Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Hfq binds ribonucleotides in three different RNA-binding sites

Victoria Murina,[a] Natalia Lekontseva[a,b] and Alexey Nikulin[a]*

[a]Institute of Protein Research, RAS, Institutskaya 4, Pushchino 142290, Moscow Region, Russian Federation, and [b]Udmurt State University, Universitetskaya str. 1, Izhevsk 426034, Russian Federation

Correspondence e-mail: nikulin@vega.protres.ru

Received 15 February 2013
Accepted 12 April 2013

PDB References: Hfq complexes, 3qui; 4j5y; 4j6w; 4j6x; 4j6y

The Hfq protein forms a doughnut-shaped homohexamer that possesses RNA-binding activity. There are two distinct RNA-binding surfaces located on the proximal and the distal sides of the hexamer. The proximal side is involved in the binding of mRNA and small noncoding RNAs (sRNAs), while the distal side has an affinity for A-rich RNA sequences. In this work, the ability of various ribonucleotides to form complexes with Hfq from *Pseudomonas aeruginosa* has been tested using X-ray crystallography. ATP and ADPNP have been located in the distal R-site, which is a site for poly(A) RNA binding. UTP has been found in the so-called lateral RNA-binding site at the proximal surface. CTP has been found in both the distal R-site and the proximal U-binding site. GTP did not form a complex with Hfq under the conditions tested. The results have demonstrated the power of the crystallographic method for locating ribonucleotides and predicting single-stranded RNA-binding sites on the protein surface.

1. Introduction

Hfq is a bacterial RNA-binding protein that plays key roles in the control of gene expression (for a recent review, see Vogel & Luisi, 2011). By facilitating the pairing of small noncoding RNAs (sRNAs) with their target mRNAs, Hfq affects the translation and turnover rates of specific transcripts and contributes to complex post-transcriptional networks (Møller *et al.*, 2002; Vassilieva *et al.*, 2002; Valentin-Hansen *et al.*, 2004; Aiba, 2007). Moreover, Hfq plays an important role in modulating mRNA degradation and possibly in RNA transcription (Sledjeski *et al.*, 2001; Basineni *et al.*, 2009; Soper *et al.*, 2010). These functions of Hfq have been attributed to its ring-like quaternary hexamer structure, which contains two non-equivalent binding surfaces that are capable of multiple interactions with RNA molecules (Fig. 1). To discriminate between them, the term 'proximal side' is usually used for the surface on which the N-terminal α-helix is located. The opposite side of the hexamer is named the 'distal side'. In the absence of strict sequence specificity, Hfq has a preference for A-rich and U-rich single-stranded RNA (ssRNA; Brennan & Link, 2007).

The crystal structure of Hfq from *Staphylococcus aureus* (SauHfq) in complex with AU_5G RNA (Schumacher *et al.*, 2002) and the recent structure of Hfq from *Salmonella typhimurium* (StyHfq) in complex with U_6 RNA (Sauer & Weichenrieder, 2011) have revealed that U-rich RNA interacts with the amino-acid residues of the proximal side. Every uracil is symmetry located in the repeating binding pockets between two neighbouring protein monomers (protomers) close to the highly conserved residues Gln8, Phe/Tyr42 and His57 (*Escherichia coli* Hfq numbering).

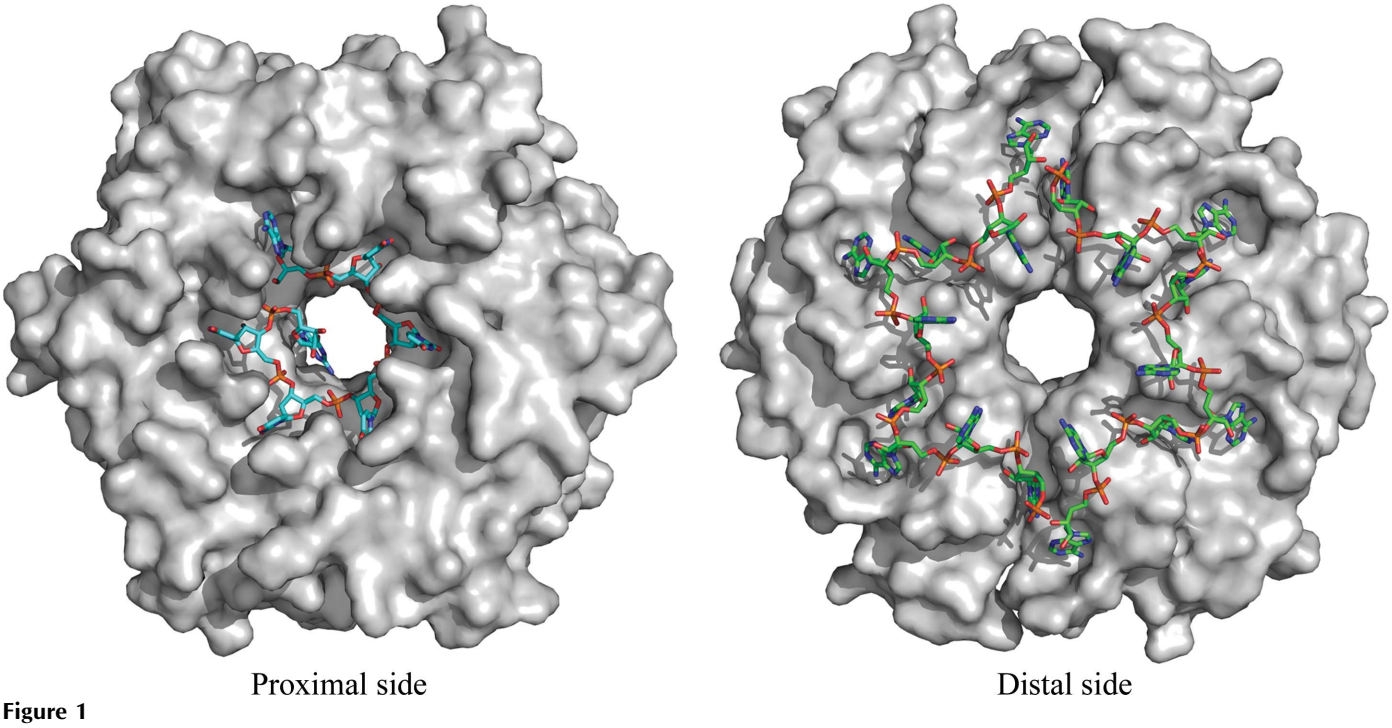

Figure 1
The Hfq hexamer surface: left, proximal side view; right, distal side view. The left figure shows Hfq from *S. aureus* in complex with AU_5G RNA (PDB entry 1kq2; Schumacher *et al.*, 2002). The right figure shows Hfq from *E. coli* in complex with poly(A) RNA (PDB entry 3gib; Link *et al.*, 2009). The proteins are represented by the solvent-accessible surface. The corresponding RNAs are represented by sticks. The figures were produced with *PyMOL* (DeLano, 2002).

