

# The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core

Cameron Mura\*<sup>†</sup>, Duilio Cascio\*, Michael R. Sawaya\*, and David S. Eisenberg\*<sup>††</sup>

\*University of California at Los Angeles—Department of Energy Laboratory of Structural Biology and Molecular Medicine, 201 Boyer Hall/Molecular Biology Institute, Box 951570, Los Angeles, CA 90095-1570; and <sup>††</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095

Contributed by David S. Eisenberg, March 1, 2001

**Sm proteins form the core of small nuclear ribonucleoprotein particles (snRNPs), making them key components of several mRNA-processing assemblies, including the spliceosome. We report the 1.75-Å crystal structure of SmAP, an Sm-like archaeal protein that forms a heptameric ring perforated by a cationic pore. In addition to providing direct evidence for such an assembly in eukaryotic snRNPs, this structure (i) shows that SmAP homodimers are structurally similar to human Sm heterodimers, (ii) supports a gene duplication model of Sm protein evolution, and (iii) offers a model of SmAP bound to single-stranded RNA (ssRNA) that explains Sm binding-site specificity. The pronounced electrostatic asymmetry of the SmAP surface imparts directionality to putative SmAP–RNA interactions.**

**E**ukaryotic pre-mRNA processing is an intricate cellular task whose many steps include intron excision. This final maturation step occurs in the spliceosome, a large ( $\approx 60$  S), transiently stable ribonucleoprotein particle that ligates two exons and releases an intron lariat. The major spliceosome contains several small nuclear ribonucleoprotein particles (snRNPs; e.g., U1, U2, and U4/U6/U5), and each U snRNP consists of a respective small nuclear RNA (snRNA; e.g., U1, U2) and many proteins (reviewed in refs. 1 and 2). The subset of proteins common to all spliceosomal U snRNPs is the Sm proteins. Discovered as a group of eight small antigens involved in autoimmune diseases such as systemic lupus erythematosus (3), these core snRNP Sm proteins have been found in most eukaryotes (4) and, recently, in a few archaeal species (ref. 5 and this report). Also, several sets of Sm-like (Lsm) proteins have been discovered in organisms with Sm proteins (4, 6). Biochemical characterization of Lsm proteins has verified their similarity to canonical Sm proteins (reviewed in ref. 7), and this report treats them as equivalent.

Together with U snRNAs, canonical Sm proteins form snRNP core complexes. The molecular structure and function(s) of Sm assemblies and Sm–RNA interactions within these core complexes are unknown, but are presumably more generic than those of snRNP-specific proteins. Cytoplasmic Sm proteins associate with exported snRNAs at short single-stranded regions that are usually flanked by stem-loop structures (8, 9). The consensus sequence for the Sm binding site is RAU<sub>4–6</sub>GR (R = purine), although this selectivity is not very stringent (10, 11). The core complex is thought to be a heteroheptamer of Sm proteins, and electron microscopic investigations of U snRNP core particles suggest that the Sm (12) and Lsm (13) cores are composed of a doughnut-shaped heteromer. The current paradigm is that seven Sm proteins (e.g., human B/B', D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, E, F, and G) assemble stepwise into a heteroheptameric ring with snRNA through various intermediates, such as D<sub>1</sub>·D<sub>2</sub> and E·F·G heteromers. This snRNP core complex is then imported to the nucleus for completion of spliceosome assembly. Functional complexes of RNA and homologous Sm protein septets, such as the Lsm 1–7 and Lsm 2–8 sets of yeast, are thought to assemble in a similar manner.

There are no atomic resolution structures of snRNP cores, although a recent electron cryomicroscopic study by Stark *et al.* (14) underscores the importance of Sm proteins in forming the core of the U1 snRNP. Crystal structure determinations of D<sub>1</sub>·D<sub>2</sub> and D<sub>3</sub>·B heterodimers by Kambach *et al.* (15, 16) provide the only known Sm structures, and show that each monomer folds as a strongly bent, five-stranded antiparallel  $\beta$ -sheet. These monomers have nearly identical three-dimensional structures, as do the heterodimers. Kambach *et al.* used their heterodimer structures to model an Sm heteroheptamer with a positively charged central hole. We now report the 1.75-Å crystal structure of Sm-like archaeal protein (SmAP)—a heptameric Sm protein from the hyperthermophilic archaeon *Pyrobaculum aerophilum*—and describe the implications of this structure for the Sm core of eukaryotic snRNPs.

## Materials and Methods

**Protein Preparation and Crystallization.** *P. aerophilum* SmAP was cloned and over-expressed in *Escherichia coli* cells. Heat treatment of lysed cells (80°C) was followed by standard chromatographic steps and proteolytic removal of a C-terminal His<sub>6</sub> tag. Final purification steps yielded full-length, wild-type SmAP with an appended glycine. Monoclinic crystals of SmAP (space group C<sub>2</sub>;  $a = 100.26$  Å,  $b = 95.74$  Å,  $c = 62.16$  Å,  $\beta = 92.69^\circ$ ;  $V_M = 2.33$  Å<sup>3</sup>/Da for seven monomers per asymmetric unit) were used to solve the structure reported here by multiwavelength anomalous diffraction (MAD) phasing of an iridium derivative (Na<sub>3</sub>IrCl<sub>6</sub>).

