Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the RuBisCO chaperone RbcX from *Synechocystis* sp. PCC6803

Shiho Tanaka,[a] Michael R. Sawaya,[b] Cheryl A. Kerfeld[c] and Todd O. Yeates[a,b,d]*

[a]UCLA Department of Chemistry and Biochemistry, 611 Charles Young Drive East, Los Angeles, CA 90095-1569, USA, [b]UCLA–DOE Institute for Genomics and Proteomics, Los Angeles, CA 90095-1570, USA, [c]DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, and [d]UCLA Molecular Biology Institute, Los Angeles, CA 90095-1570, USA

Correspondence e-mail: yeates@mbi.ucla.edu

Received 6 July 2007
Accepted 29 August 2007

PDB Reference: RbcX, 2py8, r2py8sf.

In some cyanobacteria, the genes for the large and small subunits of the enzyme RuBisCO are separated on the bacterial chromosome by the insertion of a gene coding for a protein designated RbcX, which acts as a chaperone for RuBisCO. A recent structural study [Saschenbrecker *et al.* (2007), *Cell*, **129**, 1189–1200] has shed light on the mechanism by which RbcX assists RuBisCO assembly. Here, the crystal structure of RbcX from another cyanobacterium, *Synechocystis* sp. PCC6803, is reported, revealing an unusually long protruding C-terminal helix, as well as a bound polyethylene glycol molecule in the protein substrate-binding site.

1. Introduction

While many proteins are able to fold to their native configurations without assistance, others require the participation of chaperone proteins. A variety of chaperone systems have been characterized. Some exhibit broad substrate specificities, while others have apparently evolved to fold specific target proteins (Bochkareva *et al.*, 1988; Arie *et al.*, 2001). One protein that requires a chaperone in order to fold properly is ribulose bisphosphate carboxylase/oxygenase (RuBisCO), which is one of the most abundant enzymes on the planet. In plants, the chaperone role is played by RuBisCO-binding protein cpn60 (Ellis & van der Vies, 1988; Gatenby & Ellis, 1990). In the case of cyanobacteria, protein-expression studies in *Escherichia coli* have shown that the assembly of cyanobacterial RuBisCO is assisted by DnaK and GroEL/GroES (Checa & Viale, 1997; Goloubinoff *et al.*, 1989). However, in some cyanobacteria (particularly the β-cyanobacteria, which contain form 1B RuBisCO), a distinct protein, RbcX, appears to act as a chaperone to assist RuBisCO folding. RbcX is a protein of 15–16 kDa. In β-cyanobacteria, it is typically encoded on the bacterial chromosome between the genes for the large and small RuBisCO subunits RbcL and RbcS (Onizuka *et al.*, 2004; Li & Tabita, 1997), although it is encoded at a distant chromosomal location in at least one exceptional case, *Synechococcus* sp. PCC7942 (Emlyn-Jones *et al.*, 2006). A recent structural and functional study of RbcX from *Synechococcus* sp. PCC7002 has revealed that RbcX assists RuBisCO folding and assembly by binding to the C-terminus of partially unfolded RbcL subunits before they assemble with RbcS subunits into the native hexadecameric L_8S_8 configuration (Saschenbrecker *et al.*, 2007). Here, we report the crystal structure of RbcX from *Synechocystis* sp. PCC6803. This new RbcX structure includes a longer C-terminal helix and a molecule of PEG in the binding site. Structural features reminiscent of other known chaperones are noted.

2. Experimental procedures

2.1. Cloning, expression, purification and crystallization of RbcX

The *Synechocystis* sp. PCC6803 RbcX gene was cloned into the pET22b expression vector (Novagen). The native protein and a selenomethionine (SeMet) derivative of RbcX were expressed in BL21-Gold(DE3) *E. coli* cells (Stratagene) and purified using an Ni–NTA column (Qiagen). Crystals of the native RbcX protein were grown using the hanging-drop vapor-diffusion method in 30% PEG 400, 0.1 *M* potassium/sodium phosphate pH 6.4 and 0.2 *M* NaCl.

Table 1
Summary of data-collection and refinement statistics.

Values in parentheses are for the last shell.

