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

Cloning, expression, crystallization and preliminary crystallographic analysis of a pentapeptide-repeat protein (Rfr23) from the bacterium *Cyanothece* 51142

Garry W. Buchko,[a]* Howard Robinson,[b] Shuisong Ni,[a]‡ Himadri B. Pakrasi[c] and Michael A. Kennedy[a]*‡

[a]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA, [b]Biology Department, Brookhaven National Laboratory, Upton, NY 11973-5000, USA, and [c]Department of Biology, Washington University, St Louis, MI 63130-4899, USA

‡ Current address: Department of Biochemistry, Miami University, Oxford, OH 45056, USA.

Correspondence e-mail: garry.buchko@pnl.gov, michael.kennedy@muohio.edu

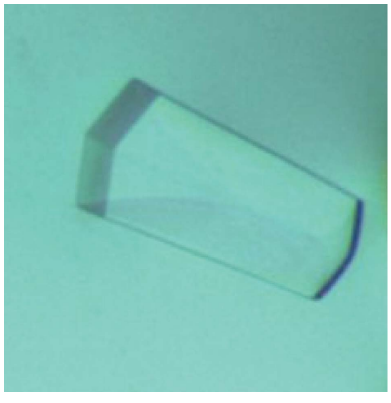
Received 28 September 2006
Accepted 6 November 2006

A unique feature of cyanobacteria genomes is the abundance of genes that code for hypothetical proteins containing tandem pentapeptide repeats approximately described by the consensus motif A(N/D)L*XX*. To date, the structures of two pentapeptide-repeat proteins (PRPs) have been determined, with the tandem pentapeptide-repeat sequences observed to adopt a novel type of right-handed quadrilateral β-helix, or Rfr-fold, in both structures. One structure, *Mycobacterium tuberculosis* MfpA, is a 183-residue protein that contains 30 consecutive pentapeptide repeats and appears to offer antibiotic resistance by acting as a DNA mimic. The other structure, *Cyanothece* 51142 Rfr32, is a 167-residue protein that contains 21 consecutive pentapeptide repeats. The function of Rfr32, like the other 35 hypothetical PRPs identified in the genome of *Cyanothece*, is unknown. In an effort to understand the role of PRPs in cyanobacteria and to better characterize the structural properties of Rfr-folds with different amino-acid sequences, a second PRP from *Cyanothece* 51142, Rfr23, has been cloned, expressed and purified. Selenomethione-substituted protein was crystallized by vapor diffusion in hanging drops. Nearly complete SAD and native diffraction data sets were collected from these crystals to 2.5 and 2.1 Å resolution, respectively, using synchrotron radiation. The crystals belonged to space group $I4_1$, with unit-cell parameters $a = b = 106.61$, $c = 53.37$ Å, and one molecule per asymmetric unit. Preliminary analysis of the electron-density map from the SAD data shows that Rfr23 contains an Rfr-fold.

1. Introduction

The genome of the diurnal cyanobacterium *Cyanothece* sp. strain ATCC 51142 (*Cyanothece* 51142; contig. 83.1_1_243_746) has recently been sequenced and observed to contain 35 proteins with tandem pentapeptide repeats of general consensus sequence A(D/N)L*XX* (Buchko *et al.*, 2006). Pentapeptide-repeat proteins (PRPs) are present throughout the prokaryotic and eukaryotic kingdoms; however, they appear clustered in cyanobacteria. For example, in addition to the 35 PRPs discovered in *Cyanothece* 51142, 16 have been identified in *Synechocystis* sp. strain PCC6803 (Bateman *et al.*, 1998) and 40 in *Nostoc punctiforme* (Vetting *et al.*, 2006). The sheer number of PRPs in photosynthethic cyanobacteria, coupled with their predicted location in all the cyanobacteria cellular compartments, argues for an important, as yet unknown, physiological function.

