crystallization papers

Coexpression, purification, crystallization and preliminary X-ray characterization of glycine decarboxylase (P-protein) of the glycine-cleavage system from *Thermus thermophilus* HB8

Tadashi Nakai,[a]* Noriko Nakagawa,[a] Nobuko Maoka,[a] Ryoji Masui,[a,b] Seiki Kuramitsu[a,b] and Nobuo Kamiya[a]

[a]RIKEN Harima Institute/SPring-8, 1-1-1 Kouto, Mikazuki, Sayo-gun, Hyogo 679-5148, Japan, and [b]Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

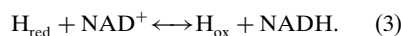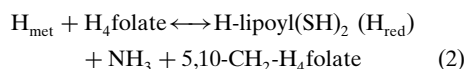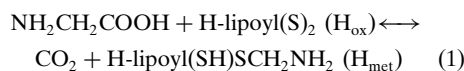Correspondence e-mail: nakaix@spring8.or.jp

Thermus thermophilus (*Tth*) HB8 glycine decarboxylase (P-protein) is an $\alpha_2\beta_2$ tetrameric enzyme with a total molecular mass of 200 kDa. The α- and β-subunits of the *Tth* P-protein have been coexpressed in *Escherichia coli* and purified as a stable complex. Dynamic light-scattering measurements indicated the recombinant protein to be monodisperse and its size to be consistent with an $\alpha_2\beta_2$ tetrameric composition. Crystals of the protein have been grown in polyethylene glycol 3350 using the vapour-diffusion method at 291 K. Synchrotron radiation from BL45XU at SPring-8 was used to measure a complete native data set to 2.4 Å resolution. The crystals belong to the trigonal space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 89.5$, $c = 371.0$ Å. Estimation of the crystal packing ($V_M = 2.15$ Å^3 Da^{-1}) and self-rotation function analysis suggest the presence of one $\alpha_2\beta_2$ tetramer per asymmetric unit, with the molecules related by non-crystallographic twofold symmetry.

1. Introduction

The glycine-cleavage system (GCS) is a multienzyme complex composed of four different components (P-, H-, T- and L-proteins). This complex catalyzes the oxidative cleavage of glycine in a multistep reaction. Firstly, the P-protein catalyzes the decarboxylation of the glycine molecule concomitantly with the transfer of the residual methylamine group to the distal S atom of the lipoyl group of the oxidized H-protein (H$_{ox}$), generating methylamine-loaded H-protein (H$_{met}$) (1). Next, the T-protein catalyzes the transfer of a methylene group from H$_{met}$ to tetrahydrofolate (H$_4$folate), resulting in the release of NH$_3$ and the generation of reduced H-protein (H$_{red}$) (2). Finally, the dihydrolipoyl group of H$_{red}$ is oxidized by the L-protein and H$_{ox}$ is regenerated (3), thereby completing the catalytic cycle (for a review, see Motokawa *et al.*, 1995).

$$NH_2CH_2COOH + \text{H-lipoyl}(S)_2 \ (H_{ox}) \longleftrightarrow$$
$$CO_2 + \text{H-lipoyl}(SH)SCH_2NH_2 \ (H_{met}) \quad (1)$$

$$H_{met} + H_4\text{folate} \longleftrightarrow \text{H-lipoyl}(SH)_2 \ (H_{red})$$
$$+ NH_3 + 5,10\text{-}CH_2\text{-}H_4\text{folate} \quad (2)$$

$$H_{red} + NAD^+ \longleftrightarrow H_{ox} + NADH. \quad (3)$$

The GCS is widely distributed in bacteria and in the mitochondria of plants and mammals and has been studied extensively (for reviews, see Perham, 2000; Douce *et al.*, 2001). In humans, its absence owing to a genetic deficiency leads to a dramatic accumulation of glycine in the blood, resulting in severe neurological diseases, termed non-ketotic hyperglycinaemia (Tada & Hayasaka, 1987). Although H-protein (Pares *et al.*, 1994, 1995; Cohen-Addad *et al.*, 1995; Faure *et al.*, 2000) and L-protein (Mattevi *et al.*, 1991, 1992, 1993; Mande *et al.*, 1996; Li de la Sierra *et al.*, 1997; Toyoda, Kobayashi *et al.*, 1998; Toyoda, Suzuki *et al.*, 1998; Faure *et al.*, 2000) have been studied extensively by X-ray crystallography, the three-dimensional structures of the P- and T-proteins have not yet been reported.

