revealed the “doughnut-shaped” ultrastructure of Sm and Lsm cores (Kastner et al. 1990; Achsel et al. 1999). The realization that Sm and Lsm proteins occur in groups of at least seven paralogs within the genome of a given organism suggests that snRNP cores are formed from Sm heteroheptamers, and two recent results verify this. First, Stark et al. (2001) reconstructed a 10 Å-resolution map of the U1 snRNP by cryo-EM and found that a model of the Sm heptamer could be docked into the ring-shaped body of the snRNP. Next, the *in vivo* stoichiometry of Sm proteins in yeast spliceosomal snRNPs was determined by a differential tag/pull-down assay, showing that the snRNP core domain contains a single copy of each of the seven Sm proteins (Walke et al. 2001). Stable subheptameric Sm complexes have been suggested as intermediates along the snRNP core assembly pathway (e.g., a D1•D2•E•F•G complex that binds snRNA; Raker et al. 1996), and ultracentrifugation and EM show that some of these oligomers can form ring-like structures that resemble intact, heptameric snRNP cores (e.g., a (E•F•G)₂ hexamer in Plessel et al. 1997). Such findings emphasize the importance of cyclic Sm heptamers in the snRNP core, and raise the possibility of other oligomeric states.

There is no atomic-resolution structure of a eukaryotic snRNP core. Nonetheless, the crystal structures of Sm-like archaeal proteins from *Afu* (Toro et al. 2001), *Pae* (Mura et al. 2001), and *Mth* (Collins et al. 2001) reveal a cyclic Sm homoheptamer and provide a model for snRNA binding in the snRNP core. Sm monomers fold as strongly bent, five-stranded antiparallel β-sheets (Kambach et al. 1999a) and form toroidal heptamers that surround a conserved cationic pore. The inner surface of this pore appears to be the oligouridine-binding site. The similarity between SmAP1 monomer and dimer structures and the nearly identical human Sm D1•D2 and D3•B heterodimers (Kambach et al. 1999b) supports SmAP-based models for the heptameric snRNP core.

Results

Crystallization and determination of the Pae and Mth SmAP1 structures

Crystallization of wild-type (*wt*) *Pae* SmAP1 was not straightforward, requiring dithiothreitol (DTT) for the formation of high-quality crystals. Identical crystallization buffers that lacked DTT failed to produce crystals, and presumably this additive is essential because it reduces the seven disulfide bonds that form between Cys8 residues in the *Pae* 14-mer (which can therefore be thought of as a dimer of heptamers rather than as a heptamer of dimers). Other reductants (e.g., β-mercaptoethanol) can substitute for DTT to yield crystals, although such crystals are of

poorer quality. Apparently, reduction of the disulfides frees heptamers to crystallize independently in orientations that relax crystal lattice strain, even when the 14-mer persists in the crystal (as in the $C222_1$ form reported here). The only other notable additive (uridine-5'-monophosphate, UMP) was unnecessary for *Pae* SmAP1 crystallization. Diffraction data extended to at least 2.05 Å-resolution for the $C222_1$ -form *Pae* SmAP1 crystals (Table 1). Previously we determined the crystal structure of *Pae* SmAP1 in spacegroup $C2$ by multiple-wavelength anomalous dispersion phasing (Mura et al. 2001). Thus, the $C222_1$ structure reported here was solved by molecular replacement. Due to poor electron density, only the uridine fragment of UMP was built into the final refined model. The final structure was refined to an R/R_{free} of 18.2%/22.6%, with acceptable model geometry (Table 1) and no outliers in a Ramachandran plot.

Mth SmAP1 was crystallized in three forms ($P1$, $P2_1$, $P2_12_12_1$) under three dissimilar conditions (a fourth form was reported by Collins et al. 2001). It is notable that *Mth* SmAP1 crystallized in the $P2_1$ form only in the presence of a single-stranded DNA (ssDNA) to which it was thought to bind, even though ssDNA was not found in the crystal structure. Because diffraction data were obtained from the $P1$ form before the first *Mth* SmAP1 structure was reported by Collins et al. (2001), we solved the $P1$ *Mth* structure by a combination of molecular replacement and free-atom model

refinement. Briefly, a homology model of *Mth* SmAP1 was built from the *Pae* SmAP1 structure. An unambiguous molecular replacement solution was found for this search model with the *Mth* $P1$ data. In order to reduce *Pae* model bias, this solution was converted to polyalanine, and phases from the initial model were used to auto-build an entirely new model with the ARP/wARP program. Initial phases for the $P2_1$ and $P2_12_12_1$ *Mth* data were obtained by molecular replacement with the refined $P1$ model (as summarized in Table 1). Electron density for the UMP-binding sites was more interpretable in *Mth* SmAP1 than in the *Pae* structure, and permitted model building of six complete UMPs (only uridine fragments were built for the other eight UMPs in the *Mth* 14-mer). All three *Mth* structures were refined to acceptable values of R/R_{free} and model geometries (Table 1).

Comparisons of known SmAP monomer, dimer, and heptamer structures

Several structures of Sm proteins and SmAPs are now available, making possible the comparative structural analyses of these proteins, and revealing the strict conservation of the Sm fold. The *Mth* heptamer structures reported here are virtually identical to the *Mth* SmAP1 structure reported by Collins et al. (2001), for example, 0.65 Å RMSD for superimposition of the $P1$ heptamer using main chain atoms.

Table 1. X-ray data collection and refinement statistics for several crystal forms

Crystal form	$P1$ (<i>Mth</i>)	$P2_1$ (<i>Mth</i> with UMP)	$P2_12_12_1$ (<i>Mth</i>)	$C222_1$ (<i>Pae</i> with UMP)
Data collection				
X-ray wavelength (Å)	1.1000	1.1000	1.5418	1.5418
Resolution range (Å)	90.0 – 1.85	100.0 – 1.90	100.0 – 2.80	100.0 – 2.05
# reflections (total/unique)	145,416/44,472	337,336/89,378	329,838/28,487	330,687/40,722
Completeness (%)	93.8 [67.4]	97.0 [92.4]	99.0 [92.5]	97.4 [95.5]
$I/\sigma(I)$	20.4 [4.1]	25.8 [2.3]	19.3 [3.3]	17.9 [4.1]
R_{merge} (%)	5.4 [25.5]	4.8 [56.5]	11.2 [38.7]	11.5 [50.9]
Model refinement				
Search model	Homology model of <i>Pae</i> SmAP1 structure (118F)	Refined $P1$ <i>Mth</i> SmAP1 structure (1JBM) <i>Face-face</i> 14-mer per a.u. (pseudo-72 point group symmetry)	Refined $P1$ <i>Mth</i> SmAP1 structure (1JBM) <i>Edge-edge</i> 14-mer per a.u.	Refined $C2$ <i>Pae</i> SmAP1 structure (118F) Heptamer per a.u.; <i>face-face</i> 14-mer in crystal (72 point group symmetry)
Crystal packing	Heptamer per a.u.			
Resolution range (Å)	20.0 – 1.85	20.0 – 1.90	15.0 – 2.80	20.0 – 2.05
<B-factor> (prot/water, Å ²)	27.9/37.2	41.7/48.6	50.9/40.1	25.1/34.8
Number of solvent or ligand molecules included in model:	water 273 ethylene glycol 11 acetate 5	water 387 MPD 14 UMP 14	water 86 ethylene glycol 13 chloride 3	water 325 glycerol 10 acetate 2 UMP 7
RMSDs: bonds (Å)	0.021	0.014	0.007	0.013
angles (°)	2.06	1.85	1.44	1.61
$R_{\text{cryst}}/R_{\text{free}}$ (%)	19.6/23.8	20.7/25.0	19.9/29.0	18.2/22.6
PDB code submitted	1JBM	1LOJ	1JRI	1LNX

Crystallographic statistics are given for the *Mth* and *Pae* SmAP1 structures in different space groups (with various packing geometries) and with bound ligands (UMP, MPD, etc.). Data were collected either in-house ($\lambda = 1.54$ Å) or at the NSLS synchrotron ($\lambda = 1.10$ Å). Statistics for the highest resolution shell are given in []. $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, and R_{free} was computed identically, except that 5% of the reflections were omitted as a test set. Nonprotein molecules were added based on sufficiently strong $|F_{\text{calc}}| - |F_{\text{obs}}|$ density ($>3\sigma$) and the chemical composition of the crystallization conditions.

Pairwise comparisons of the *Pae*, *Mth*, and *Afu* SmAP1s show that the compact, ~80-amino acid SmAP1 monomer structures are nearly identical (Fig. 1, Table 2). The most similar monomer structures are the *Afu/Mth* pair (0.51 Å RMSD), and the most dissimilar are *Mth/Pae* SmAP1 (1.02 Å RMSD). These values do not correlate to pairwise sequence similarities. The overall structure of the dimer interface is strictly conserved between SmAPs and human Sm heterodimers, as emphasized by the view in Figure 1. Greater RMSDs for heptamer compared to dimer alignments (and dimer compared to monomer alignments) suggests that a large fraction of the structural variation in higher-order SmAP oligomers is due to rigid-body reorientation of monomers with respect to one another. Mapping of the phylogenetic conservation of SmAP residues onto the *Pae*, *Mth*, or *Afu* heptamer structures shows that most of the conserved residues cluster about the cationic pore region (data not shown). The calculated electrostatic potential of the *Mth* SmAP1 surface reveals a strongly acidic loop-4 (L4) face, as found for *Pae* SmAP1 (Mura et al. 2001).

Various oligomeric states of SmAP1, including subheptamers and 14-mers

Biophysical characterization of *Pae* and *Mth* SmAP1 by a variety of methods reveals complex oligomerization properties that are consistent with nonheptameric SmAP oligomers. These methods include mass spectrometry, size exclusion HPLC, native polyacrylamide gel electrophoresis, and analytical ultracentrifugation. Sedimentation velocity ultracentrifugation reveals that *wt Pae*: (1) is monodisperse in solution; (2) has a symmetric and narrow Gaussian-shaped distribution of sedimentation coefficients, with a coefficient at 20°C of $S_{20,w} = 6.49$ S; and (3) has a frictional coefficient ratio close to one ($f/f_o = 1.2$, where f = experimentally derived frictional coefficient and

f_o = ideal frictional coefficient for a sphere with the M_r of SmAP1). These preliminary results suggested a roughly spherical, high-order *wt Pae* oligomer (SmAP1) $_n$, with $n \sim 12 \pm 2$ (data not shown).

