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

The 1.7 Å resolution structure of At2g44920, a pentapeptide-repeat protein in the thylakoid lumen of *Arabidopsis thaliana*

Shuisong Ni,[a] Michael E. McGookey,[a] Stuart L. Tinch,[a] Alisha N. Jones,[a] Seetharaman Jayaraman,[b] Liang Tong[b] and Michael A. Kennedy[a]*

[a]Department of Chemistry and Biochemistry, Miami University, Oxford, OH 45056, USA, and [b]Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Correspondence e-mail: kennedm4@muohio.edu

Received 15 August 2011
Accepted 14 September 2011

PDB Reference: At2g44920, 3n90.

At2g44920 belongs to a diverse family (Pfam PF00805) of pentapeptide-repeat proteins (PRPs) that are present in all known organisms except yeast. PRPs contain at least eight tandem-repeating sequences of five amino acids with an approximate consensus sequence (STAV)(D/N)(L/F)(S/T/R)(*X*). Recent crystal structures show that PRPs adopt a highly regular four-sided right-handed β-helical structure consisting mainly of type II and type IV β-turns, sometimes referred to as a repeated five-residue (or Rfr) fold. Among sequenced genomes, PRP genes are most abundant in cyanobacteria, leading to speculation that PRPs play an important role in the unique lifestyle of photosynthetic cyanobacteria. Despite the recent structural characterization of several cyanobacterial PRPs, most of their functions remain unknown. Plants, whose chloroplasts are of cyanobacterial origin, have only four PRP genes in their genomes. At2g44920 is one of three PRPs located in the thylakoid lumen. Here, the crystal structure of a double methionine mutant of residues 81–224 of At2g44920, the naturally processed fragment of one of its full-length isoforms, is reported at 1.7 Å resolution. The structure of At2g44920 consists of the characteristic Rfr fold with five uninterrupted coils made up of 25 pentapeptide repeats and α-helical elements capping both termini. A disulfide bridge links the two α-helices with a conserved loop between the helical elements at its C-terminus. This structure represents the first structure of a PRP protein whose subcellular location has been experimentally confirmed to be the thylakoid lumen in a plant species.

1. Introduction

Pentapeptide-repeat proteins (PRPs) contain at least eight tandem-repeating sequences of five amino acids with an approximate consensus sequence (STAV)(D/N)(L/F)(S/T/R)(*X*) (Bateman *et al.*, 1998). Bioinformatic analysis has shown that there are about 500 unique PRP proteins in this diverse family, although several thousand PRP-containing genes have been identified in Pfam00805 before redundancy is excluded (Vetting *et al.*, 2006). PRP proteins are present in almost all organisms, with the exception of yeast. Most organisms have only a few PRPs, but cyanobacterial species contain far more PRPs, sometimes dozens, implying their importance in the unique lifestyle of photosynthetic cyanobacteria. Recent crystal structures show that PRPs adopt a highly regular four-sided right-handed β-helical structure that can be described as a collection of type II and/or type IV β-turns, sometimes referred to as a repeated five-residue or Rfr fold (Vetting *et al.*, 2006; Buchko *et al.*, 2006).

The biochemical functions of most PRPs in this family remain largely unknown, but their functions are believed to be diverse. The first experimentally identified PRP was HglK (all0813) from the filamentous cyanobacterium *Nostoc* sp. 7120 (Black *et al.*, 1995). HglK, one of more than 30 PRPs in this organism, is an integral membrane protein, with its C-terminal domain predicted to contain 36 tandem pentapeptide repeats. This PRP-containing domain is essential for localizing or transporting the glycolipid components required for maturation of heterocysts, specialized cyanobacteria cells that terminally differentiate to carry out fixation of atmospheric N_2 when there are insufficient sources of soluble nitrogen-containing compounds in the growth environment. Overexpression of another PRP, HetL (all3740), has also been shown to be involved in the