P-protein, also known as glycine decarboxylase [glycine:lipoylprotein oxidoreductase (decarboxylating and acceptor-aminomethylating); EC 1.4.4.2], is a pyridoxal 5′-phosphate (PLP) dependent enzyme. P-proteins from many species have been classified into two types on the basis of their subunit composition. Those from human (Kume *et al.*, 1991), pea (Bourguignon *et al.*, 1988; Turner *et al.*, 1992), *Escherichia coli* (Okamura-Ikeda *et al.*, 1993) and other species are in an α_2 homodimeric form, while those from *Clostridium acidiurici* (Gariboldi & Drake, 1984), *Eubacterium acidaminophilum* (Freudenberg & Andreesen, 1989) and *Bacillus subtilis* (Mizuno *et al.*, 1996) are in an $\alpha_2\beta_2$ tetrameric form. In the genome of *Thermus thermophilus* (*Tth*) HB8 (Yokoyama *et al.*, 2000), we found two open reading frames which are respectively homologous to genes for α- and β-subunits of the latter type P-protein. Therefore, we expected *Tth* P-protein to be in the $\alpha_2\beta_2$ tetrameric form.

The *Tth* P-protein used in this study has a total molecular mass of 200 kDa (α-subunit of 47.1 kDa containing 438 residues and β-subunit of 52.7 kDa containing 474 residues).

Coexpression of the α- and β-subunits led to soluble and thermostable recombinant protein, whereas the α- and β-subunits could be expressed separately only in insoluble form. The α- and β-subunits of the *Tth* P-protein have 31 and 37% amino-acid sequence identities with the N-terminal and C-terminal halves of the human P-protein, respectively, suggesting that the *Tth* and human P-proteins have the same folding topology and similar three-dimensional structures. Therefore, the structure of the *Tth* P-protein might be used to model the human P-protein in investigations focusing on the molecular basis of non-ketotic hyperglycinaemia. Moreover, determination of the *Tth* P-protein structure should greatly facilitate studies aimed at understanding the roles of active-site residues, the reaction mechanisms and the structural architecture of these multienzyme complexes. Here, we report the crystallization and preliminary X-ray analysis of the *Tth* P-protein.

2. Materials and methods

2.1. Protein coexpression

Several methods for the coexpression of multisubunit proteins in bacteria have been reported (Fribourg *et al.*, 2001, and references therein) in which two or three subunits are expressed simultaneously, resulting in a functional protein assembly. However, as application of the known methods to *Tth* P-protein was unsuccessful, we have developed a new method for coexpression. Recombinant DNA techniques were performed essentially as described by Sambrook *et al.* (1989). General construction of the plasmids was carried out in *E. coli* strain DH5α.

The recombinant plasmids (pT7Blue-*gcvP1* and pT7Blue-*gcvP2*, containing the genes for the α- and β-subunits of the *Tth* P-protein, respectively) were supplied by the RIKEN Structural Genomics Initiative (Yokoyama *et al.*, 2000). Each recombinant plasmid was digested with *Nde*I and *Bgl*II and the resulting DNA fragments containing the genes for each subunit were inserted between the *Nde*I and *Bam*HI sites in pET-11a (Novagen), generating expression plasmids pET11a-*gcvP1* and pET11a-*gcvP2*. pET11a-*gcvP1* was cut with *Bgl*II, blunt-ended with T4 DNA polymerase and digested with *Eco*RI, generating a 1.7 kbp DNA fragment encoding the T7 promoter, *gcvP1* and the T7 terminator. The resulting fragment was then inserted between (blunt-ended) *Hin*dIII and *Eco*RI sites in pET11a-*gcvP2*, generating the coexpression plasmid

pET11a-*gcvP1*-*gcvP2*. *E. coli* BLR(DE3)-pLysS (Novagen) was transformed with pET11a-*gcvP1*-*gcvP2* and each colony obtained was tested for coexpression of both subunits. Cultivation and target-gene expression were performed according to the pET system manual (Novagen). Cells were grown at 291 K in Luria–Bertani medium containing 50 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol. These cultures were induced at an optical density (at 600 nm) of 0.4–0.6 using 1.0 mM isopropyl-β-D-thiogalactopyranoside and grown for a further 24 h (at 291 K); the cells were then harvested and stored at 253 K.

