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

Rational 'correction' of the amino-acid sequence of RbcX protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus* dramatically improves crystallization

Miroslaw Tarnawski,[a] Szymon Krzywda,[b] Andrzej Szczepaniak[a]* and Mariusz Jaskolski[b,c]*

[a]Department of Biophysics, Faculty of Biotechnology, University of Wroclaw, Przybyszewskiego 63/77, 51-148 Wroclaw, Poland, [b]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland, and [c]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 60-704 Poznan, Poland

Correspondence e-mail: andrzej.szczepaniak@ibmb.uni.wroc.pl, mariuszj@amu.edu.pl

Received 4 July 2008
Accepted 19 August 2008

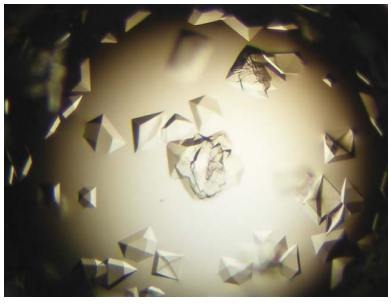
RbcX is a dimeric protein found in cyanobacteria that assists in the assembly of the oligomeric RuBisCO complex. RbcX from the thermophile *Thermosynechococcus elongatus* (TeRbcX) contains an unusual Cys103 residue in its sequence and when expressed recombinantly the protein aggregates and cannot be crystallized. Site-directed mutagenesis of Cys103 to either Arg or Ala produced non-aggregating proteins that could be readily crystallized in several crystal forms. Synchrotron-radiation X-ray diffraction data were collected to 1.96 Å resolution and formed the basis of crystal structure analysis of TeRbcX.

1. Introduction

RuBisCO is the enzyme responsible for CO_2 fixation during photosynthesis. It catalyzes the addition of CO_2 to ribulose-1,5-bisphosphate, a process that under some conditions is the slowest, rate-limiting step of photosynthesis (Cleland *et al.*, 1998). The importance of RuBisCO is difficult to overestimate since this reaction provides the only quantitatively significant link between the pool of inorganic carbon dioxide in the atmosphere and organic matter in the biosphere (Ellis, 1979; Witmarsh & Govindjee, 1999).

In terrestial plants, green algae and cyanobacteria, RuBisCO exists as a holoenzyme composed of eight large subunits (RbcL) and eight small subunits (RbcS). The assembly process requires not only chaperone factors that mediate polypeptide chain folding, but also specific assembly chaperones (Tabita, 1999; Andersson & Taylor, 2003).

The RuBisCO operon of some cyanobacteria contains three co-transcribed genes: *rbc*L, *rbc*X and *rbc*S (Larimer & Soper, 1993). The 15 kDa RbcX protein has no sequence homologues that might indicate its function. It has been suggested that the position of *rbc*X is not random and that the RbcX protein could be an assembly chaperone for RuBisCO. It has been reported that the *rbc*X gene product increases the formation of active RuBisCO when it is co-expressed with the *rbc*L and *rbc*S genes in *Escherichia coli* (Li & Tabita, 1997). Studies with the mesophile *Synechococcus* sp. PCC7002 showed that the introduction of a translational frame shift into the *rbc*X gene resulted in a significant decrease in the production of RuBisCO both in *Synechococcus* and in *E. coli* (Onizuka *et al.*, 2004). The authors also demonstrated that complete loss of the *rbc*X gene hampers the viability of the cyanobacterial cells. However, in *Synechococcus* sp. PCC7942 the *rbc*X gene is located about 100 kbp away from the *rbc*LS operon. A fully segregated *Synechococcus* sp. PCC7942 *rbc*X deletion mutant showed no perturbation in the growth rate or RuBisCO content and activity. Those conflicting observations were additionally complicated when co-expression studies of RuBisCO assembly in *E. coli* showed that RbcX from *Synechococcus* sp. PCC7942 and that from *Synechococcus* sp. PCC7002 are functionally interchangeable and can stimulate the assembly of the PCC7942 and



PCC7002 RuBisCO complexes (Emlyn-Jones *et al.*, 2006). The latest studies of RbcX from *Synechococcus* sp. PCC7002 have confirmed that cyanobacterial RbcX is a RuBisCO assembly chaperone, promoting the formation of the $RbcL_8$ core complex. The crystal structures of RbcX from *Synechococcus* sp. PCC7002 and *Anabaena* sp. CA have revealed that the RbcX protein forms a noncovalent intertwined homodimer possessing two cooperating RbcL-binding regions. A central cleft binds the exposed C-terminal end of the RbcL subunit, enabling RbcX to mediate $RbcL_8$ assembly. Owing to the dynamic nature of these interactions, RbcX is readily displaced from the $RbcL_8$ complex by RbcS, resulting in the production of the active complex (Saschenbrecker *et al.*, 2007). Very recently, the crystal structure of RbcX from another mesophilic cyanobacterium, *Synechocystis* sp. PCC6803, has been reported (Tanaka *et al.*, 2007).