Bateman *et al.* (1998) first discovered the novel family of proteins containing tandem pentapeptide repeats through analyses of the complete genome sequences that were then becoming widely available for many organisms. Currently, the Pfam database (Bateman *et al.*, 2000) lists over 2000 pentapeptide-repeat proteins (PF00805). Despite the large number of PRPs, only two PRP structures have been determined: the first was reported in 2005 (Hegde *et al.*, 2005) and the second in 2006 (Buchko *et al.*, 2006). The first structure was obtained from crystals of MfpA, a 183-residue protein associated with mycobacterial fluoroquinoline resistance in *Mycobacterium tuberculosis* (Hegde *et al.*, 2005). It was targeted for study because it was identified as a homolog (67% identity) of a newly discovered 193-residue protein in *M. smegmatis* that was shown to be responsible for



fluoroquinolone resistance in this fast-growing mycobacterium (Montero *et al.*, 2001). The second structure was obtained from crystals of Rfr32, a 167-residue protein from *Cyanothece* 51142 (Buchko *et al.*, 2006). It was targeted for study in order to obtain structure-based insights into the function of PRPs in cyanobacteria. MfpA and Rfr32 contain 30 and 21 consecutive pentapeptide repeats, respectively, and the crystal structures reveal that they adopt a novel type of right-handed quadrilateral β-helix, a repeated five-residue fold (Rfr-fold) reminiscent of a square tower with four distinct faces (Hegde *et al.*, 2005; Buchko *et al.*, 2006). As shown in Fig. 1, each pentapeptide repeat occupies one face of the Rfr-fold, with four consecutive pentapeptide repeats completing a coil that makes a revolution with an ~4.8 Å rise every 20 residues. The regular shape of the tower is maintained by two distinct four-residue type II and type IV β-turns, which may be universal motifs that shape the Rfr-fold in all pentapeptide-repeat proteins (Buchko *et al.*, 2006).

There is compelling evidence that the biochemical function of MfpA, and perhaps other PRPs expressed from bacterial plasmids, is to provide resistance to fluoroquinolones and other antibiotics. Two MfpA molecules interact at the C-termini to form a linear dimer that exhibits characteristics similar to those of B-form DNA, including size, shape and predominately electronegative surface-potential distribution. Indeed, the MfpA structure can be docked readily onto the crystal structure of an N-terminal construct of *Escherichia coli* DNA gyrase A subunit (Morais Cabral *et al.*, 1997), a protein with a large electropositive potential at the position where DNA is believed to bind, acting as a DNA mimic. This structural data was supported by biochemical data showing that MfpA inhibits the supercoiling and relaxing activity of *E. coli* DNA gyrase (Hegde *et al.*, 2005). The DNA of at least two other bacterial plasmids encode PRPs that offer antibiotic resistance: the *E. coli* McbG protein (Garrido *et al.*, 1988) and the *Bacillus magaterium* oxetanocin A resistance factor OxrA (Morita *et al.*, 1999). Because McbG and OxrA contain 13 and nine tandem pentapeptide repeats, respectively, it has been suggested that they may provide resistance in a mechanism similar to that proposed for MfpA and fluoroquinolones by acting as a DNA mimic for the antibiotic's target enzyme (Vetting *et al.*, 2006).

The plasmid DNA of MfpA and other PRPs that offer antibiotic resistance are likely to be PRP genes that had their origins in chromosomal DNA. However, little is known about the original biochemical function of the pentapeptide-repeat domain in these chromosomal PRP gene products and nothing is known about their mechanism of action. In order to gain a better understanding of the molecular function of PRPs in cyanobacteria and to further characterize the effect of sequence variation on structural features of the Rfr-fold, we have crystallized a second PRP from *Cyanothece* 51142, Rfr23. Unlike Rfr32, which is predicted to be transported into the thylakoid lumen, SOSUI signal analysis (Gomi *et al.*, 2004) predicts that Rfr23 contains an N-terminal transmembrane region that anchors it onto the surface of a membrane. In this paper, we discuss the cloning, expression, crystallization and preliminary crystallographic data for a truncated version of Rfr23, Thr27–Asp174, constructed with its membrane-spanning region removed.

