doi:10.1016/j.jmb.2005.11.010 J. Mol. Biol. (2006) 356, 86–96

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Three-dimensional Structures of Fibrillar Sm Proteins: Hfq and Other Sm-like Proteins

Véronique Arluison[1]*†, Cameron Mura[2]†, Maria Romero Guzmán[3] Jean Liquier[3], Olivier Pellegrini[1], Mari Gingery[4], Philippe Régnier[1] and Sergio Marco[5]

[1]*Institut de Biologie Physico-Chimique, CNRS UPR 9073 conventionnée avec l'université Paris 7, 13 rue P. et M. Curie 75005 Paris, France*

[2]*Department of Chemistry and Biochemistry and Center for Theoretical Biological Physics University of California, San Diego; La Jolla, CA 92093-0365 USA*

[3]*Laboratoire BioMoCeTi UMR CNRS 7033-Université Paris 13, UFR Médecine Santé Biologie humaine, 74 rue Marcel Cachin, 93017 Bobigny, France*

[4]*UCLA-DOE Institute for Genomics and Proteomics Los Angeles, CA 90095-1570 USA*

[5]*Institut Curie CNRS UMR 168, 11 rue P. et M. Curie 75005 Paris, France*

**Corresponding author*

Hfq is a nucleic acid-binding protein that functions as a global regulator of gene expression by virtue of its interactions with several small, non-coding RNA species. Originally identified as an *Escherichia coli* host factor required for RNA phage Qβ replication, Hfq is now known to post-transcriptionally regulate bacterial gene expression by modulating both mRNA stability and translational activity. Recently shown to be a member of the diverse Sm protein family, Hfq adopts the OB-like fold typical of other Sm and Sm-like (Lsm) proteins, and also assembles into toroidal homo-oligomers that bind single-stranded RNA. Similarities between the structures, functions, and evolution of Sm/Lsm proteins and Hfq are continually being discovered, and we now report an additional, unexpected biophysical property that is shared by Hfq and other Sm proteins: *E. coli* Hfq polymerizes into well-ordered fibres whose morphologies closely resemble those found for Sm-like archaeal proteins (SmAPs). However, the hierarchical assembly of these fibres is dissimilar: whereas SmAPs polymerize into polar tubes (and striated bundles of such tubes) by head-to-tail stacking of individual homo-heptamers, helical Hfq fibres are formed by cylindrical slab-like layers that consist of 36 subunits arranged as a hexamer of Hfq homo-hexamers (i.e. protofilaments in a 6×6 arrangement). The different fibrillar ultrastructures formed by Hfq and SmAP are presented and examined herein, with the overall goal of elucidating another similarity amongst the diverse members of the Sm protein family.

Keywords: RNA-binding protein; Sm-like (Lsm) protein; helical fibre; electron microscopy; FTIR

Introduction

Hfq is a phylogenetically conserved RNA-binding protein that was recently found to function as a key post-transcriptional regulator of bacterial gene expression. Discovered as an *Escherichia coli* host factor required for activity of RNA bacteriophage Qβ replicase,[1] the highly abundant Hfq protein (≈90 μM intracellularly)[2] coordinates several aspects of bacterial RNA metabolism (e.g. mRNA decay pathways), with its function typically being of a passive rather than active/catalytic nature (e.g. binding poly(A) tails and protecting mRNA from RNase E digestion[3]). Hfq also has been found to be a nucleoid-associated protein, thus suggesting a possible role in DNA packaging.[4,5] The regulatory activities of Hfq are frequently mediated by small, non-coding RNAs (sRNA), which can be considered as bacterial analogs of the diverse families of microRNAs that were recently uncovered as crucial regulatory elements in eukaryotic gene expression.[6]

† V. A. and C. M. contributed equally to this work.

Abbreviations used: ATR, attenuated total reflectance; FT, Fourier transform; IR, infrared; EM, electron microscopy; *Mth*, *Methanobacterium thermautotrophicum*; S(*f*), power spectrum; SmAP, Sm-like archaeal protein; Lsm, Sm-like protein; sRNA, small (non-coding) RNA; ssRNA, single-stranded RNA; β_{Π}, aggregated β-sheet.

E-mail address of the corresponding author: veronique.arluison@ibpc.fr

0022-2836/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.

The pleiotropic effects of Hfq upon bacterial gene expression are illustrated by the fact that disruption of the *E. coli hfq* gene affects cell viability in numerous ways, including heightened UV sensitivity and decreased growth rates.[7] These deleterious phenotypes are linked to the role of Hfq in modifying the gene expression patterns of bacteria subjected to environmental stresses. In particular, Hfq is required for the regulatory function of various sRNA "riboregulators" (such as DsrA, OxyS, and RprA) that mediate the production of an RNA polymerase subunit (σ^s) specific to stationary-phase survival by virtue of base-pairing between the sRNAs and the corresponding σ^s transcript (i.e. *rpoS* mRNA).[8] Indeed, this demonstrates the general principle that the global/downstream effects of Hfq are often achieved *via* interactions with sRNAs. Other examples of protein pathways regulated by Hfq-mediated coupling of a regulatory sRNA to a particular target RNA (typically an mRNA such as *rpoS*) include the combinations OxyS···Hfq···*fhlA*, {DsrA, OxyS, RprA}···Hfq···*rpoS* (σ^s), and Spot42···Hfq···*galK*, where {braces} indicate a set of multiple possible mRNA or sRNA binding partners, mRNA targets are italicized, and regulatory sRNAs are underlined (reviewed by Masse *et al.*[9]); note that Hfq may also operate by directly interacting with upstream regulatory elements of mRNA, as in the case of the 5′ untranslated region of *ompA* mRNA.[10] Hfq fulfils these sRNA-related functions by various mechanisms. For example, Hfq can alter the stability of mRNA transcripts, either secondary structural elements within those transcripts[11] or the degradation of the transcripts themselves,[3] and can also increase or decrease the translational activity associated with certain mRNAs, possibly *via* direct interactions with 30 S ribosomal proteins.[12] Likewise, Hfq also modulates sRNA stabilities.[13] Despite the wealth of recent information about the roles of Hfq in bacterial physiology, many details of the molecular mechanisms underlying its biological activities are unclear.