2.2. Protein purification

Unless noted otherwise, proteins were purified at room temperature. Frozen cells (22 g from 5 l culture) were thawed, suspended in 20 mM Tris–HCl buffer pH 8.0 containing 5 mM 2-mercaptoethanol and 50 mM NaCl and then disrupted by sonication. The cell lysate was incubated at 343 K for 10 min, kept on ice for 12 min and then ultracentrifuged (200 000g) for 60 min at 277 K. Ammonium sulfate was added to the resulting supernatant to a final concentration of 1.5 M, after which the solution was applied to a Resource ISO column (6 ml; Amersham Biosciences) equilibrated with 50 mM sodium phosphate buffer pH 7.0 containing 1.5 M ammonium sulfate. Protein was eluted with a linear gradient of 1.5–0.0 M ammonium sulfate in 50 mM sodium phosphate. Fractions containing the target protein were collected, desalted on a HiPrep 26/10 desalting column (Amersham Biosciences) and equilibrated with 20 mM Tris–HCl buffer pH 8.0. The sample was applied to a Resource Q column (6 ml; Amersham Biosciences) equilibrated in the same buffer; protein was then eluted with a linear gradient of 0–500 mM NaCl. Fractions containing the target protein were desalted and applied to a CHT2-I hydroxyapatite column (2 ml; Bio-Rad) that was previously equilibrated with 10 mM sodium phosphate buffer pH 7.0. Protein was eluted with a linear gradient of 10–250 mM sodium phosphate. Fractions containing the target protein were then loaded onto a HiLoad 16/60 Superdex 200 prep-grade column (Amersham Biosciences) equilibrated with 20 mM Tris–HCl buffer pH 8.0 and 150 mM NaCl and eluted with the same buffer. After addition of 1 mM dithiothreitol, the peak fractions were concentrated and stored at 277 K. At each step, the fractions were analyzed by SDS–PAGE with a 12%(w/v) acrylamide gel.

2.3. Dynamic light-scattering measurements

The dynamic light-scattering characteristics of the protein solution were analyzed using a DynaPro-801 molecular-sizing detector (Protein Solutions). 1 mg ml^{-1} protein samples were prepared in solutions of 20 mM Tris–HCl buffer pH 8.0 containing 150 mM NaCl, 1 mM dithiothreitol and 100 µM pyridoxal phosphate (PLP). Prior to the experiment, the protein samples were filtered through Ultrafree-MC 0.22 µm filters (Millipore). All data were analyzed using *DYNAMICS* software (Protein Solutions).

2.4. Crystallization

Initial screening for *Tth* P-protein crystallization conditions was performed using the hanging-drop vapour-diffusion method at 291 K and Crystal Screens 1 and 2 and the PEG/Ion Screen from Hampton Research. The initial protein concentration was 10 mg ml^{-1} in 20 mM Tris–HCl buffer pH 8.0 containing 150 mM NaCl, 1 mM dithiothreitol and 100 µM PLP. Drops were prepared by mixing 2 µl of the protein solution with 2 µl of the reservoir solution and were then equilibrated against 400 µl of reservoir solution.

2.5. Data collection

Prior to data collection, a crystal of *Tth* P-protein was soaked for a few seconds in a cryoprotectant solution containing 35%(w/v) polyethylene glycol 3350 and 300 mM potassium thiocyanate (KSCN) and then transferred to a nylon CryoLoop (Hampton Research) and flash-cooled in a nitrogen-gas stream at 100 K. X-ray diffraction data were collected at 100 K on a Rigaku R-AXIS V imaging-plate detector using synchrotron radiation with a wavelength of 0.90 Å from BL45XU at SPring-8 (Yamamoto *et al.*, 1998, 2001). The data processing was completed using *HKL*2000 (Otwinowski & Minor, 1997).

