ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 13, pp. 1434-1449. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © V. N. Murina, A. D. Nikulin, 2011, published in Uspekhi Biologicheskoi Khimii, 2011, Vol. 51, pp. 133-164.

REVIEW

RNA-Binding Sm-Like Proteins of Bacteria and Archaea. Similarity and Difference in Structure and Function

V. N. Murina and A. D. Nikulin*

Institute of Protein Research, Russian Academy of Sciences, Institutskaya ul. 4, 142292 Pushchino, Moscow Region, Russia; E-mail: nikulin@vega.protres.ru

Received April 8, 2011
Revision received May 6, 2011

Abstract—RNA-binding proteins play a significant role in many processes of RNA metabolism, such as splicing and processing, regulation of DNA transcription and RNA translation, etc. Among the great number of RNA-binding proteins, so-called RNA-chaperones occupy an individual niche; they were named for their ability to assist RNA molecules to gain their accurate native spatial structure. When binding with RNAs, they possess the capability of altering (melting) their secondary structure, thus providing a possibility for formation of necessary intramolecular contacts between individual RNA sites for proper folding. These proteins also have an additional helper function in RNA–RNA and RNA–protein interactions. Members of such class of the RNA-binding protein family are Sm and Sm-like proteins (Sm-Like, LSm). The presence of these proteins in bacteria, archaea, and eukaryotes emphasizes their biological significance. These proteins are now attractive for researchers because of their implication in many processes associated with RNAs in bacterial and archaeal cells. This review is focused on a comparison of architecture of bacterial and archaeal LSm proteins and their interaction with different RNA molecules.

DOI: 10.1134/S0006297911130050

Key words: Hfq, Sm-like protein, quaternary structure, RNA–protein interaction

RNA-binding proteins play an important role in many metabolic processes of RNA, such as RNA splicing and processing, regulation of DNA transcription and RNA translation, etc. Among the many RNA-binding proteins, so-called RNA-chaperones occupy an individual niche; they were so-named for their ability of helping RNA molecules to gain their proper native spatial structure [1]. These proteins not only bind RNA, but are also capable of altering (melting) the secondary structure of RNA molecules, thus facilitating appearance of necessary intramolecular contacts between certain RNA sites for proper folding. Such proteins also have an additional helper function in various RNA–RNA or RNA–protein interactions. The members of this class of RNA-binding proteins are Sm and Sm-like (Sm-Like, LSm) proteins [2]. The presence of these proteins in both bacteria and archaea emphasizes their biological importance [3-6].

The history of the study of LSm proteins begins from 1959, when the cause of a complex autoimmune disease – systemic lupus erythematosus – was found in 22-year-old Stephanie Smith. The patient's body produced antibodies against antigens representing ribonucleoprotein (RNP) called "smith antigen" or Sm [7]. Later, it was found that Sm-antigen comprises six small nuclear RNAs (snRNAs) (U1a, U1b, U2, U4, U5, and U6 [8]) and up to 40 different proteins [9]. These RNPs were also called small nuclear RNPs (snRNPs). The snRNPs were divided into two groups. The first includes the proteins specific for each individual snRNP, whereas the second includes other proteins resembling their primary structure. They were called Sm proteins [10]. Sm proteins are present in each snRNP particle and are necessary for snRNP biogenesis. In cells snRNPs fulfill different functions: to all appearances they are important modulators of RNA functions [11] and essential components of eukaryotic spliceosomes [12, 13]. Later, conserved Sm-like proteins were found in archaea, and a homology was revealed between eukaryotic Sm proteins and bacterial Hfq proteins. These proteins now attract growing interest for their implication in many processes occurring in bacterial and

Abbreviations: H-NS, histone-like heat-stable nucleoid-structuring protein; K_d, apparent dissociation constant; LSm, Sm-like protein; snRNA, small nuclear RNA; sRNA, small regulatory RNA.

* To whom correspondence should be addressed.

archaeal cells. This review is focused on structural aspects of bacterial and archaeal Sm-like proteins and on their interaction with various RNAs.

STRUCTURE AND FUNCTION OF BACTERIAL Sm-LIKE Hfq PROTEINS

Primary structure of Hfq proteins. The Hfq protein was discovered in 1968 as *E. coli* cell protein HF-I (Host factor I) involved in the replication cycle of bacteriophage Qβ *in vivo* [14, 15]. Somewhat later HF-I was purified using heating of the cell lysate to 85°C, and the active form of HF-I was found to be a homohexamer composed of subunits with molecular mass of about 12.5 kDa, and the stoichiometry of the protein/RNA interaction was 1 : 1 [16]. In 1991 the *E. coli* gene *hfq* encoding HF-I was identified and cloned and its sequence was determined, as well as the sequence of its 102-amino acid residue product [17]. Further experiments *in vitro* showed that Hfq melts the 3′-end of the "+" chain of phage RNA and provides interaction of Qβ-replicase with this RNA region to initiate synthesis of the "–" chain [18]. In uninfected bacterial cells Hfq is a multifunctional regulator of translation of a great number of mRNAs, including those of DNA repair proteins [19].

A comparison of Hfq amino acid sequences (Fig. 1) has demonstrated that they have a conserved central region, about 80 amino acid residues in length, and a C-terminal region varying in size and composition [20]. The Hfq proteins contain a conserved Sm1 motif characteristic of eukaryotic Sm proteins [10], so bacterial Hfq proteins and eukaryotic Sm/LSm proteins share homology [20, 21]. This suggests the hypothesis of a common ancestor of bacterial Hfq and eukaryotic Sm/LSm proteins [22].

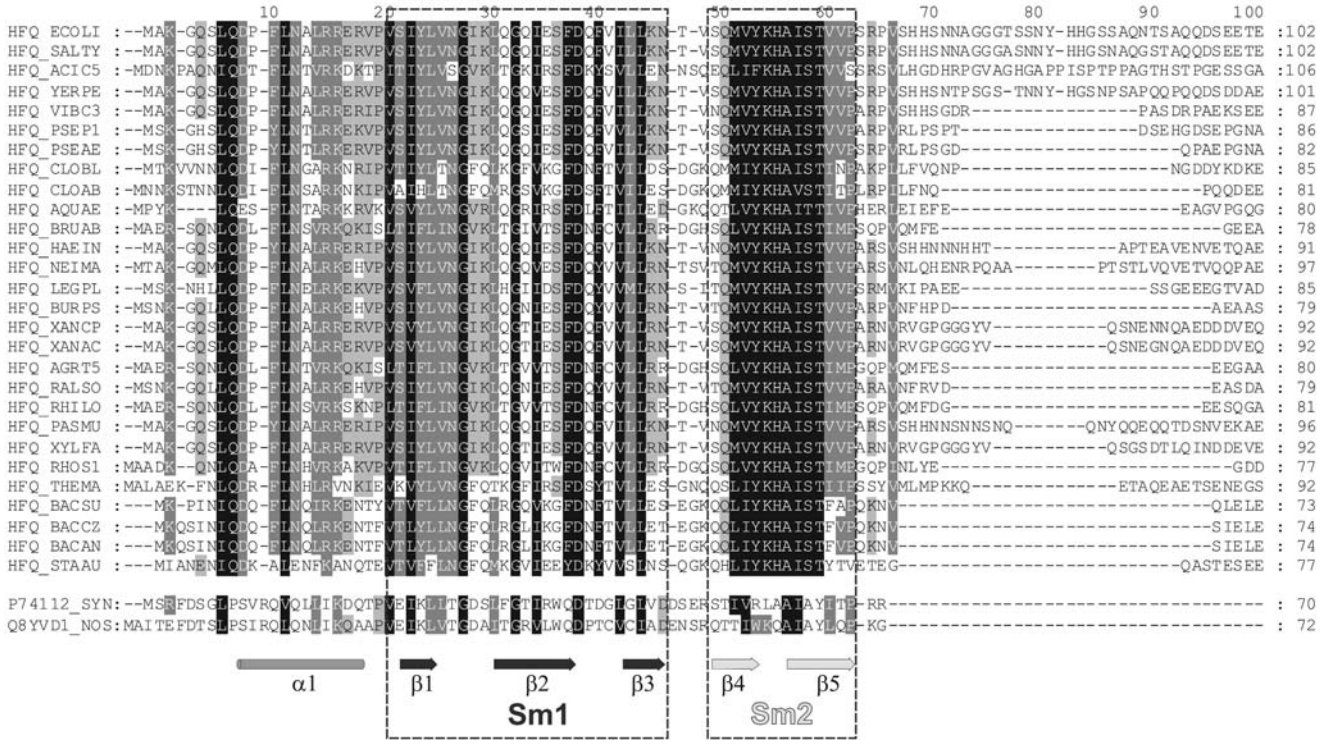

Fig. 1. Amino acid sequences of Hfq proteins from *Escherichia coli* (HFQ_ECOLI), *Salmonella typhimurium* (HFQ_SALTY), *Acidobacterium capsulatum* (HFQ_ACIC5), *Yersinia pestis* (HFQ_YERPE), *Vibrio cholerae* serotype O1 (HFQ_VIBC3), *Pseudomonas putida* (HFQ_PSEP1), *Pseudomonas aeruginosa* (HFQ_PSEAE), *Clostridium botulinum* (HFQ_CLOBL), *Clostridium acetobutylicum* (HFQ_CLOAB), *Aquifex aeolicus* (HFQ_AQUAE), *Brucella abortus* (HFQ_BRUAB), *Haemophilus influenzae* (HFQ_HAEIN), *Neisseria meningitidis* (HFQ_NEIMA), *Legionella pneumophila* (HFQ_LEGPL), *Burkholderia pseudomallei* (HFQ_BURPS), *Xanthomonas campestris* pv. *campestris* (HFQ_XANCP), *Xanthomonas axonopodis* (HFQ_XANAC), *Agrobacterium tumefaciens* (HFQ_AGRT5), *Ralstonia solanacearum* (HFQ_RALSO), *Rhizobium loti* (HFQ_RHILO), *Pasteurella multocida* (HFQ_PASMU), *Xylella fastidiosa* (HFQ_XYLFA), *Rhodobacter sphaeroides* (HFQ_RHOS1), *Thermotoga maritima* (HFQ_THEMA), *Bacillus subtilis* (HFQ_BACSU), *Bacillus cereus* (HFQ_BACCZ), *Bacillus anthracis* (HFQ_BACAN), *Staphylococcus aureus* (HFQ_STAAU), *Nostoc* sp. (Q8YVD1_NOS), and *Synechocystis* sp. (P74112_SYN). The numbering on the top corresponds to the *E. coli* protein. Identical amino acids or conserved substitutions among all sequences except for proteins from cyanobacteria *Nostoc* sp. and *Synechocystis* sp. are drawn in black, areas sharing no less than 80% identity are in dark gray, and no less than 60% identity are in light gray. Elements of secondary structure and conserved motifs Sm1 and Sm2 are denoted on the bottom. The sequences were compared using the T-Coffee software (http://www.tcoffee.org/) and arranged using the Genedoc package (http://www.nrbsc.org/gfx/genedoc/).

