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

Overexpression, purification, characterization and preliminary crystallographic study of phosphoglycolate phosphatase from *Shigella flexneri* 2a strain 301

Heli Liu, Huina Zhou, Deyu Zhu and Ruchang Bi*

Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, People's Republic of China

Correspondence e-mail: rcbi@ibp.ac.cn

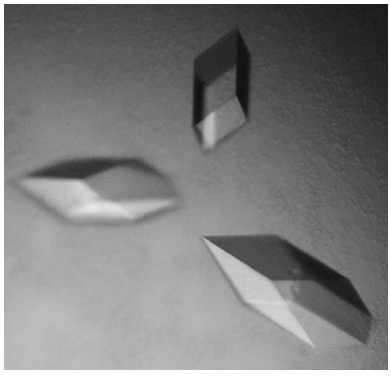
Received 14 September 2008
Accepted 21 November 2008

Phosphoglycolate phosphatase has a salvage function in the metabolism of the 2-phosphoglycolate formed during bacterial DNA repair. In order to better understand its dimerization behaviour, the influence of metal ions on its activity and its catalytic mechanism at the molecular level, recombinant phosphoglycolate phosphatase from *Shigella flexneri* was overexpressed, purified, characterized and crystallized by the hanging-drop vapour-diffusion method at 291 K using polyethylene glycol 3500 as a precipitant and zinc acetate as an additive. The crystals belonged to space group *R*3, with unit-cell parameters a = 88.1, b = 88.1, c = 259.2 Å, corresponding to the presence of two molecules in the asymmetric unit. SeMet-labelled protein was also prepared and crystallized for use in phase determination. Initial structure determination using the multiwavelength anomalous dispersion (MAD) method clearly revealed that SfPGPase bears an α-helical cap domain that differs from that of a previously reported orthologue.

1. Introduction

Phosphoglycolate phosphatase (PGPase; EC 3.1.3.18) is an enzyme that catalyzes the hydrolysis of phosphoglycolate to glycolate and phosphate. In the Gram-negative bacterium *Escherichia coli*, the gene encoding this enzyme was found in the same operon as the *dam* gene that encodes DNA adenine methyltransferase (Lyngstadaas *et al.*, 1995). The DNA breaks produced by ionizing radiation create terminal 3′-phosphate or 3′-phosphoglycolate overhangs that are inhibitors of the DNA-repair process and must be removed for DNA polymerase to function (Inamdar *et al.*, 2002; Kim *et al.*, 2004). PGPase plays a housekeeping role in the dissimilation of the intracellular 2-phosphoglycolate formed in the DNA repair of 3′-phosphoglycolate ends (Pellicer *et al.*, 2003). PGPase also participates in CO_2 assimilation in chemoautotrophic organisms (Shively *et al.*, 1998). In mammalian red blood cells, PGPase modulates the affinity of haemoglobin for oxygen by modification of the biophosphoglycerate shunt (Rose *et al.*, 1986).

PGPases from a wide variety of species have been extensively studied (Christeller & Tolbert, 1978; Husic & Tolbert, 1984; Hardy & Baldy, 1986; Zecher & Wolf, 1980; Mamedov *et al.*, 2001; Pellicer *et al.*, 2003; Kim *et al.*, 2004). The PGPases from human erythrocytes (Zecher & Wolf, 1980), *Chlamydomonas reinhardtii* (Mamedov *et al.*, 2001) and *Thermoplasma acidophilum* (Kim *et al.*, 2004) are dimeric molecules, while the PGPase from *E. coli* is a monomer (Pellicer *et al.*, 2003). The dimerization state of PGPase seems to be species-dependent. The activity of PGPases seems to be dependent on or activated by divalent cations such as Mg^{2+}, Co^{2+}, Mn^{2+} and Zn^{2+} (Husic & Tolbert, 1984; Zecher & Wolf, 1980; Kim *et al.*, 2004). However, another alkaline earth metal ion, Ca^{2+}, has been reported to strongly inhibit the activity of the PGPases from *C. reinhardtii* (Mamedov *et al.*, 2001) and *T. acidophilum* (Kim *et al.*, 2004). The crystal structure of PGPase from *T. acidophilum* (TaPGPase) revealed that this enzyme is composed of a large Rossmann hydrolase-



fold domain, as shared by the members of the haloacid dehalogenase (HAD) superfamily, and a small mixed α/β-cap domain, which further classifies TaPGPase into the type II HAD subfamily (Kim *et al.*, 2004; Lu *et al.*, 2005; Selengut, 2001).