Current molecular-level understanding of Hfq activity stems from the recent finding that it is an Sm-like protein. Other members of the phylogenetically well-conserved (but functionally diverse) Sm family include the Sm-like archaeal proteins (SmAPs) and the canonical eukaryotic Sm and Lsm proteins that function in such tasks as pre-mRNA splicing,[14] mRNA decapping and decay,[15] and telomere maintenance.[16] The relationship between Hfq and Sm proteins was first suggested by weak sequence similarities between the N-terminal regions of the ≈80–120 residue Hfq and Sm proteins, was further corroborated by phylogenetic, biophysical, fold recognition, and homology modelling studies of *E. coli* Hfq,[17–20] and has been definitively established by crystal structures of Hfq hexamers from *Staphylococcus aureus*,[21] *E. coli*,[22] and *Pseudomonas aeruginosa*.[23] As with Sm proteins, Hfq monomers adopt an OB-like fold consisting of five strongly bent, antiparallel β-strands capped by an N-terminal α-helix, and individual protomers assemble into cyclic quaternary structures perforated by a cationic pore (Figure 1). Whereas non-bacterial Sm proteins

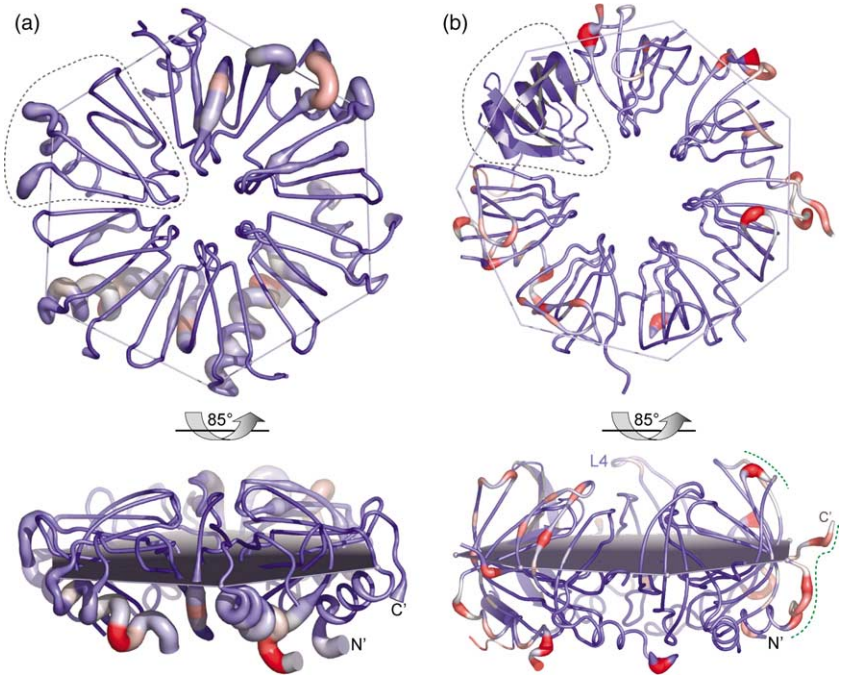

Figure 1. Tertiary and quaternary structures of Hfq and other Sm-like proteins. Nearly perpendicular views of the 3D structures of hexameric (a) *E. coli* Hfq protein and (b) heptameric Sm-like archaeal protein *Mth* SmAP1 are illustrated as tubular splines that trace the C^α backbones (monomeric subunits are delimited by broken grey lines). In these renditions, residue-specific tube diameters vary inversely with the minimum distance between a given residue and all of its nearby oligomeric neighbors within the 3D structures of the Hfq fibres and SmAP tubes (likewise, color is graded from blue→white→red). Thus, besides providing an overview of the similarities and differences between the tertiary and quaternary structures formed by Hfq and SmAP monomers and oligomers, this Figure also indicates that the majority of inter-oligomeric contacts are mediated by residues lying on both faces of the disc-shaped Hfq hexamer. In addition, residues forming the equatorial rim of the SmAP1 heptamer participate in inter-heptamer contacts, either within the asymmetric unit of the 1JRI structure (broken green lines) or between the pair of SmAP tubes that associate along the crystallographic $\vec{c}$ axis (see also Figure 4(e) and (f)). For the sake of clarity, a single SmAP1 monomer is illustrated in (b) in the more common style of a ribbon cartoon, and equatorial planes are drawn as filled (a) hexagons or (b) heptagons.

form toroidal heptamers and hexamers (homomeric or heteromeric), all known Hfq structures assemble as disc-shaped homohexamers.

Besides revealing a eubacterial branch of the Sm protein family, identification of Hfq as a member of this family implies that additional features may be conserved between Hfq and other Sm proteins. Emerging similarities in terms of sequence, structure, and function include (i) the aforementioned identity of the Hfq and Sm folds, (ii) the conserved property of assembling into characteristic ring-shaped oligomers, and (iii) the fact that Hfq preferentially binds to adenine-rich and uridine-rich single-stranded RNA (e.g. the UAU_4 tract just upstream of the Shine–Dalgarno sequence in *ompA* mRNA[24]). As with oligouridine binding sites for Sm proteins, these Hfq recognition sites are only weakly sequence-specific, and are generally flanked by at least one stem–loop secondary structural element. Apart from minor variations, the general mode of ssRNA-binding within the central cationic pore appears to be conserved between the *S. aureus* Hfq hexamer[21] and SmAP heptamers.[22,25] Unlike traditional Sm proteins, Hfq also binds poly(A) RNA,[3] probably by presenting different binding sites for the various classes of RNA (mRNA, sRNA, poly(A)) with which it interacts.[26] The wide variety of Hfq···ssRNA interactions and diversity of its RNA-related activities suggest that the primary biochemical function of Hfq may be as an RNA chaperone. Consistent with this are the surprising discovery of an ATPase activity for *E. coli* Hfq,[12] the potential of Hfq to unfold local regions of RNA,[11,24] and the requirement of Hfq for the activities of a broad range of sRNAs. Similar functionality as a generic, single-stranded nucleic acid-binding protein has been conjectured for the ancestral SmAPs.[27]

The work reported here further explores similarities and differences between Hfq and other Sm proteins by describing an unexpected biophysical property shared between these proteins: *E. coli* Hfq polymerizes into well-ordered fibres that closely resemble those found for SmAPs.[27] However, Hfq and SmAP fibres are assembled in different manners; whereas SmAPs polymerize into polar tubes *via* head-to-tail stacking of homo-heptamers, helical Hfq fibres are formed by slab-like layers that consist of 36 monomers hierarchically arranged as a hexamer of homo-hexamers. Combined with existing Hfq and SmAP crystal structures, examination of these fibrillar ultrastructures by infrared (IR) spectroscopy and electron microscopy (EM)-based image analysis has enabled construction of the 3D structural models presented herein.

Results

FTIR spectroscopy of hexameric Hfq reveals fibrillar aggregates rich in β-sheet content

Biophysical properties of the Hfq hexamer shown in Figure 1 have been characterized by a variety of means, most recently including vibrational spectroscopy.[28] As a continuation of such efforts, the secondary structural content of the full-length *E. coli* Hfq hexamer was probed *in vitro* by attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy, and a representative example of these spectra over the absorption range most sensitive to protein secondary structural features (i.e. the conformationally sensitive amide I region) is shown in Figure 2 (spectra were recorded in ^2H_2O so as to alleviate interference from overlapping H_2O bands). The broad absorption band in the amide I region lies at roughly $1730 \leftarrow 1530\ cm^{-1}$, and arises from the numerous absorption bands associated with individual secondary structural components, as well as vibrational coupling between such components. Thus, further analysis of Hfq IR spectra necessitated resolution of this broad band into its various components. Prior to doing so by spectral deconvolution, an initial estimate of the number and relative contributions of secondary structural components was made by computing the second derivative spectrum (not shown); peak positions were subsequently assigned to different types of secondary structural elements as indicated in the legend to Figure 2 (see also Table 1).[29–31]