2. Materials and methods

2.1. Cloning, expression and purification

The *Rfr23* gene minus the N-terminal 26 residues containing the predicted transmembrane region was amplified using the genomic DNA of *Cyanothece* sp. strain ATCC 51142 and the oligonucleotide primers 5′-ATCGAGGTCTCACATGACTAATAATTTGTACCGC-TTG-3′ and 5′-TGAGGTCTCCTCTCGAGCTAATCCCAAACAG-TATTTTTATCAG-3′ (Midland, Midland, TX, USA). The amplified gene, *Rfr23*, was inserted into the *Nco*I/*Xho*I-digested expression vector pET30b (Novagen, Madison, WI, USA) such that a 43-residue tag containing six consecutive histidine residues was added to the N-terminus of the gene product in order to assist in protein purification. The recombinant plasmid was transformed into *E. coli* BL21(DE3) cells (Novagen, Madison, WI, USA). The selenomethionine-substituted protein was expressed following a protocol that inhibited the methionine-biosynthesis pathway (Van Duyne *et al.*, 1993; Doublié, 1997). In brief, after growing the cells at 310 K to mid-log phase ($OD_{600\,nm} \simeq 0.8$) in M9 minimal medium supplemented with 34 μg ml^{-1} kanamycin, 30 μg ml^{-1} chloramphenicol, 120 μg ml^{-1} $MgSO_4$, 11 μg ml^{-1} $CaCl_2$, 10 ng ml^{-1} Fe_2Cl_3, 50 μg ml^{-1} NaCl and 4 mg ml^{-1} glucose, the temperature was lowered to 298 K. At this point, lysine (0.1 mg ml^{-1}), phenylalanine (0.1 mg ml^{-1}), threonine (0.1 mg ml^{-1}), isoleucine (0.05 mg ml^{-1}), valine (0.05 mg ml^{-1}) and selenomethionine (SeMet; 0.06 mg ml^{-1}) were added, followed by the induction of protein expression ~15 min later with isopropyl β-D-1-thiogalactopyranoside (0.026 mg ml^{-1}). Approximately 4–6 h later, the cells were harvested by mild centrifugation and then frozen at 193 K. Thawed cells from a 750 ml culture were resuspended in ~35 ml lysis buffer (0.3 *M* NaCl, 50 m*M* sodium phosphate, 10 m*M* imidazole pH 8.0) and brought to 0.2 m*M* in PMSF prior to three passes through a French press (SLM Instruments, Rochester, NY, USA). Following sonication for 30 s, the cell debris was spun at 25 000*g* for 1.5 h in a JA-20 rotor (Beckman Instruments, Fullerton, CA, USA). After passage through a 0.45 μm SFCA syringe filter (Corning Incorporation, Corning, NY, USA), the supernatant was loaded onto a 20 ml Ni–NTA affinity column (Qiagen, Valencia, CA, USA) and washed stepwise with 50 ml buffer (0.3 *M* NaCl, 50 m*M* sodium phosphate pH ~8.0) containing increasing concentrations of imidazole (5, 10, 20, 50 and 250 m*M*). The fraction containing Rfr23 eluted primarily with the 250 m*M* imidazole fraction

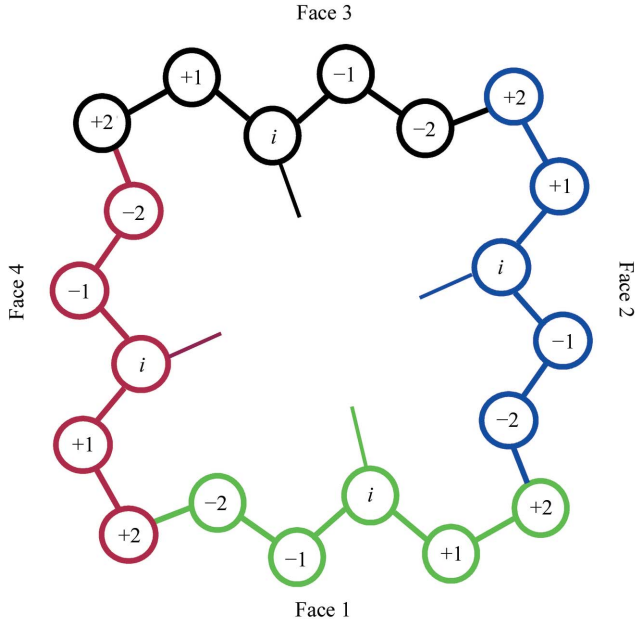


Figure 1
Stylized representation of a complete coil in an Rfr-fold, the higher order structural unit adopted by four tandem pentapeptide repeats. Each pentapeptide repeat is colored differently and is labeled relative to the central residue *i*. The coils stack on top of each other with a rise of ~4.8 Å.