The crystal structure of Hfq from *E. coli* (EcoHfq) in complex with poly(A) RNA demonstrated that poly(A) RNA binds to the distal side of the protein (Link *et al.*, 2009). In contrast to the compact uridine-binding site, the poly(A)-binding area of Hfq consists of three consecutive RNA-binding sites called A, R and E. The adenosine-specific A-site is composed of residues from β-strands 2 and 4. It has been suggested that matching between the protein residues and the adenine ensures the specificity of the site and concomitant discrimination against guanine (Link *et al.*, 2009). The second adenine-binding site (R-site) is located between the β-sheets of two neighbouring Hfq subunits in the rather deep cavity organized by Tyr25 and several hydrophobic residues. On the basis of the protein-atom arrangement it was also suggested that this site possesses purine-nucleotide selectivity. The third adenine exits above the surface of the distal face (E-site or N-site) and makes no contact with the protein. The structure of Hfq from *Bacillus subtilis* (BsuHfq) in complex with the SELEX-derived RNA aptamer $(AG)_3A$ was subsequently determined (Someya *et al.*, 2012). It was found that BsuHfq did not bind AU_5G and had a weaker affinity for A_{18} RNA compared with EcoHfq, while it had the highest affinity for AGAGAGA repeats. In the BsHfq–$(AG)_3A$ complex each adenine was symmetrically located in the previously proposed R-site and the guanines were located on the protein surface. No adenines were found in the A-sites.

In addition to its role in poly(A) binding, the R-site was proposed to be a binding pocket for ATP, as residue Tyr25 has been implicated in such binding (Arluison *et al.*, 2007; Lazar *et al.*, 2010). Nevertheless, Hfq does not utilize ATP to perform its RNA chaperone activity (Hämmerle *et al.*, 2012). The recent structures of Hfq from *E. coli* in complex with AU_6A and ADP (Wang *et al.*, 2011), and with ATP (Hämmerle *et al.*, 2012) have demonstrated similar binding properties for the ADP and ATP found in the R-site. The identification of the binding site for the adenine of RNA and individual ATP (and ADP) molecules have stimulated us to verify the ability of Hfq to bind other ribonucleotides.

In this article, we present the results of the structure determination of Hfq from *Pseudomonas aeruginosa* in complex with ADPNP, ATP, CTP and UTP. Soaking of the Hfq crystals in appropriate ribonucleotide solutions was used to obtain the nucleotide–protein complexes. The quality of the crystals was sufficiently high that a rotating-anode X-ray source could be used to collect high-resolution diffraction data in most cases. The results obtained reveal differences in the affinity of Hfq for various ribonucleotides and the potential of the protein–ribonucleotide complexes for the identification of the protein RNA-binding sites.

2. Methods

2.1. Protein purification and crystallization

Hfq from *P. aeruginosa* (PaeHfq) was obtained and purified as described previously (Nikulin *et al.*, 2005). Hfq crystals were obtained by the hanging-drop vapour-diffusion technique at 295 K. All drops were set up by mixing one volume of the protein stock solution (30 mg ml^{-1} protein in 300 m*M* NaCl, 200 m*M* ammonium sulfate, 50 m*M* Tris–HCl pH 8.0) with one

Table 1
Data-collection and refinement statistics.

Throughout refinement, 5% of the total reflections were kept aside for the calculation of R_{free}. Values in parentheses are for the highest resolution shell.

Complex	Hfq–ADPNP (soaking)	Hfq–ATP (soaking)	Hfq–CTP (soaking)	Hfq–UTP (soaking)	Hfq–GTP (cocrystallization)
PDB code	3qui	4j5y	4j6w	4j6x	4j6y
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 61.08, b = 75.86, c = 93.44, $\alpha = \beta = \gamma$ = 90	a = 61.45, b = 73.68, c = 107.56, $\alpha = \beta = \gamma$ = 90	a = 61.28, b = 75.55, c = 109.61, $\alpha = \beta = \gamma$ = 90	a = 61.58, b = 73.52, c = 107.79, $\alpha = \beta = \gamma$ = 90	a = 61.56, b = 72.86, c = 107.28, $\alpha = \beta = \gamma$ = 90
Crystallographic data					
Source	BESSY 14.1	X8 PROTEUM	BESSY 14.1	X8 PROTEUM	X8 PROTEUM
Wavelength (Å)	0.918410	1.5418	0.918410	1.5418	1.5418
Resolution range (Å)	50–1.93 (2.04–1.93)	20–2.095 (2.21–2.095)	50–1.70 (1.79–1.70)	25–2.22 (2.34–2.22)	15–2.14 (2.26–2.14)
No. of observations	95323 (5931)	112683 (14016)	188847 (17848)	101962 (12854)	156587 (20184)
No. of unique reflections	29725 (3127)	28097 (3877)	54468 (6883)	24556 (3419)	25623 (3272)
R_{merge} (%)	7.6 (45.1)	9.1 (33.9)	5.8 (37.9)	9.8 (38.4)	7.9 (38.5)
Mean $I/\sigma(I)$	9.44 (1.97)	13.02 (3.33)	11.56 (2.77)	13.12 (3.12)	16.23 (5.53)
Completeness (%)	89.8 (66.2)	95.7 (88.44)	97.78 (81.93)	99.18 (94.01)	94.06 (79.62)
Multiplicity	3.2 (1.9)	4.0 (3.6)	3.5 (2.6)	4.1 (3.8)	6.1 (6.2)
Wilson B factor (Å^2)	21.2	22.7	19.1	24.8	28.2
Refinement					
Resolution range (Å)	50–1.93 (2.04–1.93)	20–2.095 (2.17–2.095)	50–1.80 (1.84–1.80)	25–2.22 (2.34–2.22)	15–2.14 (2.23–2.14)
R_{work} (%)	21.3 (29.3)	18.4 (25.5)	19.3 (36.5)	18.3 (28.5)	17.1 (19.8)
R_{free} (%)	28.7 (33.5)	25.2 (32.2)	24.1 (42.3)	25.6 (36.6)	21.8 (24.8)
No. of nonsolvent atoms	3478	3476	3571	3413	3226
No. of solvent atoms	210	285	298	290	129
R.m.s.d.s from ideal geometry					
Bond lengths (Å)	0.007	0.008	0.009	0.007	0.008
Bond angles (°)	1.12	1.30	1.48	1.30	1.15
Chirality angles (°)	0.074	0.078	0.119	0.085	0.079
Planarity angles (°)	0.004	0.006	0.006	0.005	0.005
Dihedral angles (°)	16.51	15.6	19.13	22.08	13.9
Average B factors (Å^2)	30.4	17.5	25.2	12.6	28.8
Ramachandran plot					
Favoured (%)	94.0	95.2	94.9	93.6	95.9
Allowed (%)	4.5	3.0	3.6	5.4	2.8
Outliers (%)	1.5	1.8	1.5	1.0	1.3

volume of reservoir solution (400 m*M* ammonium chloride, 24% PEG 4000, 100 m*M* Tris–HCl pH 8.5). The crystals appeared within one week and reached maximum dimensions of 300 × 100 × 50 µm within one month.

2.2. Obtaining Hfq–ribonucleotide crystals and data collection

The Hfq–ribonucleotide crystals were obtained in two different ways: by cocrystallization of the protein with a ribonucleotide or by soaking the Hfq crystals in a ribonucleotide solution.

For crystallization of Hfq in the presence of GTP, the crystallization drops were prepared by mixing 2.0 µl protein stock solution (30 mg ml^{-1}), 0.2 µl ribonucleotide stock solution (100 m*M*) and 2.0 µl reservoir solution. The final concentration of the nucleotide was 4.75 m*M*. The crystals appeared within one week and reached maximum dimensions of 300 × 100 × 50 µm within one month. Before cooling, the crystals were transferred into 15% PEG 4000, 15% PEG 400, 200 m*M* ammonium chloride, 50 m*M* Tris–HCl pH 8.5.

For soaking in ADPNP, ATP, CTP, UTP or GTP solutions, the Hfq crystals were placed into 2 µl of a solution consisting of 50 m*M* of the appropriate ribonucleotide, 20% PEG 400, 50 m*M* Tris–HCl pH 8.5. Since the composition of the soaking solutions was optimized for crystal cooling, the crystals were used for data collection without any additional procedure after incubation.

