Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, crystallization and preliminary crystallographic analysis of RNA-binding protein Hfq (YmaH) from *Bacillus subtilis* in complex with an RNA aptamer

Seiki Baba,[a] Tatsuhiko Someya,[b] Gota Kawai,[c,d] Kouji Nakamura[b] and Takashi Kumasaka[a]*

[a]Japan Synchrotron Radiation Research Institute/SPring-8, Japan, [b]Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan, [c]Department of Life and Environmental Sciences, Faculty of Engineering, Chiba Institute of Technology, Japan, and [d]RIKEN SPring-8 Center, Japan

Correspondence e-mail: kumasaka@spring8.or.jp

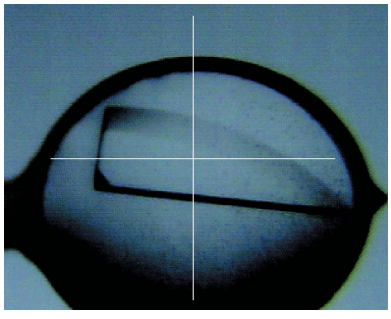
Received 15 February 2010
Accepted 17 March 2010

The Hfq protein is a hexameric RNA-binding protein which regulates gene expression by binding to RNA under the influence of diverse environmental stresses. Its ring structure binds various types of RNA, including mRNA and sRNA. RNA-bound structures of Hfq from *Escherichia coli* and *Staphylococcus aureus* have been revealed to have poly(A) RNA at the distal site and U-rich RNA at the proximal site, respectively. Here, crystals of a complex of the *Bacillus subtilis* Hfq protein with an A/G-repeat 7-mer RNA (Hfq–RNA) that were prepared using the hanging-drop vapour-diffusion technique are reported. The type 1 Hfq–RNA crystals belonged to space group *I*422, with unit-cell parameters $a = b = 123.70$, $c = 119.13$ Å, while the type 2 Hfq–RNA crystals belonged to space group *F*222, with unit-cell parameters $a = 91.92$, $b = 92.50$, $c = 114.92$ Å. Diffraction data were collected to a resolution of 2.20 Å from both crystal forms. The hexameric structure of the Hfq protein was clearly shown by self-rotation analysis.

1. Introduction

Hfq is an important component in the regulation of gene expression according to changes in the growth environment and is found in many bacteria. This protein was first identified as a host factor that was required for the replication of RNA phage Qβ in *Escherichia coli* (Franze *et al.*, 1972). The function of Hfq as a post-transcriptional regulator has been well studied in Gram-negative bacteria, especially *E. coli*; its regulation is controlled by interaction with various RNAs. *E. coli* Hfq (EcHfq) has been shown to interact with several small untranslated RNAs such as OxyS, DsrA and Spot42 (Zhang *et al.*, 2002; Majdalani *et al.*, 1998; Møller *et al.*, 2002). Moreover, EcHfq can interact with many targets as an RNA chaperone for mRNA (Moll *et al.*, 2003). In addition, EcHfq is also involved in the stabilization of RNA by protecting it from degradation by two exoribonucleases: RNase II and polynucleotide phosphorylase (Afonyushkin *et al.*, 2005; Mohanty *et al.*, 2004). In terms of gene regulation, EcHfq has been shown to interact with two RNA sequences: poly(A) tails of mRNA and unstructured A/U-rich sequences near stem-loop structures (Hajnsdorf & Régnier, 2000; Folichon *et al.*, 2005; Brescia *et al.*, 2003; Geissmann & Touati, 2004; Brennan & Link, 2007). Its RNA-binding sites have been identified by site-directed mutagenesis (Mikulecky *et al.*, 2004; Ziolkowska *et al.*, 2006).