(~40 mg per litre of medium). The protein fraction was dialyzed (8 kDa molecular-weight cutoff) overnight at 277 K in 4 l enterokinase buffer (50 m*M* NaCl, 20 m*M* Tris–HCl pH 7.4). Following volume reduction to ~1.5 ml (Amicon Centriprep-10), the N-terminal tag was removed at room temperature with ~1 unit of enterokinase (GenScript Corp., Piscataway, NJ, USA) per 2 mg protein and 1.5 µl 1.0 *M* $CaCl_2$. After cleavage of the ~5 kDa tag, uncut protein and other impurities were removed by passage through a Superdex75 HiLoad 26/60 column (Amersham Pharmacia Biotech, Piscataway, NJ, USA) that simultaneously exchanged Rfr23 into the buffer used for crystallization (100 m*M* NaCl, 20 m*M* Tris, 1.0 m*M* dithiothreitol pH 7.1). Using a flow rate of 2.5 ml min^{-1}, cleaved Rfr23 eluted with a retention time of 76 min, a value characteristic of a protein with a calculated monomeric molecular weight of 16 410 Da (after enterokinase cleavage, Rfr23 contains an alanine prior to the starting methionine). SDS–PAGE showed the proteins to be greater than 97% pure.

Note that all crystals discussed here were grown using a truncated version of native Rfr23, Thr27–Asp174, that contained an additional two amino-acid residues (Ala25*Met26*-) at the N-terminus (asterisks denote the two non-native residues) after enterokinase digestion.

2.2. Crystallization

Vapor-diffusion crystallization trials using hanging drops were set up using SeMet-labeled Rfr23 at room temperature (~295 K) using screens from Hampton Research (Aliso Viejo, CA, USA). Crystals began appearing 24 h later under a number of conditions, which were then optimized. The crystals shown in Fig. 2 were harvested 3–4 d after mixing 2 µl protein solution (~8 mg ml^{-1}) with 2 µl reservoir buffer containing 0.192 *M* magnesium acetate, 0.096 *M* sodium cacodylate, 10%(*w*/*v*) polyethylene glycol 8000, 10%(*v*/*v*) glycerol pH 6.5. The 'blade-like' crystal shown in the inset of Fig. 2 was grown similarly using reservoir buffer containing 0.185 *M* ammonium sulfate, 0.0925 *M* sodium cacodylate, 20.8%(*w*/*v*) polyethylene glycol 8000, 7.5% glycerol pH 6.5. Crystals grown under the latter conditions were directly mounted in nylon CryoLoops (Hampton Research), flash-frozen in liquid nitrogen and stored under liquid nitrogen until transport to the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory.

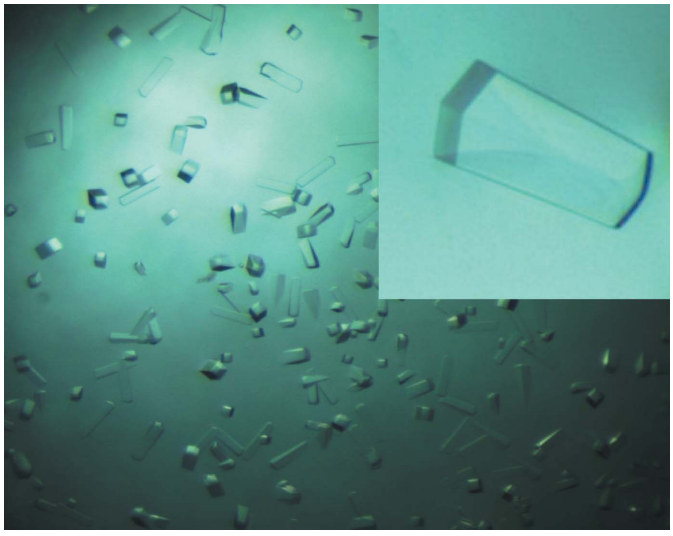


Figure 2
Crystals of SeMet-substituted Rfr23 from *Cyanothece* 51142 grown by mixing 2 µl protein solution (~8 mg ml^{-1}) with 2 µl reservoir buffer. The reservoir buffer used for the crystals in the main figure contained 0.192 *M* magnesium acetate, 0.096 *M* sodium cacodylate, 10%(*w*/*v*) polyethylene glycol 8000, 10%(*v*/*v*) glycerol pH 6.5. The reservoir buffer used for the crystal in the inset contained 0.185 *M* ammonium sulfate, 0.0925 *M* sodium cacodylate, 20.8%(*w*/*v*) polyethylene glycol 8000, 7.5% glycerol pH 6.5. SAD and native data sets reported here were collected from crystals of the type shown in the inset.

Table 1
Summary of the diffraction data-collection parameters and statistics for Rfr23.

Values in parentheses are for the highest resolution shell.