X-ray diffraction data were collected from the crystals using an X8 PROTEUM diffractometer (Bruker AXS) equipped with a Bruker PLATINUM 135 CCD detector or on beamline 14.1 at BESSY, HZB, Berlin, Germany using a MAR CCD 225 detector and were processed using the *XDS* program (Kabsch, 2010) in combination with *SCALA* (*CCP*4; Winn *et al.*, 2011) or *SADABS* (Bruker AXS) for merging raw data. Detailed data-collection statistics are shown in Table 1.

2.3. Refinement of the structures

The position of Hfq was easily determined by *Phaser* using PaeHfq as a search model and was refined using the *phenix.refine* procedure from the *PHENIX* suite (Adams *et al.* 2010). Model rebuilding was carried out using the program *Coot* (Emsley *et al.*, 2010). The positions of the nucleotides were usually found by the 'ligand search' procedure implemented in *Coot* (Emsley & Cowtan, 2004) and were checked using the $F_o - F_c$ map. After several cycles of simulated-annealing refinement, water molecules were added to the models. The final refinement statistics for all of the obtained complexes are given in Table 1.

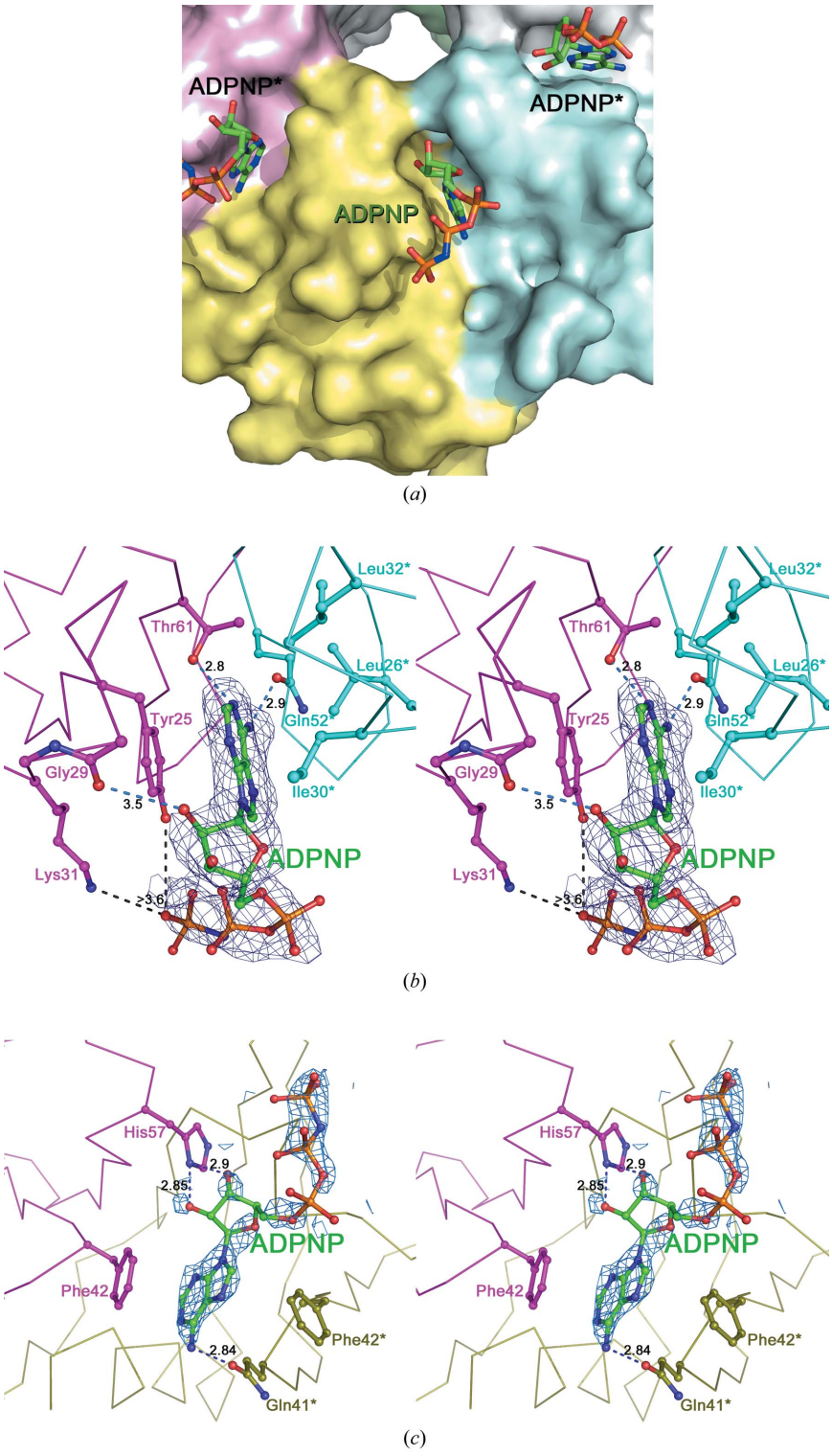

Figure 2
The structure of the PaeHfq–ADPNP complex. (*a*) The distal side of PaeHfq. The protein is represented by the solvent-accessible surface. Hfq monomers are indicated in different colours. ADPNP molecules are represented as sticks. (*b*) A stereoview of an ADPNP molecule in the R-site. This site is formed by residue Tyr25 of one protein monomer (magenta) and by residues Leu26, Ile30 and Leu32 of the adjacent monomer (cyan). Hydrogen bonds are shown as dotted lines. (*c*) A stereoview of an ADPNP molecule in the central part of PaeHfq (the U-binding site). It is formed by residues Phe42 of the two neighbouring protein monomers. Hydrogen bonds are shown as dotted lines. The $2F_o - F_c$ map showing the electron density for ADPNP molecules in the R-site (*b*) and in the U-binding site (*c*) contoured at the 1.0σ level.

3. Results

Most of the Hfq–nucleotide complexes were obtained by incubation of PaeHfq crystals in stabilization solution containing the appropriate nucleotide at a concentration of about 50 m*M*. The soaking method has been widely used to prepare heavy-atom derivatives for protein crystallography and to obtain protein–ligand complexes. Our results demonstrate that the soaking method can easily be applied to directly obtain complexes of a protein with ribonucleotides in the crystal. In all of the solved complexes the structure of Hfq did not change, and the r.m.s.d. between C^{α} atoms (residues 7–67) of all of the complexes was not greater than 0.38 Å when compared with the isolated PaeHfq. The N- and C-terminal amino-acid residues of the protein were partly traced owing to their high flexibility. The Asp40 and Asn48 residues were usually outside the expected Ramachandran region, as found in other Hfq protein structures.

3.1. Structures of the PaeHfq–ADPNP and PaeHfq–ATP complexes

Since it was postulated that Hfq possesses ATPase activity, ADPNP, a noncleavable analogue of ATP, was initially used. The PaeHfq–ADPNP complex was obtained by soaking the protein crystals in an ADPNP solution for 4 h. In addition, the PaeHfq–ATP complex was obtained by soaking the protein crystals in an ATP solution for 16 h.

