Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, crystallization and preliminary characterization of a β-carbonic anhydrase from *Escherichia coli*

Jeff D. Cronk,[a] Jason W. O'Neill,[a] Michelle R. Cronk,[a] James A. Endrizzi[b] and Kam Y. J. Zhang[a]*

[a]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA, and [b]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3206, USA

Correspondence e-mail: kzhang@fhcrc.org

Received 24 March 2000
Accepted 16 June 2000

Carbonic anhydrases are zinc metalloenzymes that fall into three distinct evolutionary and structural classes, α, β and γ. Although α-class enzymes, particularly mammalian carbonic anhydrase II, have been the subject of extensive structural studies, for the β class, consisting of a wide variety of prokaryotic and plant chloroplast carbonic anhydrases, the structural data is quite limited. A member of the β class from *E. coli* (CynT2) has been crystallized in native and selenomethionine-labelled forms and multiwavelength anomalous dispersion techniques have been applied in order to determine the positions of anomalous scatterers. The resulting phase information is sufficient to produce an interpretable electron-density map. A crystal structure for CynT2 would contribute significantly to the emerging structural knowledge of a biologically important class of enzymes that perform critical functions in carbon fixation and prokaryotic metabolism.

1. Introduction

Carbonic anhydrase (carbonate hydrolase; E.C. 4.2.1.1) has long been one of the most well characterized enzymes. The reaction catalyzed,

$$CO_2 + H_2O \longleftrightarrow H^+ + HCO_3^-,$$

the interconversion of carbon dioxide and bicarbonate, is of fundamental physiological and biochemical importance, relating in various *in vivo* contexts to pH or ionic balance and metabolic flux (Henry, 1996). The phylogenetic distribution of carbonic anhydrase (CA) is quite broad, but the best studied have been the enzymes from mammalian sources. Since the first high-resolution CA crystal structure, that of human carbonic anhydrase II (Kannan *et al.*, 1971), the structures of other CA isozymes have been determined and the enzyme has been the subject of a wide range of structural investigations, encompassing mechanistic, inhibition, folding and design studies (for examples, see Christianson & Fierke, 1996; Lindskog, 1997, and references therein). All CAs from terrestrial vertebrates are zinc metalloenzymes in which the coordination sphere is comprised of three histidine side chains and an activated water molecule. The polarizing effect of coordination to the Zn^{2+} cation lowers the pK_a of the water, which ionizes to hydroxide, the latter being the nucleophile that attacks the substrate carbon (Christianson & Cox, 1999). These enzymes also share a core architecture of a ten-stranded, highly curved and mostly antiparallel β-sheet.

When sequences of carbonic anhydrase enzymes from plant chloroplasts became available, it was evident that they bore no relationship to those of the mammalian enzymes (Burnell *et al.*, 1990; Roeske & Ogren, 1990; Fawcett *et al.*, 1990; Majeau & Coleman, 1991). Instead, it was found that prokaryotic and plant CAs form a widely distributed and evolutionarily distinct class, the β-class CAs (Hewett-Emmett & Tashian, 1996). Carbonic anhydrases have been shown to have important and even essential roles in prokaryotic and plant metabolism (Fukuzawa *et al.*, 1992; Kozliak *et al.*, 2000). For instance, the function of these enzymes has been implicated as rate-limiting for carbon fixation in most plants (Badger & Price, 1994). In sharp contrast to the well known mammalian CAs, all belonging to the α class (Hewett-Emmett & Tashian, 1996), and the more recently characterized γ class (Alber & Ferry, 1994; Kisker *et al.*, 1996), the β-CAs have significant α-helical character and have two cysteines and one histidine as Zn^{2+} ligands (Rowlett *et al.*, 1994; Hiltonen *et al.*, 1998). Structures of two β-class members from *Porphyridium purpureum* and *Pisum sativum* have recently been reported, revealing a third structural prototype for carbonic anhydrases (Mitsuhashi *et al.*, 2000; Kimber & Pai, 2000). However, the catalytic zinc ion is coordinated by a conserved aspartate residue in addition to the two cysteines and one histidine residues in *P. purpureum* β-CA, whereas the zinc ion is coordinated by an acetate ion as well as the three residues expected as ligands in *P. sativum* β-CA. The functional significance of this

difference in Zn^{2+} coordination remains to be discovered.