The main objective of the present project is to elucidate the structural determinants of the thermostability of TeRbcX, the RbcX protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. Amino-acid sequence analyses of RbcX proteins (Fig. 1) reveal moderate to high sequence conservation but do not provide any clue to the increased thermal stability of TeRbcX, although they do highlight the presence in the TeRbcX sequence of an unusual single Cys residue at position 103 that is found only in two other thermophilic RbcX proteins, those from *Phormidium laminosum* and *T. vulcanus*, the latter being indistinguishable from TeRbcX. This residue maps onto a surface Arg residue of the known RbcX crystal structures (PDB codes 2peq, 2peo and 2py8) but at a position that precludes its involvement either in creation of the biologically relevant dimer or in interactions with RbcL during $RbcL_8$ assembly. However, recombinant TeRbcX protein expressed in *E. coli* cells could not be crystallized despite numerous trials. The main problems were protein instability and uncontrolled aggregation, which was visible in native PAGE gels (Fig. 2). The aggregation was inhibited by DTT, indicating the involvement of Cys103 in the process. In an attempt to prevent aggregation *via* intermolecular disulfide-bond formation, we mutagenized TeRbcX, creating two variants, C103R and C103A. In both cases, the protein did not show any tendency to aggregate and could be easily crystallized producing several crystal forms. In this paper, we describe the crystallization conditions for the C103R and C103A TeRbcX proteins and demonstrate that the crystals diffract X-rays to at least 1.96 Å resolution.

2. Materials and methods

2.1. Cloning, mutagenesis, expression and purification

The coding sequence of the *rbcX* gene form *T. elongatus* BP-1 (GeneID 1011146) was amplified by polymerase chain reaction (PCR) using plasmid pUC18rbcLXSTh.e containing the entire RuBisCO operon (Gubernator *et al.*, 2008) as template and the following primers: forward 5′-AGACGTCTC**CCATGG**ATGTCA-AGCACA-3′ and reverse 5′-AGACGTCTC**GGATCC**TATTCAGG-GGTTTC-3′, containing *Nco*I and *Bam*HI restriction sites, respectively (bold). The PCR product was digested with *Nco*I and *Bam*HI restrictases and ligated into a pET-16b (Novagen) vector digested with the same enzymes. Clones were screened by restriction digestion and confirmed by sequencing. This construct was used as a template to prepare vectors for the expression of mutant TeRbcX proteins. Two RbcX mutants were prepared by site-directed mutagenesis using the QuikChange (Stratagene) method. In the first mutant Cys103 was changed to Arg (C103R) and in the second to Ala (C103A). Incorporation of the correct mutations and the absence of any other undesired changes was verified by sequencing. The resulting constructs were transformed into *E. coli* BL21 (DE3) cells for overnight cultivation at 310 K in LB medium in the presence of 100 µg ml^{-1} ampicillin. An inoculum at 1:50 dilution was used for the expression culture. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 m*M* when the OD_{600} reached ∼1. 3 h after induction (310 K), the cells were harvested by centrifugation at 5000*g* for 10 min at 277 K and the pellet was resuspended in lysis buffer (20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 1 m*M* EDTA) and frozen at 253 K. After thawing, the cells were disrupted by sonication. The crude lysate was centrifuged at 20 000*g* for 30 min at 277 K to pellet the cell debris. Proteins from the resulting supernatant were precipitated with ammonium sulfate on ice (40% saturation) and centrifuged at 20 000*g* for 15 min at 277 K. The pellet was dissolved in 20 m*M* Tris–HCl pH 8.0 and the clear solution was incubated for 20 min at 353 K and then chilled on ice for another 20 min. Denatured proteins were centrifuged at 20 000*g* for 20 min at 277 K and then at 40 000*g* for 30 min at 277 K. The supernatant, which mainly contained the RbcX protein, was concentrated using Amicon Ultra-15 centrifugal devices (5000 Da molecular-weight cutoff, Millipore) and dialyzed against 20 m*M* Tris–HCl pH 8.0 at 277 K.