regulation of heterocyst formation in *Nostoc* sp. 7120 (Liu & Golden, 2002). The physiological function of another cyanobacterial PRP, RfrA from the nonfilamentous *Synechocystis* sp. PCC 6803, has been determined to be a role in manganese uptake (Chandler *et al.*, 2003). Perhaps the most intriguing function of a PRP was elucidated for MfpA (Hegde *et al.*, 2005) from *Mycobacterium tuberculosis*; not only did the crystal structure establish the precise nature of the PRP fold as a right-handed quadrilateral β-helix, correcting the original prediction that this family of proteins would adopt a triangular-shaped right-handed β-helical structure, but the associated experimental studies also provided direct evidence of the important biochemical function of this PRP *in vivo*. It has been shown that the antibiotic resistance of *M. tuberculosis* results from the function of MfpA, which inhibits DNA gyrase by acting as a DNA mimic that binds to DNA gyrase and thus confers antibiotic resistance against the fluoroquinolone class of antibiotics, which are normally quite effective bactericidal agents. The dimeric MfpA is an effective DNA mimic because its right-handed quadrilateral β-helical structure exhibits a size, shape and electrostatic surface charge similar to those of right-handed DNA. Several PRP proteins that carry out similar DNA-mimicking functions have been characterized (Vetting *et al.*, 2009, 2011; Hegde *et al.*, 2011).

However, DNA mimicking does not seem to be the function of most cyanobacterial PRPs based on their varied predicted subcellular locations, leaving the function of most PRPs poorly understood or not understood at all. The increasing number of X-ray crystal structures of PRPs from *Cyanothece* 51142 and *Nostoc* sp. PCC 7120 (Ni *et al.*, 2009; Buchko *et al.*, 2006, 2008) makes it possible to refine our understanding of the precise nature of the Rfr fold, as well as the types of sequence and structural variations that occur in PRPs; however, these structures have not yielded insight into the precise biochemical functions of these proteins in cyanobacteria. Unfortunately, the abundance of PRPs present in cyanobacteria renders genetic functional analysis less effective because attempts to identify the function of PRPs using gene-knockout approaches are complicated by the potential for complementation by other PRPs within the same organism. This might be particularly true for single-domained PRPs. HetL knockout mutants, for instance, do not show any distinct phenotypes in heterocyst differentiation (Liu & Golden, 2002).

Plants, whose chloroplasts are of cyanobacterial origin, only retain four PRP genes in their entire genomes. Their functions are unknown, but all of them are highly conserved in the plant genomes sequenced to date. Three of these PRPs in *Arabidopsis thaliana* (At2g44920, At5g53490 and At1g12250) are located in the chloroplast thylakoid lumen (Kieselbach, Mant *et al.*, 1998; Schubert *et al.*, 2003), which refers to the interior aqueous space enclosed by the elaborate continuous thylakoid membrane that is organized into granna and lamella, where oxygenic photosynthesis occurs (Kieselbach, Hagman *et al.*, 1998). In general, the lumen is believed to play an important role in accumulating protons and balancing ion currents for establishment of the proton motive force over the thylakoid membrane during photosynthesis (Peltier *et al.*, 2002). While protein complexes of thylakoid membrane directly involved in photosynthesis have been studied in great detail, the soluble nonmembrane lumen proteins have received little attention. To date, about 80 proteins have been experimentally identified to function in the thylakoid lumen (van Wijk, 2004; Zybailov *et al.*, 2008). As more lumenal proteins have been discovered, their functions have expanded to include assistance in folding and proteolysis of thylakoid proteins as well as protection against oxidative stress (Sirpiö *et al.*, 2007). All three thylakoid-localized PRP genes are nucleus-coded and their precursor proteins are targeted for the chloroplast and then translocated into thylakoid lumen *via* the SEC protein-translocation pathway (Zybailov *et al.*, 2008; Agrawal *et al.*, 2005). Microarray data have shown that these PRPs are primarily expressed in the leaves. *At2g44920* gene expression is closely correlated with that of *At5g53490* and their expression is regulated by alternative splicing to generate isoforms (Barbazuk *et al.*, 2008). Two isoforms of *At2g44920* (*At2g44920.1* and *At2g44920.2*) have been identified (Agrawal *et al.*, 2005). Here, we report the crystal structure of At2g44920.2, one of the three lumenal PRPs in *A. thaliana*.

