RIKEN–UK structural genomics

Richard W. Strange,[a*]
Svetlana V. Antonyuk,[a] **Mark J. Ellis,**[b] **Yoshitaka Bessho,**[c,d] **Seiki Kuramitsu,**[d,e] **Akeo Shinkai,**[d] **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
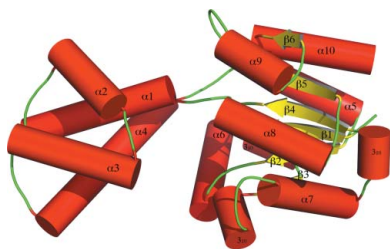
Structure of a putative β-phosphoglucomutase (TM1254) from *Thermotoga maritima*

The structure of TM1254, a putative β-phosphoglucomutase from *T. maritima*, was determined to 1.74 Å resolution in a high-throughput structural genomics programme. Diffraction data were obtained from crystals belonging to space group $P22_12_1$, with unit-cell parameters $a = 48.16$, $b = 66.70$, $c = 83.80$ Å, and were refined to an R factor of 19.2%. The asymmetric unit contained one protein molecule which is comprised of two domains. Structural homologues were found from protein databases that confirmed a strong resemblance between TM1254 and members of the haloacid dehalogenase (HAD) hydrolase family.

1. Introduction

As part of the RIKEN Structural Genomics Initiative, proteins from *Thermotoga maritima* MSB8 were targeted for high-throughput structure determination (Sugahara *et al.*, 2008). Using the Pfam database (Bateman *et al.*, 2004), the protein TM1254 was predicted to function as a β-phosphoglucomutase involved in energy metabolism, catalysing the interconversion of β-D-glucose 1-phosphate (G1P) and β-D-glucose 6-phosphate (G6P). Here, we report the crystal structure of the apo form of TM1254 at 1.74 Å resolution.

2. Materials and methods

2.1. Cloning, expression and purification

The gene encoding the TM1254 protein (gi:15644010) was amplified *via* PCR using *T. maritima* MSB8 genomic DNA and was cloned into the pET-21a expression vector (Merck Novagen, Darmstadt, Germany). The expression vector was introduced into *Escherichia coli* Rosetta (DE3) strain (Merck Novagen, Darmstadt, Germany) and the recombinant strain was cultured in 3 l minimal medium containing 25 µg ml^{-1} selenomethionine, 30 µg ml^{-1} chloramphenicol and 50 µg ml^{-1} ampicillin. The harvested cells (8.3 g) were lysed by sonication in 20 mM Tris–HCl buffer pH 8.0 containing 50 mM NaCl on ice. The cell lysate was heat-treated at 363 K for 10 min and centrifuged at 200 000g for 60 min. The supernatant was applied onto a Toyopearl SuperQ-650M column (Tosoh, Tokyo) equilibrated with 20 mM Tris–HCl buffer pH 8.0 and eluted with a linear gradient of NaCl. The target sample, which eluted in the 0.19 M NaCl fraction, was then applied onto a Resource Q column (GE Healthcare Biosciences) equilibrated with 20 mM Tris–HCl buffer pH 8.0 and eluted with a linear gradient of NaCl. The target sample, which eluted in 0.06 M NaCl, was collected and applied onto a HiLoad 16/60 Superdex 75 pg column (GE Healthcare Biosciences) equilibrated with 20 mM Tris–HCl buffer pH 8.0 containing 150 mM NaCl. The protein sample was analyzed by SDS–PAGE and was confirmed by N-terminal amino-acid sequencing. After concentration to 8.6 mg ml^{-1} by ultrafiltration, the protein yield was 64.4 mg from 8.3 g of cells.

2.2. Crystallization

Crystallization was performed by the sitting-drop vapour-diffusion method at 293 K. Each drop, consisting of 1.0 µl 8.6 mg ml^{-1} protein

solution and 1.0 µl reservoir solution (0.1 *M* trisodium citrate pH 5.0 containing 1.8 *M* ammonium sulfate), was allowed to equilibrate against 100 µl reservoir solution. Crystals suitable for X-ray data collection appeared within 3 d and reached final dimensions of $0.3 \times 0.06 \times 0.02$ mm. The crystals (Fig. 1) were flash-cooled in a nitrogen-gas stream at 100 K using the mother liquor supplemented with 20%(*v*/*v*) glycerol as a cryoprotectant for transport to the synchrotron.