The most conserved part of Hfq proteins is an Sm2 motif occupying the end of β4 and beginning of β5 strands with a short loop between them (Fig. 1). Sm2 comprises a [Y/F]KHAI consensus, which is present in virtually all bacterial Hfq proteins. Amino acid residues of the Sm2 motif are implicated in RNA–protein interaction [23] and, most probably, determine the quaternary structure of the protein and its high thermostability [24-26].

The search for Hfq homologs among known in 2002 bacterial genomes was successful for about half of the species [22]. Since homologs were not found in some proteobacteria and in Gram-positive bacteria, it was supposed that the gene of this protein was lost in some phylogenetic branches [22]. However, using advanced homolog search algorithms and amino acid patterns, Hfq orthologs were later found in additional series of genomes in which they were not previously identified [6]. This result suggests Hfq proteins are much more widely distributed and can significantly vary in their primary structure.

Demonstrative examples of distant Hfq orthologs are proteins from cyanobacteria *Synechocystis* sp. PCC 6803 and *Anabaena* PCC 7120, which were placed in a new group of Hfq proteins [27]. The members of the group have 65% similarity in primary structure and 30% identity of amino acid residues. Their divergence is confirmed by comparison with other Hfq proteins (Fig. 1). In particular, a conserved part of the Sm2 motif in cyanobacterial proteins differs from the [Y/F]KHAI consensus, being substituted by RLAAI in the protein from *Synechocystis* or by WKQAI in the protein from *Anabaena* [27]. This difference in primary structure results in less affinity to RNA of Hfq proteins from *Synechocystis* sp. PCC 6803 and *Anabaena* PCC 7120 compared to the *E. coli* Hfq. They cannot regulate gene expression in *E. coli* cells *in vivo* [27], but *hfq* gene knockout in *Synechocystis* leads to the loss of cell motility and considerably decreases the level of synthesis of a series of mRNAs [28], thus suggesting that these proteins are also global regulators of transcription.

Hfq spatial structure. The first data on Hfq structure were obtained using electron microscopy [20]. EM images show that the protein forms symmetric hexamers 70-80 Å in diameter. Determination of spatial structures of *Staphylococcus aureus* Hfq [23], core part (residues 1-72) of *E. coli* protein [24], and the protein from *Pseudomonas aeruginosa* [26] confirmed a hexameric quaternary structure of Hfq proteins (Fig. 2). They have a doughnut shape with external diameter of about 70 Å, and the central hole about 8-10 Å.

The Hfq quaternary structure is organized due to the intersubunit continuous β-sheet formed by adjacent β-strands of neighboring protein monomers. The hydrogen bonds formed by backbone atoms of adjacent subunits determine orientation of monomers in the oligomer relative to each other. Antiparallel β-strands are divided into two conserved motifs, Sm1 (strands β1, β2, and β3) and

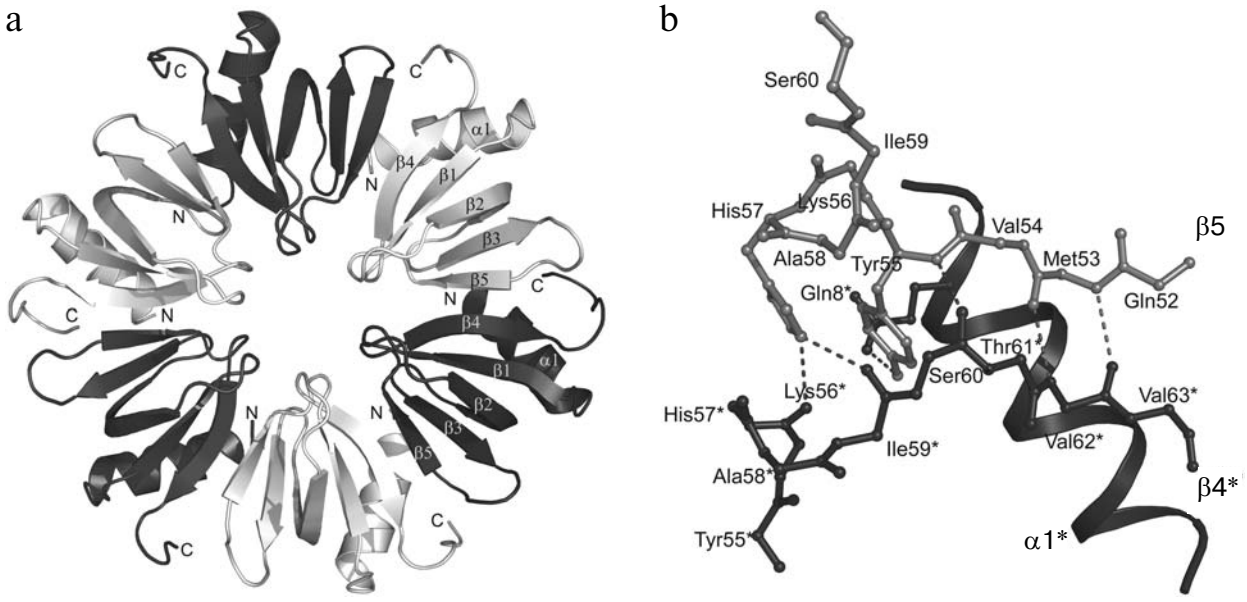

Fig. 2. Spatial structure of *P. aeruginosa* Hfq (PDB ID 1U1S). a) Quaternary structure of the protein is hexamer. Neighboring monomers are drawn in light and dark gray. Elements of secondary structure are shown on two monomers. b) Interface of two adjacent monomers. The atoms of the main chains of β4 and β5 strands, α-helix of one monomer, and side chains of residues forming intermolecular hydrogen bonds are shown. Hydrogen bonds are drawn as dotted lines. The amino acid numbers with and without asterisks belong to adjacent monomers. The spatial structures are drawn using the PyMOL software (http://www.pymol.org/) on the basis of coordinates from the RCSB database (http://www.rcsb.org/).

Sm2 (strands β4 and β5) [29]. All of the strands form a structured Sm fold, which is a characteristic feature of all members of the LSm protein family. The structure of the Sm fold is very similar to the OB (Oligonucleotide Binding) fold [30] and has SH3-like topology [31].

The N-terminal end of the Hfq molecule contains an α-helix (residues 7-18 for the *E. coli* protein) that is present in all Sm/LSm proteins but not included in the Sm fold. In the hexamer, the α-helices are located at one side of the toroid and shield the intersubunit interface from the solvent [32]. This side of the hexamer is called proximal, and opposite side is called distal.

The N-terminal α-helix with β-strands of the Sm fold comprise the Hfq core (residues 7-66), which is the most conserved part among all known Hfq proteins [24]. The following C-terminal part of the protein has the length of seven (in *Bacillus subtilis*) to 36 (in *E. coli*) residues. They have no explicit homology but most of the residues are positively charged (Fig. 1). It was reported that *E. coli* Hfq shortened to 66-72 residues from the C-terminal end retains the ability to form hexamers and bind RNAs [33, 34], but it is less compact and stable compared to the wild-type protein [35]. There are many examples of the natively shortened Hfq proteins, which nevertheless possess their normal functions. Among them are Hfq proteins from *Synechocystis* sp. PCC 6803 and *Anabaena* PCC 7120 [27], *Aquifex aeolicus* [36], *Listeria monocytogenes* [37], and a single example of the archaeal Hfq from *Methanococcus jannaschii* [38].

There are contradictory data on the functionality of *E. coli* Hfq lacking the C-terminal region. On one hand, the *E. coli* Hfq shortened to 65 residues from the C-terminal end was reported to bind noncoding RNAs and alter their secondary structures, but cannot bind mRNAs and regulate expression of a series of genes *in vivo*, including autoregulation of *hfq* gene expression [39]. On the other hand, data about negative and positive control over expression of some genes *in vivo* by *E. coli* Hfq shortened from the C-end to 65, 66, 69, and 72 residues exist [40]. Nonetheless, the two research groups agree that the shortened Hfq variants can bind with a significant number of different RNAs, and the discordance is in sustaining ability for regulation of expression of some genes.

Hfq protein as a transcription regulator. As mentioned above, Hfq is a pleiotropic multifunctional post-transcriptional regulator that can play a role of positive and negative regulator of translation of many bacterial mRNAs [32, 41]. When binding mRNA, Hfq alters its secondary structure, thus altering mRNA stability or its translation level [42-44]. This can lead to increase or decrease in expression of more than 40 genes during the stationary phase of *E. coli* cell growth [41, 45].

Hfq is also implicated in the interaction of small regulatory RNAs (sRNAs) with mRNAs [6, 41, 46], providing the melting of RNA secondary structure and facilitating the interaction. The most well studied examples are regulation of translation of *galK* (galactokinase) mRNA by Spot42 sRNA [47] and *fhlA* (DNA-binding transcription activator FhlA) mRNA by OxyS sRNA [48, 49], repression of translation of *hns* (nucleoid-associated protein H-NS) mRNA by regulatory DsrA sRNA [44], repression of translation of *rpoS* (bacterial RNA-polymerase subunit σ^s) mRNA by OxyS sRNA [50, 51], and stimulation of *rpoS* mRNA translation by DsrA sRNA [42, 52-54].

The *rpoS* gene encodes bacterial RNA-polymerase subunit σ^s [55, 56], which, in turn, regulates expression of many genes during cell growth [57], regulates expression of proteins required under stress conditions [58-62], as well as synthesis of virulence factors of some bacterial pathogens such as *P. aeruginosa*, *Salmonella typhimurium*, *Yersinia enterocolitica*, etc. [5, 63-67]. Expression of *rpoS* is controlled on the levels of transcription, translation, and protein stability [68]. Known regulators of *rpoS* translation are H-NS [69] and Hfq [70], as well as two regulatory sRNAs: OxyS [50] and DsrA [53]. The general scheme of regulation of σ^s translation shows dependence of this process on a series of external and internal factors.

Hfq is a key protein ensuring effective translation of the σ^s subunit. At the stationary growth phase *Escherichia coli* cells with inactivated *hfq* gene have lowered growth rate, shortening, and increase in osmo- and UV-sensitivity compared to the normal cells [33]. This phenotype largely resembles that of cells with inactivated *rpoS* gene, which cannot effectively adapt to stringent conditions of the stationary growth phase because of the absence of the transcription factor σ^s. The important role of Hfq is emphasized by a fact that inactivation of its gene is epistatic relative to inactivation of genes encoding other factors implicated in regulation of σ^s expression. Thus, suppression of *hfq* activity leads to the loss of the mechanism regulating σ^s expression using H-NS protein, OxyS RNA, DsrA RNA, and some external factors [42, 57].

Like in the case of Q_β RNA, the role of Hfq in stimulation of *rpoS* mRNA translation initiation is through altering of the RNA structure. Hfq can bind with the translation initiation area of *rpoS* mRNA to alter its secondary structure, thus enabling recognition of this area by the ribosomal 30S subunit [55].

Functional activity and transcription level of Hfq are controlled by H-NS (Histone-like Nucleoid Structuring protein). It is a thermostable DNA-binding protein regulating expression of a great number of *E. coli* genes. H-NS serves as an *hfq* transcription repressor [55]; moreover, it can directly bind with Hfq to inactivate this protein [17]. Besides, *hfq* expression undergoes autoregulation at the translational level. The 5′-UTR of *hfq* mRNA has two hairpins with which the Hfq interaction results in inhibition of translation initiation complex formation. Experiments, both *in vitro* and *in vivo*, have demonstrated that only simultaneous protein binding with the two hairpins results in effective repression of the *hfq* mRNA translation [71].

