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

Structure of the *Mycobacterium tuberculosis* soluble inorganic pyrophosphatase Rv3628 at pH 7.0

Stefano Benini[a]* and Keith Wilson[b]

[a]Faculty of Science and Technology, Free University of Bolzano, Piazza Università 5, Bolzano 39100, Italy, and [b]York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, England

Correspondence e-mail: stefano.benini@unibz.it

Received 11 April 2011
Accepted 15 June 2011

PDB Reference: Rv3628, 1wcf.

The 1.5 Å resolution crystal structure of the *Mycobacterium tuberculosis* soluble inorganic pyrophosphatase Rv3628 at pH 7.0 is reported. The *M. tuberculosis* and *M. leprae* genomes include genes for the only two family I inorganic pyrophosphatases known to contain two histidines in the active site. The role of these two residues in catalysis is not fully understood. Mutational and functional studies of the *M. tuberculosis* enzyme showed that His21 and His86 are not essential for pyrophosphate hydrolysis, but are responsible for a shift in the optimal pH for the reaction compared with the *Escherichia coli* enzyme. Comparison with the structure previously reported at pH 5.0 provides further insight into the role of the two histidines. Two potassium-binding sites are found as a result of the high potassium concentration in the mother liquor

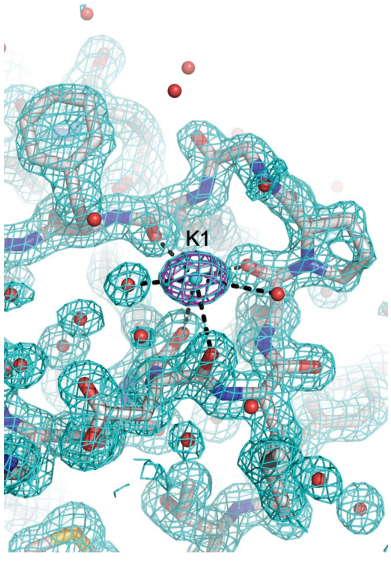
1. Introduction

Inorganic pyrophosphate (PP_i) is produced by all living cells as a byproduct of many biosynthetic reactions (*e.g.* DNA, RNA, polysaccharide and protein biosynthesis; Kornberg, 1962). The hydrolysis of PP_i carried out by soluble pyrophosphatase (PPase) is of key importance in maintaining cell viability, as its accumulation in the cell would inhibit biosynthetic reactions, eventually leading to cell death. PPases have been demonstrated to be essential in the bacterium *Escherichia coli* (Chen *et al.*, 1990) and in the yeast *Saccharomyces cerevisiae* (Lundin *et al.*, 1991).

Soluble PPases are ubiquitous enzymes that have been classified into two families: I and II. Family I PPases are present in eukaryotes, eubacteria and archaebacteria, and share a common core structure. The eukaryotic PPases are homodimers with a long N-terminal extension and a subunit size of 31–34 kDa, while the prokaryotic PPases are homohexamers with a subunit size of 19–22 kDa. Two monomeric forms of eukaryotic family I PPases have been characterized: one in the chloroplasts and one in the mitochondria of the microalga *Chlamydomonas reinhardtii* (Gómez-García *et al.*, 2006). The chloroplastic form is closely related to the eukaryotic PPases, while the mitochondrial form is related to the bacterial soluble PPases (Gómez-García *et al.*, 2006). Eukaryotic and prokaryotic family I PPases have a highly conserved active site with 13 conserved and functionally important residues (Sivula *et al.*, 1999) and share the same reaction mechanism (Heikinheimo *et al.*, 1996; Harutyunyan *et al.*, 1997).

Mg^{2+} ions are required for catalysis in both eukaryotic and prokaryotic family I enzymes. Several crystal structures of family I PPases have been determined to date from the eukaryote *S. cerevisiae* (Heikinheimo *et al.*, 1996), the prokaryote *E. coli* (Oganessyan *et al.*, 1994) and other bacteria (Teplyakov *et al.*, 1994; Leppänen *et al.*, 1999; Liu *et al.*, 2004; Wu *et al.*, 2005; Tammenkoski *et al.*, 2005; Rodina *et al.*, 2008), while several others have been deposited in the PDB as the result of structural genomics initiatives.