X-ray source	X29
Detector	ADSC Q315 CCD
X-ray wavelength (Å)	0.9791
Temperature (K)	100
Data set	SAD
Space group	$I4_1$
Unit-cell parameters	
a (Å)	106.61
b (Å)	106.61
c (Å)	53.37
$\alpha = \beta = \gamma$ (°)	90
Molecules per ASU	1
Matthews coefficient (V_M; $Å^3$ Da^{-1})	4.6
Solvent content (%)	73
Resolution range (Å)	50–2.6 (2.69–2.60)
Total No. of reflections	111116 (5865)
Total No. of unique reflections	8875 (682)
Average redundancy per shell (%)	12.5 (8.6)
Completeness (%)	94.8 (74.5)
$R_{sym}I$† (%)	0.096 (0.301)
Mean $I/\sigma(I)$	43.8 (4.2)

† $R_{sym}I = \sum_{hkl}\sum_j |I_j(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_j I_j(hkl)$, where $I_j(hkl)$ is the intensity of the *j*th symmetry-related reflection of $I(hkl)$ and $\langle I(hkl)\rangle$ is its average.

2.3. X-ray data collection and processing

X-ray diffraction data were collected at the X29A beamline using an ADSC Q315 CCD detector. Owing to the presence of 20.8%(*w*/*v*) PEG 8000 and 7.5% glycerol, which serve as cryoprotectants, the precipitant solution worked well when cryocooling the crystals in a nitrogen stream (100 K) during data collection. A SAD data set was collected on the SeMet-labeled crystals (Fig. 2, inset) to a resolution of 2.5 Å. A second native set was collected on a fresh crystal to a resolution of 2.1 Å. The space group of these crystals proved to be $I4_1$, with unit-cell parameters $a = b = 106.61$, $c = 53.37$ Å. The images for the SAD data set were integrated and scaled with the *HKL*-2000 suite (Otwinowski & Minor, 1997). Heavy-atom sites were located using the *SHELX* program suite (Schneider & Sheldrick, 2002; Sheldrick 2002) and *HKL2MAP* (Pape & Schneider, 2004). The peak wavelength data of the SAD data set was used with the program *SOLVE* (Terwilliger & Berendzen, 1999) to refine the position of the selenium heavy-atom sites. A refined electron-density map was generated using the program *RESOLVE* (Terwilliger, 2000), with the best map obtained using a solvent content of 70%. All data-collection statistics are summarized in Table 1.

3. Results and discussion

SeMet-labeled truncated recombinant Rfr23 (Thr27–Asp174) with an ~5 kDa N-terminal tag containing six consecutive histidines was expressed in *E. coli* BL21(DE3) as a soluble fraction in high yield (~40 mg per litre of medium). The tag was removed by enterokinase to generate a 16.4 kDa protein that behaved as a monomer in solution according to size-exclusion chromatography. Rfr23 readily crystallized under a number of screening conditions, with two of the better, optimized conditions described in the caption for Fig. 2. Nearly complete SAD and native data sets that diffracted to 2.5 and

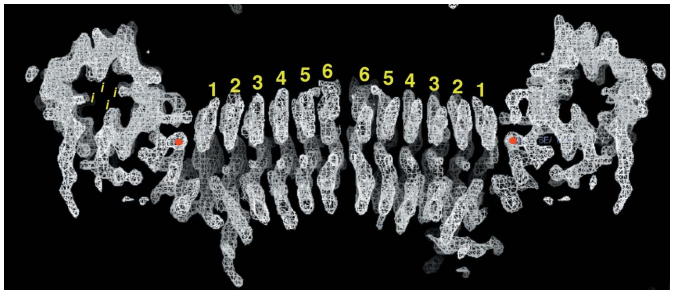


Figure 3
The F_o electron-density map for Rfr23 contoured at 1.9σ. One significant heavy-atom site per molecule was identified by the program *SOLVE* and is shown in red for two horizontally orientated molecules that pack end-to-end in the crystal lattice. Six complete or partial coils, numbered 1–6, characteristic of Rfr-folds are clearly visible for both molecules. At both corners of the figure splices through two molecules packed perpendicularly to the two horizontal molecules are shown. These splices illustrate the characteristic four-faced nature of the Rfr-fold. For the splice on the left, the central bulky side chain on each face is indicated by an '*i*'.

2.1 Å resolution, respectively, were collected from the blade-like crystals shown in the inset of Fig. 2.