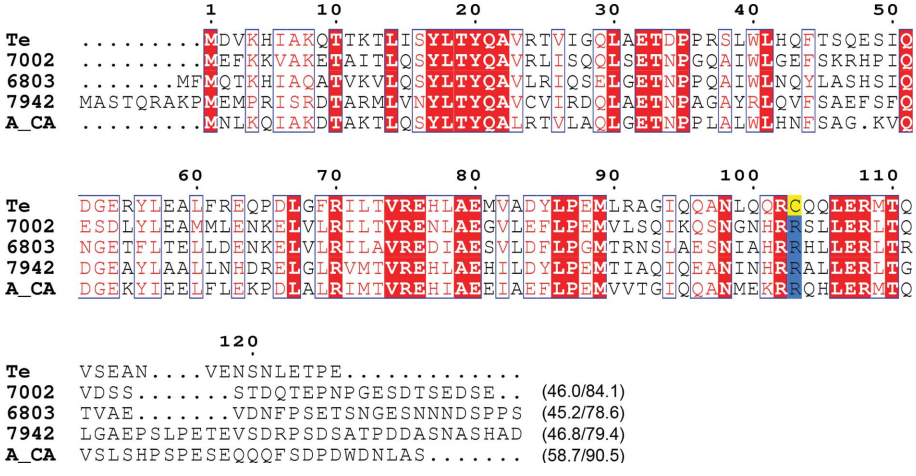


Figure 1
Sequence alignment of RbcX sequences from *T. elongatus* (Te; GenBank accession No. NP_682295), *Synechococcus* sp. PCC7002 (7002; BAA03077), *Synechocystis* sp. PCC6803 (6803; BAA10191), *Synechococcus* sp. PCC7942 (7942; ABB57565) and *Anabaena* sp. CA (A_CA, AAA63603). The sequences were aligned using *ClustalX* (Thompson *et al.*, 1997) and visualized with *ESPript*. Residues with strict identity are shown as white characters on a red background, whereas those with high similarity are shown as red characters and framed. The Cys103 residue of TeRbcX and the corresponding conserved arginine residue are highlighted with yellow and blue backgrounds, respectively. Sequence identity/similarity (%) with respect to *T. elongatus* RbcX is shown for all the remaining sequences.

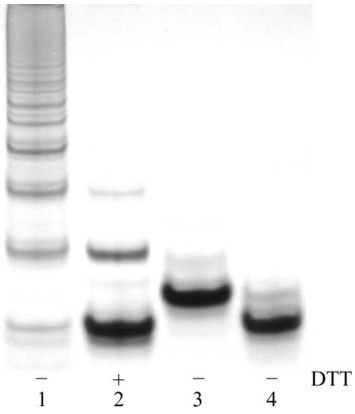


Figure 2
Native PAGE analysis of purified TeRbcX. Lane 1, wild-type TeRbcX purified in the absence of DTT; lane 2, wild-type TeRbcX purified in the presence of DTT; lane 3, C103R mutant protein purified in the absence of DTT; lane 4, C103A mutant protein purified in the absence of DTT.

Further purification steps were performed using an Äkta Purifier FPLC system (GE Healthcare). The protein solution was loaded onto a Q-Sepharose HP anion-exchange column (HiTrap 5 ml, GE Healthcare) equilibrated with 20 m*M* Tris–HCl pH 8.0. After washing off the unbound protein fraction, TeRbcX was eluted using a linear gradient of 0–0.5 *M* NaCl in 20 m*M* Tris–HCl pH 8.0. The protein was finally purified by size-exclusion chromatography with a Superdex 75 prep-grade column (HiLoad 16/60, GE Healthcare) in 20 m*M* Tris–HCl pH 8.0 and 200 m*M* NaCl.

Gel filtration of wild-type TeRbcX revealed several oligomeric forms. Calibration of the column was carried out with four proteins of known molecular weight, but the elution profile did not allow reliable estimation of the molecular weight of the oligomers. The elution profile improved when denaturing and reducing conditions were used, resulting in a single peak corresponding to the monomeric protein. When only a denaturing agent was included (6 *M* urea), two peaks were present, corresponding to TeRbcX monomers and S—S covalent dimers. Use of a reducing agent alone gave no visible improvement. Both TeRbcX mutant proteins eluted from a gel-filtration column as dimers under native conditions.

