PROTEINS: Structure, Function, and Bioinformatics 61:689–693 (2005)

STRUCTURE NOTE

Crystal Structure of an Archaeal Sm Protein From *Sulfolobus solfataricus*

Turgay Kilic,[1] Stéphane Thore,[2] and Dietrich Suck[1]*
[1]*Structural and Computational Biology Programme, European Molecular Biology Laboratory, Heidelberg, Germany*
[2]*Institute for Molecular Biology and Biophysics ETH-Hönggerberg, Zurich, Switzerland*

Introduction. Members of the Sm protein family share a common bipartite sequence motif called Sm domain, which consists of the Sm1 and Sm2 motifs separated by a loop of variable length. The Sm domain contains an N-terminal α-helix and five strongly bent antiparallel β-strands. Multiple sequence alignment of Sm1- and Sm2-type proteins reveals the most conserved and highly conserved residues among the family (Fig. 1). In eukaryotes, Sm and Sm-like (Lsm) proteins associate with RNA to form the core domain of small nuclear ribonucleoprotein particles (snRNPs), which are involved in various RNA-processing events such as messenger RNA (mRNA) decapping and degradation,[1–3] pre-mRNA splicing,[3–5] and telomere replication.[6] In snRNPs that are involved in pre-mRNA splicing, Sm proteins bind to the so-called Sm site, a U-rich, single-stranded region containing the consensus sequence $PAU_{4-6}GP$, where P is any purine. Through database searches, Sm-related proteins were also identified in archaea and eubacteria.[7,8] The function of Sm proteins in archaea is currently not clear. However, *Archaeoglobus fulgidus* Sm1- and Sm2-type proteins (AF-Sm1 and AF-Sm2) have been shown to interact with ribonuclease (RNase) P RNA in vivo,[9] which seems consistent with the earlier immunoprecipitation results of the yeast Lsm2–7 complex with pre-RNase P RNA.[7] These findings suggest that archaeal Sm proteins may have a function in transfer RNA (tRNA) processing. Archaeal Sm proteins also exhibit similar oligomerization and RNA binding properties as their eukaryotic counterparts. Archaeal genomes encode at most three Sm-related proteins forming homoheptameric or hexameric complexes.[10–12] Sm3-type proteins contain an additional C-terminal domain, which is highly conserved in the Sm3 subfamily.[10] The crystal structure of AF-Sm1 complexed with oligo U_5 represents the first high-resolution picture of binding of RNA to the archaeal Sm core.[9] Here we report the X-ray structure of the *Sulfolobus solfataricus* Sm1-type protein (SS-Sm1) at 1.68 Å resolution. The SS-Sm1 structure forms a heptameric ring ~68 Å in diameter with a central cavity of ~14 Å. Both the monomer fold and the heptameric ring structure of SS-Sm1 are similar to that of other archaeal Sm proteins. The heptameric structure is maintained through interactions between β4- and β5-strands of adjacent monomers. Band shift experiments indicate that SS-Sm1 binds to short oligo U stretches in vitro (data not shown). The SS-Sm1 heptamer shows a different charge distribution on the RNA binding side of the ring compared to the AF-Sm1 and *Pyrococcus abyssi* Sm1 (PA-Sm1) proteins.

Materials and Methods. *Cloning, expression and purification:* The SS-Sm1 protein was purified from pET24 transformed *Escherichia coli* BL21/DE3^{star} cells. The pET24 vector contains an upstream sequence coding for a 6xHis-tag followed by a tobacco etch virus protease site (TEV site). Polymerase chain reaction (PCR) amplified SS-Sm1 DNA fragment was digested and inserted into the pET24 expression vector. Upon sequence confirmation, the correct transformant was incubated in 6 liters of Luria–Bertani (LB) medium at 37°C until the OD_{600} reached ~0.6. Cells were induced by adding isopropylthio-β-D-galactoside (IPTG) at a final concentration of 1 m*M* for 3 h and harvested by centrifugation. Cell lysis was carried out using a French Press in 50 m*M* Tris-HCl, pH 8.0, 0.3 *M* NaCl, 0.5% IGEPAL, 15 m*M* imidazole, and 10 m*M* β-mercaptoethanol (added fresh). Cell lysate was loaded onto a Ni-NTA column for initial purification. A Sephadex g25 column was used for buffer exchange to 50 m*M* Tris-HCl, pH 8.0, 0.5 *M* ethylenediaminetetraacetic acid (EDTA), 1 m*M* diothreitol (DTT) for TEV site cleavage by TEV protease. After cleavage, the protein solution was heated to 76°C for 10 min to yield homogeneous SS-Sm1 protein. Denatured proteins were removed by centrifugation. The supernatant containing SS-Sm1 protein was passed through a Superdex 75 gel-filtration column for homogeneity and exchange of buffer to 0.2 *M* NaCl and 50 mM Tris-HCl, pH 8. Fractions containing the proteins were further analyzed by 18% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), concentrated to 13.8 mg/mL and stored at −20°C for crystallization.

