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

# The structure of an archaeal ribose-5-phosphate isomerase from *Methanocaldococcus jannaschii* (MJ1603)

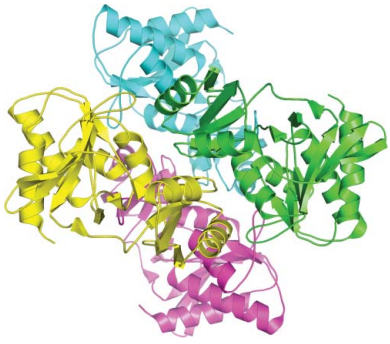
**Richard W. Strange,[a]* Svetlana V. Antonyuk,[a] Mark J. Ellis,[b] Yoshitaka Bessho,[c,d] Seiki Kuramitsu,[d,e] Shigeyuki Yokoyama[c,d,f] and S. Samar Hasnain[a]***

[a]Molecular Biophysics Group, School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, England, [b]STFC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England, [c]Systems and Structural Biology Center, Yokohama Institute, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama 230-0045, Japan, [d]RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, [e]Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka 560-0043, Japan, and [f]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: r.strange@liverpool.ac.uk, s.s.hasnain@liverpool.ac.uk

Received 25 June 2009
Accepted 27 October 2009

**PDB Reference:** ribose-5-phosphate isomerase, 3ixq, r3ixqsf.





Ribose-5-phosphate isomerase is a ubiquitous intracellular enzyme of bacterial, plant and animal origin that is involved in the pentose phosphate cycle, an essential component of cellular carbohydrate metabolism. Specifically, the enzyme catalyses the reversible conversion of ribose 5-phosphate to ribulose 5-phosphate. The structure of ribose-5-phosphate isomerase from *Methanocaldococcus jannaschii* has been solved in space group $P2_1$ to 1.78 Å resolution using molecular replacement with one homotetramer in the asymmetric unit and refined to an *R* factor of 14.8%. The active site in each subunit was occupied by two molecules of propylene glycol in different orientations, one of which corresponds to the location of the phosphate moiety and the other to the location of the furanose ring of the inhibitor.

## 1. Introduction

Ribose-5-phosphate isomerases (Rpis) are essential enzymes in the pentose phosphate pathway and in the Calvin cycle of plants. Loss of Rpi activity in humans has been linked to the disease leukoencephalopathy (Huck *et al.*, 2004). Two evolutionary distinct forms of the enzyme, RpiA and RpiB, with different amino-acid sequences and molecular weights have been identified. Both forms catalyse the reversible conversion of ribose 5-phosphate to ribulose 5-phosphate (Fig. 1). RpiA is found in the bacterial, plant and animal kingdoms, whereas RpiB is less widespread and is found in bacterial sources; genes encoding both enzymes are found in *Escherichia coli* (David & Wiesmeyer, 1970). Crystal structures of RpiA have been reported for *E. coli* (Rangarajan *et al.*, 2002; Zhang *et al.*, 2003), *Plasmodium falciparum* (Holmes *et al.*, 2006) and *Pyrococcus horikoshii* (Ishikawa *et al.*, 2002). Proteins from the archaea *Methanocaldococcus jannaschii* have been targeted for high-throughput structure determination as part of the RIKEN Structural Genomics Initiative (Sugahara *et al.*, 2008). Here, we report the crystal structure of RpiA from *M. jannaschii* at 1.78 Å resolution.

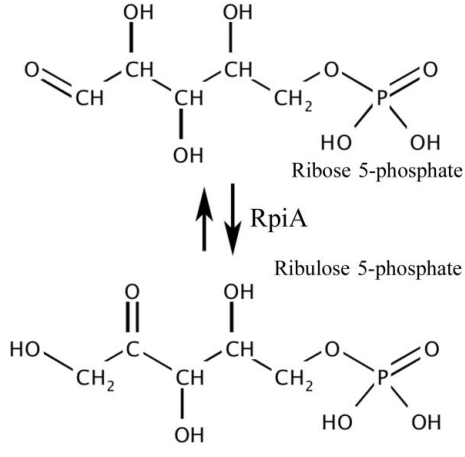



**Figure 1**
The reaction catalysed by RpiA. The open configuration of ribose 5-phosphate is required for isomerization and is reversibly converted to ribulose 5-phosphate. It is possible that the enzyme also operates on the closed form of the sugar, opening the ring prior to isomerization (Jung *et al.*, 2000).