The fractions containing TeRbcX were pooled, concentrated and dialyzed against 10 m*M* Tris–HCl pH 7.5. The final yield of the recombinant TeRbcX mutant proteins after purification was ∼15 and ∼45 mg per litre of liquid culture for C103A and C103R, respectively. Protein purity and homogeneity were examined by SDS–PAGE (Schägger & von Jagow, 1987) and native PAGE (Davis, 1964). Protein concentration was determined by the BCA method (Smith *et al.*, 1985) using bovine serum albumin as a standard.

2.2. Crystallization

Screening for crystallization conditions was performed in 24-well VDX (Hampton Research) plates using the hanging-drop vapour-diffusion method against 1 ml reservoir solution at 292 K. The Structure Screen 1 and Structure Screen 2 (Molecular Dimensions Ltd) sparse-matrix kits were used to find initial crystallization conditions. The concentration of both of the mutant proteins used for screening was 25 mg ml^{-1} and the crystallization drops were formed by mixing equal volumes (1 µl) of the precipitant and protein solutions. Initially, crystals were obtained in Structure Screen 2 conditions No. 15 (0.1 *M* HEPES pH 7.5, 2.0 *M* ammonium formate) and 37 (0.01 *M* $CoCl_2$, 0.1 *M* acetate buffer pH 4.6, 1.0 *M* 1,6-hexanediol). Fine-tuning of the crystallization conditions was performed by changing the concentration of the precipitants and the buffer pH. Finally, diffraction-quality crystals of the C103A mutant were obtained using 0.1 *M* sodium acetate pH 4.2 and 0.8 *M* 1,6-hexanediol (form A-1) or using 0.1 *M* HEPES pH 6.5 and 2.0 *M* ammonium formate (form A-2). Crystals of the C103R mutant were obtained using a reservoir solution containing 0.1 *M* sodium acetate pH 4.4 and 0.8 *M* 1,6-hexanediol (form R-1) or using 0.1 *M* HEPES pH 7.0 and 1.5 *M* ammonium formate (form R-2) (Fig. 3).

2.3. Data collection and processing

Crystals were harvested in cryoloops and before flash-vitrification in a stream of nitrogen gas were briefly soaked in a cryoprotectant solution consisting of the reservoir solution with 1,6-hexanediol

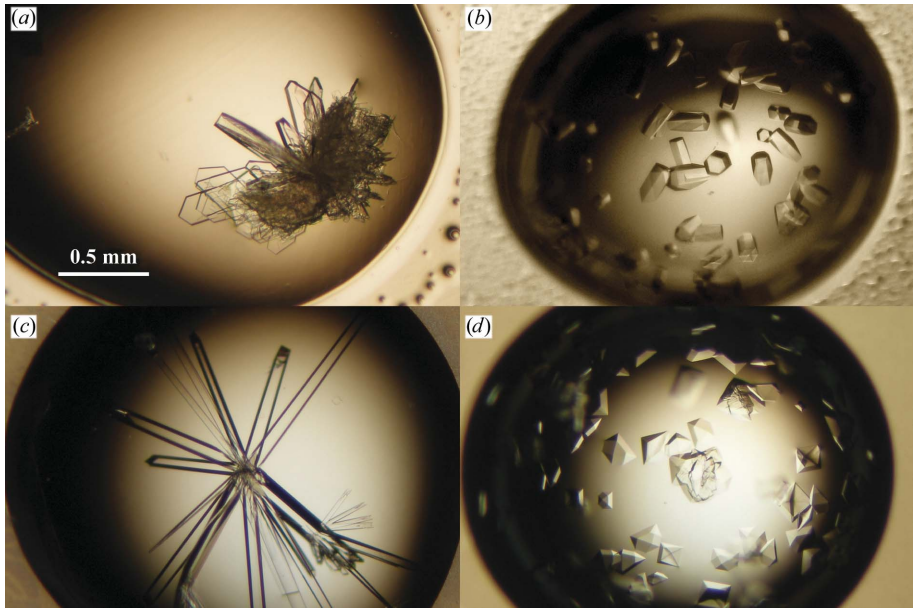


Figure 3
Crystals of TeRbcX: (*a*) form A-1, (*b*) form A-2, (*c*) form R-1, (*d*) form R-2.