*Correspondence to: Dietrich Suck, Structural and Computational Biology Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117, Heidelberg, Germany. E-mail: suck@embl.de

Received 18 March 2005; Accepted 29 April 2005

Published online 23 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20637

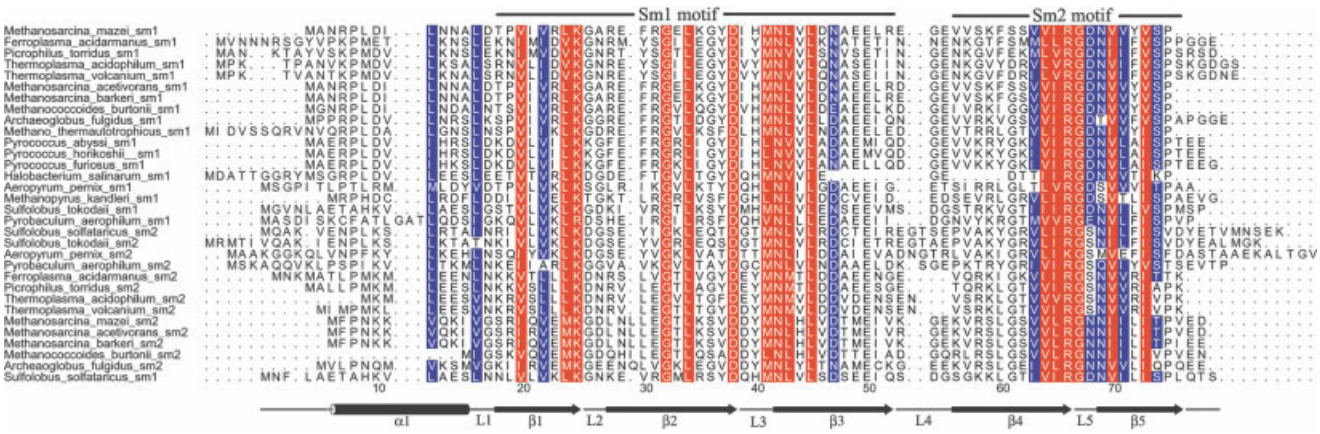

Fig. 1. Sequence alignment of all presently known archaeal Sm1- and Sm2-type proteins. Numbering and structure assignment are depicted for *S. solfataricus* Sm1 (shown at the bottom). All residue numbers throughout this article refer to the numbers assigned in this figure. Fully conserved (e.g., D38 and N42) or almost fully conserved residues are shown in red, and highly conserved residues, in blue. Sm3-type proteins that contain an additional C-terminal domain are not included for simplicity. The figure was made using ClustalX v.1.81 and Alscript v.2.05.