3. Results and discussion

Judging from SDS–PAGE analysis (Fig. 1), *Tth* P-protein was purified to homogeneity and the α- and β-subunits were produced in a 1:1 molar ratio. Their molecular masses were also estimated by SDS–PAGE and these estimates were consistent with the masses predicted from the DNA sequence. The final amount of purified protein was 14 mg and the yield was 0.65 mg of protein per gram of cells. Dynamic light-scattering results indicated the protein to be mono-
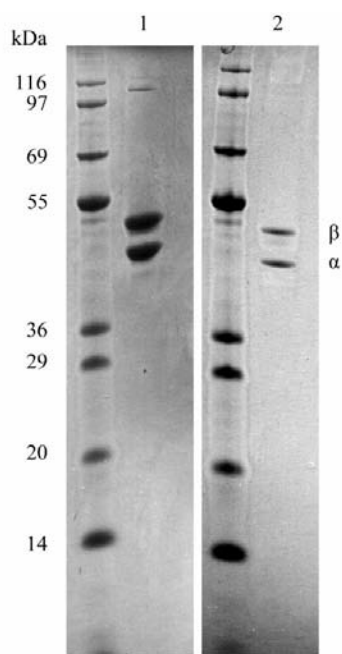
Figure 1
SDS–PAGE analysis of purified protein, and washed and dissolved crystals. Lane 1, protein solution used for crystallization; lane 2, dissolved crystals.

disperse and to have a hydrodynamic radius of 5.5 nm and an apparent molecular mass of 180 kDa, suggesting that *Tth* P-protein exists as an $\alpha_2\beta_2$ tetramer in solution. The purified protein exhibited an absorption maximum at 428 nm, which arose from the PLP molecule attached to the enzyme protein. When an excess amount of glycine, the substrate of P-protein, was added to the protein solution, the absorption maximum shifted toward shorter wavelengths (data not shown). These properties are essentially the same as those of mitochondrial P-protein as described by Kikuchi & Hiraga (1982), confirming that the purified protein has the native enzymatic properties of P-protein.

In the search for crystallization conditions for *Tth* P-protein, several crystal forms were obtained; the most promising set of crystallization conditions, corresponding to the
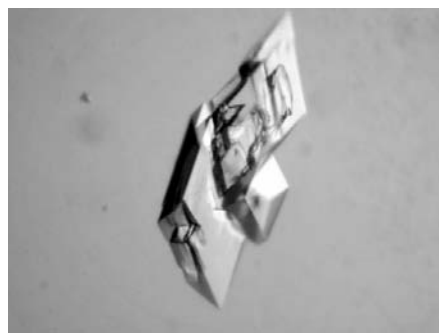

Figure 2
Crystals of P-protein from *T. thermophilus* HB8. The largest single crystal in this cluster is about 0.1 × 0.1 × 0.3 mm.

PEG/Ion Screen Kit solution No. 14, was optimized. A droplet containing 2 µl protein solution was mixed with an equal volume of reservoir solution [30%(*w/v*) polyethylene glycol 3350 and 300 m*M* KSCN] and equilibrated to give crystals of *Tth* P-protein (Fig. 2). The crystals usually appeared within a few months and continued to grow for several more weeks. Although the protein had a tendency to grow clusters of crystals, single crystals were easily obtained by picking them out with a pointed probe (Crystal Probe, Hampton Research). SDS–PAGE analysis of thoroughly washed crystals (Fig. 1) confirmed the presence of both subunits in the crystallized material.

The crystal diffracted to approximately 2.2 Å resolution with synchrotron radiation from BL45XU at SPring-8 (Fig. 3). A complete native data set with 68 490 unique reflections was collected from a single crystal, giving a data set with a completeness of 99.3% in the resolution range 50.0–2.4 Å and an R_{merge} of 8.5% (Table 1). These data indicate suitably good quality of the crystals for X-ray structural analysis. The crystals showed no significant decay upon X-ray exposure. The crystals belong to the trigonal space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 89.5$, $c = 371.0$ Å and with a unit-cell volume of 2.58×10^6 Å³.