Two members of the β class of carbonic anhydrases have been identified in *E. coli*. The first to be characterized is CynT, which is encoded by a part of the *cyn* operon (Sung & Fuchs, 1988), whose function allows *E. coli* to utilize cyanate as a nitrogen source (Guilloton *et al.*, 1992, 1993). The second member was found as a result of efforts to sequence the *E. coli* genome (Fujita *et al.*, 1994). Designated as *cynT2* (Hewett-Emmett & Tashian, 1996) or *yadF*, the translated open reading frame displayed 30% homology to CynT. No biochemical or functional studies have been reported for the CynT2 protein, but our observations indicate that its expression is up-regulated under stress conditions arising in standard liquid-culture growth. A crystal structure would greatly facilitate our understanding of the function of this protein.

We report here the cloning, purification, and crystallization of CynT2, including the characterization of three different native crystal forms. In addition, we have expressed, purified and crystallized a selenomethionine-substituted form of the protein. Collection and analysis of a multi-wavelength data set from the latter has confirmed the presence of zinc and led to determination of the positions of the zinc cations within the crystal unit cell, which in turn led to the identification of selenium sites.

2. Experimental procedures and results

2.1. Cloning of CynT2

The coding region of the *cynT2* (*yadF*) gene (GenBank PID 1786318) was cloned by PCR using primers CATATGAAAGACATAGATACACTCATCAGC (forward) and AGATCTATTTGTGGTTGGCGTGTTTCA (reverse) that incorporate *Nde*I and *Bgl*II sites at the 5′ and 3′ ends, respectively. The template was prepared from common laboratory K12 strains. The 670 bp PCR product was cloned into pPCR-Script (Stratagene), where its presence was confirmed by PCR assay, analytical restriction digests and sequencing. The *Nde*I–*Bgl*II fragment was then excized and ligated to *Nde*I + *Bam*HI-digested pET-3a expression vector (Novagen) to form the expression construct, denoted pMRC8. The presence of the insert was again confirmed by PCR assay and restriction analysis. Sequencing was performed to confirm the correctness of the coding sequence. The construct pMRC8 was transformed into BL21(DE3) cells and expression tests performed.

2.2. Protein expression and purification

Analysis by SDS–PAGE of lysates derived from cultures of bacteria harboring the pMRC8 showed a prominent 25 kDa band corresponding to the expected molecular weight for CynT2. Expression was scaled up to 5 × 1 l batches (LB media). Harvested cells were lysed by sonication in lysis/binding buffer consisting of 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 5 m*M* imidazole and PMSF at 1 m*M*. The lysate was brought to 0.1% Triton X-100 and clarified by centrifugation and applied to a 10 ml Ni^{2+}-NTA agarose (Qiagen) column. The column was washed extensively with lysis/binding buffer, prior to elution of bound protein with 20 m*M* Tris–HCl, 500 m*M* NaCl, 50 m*M* imidazole. The combined eluate was concentrated and buffer exchanged to 30 m*M* Tris–HCl, 150 m*M* NaCl pH 8.4 (TBS) using Centriprep-10 and Centricon-10 filtration devices (Amicon). The purity and native molecular weight was assessed by analytical size-exclusion chromatography. Mass spectrometry was performed to confirm that the expressed protein is the product of the *E. coli cynT2* gene and yielded a value of 25098.5 Da compared with a calculated MW of 25096.88 Da.

For the production of CynT2 protein containing selenomethionine, the B834(DE3) strain of *E. coli* (Novagen) was transformed with pMRC8 and grown overnight in minimal media (MM) supplemented with all 20 amino acids (including Met) and several vitamins (Ramakrishnan & Graziano, 1997). A new culture was inoculated by a 1:100 dilution of the overnight culture into 0.33 l fresh MM supplemented similarly except with methionine replaced by selenomethionine. This expression culture was grown to an A_{600} of 0.57 before addition of IPTG to 0.5 m*M*, followed by continued growth to a final A_{600} of 1.5. For both expression protocols, the lysis and binding was performed as described above for the native protein.