Detailed analysis of Hfq spectra proceeded by resolution enhancement *via* Fourier self-deconvolution,[32] thereby producing a set of component peaks. Iterative least-squares curve-fitting was used to maximize the agreement between the amide I absorption region of the observed spectrum and the overall spectrum computed from these fitted Gaussian peaks, the summation being restricted to those components with maxima lying within the amide I region. A first round of iterations was performed based on the initial assignment of secondary structural elements to frequency ranges (Figure 2) in order to minimize the root-mean-square error. Minimum and maximum wave numbers were then fixed for these ranges and a second iterative round was used to optimize peak intensities (i.e. agreement between the calculated and overall observed spectrum); eight distinct bands were identified in this second stage. The fitted peaks centered at $1556\ cm^{-1}$ and $1585\ cm^{-1}$ were included in the curve-fitting but not in evaluating the relative secondary structural contents because they arise from side-chain absorption and are not part of the amide I region.[33,34] The fraction of a particular secondary structural element was calculated as the integrated intensity (area) of that peak, normalized by the overall peak area in the amide I region. Table 1 lists the peak positions, their assignments and estimated secondary structural contents. The strong peak at $1636\ cm^{-1}$ confirms the antiparallel β-sheet content of the protein. We note that the fractional secondary structural contents determined at these relatively high (fibrous) concentrations are consistent with the $34(\pm2)\%$ β-strand and $10(\pm1)\%$ α-helix values reported by Arluison *et al.*[28] for Hfq at lower

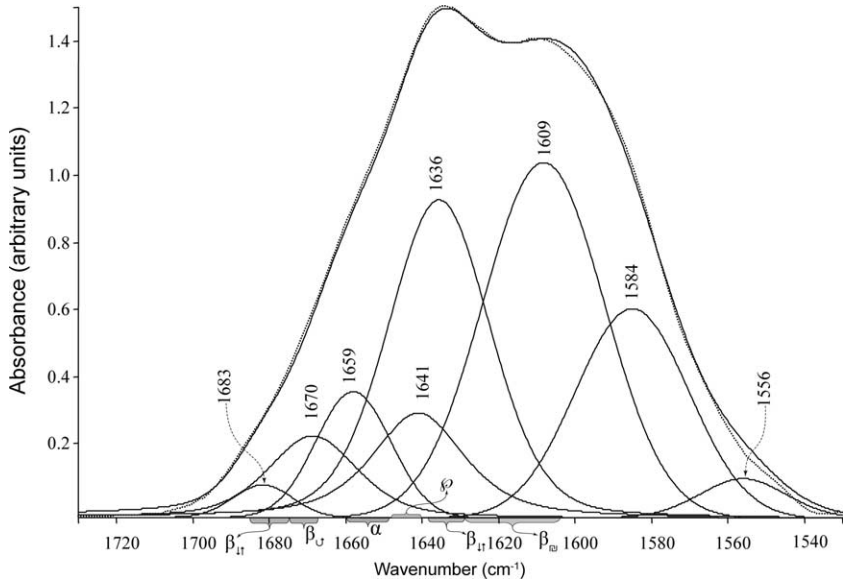

Figure 2. Secondary structure of fibrillar Hfq by FTIR spectroscopy. An ATR-FTIR spectrum of fibrillar Hfq is shown, as measured experimentally (dotted line; recorded in ^2H_2O) and as computed from the sum of components determined by least-squares curve-fitting (continuous line). The spectrum is centered near the amide I region ($\approx$1650 cm^{-1}), and positions of individual absorption bands corresponding to various secondary structural elements are indicated: α-helix (α), 1660–1650 cm^{-1}; antiparallel β-sheet ($\beta_{\Downarrow\Uparrow}$), 1636–1630 cm^{-1} and 1685–1675 cm^{-1}; β-turns ($\beta_{\cup}$), 1675–1670 cm^{-1}; irregular/random coil (℘), 1648–1640 cm^{-1}; and aggregated β-sheet ($\beta_{\Rightarrow\Rightarrow}$), 1628–1605 cm^{-1} (see Table 1 for peak positions, assignments, and estimated secondary structure contents). Note the excellent agreement between experimental and computed curves, and that the predominant feature of this spectral region is the intense shoulder at 1609 cm^{-1} that is characteristic of aggregated β-sheets found in other fibrillar proteins.

concentrations (i.e. solution samples devoid of fibres). Both the initial qualitative procedure and this more detailed deconvolution approach revealed the predominant feature of the amide I region to be an intense band centered at 1609 cm^{-1} ($\beta_{\Rightarrow\Rightarrow}$, Figure 2), consistent with the spectral features of aggregated β-sheet structures that are typically found in fibrillar proteins.[29,35,36]

A 3D model for helical Hfq fibres based on EM and image analysis

Discovery of the substantial β-aggregated signature in the FTIR spectra of soluble *E. coli* Hfq spurred further investigation of these samples by electron microscopy. The (unexpected) polymerization of Hfq into bundles of well-ordered, unbranched fibres is shown in the transmission electron micrograph in Figure 3. The fact that both Hfq and SmAP fibres formed under solution conditions not too dissimilar from physiological conditions provided additional motivation for investigating these fibrillar ultrastructures (as detailed in Materials and Methods, no denaturants, chaotropic agents, or otherwise harsh treatments were required for fibrogenesis). Negatively stained samples reproducibly displayed similarly ordered and unbranched structures, and such fibrillar morphologies are presumably a general feature of Hfq polymerization (see Discussion). Because of their highly regular appearance, image analysis techniques were used to further investigate the ultrastructure of these fibres and to provide constraints for construction of an atomically detailed 3D structural model.

Analysis of the Hfq EM images proceeded by calculation of their diffraction patterns (i.e. Fourier transformation to a power spectrum), and revealed that these Hfq fibres possess helical symmetry (Figure 3). Transforms of the EM images display roughly equivalent resolutions in the equatorial and meridional directions (Figure 3(b)), and the slightly higher resolution along the meridian (i.e. parallel with the helical fibre axis) probably results from greater intra-fibrillar order compared to lateral (inter-fibrillar) packing of the helical fibres. A diffraction resolution of 23 Å was estimated *via* a spectral signal-to-noise criterion,[37] and Figure 3(c) illustrates an isodensity surface of the Hfq fibre model (Figure 3(d) and (e)) after filtering at this resolution. Further image analysis proceeded by established methods of image alignment, clustering-based classification of subimages (as done in single-particle analysis of cryo-EM images), and class-averaging, as detailed in Materials and Methods and the legend to Figure 3. Combined with the standard theory for diffraction by helical structures,[38,39] these data revealed the helical pitch (P) to be 245 Å and the corresponding vertical spacing between consecutive helical subunits (h) to be 40 Å (Figure 3(b)). Assuming an integral number of subunits per helical turn, the number of hexameric

Table 1. Analysis of hexameric and fibrillar Hfq by FTIR spectroscopy

Absorption bands (cm^{-1})	Secondary structure component	Hfq oligomeric state (%)		
		Monoments [a]	Hexameric	Fibrillar
1670	β-turn ($\beta_{\cup}$)	7	7	8
1650-1659	α-helix (α)	14	11	9
1640-1641	irregular/random coil ($\wp$)	46	48	11
1636-1640, 1680	antiparallel β-sheet ($\beta_{\downarrow\uparrow}$)	33	34	34
1609	aggregated β-sheet (β_{Π})	ND	ND	38

ATR-FTIR spectra of both fibrillar and hexameric Hfq samples were analyzed by the deconvolution and curve-fitting procedure described in Method, with secondary structural elements assigned to particular frequency bands within the amide I absorption region as given in the first two columns. Narrow ranges (or single peaks) of absorption maxima are listed. and fractions of corresponding secondary structures are provided as percentages in the last two columns (ND, not detectable). The major difference between hexameric and fibrillar states is an increase in the spectral component characteristic of aggregated β-sheets (1609 cm^{-1} in the fiber, with an attendant decrease in the percentage of irregular/random coil component (see also Figure 1).