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. A XANES spectrum was recorded at the Se *K* edge using a low-profile monolithic four-element Ge X-ray fluorescence detector (Derbyshire *et al.*, 1999) and the optimum wavelength (0.979 Å) for the collection of f'' anomalous data was selected. Diffraction data were recorded to 2.14 Å resolution at 100 K from a single crystal at this wavelength. This data set was used for initial heavy-atom structure solution and model building in space group *P*2. A second data set was subsequently recorded on a second SeMet-TM1254 crystal at an improved resolution of 1.74 Å using an X-ray wavelength of 0.92 Å. Images were recorded using a MAR Mosaic 225 CCD detector and processed (indexed, integrated and scaled) using *HKL*-2000 (Otwinowski & Minor, 1997). The crystal was found to belong to space group $P22_12_1$, with unit-cell parameters $a = 48.16$, $b = 68.70$, $c = 83.80$ Å. The estimated solvent content was 55.6% for one molecule in the asymmetric unit.

2.4. Structure solution and refinement

Initial efforts to solve the crystal structure of SeMet-TM1254 were performed in space group *P*2, corresponding to two monomers containing a total of 12 Se atoms per asymmetric unit, all of which were located by heavy-atom searching using *SHELXD*. This solution was used for phasing with *SHELXE* (Sheldrick, 2008) and subsequent model building. Attempts to solve and phase the heavy-atom substructure using six Se atoms for a single monomer in an alternative higher symmetry space group ($P22_12_1$) were less successful. Phase improvement by *DM* (Cowtan, 1994) was followed by automated iterative model building to 2.14 Å resolution using *ARP/wARP* (Perrakis *et al.*, 1999) and *REFMAC*5 (Murshudov *et al.*, 1997).
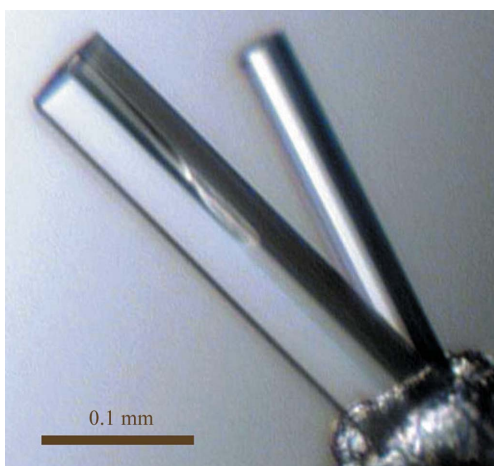

Figure 1
Crystals of *T. maritima* TM1254 prior to flash-cooling in glycerol cryoprotectant.

Table 1
Data-collection and refinement parameters.

Values in parentheses are for the highest resolution shell (1.79–1.74 Å).

Space group	$P22_12_1$
Unit-cell parameters (Å)	$a = 48.16$, $b = 66.70$, $c = 83.80$
Resolution (Å)	31.6–1.74
Unique reflections	27516
Completeness (%)	95.3 (70.8)
Redundancy	7.1 (3.3)
R_{merge}† (%)	4.6 (25.2)
$I/\sigma(I)$	37.8 (3.7)
R factor‡ (%)	19.2
R_{free}‡ (%)	23.9
B factors (Å^2)	
Wilson plot	25.1
Protein	24.9
Water	38.2
Sulfate ions	36.2
Glycerol	36.6
R.m.s. deviations	
Bond distances (Å)	0.015
Bond angles (°)	1.523
ESU§ (Å)	0.078

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

ARP/wARP was able to build 209 of 216 protein residues for each monomer in the *P*2 asymmetric unit, together with 263 solvent atoms. Manual rebuilding and addition of the 14 residues missed by *ARP/wARP* was performed using *Coot* (Emsley & Cowtan, 2004). The resulting structure was then used as the starting model for molecular replacement and further refinement and rebuilding to 1.74 Å resolution using the second data set. Molecular-replacement solution with one monomer in the asymmetric unit was successful in this case for the high-resolution data set scaled in space group $P22_12_1$. The stereochemistry of the refined structure was checked using *MolProbity* (Davis *et al.*, 2007). The Ramachandran plot reported by *PROCHECK* (Laskowski *et al.*, 1993) gave 93.7% of residues in the core region and 6.3% in the additionally allowed regions. The *R* factor and R_{free} for the final model were 19.2% and 23.9%, respectively. Table 1 summarizes the data-collection and refinement parameters. The structure was deposited in the PDB under accession code 3kbb.