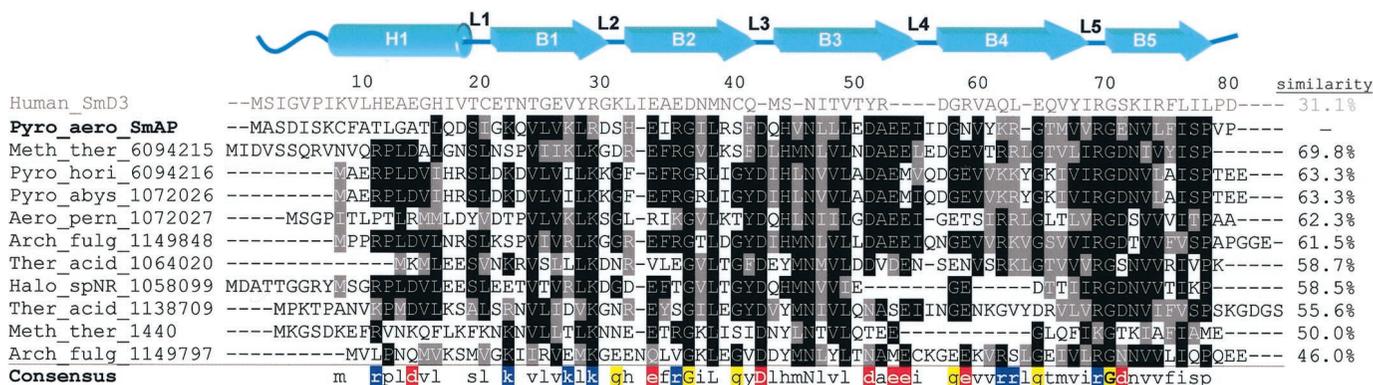
**Data Collection, Phasing, and Refinement.** Synchrotron data were collected at 105 K. Data were processed with DENZO/SCALEPACK (17), and MAD phasing proceeded by the usual methods of heavy atom location [SHELXD (<http://shelx.uni-ac.gwdg.de/SHELX/>) and SOLVE (18)], maximum likelihood phase refinement [MLPHARE (19)], and density modification [DM (20)]. Phase extension to 1.75 Å permitted automated model building for most of the protein with WARP/ARP (21). Model building and refinement were done in O (22) and CNS (23), respectively. Averaging with sevenfold noncrystallographic symmetry led us to conclude that the only significant asymmetry within the heptameric ring is attributable to side-chain rotameric variation (each monomer of the final model was refined independently). Refinement rounds ended with inspection of the model and  $\sigma_A$ -weighted  $2F_o - F_c$ ,  $F_o - F_c$ , and simulated annealing omit

Abbreviations: snRNP, small nuclear ribonucleoprotein; snRNA, small nuclear RNA; Lsm, Sm-like; SmAP, Sm-like archaeal protein; MAD, multiwavelength anomalous diffraction; ssRNA, single-stranded RNA; rmsd, root mean square deviation.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, [www.rcsb.org](http://www.rcsb.org) (PDB ID code 1IBF).

<sup>†</sup>To whom reprint requests should be addressed. E-mail: [david@mbl.ucla.edu](mailto:david@mbl.ucla.edu).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



**Fig. 1.** Sequence analysis of SmAPs. A multiple sequence alignment of SmAPs is shown, along with the sequence of the most similar known Sm structure (human D<sub>3</sub>) as a reference point (residue numbering is for the *P. aerophilum* sequence). Pairwise sequence similarity scores between SmAP and its homologs are provided in the last column. The top diagram depicts the SmAP monomer in terms of secondary structure elements. Glycine (yellow), acidic (red), and basic (blue) residues are highlighted in the consensus sequence and are referred to in the text and in Fig. 2c.

maps (the latter only as necessary). The final model consists of a SmAP heptamer (seven chains, labeled A–G), 130 waters, and five glycerols. Each monomer is complete except for (i) the absence of 4 (monomer F) to 13 (monomer C) N-terminal residues per monomer, and (ii) an average of 3 residues per monomer that are truncated to either alanine or glycine. Ramachandran plots (PROCHECK, ref. 24) and ERRAT (25) were used for model validation. Structure factors and atomic coordinates have been deposited in the Protein Data Bank (ID code 1I8F).

**Sequence and Structure Analyses.** PSI-BLAST (26) and CLUSTALW (27) were used for database searches and multiple sequence alignments, respectively. Pairwise alignments were calculated by the Smith–Waterman algorithm. Similar protein structures [rms deviation (rmsd) < ≈2.5 Å, e.g., human D<sub>2</sub>, D<sub>3</sub> proteins to SmAP] were easily superimposed with ALIGN (28), whereas more dissimilar structures (rmsd > ≈2.5 Å by ALIGN, e.g., human B, D<sub>1</sub> proteins) were optimally aligned by combinatorial extension (29). Electrostatic and surface area calculations were performed with GRASP (30). Figs. 2b, 2c, 3a, and 5 were produced with WEBLAB VIEWERPRO 3.7 (Molecular Simulations); Figs. 3 b and c were produced with SETOR (31, 32); and Fig. 4 was created in GRASP.

**The SmAP-ssRNA (Single-Stranded RNA) Model.** The SmAP-ssRNA model was constructed by manual docking of an ssRNA con-

taining an Sm consensus sequence (GAU<sub>4</sub>GA) through the pore of the refined SmAP heptamer, and did not require any alteration of the SmAP structure. Adjustment of only four phosphate backbone α torsion angles near the poly (U) tract sufficiently extended the RNA so that it was easily accommodated in the SmAP pore; moreover, the values were not adjusted drastically—from α ≈ −50° in A-form RNA to α ≈ −30° in the extended form shown in Fig. 5. The planes of the first and last uracil bases threaded through the pore are separated by roughly 17 Å.