Using the phase information in the SAD data, an electron-density map for Rfr23 with a clear solvent–protein boundary was generated using the program *RESOLVE* (Terwilliger, 2000). Assuming the presence of one molecule (150 residues, 16.4 kDa) per asymmetric unit, the calculated Matthews coefficient (Matthews, 1968) V_M is 4.6 Å^3 Da^{-1}, indicating a solvent content of 73%. Such a high solvent content dictates that the crystal packing of Rfr23 is not very tight and this is apparent in Fig. 3, which shows part of an F_o electron-density map at a 1.9σ contour level.

Rfr23 contains two SeMet residues, Met26* and Met74, and Fig. 3 also illustrates the position of the one significant heavy-atom site per molecule identified by the program *SOLVE* (Terwilliger & Berendzen, 1999). Because both SeMet sites are located towards the N-terminus of the protein sequence, the heavy-atom sites identify the N-termini and reveal that two molecules of Rfr23, with a twofold rotation axis, pack at the C-termini. Immediately apparent in the electron-density map is the regular shape of the Rfr-fold: regularly stacked coils with a rise of ∼4.8 Å along the helical axis. Six complete or partial coils are numbered in yellow for each of the two molecules packed horizontally in Fig. 3. Above the horizontally stacked molecules are two other molecules aligned perpendicularly, showing an electron-density slice through the Rfr-fold. These slices have the characteristic four-face Rfr-fold pattern (Vetting *et al.*, 2006) schematically illustrated in Fig. 1, with the bulky central residue, *i* (typically Leu or Phe), of the pentapeptide repeat directed towards the center of the column.

In summary, from a preliminary analysis of the electron-density map generated from SAD data for Rfr23, it is clear that the pentapeptide repeats in Rfr23 adopt an Rfr-fold similar to those observed in MfpA and Rfr32 (Hegde *et al.*, 2005; Buchko *et al.*, 2006). Model building and structure refinement are currently in progress to generate a structure using the SAD data collected to a resolution of 2.5 Å and native data collected to a resolution of 2.1 Å. Preliminary tests indicated that a molecular-replacement solution could also be found for the Rfr23 data sets using the core Rfr-fold of Rfr32 (PBD code 2g0y), illustrating that molecular replacement may be a useful alternative to SAD/MAD methods for determining future crystal structures of other PRPs.

This work is part of a Membrane Biology EMSL Scientific Grand Challenge project at the W. R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by US Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory (PNNL). PNNL is operated for the US Department of Energy by Battelle. The authors would like to thank Dr Enoch P. Baldwin at U. C. Davis for useful discussions and the assistance of the X29A beamline scientists at the National Synchrotron Light Source at Brookhaven National Laboratory. Support for beamline X29A at the National Synchrotron Light Source comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy and from the National Center for Research Resources of the National Institutes of Health. This manuscript has been authored by Battelle Memorial Institute, Pacific Northwest Division under Contract No. DE-AC05-76RL0 1830 with the US Department of Energy.

References

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). *Nucleic Acids Res.* **28**, 263–266.
Bateman, A., Murzin, A. G. & Teichmann, S. A. (1998). *Protein Sci.* **7**, 1477–1480.
Buchko, G. W., Ni, S., Robinson, H., Welsh, E. A., Pakrasi, H. B. & Kennedy, M. A. (2006). *Protein Sci.* **15**, 2579–2595.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Garrido, M. C., Herrero, R., Kolter, R. & Moreno, F. (1988). *EMBO J.* **7**, 1853–1862.
Gomi, M., Sonoyama, M. & Mitaku, S. (2004). *Chem-Bio Inform. J.* **4**, 142–147.
Hegde, S. S., Vetting, M. W., Roderick, S. L., Mitchenall, L. A., Maxwell, A., Takiff, H. E. & Blanchard, J. S. (2005). *Science*, **308**, 1480–1483.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Montero, C., Mateu, G., Rodriguez, R. & Takiff, H. E. (2001). *Antimicrob. Agents Chemother.* **45**, 3387–3392.
Morais Cabral, J. H., Jackson, A. P., Smith, C. V., Shikotra, N., Maxwell, A. & Liddington, R. C. (1997). *Nature (London)*, **388**, 903–906.
Morita, M., Tomita, K., Ishizawa, M., Tagaki, K., Kawamura, F., Takahashi, H. & Morino, T. (1999). *Biosci. Biotechnol. Biochem.* **63**, 563–566.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Vetting, M. W., Hegde, S. S., Fajardo, J. E., Fiser, A., Roderick, S. L., Takiff, H. E. & Blanchard, J. S. (2006). *Biochemistry*, **45**, 1–10.