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The gene encoding MJ1603 (gi:15669799) was amplified *via* PCR using *M. jannaschii* DSM 2661 genomic DNA and was cloned into the pET-21a expression vector (Merck Novagen, Darmstadt, Germany). The expression vector was introduced into *E. coli* Rosetta (DE3) strain (Merck Novagen, Darmstadt, Germany) and the recombinant strain was cultured in 2.5 l LB medium containing 30 µg ml$^{-1}$ chloramphenicol and 50 µg ml$^{-1}$ ampicillin. The harvested cells (8.5 g) were lysed by sonication in 15 ml 20 m*M* Tris–HCl buffer pH 8.0 containing 500 m*M* NaCl, 5 m*M* $\beta$-mercaptoethanol and 1 m*M* phenylmethylsulfonyl fluoride on ice. The cell lysate was heat-treated at 363 K for 12 min and centrifuged at 15 000*g* for 30 min at 277 K. The supernatant was desalted by fractionation on a HiPrep 26/10 column (GE Healthcare Biosciences). The sample was applied onto a Toyopearl SuperQ-650M column (Tosoh, Tokyo) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 and eluted with a linear NaCl gradient (0–0.4 *M*). The target sample, which eluted in the 0.34 *M* NaCl fraction, was desalted (HiPrep 26/10) and then applied onto a Resource Q column (GE Healthcare Biosciences) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 and eluted with a linear gradient of 0–0.4 *M* NaCl. The fractions that eluted in 0.34 *M* NaCl were further purified using a hydroxyapatite CHT20-I column (Bio-Rad Laboratories) with a linear gradient of 0.01–0.5 *M* potassium phosphate buffer pH 7.0. The target sample, which eluted in the 0.3 *M* potassium phosphate fraction, was concentrated and applied onto a HiLoad 26/60 Superdex 200 pg column (GE Healthcare Biosciences) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 200 m*M* NaCl. The protein sample was analyzed by SDS–PAGE and was confirmed by N-terminal amino-acid sequencing. After concentration to 8 mg ml$^{-1}$ by ultrafiltration, the protein yield was 44.5 mg from 8.5 g of cells.