	Native	Se (peak)	Se (inflection)†	Se (remote)
Data collection				
Wavelength (Å)	1.000	0.97949	0.97969	0.97182
Space group	$P4_32_12$			
Unit-cell parameters (Å, °)	$a = b = 129.8$, $c = 92.7$, $\alpha = \beta = \gamma = 90.0$			
Resolution range (Å)	19.496–2.45	90.0–3.3	90.0–3.3	90.0–3.3
R_{sym}‡ (%)	4.2 (46.6)	6.9 (43.4)	7.7 (50.5)	7.0 (53.7)
$I/\sigma(I)$, last shell	4.3	4.8	3.5	3.7
Total observations	228143	168367	84452	170353
Unique reflections (Bijvoet pairs separate)	29666	22720	12475	22996
Completeness (%)	99.9 (100)	100 (100)	100 (100)	100 (100)
Phase determination				
R_{cullis}§ (20–2.4 Å, acentric/centric) (%)	N/A	0.98/0.96	N/A	0.97/0.92
R_{cullis}¶ (20–2.4 Å, anomalous) (%)	N/A	0.82	0.95	0.89
Phasing power†† (20–2.6 Å, acentric/centric)	N/A	0.43/0.31	N/A	0.48/0.37
No. of sites	N/A	5		
Mean overall figure of merit (before/after DM)		0.327/0.744		
Refinement statistics				
R_{work}‡‡ (19.5–2.45 Å) (%)			20.9	
R_{free}§§ (19.5–2.45 Å) (%)			24.9	
No. of residues (protein/water)			460/99	
Mean B factor (Å^2)			46.1	
R.m.s.d. bond length (Å)			0.010	
R.m.s.d. bond angle (°)			1.1	
Ramachandran plot statistics (%)				
Residues in most favored regions			95.8	
Residues in additional allowed regions			4.2	
Residues in generously allowed regions			0.0	
Residues in disallowed regions			0.0	

† The inflection data set was treated as a reference for phasing. Statistics are reported to 3.3 Å resolution. ‡ $R_{sym} = \sum|I - \langle I\rangle|^2/\sum I^2$. § $R_{Cullis} = \sum\varepsilon/\sum|F_{PH} - F_P|$, where ε is the lack of closure. ¶ $R_{Cullis} = \sum\varepsilon/\sum|F^+ - F^-|$, where ε is the lack of closure. †† Phasing power = $\langle F_H/\varepsilon\rangle$. ‡‡ $R_{work} = \sum|F_{obs} - F_{calc}|/\sum F_{obs}$. §§ $R_{free} = \sum|F_{obs} - F_{calc}|/\sum F_{obs}$, where all reflections belong to a test set consisting of a randomly selected 5% of the data.

Crystals of SeMet RbcX protein were grown in 30% PEG 400, 0.1 *M* potassium/sodium phosphate pH 6.4, 0.2 *M* lithium sulfate at room temperature.

2.2. Data collection, processing and structure determination

The structure of RbcX reported here from *Synechocystis* sp. PCC6803 was determined independently of the publication of the *Synechococcus* sp. PCC7002 structure (Saschenbrecker *et al.*, 2007), requiring phasing by SeMet MAD methods. A standard three-wavelength anomalous dispersion data set was collected from a selenomethionyl derivative at the Advanced Light Source, beamline 8.2.2. An ADSC Quantum 315 CCD detector was used to record the data. Diffraction data extending to 2.45 Å were processed using *DENZO*/*SCALEPACK* (Otwinowski & Minor, 1997). Five selenium sites were identified using the program *SHELXD* (Sheldrick & Schneider, 2001). Phases were calculated with *MLPHARE* and *DM* (Collaborative Computational Project, Number 4, 1994). An initial model was built using the graphics program *O* (Jones *et al.*, 1991). The model was refined with *REFMAC* (Murshudov *et al.*, 1997) using strong NCS restraints throughout. Later rounds of model building were performed with the graphics program *Coot* (Emsley & Cowtan, 2004). The final model had R and R_{free} values of 20.9% and 24.9%, respectively. Protein structures were illustrated using *PyMOL* (DeLano, 2002). Data-collection and refinement statistics are given in Table 1.

