The Archael Lsm Protein Binds to Small RNAs*

Received for publication, March 1, 2010, and in revised form, July 18, 2010. Published, JBC Papers in Press, September 7, 2010, DOI 10.1074/jbc.M110.118950

Susan Fischer†, Juliane Benz‡, Bettina Späth‡, Lisa-Katharina Maier‡, Julia Straub‡, Michaela Granzow§, Monika Raabe‡, Henning Urlaub‡, Jan Hoffmann‡, Bernd Brutschy‡, Thorsten Allers**, Jörg Soppa‡, and Anita Marchfelder‡

From the Institute of Biology II, Ulm University, 89069 Ulm, Germany, the Institute of Molecular Biosciences, Goethe-University, 60438 Frankfurt, Germany, the Max-Planck-Institute for Biophysical Chemistry, Bioanalytical Mass Spectrometry Group, 37077 Göttingen, Germany, the Institute of Physical and Theoretical Chemistry, Goethe-University, 60438 Frankfurt, Germany, and the Institute of Genetics, University of Nottingham, Queen’s Medical Centre, Nottingham NG7 2UH, United Kingdom

Proteins of the Lsm family, including eukaryotic Sm proteins and bacterial Hfq, are key players in RNA metabolism. Little is known about the archael homologues of these proteins. Therefore, we characterized the Lsm protein from the haloarchaeon Haloferax volcanii using in vitro and in vivo approaches. H. volcanii encodes a single Lsm protein, which belongs to the Lsm1 subfamily. The lsm gene is co-transcribed and overlaps with the gene for the ribosomal protein L37e. Northern blot analysis shows that the lsm gene is differentially transcribed. The Lsm protein forms homoeptameric complexes and has a copy number of 4000 molecules/cell. In vitro analyses using electrophoretic mobility shift assays and ultrasoft mass spectrometry (laser-induced liquid bead ion desorption) showed a complex formation of the recombinant Lsm protein with oligo(U)-RNA, tRNAs, and an small RNA. Co-immunoprecipitation with a FLAG-tagged Lsm protein produced in vivo confirmed that the protein binds to small RNAs. Furthermore, the co-immunoprecipitation revealed several protein interaction partners, suggesting its involvement in different cellular pathways. The deletion of the lsm gene is viable, resulting in a pleiotropic phenotype, indicating that the haloarchael Lsm is involved in many cellular processes, which is in congruence with the number of protein interaction partners.

Sm and Sm-like (Lsm) proteins constitute a large family of proteins known to be involved in RNA metabolism. Representatives of this family are found in all three domains: bacteria, archaea, and eukarya. All of them share a common bipartite sequence motif, known as the Sm domain, consisting of two conserved segments separated by a region of variable length and sequence. The bacterial family member is the Hfq protein (1, 2), which has a plethora of functions (3). Hfq is a highly conserved protein encoded within many bacterial genomes (4). Although the protein does not show a high similarity to the Lsm proteins on the primary structure level, it possesses striking similarities in both function and tertiary and quaternary structure to the eukaryotic Lsm proteins (3, 5). Hfq monomers assemble to form highly stable hexamers (6), which bind preferentially to A/U-rich sequences (7, 8) but have a relaxed RNA binding specificity and participate in many stages of RNA metabolism. It was therefore proposed that Hfq is an ancient, less specialized form of the Lsm proteins (9). One of the identified functions of Hfq is its interaction with sRNAs (10). It has been proposed that the protein acts as an RNA chaperone that might simultaneously recognize the sRNA and its target and facilitate its interaction. An Escherichia coli hfq insertion mutant showed pleiotropic phenotypes including decreased growth rates and yields, increased cell sizes, and an increased sensitivity to stress conditions (11–13). These defects are at least in part a reflection of the fact that Hfq is required for the function of several sRNAs including DsrA, RprA, Spot42, OxyS, and RhyB (14–17).

Eukaryotes have the most diverse members of the Sm/Lsm protein family. They contain at least 18 different Sm and Lsm proteins involved in mRNA splicing, histone maturation, telomere maintenance, and mRNA degradation that form at least six different heteroheptameric complexes (18). The Lsm proteins alone form at least two heteroheptameric complexes: the nuclear Lsm2-8, a large fraction of which associates with U6 snRNA, and the cytoplasmic Lsm1-7, which functions in mRNA degradation (19, 20). The Lsm proteins that associate with U6 snRNA are necessary for its stability (21–23), binding to the U-rich region at the 3′ end of the U6 snRNA. Additional functions of the nuclear Lsm proteins are the involvement in processing pre-snoRNA, pre-rRNA, pre-tRNA precursor, and nuclear pre-mRNA decay (5).

The fact that Lsm proteins have been found in archaea (22–25) suggests that they were present in a common ancestor shared by archaea and eukarya. This correlates with the observation that several eukaryotic proteins clearly evolved from archaea-related precursors (26) and that snoRNAs have also been found in archaea (27). Some archaea, such as the Pyrococcus species and halophilic archaea, encode only one Lsm protein (Lsm1), whereas others encode two (Lsm1 and Lsm2) (23). The Lsm1 and Lsm2 proteins have been shown to be associated in vivo (28), so they might also form heteromeric complexes. Crenarchaeota have an additional Lsm protein, Lsm3, which

---

*This work was supported by the Deutsche Forschungsgemeinschaft through SPP 1258 Priority Program “Sensory and Regulatory RNAs in Prokaryotes” Grants DFG Ma1538/11-1 and So264/14-1.
†The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S3 and Figs. S1–S6.
‡To whom correspondence should be addressed. E-mail: anita.marchfelder@uni-ulm.de.
§The abbreviations used are: snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; LILBID, laser-induced liquid bead ion desorption; sRNA, small RNA.
Archaeal Lsm Protein

contains a traditional Sm domain fused to a second domain by a flexible linker (29, 30).