Family II PPases were first discovered in *Bacillus subtilis* (Young *et al.*, 1998; Shintani *et al.*, 1998). They are homodimeric, activated by Mn^{2+} ions and only found in prokaryotes. Their active site is characterized by the presence of three histidines, as shown by the crystal structures of the *B. subtilis*, *Streptococcus gordonii* (Ahn *et al.*, 2001) and *S. mutans* enzymes (Merckel *et al.*, 2001).



Mycobacterium tuberculosis PPase (Mt-PPase) is unique amongst those family I PPases for which structures are known in that it features two histidines in the active site (Tammenkoski *et al.*, 2005; Rodina *et al.*, 2008). Mt-PPase is constitutively expressed and is not over-regulated upon macrophage infection or by exposure to environmental stress when grown *in vitro* (Triccas & Gicquel, 2001).

The availability of the *M. tuberculosis* and *M. leprae* genome sequences (Cole *et al.*, 1998; Cole, 1998) allows a comparative genomic approach to identify those genes that are important for the survival of both mycobacteria. The *M. leprae* genome contains only 1604 protein-coding genes, compared with 3959 in *M. tuberculosis* (Eiglmeier *et al.*, 2001). In *M. leprae* gene deletion and decay have drastically reduced the size of the genome, eliminating many important metabolic pathways, including siderophore production and some oxidative and anaerobic enzymes (Eiglmeier *et al.*, 2001). However, the *M. leprae* genome retains the gene *ML0210* coding for PPase, again supporting its essentiality.

The presence of the histidines in the active site of Mt-PPase lowers the pH for optimal activity compared with *E. coli* PPase (Ec-PPase; Tammenkoski *et al.*, 2005). In the previous structure of Mt-PPase determined at pH 5.0 His86 is bound to a sulfate molecule, as is His98 in the crystal structure of the family II enzyme from *S. mutans* (Merckel *et al.*, 2001), suggesting a role in PP_i hydrolysis. In a recent study of the metal cofactor specificity of Mt-PPase, His21 and His86 were proposed to be responsible for enabling the enzyme to hydrolyse ATP and PNP in the presence of either Mg^{2+} or Mn^{2+} (Rodina *et al.*, 2008).

Here, we describe the 1.5 Å resolution structure of Mt-PPase from crystals grown at pH 7.0 (PDB entry 1wcf) and its comparison with the 1.3 Å resolution structure previously reported from crystals grown at pH 5.0 (PDB entry 1sxv; Tammenkoski *et al.*, 2005). The conformational changes observed confirm the involvement of His86 in modulating the catalytic activity as suggested by kinetic studies (Tammenkoski *et al.*, 2005; Rodina *et al.*, 2008). There are two potassium ions bound to two different binding sites in the present structure. One of these corresponds to the Mg^{2+} site in the *E. coli* holoenzyme (Harutyunyan *et al.*, 1997) and the other to the Na^+ site observed in Ec-PPase crystals obtained using NaCl as a precipitant (Samygina *et al.*, 2007).

2. Materials and methods

2.1. Gene cloning, protein overexpression and purification

These were carried out as described previously (Tammenkoski *et al.*, 2005).

2.2. Protein crystallization

Crystallization screening was performed using a Mosquito crystallization robot (TTP LabTech) in 96-well low-profile Greiner plates with sitting drops consisting of 150 nl protein solution mixed with an equal amount of precipitant. The screens used were Index and SaltRX from Hampton Research and CSS1 and CSS2 (Brzozowski & Walton, 2001) buffered at pH 7.0 with 150 m*M* Na HEPES. Hexagonal rod-like crystals grew from 1.8 *M* $NaKHPO_4$ pH 6.9. These conditions were reproduced manually in larger volumes by the hanging-drop method in Linbro plates, equilibrating 1 µl protein and 1 µl precipitant against 1 ml precipitant. Crystals growing from 1.6 *M* $NaKHPO_4$ pH 7.0 reached dimensions of about 0.3 × 0.3 × 1.0 mm.