In order to better understand the dimerization behaviour of PGPase, the influence of metal ions on its activity and its catalytic mechanism at the molecular level, we selected the phosphoglycolate phosphatase from *Shigella flexneri* 2a strain 301 (SfPGPase) as a research model. SfPGPase is composed of 252 amino acids, including five methionine residues, and has 14.5% sequence identity to TaPGPase. Here, we report the cloning, expression, purification, characterization and preliminary crystallographic study of SfPGPase.

2. Materials and methods

2.1. Cloning

The gene *gph* encoding SfPGPase was amplified by PCR from the genomic DNA of *S. flexneri* 2a strain 301. *Nde*I and *Xho*I restriction sites (bold) were introduced into the oligonucleotide primers for upstream and downstream sequences: 5′-GAGCGC**CATATG**AATA-AGTTTGAAGATATTCGCGG-3′ and 5′-CACCGA**CTCGAG**GTC-ATTTTTCGATTCCTGAT-3′, respectively. The PCR fragment was digested with *Nde*I and *Xho*I and subcloned into the pET-22b(+) vector (Novagen) for expression of protein with a 6×His tag at its C-terminus. The constructs were confirmed by DNA sequencing.

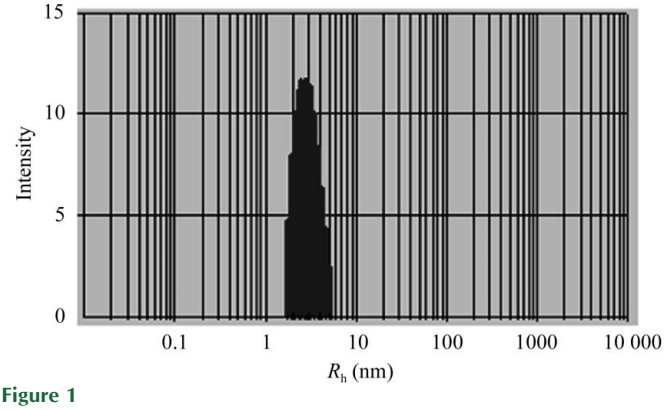


Figure 1
DLS regularization histogram analysis of SfPGPase. The vertical coordinate axis corresponds to the relative dynamic light-scattering intensity and the horizontal axis corresponds to the hydrodynamic radius (on a logarithmic scale).

2.2. Expression and purification

SfPGPase was overexpressed in *E. coli* strain BL21 (DE3) grown at 310 K in 1 l LB medium containing 50 µg ml^{-1} ampicillin. When the OD_{600} of the culture reached 0.6–0.8, SfPGPase production was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.8 m*M*. After a further 3 h of incubation, the cells were harvested and suspended in lysis buffer [25 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 10 m*M* $MgCl_2$, 10%(*v*/*v*) glycerol and 3 m*M* mercaptoethanol] supplemented with a suitable amount of phenylmethylsulfonyl fluoride (PMSF). The cells were then lysed by sonication on ice and centrifuged at 12 000*g* and 277 K for 30 min to remove cell debris. The resulting supernatant was applied onto a nickel-chelating Sepharose column (Amersham Pharmacia Biotech) previously equilibrated with lysis buffer. The column was washed with lysis buffer containing 20, 40 and 200 m*M* imidazole. The eluate obtained using 200 m*M* imidazole buffer was preserved and exchanged to low-salt non-imidazole buffer [25 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl, 10%(*v*/*v*) glycerol, 10 m*M* $MgCl_2$, 6 m*M* mercaptoethanol]. Further purification was performed on a HiLoad gel-filtration column (16/60 Superdex 75 prep-grade, Amersham Biosciences). The eluted fractions (OD_{280} > 0.26) were pooled and concentrated for characterization and crystallization.

Selenomethionine (SeMet) labelled SfPGPase was expressed in *E. coli* strain BL21 (DE3) using a metabolic inhibition method. At an OD_{600} of 0.3, the cells were harvested and resuspended in 1 l minimal medium supplemented with 3% glucose, 1% YNB culture, 60 mg ml^{-1} SeMet, 100 mg ml^{-1} lysine, 100 mg ml^{-1} threonine, 100 mg ml^{-1} phenylalanine, 50 mg ml^{-1} leucine, 50 mg ml^{-1} isoleucine and 50 mg ml^{-1} valine and then cultivated to an OD_{600} of 0.6–0.8. At this point, 0.5 m*M* IPTG was added for protein induction, which was followed by overnight incubation at 293 K. The SeMet-labelled protein was purified and concentrated using the same protocol as used for the native protein.