The archaeal Lsm1 proteins have been shown to form heptamers (28, 31–33) and bind RNA (18, 28, 31). The Lsm2 protein from *Archaeoglobus fulgidus* has been reported to form a hexameric (34) or a heptameric (31) complex. The Lsm3 protein has also been shown to form 14-mer complexes (30), a process of which some of the Lsm1 proteins are also capable (35).

Interestingly, *Methanocaldococcus jannaschii* lacks a classical Lsm gene (9, 36) but contains an Hfq-like protein. Although some data have been acquired on the structure and RNA binding characteristics of the archaeal Lsm protein, so far the function and interaction partners of the Lsm protein in archaea have not been revealed.

Here, we analyze the Lsm protein from the halophilic archaeon *Haloferax volcanii*. *H. volcanii* encodes only one Lsm protein, which makes it easier to employ genetic methods for analyzing the biological function of the archaeal Lsm proteins. Recently, it has been shown that *H. volcanii* also has an sRNA population potentially involved in gene expression regulation (37, 38). To investigate whether the *Haloferax* Lsm is involved in sRNA regulation and to clarify its biological function, we generated a deletion strain for Lsm and analyzed the *in vivo* and *in vitro* function of this protein.

**EXPERIMENTAL PROCEDURES**

**Strains and Culture Conditions—** *H. volcanii* strains H119 (Δ*pyrE2, ΔtrpA, ΔleuB*) (39) and Δ*lsm* (Δ*pyrE2, ΔtrpA, ΔleuB, Δ*lsm*) were grown aerobically at 45 °C in Hv-YPC or Hv-Ca medium (39). The term “normal growth conditions” for *H. volcanii* used here stands for aerobic growth at 45 °C and 2.1 mM NaCl.

**Generation of the pTA927 Vector—** To generate pTA927, a 131-bp KpnI fragment containing the terminator sequence of the *H. volcanii* L11e rRNA gene (40) was inserted at the KpnI site of pTA230 (39). Next, a 224-bp region of the *tnaA* promoter (41) was amplified by PCR and inserted at the Apal and ClaI sites; the reverse primer incorporated a novel NdeI site (at the NotI restriction sites NcoI and NotI, respectively). The resulting PCR product was digested with NcoI and NotI and cloned into the vector pTA927 (available upon request) and Sm2, which contained the restriction sites NcoI and NotI, respectively. The resulting PCR product was digested with NcoI and NotI and cloned into the vector pET29a (Novagen), which was previously digested with the same restriction enzymes, yielding the plasmid pET29a-Sm. pET29a-Sm was transformed into BI21-Al (Novagen), and the Lsm protein was expressed and purified according to the manufacturer’s protocol using S-protein–agarose (Novagen). For the production of antibodies, 0.5 mg of purified protein were sent to Davids Biotechnology (Regensburg, Germany).

**Western Blot Analysis and Determination of Lsm Copy Number—** For Western blot analysis cytoplasmic extracts of *H. volcanii* (20 μg) were separated by SDS-PAGE and transferred to a nylon membrane by semi-dry blotting (1.5 h with 2 mA/cm²). The membrane was blocked using skimmed milk powder, incubated with the newly generated antiserum (see above) or the preimmune serum at dilutions of 1:500, washed, and incubated with the secondary, peroxidase-conjugated goat anti-rabbit antibody. Peroxidase activity was detected with the chemiluminescence substrates luminol and para-hydroxy coumaric acid. Light emission was detected with films. The generated antiserum reacted with several bands, all but one of which also reacted with the preimmune serum. The specific band had the expected size of ~9 kDa.

For the quantification of the Lsm copy number, cytoplasmic extracts were prepared of 2.3 × 10⁸ *H. volcanii* cells. They were used for Western blot analysis alongside with 1–50-ng aliquots of purified Lsm protein. The film was scanned, and the signals were quantified using ImageJ. The aliquots of the purified Lsm protein were used to generate a standard curve, which was used to quantify the Lsm amount in cell extracts. The value was used to calculate the Lsm molecules/cell using a molecular mass of 8.25 kDa.

**Substrate Preparation and Binding of the Recombinant Protein to RNA—** Substrates for the electrophoretic mobility shift assays were prepared as follows. U₁₅⁻ and U₃₀-RNA oligonucleotides were generated by Sigma. Wheat tRNA (tRNA isolated from wheat germ, type V; Sigma) and oligo(U)-RNA were labeled at the 3′ end using [α-³²P]pCp as described (47). EMSAs were carried out as described (48) with the following modifications: 100 ng of recombinant Lsm protein was used if not indicated otherwise. For the determination of the dissociation constant of Lsm/oligo(U)-RNA (K₅), 0.1 pmol (0.6 ng) of U₁₅-RNA were incubated with increasing amounts of Lsm protein.
(0.012–24 pmol, equating 0.1–200 ng). If all of the proteins are present as homoheptameric complexes, 0.1–200 ng of protein would equate to 1.7 fmol–3.4 pmol of Lsm complexes.

