The Sm domain is an ancient RNA-binding motif with oligo(U) specificity

Tilmann Achsel, Holger Stark, and Reinhard Lührmann*

Max Planck Institute of Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Communicated by Harry F. Noller, University of California, Santa Cruz, CA, January 22, 2001 (received for review November 4, 2000)

Sm and Sm-like proteins are members of a family of small proteins that is widespread throughout eukaryotic kingdoms. These proteins form heteromers with one another and bind, as heteromeric complexes, to various RNAs, recognizing primarily short U-rich stretches. Interestingly, completion of several genome projects revealed that archaea also contain genes that may encode Sm-like proteins. Herein, we studied the properties of one Sm-like protein derived from the archaebacterium *Archaeoglobus fulgidus* and overexpressed in *Escherichia coli*. This single small protein closely reflects the properties of an Sm or Sm-like protein heteromer. It binds to RNA with a high specificity for oligo(U), and assembles onto the RNA to form a complex that exhibits, as judged by electron microscopy, a ring-like structure similar to the ones observed with the Sm core ribonucleoprotein and the like Sm (LSm) protein heteromer. Importantly, multivariate statistical analysis of negative-stain electron-microscopic images revealed a sevenfold symmetry for the observed ring structure, indicating that the proteins form a homoheptamer. These results support the structural model of the Sm proteins derived from crystallographic studies on Sm heterodimers and demonstrate that the Sm protein family evolved from a single ancestor that was present before the eukaryotic and archaeal kingdoms separated.

The family of the Sm and Sm-like proteins is characterized by a bipartite sequence motif that consists of seven highly conserved amino acids embedded in a characteristic pattern of hydrophobic and hydrophilic residues (1–3). At least 15 members of this protein family are conserved throughout the eukaryotic kingdoms and form at least three functional entities. Each of the three entities contains a set of seven distinct polypeptides (for review, see ref. 4).

Most is known about the seven human Sm proteins B, D1, D2, D3, E, F, and G. These proteins assemble onto the spliceosomal small nuclear (sn) RNAs to form the so-called Sm core ribonucleoproteins (RNPs), which are essential for biogenesis and stability of the snRNAs and, therefore, play a major role in pre-mRNA splicing. In addition to the Sm proteins, homology searches identified several proteins termed "Like Sm" (LSm) (5); these also carry the Sm motif. Eight of these LSm proteins are conserved throughout the eukaryotic kingdoms; proteins LSm 2–8 form a complex that binds U6 snRNA (6, 7) and is also found in the U4/U6 di-snRNP and the U4/U6·U5 tri-snRNP (8, 9). Like the canonical Sm proteins, LSm proteins 2–8 take part in snRNP stabilization and biogenesis (10–12). Finally, proteins LSm 1–7 associate *en bloc* with proteins required for mRNA decapping (13, 14). In addition, the yeast telomerase RNP contains a set of Sm proteins, suggesting that the Sm protein family plays additional roles in RNP metabolism (15). Moreover, LSm proteins might function in the stability or maturation of RNA polymerase III transcripts (11).

Apart from similarities in sequence, two other features are common to the Sm and LSm proteins. The first feature is their propensity to form oligomers. Sm proteins form stable complexes B·D3, D1·D2, and F·E·G and assembly of the three complexes onto the snRNA involves further protein–protein interactions (16–19). All Sm core RNPs exhibit a ring-shaped structure ≈8 nm in diameter, as observed by electron microscopy (20). Similarly, the human LSm proteins 2–8 interact to form a stable RNA-free complex similar in shape to the Sm core RNPs (6). The crystal structures of two Sm dimers, namely, D1·D2 and B·D3, revealed that the Sm motif represents an autonomously folding domain consisting of an N-terminal α-helix followed by five β-sheets (21). In both crystal structures, dimerization is brought about by an interaction of the $\beta4$ sheet of one Sm protein with $\beta5$ of the other, in line with biochemical and genetic evidence that the Sm domain is necessary and sufficient for the Sm proteins to form heteromers. Extrapolation of the Sm B·D3 and Sm D1·D2 structures, in conjunction with known interactions between the other Sm proteins, allowed the modeling of the Sm proteins into a seven-member ring (21). Although this ring still remains a model, it agrees both in size and shape with the Sm and LSm images obtained by electron microscopy (6, 20).

Another common feature of the Sm and LSm proteins is their ability to bind to RNA. However, none of the Sm heteromers suffices to bind stably to RNA, and at least the E·F·G and D1·D2 complexes are needed to form a stable RNP complex (16). The Sm protein binding site that is present on all snRNAs follows the consensus RAU$_5$GR (22), where R is any purine. Detailed analysis demonstrated that the Sm proteins recognize primarily the oligo(U) tract of the Sm site and that the conserved adenosine at the 5′ side of the oligo(U) tract is required for the high stability of the complex (16). Similarly, the LSm protein complex binds to the oligo(U) tail of the U6 snRNA (6). Thus, the Sm and LSm proteins possess, when assembled into a heteromer, a composite RNA binding site with specificity for oligo(U) stretches.