Recently, crystallographic analyses of Hfq from several different bacteria have revealed that it forms a homohexameric ring consisting of an Sm-like motif and has two possible RNA-binding sites on both sides of the torus-shaped ring (Schumacher *et al.*, 2002; Sauter *et al.*, 2003; Nikulin *et al.*, 2005; Nielsen *et al.*, 2007). The first structural determination of an Hfq–RNA complex was performed on Hfq from the Gram-positive *Staphylococcus aureus* (SaHfq) in complex with the short U-rich RNA AU5G; it revealed that the RNA is recognized by residues on the loop between β2 and β3 and β4 and β5, defined as proximal RNA-binding sites (Schumacher *et al.*, 2002). Alternatively, a structure with RNA bound to the distal site has recently been determined for EcHfq (Link *et al.*, 2009). This study showed that the poly(A) RNA binds to the (rA-rR-rN) RNA-binding site of EcHfq, on which the phosphodiester backbone traces a circular weaving



pattern. This site consists of an adenosine-specific site, a purine nucleotide-selective site and a sequence-nondiscriminating RNA-entrance/exit site.

In the Gram-positive bacterium *Bacillus subtilis*, Hfq is encoded by the *ymaH* gene. *B. subtilis* Hfq (BsHfq) can bind both SR1 sRNA and *ahr*C mRNA; however, it was not required for stabilization of SR1 sRNA or for promotion of the RNA–RNA interaction between SR1 sRNA and *ahr*C mRNA (Heidrich *et al.*, 2006, 2007). Based on its sequence similarity to SaHfq, BsHfq is expected to have a hexameric structure. On the other hand, sRNA–mRNA interactions in *S. aureus* are decisive for regulation of gene expression; they do not require Hfq (Bohn *et al.*, 2007). We identified that BsHfq recognizes the A/G-repeat RNA sequence as a novel Hfq-binding sequence using systematic evolution of ligands by exponential enrichment (SELEX) and gel-shift assays (manuscript in preparation). In order to investigate the RNA-binding mechanism, structural analysis of the BsHfq–RNA complex is in progress. Here, we report the expression, crystallization and preliminary X-ray studies of the complex of BsHfq with A/G-repeat RNA.

2. Materials and methods

2.1. Protein expression and purification

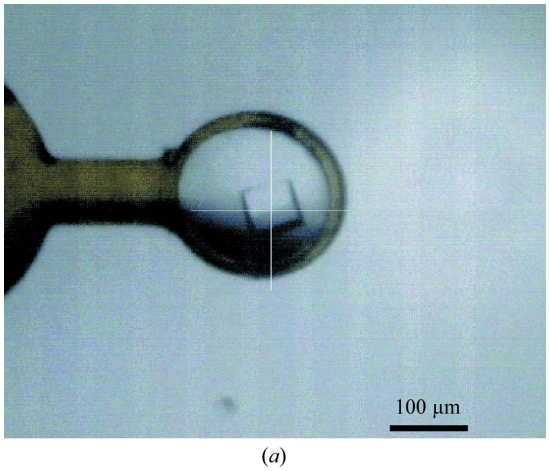
The coding sequence of Hfq from *B. subtilis* was amplified by PCR to add a *Bam*HI site at the 5′ end and an *Eco*RI site at the 3′ end. Two primers were used: a forward primer (5′-CGCACT*GGATCC*ATG-ATGAAACCGATTAATATTCAGGATC-3′) containing a *Bam*HI site (shown in italics) and a reverse primer (5′-CGCGTA*GAATTC*-TTATTCGAGTTCAAGCTGGACG-3′) containing an *Eco*RI site (shown in italics). The amplified fragment was cloned into a *Bam*HI–*Eco*RI-digested pGEX-6P-1 expression vector (GE Healthcare). GST-Hfq was expressed in *E. coli* BL21; expression was induced for 5 h in the presence of 1 m*M* IPTG. Cells were harvested by centrifugation at 5000*g* and 277 K for 15 min. The wet cells were dissolved in sonication buffer (10 m*M* Na_2HPO_4, 1.8 m*M* KH_2PO_4, 140 m*M* NaCl, 2.7 m*M* KCl, 1 m*M* DTT pH 7.4), lysed on ice by sonication and centrifuged at 5000*g* and 277 K for 15 min. The supernatant was loaded onto a GST-affinity column (Glutathione Sepharose 4 Fast Flow, GE Healthcare) and eluted with 50 m*M* Tris buffer pH 8.0 containing 10 m*M* reduced glutathione and 1 m*M* dithiothreitol (DTT). The elution pattern was monitored by 15% SDS–PAGE. GST-Hfq-containing fractions were collected and dialyzed against 50 m*M* Tris buffer pH 7.0 and the sample was loaded onto an anion-exchange column (Q Sepharose FF, GE Healthcare) and eluted with 50 m*M* Tris buffer pH 7.0 containing 500 m*M* NaCl and 1 m*M* DTT; the elution pattern was monitored by 15% SDS–PAGE. The GST-Hfq-containing fractions were collected and concentrated by ultrafiltration with an Amicon Ultra-15 (Millipore). After GST-tag cleavage with PreScission protease (GE Healthcare), which leaves a remnant sequence (GPLGS) from the tag at the N-terminus, the protein solution was loaded onto a HiTrap SP HP cation-exchange column (GE Healthcare) and eluted with a linear gradient of 100–400 m*M* NaCl in 35 ml 50 m*M* Tris buffer pH 7.0. The Hfq-containing fraction was pooled and concentrated using an Amicon Ultra-15. The protein concentration was determined by measuring the absorbance at 280 nm. The concentration of the purified protein was 8.9 mg ml^{-1} in 10 m*M* Tris–HCl pH 7.5, 10 m*M* NaCl.