[a] Calculated from the 3D structure of *E. coli* Hfq (1HK9) and listed as the average value over the six monomers in the a.u.

Hfq subunits per helical pitch repeat ($N = P/h$) is ≈36. Other computed parameters of the averaged major class image include a helical diameter of 170 Å (see also Figure 3(a)).

Putative 3D models for Hfq fibres were initially built by trial-and-error so as to be consistent with these helical parameters. The final model was constructed by rigid-body transformation of the *E. coli* Hfq hexamer crystal structure and subsequent refinement by energy minimization, and is illustrated in Figure 3(d) and (e). This structure optimizes the agreement between computed model transforms (Figure 3(b), right) and the helical diffraction patterns computed for the original EM images (Figure 3(b), left), and consists of Hfq hexamers stacked at a tilt angle of 37.7° with respect to the fibre axis (corresponding to an h of ≈40.5 Å, for which $P \approx 240$ Å). Alternative fibre models, such as one in which individual layers consist of a perfect 6-fold rotationally symmetric arrangement of Hfq hexamers, were not as consistent with the EM and image analysis data. The helical pitch is indicated as a dotted yellow line in the coarse-grained representation of Figure 3(d), and the packing of Hfq hexamers along the length of the fibre is illustrated in the axial view of Figure 3(e). A quantitative view of the interactions between Hfq hexamers within the fibre can be obtained by considering the distribution of distances of interatomic contacts between a given hexamer and its six nearest neighbors, and such a representation (Figure 1(a)) reaffirms that most of the residues that mediate inter-hexamer contacts lie along both faces of the disc-shaped hexamer.

A 3D model for tubular SmAP fibres based on EM, image analysis, and crystallography

The polymerization of SmAPs into well-ordered fibres, and the further association of these fibres into the striated bundles shown in Figure 4(a), were discovered by transmission EMs, as reported.[27] Analysis of SmAP EM images proceeded in a similar manner as described above for Hfq. Diffraction patterns obtained by Fourier transformation of fibre images revealed strong peaks on an apparently first-order layer-line, with a corresponding real-space repeat length of 42 Å (Figure 4(b)). In contrast to the Hfq fibres, a plausible starting model for these SmAP fibres already exists in the form of a 2.8 Å resolution crystal structure: of the various crystal habits of *Mth* SmAP1, heptamers are arranged in tubes within the crystal lattice of this orthorhombic form ($P2_12_12_1$; PDB code 1JRI) *via* the head-to-tail stacking of disc-shaped heptamers.

The prominent feature of the Fourier-transformed diffraction pattern resulting from such a stacked arrangement of heptamers into tubes (Figure 4(c)) is an ≈42 Å spacing between the equator and the first intense layer-line (Figure 4(d)), which corresponds closely to the ≈40 Å height of most SmAP heptamers. Transforms of the observed (Figure 4(b)) and modelled (Figure 4(d)) fibrillar lattices are in qualitative agreement, although more detailed features in the layer-lines of the crystal lattice-derived model are absent from the experimental diffraction patterns (such additional fine structure in the equatorial layer-line may correspond to, e.g., interference fringes arising from lateral packing of SmAP tubes). As done for the 3D model of Hfq fibres, interactions between heptameric subunits of the polymer were analyzed by surveying interatomic contacts between a given SmAP1 heptamer and the 14 neighbors that surround it in the quasi-hexagonal crystal lattice of Figure 4(e). Quantification of heptamer⋯heptamer contact distances within a single SmAP1 tube as well as between tubes is represented in Figure 1(b), thus illustrating such features as the *edge–edge* contacts between the two SmAP1 heptamers of a single asymmetric unit of 1JRI (broken green lines in Figure 1(b)). As with Hfq, most of the short-range contacts are mediated by facial residues.

Discussion and Conclusions

We initially detected the *in vitro* polymerization of full-length *E. coli* Hfq into soluble aggregates by virtue of characteristic features in its FTIR spectra.

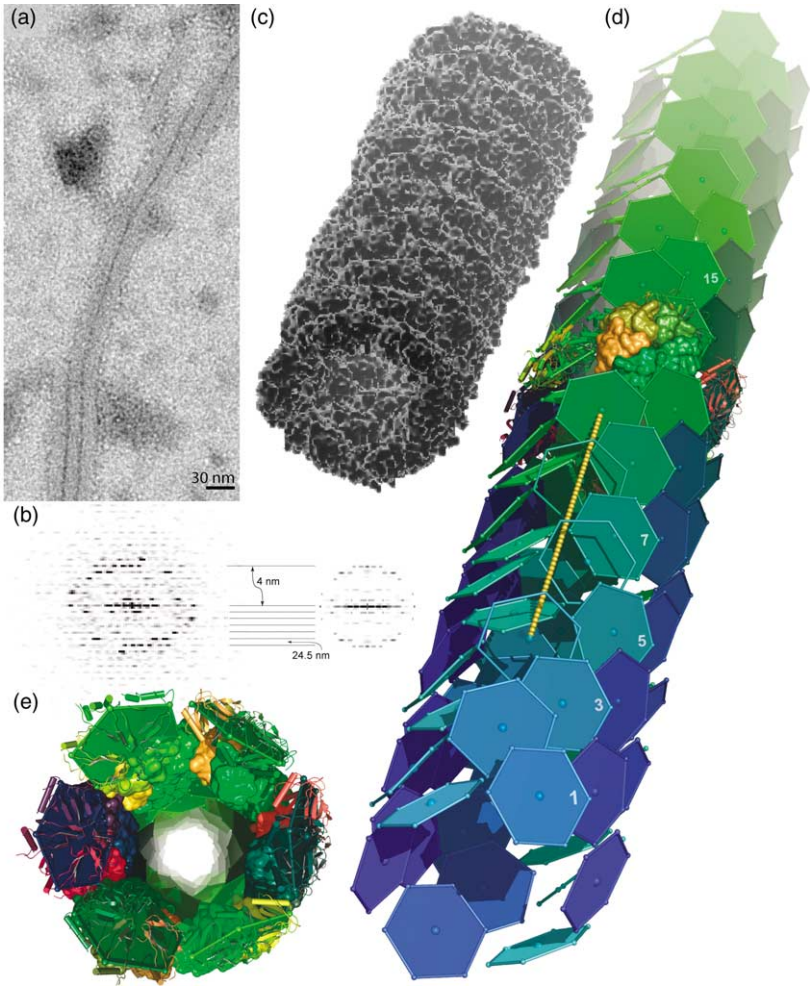

Figure 3. Structural model of fibrillar Hfq from EM and image analysis. (a) Transmission EMs of *E. coli* Hfq reveal bundles of regular, unbranched fibres. (b) Fibre images were analyzed by alignment and clustering, and averaged power spectra of a single windowed subimage from the major class of observed fibres ($\langle S(f)_{Obs} \rangle$) are shown in the left panel. Note the good agreement between this $\langle S(f)_{Obs} \rangle$ and the transform derived from resolution-filtered back-projections of the best candidate 3D structural model ((b), right panel). Putative models such as those shown in (c) were based upon layer-line spacings in these spectra, as indicated in (b). Two views of a refined, atomically detailed 3D model for Hfq fibres are shown in (d) and (e). Hfq hexamers are illustrated as hexagonal plates, with the color ramped from blue (nearest) → green (furthest). In order to convey the overall architecture of individual layers and inter-layer packing within each fibre, various hexamers from the 12th and 13th layers are rendered as solvent-accessible surface areas or as ribbon cartoons. The dotted yellow line in (e) spans a distance of ≈245 Å, and presumably corresponds to the strong layer-line spacing in (b).