**Results**

**A Family of SmAPs.** The occurrence of Sm and Sm-like proteins is not limited to eukaryotes. Along with *P. aerophilum* SmAP, we have uncovered several archaeal Sm sequences within the genomes of *Pyrococcus abyssi*, *Aeropyrum pernix*, *Thermoplasma acidophilum*, and a halobacterium (Fig. 1). Added to the initial list of five SmAPs reported by Salgado-Garrido *et al.* in 1999 (5), these archaeal sequences clearly form a well defined protein family that may be ancestral to modern eukaryotic Sm proteins. On the basis of the structure described here, we propose that *P. aerophilum* SmAP is a representative member of such a family of Sm-like archaeal proteins (SmAPs).

**SmAP Monomer and Dimer Structures.** The crystal structure of SmAP was determined by MAD phasing (Table 1), and reveals

**Table 1. Crystallographic statistics**

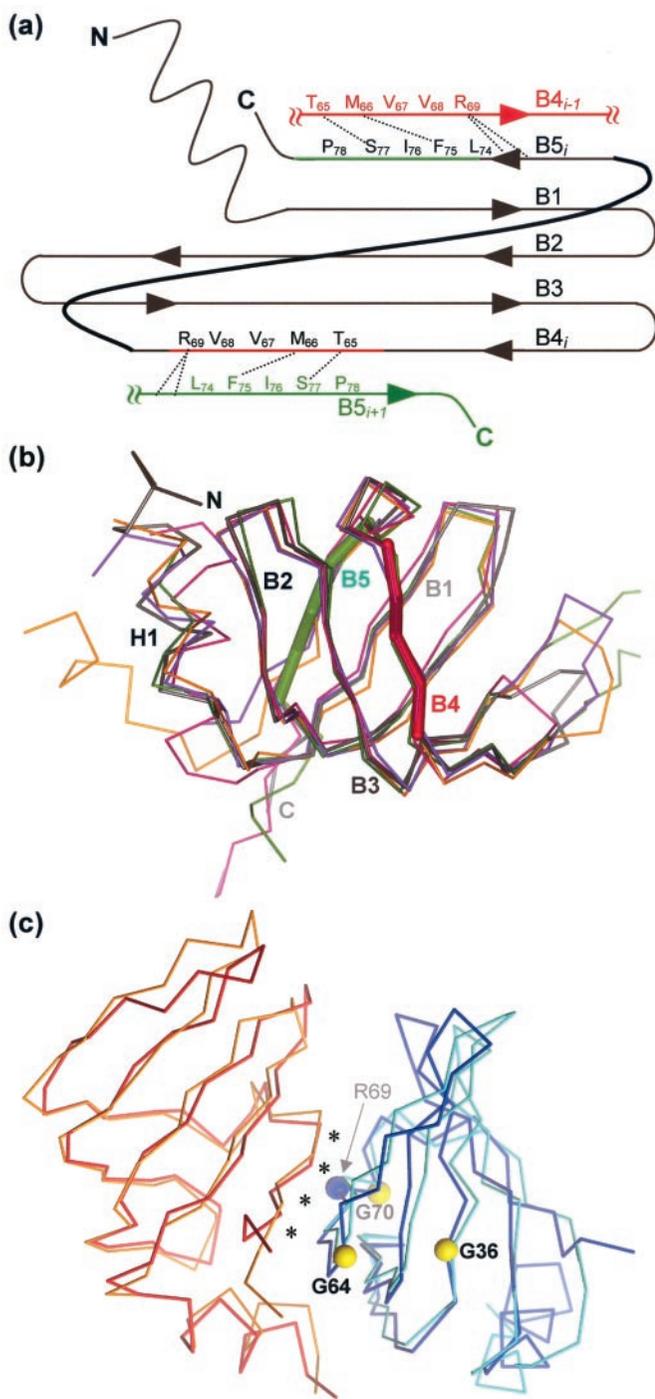
	Data collection & MAD phasing				Model refinement	
	Native	Inflection	Peak	Low-λ remote		
Wavelength, Å	0.9794	1.1058	1.1055	1.0960	Resolution range, Å	20–1.75
Resolution range, Å	100–1.71	100–1.95	100–1.95	100–1.95	No. of protein atoms	3.796
Completeness, %	98.5	97.3	97.0	95.0	No. of solvent molecules (H <sub>2</sub> O/glycerol)	130/5
I/σ(I)	44.4	24.3	25.2	22.1	R <sub>cryst</sub> /R <sub>free</sub> , % <sup>§</sup>	23.5/26.6
R <sub>merge</sub> , %*	4.0	4.8	4.8	4.0	⟨B-factor⟩ protein, Å <sup>2</sup>	38.5
# Ir sites/asymmetric unit	—	—	8	—	rmsds: bonds, Å	0.018
Phasing resolution range, Å	—	20–2.0	20–2.0	20–2.0	angles, °	1.90
R <sub>cullis</sub> <sup>†</sup> : acentric	—	—	0.71/0.42	0.83/0.70		
centric	—	—	0.64	0.82		
Figure of merit <sup>‡</sup>		0.65/0.82				

\*R<sub>merge</sub>(I) = Σ<sub>hkl</sub>(Σ<sub>i</sub>|I<sub>hkl,i</sub> - ⟨I<sub>hkl</sub>⟩)/Σ<sub>i</sub>I<sub>hkl,i</sub>.