2.3. Structural analysis

The *DALI* computer program (Holm & Sander, 1996) was used to search for proteins bearing structural similarity to RbcX.

Identification of RbcX homologues was performed with *BLAST* (Altschul *et al.*, 1990). *CD-HIT* (Li *et al.*, 2001) was used to select 11 of the most diverse homologs (70% conservation threshold) from a total of 135 homologous sequences output by a *BLAST* homology search. These 11 representative homologs were aligned using *MUSCLE* (Edgar, 2004) and displayed using *JalView* (Clamp *et al.*, 2004).

3. Results and discussion

3.1. The structure of *Synechocystis* sp. PCC6803 RbcX

The final model of *Synechocystis* sp. PCC6803 RbcX consists of four protein chains in the asymmetric unit of the crystal, arranged as a pair of dimers. The crystal structure suggested either a dimer or tetramer as the natural biological form of RbcX. Based on equilibrium sedimentation data (not shown) and the recent structure of *Synechococcus* sp. PCC7002 RbcX (Saschenbrecker *et al.*, 2007), the biological unit of RbcX was confirmed to be a dimer (Fig. 1*a*). In each chain, most of the 136 residues are visible in the electron-density map (chain *A*, 2–121; chain *B*, 2–115; chain *C*, 2–124; chain *D*, 2–104), with somewhat varying amounts of disorder in the C-termini, consistent with the earlier structural report of *Synechococcus* sp. PCC7002 (Saschenbrecker *et al.*, 2007). RbcX is nearly entirely α-helical. A particularly striking feature of its structure is the long protruding helices. The C-terminal helix extends approximately 18 Å beyond the core of the protein. The remaining 12 residues at the C-terminus are not visible in electron-density maps. This region contains primarily short polar side chains and is predicted to adopt a random-coil configuration by *PSIPRED* (McGuffin *et al.*, 2000). The protruding C-terminal helix visualized here is even longer (by approximately four residues) than that reported previously (Saschenbrecker *et al.*, 2007). This constitutes an unusual structural feature but its significance, if any, is unknown, as deleting residues 110–134 did not affect RbcX activity in an *in vitro* functional assay for RuBisCO assembly (Saschenbrecker *et al.*, 2007).

3.2. Binding sites of RbcX

In the crystal structure of *Synechocystis* sp. PCC6803 RbcX, two molecules of PEG 400 are bound in the cleft of the RbcX dimer in the same location as the C-terminus of RbcL was shown to bind in an extended conformation (Saschenbrecker *et al.*, 2007). The sequence conservation in RbcX was analyzed by multiple sequence alignment and positions conserved at a level of $\geq$90% were noted (Fig. 1*b*). Among these highly conserved residues, Tyr17 and Tyr20 are involved in binding the PEG molecules (Fig. 1*c*). These interactions mimic the biological interactions that RbcX makes with RbcL in the recent structure of *Synechococcus* sp. PCC7002 RbcX complexed with a peptide from the C-terminus of RbcL. RbcX is believed to be a specific chaperone for RuBisCO, so it is interesting to observe the fortuitous binding of the PEG molecule in the natural binding site.

The general characteristics of the RbcX cleft may be conducive to binding flexible polymer ligands.

3.3. Similar structures

Several programs were used to search the known database of protein structures for similarities to RbcX. One of these, *DALI*, identified the mitochondrial chaperone Tim9·10 as having the highest similarity. In a previous structural study, two other chaperones, prefoldin and Skp, were noted as being structurally similar to Tim9·10 (Webb *et al.*, 2006). Prefoldin binds and stabilizes nascent proteins until they are delivered to the chaperonin for completion of folding (Siegert *et al.*, 2000). Skp and Tim9·10 stabilize their substrate proteins in a soluble assembly-competent state before membrane insertion (Walton & Sousa, 2004; Webb *et al.*, 2006). Likewise, RbcX stabilizes the C-terminus of the RuBisCO large subunit prior to its oligomeric assembly (Saschenbrecker *et al.*, 2007). Although the degree of similarity between these chaperones is not high enough to