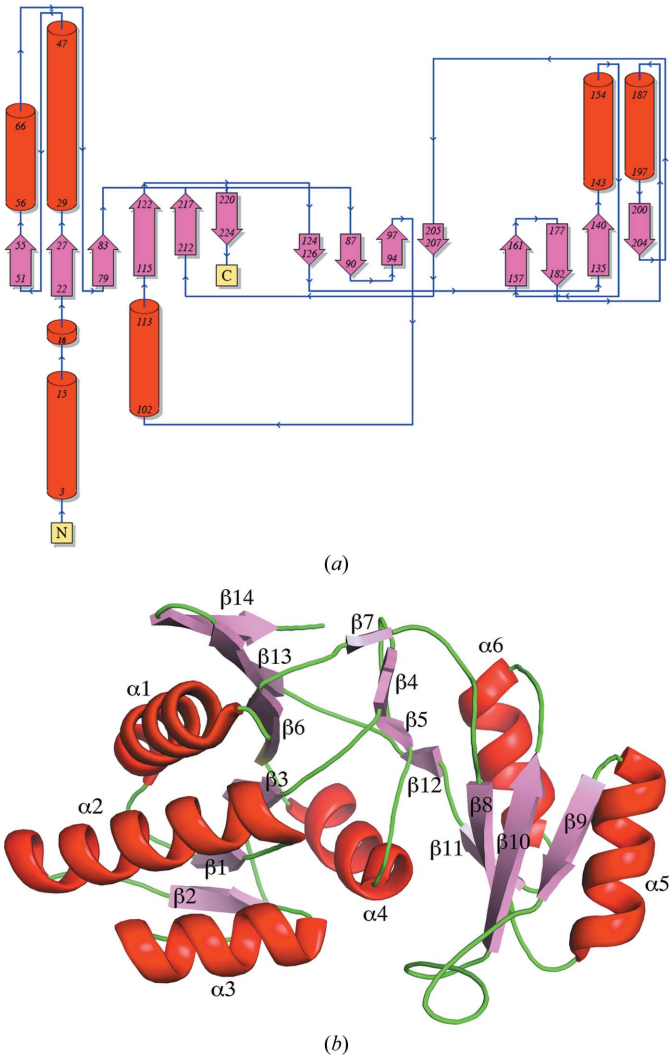



**Figure 2**
The topology and fold of a single subunit of *M. jannaschii* RpiA. (*a*) The topology diagram shows $\alpha$-helices represented as red cylinders, $\beta$-strands as pink arrows and loops as blue lines. There are two domains and an interface region, which is made up of four $\beta$-strands. The larger domain (domain I) contains four $\alpha$-helices, one $3_{10}$-helix and six $\beta$-strands and includes both the C-terminal and N-terminal ends, while the smaller domain (domain II) contains two $\alpha$-helices and four $\beta$-strands. The corresponding three-dimensional structure is shown in (*b*), with the secondary structure coloured according to topology and labelled according to the sequence. The topology diagram was obtained using the *PDBSUM* server (Laskowski, 2009). The molecular-graphics figures were all obtained using *PyMOL* (DeLano, 2008).

### 2.2. Crystallization

Crystallization was performed by the microbatch-under-oil method at 291 K. 0.5 µl crystallization reagent, consisting of 0.1 *M* acetate pH 4.5 containing 40%(*v*/*v*) 1,2-propanediol and 0.05 *M* calcium acetate (Emerald Cryo I condition No. 21, Emerald BioSystems), was mixed with 0.5 µl 8.1 mg ml$^{-1}$ protein solution. The mixture was then covered with 15 µl silicone and paraffin oil. Crystals suitable for X-ray data collection appeared within one week and reached final dimensions of 0.1 × 0.07 × 0.07 mm. The crystals were flash-cooled in a nitrogen-gas stream at 100 K without additional cryoprotectant.

### 2.3. Data collection and processing

Experiments were performed at the Daresbury Synchrotron Radiation Source (SRS) using the combined crystallography/X-ray absorption beamline 10.1, employing a Si(111) sagittally focused monochromator tuned to a wavelength of 1.074 Å. Diffraction data were recorded at 100 K from a single crystal. Images were recorded using a MAR Mosaic 225 CCD detector and were processed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The crystal was found to belong to space group $P2_1$, with unit-cell parameters $a$ = 54.86, $b$ = 100.27, $c$ = 80.46 Å, $\beta$ = 92.54°. Data-collection and processing parameters are shown in Table 1.

### 2.4. Structure solution and refinement

Structure solution was accomplished by molecular replacement with *MOLREP* (Vagin & Teplyakov, 1997) using as a search model the RpiA tetramer from *P. horikoshii* (PDB code 1lk5; Ishikawa *et al.*, 2002) obtained from the Protein Data Bank (Abola *et al.*, 1987). The search model has 52% identity and gave a molecular-replacement score and $R$ factor of 0.43 and 54.6%, respectively (the next highest score and $R$ factor were 0.26 and 60.3%, respectively). Following initial rounds of rigid-body and restrained refinement using *REFMAC* (Murshudov *et al.*, 1997), the structure was rebuilt with the correct amino-acid sequence using *ARP*/*wARP* (Perrakis *et al.*, 1999) and *REFMAC* in combination with *Coot* (Emsley & Cowtan, 2004). The $R$ factor and $R_{\rm free}$ of the final model to 1.78 Å resolution were 14.8% and 18.8%, respectively. The stereochemistry of the final model was checked using *MolProbity* (Davis *et al.*, 2007). The structure was deposited in the PDB under accession code 3ixq. The protein interfaces, surfaces and assemblies service *PISA* (Krissinel & Henrick, 2007) at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) was used to explore the structural and chemical properties of the molecule.

**Table 1**
Data-collection and refinement parameters.

Values in parentheses are for the highest resolution shell (1.83–1.78 Å).

| | |
|---|---|
| Space group | $P2_1$ |
| Unit-cell parameters (Å, °) | $a$ = 54.86, $b$ = 100.27, $c$ = 80.46, $\beta$ = 92.54 |
| Resolution (Å) | 30–1.78 |
| Unique reflections | 94575 |
| Completeness (%) | 98.5 (98.3) |
| Redundancy | 8.0 (5.1) |
| $R_{merge}$† (%) | 10.2 (39.2) |
| $I/\sigma(I)$ | 23 (4) |
| $R$ factor‡ (%) | 14.8 |
| $R_{free}$‡ (%) | 18.8 |
| $B$ factors (Å$^2$) | |
| Wilson plot | 17.0 |
| Protein | 13.2 |
| Water | 33.0 |
| Propylene glycol | 30.9 |
| $Cl^-$ | 19.7 |
| Acetate | 67.2 |
| R.m.s. deviations | |
| Bond distances (Å) | 0.015 |
| Bond angles (°) | 1.47 |
| ESU§ (Å) | 0.071 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple symmetry-related observations. ‡ $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. $R_{free}$ is the same but calculated for a test set not used in structural refinement. § Estimated standard uncertainty based on maximum likelihood as implemented in *REFMAC*.

