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

Cloning, expression, purification and preliminary X-ray diffraction studies of a putative *Mycobacterium smegmatis* thiolase

Neelanjana Janardan,[a] Anju Paul,[a] Rajesh K. Harijan,[a] Rikkert K. Wierenga[b] and M. R. N. Murthy[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560 012, India, and [b]Department of Biochemistry, Biocenter Oulu, University of Oulu, FIN-90220 Oulu, Finland

Correspondence e-mail: mrn@mbu.iisc.ernet.in

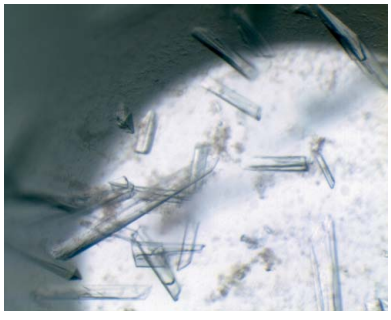
Received 18 April 2011
Accepted 21 May 2011

Thiolases are important in fatty-acid degradation and biosynthetic pathways. Analysis of the genomic sequence of *Mycobacterium smegmatis* suggests the presence of several putative thiolase genes. One of these genes appears to code for an SCP-x protein. Human SCP-x consists of an N-terminal domain (referred to as SCP2 thiolase) and a C-terminal domain (referred as sterol carrier protein 2). Here, the cloning, expression, purification and crystallization of this putative SCP-x protein from *M. smegmatis* are reported. The crystals diffracted X-rays to 2.5 Å resolution and belonged to the triclinic space group *P*1. Calculation of rotation functions using X-ray diffraction data suggests that the protein is likely to possess a hexameric oligomerization with 32 symmetry which has not been observed in the other six known classes of this enzyme.

1. Introduction

Thiolases belong to a ubiquitous group of enzymes. These enzymes are part of the β-oxidation pathway for fatty-acid degradation as well as of various biosynthetic pathways. In the β-oxidation pathway (Hiltunen & Qin, 2000; Bhaumik *et al.*, 2005) thiolases catalyze the shortening of the fatty-acid chains, whereas the biosynthetic thiolases catalyze the formation of a carbon–carbon bond *via* a thioester-dependent Claisen condensation reaction (Heath & Rock, 2002; Austin & Noel, 2003). All thiolases can catalyze both this Claisen condensation reaction as well as the reverse reaction, the degradative breakdown of a 3-ketoacyl-CoA molecule. In the latter reaction 3-ketoacyl-CoA is shortened by a two-carbon unit. Acetyl-CoA is formed as a product (Fig. 1). Thiolases exhibit variations in their oligomeric state (dimers or tetramers), cellular location (for example, in humans thiolases occur in the cytosol, peroxisomes and mitochondria) and function (degradative and biosynthetic). The smallest oligomeric state of all studied thiolases is a dimer, and thiolase tetramers are dimers of dimers. Each thiolase subunit is about 400 residues long and contains three domains: an N-terminal domain, a loop domain and a C-terminal domain (Haapalainen *et al.*, 2006). The N- and C-terminal domains of the thiolases have a similar topology (Mathieu *et al.*, 1994). Most of the catalytic loops and residues are in the C-terminal half, but the conserved reactive nucleophilic cysteine is in the N-terminal half of the protomer. Catalysis in the synthetic direction as well as in the degradative direction starts with acylation of this active-site cysteine (Haapalainen *et al.*, 2006). Sequence fingerprints have been described for each of the four important catalytic loops (Haapalainen *et al.*, 2006, 2007).

The thiolase fold is representative of the fold of a large superfamily of enzymes (Haapalainen *et al.*, 2006). The thiolases themselves form a subset of this superfamily. In humans, six classes of thiolases have been identified (T1, T2, TFE, AB, SCP2 and CT; Fukao, 2002).



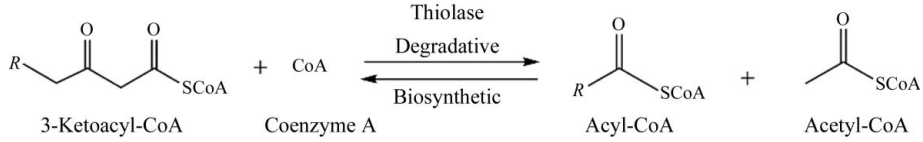


Figure 1
The reaction of thiolase in the degradative direction.