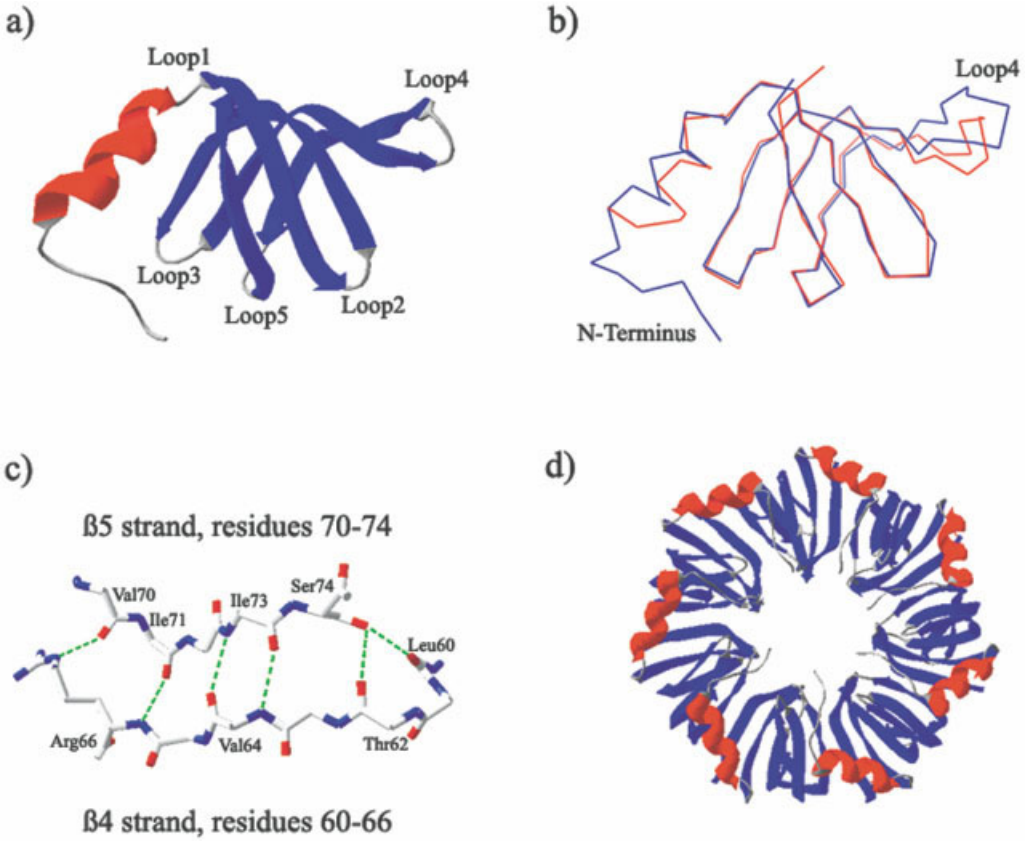

Fig. 2. (**a**) Ribbon representation of the SS-Sm1 monomer. α-helix is shown in red, β-strands are shown in blue, and loop regions are shown in gray. (**b**) Superimposed Sm1 monomers from *S. solfataricus* (blue) and *A. fulgidus* (red). RMSD (for Cα atoms) is 1.07 Å. (**c**) Hydrogen bonds between β4- and β5-strands of neighboring subunits. Residues that are involved in intersubunit hydrogen bonding are shown. (**d**) Heptameric SS-Sm1 ring structure; α helices (red) are located on the face toward the viewer. Figures were prepared using SWISS-PdbViewer v.3.7 and PyMOL v.0.97.

Crystallization, data collection, and refinement: The initial screening for crystallization conditions was carried out using the Hampton Research Index Screen with the sitting-drop vapor diffusion method. Upon finding promising crystallization conditions, drops were prepared by mixing 1 μL reservoir with 1 μL protein solution using the hanging drop method. With optimized conditions, SS-Sm1 single crystals grew in 25% (w/v) polyethylene glycol (PEG3350), 0.3 *M* lithium sulfate, and 0.1 *M* Tris-HCl, pH 8.0, in one week, with dimensions about 0.5 × 0.2 × 0.2 mm. A flash-frozen single crystal diffracted to a resolution of 1.68 Å at the PX beamline of the Swiss Light Source at Paul Scherrer Institut, Switzerland. Data were collected using a MAR charge-coupled device (CCD) detector under cryoconditions. Images were indexed and scaled using Denzo and Scalepack in the HKL package[14] (v.1.97.2). The structure was solved by molecular replacement using MOLREP[15] (v.7.2) from the CCP4 package[16] (v.5.0.2) and refined using Crystallography & NMR System (CNS)[17] (v.1.1). The R_{work} and R_{free} are 21.7% and 23.9%, respectively, after solving the structure by molecular replacement using AF-Sm1 as a search model. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with the accession code 1TH7. Data collection and refinement statistics are given in Table I.