Assuming there to be one $\alpha_2\beta_2$ tetramer per asymmetric unit, the V_M value was calculated to be 2.15 Å³ Da^{-1}, indicating a solvent content of approximately 42% in the unit cell. These values are within the range typically found for protein crystals (Matthews, 1968). This result is also consistent with the non-crystallographic peaks observed in the self-rotation function maps calculated using the *CCP*4 program *POLARRFN* (Collaborative Computational Project, Number 4, 1994). The maps of polar angle $\kappa = 180°$ sections were calculated using different resolution ranges and integration radii (Fig. 4). The highest peaks at polar angles $\omega = 90°$ and $\varphi = 0°$ represented the crystallographic twofold

Table 1
Crystal data and intensity statistics.

Values in parentheses correspond to the reflections observed in the highest resolution shell (2.49–2.40 Å).

Resolution range (Å)	50.0–2.4
No. of measured reflections	419362
No. of unique reflections	68490 (6759)
Redundancy	6.1 (4.9)
Completeness (%)	99.3 (99.0)
R_{merge}† (%)	8.5 (24.9)
Average $I/\sigma(I)$	9.7 (2.4)

† $R_{merge} = \sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i I_{hkl,i}$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity for multiple measurements.
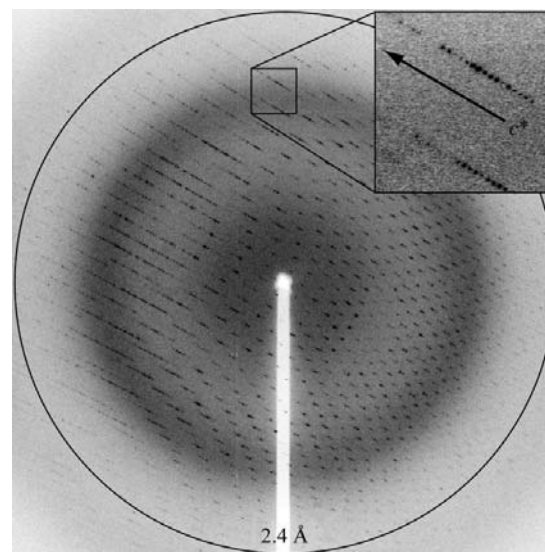

Figure 3
Diffraction pattern with 1.0° oscillation from a *Tth* P-protein crystal. A higher magnification view is shown at the top right of the image. The crystal-to-detector distance, width of the imaging plate and synchrotron-radiation wavelength were 500 mm, 400 mm and 0.9 Å, respectively.
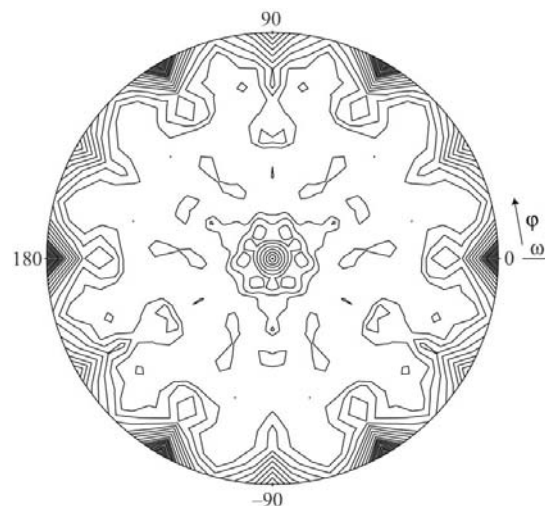

Figure 4
$\kappa = 180°$ section of the self-rotation function calculated from the native data set of a *Tth* P-protein crystal. The resolution of the data used was 15–4 Å, with a radius of integration of 30 Å. The highest non-crystallographic peak at (ω, φ) = (90, 30°) corresponds to the direction of the non-crystallographic twofold axis of the dimer and its height is 46% of the origin.

symmetry axis. The second highest peaks at $\omega = 90°$ and $\varphi = 30°$ consistently appeared in all maps calculated with the different parameters, suggesting that the two $\alpha\beta$ dimers are related by a non-crystallographic twofold axis and that this symmetry axis corresponds to their orientation within the $\alpha_2\beta_2$ tetramer in the asymmetric unit.