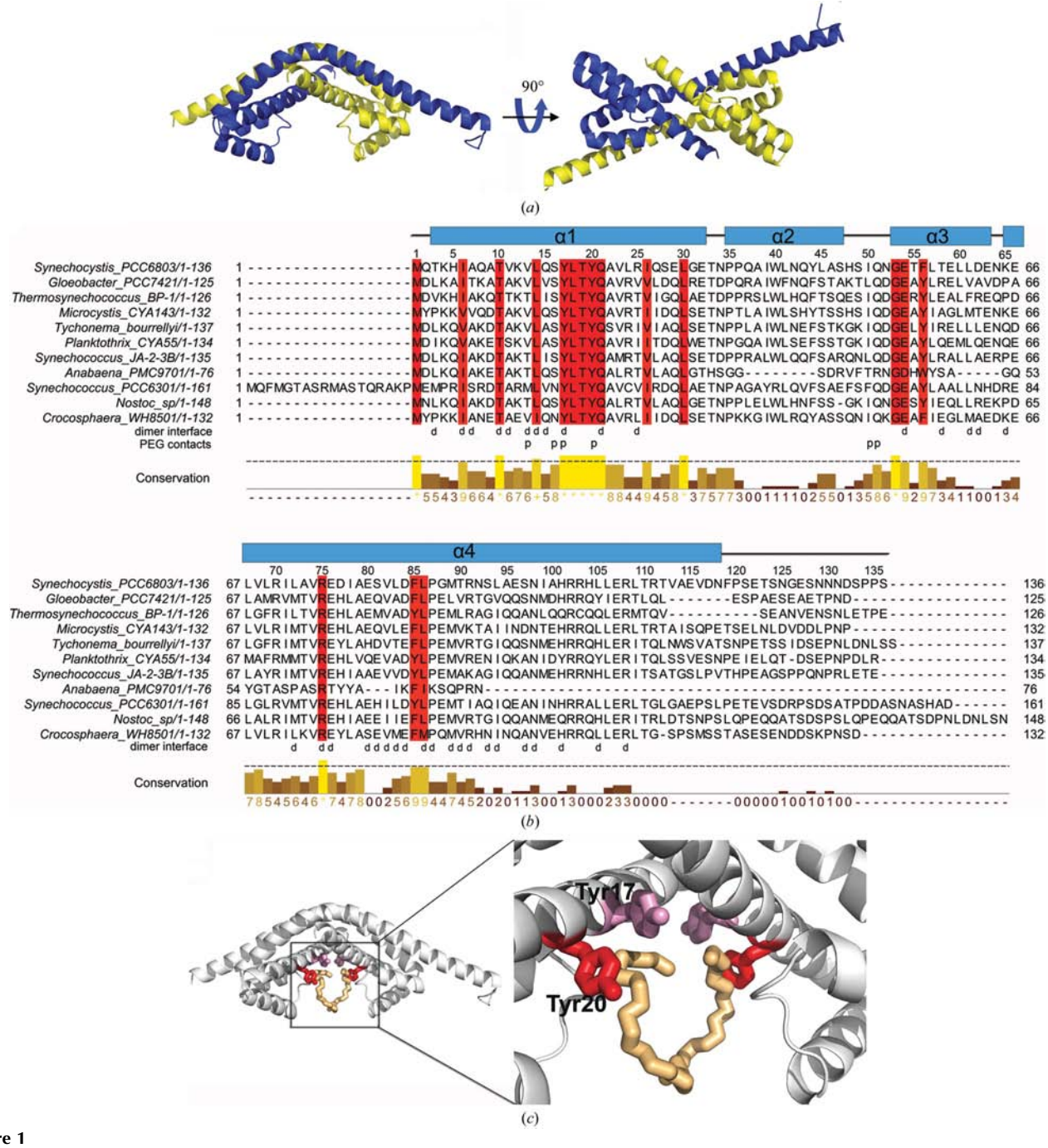

Figure 1
Overall structure of the chaperone protein RbcX and its binding site. (*a*) Cartoon diagram of the RbcX dimer shown in two different orientations to highlight the long protruding α-helices. (*b*) Multiple sequence alignment of 11 RbcX representatives. Conserved residues ($\geq$90%) are highlighted in red. Residues involved in dimerization and interactions with bound PEG molecules are indicated below the sequence alignment by 'd' and 'p', respectively. The plot below the alignment indicates the degree of conservation. The dotted line indicates the threshold of 90% conservation. (*c*) Residues in the RbcX dimer cleft that interact with bound PEG molecules are shown in a stick representation. PEG molecules are colored yellow, Tyr20 is colored red and Tyr17 is colored pink.