2.2. RNA synthesis, purification and preparation

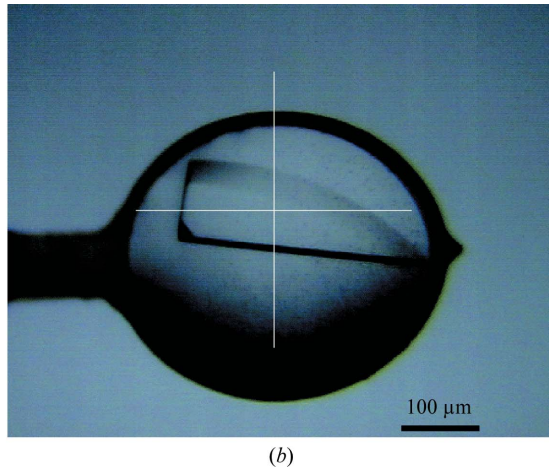
The RNA sample (rArGrArGrArGrA) was chemically synthesized using a DNA/RNA synthesizer (Expedite 8909, PerSeptive). The sample was purified using 20% PAGE under denaturing conditions with 8 *M* urea and was concentrated after desalting by ethanol precipitation. The solvent was adjusted to 10 m*M* Tris–HCl pH 7.5, 10 m*M* NaCl by adding concentrated buffer.

2.3. Crystallization

Crystallization was performed using the hanging-drop vapour-diffusion method at 293 K. Initial screening was performed with the commercial sparse-matrix crystallization kits Crystal Screen 1, Crystal Screen 2 and Natrix (Hampton Research). A 1 µl volume of Hfq–RNA solution was mixed with an equal amount of reservoir solution and the droplet was allowed to equilibrate against 300 µl reservoir solution in a sealed VDXm plate (Hampton Research). In the initial trial, crystals appeared after two weeks. After further optimization of the conditions, we obtained two crystallization conditions for Hfq–RNA: type 1 and type 2. In the crystallization condition for type 1 Hfq–RNA (714 µ*M* Hfq protein:119 µ*M* RNA) the reservoir solution consisted of 0.015 *M* cobalt(II) chloride, 0.1 *M* MES pH 6.5 and 1.8 *M* ammonium sulfate and the droplets were allowed to equilibrate against 100 µl reservoir solution in a sealed VDX48 plate (Hampton Research) for three weeks. For type 2 Hfq–RNA (735 µ*M* Hfq protein:250 µ*M* RNA) the reservoir solution consisted of 0.01 *M* cobalt(II) chloride, 0.2 *M* MES pH 6.5, 1.8 *M*



(*a*)



(*b*)

Figure 1
(*a*) Type 1 and (*b*) type 2 Hfq–RNA complex crystals.

ammonium sulfate and the droplets were allowed to equilibrate against 300 µl reservoir solution in a sealed VDXm plate (Hampton Research) for two weeks. The molar ratios of the Hfq protein to the RNA aptamer in the crystallization conditions for type 1 and type 2 Hfq–RNA were 6:1 and 6:2, respectively.