†R<sub>cullis</sub> = (Σ<sub>hkl</sub>||F<sub>PH</sub> ± F<sub>P</sub>| - F<sub>H,calc</sub>)/Σ<sub>hkl</sub>||F<sub>PH</sub> ± F<sub>P</sub>|. Statistics for acentric reflections are given as isomorphous/anomalous.

‡Values are given before/after density modification.

§R<sub>cryst</sub> = Σ<sub>hkl</sub>||F<sub>obs</sub>| - |F<sub>calc</sub>||Σ<sub>hkl</sub>||F<sub>obs</sub>|. R<sub>free</sub> was computed identically, except that 5% of the reflections were omitted as a test set.



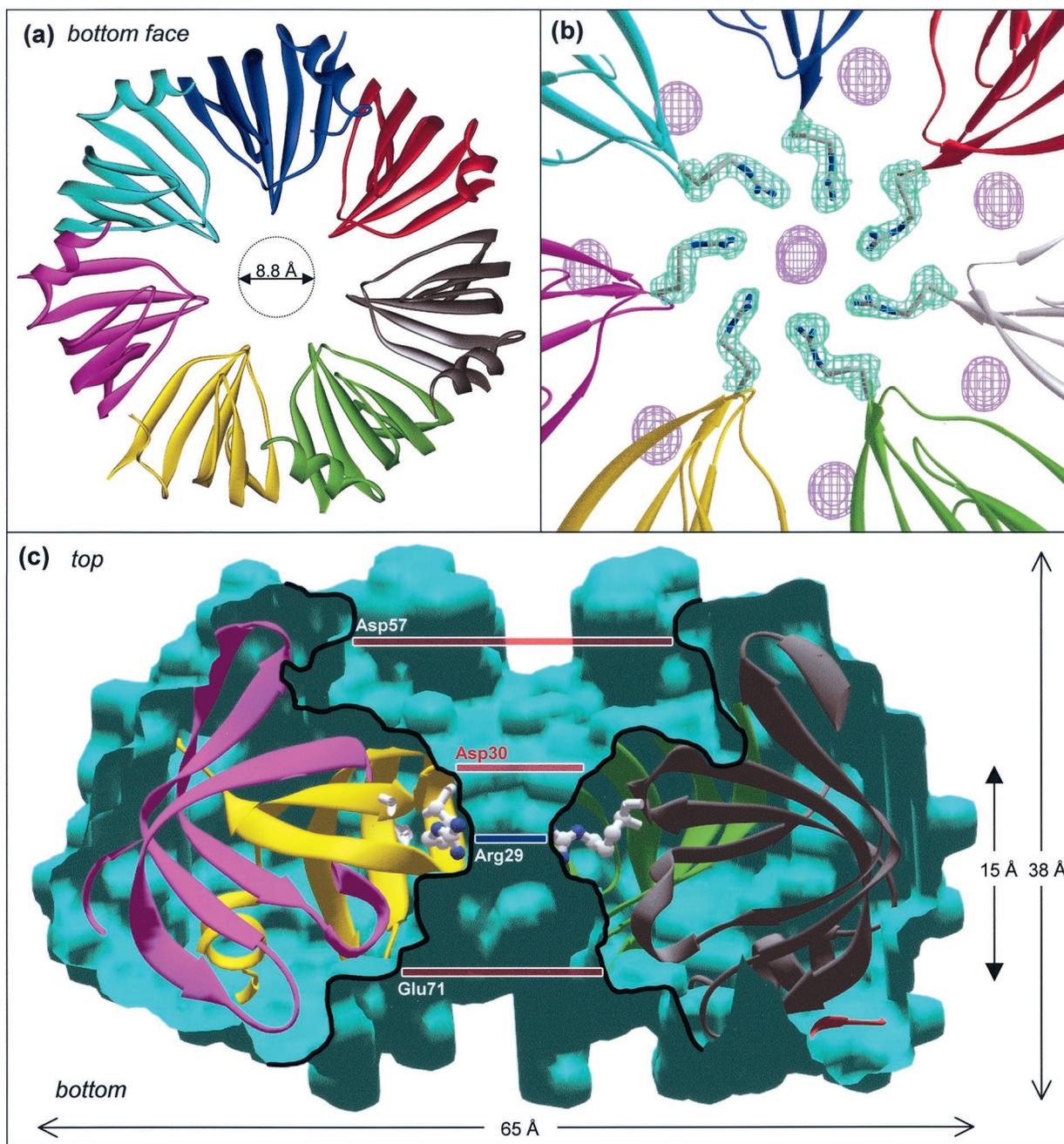
**Fig. 2.** SmAP monomer and dimer structures. (a) A cartoon of the SmAP fold, along with the B4 and B5 strands of the two neighboring monomers. Dashed lines indicate side-chain interactions that supplement the backbone hydrogen bonding of the B4–B5 pairs. (b) A depth-cued illustration of the structure of one of the SmAP monomers (gray) is superimposed on the  $C_{\alpha}$  traces of four aligned structures: human Sm D<sub>3</sub> (violet, 1.0 Å rmsd), B (green, 1.1 Å rmsd), D<sub>2</sub> (magenta, 1.2 Å rmsd), and D<sub>1</sub> (orange, 1.7 Å rmsd). The extensive L4 loop of Sm B has been truncated for clarity, and the segments of strands B4 and B5 that bind adjacent monomers are colored as thick green and red lines as in a. This orientation illustrates the strongly bent five-stranded antiparallel  $\beta$ -sheet that forms the Sm structures. (c) The human D<sub>3</sub>:B heterodimer (cyan-orange) is superimposed on an SmAP homodimer (red-blue). Asterisks mark the conserved dimer interface, and colored balls give the positions of conserved residues shown in Fig. 1. Note that the archaeal homodimer, taken directly from the SmAP heptamer crystal structure, has essentially the same structure as the human heterodimer (1.4 Å rmsd over main-chain atoms).