3. Results and discussion

Size-exclusion chromatography shows that TM1254 is a monomer in solution. The association of monomers in the asymmetric unit arises only from the crystallographic packing. Each monomer consists of 1763 protein atoms and 235 water molecules. The asymmetric unit also contains four sulfate ions and four glycerol molecules, presumably from the contents of the crystallization medium and cryoprotectant solutions that were used. The structure consists of a 'core' domain comprising six β-strands and six α-helices and a 'cap' domain comprising four α-helices (Fig. 2).

A sequence search using *BLAST* (Lopez *et al.*, 2003) showed that TM1254 has an overall sequence identity of ∼35% to the haloacid dehalogenase (HAD) hydrolase superfamily, with *T. pertophilia* Tpet_1517, *Thermotoga* sp. (strain RQ2) TRQ2_1566 and *Marinitoga piezophila* MPKA3_124 having 93% identity. A search using the *DALI* server (Holm & Sander, 1994) found several structural homologues with 29% identity to TM1254 that belong this family (Koonin & Tatusov, 1994; Allen & Dunaway-Mariano, 2004). The top

RIKEN–UK structural genomics

hits include a putative β-phosphoglucomutase from *Streptococcus thermophilus* (PDB code 3e58, $Z = 23$), a hypothetical protein from *Pyrococcus horikoshii* (PDB code 1x42, $Z = 21.8$; Arai *et al.*, 2006), a conserved HAD-like protein from *Agrobacteruim tumefaciens* strain C58 (PDB code 2fdr, $Z = 21.6$) and a β-phosphoglucomutase from *Lactococcus lactis* (PDB codes 1lvh, $Z = 22.7$; 1zol, $Z = 21.4$; 1o03, $Z = 20.7$; 1z4n, $Z = 20.7$; Lahiri *et al.*, 2002, 2003; Zhang *et al.*, 2005; Tremblay *et al.*, 2005). Superposition of these structures shows that the core domain (residues 1–14 and 86–216) is structurally well conserved, with a root-mean-square deviation of 0.83–1.07 Å relative to the core domain of the TM1254 monomer. The conformation of the cap domain varies from 'open' in the case of Mg^{2+} β-phosphoglucomutase from *L. lactis* and the *A. tumefaciens* protein to 'closed' in the phosphorylated form of the *L. lactis* enzyme and in the *P. horikoshii* protein. The *S. thermophilus* structure contains molecules in both open and closed conformations (PDB code 3e58; C. Chang, C. Tesar, K. Souvong, G. Cobb & A. Joachimiak, unpublished work). The open conformation in these structures exposes the Mg^{2+}-binding site at the interdomain interface to solvent and makes it accessible to ligands. Closure appears to be induced by ligand binding and results in desolvation of the active site and stabilization of the cap domain (Zhang *et al.*, 2005). In the TM1254 structure, with a water molecule in place of Mg^{2+} at the active site, the conformation is wider open than when the metal is bound (Fig. 3). The total solvent-exposed surface area in the cleft between the two domains in TM1254 is 1120 Å^2, enclosing a volume of 3530 Å^3, compared with 830 Å^2 and 1470 Å^3 in the open form of Mg^{2+} β-phosphoglucomutase from *L. lactis*, as calculated using the *CASTp* server (Dundas *et al.*, 2006). Three sulfate ions are found in the interdomain cleft of TM1254, one of which mimics the substrate by occupying the position normally adopted by the bound phosphate group of β-G1P or β-G6P (Fig. 4).