2. Materials and methods

2.1. Cloning, expression and purification

The *At2g44920.2* gene was amplified from cDNA reverse transcribed (cDNA synthesis kit, Roche Inc.) from total mRNA isolated from the leaves of *A. thaliana* Columbia using the following primers: 5′-ATCGATCGCATATGGTAATTCTCAGCAACGTCTC and 5′-TGACTCTCGAGGTTGCAAAGCAATGTGTCCCG. The resulting PCR product was digested with *Nde*I and *Xho*I and cloned into the expression vector pET28b (Novagen Inc.). The double methionine mutant (V173M/T174M), hereafter simply referred to as At2g44920, was generated using a QuikChange II XL kit (Stratagene, Agilent Technologies, California, USA) according to the manufacturer's protocol. This construct allowed the expression of At2g44920 with C-terminal 6×histidine (6×His) tags at both the N- and C-termini. The plasmid was then transformed into the host strains *Escherichia coli* BL21 (DE3) and B834 (DE3) (Novagen) for the expression of native and selenomethionine-labeled protein, respectively. The culture was grown at 310 K with vigorous shaking to an OD_{600} of ∼0.8 in 1 l ZYP-5052 auto-induction medium supplemented with 30 µg ml^{-1} kanamycin. For selenomethionine (SeMet) labeling, M9 minimal medium was supplemented with a mixture of 17 amino acids (no Cys, Tyr or Met). Filter-sterilized SeMet was added just prior to inoculation. The final concentration of each amino acid was 100 mg l^{-1}. Protein expression was induced at 301 K overnight. The cells were then harvested and stored at 193 K. Thawed cells were resuspended in 25 ml lysis buffer (0.25 *M* NaCl, 20 m*M* Tris pH 7.8, 10% glycerol). Phenylmethylsulfonyl fluoride was added to the cell suspension to a final concentration of about 0.2 µ*M* immediately prior to cell lysis by three passes through a French press (Thermo Inc.). The cell lysate was spun at 24 000*g* for 60 min. The supernatant was loaded onto a 10 ml Ni–NTA affinity column (Qiagen) and washed stepwise with 50 ml buffer (0.25 *M* NaCl, 20 m*M* Tris pH 7.8, 10% glycerol) containing 0 and 30 m*M* imidazole. At2g44920 protein was eluted from the Ni–NTA column using 300 m*M* imidazole in the starting buffer. Purified At2g44920 was concentrated to about 10 mg ml^{-1} with an Amicon Ultra-10 (Millipore) and further purified on a Superdex 200 HiLoad (GE Life Science) size-exclusion column equilibrated with a buffer consisting of 0.25 *M* NaCl, 20 m*M* Tris pH 7.8 and 10% glycerol. The eluted fractions were combined and concentrated to 5.0, 7.5, 10 and 15 mg ml^{-1} for crystallization screening.

2.2. Crystallization, data collection, structure solution, phasing and refinement

The hanging-drop vapor-diffusion method was used for all crystal screenings. Drops were set up by mixing 1 µl purified At2g44920 with 1 µl of each reservoir buffer from Hampton Research screening kits on a 48-well plate and incubated at room temperature (about 295 K). The final experimental crystal was obtained in buffer consisting of 2.5 *M* ammonium sulfate and 0.5 *M* sodium acetate pH 5.2 and was

briefly dipped into the same buffer containing 20% glycerol before being flash-cooled in liquid nitrogen. The crystals were screened and initially characterized using Cu $K\alpha$ radiation from a Bruker Microstar rotating-anode generator equipped with Montel optics and a SMART 6000 CCD. Anomalous data were collected from SeMet-labeled crystals at the Se peak on beamline 24IDE equipped with an ADSC Quantum 315 CCD at the Advanced Photon Source (APS), Argonne National Laboratory. 360 images were collected using 0.5° oscillation during 20 s exposures. The images were integrated and scaled with *HKL*-2000 (Otwinowski & Minor, 1997). The program *SOLVE* (Terwilliger & Berendzen, 1999) was used to find the heavy-atom locations, which were used to generate a SAD map at 1.7 Å resolution from the Se anomalous data set. Iterative model building and refinement were performed manually using *XtalView*/*Xfit* (McRee, 1999) and *REFMAC* (v.5.2.0019; Murshudov *et al.*, 2011). The resolution cutoff for final structure refinement was fixed at 1.7 Å, where the completeness in the highest resolution shell exceeded 90%. The structure quality was assessed using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010). The structure was submitted to the Protein Data Bank (PDB entry 3n90).