This was possible because of the general principle that the highest amplitude normal modes of vibration of protein backbones are sensitive to local conformational and structural features (e.g. hydrogen bonding strength and directionality of amide groups), making these transitions useful as sensitive gauges of such properties as secondary structural composition and orientation.[39] As with other molecules, transitions between ground state vibrational levels of proteins occur in the IR region, and the amide I absorption band (lying at roughly $1730 \leftarrow 1530\ cm^{-1}$) is particularly useful in resolving different types of protein secondary structures, including α-helices, parallel and antiparallel β-sheets, β-turns, and irregular/random coil regions (Table 1). Moreover, higher-order structures formed from interactions between secondary structural motifs (e.g. intermolecular β⋯β-strand contacts) often can be distinguished by their unique IR spectral features, which ultimately arise from vibrational couplings between subunits of an oligomer that are non-existent in the monomeric state (β_{Π} regions can be distinguished due to slight differences between inter- and intramolecular β⋯β-strand contacts in terms of hydrogen bonding strength and the typical number of residues involved in such interactions[35,36,40,41]). As an example of this, IR spectra of specimens containing aggregated β-sheets typically exhibit a characteristic component centered around $1609\ cm^{-1}$.

Analysis of the amide I absorption band of soluble Hfq by spectral deconvolution and curve-fitting reveals that the higher-frequency shoulder of the overall peak arises from such an aggregated β-sheet component (Figure 2). Quantification of the secondary structural components (Table 1) shows that the same overall structure is maintained in the fibrillar state, leading us to conclude that soluble Hfq self-assembles into fibres without disruption of secondary structure. This property is in contrast to the likely mechanism underlying the formation of other β-rich polymers (e.g. amyloid fibrils), which may self-assemble *via* large conformational changes in the constituent proteins. Indeed, it has become clear that a common feature of the assembly pathways for many β-rich protein fibrils and aggregates involves conversion of a helical segment of the monomeric protein into a β-strand that then helps mediate the self-association of monomers into a highly ordered, regular structure,[42] although other proteins predisposed to form β-rich fibres

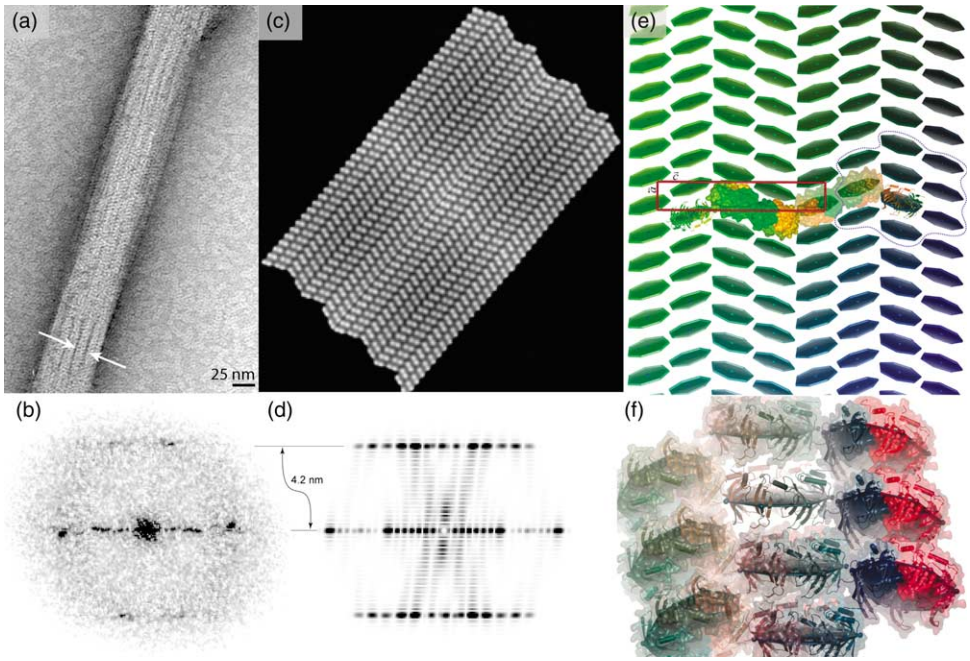

Figure 4. Confirmation of a crystallographically derived model for SmAP fibres. The polymerization of SmAPs into bundles of well-ordered, unbranched fibres was revealed by negative-stain transmission EMs such as those shown in (a). White arrows demarcate an individual SmAP fibril, and individual heptamers are visible in the background of the EM field. Agreement between (b) the averaged power spectra computed from the EM data (c) and from a putative structural model can be assessed by comparing (b) (observed) and (d) (model). The layer-line spacing is consistent with the model shown in (c), which in turn is identical with the orthorhombic crystal lattice formed by *Mth* SmAP1 (1JRI) and illustrated in (e). Unit cell edges (red) contain eight SmAP1 heptamers, and their head-to-tail stacking is illustrated by the zoomed-in portion (broken blue line), which is rendered as transparent surfaces in (f).

may do so *via* rapid growth from an initial cross-β spine nucleus.[43] Thus, although the intense β_{IR} band near 1620–1610 cm^{-1} is a shared feature of the FTIR spectra of Hfq fibres and amyloid-like fibrils, the overall 3D architecture and mode of assembly of these fibrillar structures is entirely different, and the β_{IR} band is an apparently generic feature of the FTIR spectra of β-rich protein assemblies.

Our finding that hexameric Hfq self-assembles into higher-order fibrillar structures corroborates the recent report of well-ordered tubular filaments formed by SmAP heptamers.[27] However, further investigation of the morphologies of these fibres *via* EM and image analysis reveals that the polymerization properties of Hfq and SmAP result in vastly different architectures: whereas SmAPs polymerize into polar tubes (and striated bundles of such tubes) by head-to-tail stacking of individual homo-heptamers (Figure 4), helical Hfq fibres hierarchically assemble from cylindrical slab-like layers that consist of 36 subunits arranged as a hexamer of Hfq homo-hexamers (Figure 3). Also, we note that the packing of hexamers within the lattice of the *E. coli* Hfq crystal structure (1HK9, solved in hexagonal space group $P6_1$ with one hexamer per a.u.) results in tubes of Hfq hexamers that are similar to those formed in an orthorhombic crystal structure of *Mth* SmAP1 heptamers (1JRI), the primary difference being that Hfq hexamers are not nearly as tilted with respect to the overall tube axis. Such large-scale ultrastructural differences between Hfq fibres and other Sm-like protein assemblies suggest (but certainly do not establish) significant differences in any *in vivo* functions for these polymers. As would be expected from purely geometric considerations for the packing of ellipsoidal, disc-shaped objects, most of the inter-oligomeric contacts for both Hfq and SmAP are mediated by residues that form the apical faces of these oligomers (Figure 1). Because such packing in Hfq fibres and SmAP tubes occludes most of the known functionally relevant surface patches (e.g. RNA-binding Hfq residues discussed by Mikulecky *et al.*[26]), it is likely that these Sm polymers do not retain the same biological activities as their constituent hexamers and heptamers.