The positions and orientations of the adenines in both the PaeHfq–ADPNP and the PaeHfq–ATP complexes were equivalent to those of the R-site in the EcoHfq–poly(A) RNA complex (Link *et al.*, 2009). In PaeHfq–ADPNP, six nucleotides were found in the pockets between each two neighbouring protein monomers (protomers; Fig. 2). Each pocket is formed by residue Tyr25 of one protomer and residues Leu26*, Ile30* and Leu32* of the second protomer (where the asterisk indicates residues from the neighbouring protomer) (Fig. 2*b*). The adenine was stacked with Tyr25. The N6 and N1 adenine atoms formed hydrogen bonds to the side-chain O atoms of Gln52* and Thr61, respectively. The ribosyl O2′ atom formed a hydrogen bond to the carbonyl O atom of residue Gly29, thus contributing to the preference for RNA over DNA. The phosphate group of the ribonucleotides was very flexible and

did not have a specific position in the complex. The full-sized adenosines were traced in three binding sites, but the γ-phosphate only formed hydrogen bonds to the hydroxyl group of Tyr25 and the NE atom of Lys31 in two positions. It is important to note that the 5′-triphosphate was not identified in the previously proposed cleft (Arluison *et al.*, 2007; Lazar *et al.*, 2010).

An additional ADPNP molecule was found at the centre of the PaeHfq hexamer between the Phe42 residues of two adjacent protomers (Fig. 2*c*). The position of the adenine was fixed by the formation of hydrogen bonds from its N6 atom to the OE1 atom of Gln41* and from the O2′ and O3′ atoms of the sugar to the side-chain N atom of His57. The occupancy of the ADPNP molecules was 0.43 as refined by the *phenix.refine* procedure. This location of the ADPNP complied with the positions of the 5′-adenine of AU_5G in the complex with SauHfq (Schumacher *et al.*, 2002) and of the 3′-adenine of AU_6A in the complex with EcoHfq (Wang *et al.*, 2011) and represents a possible position of the adenine in the internal U-binding site.

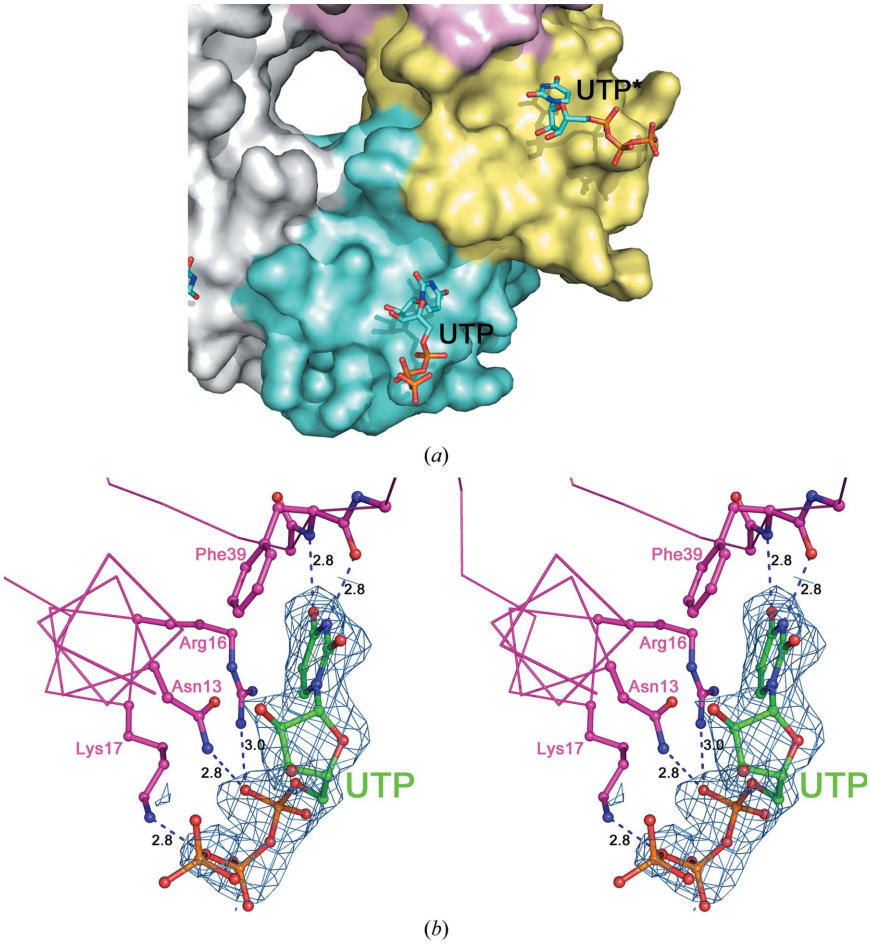

Figure 3
The structure of the PaeHfq–UTP complex. (*a*) The proximal side of PaeHfq. The protein is represented by the solvent-accessible surface. Hfq monomers are indicated in different colours. UTP molecules are represented as sticks. (*b*) A stereoview of a UTP molecule in the PaeHfq cleft close to residue Phe39. Hydrogen bonds are shown by dotted lines. A fragment of the $2F_o - F_c$ map showing the electron density for UTP is contoured at the 1.0σ level.

In the PaeHfq–ATP complex six nucleotides were arranged in the R-sites. The contacts between the adenine and the protein were identical to those in the PaeHfq–ADPNP complex. The main difference between them is that the 5′-phosphates of ATP were located closer to the OH group of Tyr25 and the NE atom of Lys31, thus providing better contacts with the protein. Another difference is that there was no ATP in the centre of PaeHfq despite the comparable conditions used for complex formation.

3.2. Structure of the PaeHfq–UTP complex

The PaeHfq–UTP complex was obtained by soaking the protein crystals in 50 m*M* UTP solution for 4 h. Six UTP molecules were located on the Hfq lateral surface close to the β2 strand of each protomer (Fig. 3). The O4 and N3 atoms of the uracil formed hydrogen bonds to the amide N and the carboxyl O atoms of Phe39, respectively. The hydrogen bonds between the peptide backbone and the nucleotide base atoms ensured the uridine specificity and the concomitant discrimination against cytidine. An additional PaeHfq–UTP interaction included base stacking against the side chain of Phe39. Both the O2′ and O3′ sugar atoms made no contacts with protein atoms. The triphosphate formed hydrogen bonds to the Asn13, Arg16 and Lys17 side chains, firmly attaching the 5′-phosphate group to the N-terminal α-helix.

3.3. Structure of the PaeHfq–CTP complex

The PaeHfq–CTP complex was obtained by soaking the protein crystals in 50 m*M* CTP solution for 4 h. Six CTP molecules were found in the classical poly(U) binding sites (Figs. 4*a* and 4*c*). They were located in the binding pockets between two neighbouring protomers. The cytosine position was defined by its stacking with the aromatic ring of Phe42 and the main-chain carboxyl O atom of Gln41*. The cytosine O2 atom formed a hydrogen bond to the NE atom of Gln8*. The distance between the cytosine N4 amino group and the OE atom of Gln41* varied from 3.0 to 3.8 Å for the different ribonucleotides. The O2′ and O3′ ribose atoms competed for hydrogen-bond formation with the His57 side chain, with the O3′–ND1 contact being prevalent.

Five additional CTP molecules were identified in the adenine-binding R-site of the protein (Figs. 4*b* and 4*d*). The sixth R-site had negligible electron density, so CTP was not placed there. The cytosine made contacts with the protein through two

water molecules. The first water molecule formed a bridge between the N4 atom of cytosine, Thr61 OG1 and Gln52* OE1. The second water molecule was the centre of contacts between the O2 and N3 atoms of the nucleotide, Ser60 OG and the carboxyl O atom of Leu26. Thr61 OG1 and Asn28* ND2 were located at about 3 Å from the N4 and O2 atoms of cytosine, so they also could participate in hydrogen bonds. The contact of the Tyr25 hydroxyl groups with the β-phosphate O atom restricted the flexibility of the 5′-phosphate tail, but the γ-phosphates made no contacts with the protein and their positions were therefore completely random.