The results of equilibrium sedimentation analyses of *wt Mth*, *wt Pae*, and the C8S mutant of *Pae* SmAP1 reveal the oligomeric states of these SmAP1s in solution, as shown in Figure 2. Molecular weights were estimated by fitting experimental curves to single exponential models. The calculated molecular weight of *wt Pae* (Fig. 2A) suggests that it exists as a 14-mer. Because other data also suggested a disulfide-bonded 14-mer, the single cysteine of *Pae* SmAP1 was mutated to serine to give the C8S mutant of *Pae* SmAP1. Sedimentation results with this mutant can be fit only by species with molecular weights much less than that of a heptamer (e.g., the 46.7-kD species shown in Fig. 2B), suggesting a pentamer ($n = 5$ gives a M_r of ~45 kD). The monodispersity of the data in Figure 2B suggests a single, stable subheptameric complex, although a rapidly exchanging mixture of several states (e.g., tetramers, pentamers, and hexamers) cannot be ruled out. In contrast to *Pae*, sedimentation equilibrium data for *Mth* SmAP1 show that it only forms a stable, monodisperse heptamer (Fig. 2C). The concentration dependence of the experimentally calculated M_r s (not shown), as well as the slight upward concavity of the residuals in Figure 2B,C, provide additional evidence for *Pae* and *Mth* SmAP1 monomer \leftrightarrow oligomer association reactions.

Polymerization of SmAP1 into polar fibers

The polymerization of both *Pae* and *Mth* SmAP1 into well ordered fibers is shown in the transmission electron micrographs (EMs) of Figure 3. Protein samples were in standard buffers (e.g., 25 mM Tris pH 7.5, 30 mM NaCl for *Pae* SmAP1), and reproducibly formed the striated bundles of

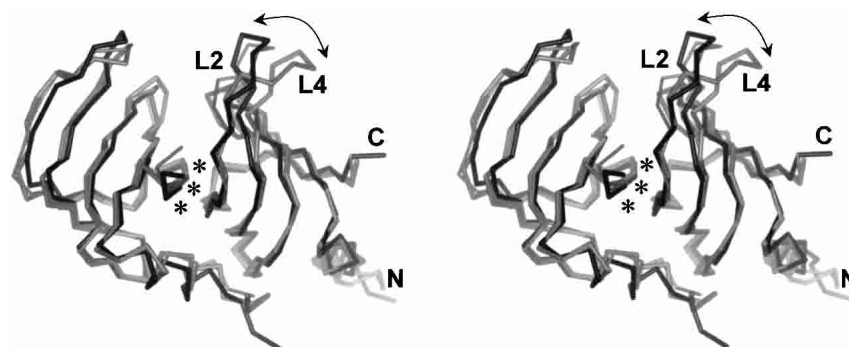


Figure 1. Similar structures of *Pae*, *Mth*, and *Afu* SmAP1 dimers shown by 3D structural alignment. A depth-cued stereoview is shown of the C α traces for aligned *Pae* (black), *Mth* (gray, thick lines), and *Afu* (gray, thin lines) SmAP1 dimers. N- and C-termini as well as loops L2 and L4 are indicated. The greatest structural variation is in the positions of these two pore-forming loops, whereas the dimer interface is structurally conserved (*). The difference in the width of the heptameric pores in *Pae* (~8–9 Å diameter) vs. *Afu* and *Mth* (~12–15 Å diameter) is due primarily to backbone variation in loops L2 and L4 (see arrows).

Table 2. Sequence and structure similarity between *Pae*, *Mth*, and *Afu* SmAPI

	<i>Afu</i>	<i>Mth</i>	<i>Pae</i>
<i>Afu</i>	—	0.51 Å (monomer) 0.61 Å (dimer) 0.81 Å (heptamer)	0.90 Å (monomer) 1.02 Å (dimer) 1.96 Å (heptamer)
	63.8% (ide) 78.3% (sim)	—	1.02 Å (monomer) 1.19 Å (dimer) 1.90 Å (heptamer)
	46.2% (ide) 61.5% (sim)	51.5% (ide) 69.7% (sim)	—

RMSDs are given in the upper triangle for pairwise 3D alignments of *Pae*, *Mth*, and *Afu* SmAPI monomers, dimers, and heptamers (using mainchain atoms only). The *Afu*–*Mth* pair superimposes best, whereas the *Pae*–*Mth* monomer structures are most dissimilar. Pairwise sequence identities (ide) and similarities (sim) are provided in the lower triangle.

fibers seen in these EMs. Measurement of the fiber dimensions, together with the diameters of SmAPI heptamers from crystal structures (~70–75 Å), suggests a model in which the fibers are formed by head-to-tail stacking of heptamers, with the SmAPI sevenfold axis roughly parallel to the fiber axis (see white arrows in Fig. 3B). Several fibers may associate laterally into bundles or sheets, as seen most clearly in Figure 3A,B.

In order to test this head-to-tail stacking model, we assayed fiber formation by *wt Pae* and the C8S mutant. Under oxidative conditions, *wt Pae* SmAPI forms disulfide-bonded 14-mers in which the highly acidic L4 faces are exposed at both ends of the barrel-shaped structure (see the *Pae* 14-mer in Fig. 4A). Such a 14-mer would be constrained to form only head-to-head interfaces (i.e., L4 face-to-L4 face) in a fiber, and would probably not do so because of the unfavorable electrostatic cost of closely apposing these anionic faces (at least not at the neutral pHs or low ionic strength conditions in which the SmAPIs were buffered). As expected, *wt Pae* forms only ring-shaped structures under oxidative conditions (Fig. 3C). However, when the seven disulfide bonds that covalently link heptamers into 14-mers are eliminated, *Pae* SmAPI assembles into fibers with roughly similar morphologies as *Mth* fibers. Polymerization can be achieved either by addition of a reducing agent (as in Fig. 3D) or by mutation of the cysteine (C8S mutant in Fig. 3E). Such fiber formation has been hitherto unreported for Sm proteins.

Packing of *Mth* and *Pae* SmAPI heptamers in four crystal forms

The fortuitous crystallization of *Mth* and *Pae* SmAPI in several forms with different heptamer packing geometries allows us to rationalize the oligomerization results described above. The *Pae* SmAPI C222₁ structure differs

from the original C2 form in that heptamers pack head-to-head in the orthorhombic lattice to give a 14-mer with point group symmetry 72, as shown in Figure 4A. This 14-mer is likely to be significant because: (1) it is consistent with the oligomerization described above on the basis of biophysical characterization, (2) it persists in the C222₁ lattice despite the requirement of DTT for crystallization (the cysteine sulfhydryls are separated by >8–9 Å), (3) the heptamer–heptamer interface occludes 7550 Å² of surface area, and (4) it is corroborated by an *Mth* 14-mer in the asymmetric unit of the P2₁ form. The surface area occluded in the heptamer–heptamer interface of the P2₁ *Mth* 14-mer (3000 Å²) is probably significant too, although less than half as much as in the *Pae* interface.

The *Mth* P2₁2₁2₁ crystal structure provides a model for the atomic structure of SmAPI fibers. In the *Mth* P1 and P2₁2₁2₁ lattices, SmAPI heptamers form quasihexagonal layers that stack upon one another to give a crystal. In the P1 form these layers are staggered; however, in the P2₁2₁2₁ form these layers are in register. Figure 4B shows how the head-to-tail stacking of SmAPI heptamers in this crystal form produces cylindrical tubes. A slight tilt of each heptamer with respect to the tube axis (~15°) results in the SmAPI sevenfold axes being parallel, but not coaxial. Because they are formed by head-to-tail stacking of asymmetric heptamers, these tubes have a defined polarity, and, when rendered as molecular surfaces, they bear a striking resemblance to the EM fibers shown in Figure 3. The tubes are also consistent with EM fiber dimensions. In addition to providing insights into polymerization and oligomerization states, two of the crystal forms (*Pae* C222₁ and *Mth* P2₁) were used to investigate the ligand-binding properties of SmAPIs.

Crystal structures of *Mth* and *Pae* SmAPI bound to various ligands

The 1.90 Å-resolution crystal structure of *Mth* SmAPI bound to uridine-5'-monophosphate (UMP) is shown in Figure 5. The protein was cocrystallized with this ribonucleotide in an effort to determine its likely RNA-binding site (cocrystallization efforts were unsuccessful with single-stranded DNA or RNA oligonucleotides). As shown in Figure 5A, SmAPI binds UMP with a 1:1 stoichiometry, so that 14 UMPs are bound to the 14-mer near the pore region. The UMPs bind near the flat face of the *Mth* heptamer, opposite the highly acidic loop L4 face. The structure of the SmAPI•UMP complex is shown in more detail in Figure 5B, which shows that the binding site is well defined by electron density. The uracil ring intercalates between the guanidinium group of Arg72 and the imidazole ring of His46 (both of these residues are highly conserved in SmAPIs). The planes of these three moieties are spaced ~3.6 Å apart, as expected for energetically favorable stacking