TABLE I. Summary of Data Collection and Refinement Statistics for *Sulfolobus solfataricus* Sm1

Data Collection*	
Resolution range (Å)	40.0–1.68
Wavelength (Å)	0.9793
Number of total reflections	1797673
Number of unique reflections	132245
% completeness	98.9 (98.2)
I/σI	22.68 (4.0)
R_{sym}**	8.6 (61.6)
Refinement*	
Resolution range for refinement (Å)	36.0–1.68
Unit cell parameters	a = 71.18 Å, b = 66.81 Å, c = 129.01 Å, α = γ = 90.0°, β = 104.19°
Space group	$P2_1$
Protein MW (81 amino acids, monomer)	8807 Daltons
Protein molecules per A.U.	14 (2 heptamers)
Matthew's coefficient (Å³/Dalton)	2.4
Solvent content (%)	48.6
B_{work} (%)	21.7 (24.2)
Number of reflections (working set)	125561 (20547)
R_{free} (%)	23.9 (25.9)
Number of reflections (test set)	6658 (1058)
Number of protein residues	1092
Number of water molecules	474
Mean B factor (Å²)	23.4
RMSD for bond lengths (Å)	0.004
RMSD for bond angles (Å)	1.30
Ramachandran plot statistics	
Most favored region (%)	88.5
Additional allowed regions (%)	11.5
Generously allowed regions (%)	0.0
Disallowed regions (%)	0.0

*Values in parentheses refer to the statistics in the highest resolution shell (1.74–1.68 Å).
**R_{sym} = Σhkl|I − ⟨I⟩|/Σhkl I, where I is the observed intensity for a reflection of index hkl and ⟨I⟩ is the mean intensity.

Results and Discussion. *SS-Sm1 monomer has the typical Sm fold:* Figure 1 shows the multiple sequence alignment of SS-Sm1 with the other currently available archaeal Sm1 and Sm2 proteins revealing the most conserved (in red) and highly conserved residues (in blue). β-strands 1, 2, and 3 constitute the first motif and β-strands 4 and 5 constitute the second motif of the Sm domain. In addition to the sequence similarity, the SS-Sm1 monomer exhibits the characteristic folding pattern of Sm proteins [Fig. 2(a)]: Five strongly bent β-strands are preceded by an N-terminal α-helix. Therefore, it was possible to solve the structure of SS-Sm1 monomer using the AF-Sm1 monomer as a model by molecular replacement. Superimposition of SS-Sm1 and AF-Sm1 results in a root-mean-square deviation (RMSD) of 1.07 Å for Cα atoms revealing significant structural differences only at the N-terminus and the variable loop 4 region [Fig. 2(b)]. The more extended N-terminus of SS-Sm1 is a consequence of the cloning strategy which resulted in two additional residues (Gly and Ala) at the N-terminus. In addition, there are two residues (Met and Pro) at the N-terminus of AF-Sm1 that are not present in the structure file. Loop 4 is a variable region among all the Sm proteins in length and conformation.

SS-Sm1 forms a heptameric ring structure: Interactions between SS-Sm1 monomers are provided through the hydrogen bonds between β4- and β5-strands of the adjacent monomers [Fig. 2(c)]. Additional hydrogen bonds and ionic interactions between the monomers provide further stability to the SS-Sm1 heptamer. A comparison of these intersubunit contacts in the AF-Sm1, PA-Sm1, and SS-Sm1 heptamers shows little difference. Some of the residues interacting through hydrogen bonds or ionic interactions (e.g., Arg66–Val70 and Lys25–Asp68) are in common for all three archaeal heptamers. Some of the hydrogen bonds show changes in their acceptors (e.g., Ser74 is hydrogen-bonded to Ser62 in AF-Sm1, to Lys62 in PA-Sm1, and to Thr62 in SS-Sm1), while some of the noncovalent interactions are present in one heptamer only. Through these interactions, seven SS-Sm1 monomers form a stable ring structure of ~68 Å diameter with a central cavity of ~14 Å [Fig. 2(d)]. Gel filtration analysis under nondenaturing conditions at pH 8.0 reveals minor amounts of SS-Sm1 monomers as well. However, SS-Sm1 is predominantly in the heptameric state as seen by the gel filtration and native-PAGE (data not shown). Formation of this ring structure is essential for the binding of the Sm protein to the oligo U stretches.