3. Results and discussion

Mature At2g44920 was detected in the thylakoid lumen in *A. thaliana* as a 15 kDa protein in recent proteomic studies (Kieselbach, Mant *et al.*, 1998; Kieselbach, Hagman *et al.*, 1998; Schubert *et al.*, 2003). The cytosolic precursor of At2g44920 (23.8 kDa) carries a cleavable N-terminal chloroplast transit peptide signal that is recognized and removed by the protein-transport machinery during translocation. The cleavage site is predicated to be Ala81. This prediction has been confirmed by Edman N-terminal sequencing of the mature protein isolated by two-dimensional gel electrophoresis (Peltier *et al.*, 2002). However, proteomic analysis identified the cleavage site at Leu128 (Zybailov *et al.*, 2008). Without knowing the exact cleavage site, we initially cloned the full-length gene as an *Nde*I–*Xho*I insert into pET28b expression vector with and without the C-terminal stop codon. The expression construct that carries the stop codon allowed expression of At2g44920 with only an N-terminal 6×His tag, while the other construct without the stop codon produced the protein with 6×His tags at both ends. Both proteins were expressed in *E. coli* with similar yield. However, At2g44920 with the N-terminal His tag did not bind to the Ni resin during purification, indicating that the tag was removed, as expected, together with the signal peptide. Thus, the expression construct without the stop codon was used to express C-terminally His-tagged At2g44920 throughout this study. Interestingly, the crystal structure analysis later showed that the structured part of this protein started at Ala81, which is in good agreement with the above-mentioned Edman analysis.

Wild-type native crystals of At2g44920 typically diffracted to about 2.2–2.5 Å resolution with a mosaicity of about 0.7°. Indexing showed an orthorhombic lattice, with unit-cell parameters a = 58.7, b = 61.1, c = 75.9 Å, $\alpha = \beta = \gamma = 90°$. Highly redundant sulfur anomalous scattering data sets were collected. Owing to the low sulfur content (two cysteines and no internal methionines), initial phase determination by sulfur SAD failed. Molecular replacement with the crystal structure of RFR32, a predicted homolog from *Cyanothece* 51142, also did not yield an interpretable map. However, amino-acid sequence alignment against RFR32 revealed potential mutation sites (Val173 and Val174) for the introduction of methionine residues into the predicted structured region of the protein. It turned out that the double Met mutations (V173M/T174M) not only made SeMet labeling possible, but also improved the resolution of crystal diffraction to 1.5–1.7 Å. The SeMet-labeled mutant crystals retained the same orthorhombic crystal lattice, with similar unit-cell parameters, compared with the wild-type native crystals (Table 1).