## 3. Results and discussion

Size-exclusion chromatography shows that a tetramer is the most probable biological unit in the solution state, which is in agreement with the crystallographic data. Each subunit of the tetramer was modelled with only the first residue at the N-terminus missing from the electron-density map. The topology and fold of one subunit is shown in Fig. 2. It comprises six $\alpha$-helices, one $3_{10}$-helix and 14 $\beta$-strands divided into two domains, with domain I containing helices $\alpha$1–$\alpha$4 and $\beta$-strands 1–3, 6, 13 and 14, and domain II containing helices $\alpha$5 and $\alpha$6 and $\beta$-strands 8–11. The remaining four $\beta$-strands (4, 5, 7 and 12) form the interface between the two domains. The tetramer buries approximately 27% of the solvent-accessible surface area of each subunit. The residues involved in intersubunit interactions (Fig. 3), using subunit *A* as an example, include Arg161, Leu162, Asp164, Arg165, Lys166, Arg167 and Asp173, all of which are on the loop linking $\beta$9 and $\beta$10 in domain II, and Glu58 and Met64 on helix $\alpha$3 and Gly100 on the loop linking $\beta$5 and helix $\alpha$3, all of which belong to domain I. These residues make up to 13 hydrogen bonds or salt bridges with their symmetric counterparts in subunit *B*. The interface between subunits *A* and *D* is formed with 18 potential hydrogen bonds and salt bridges and involves the symmetrically related residues Ser142, Tyr144, Arg145 and Arg149, all of which are on helix $\alpha$5 of domain II, and Leu73, Asp74, Tyr76 and Asp77, all of which are on the loop linking $\alpha$3 and $\beta$3, and Gln105 on $\alpha$4 of domain I. The interaction between subunits *A* and *C* is less extensive, burying only 4% of the surface area of each, comprising three hydrogen bonds and one salt bridge and involving the symmetrically equivalent residues Lys60 on helix $\alpha$3 and Leu70, Asp74 and Glu75, all of which are in domain I. The biologically active assembly of RpiA in *E. coli* is a dimer (Essenberg & Cooper, 1975; Rangarajan *et al.*, 2002). The corresponding substructure of the *M. jannaschii* tetramer is identified as the dimer *AD*, giving an r.m.s.d. of 1.04 Å for 296 $C^\alpha$ atoms.

The active site is located at the interface between the two domains and the residues involved are known to be highly conserved among RpiA proteins from different sources. In *M. jannaschii*, each of the four subunits of the carbohydrate-binding pocket was found to contain two molecules of propylene glycol, which was used in the crystallization medium. The active site comprises residues from helix $\alpha$2 and $\alpha$4 of domain I and residues from the domain-interface $\beta$-strands. The propylene glycol molecules are bound at the same location as that occupied by arabinose 5-phosphate, an inhibitor ligand found to bind at the active site in *E. coli* RpiA (PDB code