Self-association of Sm-like proteins such as Hfq and SmAPs into cyclic oligomers is an established property of this protein family and is strongly linked to the various RNA-related physiological functions of Sm proteins. The higher-order assembly of these closed hexamers or heptamers into extended fibrillar (Hfq) and tubular (SmAP) ultrastructures now appears to be a generic feature of these proteins too, although the biological significance of these polymers is less clear. Lack of definitive *in vivo* relevance may explain why the polymerization of Sm-like proteins has been a relatively obscure and hitherto overlooked property of this protein family, but we note that fibre-formation is reconcilable with several possible physiological functions (e.g. sequestering mRNAs

or the store of Hfq itself, creating high local densities of cognate RNAs, etc.). In this regard, we also note that: (i) fibre formation in our experiments is compatible with known intracellular concentrations of some Sm-like proteins (e.g. ≈10–20 μM hexamer for *E. coli* Hfq;[44]); and (ii) polymerization occurs in the absence of RNA, and is therefore an intrinsic feature of these proteins (not an RNA-induced artefact). We have found that higher-order Hfq oligomers (possibly fibrillar) are capable of RNA-binding as assayed by electrophoretic mobility shift assays (EMSA) (unpublished data and Supplementary Figure 1), so any physiological roles for Hfq and Sm polymers are probably linked to their RNA-binding properties *in vivo*. Also in connection with putative *in vivo* functions, it is known that nucleoid-associated Hfq may influence DNA packaging by polymerizing onto regions of DNA rich in A-tracts.[5] Finally, we note that Sm-like proteins are not the sole example of RNA-binding proteins that polymerize into fibres; similar biochemical (RNA-binding) and biophysical (polymerization) properties have been reported for the HIV Rev protein,[45–47] which is thought to form helical polymers that function in coating viral nucleic acids. Further elucidation of questions regarding the *in vivo* existence of Hfq and Sm fibres, their potential biological activities, and mechanistic aspects of their polymerization (e.g. kinetics, reversibility) will significantly advance our preliminary state of knowledge about these higher-order assemblies of Hfq and Sm-like proteins.

Materials and Methods

Unless specified otherwise, all reagents were purchased from Sigma or Merck-Biochemicals.

Preparation of Hfq and SmAP fibres

E. coli Hfq was over-expressed as described,[28] and purified by affinity chromatography as follows: cells from post-induction cultures were resuspended in 10 ml of a buffer containing 20 mM Tris–HCl (pH 7.8), 0.5 M NaCl, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100, 10 μl DNase I (10 mg/ml), and a protease inhibitor cocktail (EDTA-free Complete mini, Roche Diagnostic) at 4 °C. The suspension was passed through a French press (1200 bar, 20,000 psi) and lysed cells were cleared by centrifugation at 15,000*g* for 30 min. The supernatant was heated at 80 °C for 15 min, followed by centrifugation at 15,000*g* for 30 min. Imidazole–HCl (pH 7.8) was added to a final concentration of 1 mM, and the resulting solution was applied to a 1 ml Ni^{2+}-NTA column (Qiagen). The resin was sequentially washed with approximately 15 column volumes of (i) 20 mM Tris–HCl (pH 7.8), 0.3 M NaCl, 20 mM imidazole, and then (ii) 50 mM sodium phosphate (pH 6.0), 0.3 M NaCl. The Hfq protein was then eluted in a buffer containing 20 mM Tris–HCl (pH 7.8), 0.3 M NaCl, 0.5 M imidazole. Hfq-containing fractions were dialyzed against poly(A) buffer (50 mM Tris–HCl (pH 7.5), 1 mM EDTA, 5% (v/v) glycerol) supplemented with 50 mM NH_4Cl. After centrifugation, the supernatant was applied to a 5 ml poly(A) Sepharose column (Amersham Biosciences), and the resin was washed with a poly(A) rinse buffer (poly(A) buffer +1.0 M NH_4Cl); Hfq was eluted in a gradient of the rinse buffer supplemented with 8.0 M urea. Hfq-containing fractions were then dialyzed against poly(A) buffer containing 50 mM NH_4Cl and 0.1% (v/v) Triton X-100, and the protein was stored at 4 °C (typically at 1–2 mg/ml). Fibrous Hfq was produced by dialysis of protein samples (at a concentration of ≈5 μM hexamer) into a buffer consisting of 5 mM Tris–HCl (pH 8), 5 mM NaCl, and 0.005% (w/v) dodecyl-β-D-maltoside (buffer HfqIR), followed by lyophilization and re-suspension in water.

The expression and purification of *Methanobacterium thermautotrophicum* strain ΔH (*Mth*) SmAP1 also has been described in detail;[27] fibrous *Mth* SmAP1 was obtained in the course of preparing 0.5 mg/ml (≈9 μM heptamer) protein samples in 10 mM Tris (pH 7.5), 60 mM NaCl for electron microscopy on carbon-coated parlodion film. Hfq was found to be soluble at concentrations of at least 30 mg/ml (0.4 mM hexamer) in most buffers, and, similarly, *Mth* SmAP1 could be concentrated to greater than 50 mg/ml (0.8 mM heptamer) without visible precipitation. Finally, we note that solution conditions for Hfq and SmAP fibrogenesis were not optimized in terms of yield or fibre quality/homogeneity.

Infrared spectroscopy

Fourier transform IR (FTIR) spectra of Hfq were recorded with a Perkin-Elmer system 2000 infrared spectrophotometer equipped with a multiple-reflection attenuated total reflectance (ATR) system. Samples of Hfq protein that had been dialyzed into buffer HfqIR were lyophilized and then dissolved in ^2H_2O (<99.8% purity; from Euriso-Top CEA) at a protein concentration of 15 mg/ml. An 18 μl sample of the protein solution was deposited under dry air on the ZnSe crystal of the ATR device (12 internal reflections) and covered with another ZnSe plate. Twenty-five scans were typically accumulated at a nominal resolution of 2 cm^{-1}, and buffer HfqIR spectra recorded under conditions identical with those of the protein sample were automatically subtracted. Data treatment consisted of two-point baseline correction, Fourier self-deconvolution, and second derivative calculation, and was performed with the Perkin-Elmer Spectrum software. Iterative least-squares curve-fittings of spectral components (modelled as Gaussian line shapes) were performed using Grams (Galactic Software). The method used to estimate the secondary structure content of Hfq was similar to that described,[48] in which the fraction of a particular secondary structural element is taken as the sum of integrated intensities of the corresponding Gaussian components (i.e. bands with maxima falling within the frequency range assigned to the structural element of interest) divided by the total area of all bands over the entire frequency range. Absorption bands were assigned as described by Krimm *et al.*[40] and Barth & Zscherp.[41] In brief, the band centered at 1653 cm^{-1} is due to the amide I vibrational mode (essentially the symmetric ←C=O→ stretching of backbone carbonyl amides), while the band centered at 1550 cm^{-1} is attributed to amide II vibrations (essentially a deformation mode resulting from a combination of amide N–H in-plane bending and C−N stretching motions). The fine structures of both bands are sensitive to local protein conformation, with the amide I band being most frequently used in determining protein secondary structure content.