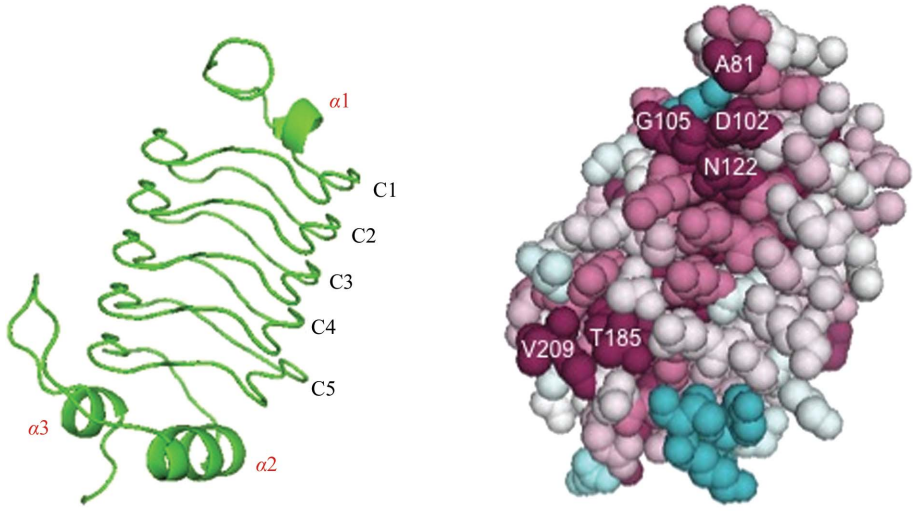
Table 1
Summary of data-collection and structure-refinement statistics for At2g44920 double methionine mutant (PDB entry 3n90).

Values in parentheses are for the highest resolution shell.

Data collection	
Crystal	SeMet At2g44920 V173M/T174M mutant
Data set	SAD
Space group	$C222_1$
Unit-cell parameters (Å, °)	a = 59.064, b = 75.444, c = 59.901, $\alpha = \beta = \gamma = 90$
Matthews coefficient (Å^3 Da^{-1})	2.2
Solvent content (%)	44.7
X-ray source	APS 24IDE
Temperature (K)	100
Resolution limits (Å)	50–1.7 (1.79–1.70)
Mosaicity (°)	0.66
Wavelength (Å)	0.979
No. of unique reflections	14215 (2143)
Multiplicity	1.87 (1.92)
R_{merge} (%)	8.1 (3.3)
Completeness (%)	99.2 (100.0)
$\langle I/\sigma(I)\rangle$	40.84 (18.9)
Structure refinement	
R_{conv} (%)	22.6
R_{free} (%)	25.85
Protein atoms	1180
No. of water molecules	103
No. of amino-acid residues	145
Average B values (Å^2)	
Main-chain protein atoms	13.8
Side-chain and solvent atoms	21.2
Overall	20.3
R.m.s. deviations from ideal	
Bond lengths (Å)	0.015
Bond angles (°)	1.6
Ramachandran plot analysis (*PROCHECK*)	
Most favored regions (%)	80.5
Additionally allowed regions (%)	19.5
Generously allowed regions (%)	0.0
Disallowed regions (%)	0.0
Ramachandran plot analysis from Richardson laboratory	
Most favored regions (%)	97.2
Allowed regions (%)	2.8
Disallowed regions (%)	0.0
PROCHECK G factor (φ/ψ only)	−0.53
PROCHECK G factor (all dihedral angles)	−0.22
Verify3D	0.49
Prosa II (−ve)	0.40
MolProbity clash score	6.60

The overall structure of At2g44920 (Fig. 1*a*), which was of high quality based on standard structure-quality factors and *MolProbity* (Table 1), adopted the characteristic Rfr-fold core, as expected based on its PRP sequence pattern (Supplementary Fig. S1[1]), consisting of a right-handed quadrilateral β-helix and α-helical elements capped at both termini. The core Rfr fold has five uninterrupted coils that are made up of 25 pentapeptide repeats (Supplementary Fig. S1[1]). As observed for other reported PRP structures, the interior of the Rfr fold in At2g44920 is highly hydrophobic, with stacked hydrophobic side chains (Fig. 2) pointing inwards from the central amino-acid residue, often designated as the i residue, and the first amino acid, designated as $i - 2$, of each pentapeptide repeat. The i position is mostly occupied by Leu and Phe, with occasional occurrence of Val, Ile or Met, while the $i - 2$ position is mostly occupied by Ala. Although the stacking of these hydrophobic residues in the interior undoubtedly plays an important structural role in stabilizing the Rfr

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5196).