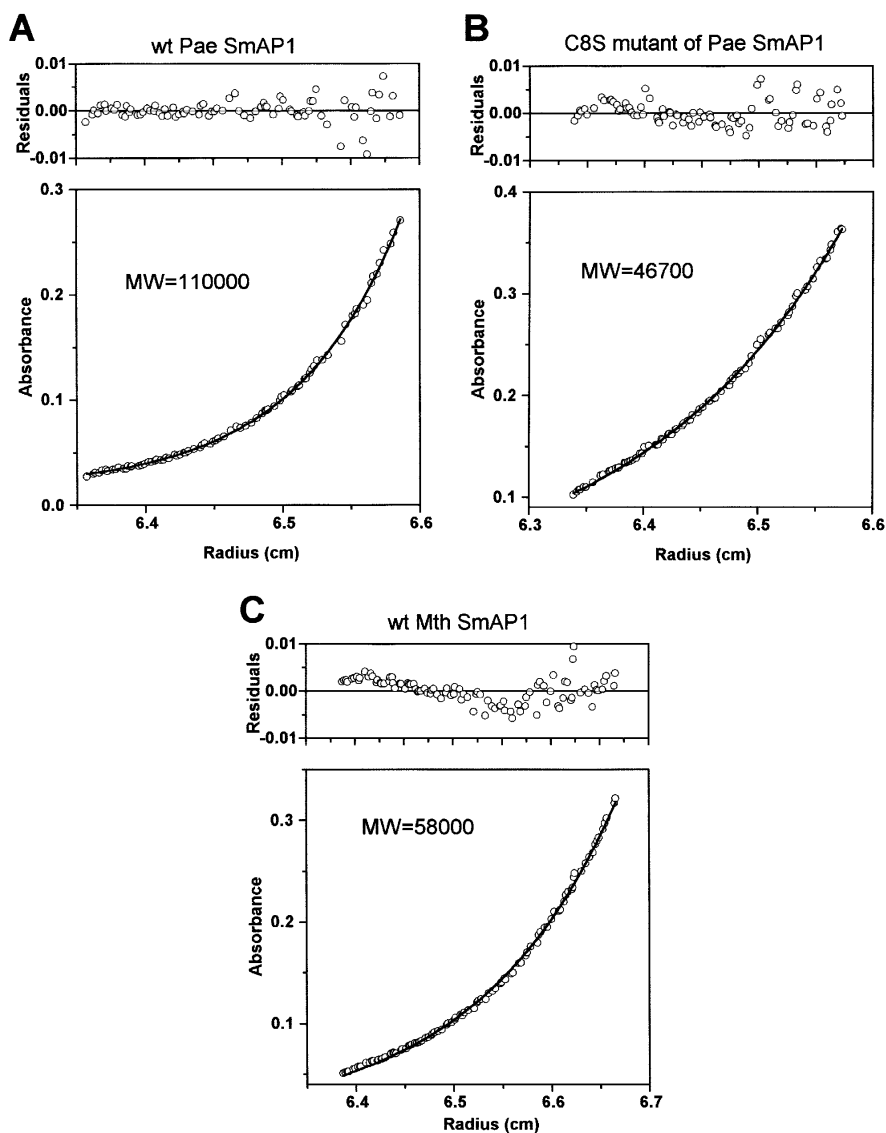


Figure 2. The oligomeric states of *Pae* and *Mth* SmAP1 in solution shown by equilibrium sedimentation. Representative sedimentation results for analytical ultracentrifugation of *wt Pae* SmAP1 (A), the C8S mutant of *Pae* SmAP1 (B), and *wt Mth* SmAP1 (C) are shown. Data were collected at 20°C, at a rotor speed of 12,500 rpm, with absorbance measured at 280 nm. Protein concentrations were 0.69 mg/mL (A), 1.26 mg/mL (B), and 0.85 mg/mL (C). Weight-average molecular weights (given in kD) were determined by fitting experimental data (circles) with a single exponential (solid line), and include roughly 2%–3% error (residuals are in top panels); note that the protein samples are monodisperse. The molecular weight of the *wt Pae* protein suggests that it exists as a 14-mer, whereas the *wt Mth* data closely fit a heptamer. The molecular weight of the C8S mutant is significantly less than that of a heptamer, suggesting lower oligomerization states (4-, 5-, or 6-mers). Such “subcomplexes” have been found for eukaryotic Sm proteins (see text).

interactions between conjugated π -systems. Individual protein-UMP contacts are discussed in greater detail below.

In addition to the expected UMP binding, we found that each *Mth* SmAP1 monomer binds a molecule of MPD. The MPD binding site is somewhat solvent-exposed, near the periphery of the SmAP1 ring (Fig. 5A). Protein-MPD recognition is the same in each of the 14-monomers, and the primary contact is hydrogen bonding between the Ser21 hydroxyl and MPD. There also are several water-mediated

SmAP \cdots H $_2$ O \cdots MPD contacts. The cryoprotectant for the *P1* and *P2*₁₂₁ *Mth* SmAP1 crystals was ethylene glycol (Table 1), and in these structures some of the SmAP1 monomers bind ethylene glycol in the same site as MPD.

A UMP binding site was found in the *Pae* SmAP1•UMP co-crystal structure as well, but it is not as clearly defined by electron density as is *Mth* SmAP1•UMP. The *Pae*•UMP structure, which was refined to a resolution of 2.05 Å, is shown in Figure 5C. UMPs bind to the same face of the

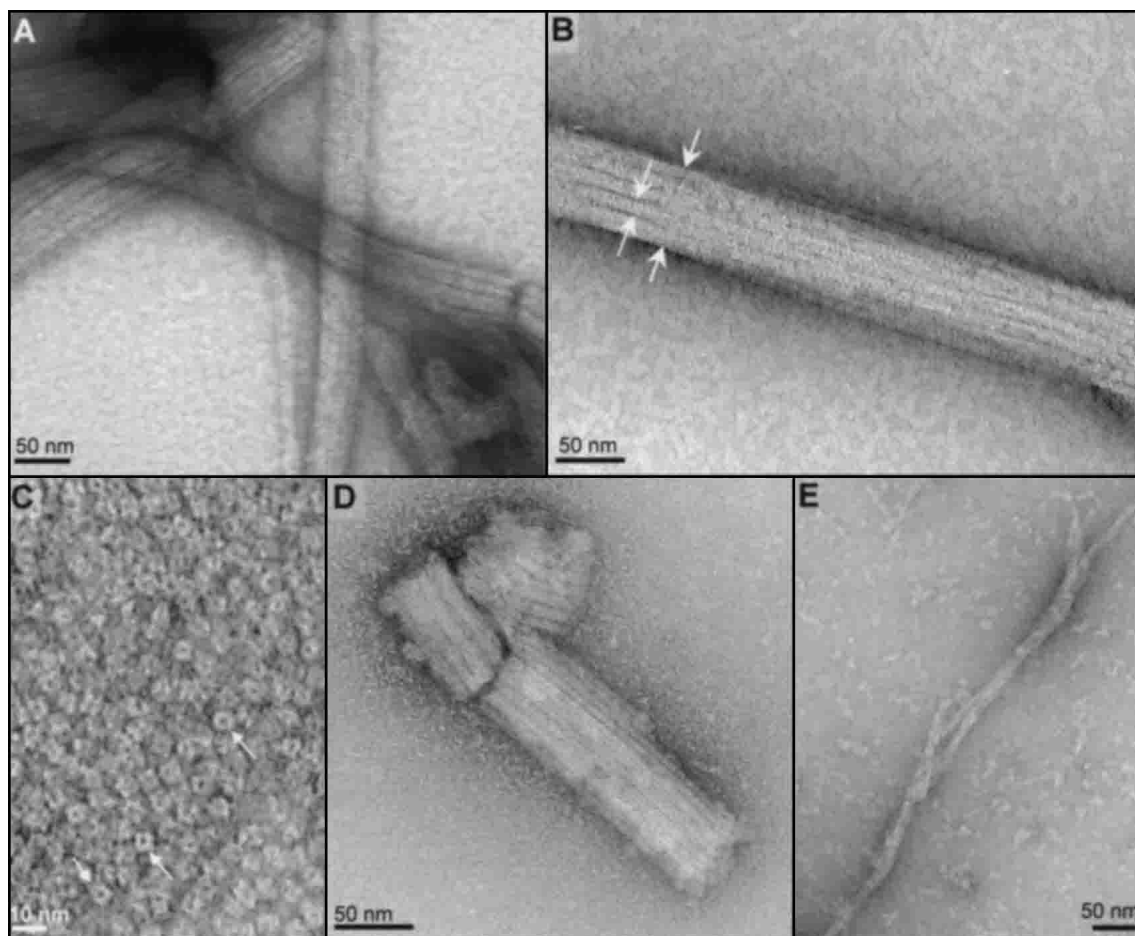


Figure 3. Polymerization of SmAP1s into polar fibers. Transmission EMs are shown for *wt Mth* (A,B), *wt Pae* (C, oxidized; D, reduced), and the C8S mutant of *Pae* SmAP1 (E). Scale bars represent 10 nm for panel (C), and 50 nm for all other panels. The striated bundles formed by *Mth* SmAP1 (A,B) and nondisulfide-bonded *Pae* SmAP1 (D,E) are extremely well ordered. The distance between the inner arrow tips in (B) corresponds to ~ 8.3 nm (in agreement with the heptamer diameters from crystal structures), and suggests that the fiber axis is parallel to the heptameric sevenfold. The ~ 50 -nm distance between the outer white arrows in (B) corresponds closely to six heptamer widths. Together with heptamer packings in various crystal forms, these EMs suggest that SmAP1 fibers form by head-to-tail stacking of heptamers (see Fig. 4). Doughnut-shaped SmAP1s are visible in the backgrounds of these EMs (most clearly for the *wt Pae* sample in panel C).

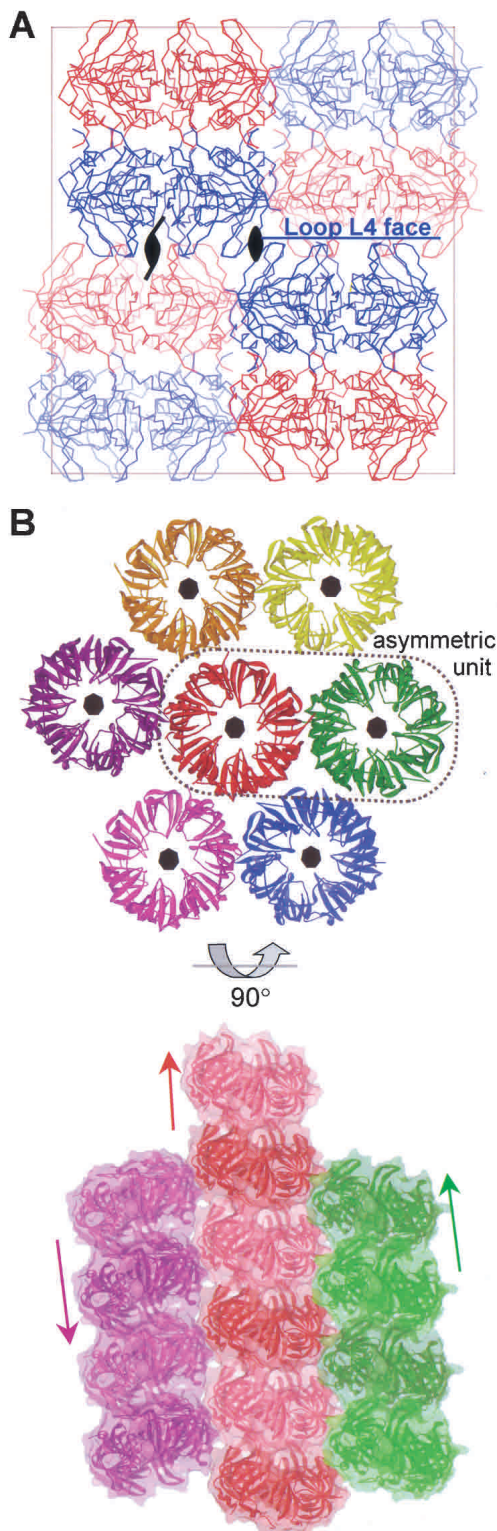
heptamer as in *Mth* (i.e., the “flat face” opposite L4), but are much more distant from the pore. Only the planar uracil fragment of UMP is clearly defined in $2|F_o| - |F_c|$ electron density maps, and protein-UMP contacts are scarce in this binding site (*Pae* SmAP1 residues in the region of this uridine are not very conserved). The only close UMP contact is made by the side chain of Asn46, but the geometry of the Asn46...UMP interaction does not satisfy standard hydrogen bond criteria (in terms of both distances and angles), and favorable interactions probably do not exist between the UMP O4 oxygen and the amide nitrogen of the Asn46 side chain, or between the UMP N3 nitrogen and the amide oxygen of Asn46. Also, there are no aromatic side chains in this region to participate in π -stacking interactions with the uracil base. As in *Mth* SmAP1, additional small-molecule

binding sites exist in *Pae* SmAP1: Many of the modeled glycerol molecules are bound identically near the loop L4 faces (Fig. 5C).