The common features of the Sm and LSm proteins raise the question of whether this diverse family has arisen from a single ancestral protein. If this is indeed the case, one may expect (*i*) that this ancestor had the ability to form homooligomers and (*ii*) that these bound to RNA, probably with oligo(U) specificity. The completion of several archaeal genome projects has revealed the presence of ORFs that may encode Sm-related proteins. None of these genomes, however, has seven Sm-related genes; instead, only one Sm sequence is conserved throughout the archaeal kingdoms, and a second ORF is found in some of the archaea. For example, the genome of *Archaeoglobus fulgidus*, a hyperthermophilic archaeon, contains two Sm-related ORFs (23). Alignment of their deduced amino acid sequences (Fig. 1) shows that both proteins contain a perfectly conserved Sm motif. In fact, each of these proteins consists entirely of an Sm domain, and additional domains as found in Sm and LSm proteins are missing in these proteins.

Herein, we describe the isolation of the *A. fulgidus* Sm-like protein Sm2 by overexpression in *Escherichia coli* and demonstrate that Sm2 and a functional Sm protein have similar properties of oligomerization and RNA binding. Moreover, in

Abbreviations: RNP, ribonucleoprotein; sn, small nuclear.

*To whom reprint requests should be addressed. E-mail: reinhard.luehrmann@mpi-bpc.mpg.de.

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.

www.pnas.org/cgi/doi/10.1073/pnas.071033998

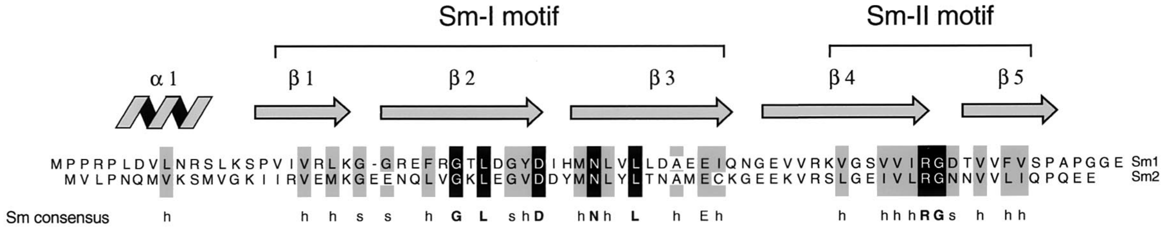
Sm-I motif
Sm-II motif
α 1
β 1
β 2
β 3
β 4
β 5
MPPRPLDVLNRSLKSPVIVRLKG-GREFRGTLDGYDIHMNLVLLDAEEIQNGEVVRKVGSVVIRGDTVVFVSPAPGGE Sm1
MVLPNQMVKSMVGKIIRVEMKGEENQLVGKLEGVDDYMNLYLTNAMECKGEEKVRSLGEIVLRGNNVVLIQPQEE Sm2
Sm consensus
h h h s s h G L s h D h N h L h E h h h h h R G s h h h

Fig. 1. Two ORFs in the genome of *A. fulgidus* are related to the eukaryotic Sm proteins. The deduced amino acid sequences of the ORFs Sm1 (GenBank accession no. O29386) and Sm2 (GenBank accession no. B69295; lower sequence) are shown. The consensus sequence of the eukaryotic Sm motifs (2, 6) is shown below. "h" indicates a bulky, hydrophobic (V, I, L, M, F, Y, or W) and "s" a small, polar (G, S, D, or N) residue. Positions that are identical or conserved in most Sm motifs are highlighted by solid and shaded bars, respectively. At the top, the secondary structure of the Sm domain (21) is indicated.

electron micrographs, the RNP complex formed by Sm2 and oligo(U) exhibits the characteristic seven-member ring structure of Sm protein complexes, data that support the model proposed by Kambach *et al.* (21).

Materials and Methods

Expression and Purification of Sm2 Protein. An artificial ORF encoding the same 76 amino acids as *A. fulgidus* Sm2 (GenBank accession no. B69295) was cloned into the *Nde*I and *Bam*HI sites of pET28c (Novagen). The protein was expressed in BL21/pLysS cells by induction with 0.4 mM isopropyl β-D-thiogalactoside for 12 h. Cells from a 2-liter culture were harvested, resuspended in 160 ml of TNEβ (20 mM Tris·HCl, pH 8.0/150 mM NaCl/0.2 mM EDTA/10 mM 2-mercaptoethanol), heated at 65°C for 30 min, and centrifuged. Ammonium sulfate was added to the supernatant to 25% saturation, and the solution was stirred for 60 min and centrifuged. The clear supernatant was loaded onto a 20-ml phenyl-Sepharose 6 Fast Flow column (high sub, Amersham Pharmacia). The column was washed with a buffer containing 25 mM sodium phosphate, 1 M ammonium sulfate, and 10 mM 2-mercaptoethanol (pH 7.0), and Sm2 protein was eluted with the same buffer but containing 0.5 M ammonium sulfate. The peak fractions (detected by A_{280}) were further fractionated on a 300-ml Sephacryl S200 column (Amersham Pharmacia) with TNEβ buffer. Protein concentrations were determined by the dye-binding assay (24).