**Laser-induced Liquid Bead Ion Desorption-MS—LILBID** is a novel mass spectrometry method that allows an exact mass determination of single macromolecules dissolved in droplets of solution containing an adequate buffer, pH, ion strength, etc., as described previously (49). Briefly, droplets of solution of analyte are ejected by a piezo-driven droplet generator and transferred into a high vacuum. There, they are irradiated droplet by droplet (d = 50 μm, V = 65 pl, 10 Hz) by a pulsed IR laser tuned to the stretching vibration of water at 2.9 μm. By laser ablation the droplets explode, ejecting preformed biomolecular ions into the vacuum. The total volume of solution required for the mass determination is only few microliters in typically micromolar concentration. The method is ideal for studying biomolecules of low availability (49). The amount of energy transferred into noncovalent complexes by the IR desorption/ablation process can be controlled in a wide range, starting from ultrasoft to harsh conditions, just by varying the laser intensity (50). At ultrasoft conditions large macromolecules can be detected in their native stoichiometry. The complexes are detected in different charged states, preferentially as anions. The number of charge states observed increases with the size of the molecules but is less than those observed in electro spray ionization and considerably more than in MALDI.

To investigate the quaternary structure of the Lsm protein, we dialyzed the recombinant protein against a buffer (50 mM NaCl, 10 mM of Tris/HCl, pH 7.5). Complexes were analyzed using LILBID-MS. To analyze the binding of Lsm proteins to oligo(U)-RNA, 8 μM oligo(U)-RNA (U₃₀) were incubated at room temperature for 30 min with 4 μM of heptameric Lsm complexes in a buffer containing 20 mM NaCl, 2 mM MgCl₂, and 10 mM Tris/HCl, pH 7.5. The resulting complexes were analyzed using LILBID-MS. To investigate the binding to sRNAs, 4 μM of Lsm heptamers were incubated with 8 μM of sRNA₃₀ at room temperature for 30 min. Again the resulting complexes were analyzed using LILBID-MS.

**Generation of the lsm Deletion Strain**—The lsm reading frame was completely removed in the *H. volcanii* strain H119 using the pop-in/pop-out method (39, 51). The upstream and downstream regions of the *lsm* gene were amplified by PCR using chromosomal DNA from *H. volcanii* and primers SmKO/FLAG1, SmKO3, SmKO/FLAG2, and SmKO4, respectively, yielding fragments Sm1 and Sm2, both ~1 kb long. PCR primers contained different restriction sites: ApaI (SmKO/FLAG1), EcoRV (SmKO3), EcoRV (SmKO/FLAG2), and XbaI (SmKO4). Both PCR fragments were first cloned into pBluescriptII (Stratagene), yielding plasmids pblue-Sm1 and pblue-Sm2 and subsequently subcloned into the integrative vector pTA131 containing the *pyrE2* marker (39), yielding pTA131-Sm1/2. This plasmid was integrated into the chromosomal DNA of *H. volcanii* (strain H119, pop-in). The plasmid containing the *pyrE2* marker was forced out by plating the cells on 5-fluoro-orotic acid (pop-out). Southern blot analysis was carried out as described in Ref. 52 with the following modifications. Chromosomal DNA was isolated from wild type and knock-out strains and digested using Xhol. 10 μg of digested DNA was separated on a 0.8% agarose gel and transferred to a nylon membrane (HybondTM-N; GE Healthcare). Hybridization probe Sm1 was generated by PCR using primers SmKO/FLAG1 and SmKO3 on template pblue-Sm1, yielding a 1-kb fragment, which was subsequently radioactively labeled using the random prime kit ReadiprimeTM II (GE Healthcare).

**Co-immunoprecipitation**—To isolate an S100 extract, the cells were grown to stationary phase in Hv-Ca⁺ broth including 0.25 mM tryptophane and harvested at OD₆₅₀ = 2.8. The cells were pelleted, and the resulting pellets were washed with enriched PBS (2.5 mM NaCl, 150 mM MgCl₂, 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂PO₄, 2 mM KH₂PO₄, pH 7.4)). The cells were again pelleted, resuspended in enriched PBS containing 1% formaldehyde, and incubated for 20 min at 45 °C. To stop the cross-linking reaction, glycine was added to a final concentration of 0.25 M and incubated for 5 min at 45 °C. The cells were washed twice with enriched PBS at 4 °C, and then lysis buffer (50 mM Tris, pH 7.4, 1 mM EDTA, 10 mM MgCl₂, 1 mM CaCl₂) containing 150 μl of proteinase inhibitor (Sigma) was added. After ultracentrifugation (100,000 × g for 30 min) RNase A was added to a final concentration of 400 μg/ml extract, and the mixture was incubated for 30 min at 37 °C. Subsequently, NaCl was added to a final concentration of 150 mM, and the lysate was frozen at −80 °C. For affinity purification, 1.6 ml of anti-FLAG M2 affinity gel (Sigma) was washed 10 times with 10 ml of ice-cold washing buffer (50 mM Tris/HCl, pH 7.4, 150 mM NaCl) before the lysate was added. After incubation overnight (14–16 h) at 4 °C, anti-FLAG M2 affinity gel was washed eight times with 10 ml of ice-cold washing buffer. The elution of the FLAG fusion protein was performed by using 4 ml of washing buffer, to which 3× FLAG peptide was added (final concentration, 150 ng/μl). The samples were incubated at 4 °C with gentle shaking. In a final elution step the affinity gel was rinsed with 2 ml of washing buffer. For the isolation of co-precipitated RNA, the cross-link reaction was released by incubating the samples at 95 °C for 20 min. The fraction was treated with 20 μg of proteinase K for 30 min at 37 °C in 100 μl of buffer (100 mM Tris/HCl, pH 7.5, 12.5 mM EDTA, 150 mM NaCl, 0.2% SDS). The solution was extracted with phenol-chlorform-isoamylalcohol. The aqueous phase containing RNA was precipitated, and the resulting pellet was dissolved in water. An aliquot of the RNA fraction was 3'-labeled with [α-³²P]pCp as described (53).