The Hfq and H-NS proteins interact with posttranscriptional regulator DsrA. It is an RNA, 87 nucleotides in length, implicated in regulation of expression of genes *hns* and *rpoS* encoding H-NS and RNA polymerase subunit σ^s, respectively [52]. DsrA blocks initiation of H-NS translation by binding with an mRNA *hns* region localized direct after the ATG start codon. This leads to suppression of H-NS synthesis, elevation of *hfq* expression, and, as a result, activation of *rpoS* mRNA translation. On the other hand, DsrA can directly upregulate *rpoS* expression. At low temperature, translation of the *rpoS* mRNA is blocked because 5′-UTR and RBS (Ribosome Binding Site) of mRNA form stable secondary structure. DsrA binds with this mRNA site to release the RBS for translation initiation [45]. There are data that only in complex with Hfq the DsrA possesses the melting activity towards the secondary structure of *rpoS* mRNA RBS [44].

The number of Hfq molecules required for initiation of *rpoS* mRNA translation can depend on regulatory OxyS sRNA, 102 nucleotides in length, which specifically interacts with free Hfq protein, thus disabling its association with other RNAs. Being associated with OxyS, Hfq cannot serve as a positive effector of σ^s translation initiation. Thus, OxyS exerts control over expression of all genes controlled by RNA polymerase subunit σ^s [50].

Hfq regulates stability of some mRNAs. Altering stability of mRNAs is a way to control gene expression, allowing quick adaptation of the cell to altering environmental conditions. Stability of individual transcript or pool of mRNAs can depend on the effect of antibiotics, nutrition, transition from one growth phase to another, and some factors influencing the growth rate. Site-specific endoribonucleases, exoribonucleases, and poly(A)-polymerase are involved in *E. coli* mRNA processing and decay [72].

Polyadenylation, synthesis of poly(A) tails at the 3′ end of mRNAs by PAP I poly(A)-polymerase to a distinct length, is one of mechanisms providing controlled degradation of various transcripts. Decrease in stability of polyadenylated mRNA is due to increase in its accessibility to ribonuclease E, which binds to the poly(A) tail at the 3′ end of mRNA, thus favoring its interaction with inner specific recognition sites of mRNA [73].

Hfq protein can participate in these processes at two stages. First, it binds effectively to A-enriched mRNA sites and elevates activity of PAP I enzyme, which leads to 3′-poly(A) elongation and mRNA destabilization *in vivo*. The mechanism of Hfq action is not well understood, but Hfq is found to bind initially to small (10-15 nucleotides in lengths) poly(A) tail of mRNA and then increase its accessibility to PAP I. Hfq therewith increases processivity of PAP I, enabling formation of poly(A). Second, Hfq can protect the recognition sites for RNases, such as polynucleotide phosphorylase, RNase E, and RNase II, from binding with these enzymes. Usually these are 3′-poly(A) sites and 5′-AU-enriched UTRs. The *rpsO* mRNA encoding S15 ribosomal protein represents a striking example for the protective effect of Hfq from RNases [74].

Regulation of *ftsZ* mRNA translation, with the translation product FtsZ protein directly participating in cell division, represents yet another example for Hfq-dependent regulation of mRNA stability. Excess FtsZ results in production of miniature cells, an effect used for revealing of the role of Hfq in the regulation of *FtsZ* translation [75]. In mutants with deleted *hfq* gene the level of FtsZ substantially elevates in the stationary growth phase, resulting in an increase of miniature cell percentage. It is very likely that the mechanism of Hfq action is an increase in accessibility of RNase E recognition sites within the *ftsZ* mRNA due to rearrangement of the RNA secondary structure.

Hfq is implicated in cell protection from stress caused by elevation of hydroxyl radical concentration during the stationary growth phase. The stress may be caused by enhanced expression of FepA and FhuE proteins responsible for Fe^{2+} assimilation. In log phase the cells actively accumulate iron ions required for proliferation. In stationary growth phase the cell does not need so much iron, and its assimilation is blocked by elevated Hfq concentration. Thus the stress induced by excess of Fe^{2+} resulting in intensive production of hydroxyl radicals is prevented. Hfq protein can influence *fepA* and *fhuE* gene expression, and this mechanism is likely associated with regulation of mRNA stability [76].

Hfq also participates in regulation of mRNA stability by yet another mechanism: facilitating interaction between regulatory RNAs and mRNAs. For example, it is involved in degradation of *ompA* mRNA responsible for synthesis of OmpA, the major protein of the outer cell membrane. Stability of this transcript depends on bacterial growth rate and is determined by the interaction with ribonuclease E, which splits its 5′-UTR.

The 5′-UTR of *ompA* mRNA consists of two hairpin structures adjacent to AU-enriched regions and providing high resistance to ribonuclease. Initially, Hfq was thought to cause a local change of the *ompA* mRNA secondary structure resulting in opening of endoribonuclease E recognition sites and initiation of mRNA degradation. In fact, *ompA* mRNA stability correlates with the Hfq content in the cell – the more abundant the protein the less stable is the mRNA. This happens when rates of growth processes are slowing, particularly when entering the stationary growth phase [72]. More detailed studies *in vitro* have demonstrated that the degradation mechanism in more complex, and RNA hairpin-like secondary structures themselves are not protective against ribonuclease E. The 30S ribosomal subunits bound to the 5′-UTR and covering the RNase recognition sites seem to play the key role in mRNA stabilization. Hfq was supposed to be a competitor for 30S ribosomal subunits displacing them from the UTR, thus opening recognition sites for RNase E [77].

However, more recent studies have demonstrated that even this scheme is not completely correct. A small regulatory RNA *MicA* was discovered that regulates degradation of *ompA* mRNA. A Δ*micA E. coli* strain was constructed in which *micA* gene was deleted but the wild type *hfq* gene was presented [78]. This strain was found to lose its ability to control *ompA* mRNA expression. On the basis of this data a new model was proposed suggesting regulation of *ompA* stability through the competitive binding of *MicA* sRNA with initiating ribosomes. In the exponential growth phase, when *MicA* sRNA content in the cell is low, ribosomes can successfully bind to the *ompA* mRNA initiation site, the synthesis of OmpA normally occurs, and mRNA is defended from RNase E by 30S ribosome subunits. In the stationary growth phase, *MicA* sRNA accumulates in the cell and binds with the 5′-UTR of *ompA* mRNA. It is likely that at this stage Hfq stabilizes *MicA* sRNA or facilitates its binding with the 5′-UTR of *ompA* mRNA. These events result in removal of ribosomal subunits from *ompA* mRNA and arrest of OmpA protein synthesis. Since ribosomes are removed from mRNA, its 5′-UTR becomes accessible for RNase E, and mRNA degradation occurs [78]. In conclusion, it should be mentioned that the lifetime of *hfq* mRNA, like *ompA* mRNA, is also determined by the presence of the Hfq protein [71].

Binding sites for mRNAs and regulatory RNAs. The first structural data on the interaction between Hfq and RNA was the crystal structure of a complex between *Staphylococcus aureus* Hfq and AU_5G ribooligonucleotide reported in 2002 [23] (Fig. 3). Nucleotides are localized in the area of the proximal site of the Hfq hexamer, each interacting with amino acids of two subunits, except for the 3′-guanosine that exits the central pore of the hexamer and does not form specific contacts with the protein.

The 5′-adenosine and all uridines bind in a similar way. Nucleotide bases stack with side chains of Lys41 and Tyr42 of two adjacent monomers. Nitrogen atom of the Lys57 side chain is bound to the O2 uracil atom by hydrogen bond, and the His58 side chain interacts with phosphate group and with the O2′ atom of the ribose ring. Surprisingly, Gln8, which does not belong to the Sm fold but is located in additional N-terminal α-helix, forms hydrogen bonds with the uracils. The authors pointed out the existence of direct contacts of Lys41 with uridine bases, but this phenomenon is rather apparent and observed in only one of six nucleotides.

Structural analysis of the complex demonstrated that the positions of Gln8, Lys41, and Lys57 side chains hamper the binding of guanidine and cytidine bases but enable uracils and allow adenines. On the other hand, Lys57 and His58, which contact ribose, determine preference for RNA over DNA binding. Thus, the structure of Hfq–AU_5G complex gives the key to understanding the principles of interaction between Hfq protein and short AU-rich single-stranded RNA fragments. However, the question of how Hfq protein binds extended (more than 100 nucleotides) mRNAs is still opened [32].

Binding area for poly(A) RNA. Hfq binding to poly(A) RNA tails and dependence of the binding con-

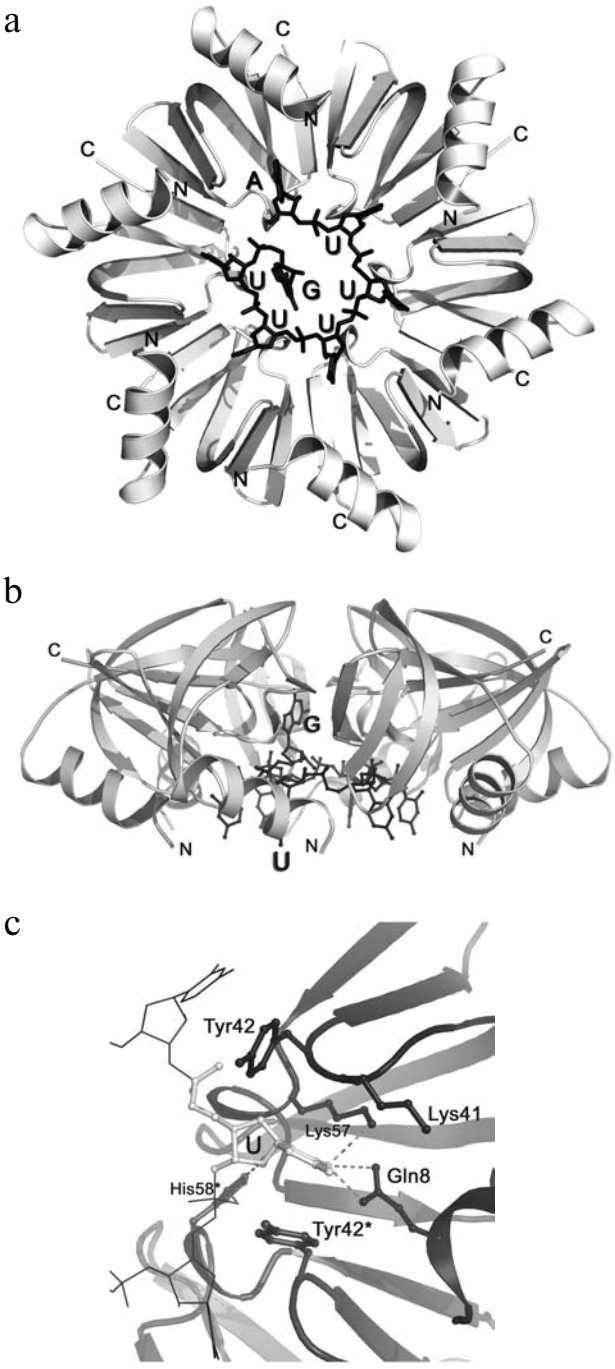

Fig. 3. a) The crystal structure of the complex between *S. aureus* Hfq and AU_5G ribooligonucleotide (drawn in black) (PDB ID 1KQ2). b) Side view of the toroid. c) Uracil-binding pocket with amino acids interacting with the nucleotide. Hydrogen bonds are shown as dotted lines. Asterisks denote amino acids of another monomer.