Electron microscopy and image analysis

Aliquots of Hfq that had been analyzed by FTIR spectroscopy in order to confirm the presence of absorption bands characteristic of aggregated β-sheets ($\approx$1610 cm^{-1}; see Table 1) were diluted 100-fold, adsorbed to 300-mesh carbon-coated grids, and then stained with 1% (w/v) uranyl acetate. Samples were visualized in a Philips CM120 transmission electron microscope (Philips; Eindhoven, Netherlands) operating at an accelerating voltage of 120 kV. Twenty images were recorded at a nominal magnification of 60,000× and average defocus of −1300 nm by using a 1024 byte Gatan ssCCD. A total of 122 fibre-containing subimages (each of 256×256 pixels) were windowed from the original 1024×1024 dataset, and discrete, clearly defined fibrous specimens from each image were masked and registered to a horizontal reference before computing their Fourier transforms (FTs). Power spectral images (S(f)) were calculated from the complex modulus of these transforms, and such images were then rotationally aligned and classified into groups *via* an unsupervised self-organizing mapping algorithm.[49] In order to increase signal intensities of Bragg reflections, an averaged power spectrum ⟨S(f)⟩ was computed from the power spectra of 87 single images belonging to the major class, as proposed by Watts *et al.*[45] The helical pitch (P) and inter-subunit spacing (h) were then calculated from ⟨S(f)⟩ *via* the standard theory for diffraction from a helical (one-dimensional) point lattice.[38,50] An improved, averaged image was computed from sub-images of fibres corresponding to individual major-class images after rotational and translational alignment, and final helical parameters were determined from this final image. This averaged image was also used to estimate a helical radius (r) of $\approx$8.5 nm and a resolution value of 2.3 nm by using a spectral signal-to-noise criterion.[37] All steps of this single-particle image analysis procedure were performed within the Xmipp software package.[51]

Candidate 3D models of Hfq fibres were generated by positioning the X-ray crystallographic structure of hexameric *E. coli* Hfq (PDB code 1HK9[22]) into 3D locations, which increased the correlation between power spectra calculated from projection images of the model ($S(f)_{\mathrm{Model}}$) and the averaged experimentally observed spectra ($\langle S(f)_{\mathrm{Obs}}\rangle$). More specifically, positioning of Hfq hexamers and calculation of synthetic EM projection images (using the Spider software[52]) was followed by filtering of the synthetic images to the resolution limit computed from the experimental EM images. Computed power spectra of these filtered images ($S(f)^{\mathrm{filt}}_{\mathrm{Model}}$) were then compared to corresponding experimental spectra ($S(f)_{\mathrm{obs}}$), and the 3D model was manually refined by rigid-body transformations of the hexamer until more optimal agreement was obtained (i.e. minimization of the difference between $S(f)_{\mathrm{obs}}$ and $S(f)^{\mathrm{filt}}_{\mathrm{Model}}$ subject to the constraint of consistency with the helical parameters P, h, and r). This heuristic procedure was iteratively applied until only minimal model improvement was attained.

Model refinement and analysis

A final stage of model refinement included gross geometric/stereochemical correction of the crystallographically derived fibre structure (e.g. de-bumping atomic clashes), followed by more careful potential energy minimization to relax Hfq monomers into a conformation more likely to be adopted in the fibrillar state. The energy of the model was reduced *in vacuo* with respect to the Amber FF03 forcefield (12 Å cut-off for non-bonded interactions) over the course of >5000 steps of steepest descent (earlier cycles) and conjugate gradient (later cycles) minimization in Amber v8.0.[53] The starting structure consisted of three layers of the fibre (corresponding to 6×6×3=108 independent 1HK9 monomers), and a final multi-layered fibre model was built by superimposing the relaxed/minimized middle layer upon adjacent layers in both directions along the helical axis (rather than imposing one-dimensional periodic boundary conditions along the fibre axis). Implicit in this method is the assumption that only interactions within ±1 layer are energetically significant (layers are separated by >50 Å, so this is likely to be a valid approach).

A 3D model for SmAP fibres was derived by two approaches: (i) using the quasi-hexagonal packing of *Mth* SmAP1 heptamers in the $P2_12_12_1$ crystal form described by Mura *et al.*,[27] and (ii) by applying the same EM and image analysis-based approach described above for Hfq. The only difference in the case of SmAP1 was that starting images used to construct this model were extracted from the electron micrographs of *Mth* SmAP1 described by Mura *et al.*[27] by digitization using a LeafScan 45 linear CCD scanner with a 25 μm window spacing.

The molecular modelling toolkit[54] and Python scripts were utilized for coordinate transformations and manipulations involving large numbers of subunits (e.g. a single fibre layer of 36 Hfq monomers), and the programs Pymol† and VMD[55] were utilized along with their respective Python and Tcl application programming interfaces in order to compute and analyze the crystal lattices and intermolecular contact distances used in creating Figure 1. Pymol was also used to generate all Figures involving illustrations of protein structures (i.e. Figures 1, 3 and 4).

Acknowledgements

We gratefully acknowledge the CNRS (UPR 9073) and University Denis Diderot-Paris 7 for financial support, as well as funding from an Alfred P. Sloan/DOE postdoctoral fellowship (C.M.) and a Conacyt doctoral fellowship (M.R.G.). We also thank the Institut Curie (Paris) for providing EM and image analysis facilities, Dr Claudine Mayer (U. Paris 6) for her assistance in constructing models of Hfq fibres, M. Folichon (IBPC) for EMSA work, and Professor Eliane Taillandier (U. Paris 13) for constant interest in this work. Drs Linda Columbus (Scripps), David Eisenberg (UCLA), and Andy McCammon (UCSD) are thanked for reviewing the manuscript.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.11.010