imply homologous relationships, the proteins do share some structural features, such as elongated helices and clamp-like shapes (Stirling *et al.*, 2006). Additionally, all four of these chaperones, including RbcX, have pronounced hydrophobic patches on the inward-facing parts of the clamp-like structures (Fig. 2). The roles of the hydrophobic patches have been studied extensively in prefoldin, where they were shown to be important in chaperone activity (Lundin *et al.*, 2004). The various chaperones discussed here seem to stabilize unfolded regions of their substrate proteins using common structural features, despite their different substrates and their apparently independent origins.

The authors thank Duilio Cascio for crystallographic assistance, Manfred Weiss at EMBL Hamburg for helpful discussions and the staff at Advanced Light Source beamline 8.2.2 for technical assistance. This work was funded by grant 2004-35318-14929 from the National Research Initiative program of the USDA and by the Biological and Environmental Research program of the Department of Energy Office of Science.

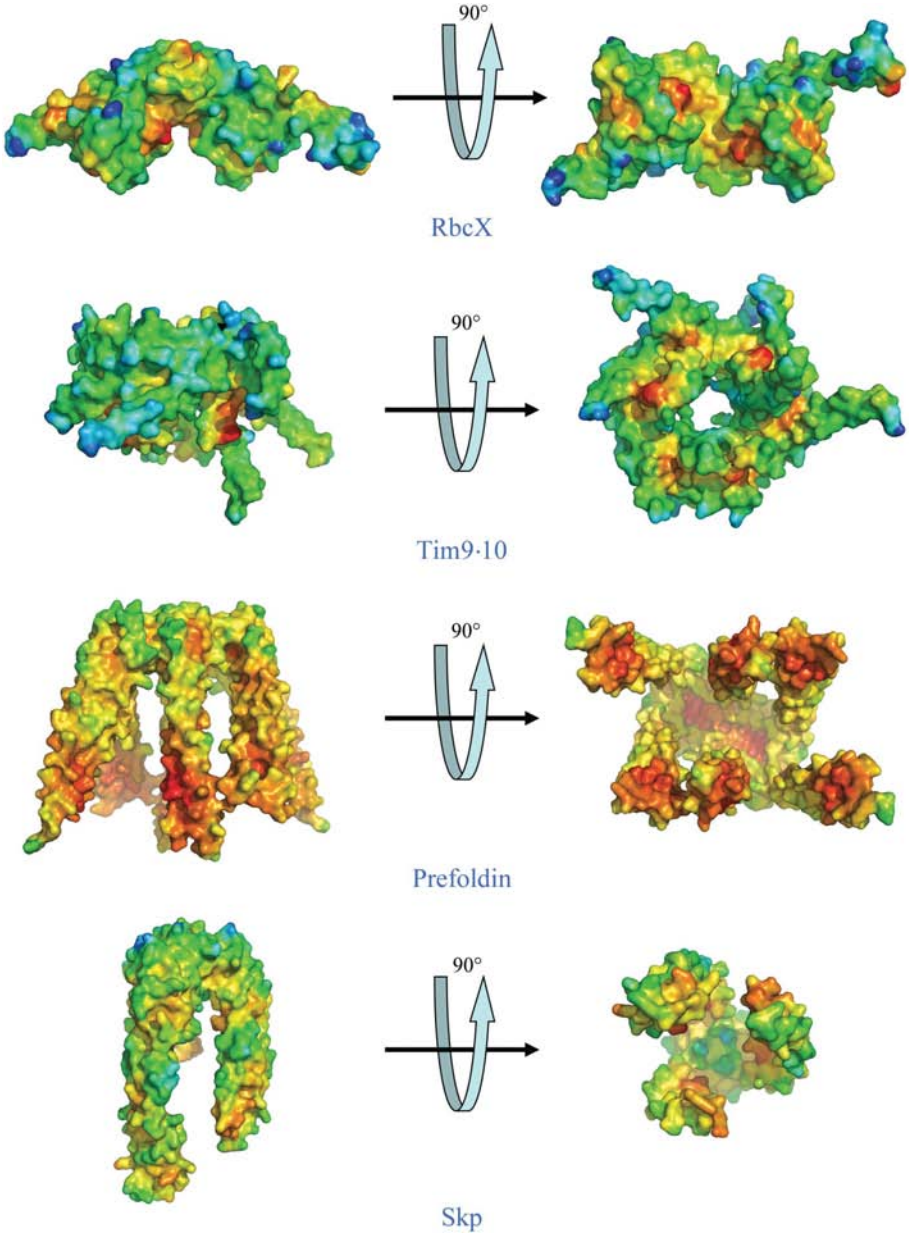

Figure 2
Common features in various chaperone structures. Four different chaperone structures are shown in which extended helices and hydrophobic patches (Pettit & Bowie, 2007) are evident. The color gradient indicates greater hydrophobicity in red and lower hydrophobicity in blue.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Arie, J. P., Sassoon, N. & Betton, J. M. (2001). *Mol. Microbiol.* **39**, 199–210.
Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. (1988). *Nature (London)*, **336**, 254–257.
Checa, S. K. & Viale, A. M. (1997). *Eur. J. Biochem.* **248**, 848–855.
Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). *Bioinformatics*, **20**, 426–427.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2002). *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, USA.
Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.