T1 (Gilbert *et al.*, 1981) and T2 (Haapalainen *et al.*, 2007) are mitochondrial tetrameric thiolases. AB (Antonenkov *et al.*, 1999) and SCP2 (Antonenkov *et al.*, 1997; Takeuchi *et al.*, 2004) are peroxisomal dimeric thiolases, CT (Kursula *et al.*, 2005) is expressed in cytosol and is a tetrameric thiolase and the TFE (Fould *et al.*, 2010) thiolase is a mitochondrial dimeric thiolase; it is the β-chain of the $\alpha_2\beta_2$ TFE complex (TFE is the mitochondrial trifunctional enzyme), whereas the α-chain is homologous to a peroxisomal multifunctional enzyme type I (MFE1; Kasaragod *et al.*, 2010). AB, T2 and CT are well studied thiolases; crystal structures are known of each of the human thiolases or their homologues (Mathieu *et al.*, 1997). The crystal structures of the other thiolases (T1, SCP2 and TFE) are unknown. AB has been reported to be a degradative thiolase. T2 is involved in ketone-body degradation and synthesis (Fukao *et al.*, 1997; Fukao, 2002). CT is a biosynthetic thiolase. T1, TFE and SCP2 have been proposed to be degradative thiolases (Fig. 1). The SCP2 thiolase is the N-terminal part of the SCP-x protein, whereas the C-terminal part is the SCP2 domain (Seedorf *et al.*, 1994).

From sequence and keyword searches of the *Mycobacterium smegmatis* genome, several putative thiolase genes have been found. These bioinformatics studies reveal that the distributions of thiolases in mycobacteria and humans are rather different, suggesting differences in the functionality of thiolases in humans and mycobacteria. Structural studies of mycobacterial thiolases are therefore of great biochemical interest for establishing the function of these thiolases as well as for the possible development of new antibiotics against these pathogenic microorganisms.

One of the identified genes corresponded to a polypeptide annotated as an acetyl-CoA acetyltransferase (YP_889758). The corresponding polypeptide has 507 amino-acid residues. According to the annotation, this thiolase could be classified as an SCP-x protein predicted to have an N-terminal thiolase part and a C-terminal extra domain. In this paper, we report the cloning, expression, purification and preliminary X-ray crystallographic studies of this SCP-x protein. Structure determination of this protein will provide valuable insights into the function of this putative thiolase.