RNA Binding Assays. The 5-μl assays containing 12.5 fmol of RNA oligonucleotides (labeled at their 5′ end with ^{32}P), 300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% Triton X-100, tRNA (0.1 μg/μl), and various concentrations of protein were incubated at 30°C for 15 min, fractionated on a 6% polyacrylamide (0.075% bisacrylamide) gel in 0.5× Tris-borate-EDTA (TBE) buffer at 170 V (10 V/cm) for 60–90 min. RNA bands were detected by autoradiography.

Polyuridylic acid (400 μg in 100 μl, pH 5.5) was partially hydrolyzed by adding 2 μl of 1 M NaOH and heating to 90°C for 15 min. The mixture was applied onto a 0.1-ml Mono Q column (Amersham Pharmacia), and resulting fragments were eluted by a gradient of 200–600 mM NaCl in 20 mM Tris·HCl, pH 8.0/0.2 mM EDTA. Note that all oligonucleotides obtained by this method terminate in a 2′ or 3′ phosphate.

Electron Microscopy. An Sm2·U_{10} RNP was obtained by scaling up the binding reaction (20 nmol of Sm2 and 3 nmol of U_{10}), and the complex plus unbound RNA was separated from free protein by rapid ion-exchange chromatography on a 1-ml Fractogel trimethylaminoethyl (TMAE) column (Merck); 50 mM NaCl/20 mM Tris·HCl/5 mM EDTA/0.5 mM dithioerythritol, pH 7.0, was used to wash the column and 300 mM NaCl in the same buffer was used to elute the snRNP. The sample was immediately used to prepare negatively stained electron microscope grids in 2% uranyl formate, and images were taken with a Philips CM200 FEG. We extracted 4,129 individual molecular complexes from four digitized electron micrographs. Multivariate statistical analysis (25) was carried out within the context of the Imagic-V software package (Image Science, Berlin) (26).

Results

Expression and Isolation of Pure Sm2 Protein. An artificial cDNA was assembled from three overlapping oligonucleotides and subcloned into a pET vector to yield an untagged protein with the amino acid sequence encoded by the *A. fulgidus* ORF Sm2 (GenBank accession no. B69295; Fig. 1). After overexpression in *E. coli*, most of the host proteins were precipitated by heat denaturation. The Sm2 protein was then purified to homogeneity by chromatography on phenyl-Sepharose, where it was eluted by 0.5 M ammonium sulfate, followed by gel filtration with Sephacryl S200 (Fig. 2*A*). The bulk of Sm2 eluted from the gel-filtration column later than the marker protein cytochrome *c* with a molecular mass of 13 kDa (Fig. 2*A*), which supports the theoretical molecular mass of 8.5 kDa and demonstrates that most of the Sm2 protein eluted as a monomer. A minor protein peak was also observed, with an elution volume slightly higher than that of the marker protein BSA (66 kDa). This peak was

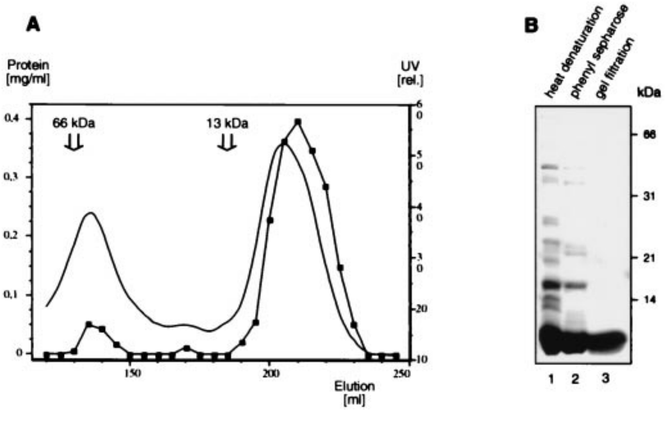

Fig. 2. Purification of Sm2 protein overexpressed in *E. coli*. (*A*) Elution profile of the final Sephacryl S200 gel-filtration column. The UV absorbance at 280 nm (continuous line) and the protein concentrations according to Bradford (connected squares) are plotted against the elution volume. The major elution peaks, at ≈60 kDa and less than 13 kDa, are, respectively, monomeric Sm2 protein and a presumed heptamer (see text). The positions of the peaks of BSA (66 kDa) and cytochrome *c* (13 kDa), used as molecular mass markers on another run of the same column, are indicated. (*B*) Purity of the preparation. Approximately 10 μg of the Sm2 protein preparation after heat denaturation of the *E. coli* proteins (lane 1), after phenyl-Sepharose chromatography (lane 2) and after gel filtration (lane 3), were fractionated by SDS/PAGE on 13% gels and stained with Coomassie blue.

shown by gel electrophoresis (data not shown) to consist largely of Sm2, but these fractions were too dilute for further investigation.