Table 1
X-ray data-collection statistics.

Values in parentheses are for the high-resolution bin.

X-ray source	ESRF ID14-1	In-house Cu $K\alpha$
Wavelength (Å)	0.93	1.54
Space group	$P6_322$	$P6_322$
Unit-cell parameters		
$a = b$ (Å)	96.90	96.82
c (Å)	103.31	103.32
Resolution range (Å)	84.52–1.54 (1.58–1.54)	20.0–1.80 (1.84–1.80)
No. of images	250/100†	254
Crystal-to-detector distance (mm)	125/266†	140
$\Delta\varphi$ per image (°)	0.4/1.0†	0.5
Raw measurements	537884	366183
Unique reflections	42745	27144
R_{merge}‡	0.085 (0.27)	0.087 (0.31)
Multiplicity	12.6 (8.5)	13.5 (8.9)
Completeness (%)	99.9 (99.8)	99.6 (98.8)
$\langle I/\sigma(I)\rangle$	30.3 (4.3)	31.8 (3.6)
Crystal mosaicity (°)	0.54	0.68
B factor from Wilson plot (Å^2)	24.7	25.7

† High resolution/low resolution. ‡ $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is an individual intensity measurement and $\langle I(hkl)\rangle$ is the average intensity for this reflection.

Table 2
Summary of refinement statistics for 1wcf.

Protein atoms	1317
Solvent atoms	286
Ligands	1 PO_4^{3-}, 2 K^+
Temperature factor for protein atoms (Å^2)	17.68
Temperature factor for solvent (Å^2)	34.67
R.m.s.d. bond lengths (Å)	0.017
R.m.s.d. bond angles (°)	1.63
Ramachandran, most favoured region (%)	96.3
Ramachandran, additional allowed region (%)	3.7
R/R_{free}† (%)	14.1/15.7

† R and $R_{free} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$; for the calculation of R_{free} a subset of reflections (5%) was randomly chosen as a test set.

2.3. Data collection and structure solution

A single crystal was soaked in 4 µl of a cryosolution consisting of 20%(*v*/*v*) glycerol, 1.8 *M* $NaKHPO_4$ pH 7.0 and vitrified in a 120 K nitrogen stream (Oxford Cryosystems Cryostream). Data to a resolution of 1.8 Å were collected in-house on a rotating-anode source with Cu $K\alpha$ radiation using a MAR345 image-plate detector (Table 1). Data were subsequently recorded to a resolution of 1.54 Å at the ESRF (Table 1). Images were integrated with *DENZO* and scaled with *SCALEPACK* (Otwinowski & Minor, 1997). The space group was $P6_322$, with one molecule in the asymmetric unit and a Matthews coefficient of 3.6 Å^3 Da^{-1} with 65.7% solvent content (Table 1). The structure of Mt-PPase at pH 5.0 (PDB entry 1sxv) stripped of water and ligands was used as an initial model. Model inspection and manual building were carried out using *Coot* (Emsley & Cowtan, 2004). Refinement was carried out with *REFMAC* (Murshudov *et al.*, 2010) coupled with *ARP*/*wARP* (Morris *et al.*, 2004) to a final R of 14.1% and R_{free} of 15.7%. The refinement statistics are summarized in Table 2. Model and structure factors have been deposited in the PDB with accession code 1wcf.

3. Results and discussion

The model 1wcf is almost complete, only lacking the C-terminal amino acid His162. The metal cofactors are absent from the active site, as in 1sxv. It has been reported for the related *E. coli* enzyme that incorporation of Mg^{2+} into the active site required cocrystallization with as much as 250 m*M* $MgCl_2$ (Harutyunyan *et al.*, 1997) or 80 m*M*