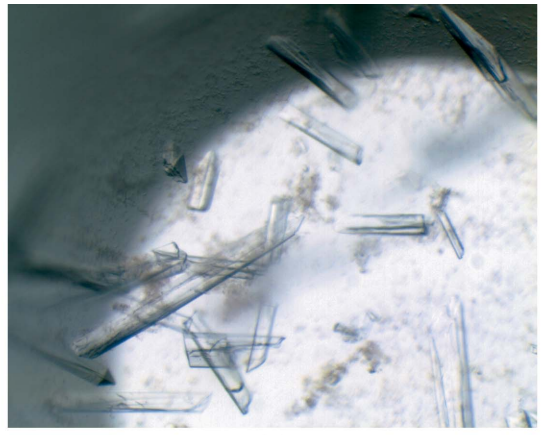
2. Materials and methods

2.1. Cloning

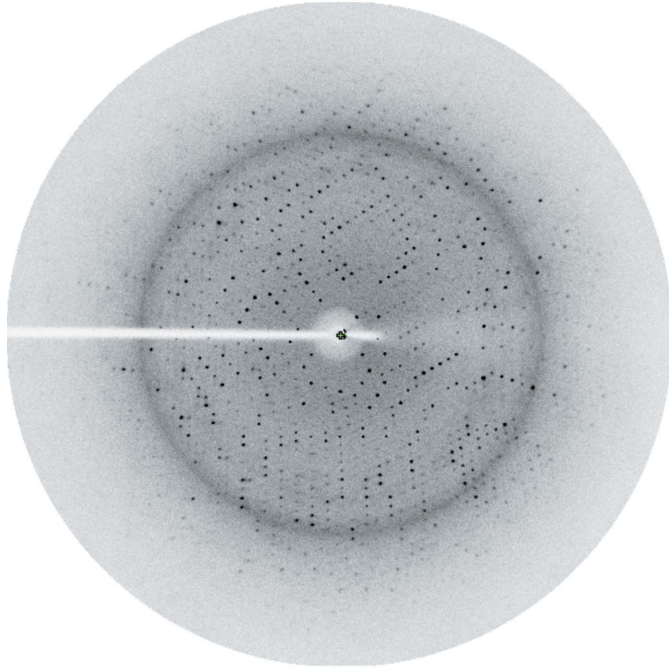
M. smegmatis strain MC2 155 genomic DNA was generously provided by Professor Dipankar Chatterji, Molecular Biophysics Unit, Indian Institute of Science, India. The gene coding for the putative SCP-x protein was PCR-amplified from *M. smegmatis* genomic DNA using Kod DNA polymerase (Merck, Germany) and specific sense (5′-CATATG**GCTAGC**ATGGTCGATCCGCGTACG-C-3′) and antisense (5′-**GGATCC**TTACTCGAGCGCCAGTACGG-CGC-3′) primers. The highlighted sequences in the sense and antisense primers represent *Nhe*I and *Bam*HI restriction sites, respectively. The PCR-amplified fragment was digested with *Nhe*I and *Bam*HI (Fermentas) and cloned into pRSETC vector (Invitrogen), an expression vector encoding an N-terminal hexahistidine tag to facilitate protein purification using Ni–NTA affinity column chromatography. The clone thus obtained was confirmed by DNA sequencing.

2.2. Overexpression and purification

Recombinant protein was overexpressed in *Escherichia coli* host strain BL21 (DE3) Rosetta. The clone was transformed into competent BL21 (DE3) Rosetta cells and transformants were selected on LB agar plates containing 100 µg ml^{-1} ampicillin. Pre-inoculum grown overnight in LB–ampicillin medium from a single cell was diluted into 3 l fresh LB medium containing 100 µg ml^{-1} ampicillin and grown at 310 K until the OD_{600} reached 0.6–0.7. Expression of the protein was induced by the addition of 0.3 m*M* IPTG and the cells were allowed to grow for an additional 12 h at 288 K. The cells recovered by centrifugation were resuspended in 150 ml cold lysis buffer consisting of 50 m*M* Tris pH 7.8, 400 m*M* NaCl, 10% glycerol, 0.1% Triton X-100 and 5 m*M* β-mercaptoethanol. The cell suspension was lysed by sonication on ice. All subsequent purification steps were performed at 277 K. Cell debris was removed from the lysate by centrifugation. Ni^{2+}–NTA beads pre-equilibrated with lysis buffer were added to the supernatant and the protein was allowed to bind for 3 h on a rotor incubator. The protein-bound matrix was packed into a column and washed with at least three column volumes of wash buffer 1 (100 m*M* NaCl, 50 m*M* Tris–HCl pH 8.0, 10% glycerol and 5 m*M* β-mercaptoethanol) and two column volumes of wash buffer 2 (100 m*M* NaCl, 50 m*M* Tris–HCl pH 8.0, 10% glycerol, 5 m*M* β-mercaptoethanol and 20 m*M* imidazole). The protein was eluted



(*a*)



(*b*)

Figure 2
(*a*) SCP-x crystals obtained from 2.5 mg ml^{-1} protein solution in 100 m*M* HEPES sodium salt pH 7.5 containing 20%(*w*/*v*) PEG 4000, 10%(*v*/*v*) 2-propanol. (*b*) A typical 0.75° oscillation image from an SCP-x crystal.

Table 1
Data-collection statistics for *M. smegmatis* SCP-x.

Values in parentheses are for the highest resolution shell.