3.4. PaeHfq did not bind GTP

Several attempts were made to obtain a complex of PaeHfq with GTP by soaking and by cocrystallization of the protein with GTP. Finally, two data sets were collected to resolutions of 2.45 Å (soaking) and 2.14 Å (cocrystallization). The data-collection statistics for the data set with higher resolution and its structure determination are given in Table 1. No GTP molecules were found in either of the crystals. The largest nonprotein electron density could only be assigned to ions.

4. Discussion

4.1. The Hfq distal R-site is the universal nucleotide-binding site

The present structures of PaeHfq–ADPNP and PaeHfq–ATP confirm the recognition features of the Hfq R-site. The R-site is a universal site for adenine binding, as has been demonstrated in the structures of the EcoHfq–A_{15} (Link *et al.*, 2009), EcoHfq–ADP (Wang *et al.*, 2011), BsuHfq–$(AG)_3A$ (Someya *et al.*, 2012) and SauHfq–A_7 (Horstmann *et al.*, 2012) complexes. The R-site is formed by several conserved amino-acid residues which form the hydrophobic pocket, thus ensuring specific contacts with the adenine (Fig. 2 and Table 2). One side of the hydrophobic pocket is formed by the highly conserved Tyr25 (or a phenylalanine in BsuHfq and SauHfq) which is stacked with the nucleotide base. The other side of the pocket is formed by Leu26, Ile30 and Leu32 of the neighbouring protomer (Fig. 5). The last two residues are not strongly conserved and could be substituted by valine, phenylalanine or even tyrosine and methionine depending on the organism. Nevertheless, these substitutions did not disturb the hydrophobic nature of the pocket.

Several conserved amino-acid residues lead to specific recognition of the nucleotide base. In EcoHfq and PaeHfq, Glu52 and Thr61 form hydrogen bonds to atoms N6 and N1 of the adenine, respectively (Table 2). The nearby Ser60 has the potential to form an alternative hydrogen bond to the

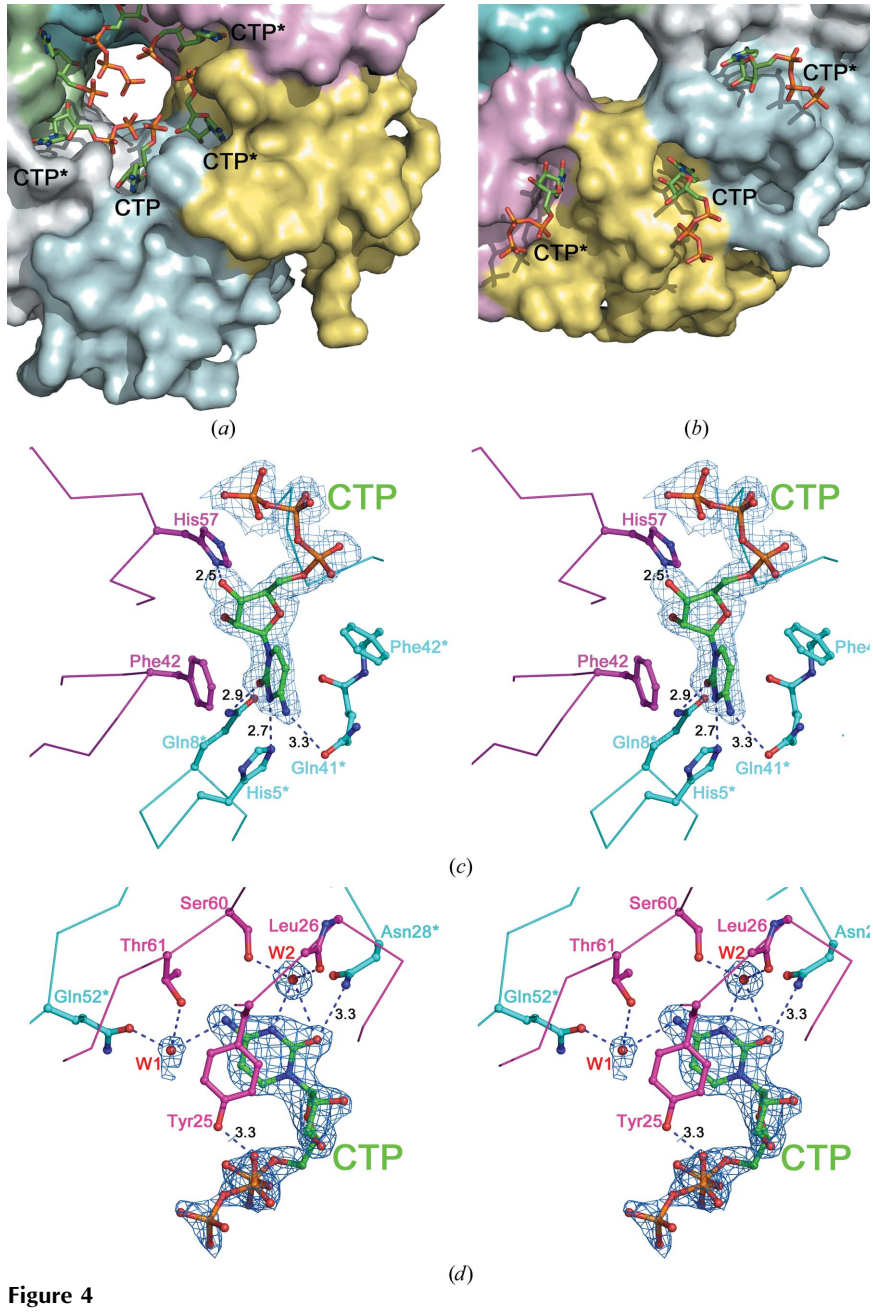

Figure 4
The structure of the PaeHfq–CTP complex. (*a*) The proximal side of PaeHfq. The protein is represented by the solvent-accessible surface. Hfq monomers are indicated in different colours. CTP molecules are represented by sticks. (*b*) The distal side of PaeHfq. (*c*) A stereoview of a CTP molecule in the proximal PaeHfq cleft (U-binding site). Hydrogen bonds are shown as dotted lines. A fragment of the $2F_o - F_c$ map showing the electron density for CTP. The contours are drawn at the 1.5σ level. (*d*) A stereoview of a CTP molecule in the distal PaeHfq cleft (R-site). A fragment of the $2F_o - F_c$ map at the 1.5σ level shows the electron density for CTP and the two adjacent water molecules. The two water molecules are represented as red spheres.

nucleotide base that is realised in the BsuHfq–$(AG)_3A$ and SauHfq–A_7 complexes, and leads to a rotation of the adenine by about 15° in the direction perpendicular to the base plane. In SauHfq the substitutions of Gln52 for histidine, Leu32 for methionine and Ile30 for phenylalanine promote displacement of the adenine (Horstmann *et al.*, 2012). This movement is stabilized by reforming the hydrogen bonds between the adenine and the protein: N6 switches to Thr62 (Thr61 in EcoHfq), N1 switches to Ser61 (Ser62 in EcoHfq) and Asn28 forms a new hydrogen bond to either the N3 or the O4′ atom of adenine (Someya *et al.*, 2012). The main reason for the adenine displacement in the BsuHfq–$(AG)_3A$ complex could be the substitution of Ile30 for phenylalanine, which promotes movement of the whole nucleotide to overcome steric hindrance. The relocation of the adenine is stabilized by the same rearrangements of the hydrogen bonds as in the SauHfq–A_7 complex. As found previously, the R-site ribosyl O2′ atom forms a hydrogen bond to the carbonyl O atom of Gly29, thus contributing to the preference of Hfq for RNA over DNA (Link *et al.*, 2009; Someya *et al.*, 2012; Horstmann *et al.*, 2012).