2.3. Crystallization and initial characterization of crystal forms

All crystallizations were performed using the hanging-drop method, using protein concentrations ranging from 11–14 mg ml^{-1}. Drops were prepared by mixing 2–4 µl protein with an equivalent volume of reservoir solution with the compositions indicated below and were equilibrated against 1 ml of reservoir solution. Initial crystallization trials employed an ammonium sulfate *versus* pH grid, as well as a sparse-matrix screen (Jancarik & Kim, 1991). Conditions from both screens produced crystals and these conditions were refined. Crystals grown from 1.6 *M* ammonium sulfate and 0.1 *M* MES pH 6.3 at 293 K are of two distinct primitive tetragonal forms, depending on whether 4% PEG 400 is

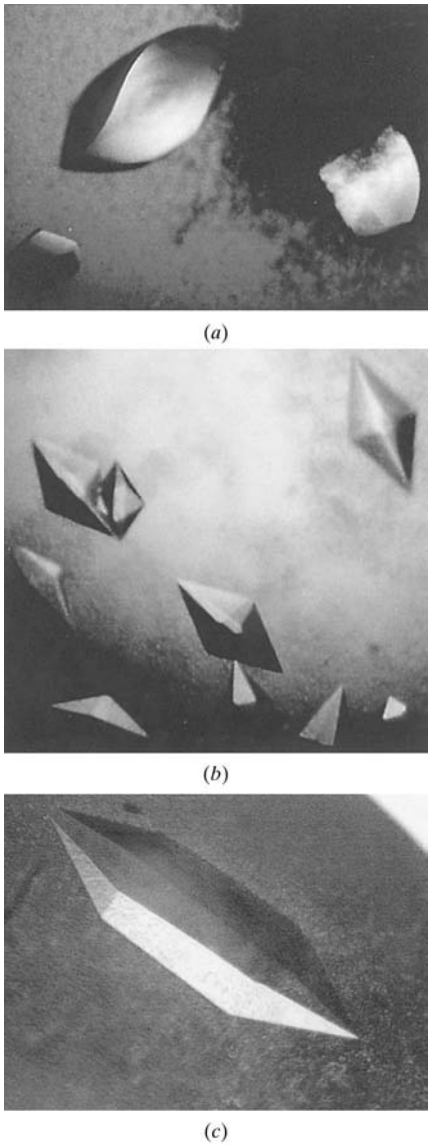

Figure 1
Three CynT2 crystal forms. (*a*) Form 1 crystals grown from 1.8 *M* $(NH_4)_2SO_4$, 0.1 *M* MES pH 6.15. Typical crystal dimensions are 0.4 × 0.5 × 0.8 mm. The protein concentration is approximately 12 mg ml^{-1}. (*b*) Form 2 crystals grown from conditions similar to those producing form 1, except for the addition of PEG 400 to 4% final concentration. Typical crystal dimensions are 0.25 × 0.3 × 0.6 mm. (*c*) Form 2 crystals of selenomethionine-substituted CynT2 grown from the same condition that produced form 2 native CynT2 crystals. Typical crystal dimensions are 0.3 × 0.4 × 0.75 mm.

Table 1
Summary of native CynT2 crystals.

Parameter	Form 1	Form 2	Form 3
d_{min} (Å)	2.0	2.89	2.29
Data-collection temperature (K)	291	100	100
Space group	$P4_22_12$	$P4_322$	$P2_1$
Unit-cell parameters (Å, °)	$a = b = 68.54$, $c = 85.88$	$a = b = 81.00$, $c = 161.98$	$a = 48.21$, $b = 140.73$, $c = 72.57$, $\beta = 99.52$
R_{merge}† (%)	6.5	7.0	4.0
$I/\sigma(I)$	20.4 (50–2.0 Å), 4.4 (2.03–2.00 Å)	17.6 (100–2.89 Å), 4.2 (2.94–2.89 Å)	13.1 (15.4–2.29 Å), 5.5 (2.33–2.29 Å)
No. of reflections measured	90500	106295	129133
Unique observations	13974	12595	42370
Completeness (%)	96.7 (50–2.0 Å), 96.9 (2.03–2.00 Å)	98.7 (100–2.89 Å), 96.3 (2.94–2.8 Å)	98.3 (15.4–2.29 Å), 99.3 (2.5–2.29 Å)
Mosaicity (°)	<0.25	0.65	0.25
Monomers per asymmetric unit	1	2	4
V_M, calc.	2.01	2.665	2.42
Solvent content (%)	39	54	49

† $R_{merge} = \sum||I_{obs}| - |\langle I\rangle||/\sum|\langle I\rangle|$.