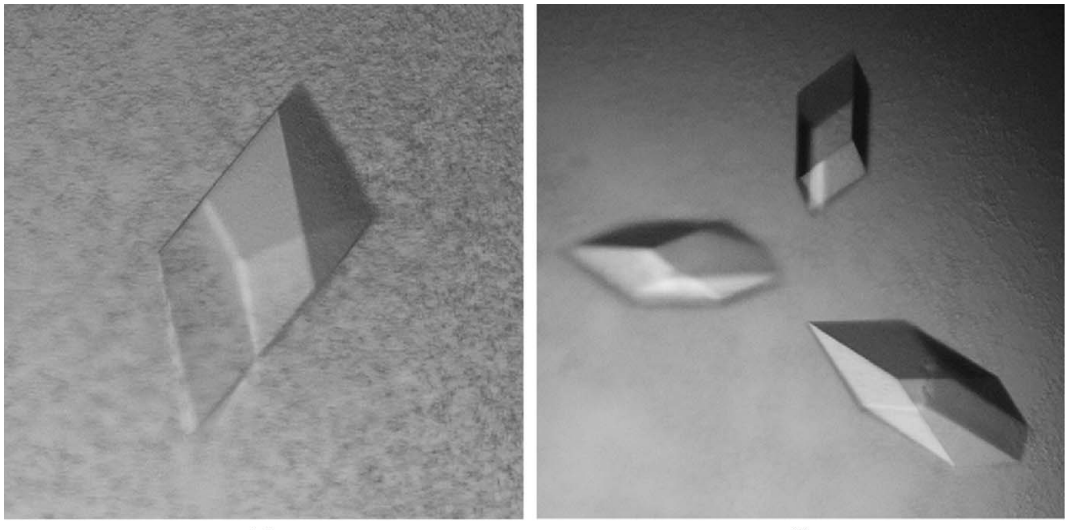


Figure 2
Typical crystals of (*a*) native and (*b*) SeMet-labelled SfPGPase.

2.3. Dynamic light scattering

All dynamic light-scattering (DLS) measurements were performed using a DynaPro DLS instrument (Protein Solutions). Photons were counted and the time-dependence of the light-intensity fluctuations was analyzed by autocorrelation. The native SfPGPase sample was diluted to a concentration of about 1 mg ml^{-1} and centrifuged for 10 min at 12 000*g* prior to scattering data collection. Experiments were conducted at 298 K and at least ten measurements were taken for each experiment. Regularization histogram analyses of samples were carried out using the *DYNAMICS* v.5.25.44 software (Protein Solutions Inc.).

2.4. Crystallization

The native protein sample was concentrated to 8 mg ml^{-1} for initial crystallization screening with 290 conditions from commercial sparse-matrix screening kits (Crystal Screen, Crystal Screen 2 and Index Screen; Hampton Research, USA). Crystallization was performed by the hanging-drop vapour-diffusion method. Each droplet, composed of 1 µl protein solution and an equal volume of the crystallization solution, was equilibrated against 450 µl reservoir solution within a sealed well. 3 d later, microcrystals were observed in most zinc-containing conditions from Index Screen. Condition No. 93 [0.05 *M* zinc acetate dehydrate, 20%(*w*/*v*) polyethylene glycol 3350] was selected and optimized. SeMet-labelled SfPGPase crystals were grown using the same optimized condition as used for the native protein.

2.5. Diffraction data collection and processing

For the native protein, diffraction data were collected using a Rigaku FR-E SuperBright X-ray generator with a wavelength of 1.5418 Å and an R-AXIS IV^{++} imaging-plate area detector. A crystal without any cryoprotectant was directly mounted on the goniometer and rapidly frozen in a nitrogen-gas stream at 100 K produced by a cryocooling device from Oxford Cryosystems. The crystal-to-detector distance was set to 200 mm. The oscillation angle per frame was 1° and the exposure time was 5 min. Three-wavelength MAD (multi-wavelength anomalous dispersion) data were collected from SeMet-labelled protein crystals on beamline BL-6A of Photon Factory in Japan. Peak, inflection and remote data sets were collected sequentially from one crystal. The exposure time for each image was 1 s and the oscillation range was 1°, with a crystal-to-detector distance of 200 mm. All data sets were indexed, integrated and scaled with *HKL*-2000 (Otwinowski & Minor, 1997) and the statistics are shown in Table 1.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution bin.