(*a*) (*b*)

Figure 1
(*a*) Ribbon diagram of the crystal structure of At2g44920. (*b*) *ConSurf* analysis plot of the crystal structure of At2g44920 shown in the same orientation as in (*a*). Some conserved residues are labeled and the location in the sequence/secondary structure can be seen in Fig. 2. The sequence alignment used for the *ConSurf* analysis is shown in Supplementary Fig. S2. Highly conserved residues are colored dark purple and labeled in the figure. A summary of the results of the *ConSurf* analysis is included in Supplementary Table S1.

fold, it is less clear whether this hydrophobic interior can serve in any important functional role such as a transport channel of some sort. Interestingly, submission of the structure to the ^{3}V *Voss Volume Voxelator* server (Voss *et al.*, 2006) failed to identify any channel in the interior of the Rfr fold. However, the characteristics of the PRP protein exterior surface presumably play an important role in its function, as shown by the role that the surface-charge distribution of MfpA plays in mimicking DNA. Such surface properties are largely determined by the side chains of the rest of the amino acids ($i - 1$, $i + 1$ and $i + 2$) within each pentapeptide repeat, all of which point outwards. The amino-acid identities at these positions are not conserved, which results in variation of surface properties in PRPs. Although At2g44920 does not have a surface-charge distribution that mimics that of double-stranded DNA like MfpA, it does have large patches of negative charge on one side and patches of positive charge on the opposite side (Supplementary Fig. S3). All PRP crystal structures reported to date have α-helical elements at either the C- or the N-terminus, although their positions and orientations relative to the core vary. In MfpA the C-terminal helix acts as the dimerization interface, providing an important functional role in DNA gyrase dimer binding. In contrast, size-exclusion chromatographic analysis indicated that At2g44920 was monomeric in solution (Supplementary Fig. S4), further supporting the assumption that the function of At2g44920 does not involve DNA mimicry. Interestingly, At2g44920 has an alternatively spliced isoform that not only lacks the final 14 amino acids, which are part of the α3 helix in the full-length protein, but also differs in the region of the α2 helix and the unstructured loop. In our full-length crystal structure, the two α-helices are linked by a cysteine disulfide bridge with an unstructured loop in between. Although it is still unknown why *Arabidopsis* expresses two splicing isoforms of At2g44920, the shorter isoform (At2g44920.1) would lack the α3 helix and possibly also the unstructured loop. *ConSurf* analysis (Ashkenazy *et al.*, 2010; Fig. 1*b*) showed that there were several conserved residues in this loop region that might be functionally important. Structural alignment using *DALI* (Holm & Rosenström, 2010) revealed that the At2g44920.2 structure is similar to those of RFR32 (r.m.s.d. of 1.3 Å) and RFR23 (r.m.s.d. of 1.6 Å), both of which are predicted to be located in the lumen of *Cyanothece* 51142. However, it differs from RFR23 in that RFR23 has a larger loop, located at its N-terminus, compared with the loop occurring at the C-terminus in RFR32 and At2g44920. Neither of the cyanobacterial homologs have the bulge at coil 4 observed in At2g44920, which arises from the presence of a one-residue insert in the standard repeat sequence. The At2g44920 structure reported here adds to the growing