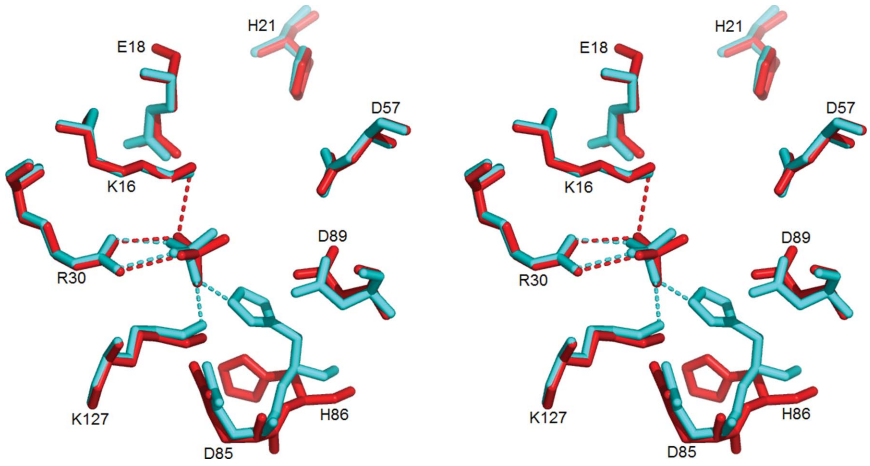


Figure 1
Stereoview of the superposition of the 1sxv (cyan) and 1wcf (red) active sites (orientation and selected residues according to Tammenkoski *et al.*, 2005). This figure was prepared with *PyMOL* (DeLano, 2002).

$MgCl_2$ (Samygina *et al.*, 2007), which is not compatible with the crystallization conditions found to date for Mt-PPase.

3.1. Structural comparison between 1wcf and 1sxv

The active sites of both structures are characterized by the absence of the metal cofactor and the presence of an anion in the catalytic pocket: phosphate for 1wcf and sulfate for 1sxv. The 1sxv crystals were obtained from 1.7 *M* ammonium sulfate and 100 m*M* sodium acetate pH 5.0 (Tammenkoski *et al.*, 2005). The two anions occupy equivalent positions (Fig. 1) and make similar interactions with the active-site residues. The site is the same as that for one of the phosphate ions in *S. cerevisiae* PPase (Y-PPase) in complex with four Mn atoms and two phosphates (Heikinheimo *et al.*, 1996) and the sulfate ion in the *E. coli* enzyme (Avaeva *et al.*, 1997; PDB entry 1jfd).

The program *LSQKAB* (Kabsch, 1976) was used to superimpose 1wcf and 1sxv, giving an r.m.s.d. of 0.65 Å, with three regions of the main chain showing the largest differences (Fig. 2). The differences between the two structures are not a consequence of crystal packing. The differences at residues 47–56 and 131–137 arise from the presence of the two potassium ions in 1wcf. Those of residues 82–89 arise from the movement of His86 in 1sxv caused by the formation of a hydrogen bond to the sulfate in the active site, which is replaced in 1wcf by a phosphate which does not interact with His86. The Cα atom of His86 moves by 1.84 Å, indicating a significant rearrangement of the catalytic residues and supporting direct involvement of His86 in modulating the catalytic properties. The different interactions with the anions in the two structures almost certainly reflect a change in the protonation state of His86 $N^{\varepsilon 2}$ (pH 5.0 for 1sxv and pH 7.0 for 1wcf, assuming a theoretical pK_a of 6.3 for the histidine).

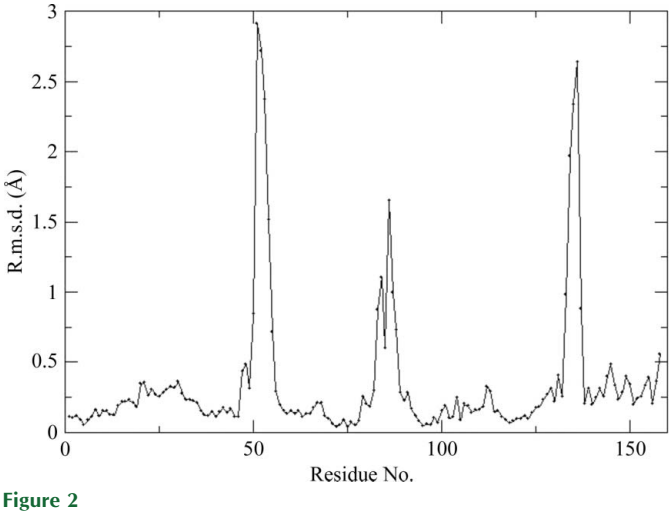


Figure 2
R.m.s.d. in Cα positions *versus* residue number for the superposition between 1sxv (cyan) and 1wcf.