Binding of Sm2 to Oligo(U). To investigate the binding of Sm2 and RNA, we used electrophoretic mobility shift assays with a synthetic U_9 RNA oligonucleotide labeled at its 5′ end with ^{32}P. The choice of oligo(U) was determined by the known binding properties of Sm proteins (see Introduction). As shown in Fig. 3 *A Left*, the mobility of oligo(U) was greatly reduced by the addition of Sm2 protein. As judged by adding increasing concentrations of Sm2 protein, about 50% of the RNA molecules were bound when the Sm2 concentration reached 0.2 μM.

To demonstrate that this RNA binding has at least some sequence specificity, we next assayed for competition by poly(A), poly(G), poly(C), and poly(U). As shown in Fig. 3 *A Right*, poly(U) RNA efficiently competed for Sm2 binding even at 20 μM (calculated with respect to monomeric uridine), whereas the other homopolymers did not appreciably compete, even at five times greater concentrations. A partial competition by poly(C), but not by poly(A) or poly(G), was observed at still higher concentrations (0.2 and 1 mM; data not shown). Thus, Sm2 binds RNA with a high preference for oligo(U), and, in this respect, Sm2 resembles the Sm and LSm proteins. Further evidence for this similarity was provided by the finding that an oligoribonucleotide containing the Sm site sequence is also bound by Sm2, whereas an oligodeoxyribonucleotide with the same sequence is not (data not shown).

To determine the minimal RNA length required for binding of the Sm2 protein, poly(U) was fragmented by limited alkaline hydrolysis, and the resulting oligonucleotides were fractionated by ion-exchange chromatography on a Mono Q column. Sm2 protein binding was tested in band-shift assays as above. As shown in Fig. 3*B*, a minimal length of five uridines was required for efficient binding of the Sm2 protein (lanes 11–15). Notably, the smallest complex, formed on a pentamer of uridines, had the same electrophoretic mobility as the complexes formed on longer oligomers; complexes with intermediate mobilities were not observed. Furthermore, the binding affinity of Sm2 did not increase greatly between U_5 and U_6 (lanes 16–20), and thereafter with U_7 and U_8, no increase was observed by eye (lanes 21–30). This result strongly indicates that the stoichiometry and stability of the Sm2 RNP are insensitive to RNA lengths from U_5 to U_8. However, a further abrupt decrease in mobility for U_{17} and above was clearly observed (Fig. 3*C*), which we interpret as being caused by the formation of a second, independent Sm2 complex on the same RNA molecule. Results, such as those shown in Fig. 3 *B* and *C,* did not reveal the stoichiometric ratio for the interaction of Sm2 and RNA molecules. However, in view of the abrupt nature of both shifts in mobility, the stoichiometry of binding is precisely defined.

Structure and Stoichiometry of the $Sm2_n$·Oligo(U) Complex. A hallmark of the Sm and LSm proteins is their oligomerization into a ring-shaped structure. Therefore, the Sm2·RNA complex was investigated by electron microscopy to determine whether this complex also forms circular structures. Complexes were obtained by scaling up the binding reaction with U_{10}. This complex was isolated, negatively stained, and examined by transmission electron microscopy. A representative view is shown in Fig. 4*A*. We extracted 4,129 individual particles from four digitized electron micrographs. Most of these particles appeared as roughly circular images with a diameter of ≈8 nm and a central accumulation of stain, similar in size and shape to the Sm core particle and the RNA-free LSm protein heteromer (6, 20). However, in contrast to these complexes, which because of the protuberances present on some of the constituent proteins are highly unsymmetrical, the homooligomeric Sm2·RNA complex was well suited for computer-aided symmetry analysis where the symmetry information of the molecular complex was extracted from a large number of noisy electron microscopic images simultaneously. This method has been established (27) and used for the determination of several macromolecular complexes (28, 29). In the first step of the analysis, the individual images were brought to a common origin. The necessary shift for each image