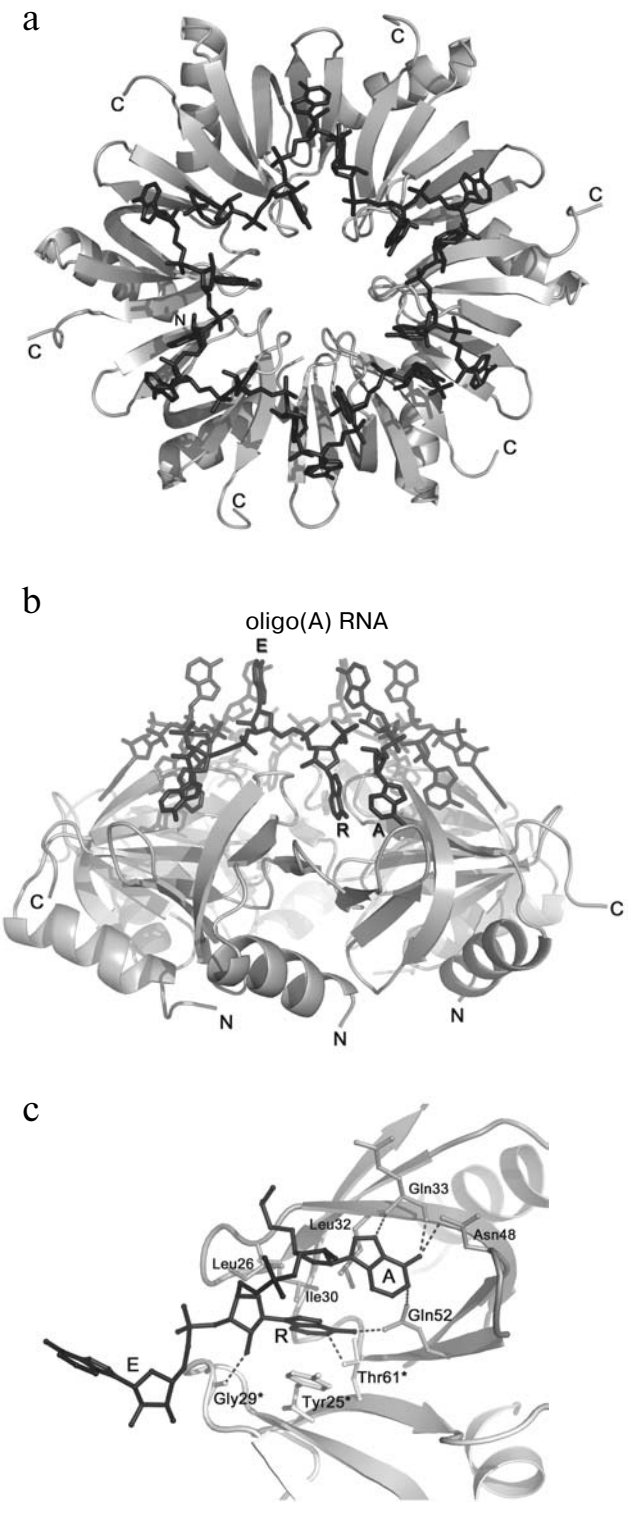

Fig. 4. Spatial structure of *E. coli* Hfq complex with poly(A)$_9$ ribooligonucleotide (drawn in black) (PDB ID 3GIB). a) Front view from the poly(A) RNA-binding area. b) Side view of the toroid. The three ribonucleotide binding sites (A, R, and E) are marked. c) The detailed view of the ribonucleotide triplet binding area. The ribbon model of the protein is drawn; the amino acid residues contacting the ribonucleotides in A, R, and E sites are shown. Hydrogen bonds are shown as dotted lines. Asterisks denote amino acids of another monomer.

stant on the length of the tail with maximum for fragment lengths of 15 to 27 nucleotides have been known since 1980 [79]. *Escherichia coli* Hfq protein residues Tyr25, Ile30, and Lys31 participate in binding of RNA poly(A) tails. Their substitution for alanine more than 100-fold decreases the affinity of Hfq to poly(A) RNA [80, 81]. These residues are localized on a distal surface of the hexamer, i.e. opposite to the site of mRNA and sRNA binding. This fact was confirmed by measurements of RNA binding constants with mutant forms of the protein. Substitutions such as Phe42Ala, Tyr55Ala, Lys56Ala, and His57Ala in the proximal RNA-binding site resulted in just a slight decrease in affinity of poly(A) RNA to the protein but decreased DsrA and poly(U) RNA binding greatly. Besides, A_{27} RNA does not compete with DsrA for Hfq binding, thus confirming the presence of two independent binding sites for poly(A) RNA and regulatory RNAs [80].

Comparison of these results with the data of stereochemical analysis and charge distribution on the *E. coli* Hfq surface has enabled establishing a hypothetical model describing the interaction between poly(A) RNA and Hfq protein [32]. According to this model, the poly(A) RNA-binding area is circular and centrally symmetric. Data on effect of amino acid substitutions allowed to identify a hydrophobic pocket composed of Tyr25, Ile30, and Leu32 and complemented by the Lys31 residue of an adjacent monomer as the recognition site for adenines.

At the end of 2009 the structure of *E. coli* Hfq complex with 15-nucleotyde poly(A) RNA was determined [82] (Fig. 4). The asymmetric crystallographic unit contains three Hfq subunits and nine adenine residues contacting the protein. The binding area and general position of poly(A) RNA are coincident with those predicted by the model proposed earlier. Each protein monomer contacts with an adenine triplet, and each triplet binds to the protein in a similar way at three sites named A, R, and E.

The authors proposed that specificity of adenine recognition by Hfq is primarily determined by the A-site localized in a groove between β2 and β4 strands. Hydrogen bonds formed by N and O atoms of the Gln33 main chain with N6 and N7 atoms of adenosine guarantee its specific recognition. The nucleotide position is additionally stabilized by Gln52, which forms a hydrogen bond with the adenosine N1 atom, and, possibly, by Asn48, which can interact with the N6 atom.

The R-site is localized in a groove between β-strands of adjacent subunits of the protein. The adenine base is enclosed by side chains of Leu26, Ile30, and Leu32 on one side and tightly fixed by side chain of Tyr25 on the other side. Adenine forms two hydrogen bonds with polar moieties of Thr61 and Gln52 side chains. The position of the Gln52 side chain does not hamper guanidine entry into the R-site, because the side chain of glutamine can easily swing to form the G(O4)–Gln52(NE) contact. It is

interesting to note the presence of the hydrogen bond between the ribose O2′ atom and the oxygen atom of the main protein chain (Gly29 residue) in this site. This likely determines why the protein binds to RNA rather than DNA.

A nucleotide base in the E-site does not interact with protein atoms, but goes out of the protein and becomes open to external contacts. Analysis of the structure has allowed prediction of preferential binding of sequences corresponding to the consensus $(5'\text{-ARN-}3')_n$, where A is adenine, R is adenine or guanine, and N is any nucleotide, at the Hfq distal surface, which is confirmed by binding experiment: Hfq affinity to $(AGG)_9$ compared to $(AAA)_9$ is tenfold less [82].

Later the structure of *Bacillus subtilis* YmaH (Hfq) complex with ribooligonucleotide AGAGAGA was deposited in the RCSB PDB protein data bank (PDB ID 3HSB) (but the paper describing the structure is not yet published). In this complex, the binding of the RNA fragment also occurs at the distal side of the protein, but alternation of adenine and guanine does not match the consensus triplet adenine-purine-N. In this case, the binding unit is the AG pair rather than the AAA triplet (Fig. 5).

All adenines of this complex are bound similarly: they are localized in a hydrophobic pocket formed by Leu25, Phe29, and Leu31 from one side and Phe24 from the opposite side. These residues are analogs of Leu26, Ile30, Leu32, and Tyr25 in *E. coli* Hfq protein. The hydrogen bond between adenine N6 atom and Thr61 remains the same, but the side chain of Gln52 is localized far from the base, not allowing any contact. Instead, additional hydrogen bonds arise between the adenine N1 atom and Ser60 side chain, and, in some binding sites, between the adenine N3 atom and Asn27 side chain. Position of ribose units of adenines is similar to that in the poly(A) RNA–Hfq complex, and a hydrogen bond between the ribose O2′ atom and the oxygen atom of Gly28 (the *E. coli* Hfq Gly29 analog) main chain also exists. Thus, despite some rearrangement of the hydrogen bonding network, the adenine-binding site of *B. subtilis* Hfq is an analog of the *E. coli* protein site R.

Guanines are bound in a site localized on the protein surface rather than in its cavity. One side of the base is completely exposed to the solvent, whereas another stacks with the Gln30 side chain. The only amino acid residue that contacts guanine is Arg32 (the *E. coli* Hfq Gln33 analog), whose terminal amine group forms hydrogen bonds with the O6 and N4 atoms of the base. This site resembles the A-site of *E. coli* Hfq complex with poly(A) RNA, but is organized quite differently. So, when complex is formed between AGAGAGA RNA and *B. subtilis* Hfq protein, only one RNA recognition site on the protein is involved, namely R-site, but the place for nucleotides at A and E sites is occupied by a single nucleotide connecting adenines across the protein surface.