that this archaeal protein—like the human Sm B, D<sub>1</sub>, D<sub>2</sub>, and D<sub>3</sub> monomers—folds as a strongly bent, five-stranded, antiparallel  $\beta$ -sheet capped by an N-terminal  $\alpha$ -helix (Fig. 2 *a* and *b*). This protein fold strongly resembles members of the oligonucleotide/oligosaccharide-binding fold family (33). The  $\beta$ -strands **B2**, **B3**, and **B4** create this peculiar structure by twisting back upon themselves to form a U-shaped core that is roughly elliptical in cross section. The amount of curvature may be measured as the distance between the N and C termini of a strand. As an example of the SmAP curvature, the distance between the **B2** strand termini ( $C_{\alpha}^{\text{Ser31}}$  and  $C_{\alpha}^{\text{Gln43}}$ ) is 23.7 Å, whereas this distance would be 39.7 Å in an un-bent, fully-extended conformation. The structural plasticity needed for such a high degree of  $\beta$ -strand curvature is apparently provided by several strictly conserved glycines that occur near the pivot points (Gly36, Gly64, Gly70 in the archaeal Sm sequences shown in Fig. 1). The segment linking **B4** and **B5** lies at the top of the U-shaped trough to close the protein into a  $\beta$ -barrel-like structure (Fig. 2*b*); it also positions **B5** for antiparallel hydrogen bonding to **B1** within the same monomer and the **B4** strand of an adjacent monomer (Fig. 2*a*).

This latter interaction—hydrogen bonding between strands **B4**...**B5**<sub>*i*+1</sub> for each monomer *i*—creates the seven dimer interfaces that orient the monomers head-to-tail around the ring. Specifically, each monomer contributes five residues from each **B4** and **B5** strand (Fig. 2*a*) to create an interface that occludes  $1,682 \pm 78 \text{ \AA}^2$  of surface area per dimer (Fig. 2*c*). The interfacial residues are predominantly apolar, and a few interesting side-chain interactions supplement the standard hydrogen bonds between the  $\beta$ -strand backbones. Such interactions include: (i) the guanidinium group of Arg-69<sub>B4</sub> in most of the monomers engages in several hydrogen bonds with main-chain and side-chain atoms from the adjacent **B5** strand, and (ii) van der Waals contact of the sulfur of Met-66<sub>B4</sub> with the aromatic ring of Phe-75<sub>B5</sub> suggests that a favorable S... $\pi$  aromatic interaction (34, 35) may stabilize the interface.

The structural superimpositions shown in Fig. 2 illustrate the strong similarity between SmAP and human Sm monomers and dimers. SmAP aligns best with Sm D<sub>3</sub> (31% sequence similarity, 1.0 Å rmsd over backbone atoms), and least well with D<sub>1</sub> (40% sequence similarity, 1.7 Å rmsd). The twisted  $\beta$ -sheet core of each Sm monomer is nearly identical (Fig. 2*b*), the main difference being the shorter L4 loop of SmAP versus three of the human structures (B, D<sub>1</sub>, and D<sub>3</sub>). A similar trend toward shorter loops has also been found in other pairs of thermophilic-mesophilic proteins (see ref. 35 and references therein). Both D<sub>3</sub>:B and D<sub>1</sub>:D<sub>2</sub> human heterodimer structures align very closely with a SmAP homodimer (1.4 Å and 1.6 Å rmsd, respectively). Superimposition of D<sub>3</sub>:B on SmAP shows that the dimer interface is conserved (Fig. 2*c*), although the amino acids in this region (**B4** residues 65–69 and **B5** residues 74–78) show greater phylogenetic variation than in the rest of the sequence (Fig. 1). This variation is explained by the fact that the dimer interface consists mainly of hydrogen bonds between the backbones of the  $\beta$ -strands. In addition to showing that human and archaeal Sm proteins belong to the same fold family, the SmAP structure is apparently an example of an ancestral homodimeric interface that evolved into several distinct, functional heterodimeric interfaces (D<sub>3</sub>:B, D<sub>1</sub>:D<sub>2</sub>, etc. interfaces in human).