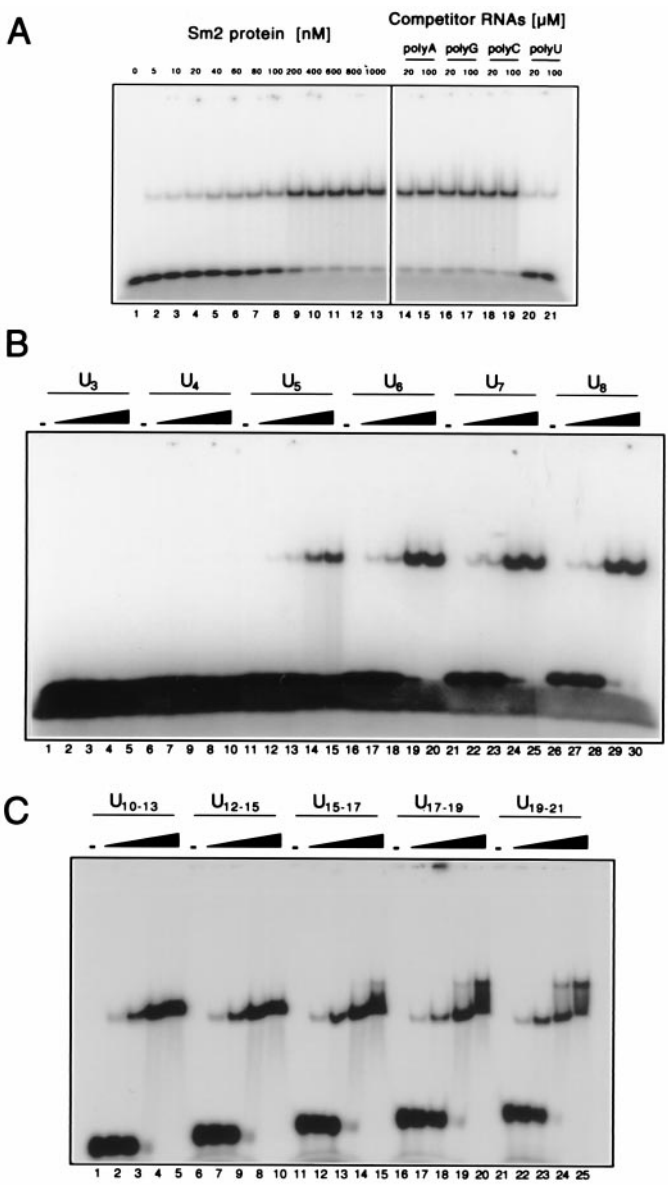

Fig. 3. Sm2 protein binds specifically to oligo(U). (*A*) Specific binding to U_9. (*Left*) Increasing concentrations of Sm2 protein, indicated above each lane, were incubated with radiolabeled U_9 and fractionated by PAGE under native conditions. The decreased mobility of the radioactive U_9 indicates the formation of a complex of U_9 and Sm2. (*Right*) Poly(A), poly(G), poly(C), or poly(U) was added to the assay containing 1 μM Sm2 protein, to a final RNA concentration (calculated for the monomer) of 20 or 100 μM. The disappearance of the immobile U_9 band indicates successful competition by the polynucleotide, as seen for poly(U) only. (*B*) Sm2 needs at least five uridines to bind. Oligo(U) of the length indicated above each block of five lanes was incubated with, respectively, 0 μM, 0.35 μM, 1.5 μM, 7 μM, or 15 μM Sm2 protein. (*C*) Two Sm2 complexes can form on longer oligo(U) RNAs. Longer oligo(U) samples were assayed as in *B*, with the same Sm2 protein concentrations. The RNA lengths used were not homogeneous, because of the limited resolution of Mono Q chromatography; the size range of the predominant oligonucleotides is given.