3.2. Potassium-binding sites

The $2F_o - F_c$ and $F_o - F_c$ electron-density maps of 1wcf showed two peaks that could be modelled as either Na^+ or K^+ with similar ligand distances and appropriate geometry. An anomalous difference Fourier calculated with the data collected in-house using Cu Kα radiation confirmed the presence of potassium (Figs. 3*a* and 3*b*). In the anomalous map the electron density observed clearly showed the K^+ ions as well as the S atoms of the methionine and cysteine residues and, less clearly, the phosphorus of the $NaKHPO_4$. One K^+ ion (K1) is bound to four main-chain carbonyl O atoms (Asp128 O$-K^+$ = 3.02 Å, Lys127 O$-K^+$ = 2.83 Å, Glu130 O$-K^+$ = 2.61 Å and Lys133 O$-K^+$ = 2.67 Å), stabilizing a loop, and to a water molecule (Fig. 3). In 1sxv there is a water molecule close to this position and there is a sodium ion at this position in Ec-PPase (Samygina *et al.*, 2007). The second K^+ site, K2, roughly corresponds to the M1 site occupied by one of the catalytic Mg^{2+} ions in the *E. coli* holoenzyme (PDB entry 1obw; Harutyunyan *et al.*, 1997). K2 is bound to three residues, Asp89 (Asp89 $O^{\delta 1}-K^+$ = 2.85 Å), Asp57 (Asp57 $O^{\delta 1}-K^+$ = 3.30 Å) and Pro55 (Pro55 O$-K^+$ = 3.59 Å) (Fig. 3), which correspond to those in *E. coli* that coordinate the catalytic Mg^{2+} ion in the M1 site (Asp102, Asp70 and Pro68 in 1obw). The K^+ ions in the structure reflect the high concentration of potassium in the crystallization buffer rather than any physiological role.

4. Conclusions

This paper describes the structure of Mt-PPase determined at pH 7.0 with a phosphate anion bound in the active site. In addition, the structure contains two unexpected potassium ions as a result of the crystallization buffer. The structure supports the direct involvement of His86 in modulating catalysis by interacting with the pyrophosphate. The ability of His86 to alternate between two neighbouring