**Mass Spectrometry**—For mass spectrometric analysis proteins associated with the FLAG only, the FLAG-Lsm (without cross-link), and the FLAG-Lsm (with cross-link) proteins were dissolved in 1× loading buffer, and cross-linked samples were incubated for 20 min at 95 °C. The samples were then loaded onto a 4–12% NuPAGE-Gel (Invitrogen). After Coomassie staining, the gel lanes were cut into 23 slices, and the proteins were in-gel digested with trypsin according to Ref. 54. Extracted peptides were analyzed by LC-MS/MS on a Q-ToF instrument (Waters) under standard conditions. Peptide fragment spectra were searched against a target decoy database for *H. volcanii* (46) using MASCOT as a search engine. Peptides with a peptide score lower than 25 were omitted from the results. Scaffold software (Proteome Software Inc., Portland, OR) was used for
Archaeal Lsm Protein

A.

B.

FIGURE 1. Genomic location of the Lsm protein gene. A, the operon containing the genes for Lsm and L37e is bordered by the gene for a potential hydrolase (HVO_2724) and the gene for an amidophosphoribosyltransferase (HVO_2721). The genomic region is given below in nucleotides. B, the reading frames for the Lsm protein (shown in light gray) and the ribosomal protein L37e (shown in dark gray) overlap by 4 nt (shown in bold type).

FIGURE 2. Differential expression of the lsm gene. A, the transcript levels of the lsm gene in cells grown under different conditions were determined by Northern blot analysis. B, the corresponding agarose gel shows that all of the lanes were loaded with the same amount of RNA. The following conditions were applied: aerobic growth in 2.1 M NaCl at 42 °C (lane 1), at 30 °C (lane 2), and at 48 °C (lane 3); anaerobic, nitrate-respirative growth (lane 4); and 1.5 M NaCl (lane 5).

data evaluation (see supplemental tables). The proteins that were co-purified with the FLAG peptide in the control reaction were subtracted from the proteins co-purified with the FLAG-Lsm protein. In Table 1 only proteins, which were present in all three FLAG-Lsm purifications with at least four MS/MS spectra in each of the three independent isolations are listed. The complete list of identified proteins is shown in supplemental Table S3.

DNA Microarray Analysis—The affinity-purified FLAG-tagged Lsm complexes and a negative control (FLAG peptide not tagged to Lsm) were used for RNA isolation as described (see “Co-immunoprecipitation” above). 1-µg aliquots of the two fractions were used for cDNA synthesis, labeling, and DNA microarray analysis as described, using a self-constructed DNA microarray for H. volcanii (55). sRNA-specific oligonucleotide probes were added to the DNA microarray to allow the analysis of sRNA gene expression. Three independent experiments were performed, including a dye swap. The analysis of DNA microarray results was performed as described (55).

RESULTS AND DISCUSSION

Little is known about the archaeal Lsm proteins and therefore we were interested in unraveling the function of the archaeal Lsm protein using in vitro and in vivo approaches.

The Lsm Reading Frame Overlaps with the Reading Frame for a Ribosomal Protein—Using a BLAST search (56) with previously described archaeal Lsm proteins (23), we identified the Lsm protein gene in the genome of H. volcanii (46). H. volcanii contains a single lsm gene (HVO_2723), which encodes a protein of 76 amino acids with a molecular mass of 8.25 kDa and a pI of 3.9. The Haloferax Lsm protein was found to belong to the Lsm1 subfamily of Lsm proteins. The lsm gene overlaps by four nucleotides with a gene annotated to encode the L37e ribosomal protein (HVO_2722; Fig. 1). To analyze the conservation of gene order and Lsm protein sequence in the domain archaea, a BLAST search was used to identify similar proteins and their genes. The gene order is highly conserved in archaea. In more than 40 archaeal genomes, the gene for the L37e proteins follows the lsm gene. In more than 30 genomes, the two genes are very closely spaced or overlap, so that co-transcription can be assumed. The multiple sequence alignment of the H. volcanii Lsm1 and 31 other archaeal Lsm1 proteins (supplemental Fig. S1A) shows that the protein is highly conserved in archaea with the exception of the regions corresponding to β-sheets 2 and 3 in the structure of the P. abyssi Lsm (18), which is variable in the whole family and especially in the six haloarchaeal Lsm proteins. It should be noted that the three residues that form a highly specific binding pocket for uridine in the P. abyssi Lsm are universally conserved, indicating specific uridine binding in all of the archaeal Lsm1 proteins.

Expression of the lsm Gene and Determination of Lsm Copy Number—Northern blot analyses were used as a first approach to analyze the expression of the lsm gene. Using a probe against the two overlapping genes, two transcripts of ~430 and ~210 nt, respectively, could be detected (Fig. 2). Gene-specific probes revealed that the smaller transcript was derived from the gene for the L37e protein, which can either be a primary transcript initiated from a promoter localized within the open reading frame of the upstream located lsm gene or originate from the processing of the bicistronic transcript. According to the genome sequence, a bicistronic transcript should be 404 nt, and a transcript encoding only L37e should be 177 nt. Using circularized RNA RTPCR (45, 57), we determined that the bicistronic transcript is leaderless and contains a 3′-UTR of 41 nt, in excellent agreement with the Northern blot results (data not shown).