2.4. Crystallographic data collection, processing and analysis

All X-ray diffraction data were collected on beamline BL38B1 of SPring-8 using a Rigaku Jupiter210 CCD detector. The type 1 and type 2 Hfq–RNA crystals were cryoprotected by soaking them in mother liquor with 24%(*v*/*v*) and 35%(*v*/*v*) glycerol, respectively. The crystals were then flash-cooled in a cold nitrogen-gas stream at 100 K. All data collections were performed at a wavelength of 1.0 Å with a total oscillation range of 180°. Diffraction images for type 1 and type 2 Hfq–RNA crystals were obtained with oscillation angles of 0.3° and 1.0°, exposure times of 40.0 and 20.0 s and crystal-to-detector distances of 150 and 176 mm, respectively. The data were integrated and scaled using the *HKL*-2000 program package (Otwinowski & Minor, 1997). *MOLREP* (Vagin & Teplyakov, 1997) was used to calculate self-rotation functions and perform molecular replacement using SaHfq (PDB code 1kq2; Schumacher *et al.*, 2002), which shares 44% sequence identity with BsHfq, as a search model. The obtained models were refined with the program *REFMAC*5 (Murshudov *et al.*,

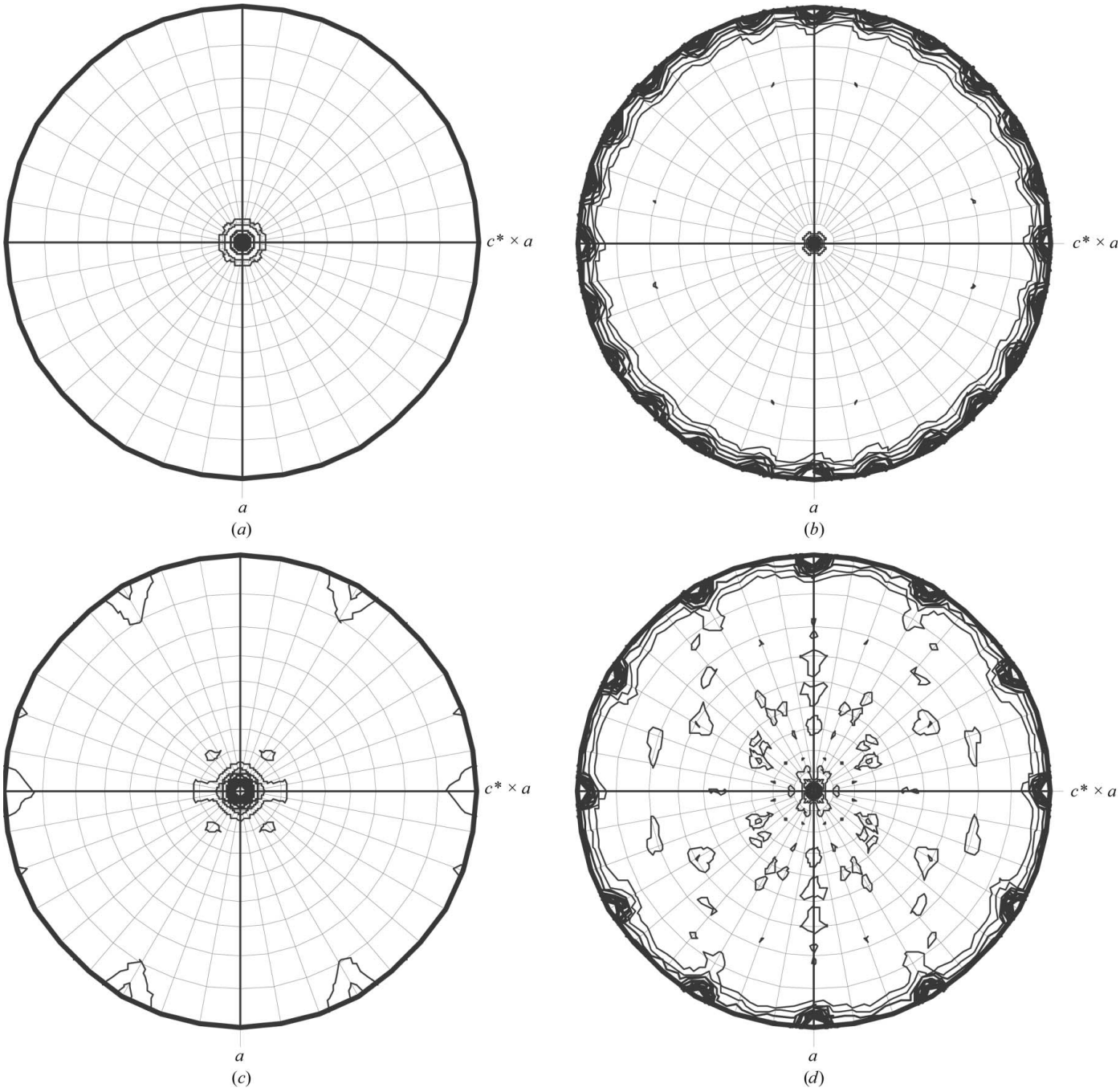


Figure 2
Self-rotation functions for (*a*, *b*) type 1 Hfq–RNA (*I*422) and (*c*, *d*) type 2 Hfq–RNA (*F*222). The sections corresponding to rotation by (*a*, *c*) $\chi = 60^\circ$ and (*b*, *d*) $\chi = 180^\circ$ are shown. This figure was produced using the program *MOLREP* (Vagin & Teplyakov, 1997).

Table 1
Diffraction data statistics for Hfq–RNA complexes.

Values in parentheses are for the highest resolution shell.

	Type 1 Hfq–RNA	Type 2 Hfq–RNA