The structure of an *Afu* SmAP1•U₃ complex was recently determined by Toro et al. (2001) and reveals a similar mode of uridine recognition in *Afu* and *Mth* SmAP1. The UMP binding site and SmAP1...UMP interactions clearly differ in *Mth* and *Pae* SmAP1, and, because the binding site was poorly resolved in the *Pae*•UMP complex, this structure was not included in the comparative analysis shown in Figure 6. In the *Mth* and *Afu* structures, the aromatic pyrimidine ring intercalates between the side chains of the highly conserved Arg/His pair, and specific recognition is achieved by hydrogen bonding of the uracil ring to the side chain of a strictly conserved asparagine residue (Asn48_{*Mth*}). The main

chain amide nitrogen of a highly conserved aspartate (Asp74_{Mth}) also participates in hydrogen bonding to a uracil carbonyl oxygen. The pattern of hydrogen bond donors/acceptors in the Asn48/Asp74_{Mth} pair makes binding spe-

cific for a uracil (if RNA) or thymine (if DNA) base. Additional specificity for uracil may be achieved by two means: (1) recognition of the 2' hydroxyl of the ribose (RNA vs. DNA discrimination) and (2) the C5 carbon of the pyrimidine ring of uracil is only 3.8 Å from the backbone carbonyl oxygen of Leu45_{Mth} from an adjacent monomer, thus providing steric and polar discrimination against the methyl on the C5 carbon of thymine. We crystallized *Pae* and *Mth* SmAP1 in the presence of various other nucleoside monophosphates (e.g., AMP, CMP, GMP), but there was no evidence for binding of these non-uridine NMPs (data not shown). The only significant differences in uridine recognition by *Mth* and *Afu* SmAP1 are highlighted by two arrows in Figure 6B. These are: (1) hydrogen bonding of an *Mth* Arg72 side chain from an adjacent monomer to the 2' hydroxyl of the ribose, and (2) hydrogen bonding between a phosphate oxygen and an imidazole nitrogen from the His46_{Mth} residue of an adjacent monomer. Overall, it appears that the mode of uridine recognition is conserved in the SmAP family.



Pae and *Mth* SmAP1 gel-shift negatively supercoiled DNA

In our initial attempts to determine the biochemical function of *Pae* SmAP1, we inadvertently found that this protein gel-shifts negatively supercoiled plasmid DNA. This activity was further investigated for both the *Mth* and *Pae* SmAP1s, and examples of it are shown in Figure 7. Migration of the negatively supercoiled plasmid “p5L1c1” is severely retarded by incubation with μM concentrations of *Mth* heptamer in Figure 7A. Interestingly, the extent of gel shift increases at higher concentrations of *Mth* SmAP1, until saturation of the effect occurs at $\sim 60 \mu\text{M}$ (cf. lanes 7 and 8 of Fig. 7A). A similar gel shift occurs to supercoiled DNA when it is incubated with *wt Pae* SmAP1, as shown in lane 4 of Figure 7B. This experiment also shows that the gel shift can be eliminated by incubation with a 26-nt single stranded DNA (ssDNA). Inhibition of the gel-shift activity is titratable, and there is no gel shift at higher concentrations of ssDNA (lane 8, Fig. 7B).

Figure 4. Various crystalline oligomers of *Pae* and *Mth* SmAP1. A unit cell of the *Pae* SmAP1 $C222_1$ crystal form is shown in (A), along with examples of crystallographic twofold and 2_1 screw axes. The asymmetric unit is a heptamer (shown as C α traces in red or blue), and a *Pae* SmAP1 14-mer with 72-point group symmetry is formed from adjacent asymmetric units (7550 Å² of surface area is buried at the heptamer–heptamer interface). Orthogonal views of the quasihexagonal packing of *Mth* SmAP1 heptamers in the $P2_12_12_1$ crystal form are shown in (B). Heptamers stack upon one another to form cylindrical tubes, thus providing a model for the structure of the EM fibrils (see text for explanation). The head-to-tail association of heptamers gives the tubes a defined polarity (colored arrows). Molecular surfaces show that the lateral packing of tubes in the crystal may generate the striated bundles seen by EM.

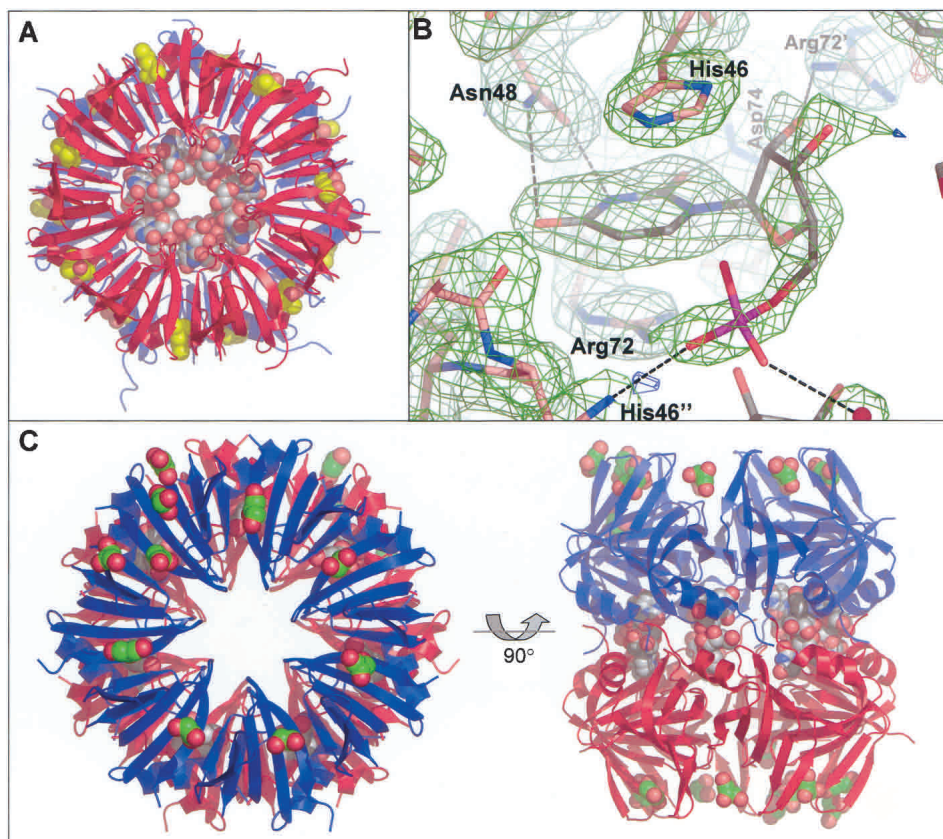


Figure 5. Ligand-binding sites in the structures of *Mth* and *Pae* 14-mers bound to UMP. The two *Mth* heptamers (red, blue) in the asymmetric unit of the $P2_1$ form are shown in (A). A single molecule of MPD binds identically to each monomer (space-filling, colored by atom type with yellow carbons). Space-filling models of the 14 UMP ligands that they bind in the pore region (colored by atom type, gray carbons). Electron density for a UMP binding site is shown in (B). The $2|F_o| - |F_c|$ density is contoured at $+1.2\sigma$ (green) and $|F_o| - |F_c|$ maps are contoured at -3.2σ (red) or $+3.2\sigma$ (blue). Conserved residues that form the UMP binding sites are labeled, and residues from different monomers are distinguished by primes. Hydrogen-bond distances are not shown, for the sake of clarity (see Fig. 6). Orthogonal views are shown in (C) for the *Pae* SmAPI 14-mer in the $C222_1$ lattice (heptamer per a.u.). Ten glycerol molecules bind to each heptamer (space-filling, green-colored carbons), and seven of them occupy identical sites. Only the uridine fragments of UMP were modeled (space-filling, gray-colored carbons), at identical sites distal to the pore region.

Related results from similar DNA gel-shift assays and control experiments have revealed that: (1) the *Pae* activity is specific for supercoiled (*sup*) plasmid DNA, whereas *Mth* SmAPI gel-shifts both *sup* and linearized plasmids; (2) *Pae* activity is eliminated by $MgSO_4$, whereas the dependence of *Mth* activity on divalent metals such as Ca^{2+} , Mg^{2+} , and Mn^{2+} is not as straightforward; (3) ssDNA of any sequence and length $> \sim 20$ nt inhibits the gel-shift activity of *Pae* and *Mth* in a concentration-dependent manner; (4) both *Pae* and *Mth* activities are nonspecific with respect to the *sup* DNA; (5) *Pae* and *Mth* are not causing a gel-shift by linearizing or otherwise cutting both strands of the *sup* DNA; (6) *Mth* gel-shift activity is not temperature-dependent at and above room temperature, whereas the extent of *Pae*-induced gel shift abruptly increases at $\sim 55^\circ$ – $60^\circ C$. All of these results come from experiments in which the migration of a large (>4000 -bp) plasmid DNA is assayed in agarose gels. Binding of *Mth* SmAPI to any one of the ssDNAs that inhibit the

sup DNA gel shift (e.g., Fig 7B) has been assayed in preliminary native PAGE experiments; these results suggest that ssDNA inhibits the *sup* DNA gel shift by directly binding to SmAPI.