To observe a potential differential regulation of transcription, Northern blot analyses were performed using RNA from cells cultivated under different conditions. During aerobic growth, the transcript levels did not change throughout the
growth curve, from early exponential to stationary phase (data not shown). It was also identical during growth at low salt (1.5 M NaCl; Fig. 2), high salt (3 M NaCl, data not shown), and high temperature (48 °C; Fig. 2). By contrast, both the bicistronic and the monocistronic transcript were undetectable in the cultures grown at a low temperature (30 °C; Fig. 2) or via nitrate-regenerative growth (Fig. 2). Taken together, both transcripts were apparently co-regulated and are present in *H. volcanii* under most but not all conditions.

For the analysis of the Lsm protein, we expressed the *lsm* gene in *E. coli* to produce a recombinant protein. The gene was efficiently expressed to yield a pure fraction of recombinant Lsm protein (supplemental Fig. S3 against which an antiserum was generated. Western blot analysis was used for the relative quantification of the protein levels in cytoplasmic extracts from cells grown at different salt concentrations (1.2, 2.5, and 3 M) either to the exponential or stationary phase. In each case, the Lsm protein levels were identical; thus, we found no indication for translational regulation (data not shown). For the absolute quantification of the intracellular protein level, a standard curve was generated using heterologously produced and purified Lsm protein (see below), revealing that *H. volcanii* contains ~4,000 Lsm molecules/cell (supplemental Fig. S2). By contrast, 50,000–60,000 copies of Hfq are present in rapidly growing *E. coli* cells in the exponential phase, but the level is rapidly down-regulated to ~20,000 copies/cell at the onset of the stationary phase (3, 58). We found no reports about intracellular copy numbers of additional Lsm proteins, neither in procaryotes nor in eukaryotes.

*The Recombinant Lsm Protein Forms Homoeptamers*—To investigate whether the *Haloferax* Lsm protein forms homoeptameric complexes, we employed ultrasoft mass spectrometry (LILBID-MS) (see “Experimental Procedures” for details) (49). This approach revealed that the protein forms homoheptamers in *vitro* (Fig. 3A and data not shown). Under harsh conditions (high laser intensity), the complex could be fragmented and masses corresponding to Lsm monomers, dimers, trimers, and tetramers were observed (Fig. 3B). Other archaeal Lsm1-type proteins also form homoheptamers (28, 31–33), in contrast to proteins of the Hfq and Lsm2 subfamilies, which form exclusively homohexamers (Hfq) or have the potential to form homohexamers (Lsm2) (6, 36). The eukaryotic Lsm proteins have been shown to form heteroheptamers (28, 31–33). Therefore, as for other archaeal proteins involved in transcription, replication, or translation, the archaeal Lsm proteins can be regarded as a closer mimic and simpler model for the eukaryotic proteins, which have added further complexity during evolution. Thus, the archaeal Lsm proteins are much better models for the eukaryotic proteins than the bacterial Hfq protein (5).

*Characterization of Lsm-RNA Interactions in Vitro*—To analyze whether the recombinant Lsm protein binds RNA, we incubated it with oligo(U)-RNA (U15- and U30-RNA) and investigated the interaction using EMSA. The gel shift analysis showed that the recombinant Lsm indeed binds U30-RNA. Using U30-RNA and increasing Lsm protein concentrations, we determined the dissociation constant *Kd* to be 72 nM (Fig. 4A and supplemental Fig. S4). Binding to oligo(U)-RNA has been shown for eukaryotic Lsm proteins (31, 59), for other archaeal ones (31), and also for Hfq (6). The physiological significance of the archaeal Lsm binding to oligo(U)-RNA is unclear because oligo(U) stretches have not been identified in the RNA population from *Haloferax* so far.

Because the *E. coli* Hfq and the yeast Lsm protein were suggested to be involved in tRNA processing and modification (60–62), we incubated the archaeal Lsm protein with tRNAs. EMSA revealed that the *Haloferax* Lsm protein also binds to tRNAs (Fig. 4B).

*Native Mass Spectrometry Confirms Lsm-RNA Interactions*—An additional approach to study RNA binding by Lsm and to unravel the stoichiometry of complex formation LILBID-MS was used. Purified Lsm protein was incubated with U30-RNA, and mass spectrometry analysis under ultrasoft conditions (low laser intensity) confirmed that one Lsm heptamer bound to U30-RNA and revealed in addition that another complex forms consisting of two Lsm heptamers bound to U30-RNA (Fig. 5A). Analysis under harsh conditions (high laser intensity) revealed that the ternary complex was very
Archaeal Lsm Protein

exhibited an extensive lag phase before the onset of growth and had a reduced growth rate (Fig. 7). Comparison of the growth capabilities of mutant and wild type under various conditions revealed that the phenotypic difference between the two strains was variable, e.g. the mutant grew nearly as well as the wild type on casamino acids, pyruvate, xylose, and arabinose (lower growth yield on arabinose) but was severely compromised on glycerine and sucrose (data not shown). Therefore, it seems that the importance of the Lsm protein for cellular physiology is different for various metabolic pathways. To gain further insight into the function of Lsm, we decided to identify its interaction partners.