Surface charge distribution of the heptamer: The electrostatic surface charge potential of the SS-Sm1 heptamer shows an accumulated positive charge in and surrounding the central hole of the heptamer (indicated in blue), similar to what is observed in the other two heptamers, AF-Sm1 and PA-Sm1 (Fig. 3, top). This positive-charge environment in the center of SS-Sm1 is produced by the side-chains of residues Lys25, Lys28, and Arg66 extending from each monomer toward the central cavity. The other side of the SS-Sm1 heptamer, where the binding of the

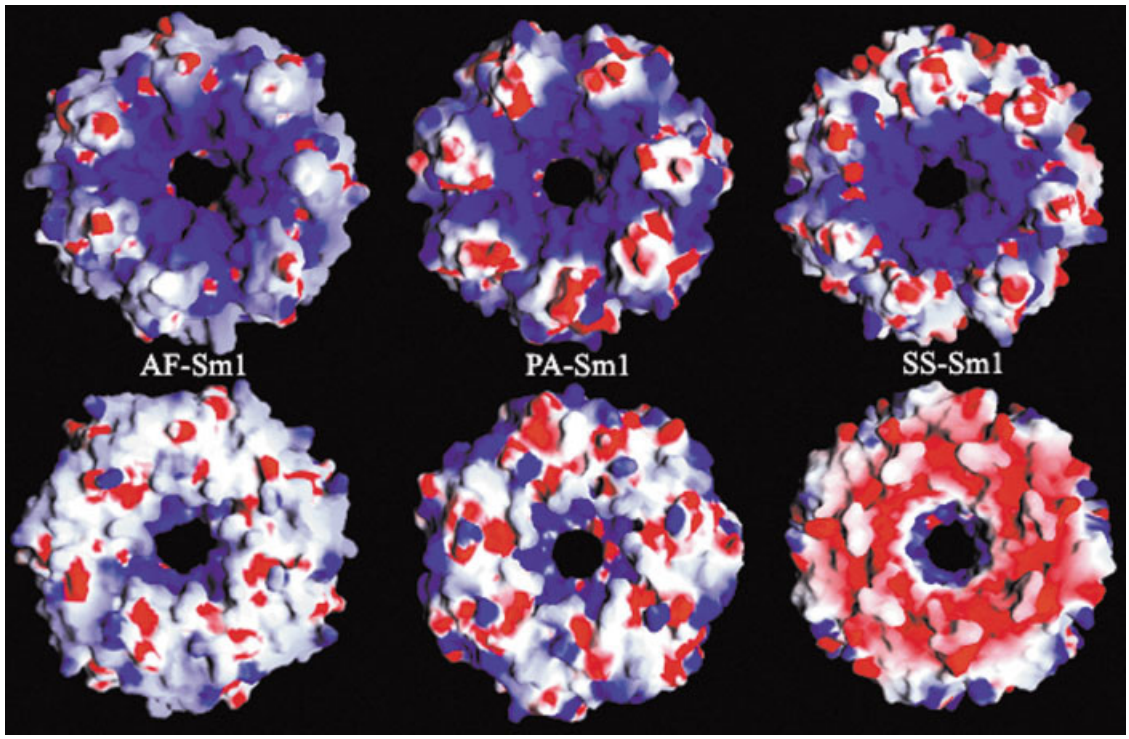

Fig. 3. Comparison of the electrostatic surface charge distribution of three archaeal Sm1-type heptamers. Accumulated positive charge in and around the central cavity (in blue) can be seen in all the three Sm1 proteins on one side. However, the opposite side, which contains the RNA-binding pocket, shows a distinct charge difference between the heptamers. *S. solfataricus* Sm1 (SS-Sm1) displays more negative surface charge (in red) on this side than *A. fulgidus* Sm1 (AF-Sm1) and *P. abyssi* Sm1 (PA-Sm1). Figure was made using Grasp v.1.3.6.