Space group	*P*1
Wavelength (Å)	1.5418
Temperature (K)	100
Resolution range (Å)	4.06–2.44 (2.57–2.44)
Unit-cell parameters (Å, °)	a = 96.30, b = 101.95, c = 102.40, α = 116.32, β = 101.20, γ = 97.71
Observed reflections	420328
Unique reflections	114169
Data completeness (%)	92.5 (74.9)
Multiplicity	3.7
$\langle I/\sigma(I)\rangle$†	16.7 (2.8)
$R_{\rm merge}$‡ (%)	5.0 (36.0)

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

using 100 m*M* NaCl, 50 m*M* Tris–HCl pH 8.0, 10% glycerol, 5 m*M* β-mercaptoethanol, 200 m*M* imidazole. The eluted protein was concentrated to 1 ml by low-speed centrifugation using a 30 kDa molecular-weight cutoff Centricon (Amicon). Further purification was performed by size-exclusion chromatography using an ÄKTA-basic 10 liquid-chromatography system (GE Healthcare, Uppsala, Sweden) with a Superose 6 HR 10/30 column. Chromatography was carried out in 50 m*M* Tris–HCl buffer pH 8.0 containing 100 m*M* NaCl, 10% glycerol, 5 m*M* β-mercaptoethanol. Purified protein was concentrated to 5 mg ml^{-1}. The protein concentration was estimated by the Bradford method (Bradford, 1976). The molecular weight and purity of the enzyme were checked on 12% SDS–PAGE (Laemmli, 1970).

2.3. Crystallization

Crystallization conditions were screened by the microbatch method at 291 K using Jena Biosciences crystallization screens. Each crystallization drop consisted of 2 µl protein solution and 2 µl well solution. Small crystals were obtained under ten different conditions. Diffraction-quality crystals were obtained on further optimization of JBS Basic II condition No. 3 [20%(*w*/*v*) PEG 4000, 10%(*v*/*v*) 2-propanol, 100 m*M* HEPES sodium salt pH 7.5]. These crystals appeared after six weeks of equilibration against the crystallization condition and grew to full size (1 mm) in two months (Fig. 2*a*).

2.4. Data collection

For X-ray diffraction data collection, the crystals were transferred to a CryoLoop and frozen at 100 K in a stream of nitrogen gas. Crystals diffracted to a resolution of 2.5 Å (Fig. 2*b*) when exposed to X-rays from a Rigaku RU-200 rotating-anode X-ray generator equipped with a 300 µm focal cup and a MAR Research image-plate detector system (diameter 345 mm). X-rays were focused with Osmic mirrors. A complete data set was collected from a single crystal. A total of 450 frames of 0.75° oscillation were recorded with an exposure time of 60 s per frame. The data were processed using *DENZO* and *SCALEPACK* from the *HKL*-2000 suite (Otwinowski & Minor, 1997). The final statistics for data collection and processing are summarized in Table 1.

3. Results and discussion

The putative thiolase gene from *M. smegmatis* was successfully cloned in the pRSETC vector with an N-terminal hexahistidine tag and purified using Ni–NTA column chromatography. Further purification was achieved by size-exclusion chromatography. The purity of the protein was ascertained by SDS–PAGE and MALDI mass spectrometry (data not shown). Crystals suitable for structural studies using X-ray diffraction were obtained in 20%(*w*/*v*) PEG 4000, 10%(*v*/*v*) 2-propanol, 100 m*M* HEPES sodium salt pH 7.5. A complete diffraction data set to 2.5 Å resolution was collected at 100 K. Data