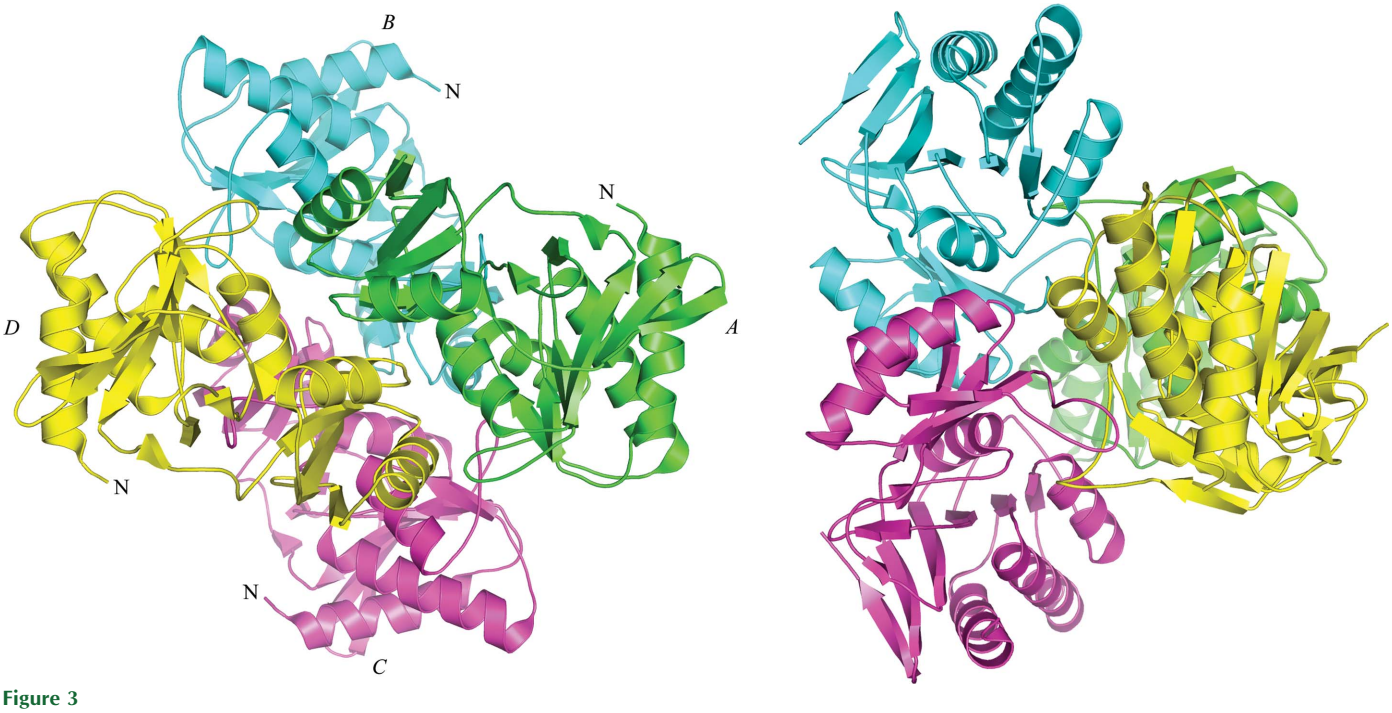



**Figure 3**
Cartoon representation of the RpiA tetramer. The major contacts in the tetramer, as described in the text, are between subunits *AD* and *BC*, which form dimers equivalent to the biologically functional dimer of RpiA in *E. coli*, and between subunits *AB* and *CD*. These interactions within the tetramer are clearly shown here by two orientations of the molecule related by a 90° rotation about the vertical axis.