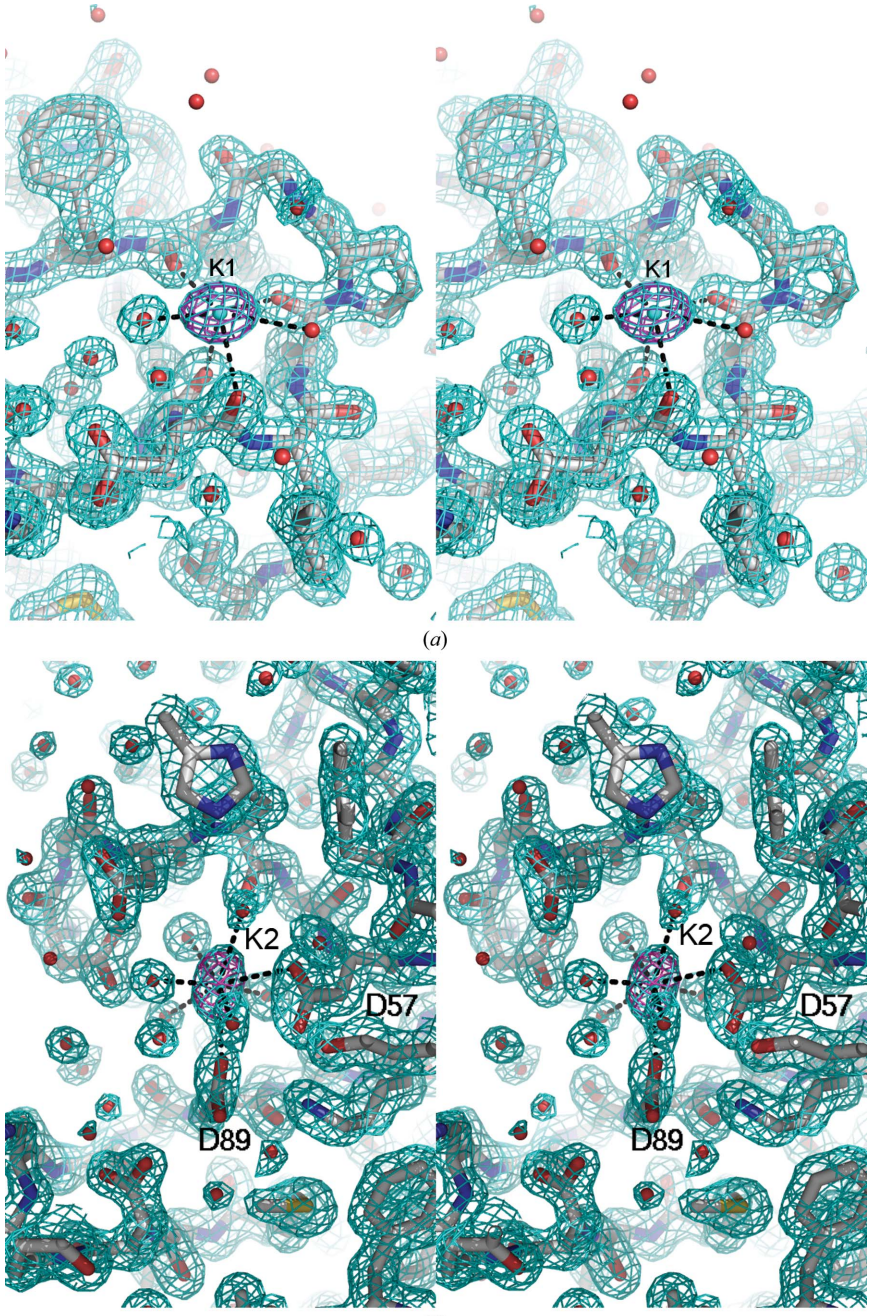


Figure 3
Stereoview of the regions around the two potassium-binding sites: $2F_o - F_c$ map contoured at 1.5σ (cyan) and anomalous difference Fourier contoured at 5σ (magenta). The latter reveals the presence of the potassium ion. (*a*) shows the K1 site and (*b*) the K2 site. This figure was prepared with *PyMOL* (DeLano, 2002).

positions with a Cα movement of 1.84 Å could favour ligand binding or product release. Unfortunately, no metal cofactors were bound in the active site and further studies with the metal cofactor and the pyrophosphate bound will be required to establish the full catalytic role of the histidines.

The authors thank the European Union for supporting access to the ESRF ID14-1 X-ray beamline and the beamline scientists for user support during data collection. The authors thank the EU for support through the grant 'Structural and Functional Genomics of *Mycobacterium tuberculosis*' (QLK2-CT-2001-02018).

References

Ahn, S., Milner, A. J., Fütterer, K., Konopka, M., Ilias, M., Young, T. W. & White, S. A. (2001). *J. Mol. Biol.* **313**, 797–811.
Avaeva, S., Kurilova, S., Nazarova, T., Rodina, E., Vorobyeva, N., Sklyankina, V., Grigorjeva, O., Harutyunyan, E., Oganessyan, V., Wilson, K., Dauter, Z., Huber, R. & Mather, T. (1997). *FEBS Lett.* **410**, 502–508.
Brzozowski, A. M. & Walton, J. (2001). *J. Appl. Cryst.* **34**, 97–101.
Chen, J., Brevet, A., Fromant, M., Lévêque, F., Schmitter, J. M., Blanquet, S. & Plateau, P. (1990). *J. Bacteriol.* **172**, 5686–5689.
Cole, S. T. (1998). *Int. J. Lepr. Other Mycobact. Dis.* **66**, 589–591.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Eiglmeier, K., Parkhill, J., Honoré, N., Garnier, T., Tekaia, F., Telenti, A., Klatser, P., James, K. D., Thomson, N. R., Wheeler, P. R., Churcher, C.,