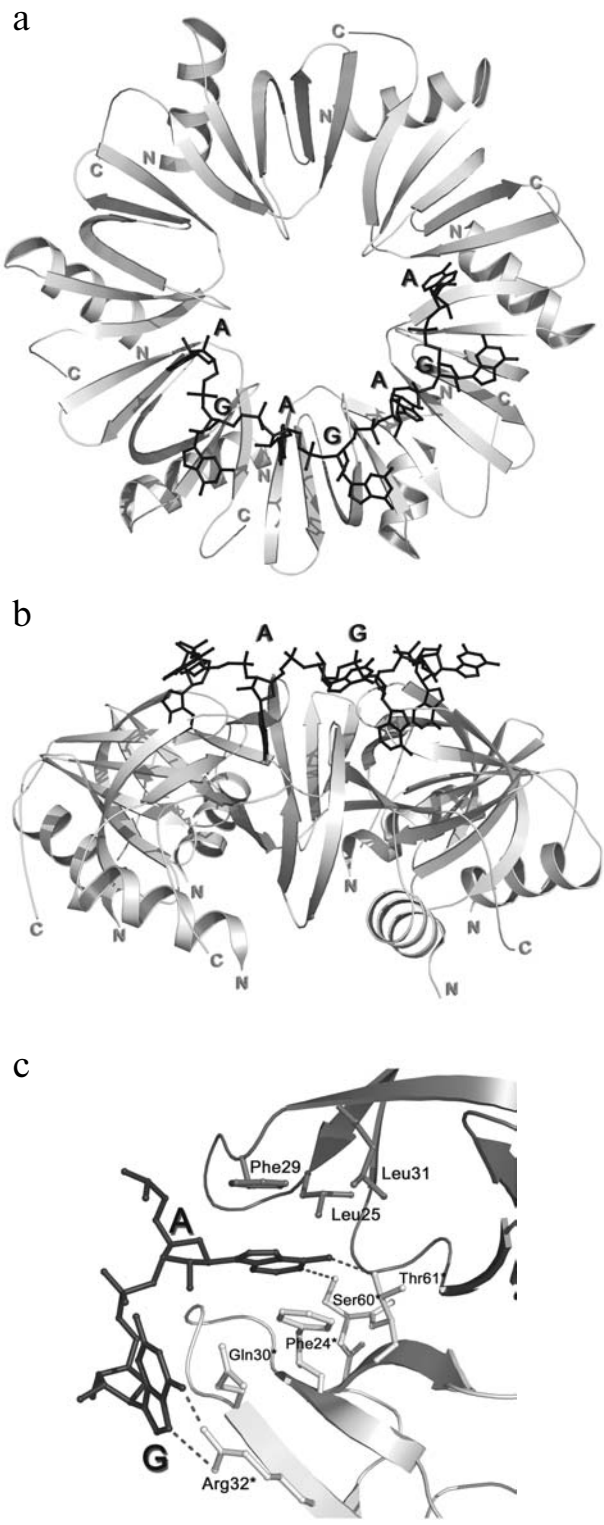

Fig. 5. Spatial structure of *B. subtilis* Hfq (YmaH) complex with AGAGAGA ribooligonucleotide (drawn in black) (PDB ID 3HSB). a) View on the ribooligonucleotide-binding side of the protein. b) Side view of the toroid. Positions of two of seven ribonucleotides, A and G, are denoted. c) The ribonucleotide-binding site on the protein surface. The ribbon model of the protein is shown with amino acid residues contacting nucleotides. Hydrogen bonds are drawn as dotted lines. Asterisks denote amino acid residues of an adjacent monomer.

Thus, amino acid residues forming the RNA-recognizing R-site at the Hfq distal surface prove to be a determining factor for recognition of poly(A) and similar RNAs.

Stoichiometry of RNA–Hfq binding. Since Hfq facilitates interaction between various RNAs, such as mRNAs and regulatory RNAs, a question arises about the binding stoichiometry between RNA and Hfq, i.e. is the protein capable of recognizing several RNA molecules simultaneously? The answers to this question are still contradictory. For instance, isothermal calorimetric titration demonstrated that the RNA/protein ratio is 1 : 1 for DsrA–Hfq_6 (Hfq hexamer) and rpoS–Hfq_6 complexes, but 1 : 2 for A_{18}–Hfq_6 [80]. The ratio 1 : 1 for the rpoS–Hfq_6 complex differs from 1 : 2 ratio found earlier for the same complex by RNA electrophoretic mobility shift assay [83]. Similar measurements of the RNA electrophoretic mobility shift done later demonstrated the ratio 1 : 2 in DsrA–Hfq_6 and A_{18}–Hfq_6 complexes [81]. Fluorescence anisotropy measurements for A_{18}–Hfq_6 complex demonstrated 1 : 2 for the RNA/protein ratio, which apparently suggests the presence of two independent but almost identical interaction acts between the protein and poly(A) RNA [81].

New data on the interaction stoichiometry was obtained in 2011 [84]. The RNA/protein ratio equal to 1 : 1 for A_{18}–Hfq_6 and $DsrA_{DII}$ (DsrA domain II)–Hfq_6 complexes was demonstrated by mass spectrometry, analytical centrifugation, fluorescence anisotropy assay, and electrophoretic mobility shift assay, and the C-terminal truncation of the protein did not alter the interaction stoichiometry. The authors also determined similar ratio (1 : 1) in the Hfq_6 complex with full-length regulatory RNAs (DsrA, RprA, and OxyS) or 18-nt OxyS RNA fragment. Moreover, the peak of Hfq_6–A_{18}–$DsrA_{DII}$ ternary complex was observed in mass spectrometric experiments, although the peak height raises some doubts about stability of such complex in solution.

What is the reason for the discrepancy in the results of study of the stoichiometry of the Hfq complexes with RNA? It is likely that this might result from difference in protein purification protocols that lead to differences in activities of the molecules or even partial contamination of the protein with RNA short fragments or single nucleotides. This, in turn, can drastically affect the interaction between the protein and RNA. The *E. coli* Hfq, which is often copurified with such admixtures, sometimes provides surprising results when studying its interaction with RNA. Thus, discrepancies in binding experiments can be explained by the difference in experimental approaches and conditions, as well as by the difference in purity of samples and homogeneity of RNA. Perhaps the final answer to the question on the interaction stoichiometry between Hfq and various RNA molecules would be structural data on corresponding RNA–protein complexes, which would enable direct determination of the protein binding sites on RNAs and stoichiometry of Hfq–RNA interaction.

Hfq is a component of the protein fraction of *Escherichia coli* chromosomal DNA. *Escherichia coli* genomic DNA is associated with approximately 10 different proteins with DNA-binding properties [85]. Quantitative and qualitative composition of a protein fraction apparently influences the character of folding and general DNA activity, namely, replication, recombination, repair, and transcription.

Hfq protein comprises the nucleoid protein fraction and is capable of binding chromosomal DNA at different stoichiometric ratios depending on the cell growth phase [86]. During the log phase, Hfq is one of the major proteins associated with chromosomal DNA, which may suggest its possible effect on *E. coli* DNA replication and transcription.

Hfq was demonstrated to possess nonspecific affinity to supercoiled and relaxed DNA forms. Plasmid DNAs bind Hfq *in vitro* and *in vivo* in *E. coli* cells [87]. As the rate of cellular growth processes changes, the protein composition of the nucleoid alters, including replacement of Hfq by other DNA-binding proteins in chromosomal DNA in the stationary growth phase. The specific role of Hfq in these processes remains unknown.

In 2010, an attempt to localize the DNA-binding region on the *E. coli* Hfq and determine a deoxyribonucleotide consensus recognized by Hfq was made [88]. Genomic DNA fragments associated with *E. coli* Hfq were found after multiple purification steps. Sequences of 41 amplified segments, of 60 to 567 bp, from the DNA fragments associated with Hfq were determined. These segments included 24 unique nucleotide sequences, and 13 of them were found two or more times (segment names rather than sequences are given in the work). A large number of the DNA segments were predicted to have significant helical axis curvature and were from genes associated with membrane proteins, characteristics unexpected for nonspecific binding. The curvature of DNA fragments was determined according to an algorithm of Vlahovicek et al. [89].

Analysis by analytical ultracentrifugation indicated that A_{18} RNA binding to Hfq disrupts Hfq–DNA interactions [88]. This suggests that the binding area for poly(A) RNA on the Hfq distal surface overlays that for DNA. Since the minimum apparent K_d for the Hfq complexes with the DNAs studied in this work was estimated as ~400 nM, and K_d for the Hfq complex with $A_{16\text{-}27}$ RNA was ~1.5 nM, the displacement of DNA by poly(A) RNA from the binding sites on the distal side of the protein seems realistic.

To confirm this hypothesis, a series of Hfq mutants was constructed with single amino acid replacements for alanine in proximal, distal, and lateral surfaces of the protein hexamer. Both mutations on the Hfq distal surface (Tyr25Ala and Lys31Ala) and one on the outer surface (Arg16Ala) inhibited Hfq binding to duplex DNA, while six of seven substitutions on the proximal (Gln8Ala,

Arg17Ala, Phe39Ala, and Phe42Ala) and lateral (Phe11Ala, and Arg19Ala) surfaces did not prevent the DNA–Hfq complex formation. Amino acid residues Tyr25 and Lys31 form the binding site for poly(A) RNA, so they are necessary for its binding.

Thus, the data confirm position identity of poly(A)- and DNA-binding sites. Interestingly, the truncation of *E. coli* Hfq to 65 and 75 amino acid residues also inhibits DNA binding, thus suggesting essentiality of the long C-end, which is unfolded in the intact protein, for the DNA–Hfq interaction [88].

Sm-LIKE PROTEINS OF ARCHAEA

Structure of archaeal LSm proteins. Archaeal LSm proteins were first identified in 1999 during systematic analysis of eukaryotic and archaeal genomes [90]. Two different Sm-like protein sequences were found in the genomes of both *Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus*, one in the genome of *Pyrococcus horikoshii* OT3, and none in the genome of *Methanococcus jannaschii*. It is worth noting that later an Sm-like protein was identified in *M. jannaschii*, which shares no homology with eukaryotic and archaeal LSm protein, but is homologous to bacterial Hfq proteins [38].

Thus, it was first demonstrated that archaeal genomes encode one or two Sm-like proteins belonging to two subfamilies named Sm1 and Sm2 [90]. Homology between Sm1 protein sequences of various species is no less than 60%, while it does not exceed 30% between the Sm1 and Sm2 proteins of the same organism [91].

Electron microscopic images of archaeal Sm-like proteins show that, unlike bacterial Hfq proteins forming hexameric rings, they form toroids composed of seven monomers [92]. Spatial structures of archaeal Sm1 proteins from *M. thermoautotrophicum* [93], *P. aerophilum* [94], and *A. fulgidus* AF-Sm1 [91] have confirmed the EM data (Fig. 6). All these proteins form homoheptameric toroids with intersubunit contacts resembling those in eukaryotic Sm dimers and bacterial Hfq hexamers, which was immediately interpreted as evolutionary conservatism of the Sm-like proteins.

Archaeal LSm proteins are composed of an N-terminal α-helix and five β-strands forming a continuous Sm-fold (Fig. 7). They have several characteristic differences from bacterial and eukaryotic LSm proteins, despite homology between the archaeal LSm Sm1 motif and that of other members of this family. Archaeal LSm proteins do not have a long C-terminal domain characteristic of both the bacterial and eukaryotic proteins. Like eukaryotic Sm proteins, they contain an RGXX consensus in the Sm2 motif (where X is a charged amino acid residue) instead of [Y/F]KHAI consensus in bacterial proteins. An important feature of archaeal proteins is a long L4 loop connecting the Sm1 and Sm2 motifs. This loop is found in eukaryotic Sm proteins and absent in bacterial Hfq proteins. Archaeal LSm and eukaryotic Sm proteins form heptamers, whilst Hfq form hexamers. However, it is worth noting that, unlike eukaryotic Sm heptamers composed of subunits differing in primary structure, archaeal LSm heptamers are composed of identical ones. Nonetheless, in general, archaeal LSm proteins are more closely related to eukaryotic Sm proteins than to bacterial Hfq proteins.