Table 1
Summary of crystal data and data-collection statistics.

Values in parentheses are for the highest resolution shell.

Data set	A-1	A-2	R-1	R-2
Crystal parameters				
Dimensions (mm)	0.4 × 0.2 × 0.05	0.3 × 0.1 × 0.1	0.7 × 0.07 × 0.07	0.2 × 0.3 × 0.1
Space group	$P2_12_12_1$	$P321$	$P2_12_12_1$	$P4_12_12$ or $P4_32_12$
Unit-cell parameters (Å)				
a	45.96	81.60	75.58	144.81
b	67.86	81.60	109.27	144.81
c	94.81	176.85	123.84	86.88
Estimated molecules per ASU	2	6	8	6
Estimated V_M (Å^3 Da^{-1})	2.5	2.0	2.2	2.6
Estimated solvent content (%)	51.6	36.8	43.7	52.6
Data collection				
Radiation source	BESSY BL 14.2	MAX-lab J 911-2	BESSY BL 14.2	BESSY BL 14.2
Wavelength (Å)	0.9184	1.0379	0.9184	0.9184
Temperature (K)	100	100	100	100
Detector	MAR 165 CCD	MAR 165 CCD	MAR 165 CCD	MAR 165 CCD
Crystal-to-detector distance (mm)	150	200/140	210	275
Oscillation angle (°)	1.0	0.75/0.5	1.0	1.0
No. of frames	180	100/165	180	160
Resolution range (Å)	50–1.96	50–2.05	50–2.64	50–3.75
No. of observed reflections	140860	274484	195075	103593
No. of unique reflections	21599	41667	28757	9090
Redundancy	6.5 (4.9)	6.6 (2.2)	6.8 (4.2)	11.4 (8.2)
Completeness (%)	98.1 (89.3)	95.3 (62.6)	93.0 (61.5)	90.0 (65.7)
$R_{\rm merge}$† (%)	6.5 (46.9)	5.6 (39.5)	11.8 (51.2)	10.8 (61.3)
$\langle I/\sigma(I)\rangle$	25.9 (3.2)	31.3 (2.2)	14.5 (2.0)	19.7 (2.3)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement of the intensity of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl.

added to a final concentration of 1.8 *M* for forms A-1 and R-1, the reservoir solution supplemented with 20% glycerol for form A-2 and the reservoir solution supplemented with 15% glycerol for form R-2. X-ray diffraction data were collected using synchrotron radiation. The data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The final statistics characterizing the four data sets are summarized in Table 1.

3. Results and discussion

The RbcX protein from the thermophilic organism *T. elongatus* BP-1 was successfully overexpressed in *E. coli* and purified. It is of note that after only two purification steps, *i.e.* ammonium sulfate precipitation and heat treatment, the purity of the protein preparation reached about 90% as judged by SDS–PAGE electrophoresis. After two additional chromatographic steps high-purity recombinant TeRbcX protein was obtained, which appeared as a single band with a molecular weight of about 14.5 kDa in SDS–PAGE under reducing conditions. However, purification of wild-type TeRbcX in the absence of a reducing agent resulted in protein aggregation, which could be detected by native PAGE (Fig. 2). Since the addition of DTT to all buffers strongly (although not completely) inhibited the aggregation process, as confirmed by native PAGE (Fig. 2), it could be deduced that the aggregation was related to cysteine oxidation. This argument seems logical since the amino-acid sequence of TeRbcX contains a Cys residue in the position of the highly conserved Arg103. Nevertheless, the recombinant wild-type TeRbcX protein could never be crystallized, even when reducing agents were present in the crystallization buffers.

Therefore, we attempted another approach, aimed at mutating the Cys103 residue to either arginine or alanine. Both mutants could be purified using the same protocol as used for the wild-type protein. However, no aggregation was observed in native PAGE tests and both mutant proteins could readily be crystallized in more than one crystal form. The crystals are of very good quality, allowing the collection of X-ray diffraction data to at least 1.96 Å resolution (Fig. 4).