In conclusion, the structure of TM1254 in the apo (Mg^{2+}-free) form has been solved. The structure and fold are similar to those of β-phosphoglucomutase enzymes found in other organisms, with a core domain and a flexible cap domain. The apo form adopts an open conformation that creates a large solvent-accessible cleft between the two domains that is available for the diffusion of ligands (in this case sulfate ions) and solvent to the active site.
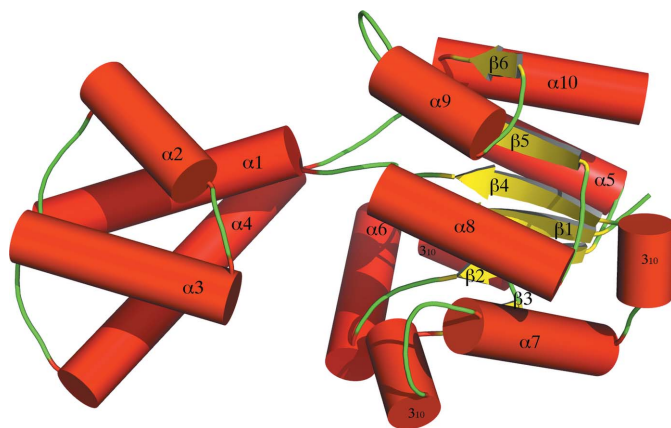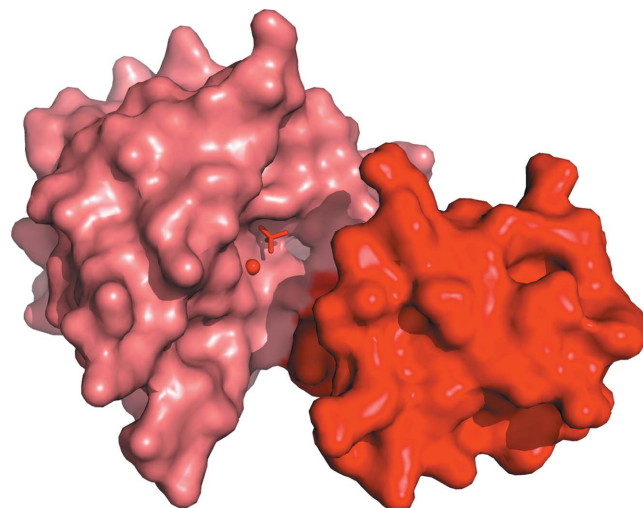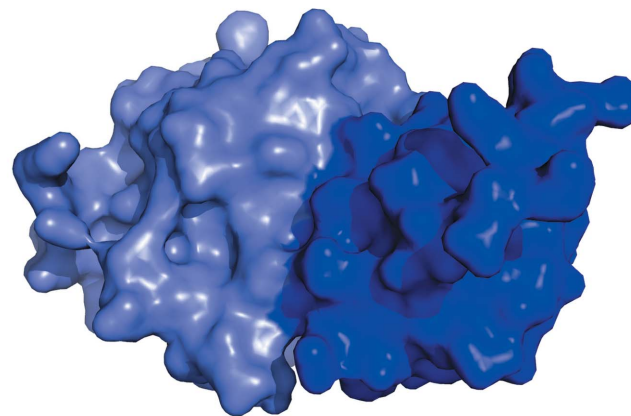
(a)

(b)

(c)

Figure 3
Surface representations of *T. maritima* TM1254 (*a*) together with two conformations of *L. lactis* Mg^{2+} β-phosphoglucomutase, shown in (*b*) in the ligand-bound closed form (PDB entry 1o08; Lahiri *et al.*, 2003) and in (*c*) in the unliganded open form (PDB entry 1zol; Zhang *et al.*, 2005). For clarity, separate colours are used for the core and cap domains, with the core domain shown on the left. Also shown is the location of the active site at the edge of the core domain, illustrated in TM1254 by the water molecule and sulfate ion (red sphere and sticks) and in *L. lactis* by the Mg^{2+} ion (black sphere). The mobile cap domain buries this site in the closed conformation (*b*).
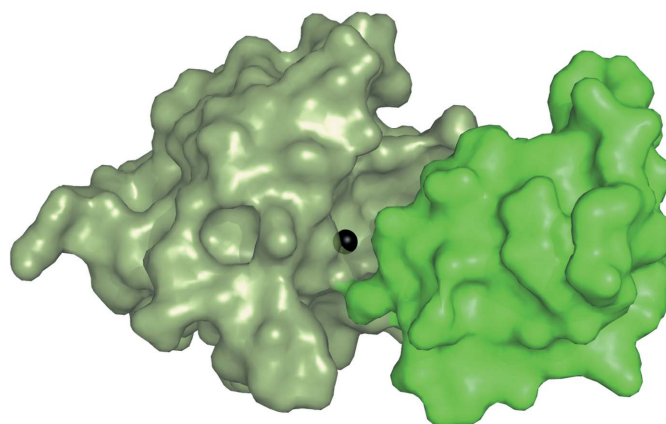