Table 2
Synchrotron data-collection statistics of selenomethionine-substituted CynT2 crystals.

Values in parentheses are for the highest resolution shell, 2.25–2.13 Å.

Source	ALS 5.0.2
Crystal form	SeMet form 2 ($P4_322$)
d_{min} (Å)	2.13
Unit-cell parameters (Å)	$a = b = 81.24$, $c = 162.14$
Mosaicity (°)	0.39

Wavelengths (Å)	$\lambda_1 = 0.96112$	$\lambda_2 = 0.97950$	$\lambda_3 = 0.97970$	$\lambda_4 = 1.07812$	$\lambda_5 = 1.28109$
Unique observations	31137	31132	31133	30857	25441
Redundancy	7.9	7.9	8.0	7.7	6.9
Completeness (%)	99.5 (99.4)	99.5 (99.4)	99.6 (99.4)	98.6 (94.5)	81.5 (38.4)
$I/\sigma(I)$	19.7 (6.4)	22.0 (6.5)	22.2 (6.4)	25.2 (8.4)	28.0 (5.7)
R_{merge} (%)	7.0 (39.1)	6.4 (37.7)	6.7 (36.3)	5.4 (25.5)	4.9 (24.1)
R_{anom}†	7.0 (21.6)	7.6 (21.3)	7.2 (21.8)	5.1 (17.3)	5.0 (23.5)

† $R_{anom} = \{\langle(|F^+| - |F^-|)^2\rangle/[(\langle|F^+|\rangle^2 + \langle|F^-|\rangle^2)/2]\}^{1/2}$.

included. Form 1 crystals, which grow to dimensions as large as 1.0 mm long and 0.5 mm thick, display a characteristic spiral morphology (Fig. 1*a*) and belong to space group $P4_22_12$ ($a = b = 68$, $c = 86$ Å). Form 2, which occurs with PEG 400, are bipyramidal and generally smaller (Fig. 1*b*) and belong to space group $P4_322$ ($a = b = 81.9$, $c = 161.9$ Å). A third form, $P2_1$, arises in crystallizations carried out at 277 K with reservoir solutions similar to those producing form 1. Preliminary characterization of diffraction from CynT2 crystal forms was performed in-house using a RU-200B rotating-anode generator (Rigaku Corp.) operating at 50 kV and 100 mA producing Cu $K\alpha$ radiation at 1.5418 Å and an R-AXIS IV image-plate detector (Molecular Structure Corp.). Form 1 crystals diffracted to 1.8 Å at room temperature, but display relatively high mosaicity after freezing by a number of methods. Form 2 crystals incubated for a short time in 40% PEG 400 and flash-frozen in liquid nitrogen display mosaic spreads of 0.6–0.7° and good-quality diffraction under cryostream cooling to about 100 K, albeit to a somewhat lower resolution compared with form 1 (Table 1 and Fig. 1).

The possible presence of a zinc cofactor suggested CynT2 as a good candidate for a multiwavelength anomalous dispersion (MAD) phasing approach. This was supported by the initial characterizations of CynT2 crystals and production of the selenomethionine-substituted protein (at five Met residues per 25 kDa monomer), which forms crystals apparently identical to native (Fig. 1*c*). Preliminary diffraction studies using synchrotron radiation directly demonstrated the suitability of the selenomethionine-containing crystals for acquisition of a complete MAD data set.

2.4. Data collection and processing

Collections of complete data sets were performed in house (as described above), at beamline X12C at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) and at the Macromolecular Crystallography Facility on beamline 5.0.2 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA). Multiwavelength X-ray diffraction data from selenomethionine-labelled CynT2 crystals were acquired at both synchrotron sources at 100 K. Fluorescence scans were used to determine the energy of the Se edge and as a test for the presence of the Zn edge. This provided the first direct evidence for CynT2 of a Zn^{2+} cofactor, which is the native catalytic metal ion for all carbonic anhydrase enzymes characterized to date.