**The SmAP Heptamer.** The heptameric organization of SmAP shown in Fig. 3 was revealed by the binding of iridium ions (used for phasing purposes, Fig. 3*b*) and by *in vitro* biophysical characterization (C.M. and D.E., unpublished results). The refined crystallographic model reveals a disk-shaped homoheptameric ring that measures  $\approx 65 \text{ \AA}$  in diameter and  $\approx 38 \text{ \AA}$  in height (Fig. 3*c*), which is consistent with the dimensions from electron microscopy of human Sm cores (12). One of the most notable features of the heptamer is the propagation of the  $\beta$ -sheet core



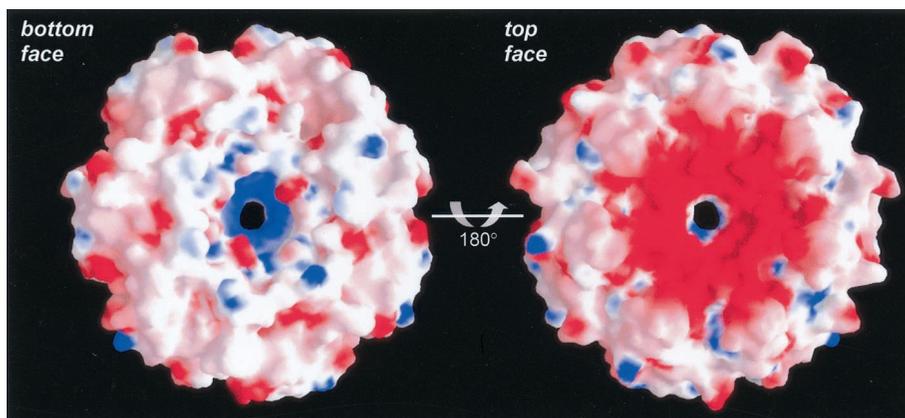
**Fig. 3.** Structure of the SmAP heptamer and the cationic pore. (a) A ribbon diagram of the SmAP heptamer, viewed from the “bottom” face, parallel to the sevenfold axis. (b) The enlarged view of experimental  $2F_o - F_c$  electron density for the Arg-29 side chains that line the pore (green, contoured at  $1.2 \sigma$ ). The anomalous difference density marks the sites of the eight bound iridium ions (indigo, contoured at  $3.5 \sigma$ ). (c) Illustration of the molecular surface for a sagittal section of the heptamer, with the sevenfold axis vertical. Arg-29 side chains are rendered as ball-and-stick models, and reveal that the most constricted section of the hourglass-shaped pore (traced in black) has a diameter of 8.8 Å [heavy, closed arrows ( $\uparrow$ ) are pore dimensions; light, open arrows ( $\uparrow$ ) are for the entire heptamer]. Horizontal bars indicate the concentric charged rings that are discussed in the text (red = anionic; blue = cationic).

of individual monomers across the ring to give a circular, 35-stranded  $\beta$ -sheet. This extended  $\beta$ -sheet does not lie in a single plane because of the strong curvature of the constituent strands. Rather, the U-shaped monomers are arranged like the blades of a turbine, rotated by  $\approx 45^\circ$  out of the horizontal plane of the heptamer. In contrast to the model proposed for the human heptamer (15), the remarkable asymmetry in the electrostatic surface of the SmAP heptamer gives the disk a large dipole moment (Fig. 4). This feature may be of functional

relevance because it imparts directionality to putative SmAP–RNA interactions.

### Discussion

**The Cationic Pore.** Interactions of SmAP with RNA (or DNA) are likely to occur within the shallow hourglass-shaped pore that runs along the sevenfold axis (Fig. 3 a and c). This cationic pore is  $\approx 15$  Å in depth and has a minimum diameter of 8.8 Å, the most constricted region being formed by the Arg-29 side chains of



**Fig. 4.** Electrostatic properties of the SmAP heptamer surface. The molecular surfaces of both faces of the SmAP heptamer are shown color-coded by electrostatic potential (red =  $-9.3$  kT; blue =  $+10.9$  kT). Note that there are two distinctive features: a ring of strong positive potential that lines the heptamer pore toward the bottom face, and a diffuse zone of intense negative potential across most of the top face. Such pronounced charge asymmetry gives the heptameric disk a large dipole moment (calculated to be 553 debye).

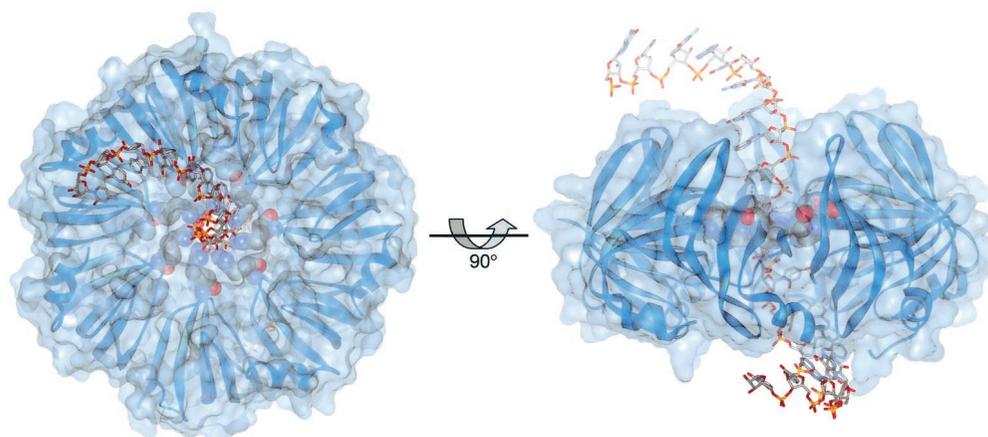
loop **L2**. The pore tapers steeply away on both sides of the narrow ring of arginines so that no other residues can be said to *line* the pore. Fig. 3c shows that four layers of charged heptagonal rings lie near to or within the pore. Starting at the bottom face of the heptamer, there is a negatively charged ring of glutamic acid residues (Glu-71, 18-Å diameter) followed by the narrow cationic ring of arginine residues (Arg-29, 8.8-Å diameter); a larger ring of aspartic acid residues (Asp-30, 11.5-Å diameter) is slightly above this; finally, there is a much wider ring of aspartic acid residues at the top face of the heptamer (Asp-57, 28.8-Å diameter). Note that unproductive SmAP-ssRNA bonding (i.e., interactions outside of the pore) would be diminished by such an arrangement of anionic rings.