		SeMet derivative		
	Native	Peak	Inflection	Remote
Space group	*R*3	*R*32		
Unit-cell parameters (Å)	a = 88.1, b = 88.1, c = 259.2	a = 89.0, b = 89.0, c = 260.2		
Wavelength (Å)	1.5418	0.9786	0.9796	0.9600
Resolution range (Å)	20.00–2.80	50.00–3.30	50.00–3.30	50.00–3.30
No. of observed reflections	40023	123774	96335	96414
No. of unique reflections	15553	6147	6140	6144
$\langle I/\sigma(I)\rangle$	17.1 (4.0)	29.6 (7.7)	24.7 (6.4)	23.5 (6.1)
Completeness (%)	89.0 (91.0)	97.9 (98.0)	97.0 (100.0)	87.8 (100.0)
$R_{\rm merge}$† (%)	5.8 (30.0)	13.4 (50.2)	13.5 (53.4)	14.2 (56.5)

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

3. Results and discussion

SfPGPase was expressed as a soluble protein with a high yield of about 16 mg protein per litre of LB culture. After purification by Ni^{2+}–NTA affinity chromatography and gel-filtration chromatography, SfPGPase showed high purity as confirmed by SDS–PAGE (data not shown). In solution, SfPGPase exhibited good homogeneity

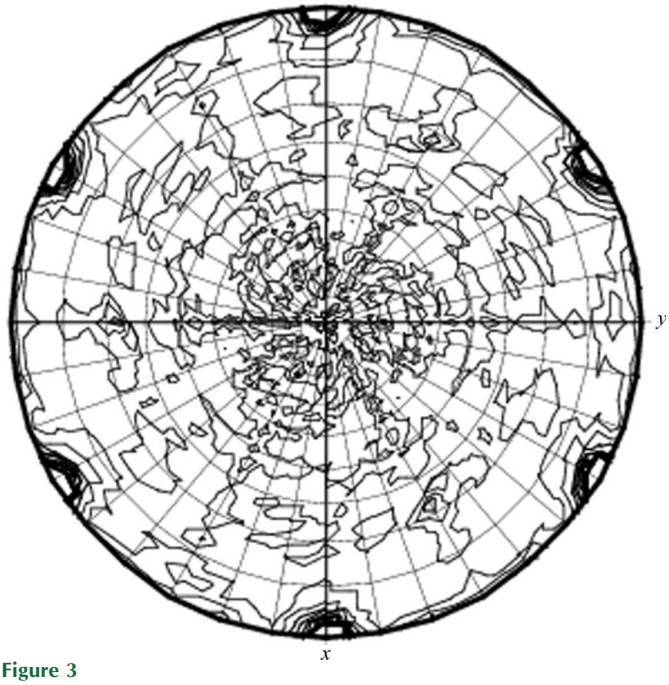


Figure 3
Self-rotation function calculated at the κ = 180° section. Six peaks are observed in directions perpendicular to the *c* axis.

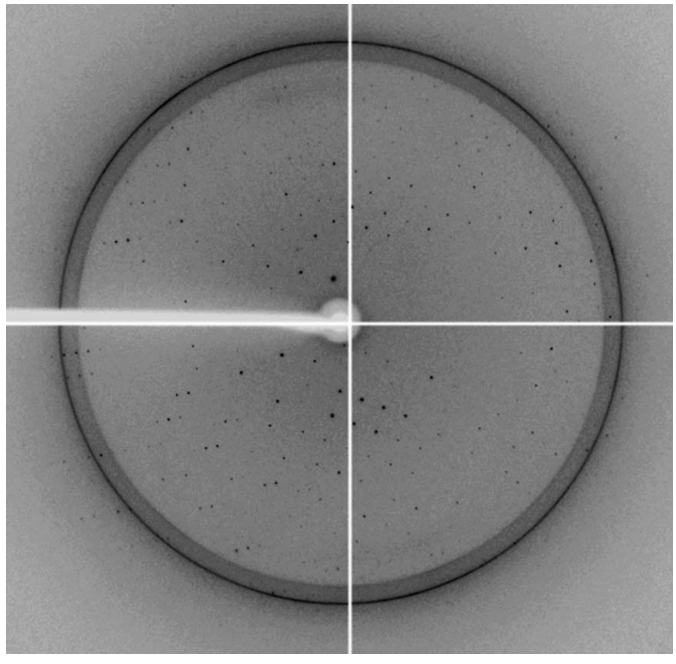


Figure 4
Snapshot of the diffraction pattern of a SeMet-labelled SfPGPase crystal.