Discussion

Comparative structural analysis of *Sm* proteins and SmAPs

SmAPs form a phylogenetically well conserved family of proteins whose sequences and structures are similar to eukaryotic Sm and Lsm proteins. The *Afu*, *Mth*, and *Pae* SmAPI monomer and homodimer structures are nearly identical to one another (Fig. 1, Table 2) and to the human Sm D3•B and D1•D2 heterodimers (Collins et al. 2001; Mura et al. 2001; Toro et al. 2001), thus qualifying the

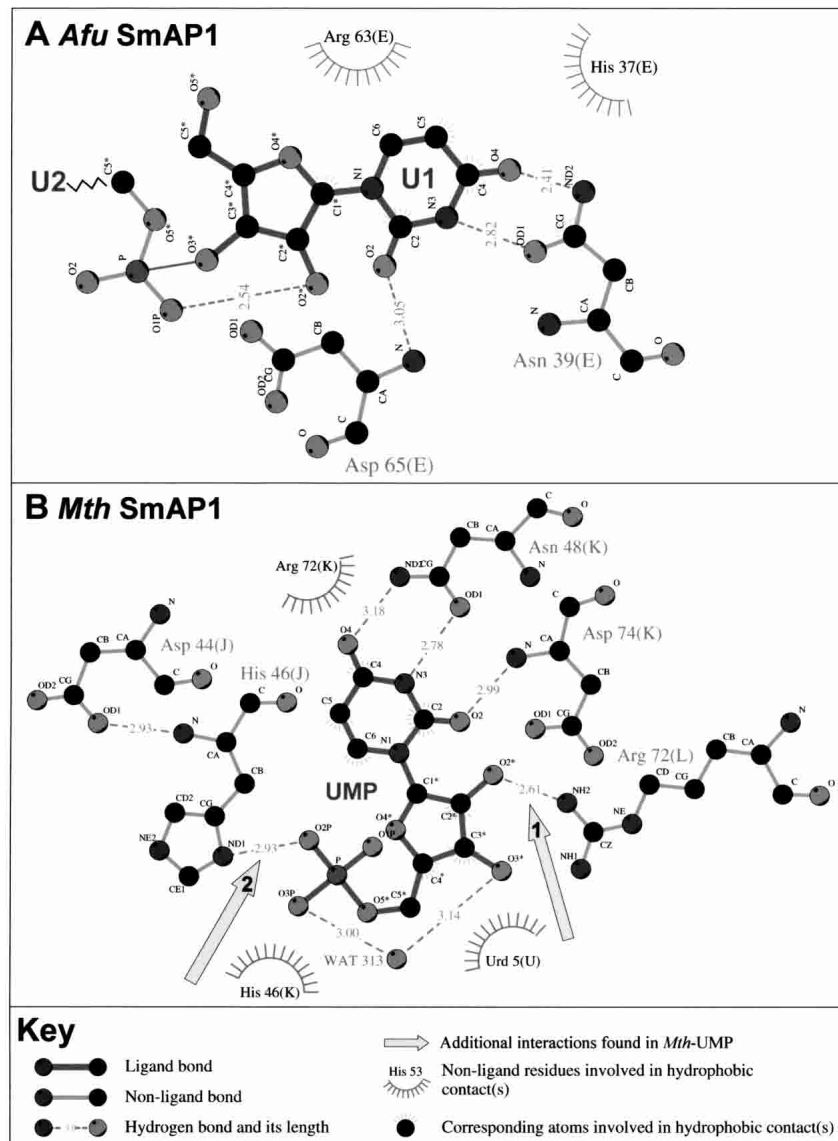


Figure 6. Conserved mode of uridine recognition by *Mth* and *Afu* SmAP1. Interactions between SmAP1 and uridine are diagrammed for *Afu* (A) and *Mth* (B). The remainder of the U_3 oligouridine from the *Afu* structure (indicated by a $U2\sim$) has been omitted in (A) for the sake of clarity. Parenthesized letters after residue labels denote individual monomers. In both structures, the aromatic uracil base intercalates between a highly conserved pair of Arg/His side chains—e.g., the guanidinium of Arg72 and imidazole of His46 for *Mth* SmAP1. Specific interactions and differences between *Afu* and *Mth* are discussed in the text. This figure was derived from LIGPLOT-generated output (Wallace et al. 1995).

SmAP1 homoheptamer as a model for the Sm heteroheptamer. Folding as antiparallel, five-stranded β -sheets capped by a short N-terminal α -helix, Sm and SmAP structures closely resemble proteins of the oligosaccharide/oligonucleotide binding (OB) fold family (Murzin 1993). The Sm–Sm interface is mostly formed by hydrogen bonding between mainchain atoms of β_4 and β_5 strands from adjacent monomers, thus explaining the reduced sequence conservation of interfacial residues. The recent solution structure of the SMN Tudor domain (Selenko et al. 2001), which interacts with Sm proteins to form snRNP cores, provides

the unexpected result that the SMN and Sm monomers have similar OB-like folds.

The *Afu*, *Mth*, and *Pae* SmAP1 heptamer structures are similar (Table 2), primarily because of the conserved interface between adjacent monomers. The central cationic pore is also highly conserved in terms of sequence and overall structure. However, one of the least conserved features of the SmAP1 heptamers is the calculated electrostatic potential of the surfaces: The L4 face of the *Afu* surface is very basic, whereas the *Pae* and *Mth* L4 faces are intensely acidic. Such differences are likely to be important for modu-

lating putative SmAP-RNA interactions near the pore. The most obvious structural difference between SmAP1 heptamers is the width of the cationic pore. Variation of pore width in *Pae* (~8–9 Å diameter) versus *Afu* and *Mth* (~12–15 Å) is due to main chain and side chain rotamer variations in the L2 and L4 loops. Besides the N- and C-termini, these pore-forming loops are the most structurally variant regions in SmAP1 monomers (Fig. 1). The most significant differences between SmAP1 heptamer structures and snRNP Sm cores will likely arise from the two largest differences between the *Pae*, *Mth*, and *Afu* SmAP1 sequences and some eukaryotic and archaeal Sm protein sequences: (1) several eukaryotic Sm proteins (e.g., human SmB/B') and some SmAPs have extended C-terminal regions with up to 70 more residues than the ~80-residue core Sm domain and (2) some eukaryotic Sm proteins may have up to 30 more amino acids in the L4 loop. Preliminary results with a *Pae* SmAP3 homolog that contains 60 additional C-terminal residues show that it too forms heptamers, and that the Sm core heptamer is conserved (C. Mura and D. Eisenberg, unpubl.).

The oligomerization properties of SmAPs

Like the Lsm (but not Sm) proteins, *Pae*, *Mth*, and *Afu* SmAP1 form heptamers in the absence of RNA. In addition to the expected heptamers, SmAP1 exhibits complex self-association properties indicative of 14-mers and subheptameric oligomers. Various oligomeric states were characterized in vitro (primarily by ultracentrifugation, Fig. 2), revealing roughly spherical disulfide-bonded *Pae* SmAP1 14-mers and a monodisperse population of *Mth* SmAP1 heptamers. Additionally, we created a cysteine-free point mutant of *Pae* SmAP (C8S) and found that it forms subheptameric states (most likely pentamers). Similar plasticity of oligomerization behavior has been reported for human Sm proteins. Lührmann et al. found that a human Sm E●F●G complex forms a stable oligomer—most likely an (E●F●G)₂ hexamer—whose ring-shaped structure resembles intact Sm heteroheptamers by EM (Raker et al. 1996; Plessel et al. 1997). One of these studies also found that stable, subheptameric complexes of human Sm proteins (e.g., a D1●D2●E●F●G pentamer) may be intermediates in the Sm-RNA assembly pathway (Raker et al. 1996). In the human Sm D3●B structure, the heterodimers pack as (D3●B)₃ hexamers in the asymmetric unit of the crystal (Kambach et al. 1999b), and *Afu* SmAP2 has been shown to form hexamers (Toro et al. 2002). The recently discovered *E. coli* Sm-like protein *Hfq* is thought to form hexamers as well (Arлуison et al. 2002).

We found that *Pae* and *Mth* SmAP1 oligomerize into 14-mers, either in vitro (*Pae*) or in various crystal forms (*Pae* and *Mth*). The highly acidic L4 faces are exposed in the barrel-shaped 14-mers, as expected from electrostatic

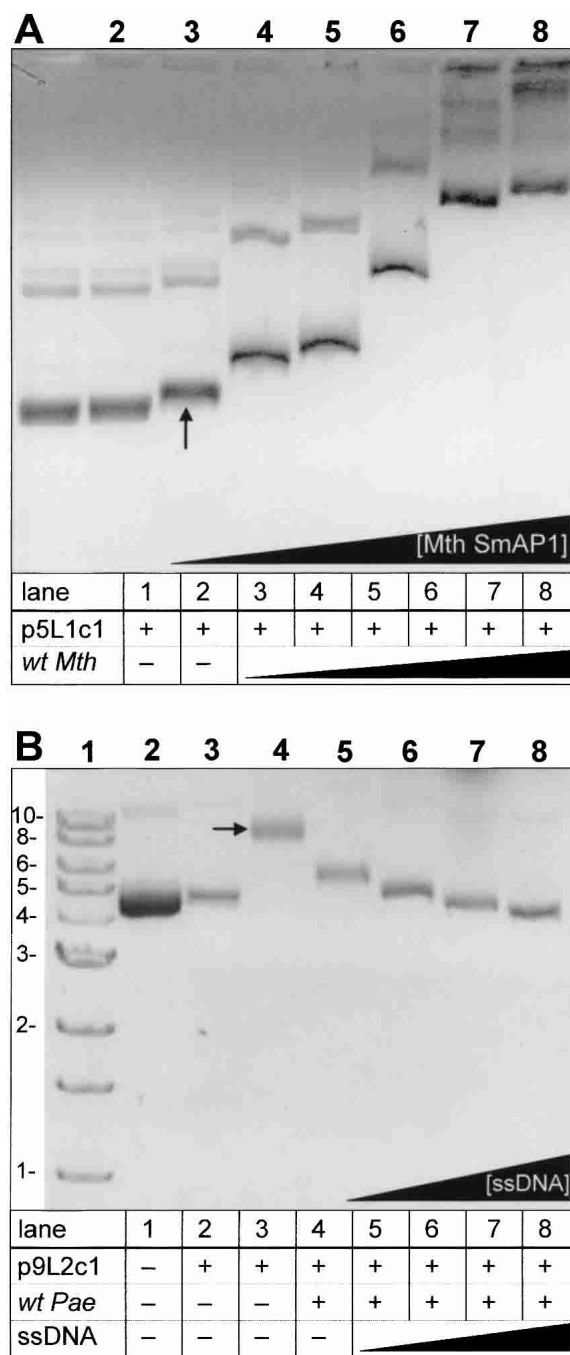


Figure 7. Gel-shift of supercoiled DNA by *Mth* and *Pae* SmAP1. The ability of *Mth* and *Pae* SmAP1 to shift the electrophoretic mobility of negatively supercoiled plasmid DNA is shown in the agarose gels of (A) and (B), respectively. In (A), increasing concentrations of *Mth* SmAP1 were incubated with a negatively supercoiled plasmid ("p5L1c1"). The first onset of gel shift is apparent at the lowest concentration of *Mth* (1.1 μM heptamer, arrow in lane 3), and saturates by the highest concentration (60 μM, lane 8). The ability of a 26-nt ssDNA to inhibit the gel shift induced by *Pae* SmAP1 is shown in (B). A DNA bp ladder is provided in lane 1 (numbers indicate 1000 bp), and negative controls are provided by freshly prepared plasmid alone (lane 2) or plasmid incubated under reaction conditions lacking SmAP1 protein (lane 3). The arrow in lane 4 shows the maximal gel shift in the absence of ssDNA.

considerations. The heptamer–heptamer interface buries much surface area in both *Pae* (7550 Å²) and *Mth* (3005 Å²), suggesting the significance of these oligomers. The crystal structure of another SmAP homolog (*Pae* SmAP3) shows that it also forms 14-mers in the asymmetric unit (C. Mura and D. Eisenberg, unpubl.). The propensity of cyclic SmAPs to crystallize as head-to-head oligomers with dihedral symmetry is shared by another single-stranded RNA binding protein that has an OB-like fold: The *trp* RNA-binding attenuation protein (TRAP) forms toroidal 11-mers that stack as both head-to-head and head-to-tail 22-mers in the crystal (Antson et al. 1999).