It has previously been suggested that the R-site in EcoHfq could also bind guanine (Link *et al.*, 2009). Indeed, the shape of the R-site cleft and the arrangement of the polar amino acids that contact the nucleotide base are favourable to accommodate both adenine and guanine. In the latter case only a simple rotation of the Gln52 side chain is required to form an efficient hydrogen bond from Glu52 NE2 to guanine O6 (Fig. 5*b*). This rotation is possible since there is no hydrogen bond between the Gln52 side chain and the rest of the protein. Moreover, the N2 atom of guanine could form an additional hydrogen bond to the hydroxyl group of Ser60 and thus additionally stabilize guanine binding. Nevertheless, a PaeHfq–GTP complex was not obtained either by soaking PaeHfq crystals in GTP solution or by cocrystallization of the protein and GTP.

An explanation of the preference for adenine over guanine has been suggested when discussing the SauHfq R-site specificity (Horstmann *et al.*, 2012). Firstly, it was supposed that an adjustment of the guanine position is needed to overcome a steric clash between the N2 guanine amino group and the C^{β} atom of Ser61 (Fig. 5*d*). Secondly, it was outlined that the Asn28 amide group interacts poorly with N1, which is a hydrogen donor in the case of guanine. In fact, this is a misprint in this paper since Asn28 contacts the adenine N3 atom rather than the N1 atom, so the introduced guanine will have no hindrance from the Asn28 side chain.

This explanation does not apply to PaeHfq and EcoHfq because the adenines in their complexes are located far enough from the asparagine and the threonine that there would consequently be no steric hindrance between an introduced guanine and the amino acids. We speculate that the divergence in the binding of adenine and guanine could be explained by the geometry and energy of the contacts between the nucleotide base atoms and their protein counterparts. The adenine forms one O—H···N and one N—H···O bond, whereas the guanine could make two N—H···O bonds, which are less energetically favourable because of the lower electronegativity of the N atom (Fig. 5*b*). The guanine N2 amino group will be too far from the hydroxyl group of Ser60 to form a hydrogen bond. The serine side chain should rotate to form it, but all other rotamers are not favourable because of steric hindrance with neighbouring amino acids. Therefore, the formation of the hydrogen bond between guanine N2 and the hydroxyl of Ser60 is not favourable. Thus, the serine should play an important role in discrimination between adenine and guanine in both SauHfq and PaeHfq.

In contrast to previous modelling (Link *et al.*, 2009; Horstmann *et al.*, 2012), the crystal structure of the PaeHfq–CTP complex revealed that the R-site could also accommodate CTP (Fig. 4). The cytosine is located in a very similar position to the adenine in the BsuHfq–$(AG)_3A$ and SauHfq–A_7 complexes. The N4, O2 and O2′ cytidine atoms form hydrogen bonds to Thr61 OG1, Asn28 ND2 and Gly29 CO, respectively. A water molecule serves as a bridge between the N3 atom of the cytidine, Ser60 OG and the carboxyl O atom of Leu26. Owing to the multiple contacts CTP is firmly attached to the protein and is found in all six R-sites of PaeHfq.

Our data confirmed that the R-site is a unique nucleotide-binding site on the Hfq distal side owing to its formation by hydrophobic protein residues in combination with the preformed arrangement of several polar amino acids. It possesses the highest affinity for poly(A) RNA thanks to the

Table 2
Contacts of adenine and protein atoms in Hfq–nucleotide complexes.

The resulting average distance (Å) and deviation are calculated for all presented nucleotides in complexes without correction for the resolution of the crystallographic data. The high deviation is caused by variations in the protein side-chain positions relative to the nucleotides. PDB codes are given in parentheses.

(*a*) PaeHfq and EcoHfq.

Contacting atoms	PaeHfq–ADPNP (3qui)	PaeHfq–ATP (4j5y)	EcoHfq–A15 (3gib)	EcoHfq–ADP (3rer)	EcoHfq–ATP (3qo3)
N1–Thr61 OG1	2.8 ± 0.10	2.7 ± 0.19	2.6 ± 0.06	2.9 ± 0.16	2.9 ± 0.11
N6–Gln52 OE1	2.9 ± 0.09	2.6 ± 0.21	2.7 ± 0.05	2.9 ± 0.09	2.9 ± 0.17
N3–H_2O	2.7 ± 0.23	2.9 ± 0.46	3.3 ± 0.08	3.0 ± 0.31	3.0 ± 0.31
O2′–Gly29 CO	3.5 ± 0.64	3.0 ± 0.32	2.8 ± 0.09	2.5 ± 0.08	2.5 ± 0.08

(*b*) BsuHfq.

Contacting atoms	BsuHfq–$(AG)_3A$ (3hsb)	BsuHfq–$(AG)_3A$ (3ahu)
N1–Ser60 OG	2.8 ± 0.05	2.9 ± 0.15
N6–Thr61 OG1	3.0 ± 0.26	3.0 ± 0.26
N3–Asn27 ND2	3.0 ± 0.17	3.4 ± 0.12
O2′–Gly28 CO	3.1 ± 0.54	3.0 ± 0.15

(*c*) SauHfq.

Contacting atoms	SauHfq–A_7 5′-end A (3qsu)	SauHfq–A_7 (3qsu)
N1–Ser61 OG	2.7 ± 0.03	2.4 ± 0.025
N6–Thr62 OG1	2.7 ± 0.08	3.1 ± 0.01
N3–Asn28 ND2	2.9 ± 0.03	3.9 ± 0.43
O2′–Gly29 CO	2.4 ± 0.02	3.8 ± 0.10

hexameric organization of the protein, but seems to have the potential to anchor other nucleotides depending on the RNA sequences.

4.2. Lateral RNA-binding site

There is a known binding area for short poly(U) RNA fragments in the central pore of Hfq (Schumacher *et al.*, 2002; Sauer & Weichenrieder, 2011). Our attempts to obtain a complex of PaeHfq with UTP in this binding site using the soaking procedure were unsuccessful. In the PaeHfq–UTP complex six UTP molecules in the *anti* conformation were found at the proximal protein side bound to residues Asn13, Arg16, Lys17 and Phe39 (Fig. 3). The UTP molecules are located rather close to the protomer N-terminal ends. Every uracil is stacked with a Phe39 phenyl ring, which acts as an adhesive surface. The opposite side of the uracil base is open to the solvent. The hydrogen bonds between the uracil O4 and