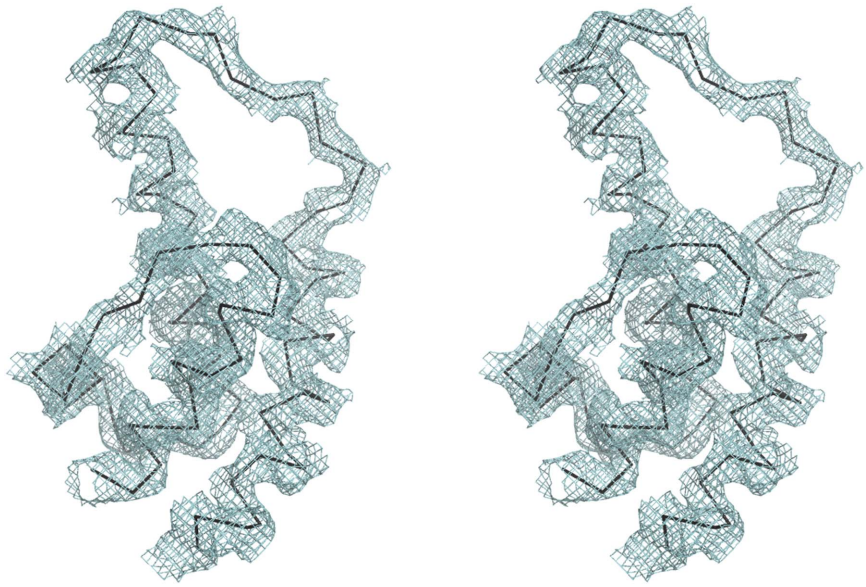


Figure 5
Stereoview of the $2F_o - F_c$ electron-density map at 2.80 Å resolution. The C^α polypeptide trace (black) is superimposed onto the electron density (cyan) contoured at 1.0σ and cut off at a distance of 2.5 Å from the backbone atoms. This figure was produced using *PyMOL* (DeLano, 2002).

and monodispersity as evidenced by a single peak with a hydrodynamic radius (R_h) of 3.02 nm in the regularization histogram (Fig. 1). Correspondingly, SfPGPase has a measured molecular weight of 27.0 kDa, approximating the value of 28.1 kDa calculated for the sequence of SfPGPase plus the C-terminal 6×His tag. This result suggested that SfPGPase mainly exists as a monomer in solution, which is consistent with the monomeric state of PGPase from *E. coli* (Pellicer *et al.*, 2003).

The crystallization of SfPGPase seemed to require zinc ion. During initial screening, microcrystals were observed after 3–5 d in condition No. 93 of Index Screen as well as other conditions containing zinc acetate. The preference for zinc ion during crystallization suggested that SfPGPase might contain zinc-binding site(s) and that zinc coordination promotes the enzyme to a stable conformation and facilitates crystallization. After a large number of optimization trials, large crystals suitable for X-ray diffraction data collection were obtained from a droplet containing 5 mg ml^{-1} protein solution, 6 m*M* zinc acetate, 10%(*w*/*v*) polyethylene glycol 3500, 0.05 *M* HEPES pH 7.5 and 0.01 *M* ammonium sulfate. The droplet was equilibrated against 500 µl reservoir solution containing 25%(*w*/*v*) polyethylene glycol 3500, 0.1 *M* HEPES pH 7.5 and 0.02 *M* ammonium sulfate. The native SfPGPase crystals (Fig. 2*a*) used for data collection typically had dimensions of 0.20 × 0.20 × 0.15 mm and diffracted to 2.8 Å resolution. The data-processing statistics are shown in Table 1. The self-rotation function calculated with *MOLREP* (Vagin & Teplyakov, 1997) showed six non-origin peaks with $\kappa = 180^\circ$, indicating the presence of noncrystallographic twofold axes (Fig. 3). This suggested that there are two molecules in the asymmetric unit, with a reasonable Matthews coefficient (V_M; Matthews, 1968) of 3.4 Å^3 Da^{-1} and a relatively high solvent content of 63.9%.