X-ray source	SPring-8 BL38B1	SPring-8 BL38B1
Detector	Jupiter210	Jupiter210
Wavelength (Å)	1.0000	1.0000
Resolution (Å)	25.00–2.20 (2.28–2.20)	30.00–2.20 (2.28–2.20)
Space group	*I*422	*F*222
Unit-cell parameters (Å)	$a = b = 123.70$, $c = 119.13$	$a = 91.92$, $b = 92.50$, $c = 114.92$
Multiplicity	14.8 (14.9)	7.4 (7.4)
No. of observations	350287	92082
No. of unique reflections	23732	12626
Completeness (%)	100 (100)	100 (100)
$R_{\rm merge}$† (%)	6.3 (32.5)	3.3 (30.7)
Average $I/\sigma(I)$	38.02 (7.07)	45.75 (4.84)
Matthews coefficient (Å^3 Da^{-1})	2.34	2.51
Solvent content (%)	48	51

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

1997). *MOLREP* and *REFMAC*5 were implemented within the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

3. Results

The type 1 crystals of Hfq–RNA (Fig. 1*a*) were found to diffract to 2.20 Å resolution and to belong to the body-centred tetragonal space group *I*422, with unit-cell parameters $a = b = 123.70$, $c = 119.13$ Å (Table 1). Assuming the presence of six protein monomers (one hexamer) in the asymmetric unit, the Matthews coefficient was 2.34 Å^3 Da^{-1}, corresponding to a solvent content of 48% (Matthews, 1968; Table 1). The type 2 crystals of Hfq–RNA (Fig. 1*b*) were found to diffract to 2.20 Å resolution and to belong to the face-centred orthorhombic space group *F*222, with unit-cell parameters $a = 91.92$, $b = 92.50$, $c = 114.92$ Å (Table 1). Assuming the presence of three protein monomers (half a hexamer) in the asymmetric unit, the Matthews coefficient was 2.51 Å^3 Da^{-1}, corresponding to a solvent content of 51% (Matthews, 1968; Table 1).