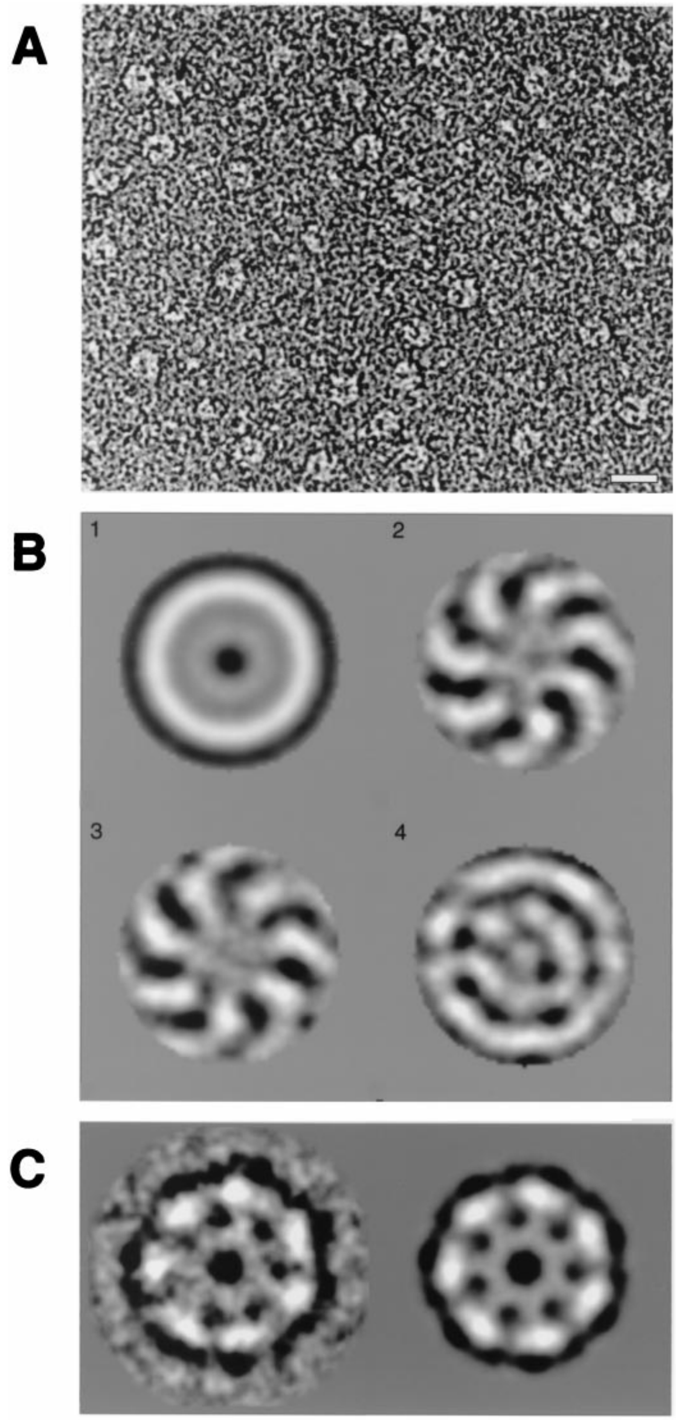

Fig. 4. Sm2·U_{10} complex exhibits, in electron micrographs, a ring structure with a 7-fold symmetry. (*A*) A typical electron micrograph field of Sm2 in the presence of U_{10}, negatively stained with 2% uranyl formate. (Bar = 1 nm.) (*B*) Symmetry analysis by multivariate statistical analysis. The first four eigenimages are depicted. The first resembles the total average of the data set. The second and third illustrate both the 7-fold symmetry and the rotational misalignment of the 7-fold symmetric component of the molecular images. The fourth and all subsequent eigenimages (not shown) do not show significant information in addition to the 7-fold symmetry. In particular, residual harmonic components with 6-fold and 8-fold symmetry were not found. (*C*) Class average of 20 individual molecular images grouped after an automated classification procedure was applied. The class average is shown without (*Left*) and with (*Right*) 7-fold symmetry imposed.

within its pixel frame was calculated by alignment to a reference image. To avoid any symmetry bias, only a translational alignment was used. The circular variations between the images were then assessed by an eigenvector–eigenvalue analysis (25).

The first four eigenvectors (which themselves were images) of this analysis are shown in Fig. 4*B*. The first eigenvector corresponds to the mass average of all images in the data set. The second and third eigenvectors reflect the symmetry properties of the Sm2·RNA complex. Because no rotational alignment was performed, eigenvectors two and three were 90° out of phase, which shows the absence of bias in the measurements. Both eigenvectors had a clear sevenfold symmetry, the only symmetry pattern found in the analysis. After the eigenvector analysis, class averages with improved signal-to-noise ratio were obtained by automated classification of the data set. One of these class averages is shown in Fig. 4*C*, with and without sevenfold symmetry imposed. Both averaged images clearly showed that the ring structure consisted of seven dense regions separated by regions with less density. This analysis therefore demonstrates beyond reasonable doubt that the Sm2·RNA ring has sevenfold symmetry.

Discussion

If the family of the Sm and Sm-like proteins evolved from a single ancestor, then it is likely that this ancestor bound RNA in a manner similar to the way in which it is bound by the Sm and LSm proteins of highly evolved organisms today. In this work we have shown that such a protein indeed exists in archaea, indicating that the ancestral Sm protein must have been present before the separation of the archaea and the eukaryotes three billion years ago.