Because of the low sequence identity and potential structural discrepancy between TaPGPase and SfPGPase, the native SfPGPase crystal could not be phased by the molecular-replacement method using the released structure of TaPGPase (PDB code 1l6r; Kim *et al.*, 2004) as a search template. Therefore, SeMet SfPGPase was prepared using the same procedure as used for the native protein, resulting in smaller crystals with dimensions of 0.15 × 0.10 × 0.06 mm (Fig. 2*b*). The SeMet protein crystals diffracted to 2.9 Å resolution (Fig. 4) and belonged to space group *R*32 with one protein molecule per asymmetric unit, corresponding to a Matthews coefficient of 3.4 Å^3 Da^{-1} and a solvent content of 63.7%. The high solvent content may account for the relatively weak diffraction of both the native and the SeMet-derivative crystals.

Using the three-wavelength MAD data sets, four Se atoms per asymmetric unit were found using the software *SOLVE* (Terwilliger & Berendzen, 1999) for initial phase calculation, leading to an overall FOM (figure of merit) of 0.52 for reflections in the 20.0–3.3 Å resolution range. Phase improvement and automatic model building was performed using *RESOLVE* (Terwilliger, 2000). Using the explicit positions of four Se atoms as well as iterative manual model building, the small domain of SfPGPase could be unambiguously discerned in the electron-density map calculated with the final improved phases (Fig. 5). The small cap domain of SfPGPase basically comprises four helixes and a long loop and stands in sharp contrast to the cap domain of TaPGPase, which is an open β-sandwich consisting of three antiparallel β-sheets flanked by two helices (Kim *et al.*, 2004).

The nature and location of the cap domain has been used to divide the HAD superfamily into three subfamilies: subfamily I with a small α-helical bundle cap domain, subfamily II with a mixed α/β domain and subfamily III with no cap domain (Lu *et al.*, 2005; Selengut, 2001). Therefore, SfPGPase is a subfamily I HAD enzyme, while TaPGPase is assigned to subfamily II. The fact that SfPGPase and TaPGPase belong to different HAD subfamilies implies that the two orthologues would have different functions of the cap in substrate recognition and in the catalytic process. Further model building and structure analysis of the whole SfPGPase protein are in progress.

We are grateful to Yi Han for his help with in-house data collection from native protein crystals. This project was supported financially by the Innovation Program of the Chinese Academy of Sciences.

References

Allen, K. N. & Dunaway-Mariano, D. (2004). *Trends Biochem. Sci.* **29**, 495–503.

Christeller, J. T. & Tolbert, N. E. (1978). *J. Biol. Chem.* **253**, 1780–1785.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Hardy, P. & Baldy, P. (1986). *Planta*, **168**, 245–252.
Husic, H. D. & Tolbert, N. E. (1984). *Arch. Biochem. Biophys.* **229**, 64–72.
Inamdar, K. V., Pouliot, J. J., Zhou, T., Lees-Miller, S. P., Rasouli-Nia, A. & Povirk, L. F. (2002). *J. Biol. Chem.* **277**, 27162–27168.
Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H., Joachimiak, A., Edwards, A. M. & Christendat, D. (2004). *J. Biol. Chem.* **279**, 517–526.
Lu, Z., Dunaway-Mariano, D. & Allen, K. N. (2005). *Biochemistry*, **44**, 8684–8696.
Lyngstadaas, A., Lobner-Olesen, A. & Boye, E. (1995). *Mol. Gen. Genet.* **247**, 546–554.
Mamedov, T. G., Suzuki, K., Miura, K., Kucho, K. K. & Fukuzawa, H. (2001). *J. Biol. Chem.* **276**, 45573–45579.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pellicer, M. T., Nuñez, M. F., Aguilar, J., Badia, J. & Baldoma, L. (2003). *J. Bacteriol.* **185**, 5815–5821.
Rose, Z. B., Grove, D. S. & Seal, S. N. (1986). *J. Biol. Chem.* **261**, 10996–11002.
Selengut, J. D. (2001). *Biochemistry*, **40**, 12704–12711.
Shively, J. M., van Keulen, G. & Meijer, W. G. (1998). *Annu. Rev. Microbiol.* **52**, 191–230.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Zecher, R. & Wolf, H. U. (1980). *Biochem. J.* **191**, 117–124.