The structure of archaeal Sm2 family member AF-Sm2 from *A. fulgidus*, which was determined in 2002, demonstrates that the archaeal LSm proteins can also form hexamers [95]. AF-Sm2 hexamers significantly differ from AF-Sm1 heptamers in surface distribution of

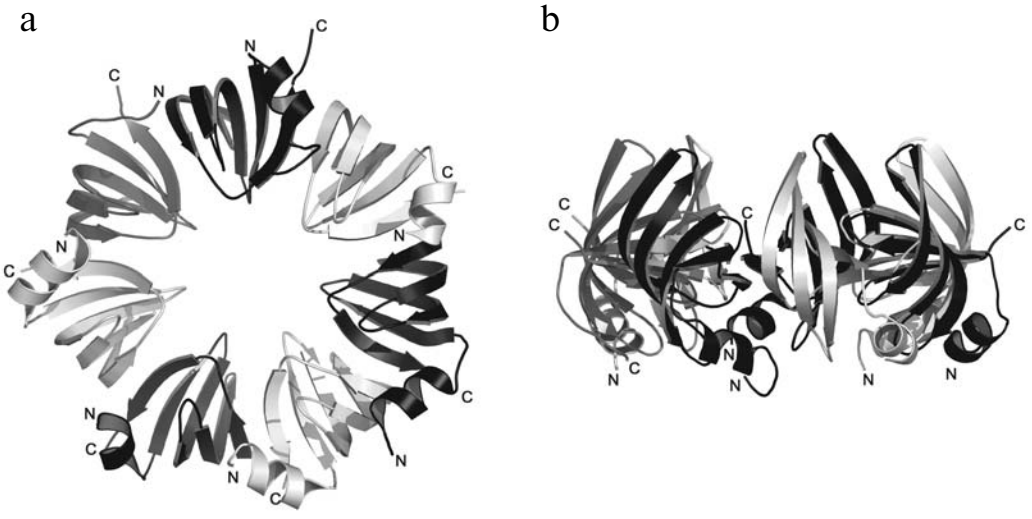

Fig. 6. Spatial structure of archaeal heptameric LSm protein from *P. aerophilum* (PDB ID 1I8F). a) View from the central pore. b) Side view of the toroid. The N- and C-ends of the protein monomers are labeled.

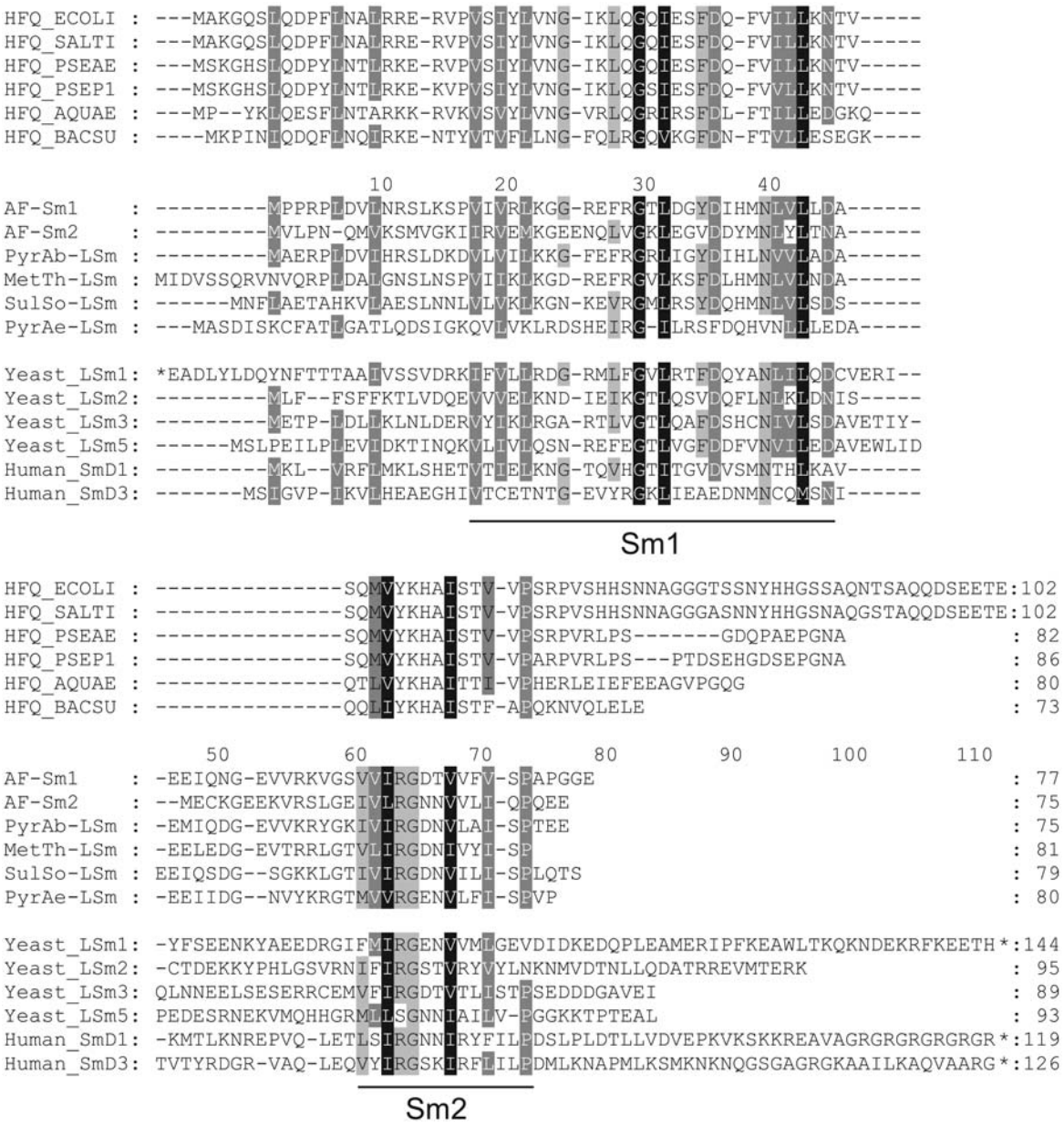

Fig. 7. Comparison of amino acid sequences of archaeal LSm, bacterial Hfq, and eukaryotic Sm proteins. The following sequences are shown: bacterial – *Escherichia coli* (HFQ_ECOLI), *Salmonella typhimurium* (HFQ_SALTY), *Pseudomonas aeruginosa* (HFQ_PSEAE), *Pseudomonas putida* (HFQ_PSEP1), *Aquifex aeolicus* (HFQ_AQUAE), and *Bacillus subtilis* (HFQ_BACSU); archaeal – *Archaeoglobus fulgidus* AF-Sm1 and AF-Sm2, *Pyrococcus abyssi* (PyrAb-LSm), *Methanobacterium thermoautotrophicum* (MetTh-LSm), *Sulfolobus solfataricus* (SulSo-LSm), and *Pyrobaculum aerophilum* (PyrAe-LSm); eukaryotic – yeast LSm (Yeast_LSm1, Yeast_LSm2, Yeast_LSm3, and Yeast_LSm5) and human Sm (Human_SmD1, Human_SmD3). The numbering corresponds to the AF-Sm1 sequence. Conserved in all sequences amino acid residues are drawn in black, conserved in 80% sequences are in dark gray, conserved among sequences are in light gray. In bacterial sequences rectangle denotes a conserved [Y/F]KHAI motif. Conserved Sm1 and Sm2 motifs are indicated at the bottom. Asterisks designate shortened regions of protein sequences.

charges, but they have similar spatial structure and intersubunit contacts. Existence of AF-Sm2 hexamers in solution was confirmed by mass spectrometry. Biochemical experiments have shown a certain dependence of AF-Sm2 3D structure on pH or presence of RNA [96]. The protein forms a stable hexamer at pH 4.5, does not undergo oligomerization at pH 8.0, and forms heptamer only in presence of RNA; the stability of the heptamer–RNA complex decreases with increase of pH.

It was supposed that this feature of AF-Sm2 might be explained by interaction of Glu19, Glu23, and Glu47 on the surfaces of adjacent monomers [95]. At pH > 7.0 the carboxylic groups of glutamic acids are deprotonated, which would destabilize the oligomer because of repul-

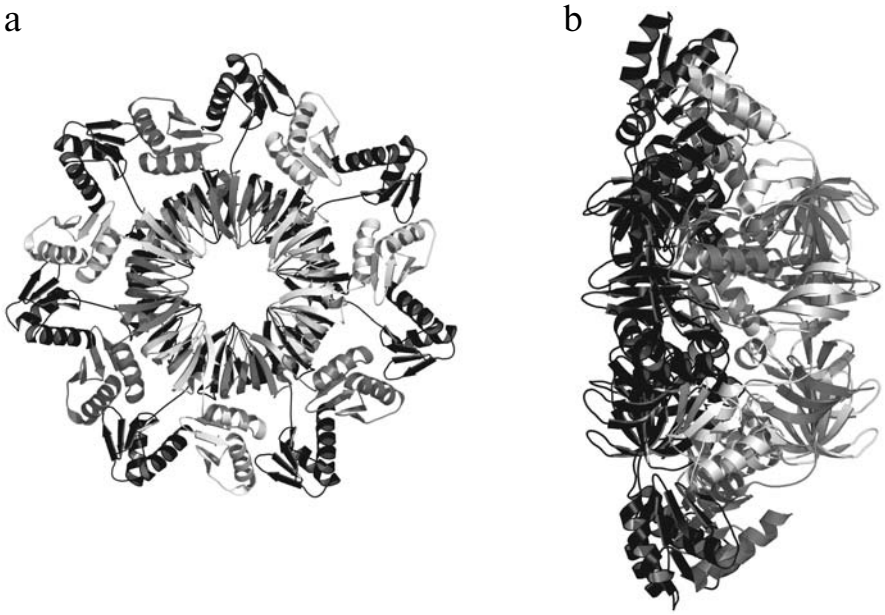

Fig. 8. Quaternary structure of 14-mer *P. aerophilum* LSm protein SmAP3 (PDB ID 1M5Q). a) Axial view. b) Lateral view. Adjacent monomers are drawn in black and gray.

sion of closely disposed negatively charged groups. As pH decreases, the glutamate side chains are protonated to eliminate repulsion of the charges, so the hexamer formation becomes possible. To verify this hypothesis, mutants of the protein with single substitutions of Glu19, Glu23, or Glu47 were constructed, and their oligomerization degree was determined at various pH values in presence or absence of RNA [96]. All of the mutants compared to the wild-type protein were characterized by increased stability of their oligomers, and addition of RNA resulted in RNA–protein complex formation.

These data indicated great significance of Glu19, Glu23, and Glu47 in quaternary structure formation, but the data on structure and function of archaeal Sm2 proteins are not yet sufficient for the statement that the above feature of AF-Sm2 is unique and independent on other factors [96].

Another interesting exception from heptameric architecture of archaeal LSm proteins is the 14-mer protein from *Pyrobaculum aerophilum*, which was classified as a member of the new family SmAP3 [97, 98] (Fig. 8). Phylogenetic analysis and search for homologs in protein databases show uniqueness of this protein. It contains an additional C-terminal domain implicated in dimerization of two heptamers. One of the heptamers occupies apical position, while the other equatorial, to form a cylinder with a central pore; their C-terminal domains contact each other in the outer space of the cylinder with formation of nonpolar contacting areas and abundance of hydrogen bonds. In solution, SmAP3 can form several oligomeric forms in which the number of monomers is divisible by seven $(SmAP3)_{n=7,14,28}$. This feature of the protein is seen in the presence of bivalent metal cations, which bind with the protein. The structure the authors prepared contained 14 cadmium (Cd^{2+})-binding sites different for apical and equatorial subunits.