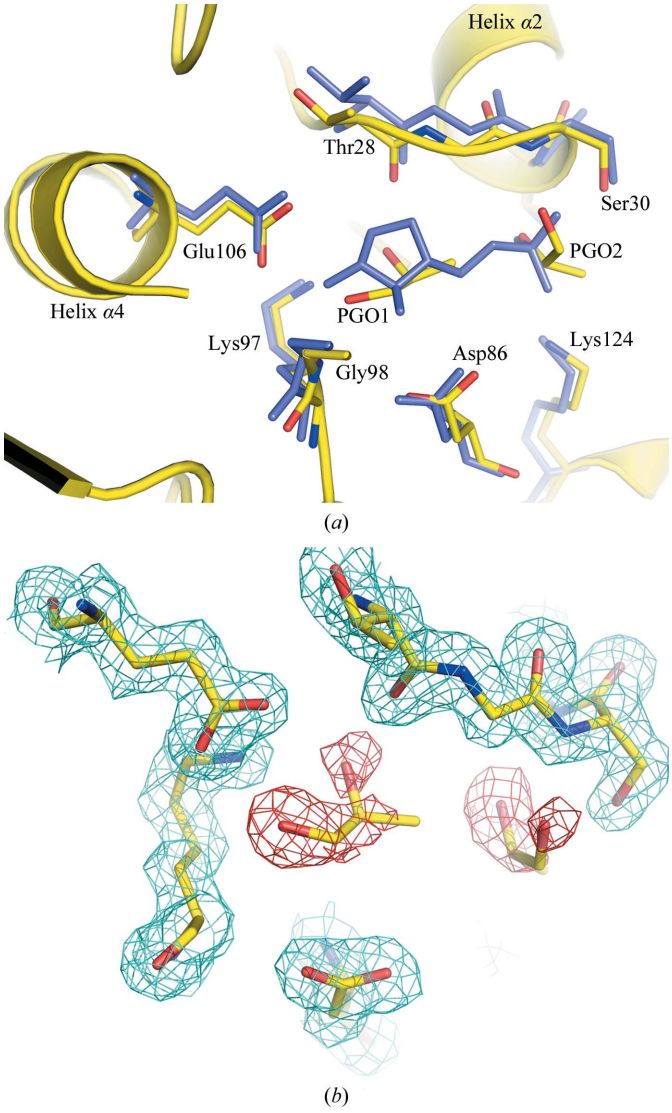



**Figure 4**
The carbohydrate-binding site of RpiA. (*a*) Subunit *D* of the *M. jannaschii* tetramer is shown in yellow together with the equivalent active site of *E. coli* RpiA (PDB code 1o8b; Zhang *et al.*, 2003), which is shown in light blue. The propylene glycol molecules found at the active site of subunit *D* are also shown, denoted by PGO. Their locations and orientations are compared with the position of the arabinose 5-phosphate ligand present in the *E. coli* structure. The amino acids are numbered according to the *M. jannaschii* sequence. (*b*) The subunit *D* active site, presented in a slightly different orientation, with electron density drawn at the 1$\sigma$ level (blue) together with a 3$\sigma$ level OMIT map (red) showing the fit of the putative PGO molecules.

1o8b; Zhang *et al.*, 2003). The propylene glycol molecule takes up two orientations at the active site (Fig. 4). In the first of these it occupies a position corresponding to that of the phosphate moiety in the inhibitor ligand, in which it is has the potential to form hydrogen bonds to the backbone carbonyl group of Thr28 and the Lys97 and Glu106 side chains. The second binding site corresponds to the location of the furanose ring of the inhibitor, in which the propylene glycol hydroxyl groups can interact with both the Ser30 and Thr31 side-chain O atoms.

We thank Mr Yoshihiro Agari and Drs Akeo Shinkai and Akio Ebihara for their help in sample preparation. This work was supported in part by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was supported by the Synchrotron Radiation Department at the Science and Technology Facilities Council, Daresbury Laboratory UK and X-ray data were collected on beamline 10.1 at the Synchrotron Radiation Source, which was supported by Biotechnology and Biological Sciences Research Council Grant BB/E001971 (to SSH and RWS).

## References

Abola, A., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, edited by F. H. Allen, G. Bergerhoff & R. Sievers, pp. 107–132. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography.
David, J. & Wiesmeyer, H. (1970). *Biochim. Biophys. Acta*, **208**, 58–67.
Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.
DeLano, W. L. (2008). *PyMOL Molecular Viewer.* DeLano Scientific, Palo Alto, California, USA. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Essenberg, M. K. & Cooper, R. A. (1975). *Eur. J. Biochem.* **55**, 323–332.
Holmes, M. A. *et al.* (2006). *Acta Cryst.* F**62**, 427–431.
Huck, J. H. J., Verhoeven, N. M., Stuys, E. A., Salomons, G. S., Jakobs, C. & van der Knapp, M. S. (2004). *Am. J. Hum. Genet.* **74**, 745–751.
Ishikawa, K., Matsui, I., Payan, F., Cambillau, C., Ishida, H., Kawarabayasi, Y., Kikuchi, H. & Roussel, A. (2002). *Structure*, **10**, 877–886.
Jung, C. H., Hartman, F. C., Lu, T. Y. & Larimer, F. W. (2000). *Arch. Biochem. Biophys.* **373**, 409–417.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–779.
Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rangarajan, E. S., Sivaraman, J., Matte, A. & Cygler, M. (2002). *Proteins*, **14**, 737–740.
Sugahara, M. *et al.* (2008). *J. Struct. Funct. Genomics*, **9**, 21–28.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Zhang, R.-G., Andersson, C. E., Savchenko, A., Skarina, T., Evdokimova, E., Beasley, S., Arrowsmith, C. H., Edwards, A. M., Joachimiak, A. & Mowbray, S. L. (2003). *Structure*, **11**, 31–42.