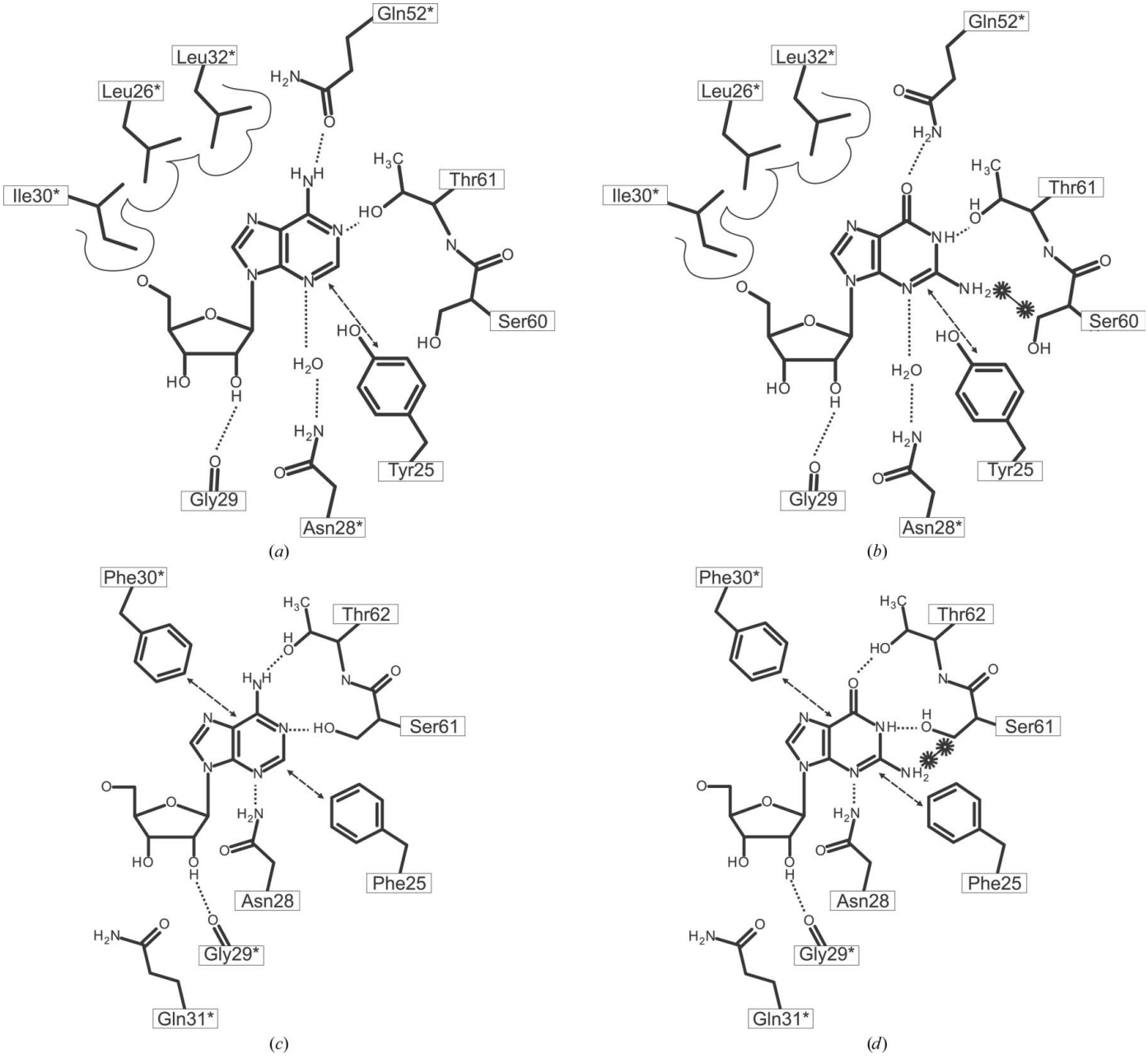

Figure 5
(*a*, *c*) Schematic representations of the interactions between Hfq and adenine in the PaeHfq–ATP complex (PDB entry 4j5y) and in the complex of SauHfq and poly(A) (PDB entry 3qsu; Horstmann *et al.*, 2012), respectively. Dashed lines and arrows indicate hydrogen-bonding and stacking interactions, respectively. The thin line marks the hydrophobic surface of the protein cleft in PaeHfq–ATP. (*b*, *d*) Scheme of the proposed contacts in complexes in which the adenine is subsituted by a guanine. In the case of PaeHfq (*b*) as well as EcoHfq, another rotamer of Ser60 is required to form a hydrogen bond to the N2 amino group of the introduced guanine. In the case of SauHfq (*d*) there is steric hindrance between the introduced N2 amino group of guanine and Ser61 C^{β}.

N3 atoms and the amino and carboxyl main-chain atoms of Phe39 ensure the uridine specificity of this nucleotide-binding site. The O atoms of the 5′-triphosphate form a net of hydrogen bonds to the Asn13, Arg16 and Lys17 side-chain atoms. The UTP α-phosphate forms a hydrogen bond to the side-chain N atoms of Arg16 in most of the six positions, thus representing contacts of the nucleotide 5′-phosphate as part of the RNA. In two positions Arg16 is slightly moved out and the neighbouring Asn13 and Lys17 play its role.

All of these results are in good agreement with the previously published biochemical data. It is known that R16A or F39A substitutions reduced the affinity of EcoHfq for several regulator sRNAs: the $DsrA_{DII}$ sRNA fragment (Sun & Wartell, 2006), the RprA sRNA and rpoS mRNA fragments (Updegrove *et al.*, 2008), OxyS, DsrA sRNA and the DsrA–rpoS-L complex (Updegrove & Wartell, 2011). Nevertheless, the L12F/F39A double mutant possesses approximately the same affinity for sRNA as wild-type EcoHfq (Updegrove & Wartell, 2011). Recently, it has been found that the complex mutation R16S/R17A/R19A/K47A in StyHfq efficiently abolished *S. typhimurinum* RprA sRNA fragment binding (Sauer *et al.*, 2012). The authors identified this area as a third RNA-binding site and proposed a model of an Hfq–sRNA interaction in which this new lateral RNA-binding site plays a crucial role. Recent *in vivo* experiments have supported the impact of EcoHfq residues Arg16, Arg17, Arg19 and Phe39 in the association of small RNAs with their mRNA targets (Zhang *et al.*, 2013).

In fact, the lateral RNA-binding site was originally recognized in the structure of PaSm1 from *Pyrococcus abyssi* in complex with U_7 RNA (Thore *et al.*, 2003). In the PaSm1–U_7 complex the 5′-uridine of the RNA is located at Tyr34, which is the structural homologue of Phe39 in PaeHfq and EcoHfq. This phenylalanine is extremely well conserved throughout all of the known bacterial sequences. The orientation of uridine in the PaSm1–U_7 complex and the contacts between uracil and the protein are identical to those in the PaeHfq–UTP complex. Taking the PaSm1–U_7 structure and the position of UTP in the complex with PaeHfq into account, it is possible to extend the previously suggested model of sRNA binding in the Hfq lateral site (Sauer *et al.*, 2012). It is likely that the 5′-end part of the sRNA is anchored to the lateral binding surface by several uridines located in the UTP-binding sites. The distance between the neighbouring uridines associated with the protein in our model corresponds to at least three or four bound nucleotides in the RNA chain which should be located in the protein α-helix area. It is obvious that additional experiments are needed to check the model of sRNA binding at the lateral binding site.

4.3. The proximal pyrimidine-binding site

Although our attempts to prepare a UTP–PaeHfq complex at the proximal RNA-binding site failed, a similar complex has been obtained by soaking PaeHfq crystals in CTP solution (Figs. 4*a* and 4*c*). The position of cytosine is identical to that of uridine in the previously obtained SauHfq–AU_5G (Schumacher *et al.*, 2002), StyHfq–U_6 (Sauer & Weichenrieder, 2011) and EcoHfq–AU_6A (Wang *et al.*, 2011) complexes. Every cytosine lies along the phenyl ring of Phe42 of Hfq. The interactions between the nucleotide base atoms and protein atoms are not numerous. In fact, only the cytidine O2 atom makes a hydrogen bond to Gln8 in all six sites. The O4 atom positions are close to Gln41 of the neighbouring protomer, but the glutamine is rather distant from the nucleotides and its side chain has poor electron density, confirming the weakness of the potential hydrogen bonding. The same situation was observed in the SauHfq–AU_5G complex, in which Lys41 (the structural homologue of Gln41) did not form a hydrogen bond to the uridine. In the StyHfq–U_6 complex, Gln41 contacted the O4 atom of uracil but the geometry of the hydrogen bond was not optimal (Sauer & Weichenrieder, 2011). In some PaeHfq protomers it is possible to trace the N-terminal end of the protein and observe that the side chain of His5 forms a hydrogen bond to the cytidine N3 atom. Nevertheless, the electron density at the PaeHfq N-terminal ends was uncertain and the position of the N-terminal amino acids could be interpreted ambiguously. In all other structures of Hfq in complex with uridine-rich RNA the uracil N3 atom makes no contacts with the protein.