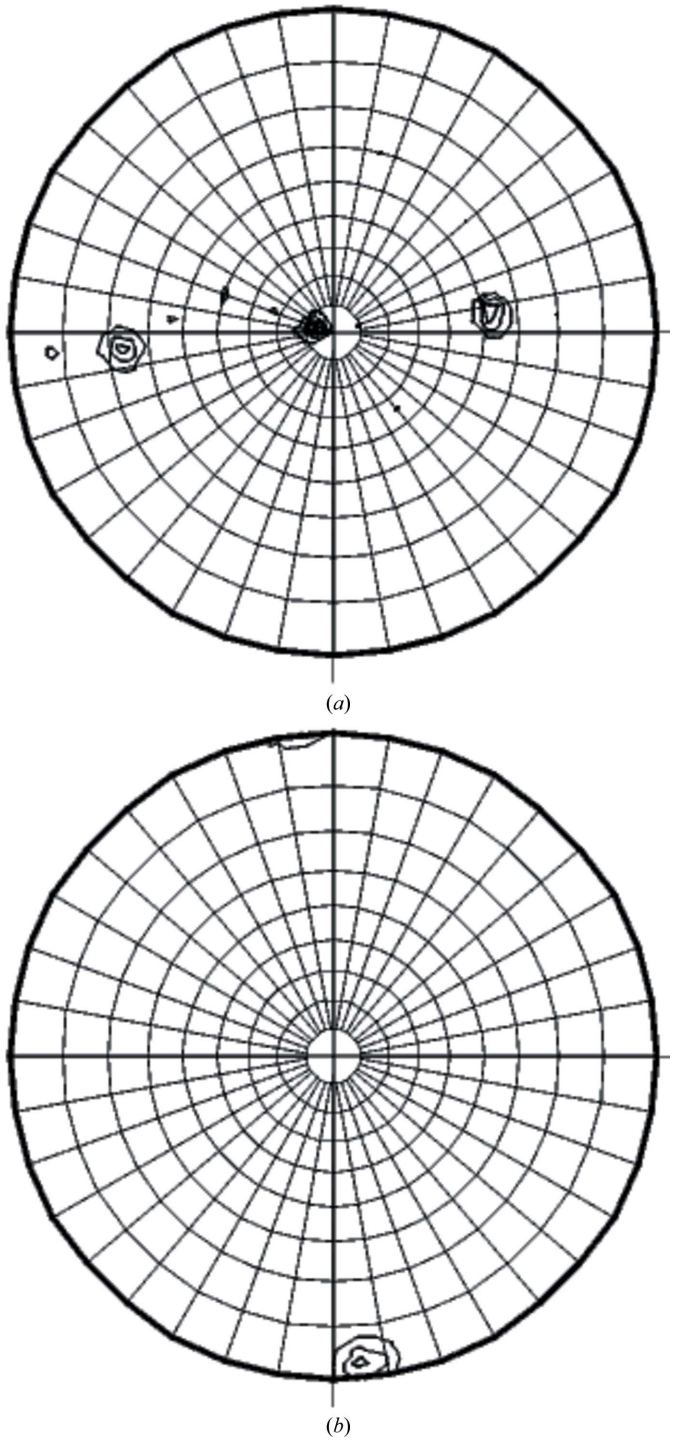


Figure 3
Self-rotation functions for the SCP-x data calculated using *MOLREP*. Data between 20.0 and 5.0 Å resolution and a radius of integration of 48 Å were used for the computation of rotation functions. (*a*) 180° hemisphere representing twofold-related peaks; (*b*) 120° hemisphere representing threefold-related peaks. The peaks are between 10.6σ and 18.5σ.

processing revealed that the crystal belonged to the triclinic space group $P1$, with unit-cell parameters a = 96.3, b = 101.9, c = 102.4 Å, α = 116.32, β = 101.20, γ = 97.71°. Calculation of the Matthews coefficient (Matthews, 1968) revealed that the unit cell is compatible with 4–8 protomers of molecular weight 54 kDa, corresponding to Matthews coefficients of 4.0–2.0 Å^3 Da^{-1}.

To examine the plausible oligomeric states of this putative thiolase, self-rotation functions were calculated using the *MOLREP* (Vagin & Teplyakov, 2010) program from the *CCP*4 suite (Winn *et al.*, 2011) for κ = 180, 120, 90 and 60° hemispheres corresponding to twofold, threefold, fourfold and sixfold noncrystallographic symmetry axes, respectively. Data between 20.0 and 5.0 Å resolution were used for rotation-function calculations with an integration radius of 48 Å. Strong peaks on the 180 and 120° hemispheres suggested the presence of noncrystallographic twofolds and a threefold, respectively (Fig. 3). The threefold peak was found to be perpendicular to the three independent twofold peaks. The values of (θ, φ, χ) (defined as in the *CCP*4 suite) for the three twofold peaks were (172.64, 77.38, 179.99°), (126.82, −81.71, 179.99°) and (113.07, 95.86, 180.00°), respectively. The values of (θ, φ, χ) for the threefold peak were (92.55, −173.52, 120.08°). These results suggest that the protein may be hexameric with 32 symmetry in the crystals.