The presence of a single Cys residue in a protein sequence can only lead to S—S-linked dimers. The fact that higher-order aggregation is observed with the native TeRbcX sequence therefore indicates that the aggregation process involves not only intermolecular S—S oxidation but also another mechanism. This conclusion is in agreement with the dimeric nature of the C103A/R TeRbcX mutants revealed by the gel-filtration results and with the oligomeric state of

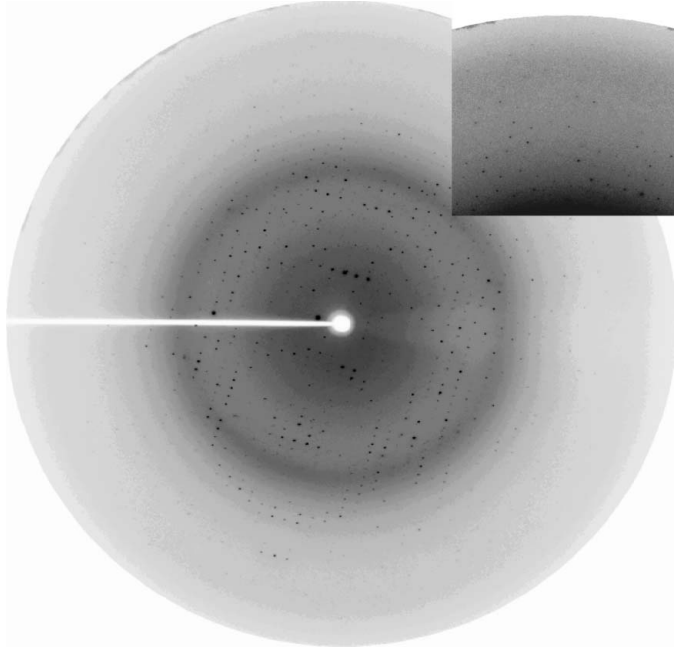


Figure 4
An X-ray diffraction pattern recorded from a form A-1 crystal of TeRbcX (oscillation range 1°). The edge of the detector (inset) corresponds to a resolution of 1.96 Å.

other RbcX proteins. Thus, the higher-order aggregates of TeRbcX can be considered as noncovalent dimers linked together by S—S covalent bonds.

This work was supported in part by a grant from the University of Wroclaw to AS.

References

Andersson, I. & Taylor, T. C. (2003). *Arch. Biochem. Biophys.* **414**, 130–140.
Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C. & Lorimer, G. H. (1998). *Chem. Rev.* **98**, 549–562.
Davis, B. J. (1964). *Ann. N. Y. Acad. Sci.* **121**, 404–427.
Ellis, R. J. (1979). *Trends Biochem. Sci.* **4**, 241–244.
Emlyn-Jones, D., Woodger, F. J., Price, G. D. & Whitney, S. M. (2006). *Plant Cell Physiol.* **47**, 1630–1640.
Gubernator, B., Bartoszewski, R., Kroliczewski, J., Wildner, G. & Szczepaniak, A. (2008). *Photosynth. Res.* **95**, 101–109.
Larimer, F. W. & Soper, T. S. (1993). *Gene*, **126**, 85–92.
Li, L.-A. & Tabita, F. R. (1997). *J. Bacteriol.* **179**, 3793–3796.
Onizuka, T., Endo, S., Akiyama, H., Kanai, S., Hirano, M., Yokoya, A., Tanaka, S. & Miyasaka, H. (2004). *Plant Cell Physiol.* **45**, 1390–1395.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Saschenbrecker, S., Bracher, A., Rao, K. V., Rao, B. V., Hartl, F. U. & Hayer-Hartl, M. (2007). *Cell*, **129**, 1189–1200.
Schägger, H. & von Jagow, G. (1987). *Anal. Biochem.* **166**, 368–379.
Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. (1985). *Anal. Biochem.* **150**, 76–85.
Tabita, F. R. (1999). *Photosynth. Res.* **60**, 1–28.
Tanaka, S., Sawaya, M. R., Kerfeld, C. A. & Yeates, T. O. (2007). *Acta Cryst.* D**63**, 1109–1112.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). *Nucleic Acids Res.* **25**, 4876–4882.
Whitmarsh, J. & Govindjee (1999). *Concepts in Photobiology: Photosynthesis and Photomorphogenesis*, edited by G. S. Singhal, G. Renger, S. K. Sopory, K.-D. Irrgang & Govindjee, pp. 11–51. Dordrecht: Narosa Publishers/ Kluwer Academic Publishers.