Function of archaeal LSm proteins. Functions of archaeal LSm proteins in the cell are still poorly understood. Most probably, archaea lack a splicing system, so archaeal LSm proteins, unlike eukaryotic Sm proteins, cannot be involved in spliceosome formation. In terms of function, they seem to be more closely related to bacterial Hfq proteins and are implicated in RNA processing or play a role of messenger in RNA–RNA contacts.

The structure of *Archaeoglobus fulgidus* AF–Sm1 complex with U_5 ribooligonucleotide first demonstrated the details of RNA interaction with heptameric LSm protein of archaea [91] (Fig. 9). The U_5 RNA binds in a central hole of heptameric ring similarly to the AU_5G RNA in a complex with *S. aureus* Hfq [23]. Each base is located between side chains of His37 and Arg63 in a pocket formed by highly conserved residues of one monomer. Since the structure resolution was 2.75 Å, the distances between atoms of nucleotides and atoms of the protein vary significantly from base to base, and interatomic distances for expected hydrogen bonds do not coincide well. The most reliable results are interactions between the uracil O4 and N3 atoms and nitrogen and oxygen atoms of Asn39 side chain, as well as between uracil O2 atom and nitrogen atom of protein backbone (Asp65). The authors suggest that such position of the Asn39 side chain is not beneficial for binding of cytosine because its methyl group meets a steric hindrance from asparagine atoms. On the other hand, on the bottom of the nucleotide-bind-

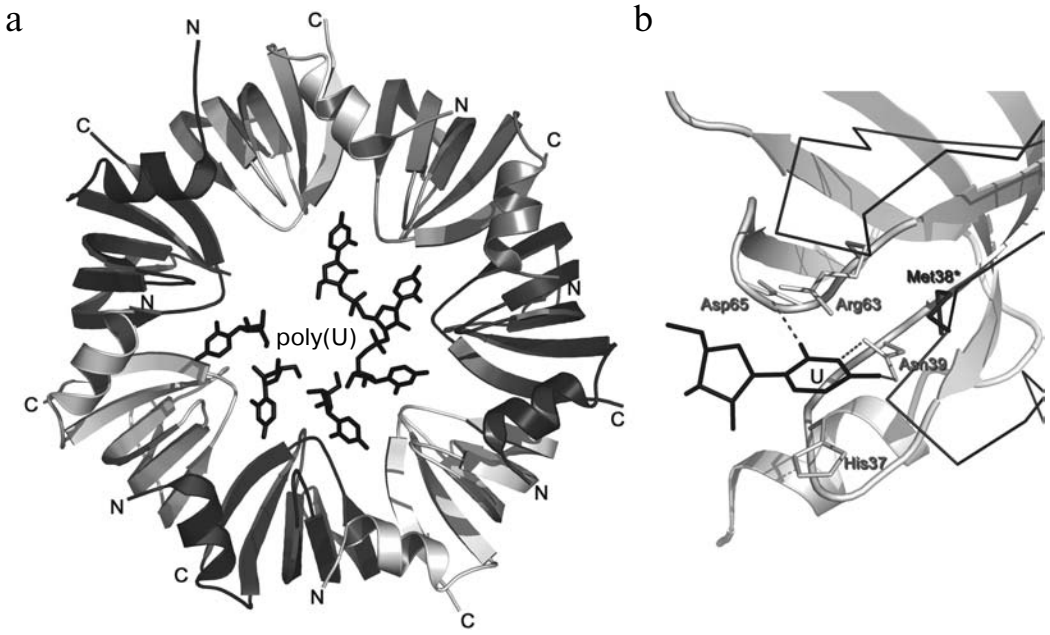

Fig. 9. Spatial structure of *A. fulgidus* Hfq protein AF-Sm1 in complex with poly(U)$_6$ ribooligonucleotide (black) (PDB ID 1I5L). a) View of the region of poly(U)$_6$ binding. b) Uridine binding site on the protein surface. One of the protein monomers is presented as ribbon model, the neighboring monomer is shown as Cα trace (dark gray). Amino acid residues that contact with nucleotides are shown. Met38* restricts the nucleotide binding pocket. Hydrogen bonds are shown as dotted lines. Asterisks denote amino acid residues of an adjacent monomer.

ing pocket Met38 should hamper the entry of a large pyrimidine base into this binding site. These data suggest that the given binding site is specific for U-rich RNA regions.

Data implicating sRNAs in regulation of gene expression in archaeon *Haloferax volcanii* was published in 2009 [99, 100]. This halophilic archaeon possesses single gene encoding LSm protein belonging to the LSm1 family. Electrophoresis mobility shift assay at non-denaturing conditions and mass spectrometric data suggest that this LSm protein forms complexes with oligo(U)$_{15-30}$ RNA, tRNAs, and several small RNAs [4]. Experiments on immune coprecipitation with *in vivo* produced FLAG-LSm protein (carrying conjugated FLAG polypeptide) have confirmed that the protein binds to more than 15 different small RNA, as well as more than 30 different proteins, most of them participating in translation regulation or RNA metabolism. The strain *H. volcanii* Δ*lsm* also obtained in this work was able to survive, but it gave colonies of smaller size and grew slower than the wild-type strain [4]. Removal of the gene encoding LSm protein resulted in prominent pleiotropic phenotype, suggesting involvement of the protein in many cell processes; however, these processes and the role of LSm protein are still to be elucidated.

Formation of fibrillar structures by LSm proteins. Besides formation of the quaternary structure in form of heptameric doughnut, archaeal LSm proteins in solution can form complex fibrillary structures. This polymerization was demonstrated for archaeal LSm proteins from *Pyrobaculum aerophilum* and *Methanobacterium thermoautotrophicum* [97].

The *M. thermoautotrophicum* SmAP1 can form fibers at protein concentration of 0.5-1.2 mg/ml in 10-25 mM Tris-HCl buffer, pH 6.5-7.5, containing 20-60 mM NaCl. The diameter of the fibril is comparable with the size of a single SmAP1 heptamer, suggesting a model for protein folding into fiber with parallel axes of heptamer and fibrilla [97]. Several protein fibers can associate side-to-side to form flocculi or sheets.

SmAP1 from *P. aerophilum* forms similar fibers, but reducing conditions are necessary for this. Under oxidizing conditions the protein forms 14-mer connected by Cys8–Cys8 disulfide bond, with charged surface Lys4 residues at both ends of the dimerized heptamer. Each 14-mer maintains the head-to-head shape, and electrostatic repulsion of 14-mer surfaces due to the presence of charged surface lysines hinders fibrilla formation. The protein forms fibers when reducing agents disrupting disulfide bridges between heptamers are added, or when mutation Cys8Ser is present [97].

Later, highly regular fibers of *E. coli* Hfq were prepared [101] under conditions which substantially differed from those for archaeal SmAP1: the protein was first dialyzed against 5 mM Tris-HCl (pH 8.0) containing 5 mM NaCl and 0.005% (w/v) dodecyl-β-D-maltoside, lyophilized, and resuspended in water. The resulting structures were analyzed by IR spectroscopy and electron

microscopy followed by reconstruction of fiber structure on the basis of known spatial structure of the protein hexamer. They proved to be different in shape and oligomer symmetry from those formed by archaeal SmAP1. SmAP1 proteins form tube-like polar fibers with head-to-tail packaging, whereas *E. coli* Hfq forms cylindrical helically coiled fibers composed of six protein hexamers arranged in a thin layer across the cylinder surface. The diameter of such cylinder is 170 Å, and the angle between the hexamer plane and the fiber axis is 37.7° (calculated data assuming the pitch of a helix of 240 Å and distance between Hfq homohexamer planes of ~40.5 Å) [101]. Thus, bacterial Hfq proteins can form fibrillar structure, but it differs from that of fibers formed by archaeal proteins.

To summarize the data on structure and function of bacterial Hfq proteins, one can say that many facts are known. Nevertheless, a number of unsolved questions remain, particularly, from what are the causes of high stability of these proteins to what are the molecular principles enabling these proteins to perform their specific functions. The mechanism regulating mRNA translation involving regulatory RNAs and Hfq protein is still only hypothetical. It is unknown how this protein binds to large RNA molecules. Substantial data scattering exists in evaluating RNA–protein complex stoichiometry. So when principles of small regulatory RNA action are under comprehensively studied these questions remain extremely important.

One can find many common items when considering structures and functions of bacterial Hfq and archaeal LSm proteins: the presence of Sm structural motif, ability to form doughnut-shaped quaternary structures composed of homological monomers, structural similarity of RNA-binding sites, high affinity to U-enriched RNA sites, etc. These facts suggest that LSm proteins in Archaea should perform similar functions. Nonetheless, virtually nothing is known about functions of archaeal LSm proteins. Further study in this field is of great interest.

This study was supported by the Program of the Russian Academy of Sciences Presidium "Molecular and Cellular Biology" and the Ministry for Science and Education (State contract No. 02.740.11.0295).

REFERENCES

1. Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., and Mayer, O. (2007) *RNA Biol.*, **4**, 118-130.
2. Wilusz, C. J., and Wilusz, J. (2005) *Nat. Struct. Mol. Biol.*, **12**, 1031-1036.
3. Tharun, S. (2009) *Int. Rev. Cell Mol. Biol.*, **272**, 149-189.
4. Fischer, S., Benz, J., Spath, B., Maier, L., Straub, J., and Granzow, M. (2010) *J. Biol. Chem.*, **285**, 34429-34438.
5. Chao, Y., and Vogel, J. (2010) *Curr. Opin. Microbiol.*, **13**, 24-33.
6. Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004) *Mol. Microbiol.*, **51**, 1525-1533.
7. Notman, D. D., Kurata, N., and Tan, E. M. (1975) *Ann. Int. Med.*, **83**, 464-469.
8. Lerner, M. R., and Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5495-5499.
9. Luhrmann, R., Kastner, B., and Bach, M. (1990) *Biochim. Biophys. Acta*, **1087**, 265-292.
10. Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., and Brahms, H. (1995) *EMBO J.*, **14**, 2076-2088.
11. He, W., and Parker, R. (2000) *Curr. Opin. Cell Biol.*, **12**, 346-350.
12. Yong, J., Golembe, T. J., Battle, D. J., Pellizzoni, L., and Dreyfuss, G. (2004) *Mol. Cell. Biol.*, **24**, 2747-2756.
13. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J., and Nagai, K. (2009) *Nature*, **458**, 475-480.
14. Shapiro, L., Franze de Fernandez, M. T., and August, J. T. (1968) *Nature*, **220**, 478-480.
15. Franze de Fernandez, M. T., Eoyang, L., and August, J. T. (1968) *Nature*, **219**, 588-590.
16. Franze de Fernandez, M. T., Hayward, W. S., and August, J. T. (1972) *J. Biol. Chem.*, **247**, 824-831.
17. Kajitani, M., and Ishihama, A. (1991) *Nucleic Acids Res.*, **19**, 1063-1066.
18. Schuppli, D., Miranda, G., Tsui, H. C., Winkler, M. E., Sogo, J. M., and Weber, H. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 10239-10242.
19. Tsui, H. C., Feng, G., and Winkler, M. E. (1997) *J. Bacteriol.*, **179**, 7476-7487.
20. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C., and Storz, G. (2002) *Mol. Cell*, **9**, 11-22.
21. Moller, T., Franch, T., Hojrup, P., Keene, D. R., Bachinger, H. P., and Brennan, R. G. (2002) *Mol. Cell*, **9**, 23-30.
22. Sun, X., Zhulin, I., and Wartell, R. M. (2002) *Nucleic Acids Res.*, **30**, 3662-3671.
23. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P., and Brennan, R. G. (2002) *EMBO J.*, **21**, 3546-3556.
24. Sauter, C., Basquin, J., and Suck, D. (2003) *Nucleic Acids Res.*, **31**, 4091-4098.
25. Moskaleva, O., Melnik, B., Gabdulkhakov, A., Garber, M., Nikonov, S., and Stolboushkina, E. (2010) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **66**, 760-764.
26. Nikulin, A., Stolboushkina, E., Perederina, A., Vassilieva, I., Blaesi, U., and Moll, I. (2005) *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 141-146.
27. Boggild, A., Overgaard, M., Valentin-Hansen, P., and Brodersen, D. E. (2009) *FEBS J.*, **276**, 3904-3915.
28. Dienst, D., Duhring, U., Mollenkopf, H., Vogel, J., Golecki, J., and Hess, W. R. (2008) *Microbiology*, **154**, 3134-3143.
29. Seraphin, B. (1995) *EMBO J.*, **14**, 2089-2098.
30. Murzin, A. G. (1993) *EMBO J.*, **12**, 861-867.
31. Musacchio, A., Gibson, T., Lehto, V. P., and Saraste, M. (1992) *FEBS Lett.*, **307**, 55-61.
32. Brennan, R. G., and Link, T. M. (2007) *Curr. Opin. Microbiol.*, **10**, 125-133.