Self-rotation function calculations for type 1 and type 2 Hfq–RNA were performed using data in the resolution ranges 36.18–4.00 and 29.39–4.00 Å, respectively (Fig. 2). High peaks are observed for both crystal forms along the *c* axis at $\chi = 60^\circ$ and 120° and correspond to sixfold and/or threefold noncrystallographic symmetry axes, respectively (Figs. 2*a* and 2*c*). By combination of the noncrystallographic and crystallographic symmetries, peaks for type 1 and type 2 Hfq–RNA around the *c* axis at $\chi = 180^\circ$ (Figs. 2*b* and 2*d*) indicate that the sixfold noncrystallographic symmetry axis is parallel to the *c* axis in both forms. These results are consistent with the experimental assumption that BsHfq forms a homohexameric C_6 ring. This was confirmed by the molecular-replacement trial, as described below.

Further molecular-replacement trials using the SaHfq structure gave good initial solutions with bound RNA. The solutions of type 1 and type 2 Hfq–RNA exhibited *R* factors of 0.335 and 0.328, and correlation coefficients of 0.631 and 0.650, respectively. Construction, revision and analysis of atomic models using the BsHfq sequence and RNA are currently in progress.

Synchrotron-radiation experiments were performed on beamline BL38B1 at SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI; proposal No. 2008B1506).

References

Afonyushkin, T., Vecerek, B., Moll, I., Bläsi, U. & Kaberdin, V. R. (2005). *Nucleic Acids Res.* **33**, 1678–1689.
Bohn, C., Rigoulay, C. & Bouloc, P. (2007). *BMC Microbiol.* **7**, 10.
Brennan, R. G. & Link, T. M. (2007). *Curr. Opin. Microbiol.* **10**, 125–133.
Brescia, C. C., Mikulecky, P. J., Feig, A. L. & Sledjeski, D. D. (2003). *RNA*, **9**, 33–43.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Folichon, M., Allemand, F., Régnier, P. & Hajnsdorf, E. (2005). *FEBS J.* **272**, 454–463.
Franze, F. M. T., Hayward, W. S. & August, J. T. (1972). *J. Biol. Chem.* **247**, 824–831.
Geissmann, T. A. & Touati, D. (2004). *EMBO J.* **23**, 396–405.
Hajnsdorf, E. & Régnier, P. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 1501–1505.
Heidrich, N., Chinali, A., Gerth, U. & Brantl, S. (2006). *Mol. Microbiol.* **62**, 520–536.
Heidrich, N., Moll, I. & Brantl, S. (2007). *Nucleic Acids Res.* **35**, 4331–4346.
Link, T. M., Valentin-Hansen, P. & Brennan, R. G. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 19292–19297.
Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T. & Gottesman, S. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12462–12467.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D. & Feig, A. L. (2004). *Nature Struct. Mol. Biol.* **11**, 1206–1214.
Mohanty, B. K., Maples, V. F. & Kushner, S. R. (2004). *Mol. Microbiol.* **54**, 905–920.
Moll, I., Leitsch, D., Steinhauser, T. & Bläsi, U. (2003). *EMBO Rep.* **4**, 284–289.
Møller, T., Franch, T., Udesen, C., Gerdes, K. & Valentin-Hansen, P. (2002). *Genes Dev.* **16**, 1696–1706.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nielsen, J. S., Bøggild, A., Andersen, C. B., Nielsen, G., Boysen, A., Brodersen, D. E. & Valentin-Hansen, P. (2007). *RNA*, **13**, 2213–2223.
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.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sauter, C., Basquin, J. & Suck, D. (2003). *Nucleic Acids Res.* **31**, 4091–4098.
Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P. & Brennan, R. G. (2002). *EMBO J.* **21**, 3546–3556.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C. & Storz, G. (2002). *Mol. Cell*, **9**, 11–22.
Ziolkowska, K., Derreumaux, P., Folichon, M., Pellegrini, O., Régnier, P., Boni, I. V. & Hajnsdorf, E. (2006). *Nucleic Acids Res.* **34**, 709–720.