Co-immunoprecipitation Reveals Several Interaction Partners—To identify the interaction partners of the Lsm protein, we constructed a FLAG-Lsm fusion protein. For that purpose we first generated the expression vector pTA927, which is based on pTA230 (39) and features the tryptophan-inducible tnaA promoter for regulatable gene expression in Haloferax (41). Subsequently, the FLAG peptide cDNA was cloned in-frame downstream and upstream, respectively, of the Lsm reading frame into the pTA927 vector. In addition, a plasmid was constructed encoding only the FLAG peptide as a negative control. Haloferax was transformed with the plasmids, and expression was analyzed using Western blots (supplemental Fig. S6A), showing that both fusion proteins were efficiently expressed in Haloferax. The lsm deletion strain 

A.

stable, and Lsm subunits were lost, whereas the complex remains otherwise intact (Fig. 5B).

LILBID-MS was also used to clarify whether sRNAs bind to Lsm. Incubation of Lsm with sRNA$_{30}$ and subsequent analysis with LILBID-MS revealed that an Lsm-sRNA$_{30}$ complex forms but, in contrast to U$_{30}$RNA ternary complexes (Lsm-sRNA$_{30}$-Lsm), were not detected (Fig. 6).

Deletion of the Lsm Frame Is Viable—To pinpoint the biological function of the archaeal Lsm protein, we generated an lsm deletion mutant using the pop-in/pop-out method (39, 51, 63). Because the overlap of the lsm and l37e genes indicated translational coupling, care was taken to generate an in-frame deletion mutant that left translational coupling intact and avoided putative polar effects. After pop-out selection, small and large colonies were observed. Southern blot analysis revealed that only the small colonies contained the lsm deletion (termed Δlsm), and the large colonies still contained the wild type lsm gene (supplemental Fig. S5). Comparison of the Δlsm deletion mutant and the wild type under standard growth conditions (see “Experimental Procedures”) revealed that the mutant

B.

FIGURE 4. Haloferax Lsm binds to RNA. A, the recombinant Lsm protein was incubated with oligo(U)-RNA and subsequently loaded onto nondenaturing PAGE. Lanes c, control reaction without proteins; lanes 200 and 400, incubation with 200 and 400 ng of recombinant Lsm protein, respectively. U15 and U30, incubation with U$_{15}$RNA and U$_{30}$RNA, respectively. RNA and complex are shown schematically at the right. B, tRNA is also bound by the Lsm protein. Incubation of Lsm with a wheat tRNA fraction shows that Lsm also binds to tRNAs. Lane c, control reaction without protein. Lane L, incubation with 100 ng of Lsm protein. RNA and complex are shown schematically on the right.
were present in all three samples (supplemental Table S1), showing that few proteins bind to the FLAG tag. The precipitation of proteins from the FLAG-Lsm sample, which was not treated with formaldehyde before cell lysis, also revealed only very few proteins. In this case, only a single protein was identified that was present in all three independent samples, indicating that without cross-linking no specific interaction partner can be isolated (supplemental Table S2) and thus that the cross-linking step is required to identify interaction partners. The comparison of FLAG-Lsm co-immunoprecipitation with and without cross-link clearly showed that the purification procedure interrupts existing complexes and that cross-linking is required to keep the complexes intact upon lowering the salt concentration from the intracellular 2.1 M KCl to 150 mM NaCl.

To identify proteins specific for Lsm co-immunoprecipitation, the proteins identified by mass spectrometry in the control (FLAG only) (supplemental Table S1) were subtracted from those identified in the FLAG-Lsm co-immunoprecipitation, resulting in proteins specific for the Lsm co-immunoprecipitation (Table 1). Therefore, the proteins listed in Table 1 are true interaction partners, because proteins from the control co-purification (FLAG only) were subtracted, and in addition an RNase digest was performed. Altogether 33 proteins were identified; a similar high number of interaction partners has been found for the bacterial Hfq (57 proteins (64)) and the eukaryotic Sm and Lsm proteins (5). Furthermore, the proteins identified here as interaction partners belong to similar functional classes as the partners identified for the bacterial Hfq and eukaryotic Lsm proteins (5): e.g. ribosomal proteins, elongation factors, tRNA synthetases, chaperones, and ribonucleases. Details such as the regions of the Lsm protein involved in the interactions remain to be analyzed, but the apparent functional conservation of the protein is striking, and the number of interaction partners confirms the versatility of these proteins.

**Archaeal Lsm Protein Interacts with sRNAs and snoRNAs**—To identify the RNA interaction partners, the cross-link was reversed, and the RNA was isolated from this fraction. An aliquot was labeled with [α-32P]pCp, revealing several RNA molecules binding to Lsm (supplemental Fig. S6C). To further identify the RNA molecules, we employed DNA microarray analysis. Labeled cDNA was generated from the RNA, which co-purified with the FLAG-Lsm protein and with the FLAG only peptide, respectively. Competitive hybridization with a self-constructed DNA microarray led to
Archaean Lsm Protein