For the ALS data set (Table 2), data were collected at four wavelengths at or near the Se edge in the vicinity of 0.98 Å, plus a fifth wavelength, 1.28 Å, corresponding to the peak anomalous scattering for zinc. Two wedges of reciprocal space, each consisting of 40 1.3° oscillation images, were recorded for each wavelength, 180° apart, for equivalent measurement of Friedel pairs.

The acquired data was reduced and scaled in preparation for phase calculations using either *ELVES* (J. M. Holton & T. Alber, personal communication), which generates shell scripts that perform intensity measurements using *MOSFLM* (Leslie *et al.*, 1986) and data-scaling procedures by *SCALA* (Collaborative Computational Project, Number 4, 1994), or *DENZO* and *SCALEPACK* (Otwinoski & Minor, 1997). The resulting scaled and merged structure-factor amplitudes were analyzed for anomalous and dispersive difference signals using the *XtalView* program package (McRee, 1992) and *CNS* (Brunger *et al.*, 1998).

3. Conclusions

Based on preliminary characterizations of native and selenomethionine-labelled CynT2 crystals, a data-collection strategy exploiting anomalous scattering by both Zn and Se atoms was pursued. The requisite multiwavelength diffraction data sets were acquired using synchrotron radiation, further analysis of which revealed the presence of a significant anomalous signal (Table 2). Given the expectation of one Zn atom *versus* five Se atoms per monomer, an anomalous difference Patterson map calculated using the data collected at 1.28 Å ought ideally to yield a clearly interpretable map. Indeed, the Harker sections displayed the expected peaks and the Harker section at $w = 1/4$ (Fig. 2) showed peaks at 6σ, which could only arise in space group $P4_322$ ($P4_122$). The screw axes could not be assigned unambiguously from the diffraction data alone owing to the limited number of

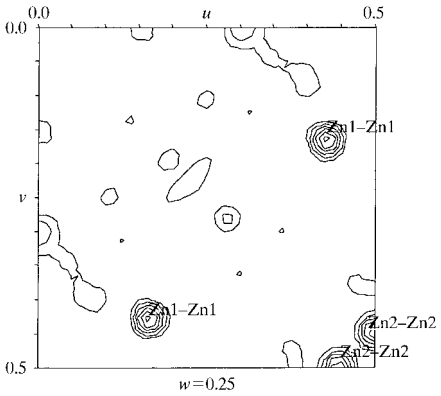

Figure 2
The scaled and merged data from 1.28109 Å, corresponding to the theoretical peak of anomalous scattering from zinc, was imported into the *XtalView* program package (McRee, 1992), which was used to calculate an anomalous difference Patterson map using data in the resolution range 15–4 Å and an outlier cutoff of 100%. The figure shows the Harker section at $w = 0.25$, with contour levels starting at 1.0σ and increasing by steps of 1.0σ. The large peaks seen in this section correspond to the two Zn atoms expected per asymmetric unit for this crystal form (form 2). Zn1–Zn1 and Zn2–Zn2 denote the positions of predicted zinc Harker peaks for a real-space solution deduced by examination of all Harker sections. The peaks on the $w = 0.25$ section confirm the presence of a 4_3 or 4_1 screw axis along c for this primitive tetragonal space group.

systematic absent reflections measured and the noise in data measurement. The Harker section from the Zn anomalous Patterson has eliminated other alternative space groups in the same Laue group, *P*4/*mmm*.

The location of the Zn atoms has been used to find six (of ten expected) selenium sites by difference Fourier methods. This more complete constellation of sites corresponding to anomalous scatterer positions has led directly to good-quality electron-density maps, from which a starting model is being constructed. It is anticipated that the structure determination of the CynT2 carbonic anhydrase will, in addition to constituting one of the first β-CA structures, provide clear direction for further crystallographic studies and biochemical investigations. Work to more fully characterize the functional and regulatory properties of the protein is in progress. These studies have the potential to make a significant contribution to a fuller understanding of β-carbonic anhydrase structure and function.