Strict conservation of Arg-29 in archaeal Sm homologs (Fig. 1) suggests that a cationic pore with similar properties exists in all archaeal Sm heptamers (this arginine is also conserved in the human D<sub>3</sub> protein). By superimposing the human heterodimers on the SmAP heptamer, we find that the cationic character of the SmAP pore is probably conserved in the human Sm heptamer. Several positively charged residues from the human proteins would lie near the SmAP pore (although not in exactly the same

location as the Arg-29 ring), including Arg-51 and Lys-67 from D<sub>3</sub>, and Lys-54 from the B monomer.

The aforementioned residues that form the pore and charged heptagonal rings all lie in loops **L2**, **L4**, and **L5**. However, recent experiments by Urlaub *et al.* (37) demonstrate that residues from loop **L3** in human Sm B and G proteins can crosslink with an RNA nonanucleotide containing the Sm binding site (AAU<sub>5</sub>GA). Assuming that snRNA binds in the pore *in vivo*, these results are not consistent with the relative locations of the pore and loop **L3** in SmAP, and suggest that either (i) the human Sm monomers adopt altered conformations in the heteroheptamer to position the **L3** loops more proximal to the pore, or (ii) the RNA nonanucleotide used by Urlaub *et al.* binds near the pore, but not within it.

**Elucidation of Sm Binding Site Specificity.** Notably, the SmAP heptamer structure offers a simple explanation for the specificity of the snRNA Sm binding site (RAU<sub>4-6</sub>GR). The depth of the pore varies depending on exactly how it is defined, and a lower bound estimate of the pore depth (15 Å) is shown in Fig. 3c. More importantly, the aperture of the cationic Arg-29 ring is clearly defined in electron density (Fig. 3b) and gives the pore a



**Fig. 5.** Model of a SmAP-ssRNA complex. Orthogonal views are shown for a model of SmAP bound to a hypothetical 20-nucleotide ssRNA (e.g., eukaryotic snRNA) that consists of three segments from 5' to 3': a random string of 6 nucleotides in the A-form conformation (ACGAUC), followed by a minimal consensus Sm binding site (GAU<sub>4</sub>GA), and ending with 6 more nucleotides in A-form geometry (ACGAUC). The SmAP heptamer is depicted as a ribbon diagram along with the solvent-accessible surface. The Arg-29 ring that forms the pore is colored by atom type and rendered in space-filling form, with the ssRNA drawn as a stick model. The steric and electrostatic environment of the pore is ideally suited to accommodate a single-stranded polypyrimidyl nucleic acid.

minimum diameter of 8.8 Å. Fig. 5 shows a model of the SmAP heptamer complexed with a 20-nucleotide ssRNA that contains the consensus Sm binding site. The minimal diameter of the pore is ideally suited for binding of a single strand of polypyrimidyl RNA, but is too narrow to accommodate a polypurine, thereby affording steric discrimination between polypyrimidine and polypurine tracts. At least four or five nucleotides are required to span the  $\approx 16$  Å length of the pore, thus explaining the minimum requirement of four to five uridines in the Sm binding site. A more difficult matter to explain is the selection of poly(U) versus the other possible polypyrimidine, poly(C). It is possible that the cationic pore favors poly(U) because it is a better hydrogen-bond acceptor than the slightly more electropositive poly(C). Most importantly, the eighth iridium peak in anomalous difference Fourier maps is located in the pore,  $\approx 3$  Å below the cationic Arg-29 plane (Fig. 3b). This peak proves that the pore is accessible to polyvalent anions (e.g., octahedral  $\text{IrCl}_6^{3-}$ ) by an approach from the bottom face of the heptamer.

**The Sm Core of Eukaryotic snRNPs.** The crystal structure of SmAP and the model of it bound to an snRNA Sm binding site have several implications for the Sm core of eukaryotic snRNPs. We note that there are two possible paths by which SmAP may bind snRNA. Either the RNA is threaded through the pore (from the bottom face), or the heptamer assembles onto the snRNA binding site. The SmAP structure shows that the latter scenario is more plausible because purine bases cannot fit in the Sm pore. *In vitro* evidence for stable, subheptamers of human Sm proteins (e.g., an E·F·G heterotrimer, ref. 8), along with the fact that a minimal complex of five Sm proteins will bind RNA, lends credence to this stepwise core snRNP assembly pathway.

The regulation of such an assembly pathway could be achieved by noting the recent finding by Brahm *et al.* (32) that the C-terminal Arg-Gly dipeptides of human Sm proteins D<sub>1</sub> and D<sub>3</sub> are symmetrically dimethylated on the arginine  $\eta^1$  and  $\eta^2$  nitrogens *in vivo*. Fig. 1 shows that a C-terminal Arg-69–Gly-70 dipeptide of SmAP is strictly conserved among archaeal Sm

proteins and aligns with one of the Arg-Gly dipeptides of human D<sub>3</sub> protein. Inspection of the structural environment of Arg-69 in the human and SmAP structures shows that it lies at the Sm dimer interface, where it makes many contacts (Fig. 2c). Dimethylation of Arg-69 guanidinium groups by a protein arginine methyltransferase is predicted to interfere sterically with this interface, preventing Sm heptamer formation. Therefore, the structure of SmAP leads us to propose that regulation of the snRNP core assembly may be achieved by arginine dimethylation.