Diverse functions

| Protein | Number of MS/MS spectra |
|---------|-------------------------|
| 21      | Betaine aldehyde dehydrogenase | 16 |
| 22      | Chlorite dismutase family protein | 12 |
| 23      | Pyruvate-ferredoxin oxidoreductase α subunit | 10 |
| 24      | Predicted hydrolase | 10 |
| 25      | Aconitase hydratase, putative | 10 |
| 26      | Coiled-coil protein | 9 |
| 27      | Fumarate hydratase class II | 9 |
| 28      | Proteosome subunit 1 | 8 |
| 29      | Putative orotate/ubiquinone oxidoreductase | 8 |
| 30      | Phosphopyruvate hydratase | 7 |
| 31      | Aspartate carbamoyltransferase | 7 |
| 32      | Short chain family oxidoreductase | 7 |
| 33      | Inosine-5-monophosphate dehydrogenase | 6 |

the identification of 20 sRNAs that co-purified with the Lsm protein (Table 2). Several of these RNAs have recently been identified as candidate sRNAs (Ref. 38 and Table 1 therein): intergenic sRNAs 30, 34, 132, 304, and 308; antisense sRNAs 25 and 144; and sense sRNAs 15 and 93. Seven of the Lsm-binding RNAs (H225.1, p12, p5, H225.2r, H230, H11.1, and H62r) had been predicted as sRNAs using bioinformatic approaches.4 The DNA microarray results show for the first time that these predicted sRNAs are indeed expressed. A snoRNA (sRNA45) that had been predicted as C/D box snoRNA4 was also identified. Interestingly, the 7 S RNA also co-purified with the Lsm protein. The binding of Lsm to sRNAs suggests a similar function of the archaean protein in the regulatory network of sRNAs as for the bacterial Hfq protein. Unfortunately, so far no targets have been identified for an archaean sRNA; thus, the influence of Lsm on sRNA/target RNA interaction remains to be determined.

The binding of the Lsm protein to a potential C/D box snoRNA is interesting because the attachment of an Lsm protein to a snoRNA has not yet been found in archaea. The archaean C/D box snoRNAs and their function have been studied in detail in *Sulfolobus solfataricus* (65). Three proteins have been identified that associate with the snoRNA to form the methylation guide complex: L7Ae, aFib, and aNop56/58. Homologues for all three proteins are also present in *H. volcanii*. Interestingly, the aNop56/58 homologue is also found as a protein interaction partner in the co-immunoprecipitation but not L7Ae and aFib (Table 1). Because the eukaryotic counterparts of the archaean Lsm protein bind to snoRNAs, it is likely that the archaean Lsm protein binds to archaean snoRNAs. The specific role of that interaction remains to be determined.

The lsm Deletion Mutant Exhibits a Pleiotropic Phenotype—Although the Lsm protein is involved in many processes, it is not essential. The mutant has severe growth defects compared with the wild type under a variety of conditions, supporting the suggestion that Lsm is involved in many different pathways. Similar observations have been made in bacteria, where Hfq is involved in several processes. Deletions of the *E. coli* *hfq* gene

---

4 J. Straub, B. Stoll, B. Tjaden, B. Voss, W. R. Hess, A. Marchfelder, and J. Soppa, in preparation.
resulted in pleiotropic physiological effects, and the lack of phenotype under specific conditions was also observed (11, 13, 66, 67). The construction of hfq deletion mutants in other bacterial species revealed a fundamental role of Hfq in the virulence of pathogenic bacteria (67–73). No apparent phenotype emerged from an hfq knock-out in Staphylococcus aureus (74). In summary, deletion mutants of prokaryotic lsm genes revealed that Lsm proteins are involved in many processes, and their absence results in pleiotropic phenotypes.

Acknowledgments—We thank Elli Bruckbauer for expert technical assistance and Katharina Jantzer for growth experiments with the lsm mutant. Furthermore, we thank members of the SPP 1258 Priority Program for helpful discussions.