The protein Sm2, derived from the hyperthermophilic archaeon *A. fulgidus* by overexpression in *E. coli*, fails to form stable homooligomers (Fig. 2; under such conditions, Sm proteins would form dimers and trimers) (17). However, Sm2 clearly forms heptamers when bound to RNA. First, the analysis of electron micrographs (Fig. 4) indicates that Sm2·RNP has a ring shape with sevenfold symmetry. From this analysis in principle, a protein-to-RNA stoichiometry of 14:1 or even 21:1 is conceivable. This is, however, highly unlikely for two reasons. (*i*) All of the complexes lie almost flat (Fig. 4*A*), indicating that they are unable to rest on an edge and are, therefore, unlikely to be more than one protein molecule thick. (*ii*) The size of the images agrees with the size expected of a complex containing seven Sm2 molecules (it would be remarkable if a second layer of these molecules were completely occluded in Fig. 4*C*). Second, 60 kDa or more, the expected molecular mass of the heptameric complex (7 × 8.5 kDa plus the bound RNA), is in good agreement with that of the minor peak observed during purification of the Sm2 protein from *E. coli* extracts (Fig. 2), presumably arising from Sm2 and a fragment of endogenous RNA from the host *E. coli* (Fig. 2*A*). The heptamer stoichiometry appears to be invariable, because no stable complexes of smaller size assemble onto short RNAs (Fig. 3*B*). Moreover, elongation of the oligo(U) leads to the assembly of a second RNP onto the same RNA rather than to a gradual increase in the mass of bound protein (Fig. 3*C*), and the minimal RNA length required for this supershift is more than double the length required for assembly of a single RNP, indicating that a spacer is needed between the two independent complexes. Importantly, the ring structure of the Sm2 RNP (Fig. 4) closely resembles the shape of the Sm core RNP and of the LSm complex observed under the electron microscope (6, 20). In contrast to the Sm and LSm proteins, Sm2 consists of a minimal Sm domain with no extra domains (Fig. 1). Therefore, our results show that the propensity to oligomerize into these characteristic ring shapes is a property of the Sm domain alone, in line with crystallographic data showing that the SmD1·SmD2 and SmB·SmD3 dimerization interfaces involve the

β4 and β5 sheets, which are highly conserved features of the Sm domain (21). The crystallographic data suggest that the Sm core is a heptamer (21), and this model agrees with the fact that the Sm core RNP (2) and the complex of LSm proteins specific for U6 SnRNA (6, 11, 12) both contain seven distinct proteins. However, the stoichiometry of these complexes has not yet been determined, leaving the possibility that one or more of the proteins is present in two copies. Our data thus lend firm experimental evidence to the stoichiometric postulate inherent in the structural model of Kambach *et al.* (21).

Like the Sm and LSm proteins, Sm2 binds preferentially to oligo(U) RNA, with only weak competition by poly(C) and none at all by poly(A) or poly(G) (Fig. 3*A*). Furthermore, the binding of RNA by the Sm2 proteins shares several features with the binding of U snRNA by the Sm proteins. First, Sm2 discriminates against DNA oligonucleotides (data not shown), indicating that Sm2, like the Sm proteins, contacts both the bases and the sugar moieties (16, 30). Second, no stable oligo(U) complexes are detected with fewer than seven Sm2 monomers (Fig. 3), suggesting that the strong RNA binding activity of the Sm2 protein is also present on multimers only. Similarly, the Sm proteins bind the RNA in a composite binding pocket that is present only on complexes containing at least five Sm proteins (16). The Sm2 protein consists of a minimal, canonical Sm domain, implying that the Sm domain is responsible for both protein oligomerization and specific RNA binding. This idea is strongly supported by the finding that homologous amino acids at the heart of the Sm motif of both SmG and SmB are in intimate contact with uracil bases of the Sm site (31).

In summary, the archaeal Sm2 protein has all of the properties expected of an ancestor of the eukaryotic Sm protein family. It therefore seems probable that Sm2 plays a role *in vivo* that is comparable to that of the Sm and LSm proteins, i.e., that it forms oligomers with a ring structure that binds to and stabilizes a target RNA. The *A. fulgidus* genome contains two ORFs encoding Sm-related proteins (23). We consider it likely that the two proteins form two distinct complexes rather than a single heteromer. First, most archaea contain only one Sm-related ORF, and this protein must, therefore, be capable of forming homooligomers. Second, nonrandom incorporation of two distinct proteins into a seven-member ring is not possible (unless the symmetry is broken by other interacting species), and a defined heteromer of two Sm proteins is, therefore, hard to envisage. At the RNA level, it is tempting to speculate that the archaeal Sm proteins bind to stable RNA(s), and the binding site(s) should contain oligo(U) stretches similar to the ones found in the Sm site (16, 22) and the LSm binding site on the U6 snRNA (6). Experiments are currently under way to determine the *in vivo* RNA target for Sm2 binding in archaea.