Harris, D., Mungall, K., Barrell, B. G. & Cole, S. T. (2001). *Lepr. Rev.* **72**, 387–398.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gómez-García, M. R., Losada, M. & Serrano, A. (2006). *Biochem. J.* **395**, 211–221.
Harutyunyan, E. H., Oganessyan, V. Y., Oganessyan, N. N., Avaeva, S. M., Nazarova, T. I., Vorobyeva, N. N., Kurilova, S. A., Huber, R. & Mather, T. (1997). *Biochemistry*, **36**, 7754–7760.
Heikinheimo, P., Lehtonen, J., Baykov, A., Lahti, R., Cooperman, B. S. & Goldman, A. (1996). *Structure*, **4**, 1491–1508.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Kornberg, A. (1962). *Horizons in Biochemistry*, edited by H. Kasha & B. Pullman, pp. 251–264. New York: Academic Press.
Leppänen, V. M., Nummelin, H., Hansen, T., Lahti, R., Schäfer, G. & Goldman, A. (1999). *Protein Sci.* **8**, 1218–1231.
Liu, B., Bartlam, M., Gao, R., Zhou, W., Pang, H., Liu, Y., Feng, Y. & Rao, Z. (2004). *Biophys. J.* **86**, 420–427.
Lundin, M., Baltscheffsky, H. & Ronne, H. (1991). *J. Biol. Chem.* **266**, 12168–12172.
Merckel, M. C., Fabrichniy, I. P., Salminen, A., Kalkkinen, N., Baykov, A. A., Lahti, R. & Goldman, A. (2001). *Structure*, **9**, 289–297.
Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. & Lamzin, V. S. (2004). *J. Synchrotron Rad.* **11**, 56–59.
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.
Oganessyan, V. Y., Kurilova, S. A., Vorobyeva, N. N., Nazarova, T. I., Popov, A. N., Lebedev, A. A., Avaeva, S. M. & Harutyunyan, E. H. (1994). *FEBS Lett.* **348**, 301–304.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rodina, E. V., Vainonen, L. P., Vorobyeva, N. N., Kurilova, S. A., Sitnik, T. S. & Nazarova, T. I. (2008). *Biochemistry*, **73**, 897–905.
Samygina, V. R., Moiseev, V. M., Rodina, E. V., Vorobyeva, N. N., Popov, A. N., Kurilova, S. A., Nazarova, T. I., Avaeva, S. M. & Bartunik, H. D. (2007). *J. Mol. Biol.* **366**, 1305–1317.
Shintani, T., Uchiumi, T., Yonezawa, T., Salminen, A., Baykov, A. A., Lahti, R. & Hachimori, A. (1998). *FEBS Lett.* **439**, 263–266.
Sivula, T., Salminen, A., Parfenyev, A. N., Pohjanjoki, P., Goldman, A., Cooperman, B. S., Baykov, A. A. & Lahti, R. (1999). *FEBS Lett.* **454**, 75–80.
Tammenkoski, M., Benini, S., Magretova, N. N., Baykov, A. A. & Lahti, R. (2005). *J. Biol. Chem.* **280**, 41819–41826.
Teplyakov, A., Obmolova, G., Wilson, K. S., Ishii, K., Kaji, H., Samejima, T. & Kuranova, I. (1994). *Protein Sci.* **3**, 1098–1107.
Triccas, J. A. & Gicquel, B. (2001). *BMC Microbiol.* **1**, 3.
Wu, C. A., Lokanath, N. K., Kim, D. Y., Park, H. J., Hwang, H.-Y., Kim, S. T., Suh, S. W. & Kim, K. K. (2005). *Acta Cryst.* D**61**, 1459–1464.
Young, T. W., Kuhn, N. J., Wadeson, A., Ward, S., Burges, D. & Cooke, G. D. (1998). *Microbiology*, **144**, 2563–2571.