An unexpected property of SmAPs is their polymerization into well-ordered fibers under physiological conditions. Three lines of evidence suggest that these polar fibers form by the head-to-tail stacking of heptamers (Fig. 3): differential fiber formation by C8S and *wt Pae* SmAP1; comparison of measured fiber dimensions with SmAP1 heptamer dimensions; and electrostatic considerations for the packing of highly charged heptameric disks. The packing of *Mth* SmAP1 heptamers in the *P2₁2₁2₁* lattice supports our head-to-tail polymerization model, and provides an atomic-resolution model for the fibers (Fig. 4). The critical role of Cys8 in preventing fiber formation by stabilizing the *Pae* 14-mer seems significant, given the potentially oxidative cytosol of *Pae* and other thermophilic archaea (Mallick et al. 2002). Such complex oligomerization properties have not been reported for eukaryotic Sm proteins, and the biological significance of SmAP1 14-mers and homogeneous, fibrillar SmAP polymers is not yet known.

The ligand-binding properties of SmAPs

Comparison of the structures of *Mth* SmAP1 bound to UMP and *Afu* SmAP1 bound to oligouridine (U₃) reveals a highly conserved mode of RNA recognition in SmAPs. UMP binds near the sevenfold axis, suggesting the pore as a putative RNA binding site. Diagrams of SmAP1...UMP interactions show that both SmAPs specifically bind the uracil base by a combination of π -stacking and hydrogen-bond interactions with strictly conserved SmAP residues (Fig. 6). Differences between UMP binding in *Mth* and *Afu* are limited to interactions with the ribophosphate moiety, and may not be significant, because *Mth* SmAP1 was cocrystallized with free UMP nucleotide, whereas *Afu* SmAP1 was crystallized with a U₃ oligouridine. The oligo(U) specificity of RNA binding to *Afu* SmAP1 is the same as the substrate specificity of eukaryotic Sm proteins (Achsel et al. 2001; Toro et al. 2001). The binding geometry of UMPs in *Mth* SmAP1 allows them to be strung together into a hypothetical oligouridine that may mimic biologically relevant RNA binding in the Sm core of snRNPs. If all SmAPs specifically bind to an oligouridine site in vivo, then geometric considerations require such an RNA-binding site to lie near the

sevenfold symmetry axis (i.e., the pore); however, the uridine-binding site in *Pae* SmAP1 is distal to the pore and not easily interpretable in electron density maps, suggesting low-affinity binding at this alternative site (Fig. 5). We note that the same UMP-binding site proximal to the pore in *Afu* and *Mth* exists in *Pae* SmAP1, and that UMP can be docked into this site with only minimal changes to side chain rotamers. Failure of other NMPs to cocrystallize with *Mth* or *Pae* SmAP1 supports the specificity of uridine binding that we infer from the *Mth* and *Afu* crystal structures. Additional sites occupied by MPD, ethylene glycol, or glycerol are clearly defined by electron density in *Mth* and *Pae* SmAP1, and many of the residues in these sites are phylogenetically conserved; however, any biological significance of these additional ligand-binding sites is unknown.

Based on the gel-shift activity of *Mth* and *Pae* SmAP1 on supercoiled DNA (Fig. 7) and the striking similarity of SmAP monomers to the OB fold, we propose that SmAPs may have a generic single-stranded nucleic acid-binding activity (e.g., as a nucleic acid chaperone). We found that SmAPs nonspecifically gel-shift a variety of negatively supercoiled DNA substrates and that ssDNA oligonucleotides of >20 nt inhibit the gel shift (Fig. 7B). Because eukaryotic Sm proteins bind to ssRNA, and because SmAP homoheptamers probably do not function identically to eukaryotic Sm heteroheptamers, it is possible that this gel-shift inhibition results from direct binding of the oligonucleotides to SmAP1. The striking resemblance of the SmAP and OB folds corroborates this idea, given that several OB-fold proteins bind to ssDNA nonspecifically. The following recently determined structures are highly similar or identical to the OB-like fold of Sm proteins: the single-stranded DNA-binding domain of replication factor A (Bockkareva et al. 2001); the S1 RNA-binding domain (Bycroft et al. 1997); the single-stranded telomeric DNA binding protein (Mitton-Fry et al. 2002); and the *Streptococcus pneumoniae* SP14.3 protein (which is fused to a domain that is homologous to ribosomal protein S3; Yu et al. 2001).

Emerging differences between SmAPs and eukaryotic Sm proteins

Eukaryotic Sm and Lsm proteins and their archaeal homologs, which we term Sm-like archaeal proteins, share a number of structural and functional features. Perhaps the most significant similarity is in their 3D and quaternary structures: The monomers are nearly identical, and the SmAP homoheptamer parallels the Sm heteroheptamer that forms snRNP cores. Also, both sets of proteins apparently bind oligouridine-containing RNA. However, several differences are emerging between SmAPs and the snRNP-based roles of canonical, eukaryotic Sm proteins. The results presented here show that SmAPs associate into many oligomeric states besides the standard heptamer (e.g., 14-

mers and subheptamers), and can polymerize into homogeneous fibers. No structural information is available for Sm proteins bound to RNA (or any other ligand), and thus it is difficult to evaluate the similarity of uridine binding by eukaryotic Sm proteins and SmAPs. Cross-linking experiments with human Sm heptamers corroborate RNA binding near the pore (Urlaub et al. 2001). The near identity of the Sm and SmAP dimer structures, as well as the strictly conserved mode of uridine recognition between *Afu* and *Mth* SmAPI, suggest that the SmAPI UMP-binding site is an accurate model for RNA binding in the snRNP core. In this model, snRNA wraps around the circumference of the pore, but does not thread through it. Further elucidation of the similarities and differences between archaeal SmAP complexes and the Sm cores of eukaryotic snRNPs will provide insight into the structures and evolution of snRNPs.

Materials and methods

Cloning, expression, and purification of Pae and Mth SmAPIs

A genomic phosmid clone that contains the *Pae* SmAPI open reading frame (ORF) and genomic DNA containing the *Mth* (strain Δ H) SmAPI ORF were kindly provided by the laboratories of Jeffrey H. Miller (UCLA) and John Reeve (Ohio State Univ.), respectively. Primers were designed based on these sequences, and PCR products were cloned into a pET-based expression vector. DNA sequencing of plasmids verified that expressed constructs would contain a C-terminal His6 \times -tag after a 10-residue serine protease-sensitive linker, that is, wild-type (*wt*) SmAPI + GR*GKLAALAEHHHHHH (* indicates intended protease site). Recombinant proteins were overexpressed in BL21(DE3) *E. coli*, with at least 120 mg of soluble protein produced per liter of cell culture. The Cys8 \rightarrow Ser mutant of *Pae* SmAPI was created in a similar manner, except that site-directed mutagenesis was achieved via overlap-extension PCR with an additional pair of primers that contained the mutant site.

Harvested cells were thawed and resuspended in a high-salt-concentration buffer, and were lysed by French-press and lysozyme treatment. Cleared supernatants were heated to \sim 80°C, followed by high-speed centrifugation (37,000g). SmAPI-His6 \times proteins were further purified by affinity chromatography on Ni²⁺-charged iminodiacetic acid-sepharose, which afforded >99% purity (as determined by SDS-PAGE and MALDI-TOF mass spectrometry). Because the His6 \times tag prevents heptamer formation for some SmAPs (C. Mura, unpubl.), the next step was proteolytic removal of the C-terminal tag and its linker (*wt Mth* SmAPI is 81 aa, with a M_r of 9029 Da; *wt Pae* SmAPI is 80 aa, with a M_r of 8800 Da). Trypsin was used for limited proteolysis, as thrombin was ineffective: Ni²⁺-column fractions were pooled and dialyzed at room temperature into a phosphate-buffered saline buffer supplemented with 15 mM EDTA (to prevent His-tag mediated aggregation). EDTA was gradually eliminated over 2–3 buffer exchanges, and porcine trypsin was added at \sim 1 mg trypsin per 100 mg SmAPI. Complete removal of the tag occurred after \sim 4 h at 37°C, as assayed by MALDI-TOF spectra of time points. Transfer to 4°C and addition of a protease inhibitor (50 mM PMSF) terminated the reaction. Isoelectric points of \sim 5.2 and 5.8 were calculated for *Mth* and *Pae* SmAPI, respectively; therefore, anion ex-

change chromatography on a quaternary ammonium matrix (UNO-Q6, BioRad) was used to separate cut (i.e., *wt*) SmAPI from trypsin, uncut protein, and any other contaminants. *Pae* SmAPI was in 20 mM Tris, pH 8.55 and *Mth* SmAPI was in 20 mM Tris pH 8.55, 30 mM EDTA pH 8.0 (EDTA was required for solubility, and did not interfere with chromatography). Both SmAPIs eluted at \sim 80 mM NaCl in the salt concentration gradient. Protein purity was assayed by SDS-PAGE and MALDI-TOF, and pure fractions were pooled and dialyzed into a buffer for crystallization.