Thus, the structure of the PaeHfq–CTP complex supports the suggestion that the proximal binding side could accommodate not only uridine but also cytidine owing to its particular binding geometry (Sauer & Weichenrieder, 2011).

5. Conclusion

In general, the crystallization of RNA–protein complexes is a difficult task because of RNA instability and problems in the selection of the required RNA fragment and in finding RNA–protein crystallization conditions. Chemical probing (Christiansen & Garrett, 1988) and SELEX (Conrad *et al.*, 1996) can assist in determining the RNA sequence and predicting the approximate location of the RNA on the protein surface. To locate the RNA-binding site on the protein surface, multiple site-specific mutagenesis experiments followed by measurements of the RNA–protein affinity are necessary.

Our results have demonstrated that it is possible to determine single-stranded RNA-binding sites on the protein surface using a soaking procedure followed by structure determination of the nucleotide–protein complexes. We showed that in the nucleotide–Hfq complexes the positions of ribonucleotides in the two classical sites and in a novel Hfq RNA-binding site correlate with the available structural and biochemical data. The structures of the complexes provide not only the location of the bound nucleotides but also details of the specific recognition of the protein by the nucleotide base.

This method should be specific for single-stranded RNA-binding sites since in the case of double-stranded RNA the proteins recognize its particular fold but not its nucleotides (Nikulin *et al.*, 2000, 2003; Nevskaya *et al.*, 2005; Tishchenko *et al.*, 2006; Gupta & Gribskov, 2011). Indeed, the ribosome protein L1 (a double-stranded RNA-binding protein) gives negative results for ATP-binding tests as does lysozyme, which

is a high positively charged protein. Therefore, the application of the soaking procedure to obtain nucleotide–protein complexes could be used to locate proposed specific RNA-binding sites for single-stranded RNA-binding proteins.

We thank Azat Gabdulkhakov for assistance during data collection and Tatiana Kuvshinkina for technical assistance. We also thank Stanislav Nikonov and Natalia Nevskaya for helpful discussions. This research was partially supported by RFBR (research project 13-04-00783-a), by the Molecular and Cellular Biology Program of RAS and by the German–Russian Cooperation for the Development and Use of Accelerator-Based Light Sources (05K10CBB).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Aiba, H. (2007). *Curr. Opin. Microbiol.* **10**, 134–139.
Arluison, V., Mutyam, S. K., Mura, C., Marco, S. & Sukhodolets, M. V. (2007). *Protein Sci.* **16**, 1830–1841.
Basineni, S. R., Madhugiri, R., Kolmsee, T., Hengge, R. & Klug, G. (2009). *RNA Biol.* **6**, 584–594.
Brennan, R. G. & Link, T. M. (2007). *Curr. Opin. Microbiol.* **10**, 125–133.
Christiansen, J. & Garrett, R. (1988). *Methods Enzymol.* **164**, 456–468.
Conrad, R. C., Giver, L., Tian, Y. & Ellington, A. D. (1996). *Methods Enzymol.* **267**, 336–367.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Gupta, A. & Gribskov, M. (2011). *J. Mol. Biol.* **409**, 574–587.
Hämmerle, H., Beich-Frandsen, M., Večerek, B., Rajkowitsch, L., Carugo, O., Djinović-Carugo, K. & Bläsi, U. (2012). *PLoS One*, **7**, e50892.
Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S. A. III & Brennan, R. G. (2012). *Nucleic Acids Res.* **40**, 11023–11035.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Lazar, P., Kim, S., Lee, Y. & Lee, K. W. (2010). *J. Mol. Graph. Model.* **29**, 573–580.
Link, T. M., Valentin-Hansen, P. & Brennan, R. G. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 19292–19297.
Møller, T., Franch, T., Højrup, P., Keene, D. R., Bächinger, H. P., Brennan, R. G. & Valentin-Hansen, P. (2002). *Mol. Cell*, **9**, 23–30.
Nevskaya, N., Tishchenko, S., Gabdoulkhakov, A., Nikonova, E., Nikonov, O., Nikulin, A., Platonova, O., Garber, M., Nikonov, S. & Piendl, W. (2005). *Nucleic Acids Res.* **33**, 478–485.
Nikulin, A., Eliseikina, I., Tishchenko, S., Nevskaya, N., Davydova, N., Platonova, O., Piendl, W., Selmer, M., Liljas, A., Drygin, D., Zimmermann, R., Garber, M. & Nikonov, S. (2003). *Nature Struct. Biol.* **10**, 104–108.
Nikulin, A., Serganov, A., Ennifar, E., Tishchenko, S., Nevskaya, N., Shepard, W., Portier, C., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S. & Dumas, P. (2000). *Nature Struct. Biol.* **7**, 273–277.
Nikulin, A., Stolboushkina, E., Perederina, A., Vassilieva, I., Blaesi, U., Moll, I., Kachalova, G., Yokoyama, S., Vassylyev, D., Garber, M. & Nikonov, S. (2005). *Acta Cryst.* D**61**, 141–146.
Sauer, E., Schmidt, S. & Weichenrieder, O. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 9396–9401.
Sauer, E. & Weichenrieder, O. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 13065–13070.
Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P. & Brennan, R. G. (2002). *EMBO J.* **21**, 3546–3556.
Sledjeski, D. D., Whitman, C. & Zhang, A. (2001). *J. Bacteriol.* **183**, 1997–2005.
Someya, T., Baba, S., Fujimoto, M., Kawai, G., Kumasaka, T. & Nakamura, K. (2012). *Nucleic Acids Res.* **40**, 1856–1867.
Soper, T., Mandin, P., Majdalani, N., Gottesman, S. & Woodson, S. A. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 9602–9607.
Sun, X. & Wartell, R. M. (2006). *Biochemistry*, **45**, 4875–4887.
Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. (2003). *J. Biol. Chem.* **278**, 1239–1247.
Tishchenko, S. V., Nikonova, E. I., Nevskaia, N. A., Nikonov, O. S., Garber, M. B. & Nikonov, S. V. (2006). *Mol. Biol. (Mosk.)*, **40**, 650–657.
Updegrove, T. B. & Wartell, R. M. (2011). *Biochim. Biophys. Acta*, **1809**, 532–540.
Updegrove, T., Wilf, N., Sun, X. & Wartell, R. M. (2008). *Biochemistry*, **47**, 11184–11195.
Valentin-Hansen, P., Eriksen, M. & Udesen, C. (2004). *Mol. Microbiol.* **51**, 1525–1533.
Vassilieva, I. M., Rouzanov, M. V., Zelinskaya, N. V., Moll, I., Blasi, U. & Garber, M. B. (2002). *Biochemistry (Mosc.)*, **67**, 1293–1297.
Vogel, J. & Luisi, B. F. (2011). *Nature Rev. Microbiol.* **9**, 578–589.
Wang, W., Wang, L., Zou, Y., Zhang, J., Gong, Q., Wu, J. & Shi, Y. (2011). *Genes Dev.* **25**, 2106–2117.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, A., Schu, D. J., Tjaden, B. C., Storz, G. & Gottesman, S. (2013). *J. Mol. Biol.*, doi:10.1016/j.jmb.2013.01.006.