The evolutionary evidence based on sequence comparisons (Fig. 1) and the biochemical evidence presented herein lead us to conclude that an ancestor of the Sm protein family existed before the archaeal and eukaryotic kingdoms diverged. From this ancestor, the eukaryotic Sm and LSm proteins evolved. The tendency to amplify the Sm proteins is also witnessed in the archaea, because only two of six species contain a second Sm protein, and these additional Sm proteins must have arisen by independent gene duplication events (data not shown). In the eukaryotes, the Sm proteins have at least three cellular functions (see Introduction). This is only possible with different sets of Sm proteins that have different RNA sequence requirements. Moreover, the Sm and LSm proteins carry different extra domains that might well modulate their oligomerization behavior and the subcellular localization signal present on the Sm ring. Finally, each Sm and LSm complex contains seven distinct proteins. This asymmetry is necessary for recognition of sequence elements flanking the oligo(U) tract. For example, the adenosine preceding the U-rich stretch on the Sm site is required for the high stability of the Sm core complex (16), and the AU dinucleotide is invariably contacted by the SmG protein (31, 32). Thus, although the Sm protein family evolved from a single ancestor that was present before the evolution of the cell nucleus, the diversification into a larger protein family has made it possible for the Sm proteins to meet the specific requirements of the eukaryotic cell.

We thank T. Hurek for helpful advice and discussions and N. Watkins for critically reading the manuscript. This work was supported by the Gottfried Wilhelm Leibniz program and a grant of the Deutsche Forschungsgemeinschaft (SFB523/A8) to R.L.

1. Cooper, M., Johnston, L. H. & Beggs, J. D. (1995) *EMBO J.* **14,** 2066–2075.
2. Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. & Lührmann, R. (1995) *EMBO J.* **14,** 2076–2088.
3. Séraphin, B. (1995) *EMBO J.* **14,** 2089–2098.
4. He, W. & Parker, R. (2000) *Curr. Opin. Cell Biol.* **12,** 346–350.
5. Fromont-Racine, M., Rain, J. C. & Legrain, P. (1997) *Nat. Genet.* **16,** 277–282.
6. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. & Lührmann, R. (1999) *EMBO J.* **18,** 5789–5802.
7. Vidal, V. P., Verdone, L., Mayes, A. E. & Beggs, J. D. (1999) *RNA* **5,** 1470–1481.
8. Stevens, S. W. & Abelson, J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 7226–7231.
9. Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R. & Fabrizio, P. (1999) *EMBO J.* **18,** 4535–4548.
10. Pannone, B. K., Xue, D. & Wolin, S. L. (1998) *EMBO J.* **17,** 7442–7453.
11. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. & Séraphin, B. (1999) *EMBO J.* **18,** 3451–3462.
12. Mayes, A. E., Verdone, L., Legrain, P. & Beggs, J. D. (1999) *EMBO J.* **18,** 4321–4331.
13. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. & Séraphin, B. (2000) *EMBO J.* **19,** 1661–1671.
14. Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D. & Parker, R. (2000) *Nature (London)* **404,** 515–518.
15. Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L. & Cech, T. R. (1999) *Nature (London)* **401,** 177–180.
16. Raker, V. A., Hartmuth, K., Kastner, B. & Lührmann, R. (1999) *Mol. Cell. Biol.* **19,** 6554–6565.
17. Raker, V. A., Plessel, G. & Lührmann, R. (1996) *EMBO J.* **15,** 2256–2269.
18. Fury, M. G., Zhang, W., Christodoulopoulos, I. & Zieve, G. W. (1997) *Exp. Cell Res.* **237,** 63–69.
19. Camasses, A., Bragado-Nilsson, E., Martin, R., Séraphin, B. & Bordonné, R. (1998) *Mol. Cell. Biol.* **18,** 1956–1966.
20. Kastner, B., Bach, M. & Lührmann, R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1710–1714.
21. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J. & Nagai, K. (1999) *Cell* **96,** 375–387.
22. Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B. & Jacob, M. (1982) *EMBO J.* **1,** 1259–1265.
23. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., *et al*. (1997) *Nature (London)* **390,** 364–370.
24. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254.
25. van Heel, M. & Frank, J. (1981) *Ultramicroscopy* **6,** 187–194.
26. van Heel, M., Harauz, G. & Orlova, E. V. (1996) *J. Struct. Biol.* **116,** 17–24.
27. Dube, P., Tavares, P., Lurz, R. & van Heel, M. (1993) *EMBO J.* **12,** 1303–1309.
28. Marco, S., Urena, D., Carrascosa, J. L., Waldmann, T., Peters, J., Hegerl, R., Pfeifer, G., Sack-Kongehl, H. & Baumeister, W. (1994) *FEBS Lett.* **341,** 152–155.
29. Yu, X., Jezewska, M. J., Bujalowski, W. & Egelman, E. H. (1996) *J. Mol. Biol.* **259,** 7–14.
30. Hartmuth, K., Raker, V. A., Huber, J., Branlant, C. & Lührmann, R. (1999) *J. Mol. Biol.* **285,** 133–147.
31. Urlaub, H., Raker, V. A., Kostka, S. & Lührmann, R. (2001) *EMBO J.* **20,** 187–196.
32. Heinrichs, V., Hackl, W. & Lührmann, R. (1992) *J. Mol. Biol.* **227,** 15–28.