Figure 2
The secondary structure and overall fold of *T. maritima* TM1254. The core globular domain, shown in the right half of the figure, consists of six β-strands (yellow arrows), six α-helices (red cylinders) and three 3_{10}-helices. The cap domain, shown on the left, consists of four α-helices and has a larger isotropic *B* factor than the core domain, signifying greater flexibility for this part of the structure. The overall fold is consistent with the identification of TM1254 as a member of the HAD hydrolase family.
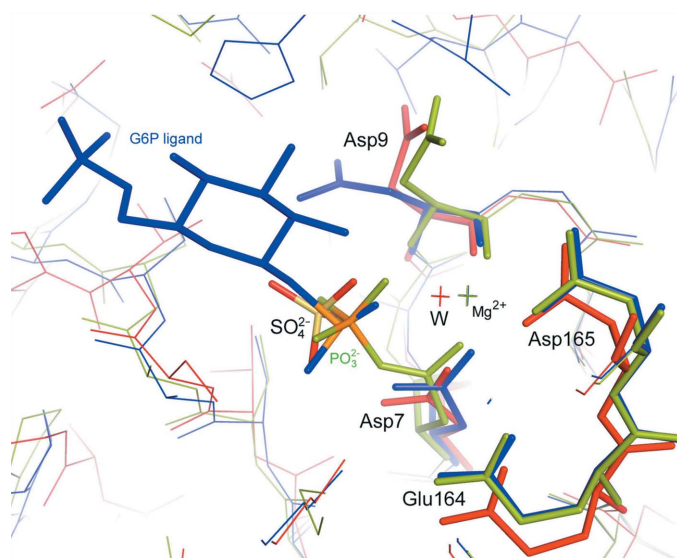
Figure 4
The active site of *T. maritima* TM1254 (red sticks) is shown together with Mg^{2+} β-phosphoglucomutase from *L. lactis*: phosphorylated (PDB entry 1lhv; green sticks; Lahiri *et al.*, 2002) and complexed with β-glucose-6-phosphate-1-phosphorane (G6P; PDB entry 1o03; blue sticks; Lahiri *et al.*, 2003). A water molecule, W, occupies the TM1254 active site in place of the Mg^{2+} ion (other water molecules at the active site are omitted for clarity). A sulfate ion is positioned in the active-site pocket at the location normally occupied by the phosphate group of the bound substrate. The lines at the top of the figure indicate the orientations of residues in the cap domains relative to the binding sites. The residues are numbered according to the TM1254 sequence.

References

Allen, K. N. & Dunaway-Mariano, D. (2004). *Trends Biochem. Sci.* **29**, 495–503.

Arai, R., Kukimoto-Niino, M., Kuroishi, C., Bessho, Y., Shirouzu, M. & Yokoyama, S. (2006). *Protein Sci.* **15**, 373–377.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

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.

Derbyshire, G., Cheung, K.-C., Sangsingkeow, P. & Hasnain, S. S. (1999). *J. Synchrotron Rad.* **6**, 62–63.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Holm, L. & Sander, C. (1994). *Proteins*, **19**, 165–173.

Koonin, E. V. & Tatusov, R. L. (1994). *J. Mol. Biol.* **244**, 125–132.

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. (2002). *Biochemistry*, **41**, 8351–8359.

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. (2003). *Science*, **299**, 2067–2071.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lopez, R., Silventoinen, V., Robinson, S., Kibria, A. & Gish, W. (2003). *Nucleic Acids Res.* **31**, 3795–3798.

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.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sugahara, M. *et al.* (2008). *J. Struct. Funct. Genomics*, **9**, 21–28.

Tremblay, L. W., Zhang, G., Dai, J., Dunaway-Mariano, D. & Allen, K. N. (2005). *J. Am. Chem. Soc.* **127**, 5298–5299.

Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W. & Allen, K. N. (2005). *Biochemistry*, **44**, 9404–9416.