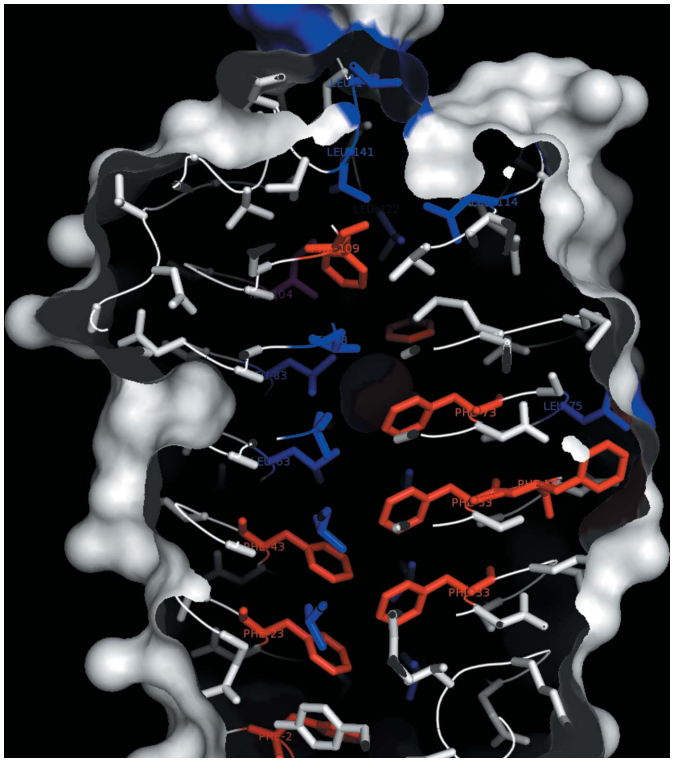


Figure 2
Cutaway cross-section of the interior structure of At2g44920, highlighting the hydrophobic side chains, rendered in stick representation, that compose the hydrophobic core of the Rfr fold. The coloring scheme is as follows: Phe, red; Leu, blue; Ile, purple. Highlighted residues are also labeled.

library of PRP structures and should eventually help to elucidate the precise biochemical function of this plant protein.

References

Agrawal, G. K., Yonekura, M., Iwahashi, Y., Iwahashi, H. & Rakwal, R. (2005). *J. Chromatogr. B*, **815**, 125–136.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). *Nucleic Acids Res.* **38**, W529–W533.

Barbazuk, W. B., Fu, Y. & McGinnis, K. M. (2008). *Genome Res.* **18**, 1381–1392.

Bateman, A., Murzin, A. G. & Teichmann, S. A. (1998). *Protein Sci.* **7**, 1477–1480.

Black, K., Buikema, W. J. & Haselkorn, R. (1995). *J. Bacteriol.* **177**, 6440–6448.

Buchko, G. W., Ni, S., Robinson, H., Welsh, E. A., Pakrasi, H. B. & Kennedy, M. A. (2006). *Protein Sci.* **15**, 2579–2595.

Buchko, G. W., Robinson, H., Pakrasi, H. B. & Kennedy, M. A. (2008). *J. Struct. Biol.* **162**, 184–192.

Chandler, L. E., Bartsevich, V. V. & Pakrasi, H. B. (2003). *Biochemistry*, **42**, 5508–5514.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Hegde, S. S., Vetting, M. W., Mitchenall, L. A., Maxwell, A. & Blanchard, J. S. (2011). *Antimicrob. Agents Chemother.* **55**, 110–117.

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.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Kieselbach, T., Hagman, Andersson, B. & Schröder, W. P. (1998). *J. Biol. Chem.* **273**, 6710–6716.

Kieselbach, T., Mant, A., Robinson, C. & Schröder, W. P. (1998). *FEBS Lett.* **428**, 241–244.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Liu, D. & Golden, J. W. (2002). *J. Bacteriol.* **184**, 6873–6881.

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Ni, S., Sheldrick, G. M., Benning, M. M. & Kennedy, M. A. (2009). *J. Struct. Biol.* **165**, 47–52.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Peltier, J. B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D. A., Söderberg, L., Roepstorff, P., von Heijne, G. & van Wijk, K. J. (2002). *Plant Cell*, **14**, 211–236.

Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schröder, W. P. & Kieselbach, T. (2003). *J. Biol. Chem.* **278**, 13590.

Sirpiö, S., Allahverdiyeva, Y., Suorsa, M., Paakkarinen, V., Vainonen, J., Battchikova, N. & Aro, E. M. (2007). *Biochem. J.* **406**, 415–425.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Vetting, M. W., Hegde, S. S. & Blanchard, J. S. (2009). *Acta Cryst.* D**65**, 462–469.

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.

Vetting, M. W., Hegde, S. S., Zhang, Y. & Blanchard, J. S. (2011). *Acta Cryst.* F**67**, 296–302.

Voss, N. R., Gerstein, M., Steitz, T. A. & Moore, P. B. (2006). *J. Mol. Biol.* **360**, 893–906.

Wijk, K. J. van (2004). *Plant Physiol. Biochem.* **42**, 963–977.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q. & van Wijk, K. J. (2008). *PLoS One*, **3**, e1994.