We are now in the process of searching for heavy-atom derivatives as well as for crystallization conditions for the selenomethionine-substituted protein, in order to determine the three-dimensional structure of *Tth* P-protein.

References

Bourguignon, J., Neuburger, M. & Douce, R. (1988). *Biochem. J.* **255**, 169–178.

Cohen-Addad, C., Pares, S., Sieker, L., Neuburger, M. & Douce, R. (1995). *Nature Struct. Biol.* **2**, 63–68.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Douce, R., Bourguignon, J., Neuburger, M. & Rebeille, F. (2001). *Trends Plant Sci.* **6**, 167–176.

Faure, M., Bourguignon, J., Neuburger, M., Macherel, D., Sieker, L., Ober, R., Kahn, R., Cohen-Addad, C. & Douce, R. (2000). *Eur. J. Biochem.* **267**, 2890–2898.

Freudenberg, W. & Andreesen, J. R. (1989). *J. Bacteriol.* **171**, 2209–2215.

Fribourg, S., Romier, C., Werten, S., Gangloff, Y. G., Poterszman, A. & Moras, D. (2001). *J. Mol. Biol.* **306**, 363–373.

Gariboldi, R. T. & Drake, H. L. (1984). *J. Biol. Chem.* **259**, 6085–6089.

Kikuchi, G. & Hiraga, K. (1982). *Mol. Cell Biochem.* **45**, 137–149.

Kume, A., Koyata, H., Sakakibara, T., Ishiguro, Y., Kure, S. & Hiraga, K. (1991). *J. Biol. Chem.* **266**, 3323–3329.

Li de la Sierra, I., Pernot, L., Prange, T., Saludjian, P., Schiltz, M., Fourme, R. & Padron, G. (1997). *J. Mol. Biol.* **269**, 129–141.

Mande, S. S., Sarfaty, S., Allen, M. D., Perham, R. N. & Hol, W. G. J. (1996). *Structure*, **4**, 277–286.

Mattevi, A., Obmolova, G., Kalk, K. H., van Berkel, W. J. & Hol, W. G. (1993). *J. Mol. Biol.* **230**, 1200–1215.

Mattevi, A., Obmolova, G., Sokatch, J. R., Betzel, C. & Hol, W. G. (1992). *Proteins*, **13**, 336–351.

Mattevi, A., Schierbeck, A. J. & Hol, W. G. (1991). *J. Mol. Biol.* **220**, 975–994.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mizuno, M., Masuda, S., Takemaru, K., Hosono, S., Sato, I., Takeuchi, M. & Kobayashi, Y. (1996). *Microbiology*, **142**, 3103–3111.

Motokawa, Y., Fujiwara, K. & Okamura-Ikeda, K. (1995). *Bithiols in Health and Disease*, edited by L. Packer & E. Cadenas, pp. 389–407. New York: Marcel Dekker.

Okamura-Ikeda, K., Ohmura, Y., Fujiwara, K. & Motokawa, Y. (1993). *Eur. J. Biochem.* **216**, 539–548.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. & Douce, R. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4850–4853.

Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. & Douce, R. (1995). *Acta Cryst.* D**51**, 1041–1051.

Perham, R. N. (2000). *Annu. Rev. Biochem.* **69**, 961–1004.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Tada, K. & Hayasaka, K. (1987). *Eur. J. Pediatr.* **146**, 221–227.

Toyoda, T., Kobayashi, R., Sekiguchi, T., Koike, K., Koike, M. & Takenaka, A. (1998). *Acta Cryst.* D**54**, 982–985.

Toyoda, T., Suzuki, K., Sekiguchi, T., Reed, L. J. & Takenaka, A. (1998). *J. Biochem.* **123**, 668–674.

Turner, S. R., Ireland, R. & Rawsthorne, S. (1992). *J. Biol. Chem.* **267**, 5355–5360.

Yamamoto, M., Kumasaka, T., Fujisawa, T. & Ueki, T. (1998). *J. Synchrotron Rad.* **5**, 222–225.

Yamamoto, M., Kumasaka, T., Yamazaki, H., Sasaki, K., Yokozawa, Y. & Ishikawa, T. (2001). *Nucl. Instrum. Methods Phys. Res. A*, **467**, 1160–1162.

Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, 943–945.