33. Tsui, H. C., Leung, H. C., and Winkler, M. E. (1994) *Mol. Microbiol.*, **13**, 35-49.
34. Sonnleitner, E., Moll, I., and Blasi, U. (2002) *Microbiology*, **148**, 883-891.
35. Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., and Seguin, J. (2004) *Eur. J. Biochem.*, **271**, 1258-1265.
36. Sittka, A., Sharma, C. M., Rolle, K., and Vogel, J. (2009) *RNA Biol.*, **6**, 266-275.
37. Nielsen, J. S., Lei, L. K., Ebersbach, T., Olsen, A. S., Klitgaard, J. K., and Valentin-Hansen, P. (2010) *Nucleic Acids Res.*, **38**, 907-919.
38. Nielsen, J. S., Boggild, A., Andersen, C. B. F., Nielsen, G., Boysen, A., and Brodersen, D. E. (2007) *RNA*, **13**, 2213-2223.
39. Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R., and Blasi, U. (2008) *Nucleic Acids Res.*, **36**, 133-143.
40. Olsen, A. S., Moller-Jensen, J., Brennan, R. G., and Valentin-Hansen, P. (2010) *J. Mol. Biol.*, **404**, 173-182.
41. Vassilieva, Yu. M., and Garber, M. B. (2002) *Mol. Biol. (Moscow)*, **36**, 785-791.
42. Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S. A. (2010) *Proc. Natl. Acad. Sci. USA*, **107**, 9602-9607.
43. Basineni, S. R., Madhugiri, R., Kolmsee, T., Hengge, R., and Klug, G. (2009) *RNA Biol.*, **6**, 584-594.
44. Sledjeski, D. D., Whitman, C., and Zhang, A. (2001) *J. Bacteriol.*, **183**, 1997-2005.
45. Nogueira, T., and Springer, M. (2000) *Curr. Opin. Microbiol.*, **3**, 154-158.
46. Aiba, H. (2007) *Curr. Opin. Microbiol.*, **10**, 134-139.
47. Gorke, B., and Vogel, J. (2008) *Genes Dev.*, **22**, 2914-2925.
48. Urban, J. H., and Vogel, J. (2007) *Nucleic Acids Res.*, **35**, 1018-1037.
49. Argaman, L., and Altuvia, S. (2000) *J. Mol. Biol.*, **300**, 1101-1112.
50. Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R., and Storz, G. (1998) *EMBO J.*, **17**, 6061-6068.
51. Altuvia, S., Zhang, A., Argaman, L., Tiwari, A., and Storz, G. (1998) *EMBO J.*, **17**, 6069-6075.
52. Lease, R. A., and Belfort, M. (2000) *Mol. Microbiol.*, **38**, 667-672.
53. Sledjeski, D. D., Gupta, A., and Gottesman, S. (1996) *EMBO J.*, **15**, 3993-4000.
54. Vecerek, B., Beich-Frandsen, M., Resch, A., and Blasi, U. (2010) *Nucleic Acids Res.*, **38**, 1284-1293.
55. Muffler, A., Fischer, D., and Hengge-Aronis, R. (1996) *Genes Dev.*, **10**, 1143-1151.
56. Brown, L., and Elliott, T. (1996) *J. Bacteriol.*, **178**, 3763-3770.
57. Dong, T., and Schellhorn, H. E. (2009) *Mol. Genet. Genom.*, **281**, 19-33.
58. Loewen, P. C., and Hengge-Aronis, R. (1994) *Annu. Rev. Microbiol.*, **48**, 53-80.
59. Hengge-Aronis, R. (2000) *Adv. Exp. Med. Biol.*, **485**, 85-93.
60. Hengge-Aronis, R. (1996) *Mol. Microbiol.*, **21**, 887-893.
61. Hengge-Aronis, R., Lange, R., Henneberg, N., and Fischer, D. (1993) *J. Bacteriol.*, **175**, 259-265.
62. Hengge-Aronis, R. (1993) *Cell*, **72**, 165-168.
63. Fang, F. C., Libby, S. J., Buchmeier, N. A., Loewen, P. C., Switala, J., and Harwood, J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11978-11982.
64. Webb, C., Moreno, M., Wilmes-Riesenberg, M., Curtiss, R. III., and Foster, J. W. (1999) *Mol. Microbiol.*, **34**, 112-123.
65. Jorgensen, F., Bally, M., Chapon-Herve, V., Michel, G., Lazdunski, A., and Williams, P. (1999) *Microbiology*, **145**, 835-844.
66. Suh, S. J., Silo-Suh, L., Woods, D. E., Hassett, D. J., West, S. E., and Ohman, D. E. (1999) *J. Bacteriol.*, **181**, 3890-3897.
67. Badger, J. L., and Miller, V. L. (1995) *J. Bacteriol.*, **177**, 5370-5373.
68. Lange, R., and Hengge-Aronis, R. (1994) *Genes Dev.*, **8**, 1600-1612.
69. Yamashino, T., Ueguchi, C., and Mizuno, T. (1995) *EMBO J.*, **14**, 594-602.
70. Muffler, A., Traulsen, D. D., Fischer, D., Lange, R., and Hengge-Aronis, R. (1997) *J. Bacteriol.*, **179**, 297-300.
71. Vecerek, B., Moll, I., and Blasi, U. (2005) *RNA*, **11**, 976-984.
72. Vytvytska, O., Jakobsen, J. S., Balcunaite, G., Andersen, J. S., Baccarini, M., and von Gabain, A. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 14118-14123.
73. Hajnsdorf, E., and Regnier, P. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 1501-1505.
74. Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P., and Hajnsdorf, E. (2003) *Nucleic Acids Res.*, **31**, 7302-7310.
75. Takada, A., Wachi, M., and Nagai, K. (1999) *Biochem. Biophys. Res. Commun.*, **266**, 579-583.
76. Wachi, M., Takada, A., and Nagai, K. (1999) *Biochem. Biophys. Res. Commun.*, **264**, 525-529.
77. Vytvytska, O., Moll, I., Kaberdin, V. R., von Gabain, A., and Blasi, U. (2000) *Genes Dev.*, **14**, 1109-1118.
78. Udekwu, K. I., Darfeuille, F., Vogel, J., Reimegard, J., Holmqvist, E., and Wagner, E. G. H. (2005) *Genes Dev.*, **19**, 2355-2366.
79. De Haseth, P. L., and Uhlenbeck, O. C. (1980) *Biochemistry*, **19**, 6146-6151.
80. Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D., and Feig, A. L. (2004) *Nat. Struct. Mol. Biol.*, **11**, 1206-1214.
81. Sun, X., and Wartell, R. M. (2006) *Biochemistry*, **45**, 4875-4887.
82. Link, T. M., Valentin-Hansen, P., and Brennan, R. G. (2009) *Proc. Natl. Acad. Sci. USA*, **106**, 19292-19297.
83. Lease, R. A., and Woodson, S. A. (2004) *J. Mol. Biol.*, **344**, 1211-1223.
84. Updegrove, T. B., Correia, J. J., Chen, Y., Terry, C., and Wartell, R. M. (2011) *RNA*, **17**, 489-500.
85. Azam, T. A., and Ishihama, A. (1999) *J. Biol. Chem.*, **274**, 33105-33113.
86. Ishihama, A. (1999) *Genes Cells*, **4**, 135-143.
87. Takada, A., Wachi, M., Kaidow, A., Takamura. M., and Nagai, K. (1997) *Biochem. Biophys. Res. Commun.*, **236**, 576-579.
88. Updegrove, T. B., Correia, J. J., Galletto, R., Bujalowski, W., and Wartell, R. M. (2010) *Biochim. Biophys. Acta*, **1799**, 588-596.
89. Vlahovicek, K., Kajan, L., and Pongor, S. (2003) *Nucleic Acids Res.*, **31**, 3686-3687.
90. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Seraphin, B. (1999) *EMBO J.*, **18**, 3451-3462.

91. Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B., and Suck, D. (2001) *EMBO J.*, **20**, 2293-2303.
92. Achsel, T., Stark, H., and Luhrmann, R. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 3685-3689.
93. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M., and Mabbutt, B. C. (2001) *J. Mol. Biol.*, **309**, 915-923.
94. Mura, C., Cascio, D., Sawaya, M. R., and Eisenberg, D. S. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 5532-5537.
95. Toro, I., Basquin, J., Teo-Dreher, H., and Suck, D. (2002) *J. Mol. Biol.*, **320**, 129-142.
96. Kilic, T., Sanglier, S., van Dorsselaer, A., and Suck, D. (2006) *Protein Sci.*, **15**, 2310-2317.
97. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M., and Eisenberg, D. (2003) *Protein Sci.*, **12**, 832-847.
98. Mura, C., Phillips, M., Kozhukhovsky, A., and Eisenberg, D. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 4539-4544.
99. Soppa, J., Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., and Fischer, S. (2009) *Biochem. Soc. Trans.*, **37**, 133-136.
100. Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Soppa, J., and Marchfelder, A. (2009) *RNA Biol.*, **6**, 281-292.
101. Arluison, V., Mura, C., Guzman, M. R., Liquier, J., Pellegrini, O., and Gingery, M. (2006) *J. Mol. Biol.*, **356**, 86-96.