† http://pymol.sourceforge.net

References

1. Franze de Fernandez, M. T., Hayward, W. S. & August, J. T. (1972). Bacterial proteins required for replication of phage Qβ ribonucleic acid. *J. Biol. Chem.* **247**, 821–824.
2. Valentin-Hansen, P., Eriksen, M. & Udesen, C. (2004). The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.* **51**, 1525–1533.
3. Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P. & Hajnsdorf, E. (2003). The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucl. Acids Res.* **31**, 7302–7310.
4. Azam, T. A. & Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from *Escherichia coli*. *J. Biol. Chem.* **274**, 33105–33113.
5. Tolstorukov, M. Y., Virnik, K. M., Adhya, S. & Zhurkin, V. B. (2005). A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucl. Acids Res*, **33**, 3907–3918.
6. Storz, G. (2002). An expanding universe of noncoding RNAs. *Science*, **296**, 1260–1263.
7. Tsui, H.-C. T., Leung, H.-C. E. & Winkler, M. E. (1994). Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol. Microbiol.* **13**, 35–49.
8. Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* **66**, 373–395.
9. Masse, E., Majdalani, N. & Gottesman, S. (2003). Regulatory roles for small RNA in bacteria. *Curr. Opin. Microbiol.* **6**, 120–124.
10. Vytvytska, O., Jakobsen, J. S., Balcunate, G., Andersen, J. S., Baccarini, M., von Gabain, A. & Host-factor, I. (1998). Hfq, binds to *Escherichia coli ompA* mRNA in a growth rate-dependent fashion and regulates its stability. *Proc. Natl Acad. Sci. USA*, **95**, 14118–14123.
11. Geissmann, T. A. & Touati, D. (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J.* **23**, 396–405.
12. Sukhodolets, M. V. & Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry*, **42**, 8022–8034.
13. Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V. R. & Blasi, U. (2003). Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA*, **9**, 1308–1314.
14. Will, C. & Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301.
15. Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D. & Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. *Nature*, **404**, 515–518.
16. Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L. & Cech, T. R. (1999). Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature*, **401**, 177–180.
17. Arluison, V., Derreumaux, P., Allemand, F., Folichon, M., Hajnsdorf, E. & Regnier, P. (2002). Structural modelling of the Sm-like protein Hfq from *Escherichia coli*. *J. Mol. Biol.* **320**, 705–712.
18. Sun, X., Zhulin, I. & Wartell, R. M. (2002). Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucl. Acids Res.* **30**, 3662–3671.
19. Moller, T., Franch, T., Hojrup, P., Keene, D. R., Bachinger, H. P., Brennan, R. G. & Valentin-Hansen, P. (2002). Hfq. A bacterial Sm-like protein that mediates RNA–RNA interaction. *Mol. Cell.* **9**, 23–30.
20. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C. & Storz, G. (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell.* **9**, 11–22.
21. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P. & Brennan, R. G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.* **21**, 3546–3556.
22. Sauter, C., Basquin, J. & Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucl. Acids Res.* **31**, 4091–4098.
23. Nikulin, A., Stolboushkina, E., Perederina, A., Vassilieva, I., Blaesi, U., Moll, I. *et al*. (2005). Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallog. sect. D*, **61**, 141–146.
24. Moll, I., Leitsch, D., Steinhauser, T. & Blasi, U. (2003). RNA chaperone activity of the Sm-like Hfq protein. *EMBO Rep.* **4**, 284–289.
25. Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B. & Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.* **20**, 2293–2303.
26. Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D. & Feig, A. L. (2004). *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nature Struct. Mol. Biol.* **11**, 1206–1214.
27. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M. & Eisenberg, D. (2003). The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). *Protein Sci.* **12**, 832–847.
28. Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J. *et al*. (2004). The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.* **271**, 1258–1265.
29. Jackson, M. & Mantsch, H. H. (1995). The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30**, 95–120.
30. Tatulian, S. A. (2003). Attenuated total reflection Fourier transform infrared spectroscopy: a method of choice for studying membrane proteins and lipids. *Biochemistry*, **42**, 11898–11907.
31. Fandrich, M. & Dobson, C. M. (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682–5690.
32. Arrondo, J. L., Muga, A., Castresana, J. & Goni, F. M. (1993). Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **59**, 23–56.
33. Chirgadze, Y. N., Fedorov, O. V. & Trushina, N. P. (1975). Estimation of amino acid residue side-chain absorption in the infrared spectra of protein solutions in heavy water. *Biopolymers*, **14**, 679–694.
34. Bendit, E. G. (1967). Infrared absorption of tyrosine side chains in proteins. *Biopolymers*, **5**, 4525–4533.
35. Kubelka, J. & Keiderling, T. A. (2001). Differentiation of beta-sheet-forming structures: ab initio-based simulations of IR absorption and vibrational CD for model peptide and protein beta-sheets. *J. Am. Chem. Soc.* **123**, 12048–12058.

36. Kubelka, J. & Keiderling, T. A. (2001). The anomalous infrared amide I intensity distribution in (13)C isotopically labeled peptide beta-sheets comes from extended, multiple-stranded structures: an *ab initio* study. *J. Am. Chem. Soc.* **123**, 6142–6150.
37. Unser, M., Trus, B. L. & Steven, A. C. (1987). A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, **23**, 39–51.
38. Klug, A., Crick, F. H. C. & Wyckoff, H. W. (1958). Diffraction by helical structures. *Acta Crystallog.* **11**, 199–213.
39. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry*, Freeman Co., San Francisco, CA.
40. Krimm, S. & Bandekar, J. (1986). Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Advan. Protein Chem.* **38**, 181–364.
41. Barth, A. & Zscherp, C. (2002). What vibrations tell us about proteins. *Quart. Rev. Biophys.* **35**, 369–430.
42. Ohnishi, S. & Takano, K. (2004). Amyloid fibrils from the viewpoint of protein folding. *Cell Mol. Life Sci.* **61**, 511–524.
43. Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R. & Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, **435**, 773–778.
44. Kajitani, M., Kato, A., Wada, A., Inokuchi, Y. & Ishihama, A. (1994). Regulation of the *Escherichia coli* hfq gene encoding the host factor for phage Q beta. *J. Bacteriol.* **176**, 531–534.
45. Watts, N. R., Misra, M., Wingfield, P. T., Stahl, S. J., Cheng, N., Trus, B. L. *et al*. (1998). Three-dimensional structure of HIV-1 Rev protein filaments. *J. Struct. Biol.* **121**, 41–52.
46. Jain, C. & Belasco, J. G. (2001). Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell.* **7**, 603–614.
47. Blanco, F. J., Hess, S., Pannell, L. K., Rizzo, N. W. & Tycko, R. (2001). Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. *J. Mol. Biol.* **313**, 845–859.
48. Cabiaux, V., Brasseur, R., Wattiez, R., Falmagne, P., Ruysschaert, J. M. & Goormaghtigh, E. (1989). Secondary structure of diphtheria toxin and its fragments interacting with acidic liposomes studied by polarized infrared spectroscopy. *J. Biol. Chem.* **264**, 4928–4938.
49. Marabini, R. & Carazo, J. M. (1994). Pattern recognition and classification of images of biological macromolecules using artificial neural networks. *Biophys. J.* **66**, 1804–1814.
50. Sherwood, D. (1976). *Crystals, X ray and Proteins*, Longman Group, London.
51. Marabini, R., Masegosa, I. M., San Martin, C., Marco, S., Fernandez, J. J., de la Fraga, L. G. *et al*. (1996). Xmipp: an image processing package for electron microscopy. *J. Struct. Biol.* **116**, 237–240.
52. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). SPIDER and WEB processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199.
53. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E. *et al*. (2004). *AMBER v8*, University of California, San Francisco.
54. Hinsen, K. (2000). The molecular modelling toolkit: a new approach to molecular simulations. *J. Comput. Chem.* **21**, 79–85.
55. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD–visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.

Edited by I. B. Holland

(Received 26 August 2005; received in revised form 2 November 2005; accepted 3 November 2005)
Available online 22 November 2005