REFERENCES

1. Moller, T., Franch, T., Hejrup, P., Keene, D. R., Bächinger, H. P., Brennan, R. G., and Valentijn-Hansen, P. (2002) Mol. Cell 9, 23–30
2. Zhang, A., Wasserman, K. M., Ortega, J., Steven, A. C., and Storz, G. (2002) Mol. Cell 9, 11–22
3. Valentijn-Hansen, P., Eriksen, M., and Udesen, C. (2004) Mol. Microbiol. 51, 1525–1533
4. Sun, X., Zihlin, U., and Wartell, R. M. (2002) Nucleic Acids Res. 30, 3662–3671
5. Wilusz, C. J., and Wilusz, J. (2005) Nat. Struct. Mol. Biol. 12, 1031–1036
6. Schumacher, M. A., Pearson, R. F., Moller, T., Valentijn-Hansen, P., and Brennan, R. G. (2002) EMBO J. 21, 3546–3556
7. Seneor, A. W., and Steitz, J. A. (1976)
8. de Haseth, P. L., and Uhlenbeck, O. C. (1980)
9. Nielsen, J. S., Bøggild, A., Andersen, C. B., Nielsen, G., Boysen, A., Brodersen, D. E., and Valentijn-Hansen, P. (2007) RNA 13, 2213–2223
10. Wasserman, K. M., Repolla, F., Rosenow, C., Storz, G., and Gottesman, S. (2001) Genes Dev. 15, 1637–1651
11. Tsui, H. C., Leung, H. C., and Winkler, M. E. (1994) Mol. Microbiol. 13, 35–49
12. Brown, L., and Elliott, T. (1996) J. Bacteriol. 178, 3763–3770
13. Muffler, A., Fischer, D., and Hengge-Aronis, R. (1996) Genes Dev. 10, 1143–1151
14. Massé, E., and Gottesman, S. (2002) Proc. Natl. Acad. Sci. U.S.A. 99, 4620–4625
15. Moller, T., Franch, T., Udesen, C., Gerdes, K., and Valentijn-Hansen, P. (2002) Genes Dev. 16, 1696–1706
16. Slodjeski, D. D., Whitman, C., and Zhang, A. (2001) J. Bacteriol. 183, 1997–2005
17. Zhang, A., Altuvi, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998) EMBO J. 17, 6061–6068
18. Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003) J. Biol. Chem. 278, 1239–1247
19. Tharun, S., He, W., Hayes, M. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000) Nature 404, 515–518
20. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Säfström, B. (2000) EMBO J. 19, 1661–1671
21. Pannone, B. K., Xue, D., and Wolin, S. L. (1998) EMBO J. 17, 7442–7453
22. Mayes, A. E., Verdon, L., Legrain, P., and Beggs, J. D. (1999) EMBO J. 18, 4321–4331
23. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Säfström, B. (1999) EMBO J. 18, 3451–3462
24. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., and Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Blaisdell, F. W., Cole, S. T., Wing, M. C., and Olsen, G. I., Fraser, C. M., and Venter, J. C. (1995) Science 277, 335–342
25. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilber, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbokowski, J., Gibson, R., Juwen, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shim, G., Goyal, A., Pirtetrokovski, S., Church, G. M., Daniels, C. J., Miao, J. L., Rice, P., Nölling, J., and Reeves, J. N. (1997) J. Bacteriol. 179, 7135–7155
26. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A., and Zillig, W. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 4586–4593
27. Omer, A. D., Lowe, T. M., Russell, A. G., Eddy, S. R., and Dennis, P. P. (2000) Science 288, 517–522
28. Törö, I., Thore, S., Mayer, C., Basquin, J., Säfström, B., and Suck, D. (2001) EMBO J. 20, 2293–2303
29. Khusial, P., Plaag, R., and Zieve, G. W. (2005) Trends Biochem. Sci. 30, 522–528
30. Mura, C., Phillips, M., Kozhukhovskaya, A., and Eisenberg, D. (2003) Proc. Natl. Acad. Sci. U.S.A. 100, 4539–4544
31. Achsel, T., and Marchfelder, A. (2003) Nat. Struct. Mol. Biol. 10, 5532–5537
32. Sauter, C., Basquin, J., and Suck, D. (2003) Nucleic Acids Res. 31, 6061–6068
33. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilber, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbokowski, J., Gibson, R., Juwen, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shim, G., Goyal, A., Pirtetrokovski, S., Church, G. M., Daniels, C. J., Miao, J. L., Rice, P., Nölling, J., and Reeves, J. N. (1997) J. Bacteriol. 179, 7135–7155
34. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A., and Zillig, W. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 4586–4593
35. Omer, A. D., Lowe, T. M., Russell, A. G., Eddy, S. R., and Dennis, P. P. (2000) Science 288, 517–522
36. Törö, I., Thore, S., Mayer, C., Basquin, J., Säfström, B., and Suck, D. (2001) EMBO J. 20, 2293–2303
37. Khusial, P., Plaag, R., and Zieve, G. W. (2005) Trends Biochem. Sci. 30, 522–528
38. Mura, C., Phillips, M., Kozhukhovskaya, A., and Eisenberg, D. (2003) Proc. Natl. Acad. Sci. U.S.A. 100, 4539–4544
39. Achsel, T., and Marchfelder, A. (2003) Nat. Struct. Mol. Biol. 10, 5532–5537
55. Zaigler, A., Schuster, S. C., and Soppa, J. (2003) Mol. Microbiol. 48, 1089–1105
56. Altschul, S. F., Madden, T. L., Scha¨ffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
57. Brenneis, M., Hering, O., Lange, C., and Soppa, J. (2007) PloS Genet 3, e229
58. Azam, T. A., and Ishihama, A. (1999) J. Biol. Chem. 274, 33105–33113
59. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührrmann, R. (1999) EMBO J. 18, 5789–5802
60. Kufel, J., Allmang, C., Verdone, L., Beggs, J. D., and Tollervey, D. (2002) Mol. Cell Biol. 22, 5248–5256
61. Lee, T., and Feig, A. L. (2008) RNA 14, 514–523
62. Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., and Gottesman, S. (2003) Mol. Microbiol. 50, 1111–1124
63. Allers, T., and Mevarech, M. (2005) Nat. Rev. Genet. 6, 58–73
64. Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Dav, M., Parkin-son, J., Greenblatt, J., and Emili, A. (2005) Nature 433, 531–537
65. Omer, A. D., Ziesche, S., Ebhardt, H., and Dennis, P. P. (2002) Proc. Natl. Acad. Sci. U.S.A. 99, 5289–5294
66. Muffler, A., Traulsen, D. D., Fischer, D., Lange, R., and Hengge-Aronis, R. (1997) J. Bacteriol. 179, 297–300
67. Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007) Mol. Microbiol. 63, 193–217
68. Christiansen, J. K., Larsen, M. H., Ingmer, H., Søgaard-Andersen, L., and Kallipolitis, B. H. (2004) J. Bacteriol. 186, 3355–3362
69. Ding, Y., Davis, B. M., and Waldor, M. K. (2004) Mol. Microbiol. 53, 345–354
70. Lucchetti-Miganeh, C., Burrowes, E., Baysse, C., and Ermel, G. (2008) Microbiology 154, 16–29
71. Robertson, G. T., and Roop, R. M., Jr. (1999) Mol. Microbiol. 34, 690–700
72. Sharma, A. K., and Payne, S. M. (2006) Mol. Microbiol. 62, 469–479
73. Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jäger, K. E., and Bläsi, U. (2003) Microb. Pathog. 35, 217–228
74. Bohn, C., Rigoulay, C., and Boulou, P. (2007) BMC Microbiol. 7, 10
75. Dyall-Smith, M. L. (2001) The Halohandbook: Protocols for Haloar-chaebal Genetics, http://www.haloarchaea.com/resources/halohandbook/ Halohandbook_2008_v7.pdf