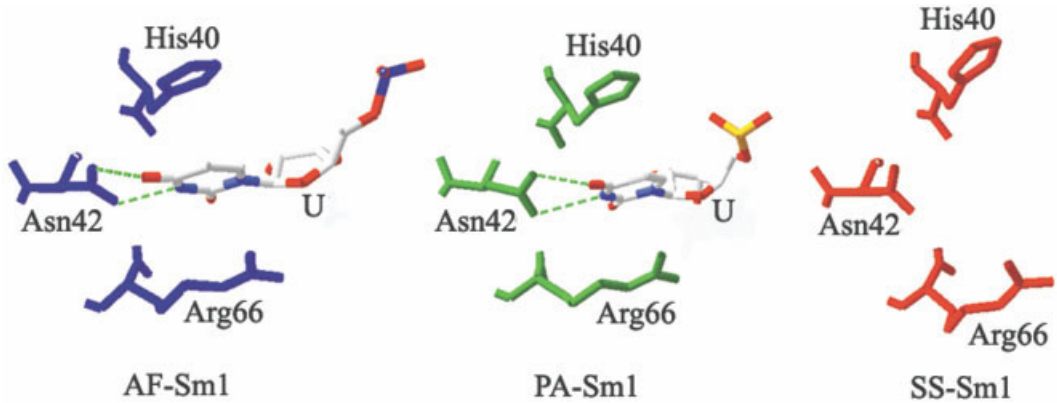

Fig. 4. Three residues, His40, Asn42, and Arg66 of *A. fulgidus* Sm1 (AF-Sm1) and *P. abyssi* Sm1 (PA-Sm1) form the RNA-binding pocket where a uracil base is hydrogen-bonded to Asn42 and stacked between side-chains of His40 and Arg66. The same binding pocket is present in *S. solfataricus* Sm1 (SS-Sm1). Residue numbers are based on the numbering in Figure 1.

RNA is expected, shows more negative-charge accumulation (indicated in red) than AF-Sm1 and PA-Sm1 (Fig. 3, bottom). This negative charge on the RNA-binding side of the SS-Sm1 heptamer is produced by residues Glu6, Glu14, and Asp38. Despite this negative-charge accumulation on the RNA-binding side, we observe the binding of oligo U by band shift experiments. The charge difference between the two faces of the SS-Sm1 heptamer may play a role in the interaction with its protein or nucleic acid binding partners.

SS-Sm1 contains the conserved uridine binding pocket: In the uridine binding pocket of AF-Sm1 and PA-Sm1, the uracil base is stacked between the side-chains of His40 and Arg66. Specific hydrogen bonds between the uracil base and the side-chain of Asn42 and the main-chain NH-group of Asp68 provide additional stability for RNA binding (Fig. 4, Asp68 not shown for clarity).[9,13] This uridine binding pocket is conserved in SS-Sm1 and contains the same residues (His40, Asn42, Arg66, and Asp68). The binding pocket of SS-Sm1 is a preformed binding site where we

expect little movement of the side-chains of the involved residues upon binding to uracil or interaction with its target nucleic acid. The secondary RNA-binding site seen in PA-Sm1/U_7 complex[13] is only partially conserved in SS-Sm1; however, the conserved tyrosine at position 37 could well be involved in a similar stacking interaction with a nucleobase as Tyr37 of PA-Sm1.

Acknowledgments. We would like to thank Dr. Eric Ennifar for his helpful suggestions on the data processing and Dr. Christian Biertuempfel, as well as the staff of SLS for help with data collection at the X06SA beamline.

REFERENCES

1. Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R. Yeast Sm-like proteins function in mRNA decapping and decay. Nature 2000;404:515–518.
2. Bouveret E, Rigaut G, Shevchenki A, Wilm M, Seraphin B. A Sm-like protein complex that participates in mRNA degradation. EMBO J 2000;19:1661–1671.
3. He W, Parker R. Functions of Lsm proteins in mRNA degradation and splicing. Curr Opin Cell Biol 2000;12:346–350.
4. Mattaj IW, Tollervey D, Seraphin B. Small nuclear RNAs in messenger RNA and ribosomal RNA processing. FASEB J 1993;7: 47–53.
5. Will CL, Luhrmann R. Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol 2001;13:290–301.
6. Seto AG, Zaug AJ, Sobel SG, Wolin SL, Cech TR. *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. Nature 1999;401:177–180.
7. Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Seraphin B. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. EMBO J 1999;18:3451–3462.
8. Moller T, Franch T, Hojrup P, Keene DR, Bachinger HP, Brennan RG, Valentin-Hansen P. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 2002;9:23–30.
9. Törö I, Thore S, Mayer C, Basquin J, Seraphin B, Suck D. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J 2001;20: 2293–2303.
10. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D. Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc Natl Acad Sci USA 2003;100:4539–4544.
11. Törö I, Basquin J, Teo-Dreher H, Suck D. Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. J Mol Biol 2002;320:129–142.
12. Collins BM, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC. Crystal structure of a heptameric Sm-like protein from archaea: implications for the structure and evolution of snRNPs. J Mol Biol 2001;309:915–923.
13. Thore S, Mayers C, Sauter C, Weeks S, Suck D. Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. J Biol Chem 2003;278:1239–1247.
14. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
15. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. J Appl Crystallogr 1997;30:1022–1025.
16. CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
17. Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.