Because SmAP is a homoheptamer, our results support a mechanism of modern eukaryotic Sm protein evolution by early gene duplication events. Presumably, an archaeal SmAP gene was duplicated and gradually accrued neutral point mutations, with the structural restraint that the independently evolving monomers still form a heteroheptamer. Such a scheme would give rise to modified/modern Sm heteroheptamers, such as the human B·D<sub>1</sub>·D<sub>2</sub>·D<sub>3</sub>·E·F·G heptamer, with modified/modern biochemical activities, such as U snRNA binding. This model implies asymmetric heteroheptamers composed of paralogous Sm proteins, and explains the diversification of Sm protein function. Our preliminary crystallographic results with another archaeal Sm protein suggest that it also assembles into a homoheptamer, thereby supporting this model. Finally, we note that these results raise several interesting questions regarding the potential RNA that *P. aerophilum* SmAP interacts with *in vivo*, the cellular role of this putative RNA binding, and the possibility of snRNP-like particles in archaeal species.

We thank Dr. Kym Faull for help with mass spectrometry, and Adam Frankel, Linda Columbus, Drs. Dan Anderson, Doug Black, Jim Bowie, Guillaume Chanfreau, Sorel Fitz-Gibbon, Celia Goulding, and Ioannis Xenarios for useful discussions. We also thank Brookhaven National Lab for the use of beamline X8C of the National Synchrotron Light Source. We gratefully acknowledge financial support from the National Institutes of Health, the Department of Energy, and a National Science Foundation predoctoral fellowship (to C.M.).

1. Yu, Y.-T., Scharl, E. C., Smith, C. M. & Steitz, J. A. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 487–524.
2. Burge, C. B., Tuschl, T. & Sharp, P. A. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 525–560.
3. Lerner, M. R. & Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5495–5499.
4. Seraphin, B. (1995) *EMBO J.* **14**, 2089–2098.
5. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. & Seraphin, B. (1999) *EMBO J.* **18**, 3451–3462.
6. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. & Seraphin, B. (2000) *EMBO J.* **19**, 1661–1671.
7. He, W. & Parker, R. (2000) *Curr. Opin. Cell Biol.* **12**, 346–350.
8. Raker, V. A., Plessel, G. & Luhrmann, R. (1996) *EMBO J.* **15**, 2256–2269.
9. Mattaj, I. W. (1986) *Cell* **46**, 905–911.
10. Jones, M. H. & Guthrie, C. (1990) *EMBO J.* **9**, 2555–2561.
11. Raker, V. A., Hartmuth, K., Kastner, B. & Luhrmann, R. (1999) *Mol. Cell. Biol.* **19**, 6554–6565.
12. Kastner, B., Bach, M. & Luhrmann, R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1710–1714.
13. Achsel, T., Brahm, H., Kastner, B., Bachi, A., Wilm, M. & Luhrmann, R. (1999) *EMBO J.* **18**, 5789–5802.
14. Stark, H., Dube, P., Luhrmann, R. & Kastner, B. (2001) *Nature (London)* **409**, 539–542.
15. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J. & Nagai, K. (1999) *Cell* **96**, 375–387.
16. Kambach, C., Walke, S. & Nagai, K. (1999) *Curr. Opin. Struct. Biol.* **9**, 222–230.
17. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276A**, 307–326.
18. Terwilliger, T. C. & Berendzen, J. (1999) *Acta Crystallogr. D* **55**, 849–861.
19. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. D* **50**, 760–763.
20. Cowtan, K. & Main, P. (1998) *Acta Crystallogr. D* **54**, 487–493.
21. Perrakis, A., Morris, R. & Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**, 458–463.
22. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991) *Acta Crystallogr. A* **47**, 110–119.
23. 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) *Acta Crystallogr. D* **54**, 905–921.
24. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993) *J. Mol. Biol.* **231**, 1049–1067.
25. Colovos, C. & Yeates, T. O. (1993) *Protein Sci.* **2**, 1511–1519.
26. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
27. Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996) *Methods Enzymol.* **266**, 383–402.
28. Cohen, G. H. (1986) *J. Mol. Biol.* **190**, 593–604.
29. Shindyalov, I. N. & Bourne, P. E. (1998) *Protein Eng.* **11**, 739–747.
30. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281–296.
31. Evans, S. V. (1993) *J. Mol. Graph.* **11**, 134–138.
32. Brahm, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L. & Luhrmann, R. (2000) *J. Biol. Chem.* **275**, 17122–17129.
33. Murzin, A. G. (1993) *EMBO J.* **12**, 861–867.
34. Reid, K. S. C., Lindley, P. F. & Thornton, J. M. (1985) *FEBS Lett.* **190**, 209–213.
35. Druhan, L. J. & Swenson, R. P. (1998) *Biochemistry* **37**, 9668–9678.
36. Thompson, M. J. & Eisenberg, D. (1999) *J. Mol. Biol.* **290**, 595–604.
37. Urlaub, H., Raker, V. A., Kostka, S. & Luhrmann, R. (2001) *EMBO J.* **20**, 187–196.