Crystallization and data collection

For crystallization, *Pae* SmAPI was in buffer “XB” (10 mM Tris pH 7.8, 5 mM EDTA pH 8.0), and *Mth* (which requires higher ionic strengths for solubility) was in “XB6 β ” (10 mM Tris pH 7.8, 5 mM EDTA pH 8.0, and 0.1 M NaCl). Protein concentrations in these buffers were increased by using Centripreps to reduce sample volumes. After initial sparse matrix screening of conditions, final, optimized *Pae* SmAPI crystals of the C222₁ form were grown by hanging-drop vapor diffusion. An 11- μ L drop [4 μ L well buffer + 5 μ L *wt* 29.6 mg/mL *Pae* SmAPI + 1 μ L 0.1 M dithiothreitol (DTT) + 1 μ L 0.1 M uridine-5'-monophosphate (UMP)] was equilibrated against an 800- μ L well [0.1 M sodium acetate pH 8.20, 0.1 M ammonium acetate, 8.6% w/v PEG-4000, and 23.8% v/v glycerol] at room temperature (\sim 19.8°C). Orthorhombic crystals reached maximum dimensions of 0.1 \times 0.1 \times 0.3 mm within 5 d. Hanging drops contained a mixture of the new C222₁ crystals and the previously reported C2 form (Mura et al. 2001).

Three forms of *Mth* SmAPI crystals were obtained by hanging-drop vapor diffusion at room temperature. For the P1 form, *Mth* SmAPI was at 56 mg/mL in buffer XB6 β . The drop was 4 μ L of protein + 4 μ L of well buffer. The well was 600 μ L of [0.1 M sodium citrate pH 5.60, 15% w/v PEG-4000, 0.2 M ammonium acetate]. Crystals grew to maximum dimensions of \sim 0.1 \times 0.1 \times 0.25 mm within 7 d. For the P2₁2₁2₁ form, *Mth* SmAPI was at 42 mg/mL in buffer XB6 β . The drop was 3 μ L of protein + 3 μ L of well buffer. The well was 600 μ L of [0.1 M Tris pH 8.50, 10% v/v isopropanol]. Crystals grew to maximum dimensions of \sim 0.3 \times 0.3 \times 0.6 mm within 3 d. For the P2₁ form, *Mth* SmAPI was at 30.3 mg/mL in a modified form of buffer XB6 β that contained a 26-nt single-stranded DNA [10 mM Tris pH \sim 7.7, 3 mM EDTA pH 8.0, 55 mM NaCl, 0.6 mM ssDNA]. Drops were 2.5 μ L of protein/ssDNA + 2.5 μ L of well buffer + 1 μ L of 0.1 M UMP. The 600- μ L well contained 55 μ L of 1.0 M sodium citrate (pH 5.6), 5 μ L of 1.0 M sodium citrate (pH 8.0), 60 μ L of 2.0 M ammonium acetate, 180 μ L of neat MPD and 300 μ L of sterile dH₂O (interestingly, 2.5 M 1,6-hexanediol could be substituted for neat MPD). Crystals grew to maximum dimensions of \sim 0.15 \times 0.15 \times 0.25 mm within 7 d.

The C222₁ *Pae* SmAPI and P2₁ *Mth* SmAPI crystals did not require additional cryosolvent, due to the 23.8% v/v glycerol or 30% v/v MPD in those drops, respectively. The other two *Mth* SmAPI crystal forms were cryoprotected as follows: (1) for the P1 form, ethylene glycol was added directly to the drop to a final concentration of \sim 20% v/v, and crystals were allowed to soak for 20 sec prior to mounting in a cryoloop; (2) for the fragile P2₁2₁2₁ crystals, the cryoprotectant was ethylene glycol (mixed with well buffer), and had to be introduced gradually over several hours (in \sim 5% v/v increments). The P2₁2₁2₁ crystals were soaked for only \sim 2–3 sec at the final ethylene glycol concentration (30% v/v). Diffraction data were collected either at the synchrotron (P1 and P2₁ form *Mth* crystals) or in-house (P2₁2₁2₁ *Mth* and C222₁ *Pae* crystals) on an ADSC Quantum-4 charge-coupled device (CCD)

detector. All crystals were mounted in a cryogenic nitrogen stream at -168°C for data collection. After autoindexing, images were indexed/integrated/reduced in DENZO, and reflections were scaled and merged in SCALEPACK (Otwinowski and Minor 1997). Complete data sets were collected from single crystals (Table 1). Unit cell dimensions for the *Pae* C222₁ form are $a = 91.83$, $b = 113.76$, $c = 126.59$ Å; for the *Mth* crystals they are: $a = 45.07$, $b = 54.08$, $c = 62.35$ Å, $\alpha = 87.58^{\circ}$, $\beta = 72.86^{\circ}$, $\gamma = 81.45^{\circ}$ (*P1*); $a = 65.25$, $b = 109.96$, $c = 83.76$ Å, $\beta = 95.81^{\circ}$ (*P2*₁); $a = 40.37$, $b = 114.70$, $c = 238.60$ Å (*P2*₁2₁2₁). The large unit cell edge of the *Mth* *P2*₁2₁2₁ crystals led to spot overlap for high-resolution reflections ($d < 3$ Å), so multiple data sets were collected at two 2θ values (0° , -12°) for two crystal alignments (related by a 45° azimuthal rotation).

Structure determination, refinement, and validation

Initial phases for the C222₁ *Pae* SmAP1 structure were determined by the stochastic evolutionarily programmed molecular replacement method (EPMR; Kissinger et al. 2001). The most reasonable Matthews coefficient ($V_M = 2.58$ Å³/Da) corresponds to a heptamer in the asymmetric unit (a.u.); therefore, the search model was the identical *Pae* SmAP1 heptamer from the C2 crystal form (Mura et al. 2001). The EPMR solution was used for manual model building in the program O (Jones et al. 1991), and model refinement in CNS (Brunger et al. 1998). Refinement in CNS proceeded by standard protocols, using the maximum-likelihood target function for amplitudes (mLf), bulk solvent correction, and anisotropic B-factor correction terms. Sevenfold noncrystallographic symmetry (NCS) was determined by calculation of a locked self-rotation function, but NCS restraints were not imposed during refinement. Solvent molecules were added as necessary (water, glycerol, acetate). Refinement of individual atomic positions, isotropic temperature factors, and simulated annealing torsion angle dynamics was performed in most rounds. Each refinement round ended with inspection of the agreement between the model and σ_A -weighted $2|F_o| - |F_c|$, $|F_o| - |F_c|$, and, when needed, simulated annealing omit maps.

Determination of the *Mth* *P1* structure proceeded in two steps. First, a homology model of the *Mth* SmAP1 heptamer was built from the *Pae* SmAP1 structure using an in-house script (C. Mura and D. Eisenberg, unpubl.), and was used as a search model for molecular replacement ($V_M = 2.29$ Å³/Da for one heptamer in the *P1* cell). Then, the EPMR solution was converted to a polyaniline model and subjected to free-atom model refinement with the ARP/wARP program (Perrakis et al. 1999) in the "molrep" mode. *Mth* SmAP1 side chains were built in the final wARP stage. The *Mth* *P1* structure was refined with CNS, as described above for the *Pae* structure. The *P2*₁ and *P2*₁2₁2₁ *Mth* structures were solved by molecular replacement with the refined *P1* *Mth* model. Self-rotation functions and $|F_o|^2$ Patterson maps were calculated to deduce the NCS between heptamers in the *P2*₁ and *P2*₁2₁2₁ forms (each of which contains 14 monomers per a.u.). Solvent was added as necessary for all structures (see Table 1), and no NCS restraints were enforced at any point in the refinements. Partial atomic occupancies (q) were restricted to a reasonable range ($0.2 < q < 1.5$) during latter refinement rounds, in which only the occupancies for atoms of UMP (not for any other ligand or protein atoms) were refined.

Refinement statistics for the *Pae* and three *Mth* structures are shown in Table 1. Each of the four protein models is complete, except for ~6–11 missing N-term residues in various models (see PDB files). The stereochemistry and geometry of each SmAP1 monomer was validated with PROCHECK (Laskowski et al. 1993) and ERRAT (Colovos and Yeates 1993), and was found to be

acceptable (e.g., no residues in the disallowed region of ϕ, ψ space for the *Pae* C222₁ model). Final model coordinates and diffraction intensity data were submitted to the PDB with ID codes 1JBM, 1LOJ, 1JRI, and 1LNX (see Table 1).

Analytical ultracentrifugation

The *wt Pae* protein in 75 mM NaCl, 10 mM Tris, pH 7.8, was examined by sedimentation velocity in a Beckman Optima XL-A analytical ultracentrifuge at 52,000 rpm and 20°C using absorption optics at 273 nm and a 12-mm pathlength double sector cell. The sedimentation coefficient distribution was determined from a $g(s)$ plot using the Beckman Origin-based software (Version 3.01). The peak sedimentation coefficient was corrected for density and viscosity to an $S_{20, \text{wat}}$ value by using a value for the partial specific volume at 20°C of 0.743 [calculated from the amino acid composition (Edsall 1943) and corrected to 20°C (Laue et al. 1992)].

Sedimentation equilibrium runs at 20°C were performed on all three proteins—*wt Mth*, *wt Pae*, and the *Pae* C8S mutant—in 150 mM NaCl, 10 mM Tris, pH 7.8, again using a Beckman Optima XL-A analytical ultracentrifuge. Each protein was examined at three different concentrations and four speeds, using 12-mm pathlength six-sector cells. Protein concentrations used were 3.4, 0.69, and 0.19 mg/mL for *wt Pae*; 5.9, 1.26, and 0.32 mg/mL for the C8S mutant of *Pae*; and 4.1, 0.85, and 0.22 mg/mL for *wt Mth*. Rotor speeds were 8,000, 10,000, 12,500, and 14,500 rpm. Protein concentration was monitored by absorption at 280 nm and, for the lowest protein concentrations, at 232 nm. A partial specific volume of 0.743, calculated as described above, was used for all three proteins. Individual scans were analyzed using the Beckman Origin-based software (Version 3.01) to perform a nonlinear least-squares exponential fit for a single ideal species, thus giving the weight-averaged molecular weight for each protein.

Transmission electron microscopy

The following protein samples were prepared for electron microscopy: (1) 0.5 mg/mL wild-type *Mth* SmAP1 in 10 mM Tris pH 7.5, 60 mM NaCl; (2) 1.2 mg/mL wild-type *Pae* SmAP1 in 25 mM Tris pH 7.5, 30 mM NaCl; (3) 1.1 mg/mL C8S mutant *Pae* SmAP1 in the same buffer as the *wt* protein; and (4) 1.2 mg/mL *wt Pae* SmAP1 in reductant buffer (25 mM Tris pH 7.5, 30 mM NaCl, 10 mM DTT). Carbon-coated parlodion support films mounted on copper grids were made hydrophilic immediately before use by high-voltage, alternating-current glow discharge. Protein samples were applied directly onto the grids and allowed to adhere for 2 min. Grids were rinsed with distilled water and negatively stained with 1% w/v uranyl acetate. Specimens were examined in a Hitachi H-7000 electron microscope at an accelerating voltage of 75 kV.