We are grateful to Dr David S. King for mass spectrometry, Dr Mike Becker and Dr Robert Sweet of Brookhaven National Laboratory, as well as Dr Thomas Earnest of Lawrence Berkeley Laboratory for assistance with data collection and James Holton for help with data processing.

References

Alber, B. E. & Ferry, J. G. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 6909–6913.
Badger, M. R. & Price, G. D. (1994). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 369–392.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burnell, J. N., Gibbs, M. J. & Mason, J. G. (1990). *Plant Physiol.* **92**, 37–40.
Christianson, D. W. & Cox, J. D. (1999). *Annu. Rev. Biochem.* **68**, 33–57.
Christianson, D. W. & Fierke, C. A. (1996). *Acc. Chem. Res.* **29**, 331–339.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Fawcett, T. W., Browse, J. A., Volokita, M. & Bartlett, S. G. (1990). *J. Biol. Chem.* **265**, 5414–5417.
Fujita, N., Mori, H., Yura, T. & Ishihama, A. (1994). *Nucleic Acids Res.* **22**, 1637–1639.
Fukuzawa, H., Suzuki, E., Komukai, Y. & Miyachi, S. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 4437–4441.
Guilloton, M. B., Korte, J. J., Lamblin, A. F., Fuchs, J. A. & Anderson, P. A. (1992). *J. Biol. Chem.* **267**, 3731–3734.
Guilloton, M. B., Lamblin, A. F., Kozliak, E. I., Gerami-Nejad, M., Tu, C., Silverman, D., Anderson, P. A. & Fuchs, J. A. (1993). *J. Bacteriol.* **175**, 1443–1451.
Henry, R. P. (1996). *Annu. Rev. Physiol.* **58**, 523–538.
Hewett-Emmett, D. & Tashian, R. E. (1996). *Mol. Phylogen. Evol.* **5**, 50–77.
Hiltonen, T., Björkbacka, H., Forsman, C., Clarke, A. K. & Samuelsson, A. (1998). *Plant Physiol.* **117**, 1341–1349.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kannan, K. K., Liljas, A., Waara, I., Bergstén, P.-C., Lövgren, B., Strandberg, U., Bengtsson, U., Carlbom, U., Fridborg, K., Järup, L. & Petef, M. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 221–231.
Kimber, M. S. & Pai, E. F. (2000). *EMBO J.* **19**, 1407–1418.
Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G. & Rees, D. C. (1996). *EMBO J.* **15**, 2323–2330.
Kozliak, E. I., Guilloton, M. B., Fuchs, J. A. & Anderson. P. M. (2000). *Bacterial Carbonic Anhydrases*. In *The Carbonic Anhydrases: New Horizons*, edited by W. R. Chegwidden, N. D. Carter & Y. H. Edwards. Basel: Birkhauser. In the press.
Leslie, A. G. W., Brick, P. & Wonacott, A. (1986). *Daresbury Laboratory Information Quarterly for Protein Crystallography*, **18**, 33–39.
Lindskog, S. (1997). *Pharmacol. Ther.* **74**, 1–20.
McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.
Majeau, N. & Coleman, J. R. (1991). *Plant Physiol.* **95**, 264–268.
Mitsuhashi, S., Mizushima, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Moriyama, H., Ueki, T., Miyachi, S. & Tsukihara, T. (2000). *J. Biol. Chem.* **275**, 5521–5526.
Otwinoski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ramakrishnan, V. & Graziano, V. (1997). *Protocol for Growth in Selenomethionine*, http://snowbird.med.utah.edu/~ramak/madms/segrowth.html
Roeske, C. A. & Ogren, W. L. (1990). *Nucleic. Acids Res.* **18**, 3413.
Rowlett, R. S., Chance, M. R., Wirt, M. D., Sidelinger, D. E., Royal, J. R., Woodroffe, M., Wang, Y.-F. A., Saha, R. P. & Lam, M. G. (1994). *Biochemistry*, **33**, 13967–13976.
Sung, Y. & Fuchs, J. A. (1988). *J. Biol. Chem.* **263**, 14769–14775.