The total length of the YP_889758 protein is 507 residues, which is significantly longer than the approximately 400 residues of the well characterized thiolases (Haapalainen *et al.*, 2006). It is possible that the C-terminal part is the extra domain as observed in the SCP-x protein. Since no structure with considerable sequence identity to this SCP-x protein is available, attempts to determine the crystal structure using single-wavelength or multiple-wavelength anomalous dispersion are now in progress. The unique sequence features of this *M. smegmatis* SCP-x protein indicates that its structure will be very interesting. Structural studies of crystals of the selenomethionine derivative have been initiated.

The intensity data were collected using the X-ray Facility for Structural Biology at the Molecular Biophysics Unit, Indian Institute of Science, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT) of the Government of India. This work was supported by an Indo–Finnish grant from the Department of Biotechnology, Government of India and the Academy of Finland (grant 1131795), Finland. We acknowledge the award of a Shyama Prasad Mukherji Fellowship from the Council for Scientific and Industrial Research (CSIR), Government of India to NJ. We thank Professor D. Chatterji for the gift of *M. smegmatis* genomic DNA.

References

Antonenkov, V. D., Van Veldhoven, P. P., Waelkens, E. & Mannaerts, G. P. (1997). *J. Biol. Chem.* **272**, 26023–26031.
Antonenkov, V. D., Van Veldhoven, P. P., Waelkens, E. & Mannaerts, G. P. (1999). *Biochim. Biophys. Acta*, **1437**, 136–141.
Austin, M. B. & Noel, J. P. (2003). *Nat. Prod. Rep.* **20**, 79–110.
Bhaumik, P., Koski, M. K., Glumoff, T., Hiltunen, J. K. & Wierenga, R. K. (2005). *Curr. Opin. Struct. Biol.* **15**, 621–628.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Fould, B., Garlatti, V., Neumann, E., Fenel, D., Gaboriaud, C. & Arlaud, G. J. (2010). *Biochemistry*, **49**, 8608–8617.
Fukao, T. (2002). *Wiley Encyclopedia of Molecular Medicine*, Vol. 5, pp. 6–9. New York: John Wiley & Sons.
Fukao, T., Song, X.-Q., Mitchell, G. A., Yamaguchi, S., Sukegawa, K., Orii, T. & Kondo, N. (1997). *Pediatr. Res.* **42**, 498–502.
Gilbert, H. F., Lennox, B. J., Mossman, C. D. & Carle, W. C. (1981). *J. Biol. Chem.* **256**, 7371–7377.
Haapalainen, A. M., Meriläinen, G., Pirilä, P. L., Kondo, N., Fukao, T. & Wierenga, R. K. (2007). *Biochemistry*, **46**, 4305–4321.
Haapalainen, A. M., Meriläinen, G. & Wierenga, R. K. (2006). *Trends Biochem. Sci.* **31**, 64–71.
Heath, R. J. & Rock, C. O. (2002). *Nat. Prod. Rep.* **19**, 581–596.
Hiltunen, J. K. & Qin, Y. (2000). *Biochim. Biophys. Acta*, **1484**, 117–128.
Kasaragod, P., Venkatesan, R., Kiema, T. R., Hiltunen, J. K. & Wierenga, R. K. (2010). *J. Biol. Chem.* **285**, 24089–24098.
Kursula, P., Sikkilä, H., Fukao, T., Kondo, N. & Wierenga, R. K. (2005). *J. Mol. Biol.* **347**, 189–201.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Mathieu, M., Modis, Y., Zeelen, J. P., Engel, C. K., Abagyan, R. A., Ahlberg, A., Rasmussen, B., Lamzin, V. S., Kunau, W. H. & Wierenga, R. K. (1997). *J. Mol. Biol.* **273**, 714–728.
Mathieu, M., Zeelen, J. P., Pauptit, R. A., Erdmann, R., Kunau, W. H. & Wierenga, R. K. (1994). *Structure*, **15**, 797–808.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Seedorf, U., Brysch, P., Engel, T., Schrage, K. & Assmann, G. (1994). *J. Biol. Chem.* **269**, 21277–21283.
Takeuchi, H., Chen, J.-H., Jenkins, J. R., Bun-Ya, M., Turner, P. C. & Rees, H. H. (2004). *Biochem. J.* **382**, 93–100.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.