Ellis, R. J. & van der Vies, S. M. (1988). *Photosynth. Res.* **16**, 101–115.
Emlyn-Jones, D., Woodger, F. J., Price, G. D. & Whitney, S. M. (2006). *Plant Cell Physiol.* **47**, 1630–1640.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gatenby, A. A. & Ellis, R. J. (1990). *Annu. Rev. Cell Biol.* **6**, 125–149.
Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. (1989). *Nature (London)*, **337**, 44–47.
Holm, L. & Sander, C. (1996). *Science*, **273**, 595–603.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Li, L. A. & Tabita, F. R. (1997). *J. Bacteriol.* **179**, 3793–3796.
Li, W., Jaroszewski, L. & Godzik, A. (2001). *Bioinformatics*, **17**, 282–283.
Lundin, V. F., Stirling, P. C., Gomez-Reino, J., Mwenifumbo, J. C., Obst, J. M., Valpuesta, J. M. & Leroux, M. R. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 4367–4372.
McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). *Bioinformatics*, **16**, 404–405.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Onizuka, T., Endo, S., Akiyama, H., Kanai, S., Hirano, M., Yokota, A., Tanaka, S. & Miyasaka, H. (2004). *Plant Cell Physiol.* **45**, 1390–1395.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pettit, F. K. & Bowie, J. U. (2007). *Ezprot Class Library for Analysis of Protein Sequence and Structure.* http://www.doe-mbi.ucla.edu/~pettit/Ezprot/ezpl_page.htm.
Saschenbrecker, S., Bracher, A., Rao, K. V., Rao, B. V., Hartl, F. U. & Hayer-Hartl, M. (2007). *Cell*, **129**, 1189–1200.
Sheldrick, G. M. & Schneider, T. R. (2001). In *Methods in Macromolecular Crystallography*, edited by D. Turk & L. N. Johnson. Amsterdam: IOS Press.
Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U. & Moarefi, I. (2000). *Cell*, **103**, 621–632.
Stirling, P. C., Bakhoum, S. F., Feigl, A. B. & Leroux, M. R. (2006). *Nature Struct. Mol. Biol.* **13**, 865–870.
Walton, T. A. & Sousa, M. C. (2004). *Mol. Cell*, **15**, 367–374.
Webb, C. T., Gorman, M. A., Lazarou, M., Ryan, M. T. & Gulbis, J. M. (2006). *Mol. Cell*, **21**, 123–133.