Gel-shift assays

Negatively supercoiled plasmid DNA was prepared by transforming the plasmid into *E. coli* BL21(DE3) cells and mini-prepping (Qiagen) it from spun-down cells that had reached stationary phase. Several different plasmids were tested, including ones derived from pUC18, pACYC, pET-22b(+) (Novagen), and pCR-Blunt (Invitrogen). Titration of plasmids with ethidium bromide was used to verify the negative superhelicity of the DNA via electrophoretic mobility changes in agarose gels. Single-stranded DNAs of various lengths and sequences were synthesized by In-

tegrated DNA Technologies (e.g., the 26-mer in Fig. 7B with the following sequence: 5'CGGATCCTCAGTAAAAAGTGCAGAA3'). Stock solutions of protein were *wt Pae* at 5.6 mg/mL in buffer XB (see above) or *wt Mth* at 5.6 mg/mL in buffer XB6 β (see above). Except as noted, buffer, DNA, and protein samples were mixed to produce 25- or 50- μ L reactions that were incubated at room temperature (generally for 30–60 min). Gel-shift of the DNA was assayed by electrophoresis at a constant voltage (120V) in 1.3% or 1.5% w/v TAE/agarose gels. Examples of typical reactions and concentration ranges are given in Figure 7. Reactions in which SmAP1 was replaced by single-stranded DNA-binding protein (Stratagene) or by arbitrary *Pae* proteins unrelated to SmAP1 (e.g., an acid phosphatase) served as positive and negative controls, respectively.

Acknowledgments

We thank Dr. Duilio Cascio and Dr. Michael Sawaya (UCLA) for data collection at beamline X8C of the National Synchrotron Light Source (Brookhaven National Lab), Dr. Lukasz Salwinski and Dr. Todd Yeates for discussions, and NIH and DOE-BER grants for funding. D.E. is an investigator for the HHMI.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Note added in proof

Thore et al. (2003) recently reported the structure of a *P. abyssi* SmAP that also forms 14-mers (in the crystal) and contains a secondary RNA-binding site (which differs from that of *Pae* SmAP1). Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. 2003. Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. *J. Biol. Chem.* **278**: 1239–1247.

References

- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, R. 1999. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.* **18**: 5789–5802.
- Achsel, T., Stark, H., and Luhrmann, R. 2001. The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proc. Natl. Acad. Sci.* **98**: 3685–3689.
- Antson, A.A., Dodson, E.J., Dodson, G., Greaves, R.B., Chen, X., and Gollnick, P. 1999. Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**: 235–242.
- Arluisson, V., Derreumaux, P., Allemand, F., Folichon, M., Hajnsdorf, E., and Regnier, P. 2002. Structural modeling of the Sm-like protein Hfq from *Escherichia coli*. *J. Mol. Biol.* **320**: 705–712.
- Bochkareva, E., Belegu, V., Korolev, S., and Bochkarev, A. 2001. Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding. *EMBO J.* **20**: 612–618.
- Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., and Luhrmann, R. 2000. The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J. Biol. Chem.* **275**: 17122–17129.
- Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Luhrmann, R. 2001. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA* **7**: 1531–1542.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.
- Burge, C.B., Tuschl, T., and Sharp, P.A. 1999. Splicing of precursors to mRNAs by the spliceosome. In *The RNA world*, 2nd ed. (eds. R.F. Gesteland et al.), pp. 525–560. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Bycroft, M., Hubbard, T.J., Proctor, M., Freund, S.M., and Murzin, A.G. 1997. The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell* **88**: 235–242.
- Collins, B.M., Harrop, S.J., Kornfeld, G.D., Dawes, I.W., Curmi, P.M., and Mabbitt, B.C. 2001. Crystal structure of a heptameric Sm-like protein complex from archaea: Implications for the structure and evolution of snRNPs. *J. Mol. Biol.* **309**: 915–923.
- Colovos, C. and Yeates, T.O. 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* **2**: 1511–1519.
- Edsall, E.J.C.J.T. 1943. Density and apparent specific volume of proteins. In *Proteins, amino acids, and peptides as ions and dipolar ions*. (ed. E.J.C.J.T. Edsall), pp. 370–381. Reinhold Publishing Corp., New York.
- Friesen, W.J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G.S., Van Duyne, G., Rappsilber, J., Mann, M., and Dreyfuss, G. 2001. The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol. Cell. Biol.* **21**: 8289–8300.
- Friesen, W.J., Wyce, A., Paushkin, S., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. 2002. A novel WD repeat protein component of the methylosome binds Sm proteins. *J. Biol. Chem.* **277**: 8243–8247.
- Fromont-Racine, M., Mayes, A.E., Brunet-Simon, A., Rain, J.C., Colley, A., Dix, I., Decourty, L., Joly, N., Ricard, F., Beggs, J.D., et al. 2000. Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast* **17**: 95–110.
- Jarmolowski, A. and Mattaj, I.W. 1993. The determinants for Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and nonidentical. *EMBO J.* **12**: 223–232.
- Jones, M.H. and Guthrie, C. 1990. Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: Genetic analysis of the Sm binding site. *EMBO J.* **9**: 2555–2561.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**: 110–119.
- Kambach, C., Walke, S., and Nagai, K. 1999a. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.* **9**: 222–230.
- Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann, R., Li, J., and Nagai, K. 1999b. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**: 375–387.
- Kastner, B., Bach, M., and Luhrmann, R. 1990. Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: Evidence for a common structure-determining principle in the major U snRNP family. *Proc. Natl. Acad. Sci.* **87**: 1710–1714.
- Kissinger, C.R., Gehlhaar, D.K., Smith, B.A., and Bouzida, D. 2001. Molecular replacement by evolutionary search. *Acta Crystallogr. D Biol. Crystallogr.* **57**: 1474–1479.
- Laskowski, R.A., Moss, D.S., and Thornton, J.M. 1993. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**: 1049–1067.
- Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical ultracentrifugation in biochemistry and polymer science*. (eds. S.E. Harding et al.), pp. 90–125. The Royal Society of Chemistry, Cambridge, England.
- Long, M., de Souza, S.J., and Gilbert, W. 1995. Evolution of the intron-exon structure of eukaryotic genes. *Curr. Opin. Genet. Dev.* **5**: 774–778.
- Mallick, P., Boutz, D.R., Eisenberg, D., and Yeates, T.O. 2002. Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proc. Natl. Acad. Sci.* **99**: 9679–9684.
- Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C., and Fischer, U. 2001. Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* **11**: 1990–1994.
- Mitton-Fry, R.M., Anderson, E.M., Hughes, T.R., Lundblad, V., and Wuttke, D.S. 2002. Conserved structure for single-stranded telomeric DNA recognition. *Science* **296**: 145–147.
- Moller, T., Franch, T., Hojrup, P., Keene, D.R., Bachinger, H.P., Brennan, R.G., and Valentin-Hansen, P. 2002. Hfq: A bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol. Cell* **9**: 23–30.
- Muller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R. 1998. A supraspliceosome model for large nuclear ribonucleopro-

- tein particles based on mass determinations by scanning transmission electron microscopy. *J. Mol. Biol.* **283**: 383–394.
- Mura, C., Cascio, D., Sawaya, M.R., and Eisenberg, D.S. 2001. The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. *Proc. Natl. Acad. Sci.* **98**: 5532–5537.
- Murzin, A.G. 1993. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for nonhomologous sequences. *EMBO J.* **12**: 861–867.
- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. In *Methods in enzymology*. (eds. R. Sweet and C.W. Carter, Jr.), pp. 307–326. Academic Press, San Diego, CA.
- Perrakis, A., Morris, R., and Lamzin, V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**: 458–463.
- Plessel, G., Luhrmann, R., and Kastner, B. 1997. Electron microscopy of assembly intermediates of the snRNP core: Morphological similarities between the RNA-free (E.F.G.) protein heteromer and the intact snRNP core. *J. Mol. Biol.* **265**: 87–94.
- Raker, V.A., Plessel, G., and Luhrmann, R. 1996. The snRNP core assembly pathway: Identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *EMBO J.* **15**: 2256–2269.
- Raker, V.A., Hartmuth, K., Kastner, B., and Luhrmann, R. 1999. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol.* **19**: 6554–6565.
- Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Seraphin, B. 1999. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.* **18**: 3451–3462.
- Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. 2001. SMN tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol.* **8**: 27–31.
- Stark, H., Dube, P., Luhrmann, R., and Kastner, B. 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**: 539–542.
- Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. 2002. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* **9**: 31–44.
- Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B., and Suck, D. 2001. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.* **20**: 2293–2303.
- Toro, I.I., Basquin, J., Teo-Dreher, H., and Suck, D. 2002. Archaeal Sm proteins form heptameric and hexameric complexes: Crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. *J. Mol. Biol.* **320**: 129–142.
- Urlaub, H., Raker, V.A., Kostka, S., and Luhrmann, R. 2001. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **20**: 187–196.
- Walke, S., Bragado-Nilsson, E., Seraphin, B., and Nagai, K. 2001. Stoichiometry of the Sm proteins in yeast spliceosomal snRNPs supports the heptamer ring model of the core domain. *J. Mol. Biol.* **308**: 49–58.
- Wallace, A.C., Laskowski, R.A., and Thornton, J.M. 1995. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**: 127–134.
- Will, C.L. and Luhrmann, R. 1997. Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**: 320–328.
- . 2001. Spliceosomal UsnRNP biogenesis, structure, and function. *Curr. Opin. Cell Biol.* **13**: 290–301.
- Yu, L., Gunasekera, A.H., Mack, J., Olejniczak, E.T., Chovan, L.E., Ruan, X., Towne, D.L., Lerner, C.G., and Fesik, S.W. 2001. Solution structure and function of a conserved protein SP14.3 encoded by an essential *Streptococcus pneumoniae* gene. *J. Mol. Biol.* **311**: 593–604.
- Yu, Y.-T., Scharl, E.C., Smith, C.M., and Steitz, J.A. 1999. The growing world of small nuclear ribonucleoproteins. In *The RNA world*, 2nd ed. (eds. R.F. Gesteland et al.), pp. 487–524. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C., and Storz, G. 2002. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell* **9**: 11–22.