Structure of the methanofuran/methanopterin-biosynthetic enzyme MJ1099 from *Methanocaldococcus jannaschii*

Thomas A. Bobik,^{a*} Erick J. Morales,^b Annie Shin,^c Duilio Cascio,^{c,d} Michael R. Sawaya,^c Mark Arbing,^c Todd O. Yeates^{c,d,e} and Madeline E. Rasche^b

^aRoy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA, ^bDepartment of Chemistry and Biochemistry, California State University, Fullerton, CA 92834, USA,

^cDepartment of Energy Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA 90095, USA,

^dMolecular Biology Institute, University of California Los Angeles, Los Angeles, CA 90095, USA, and ^eDepartment of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095, USA

Correspondence e-mail: bobik@iastate.edu

Received 4 September 2014

Accepted 25 September 2014

PDB references: MJ1099, 4u9p; with bound phosphate, 4rc1

Prior studies have indicated that MJ1099 from *Methanocaldococcus jannaschii* has roles in the biosynthesis of tetrahydromethanopterin and methanofuran, two key cofactors of one-carbon (C1) metabolism in diverse organisms including the methanogenic archaea. Here, the structure of MJ1099 has been solved to 1.7 Å resolution using anomalous scattering methods. The results indicate that MJ1099 is a member of the TIM-barrel superfamily and that it is a homohexamer. Bioinformatic analyses identified a potential active site that is highly conserved among MJ1099 homologs and the key amino acids involved were identified. The results presented here should guide further studies of MJ1099 including mechanistic studies and possibly the development of inhibitors that target the methanogenic archaea in the digestive tracts of humans and that are a source of the greenhouse gas methane.

1. Introduction

Tetrahydromethanopterin (H₄MPT, a folate analog) and methanofuran function as C1 carriers in diverse organisms (Maden, 2000; DiMarco *et al.*, 1990; de Crécy-Lagard *et al.*, 2012). Both are essential cofactors in archaeal methane production, and also function in formaldehyde metabolism during methane oxidation by methylo-trophic bacteria (DiMarco *et al.*, 1990; Möller-Zinkhan *et al.*, 1989; Vorholt *et al.*, 1999, 2000; Marx *et al.*, 2003; Chistoserdova *et al.*, 1998, 2009). Their wide use in prokaryotic C1 metabolism, especially methane cycling, underscores their environmental impact as methane is 20 times more potent as a greenhouse gas than is CO₂ (Wuebbles & Hayhoe, 2000). In addition, H₄MPT and methanofuran-biosynthetic enzymes are potential targets for anti-obesity drugs (Samuel & Gordon, 2006; Samuel *et al.*, 2007; Hansen *et al.*, 2011). The presence of methanogens in the gut has been linked to increased adiposity in mice and higher body-mass index in obese humans (Samuel & Gordon, 2006; Basseri *et al.*, 2012). H₄MPT and methanofuran are essential to the growth of methanogens but are not required by other major groups of intestinal bacteria (Hansen *et al.*, 2011; Samuel & Gordon, 2006; Samuel *et al.*, 2007). Hence, blocking the biosynthesis of these cofactors might allow selective inhibition of methanogens in the gut environment.

The biosynthesis of H₄MPT is complex and most of the enzymes involved are unrelated in sequence to those used for folate production (de Crécy-Lagard *et al.*, 2012; Grochowski & White, 2010). The synthesis of the pterin ring from GTP requires the MptA, MptB, MptD and MptE enzymes (de Crécy-Lagard *et al.*, 2012; Grochowski *et al.*, 2007; Mashhadi *et al.*, 2009). The aminophenyl-ribitol moiety is derived from β -ribofuranosylaminobenzene 5'-phosphate which is produced by a series of reactions that are only partly characterized (Bechard *et al.*, 2003; Dumitru *et al.*, 2003; Rasche & White, 1998; Scott & Rasche, 2002; White, 2011). Studies have also identified three different enzymes (DmrA, DmrB and DmrX) that catalyze the final reductive step of H₄MPT biosynthesis (the conversion of dihydromethanopterin to H₄MPT; Caccamo *et al.*, 2004; McNamara *et al.*, 2014; Wang *et al.*, 2014). Although substantial progress has been made, many of the steps in the proposed pathway of H₄MPT biosynthesis are uncharacterized (de Crécy-Lagard *et al.*, 2012; Grochowski & White, 2010; White, 2011). In *Methylobacterium extorquens*, studies of an archaeal gene cluster acquired by horizontal

transfer found six genes (ORF5, ORF9, ORF19, ORF20, ORF21 and ORF22) that are required for the synthesis of dephosphomethanopterin (an analog of H₄MPT with a shorter side chain; Chistoserdova *et al.*, 1998, 2005). Homologs of these genes are widely distributed in the methanogenic archaea and frequently cluster with other H₄MPT-biosynthetic genes supporting a conserved role in H₄MPT biosynthesis (Kalyuzhnaya *et al.*, 2005; White, 2011). However, the specific functions of these ORFs in H₄MPT biosynthesis are unknown. Surprisingly, a recently published report indicated that the ORF22 homolog from *Methanocaldococcus jannaschii* (MJ1099) is used for the synthesis of methanofuran (Miller *et al.*, 2014). Thus, genetic studies in *Methylobacterium* suggested that MJ1099 and homologs function in H₄MPT biosynthesis, while biochemical studies indicated a role in methanofuran biosynthesis. This raises the possibility that MJ1099 and homologs may be bifunctional enzymes used for the biosynthesis of two physiologically related C1 carriers. Alternatively, however, they might have divergent functions.

Here, we present the crystal structure of MJ1099 (a homolog of ORF22 from *Methylobacterium*) from *M. jannaschii* solved to 1.7 Å resolution using anomalous scattering methods. MJ1099 is a member of the TIM-barrel superfamily. Modeling predicts a substrate-binding pocket near the C-terminus of the β -barrel (as is typical of this family) and a phosphorylated substrate. The specific function of MJ1099 in H₄MPT biosynthesis was not investigated because the proposed substrates are unavailable commercially and are difficult to prepare. Nonetheless, the structure presented should help with future mechanistic studies of MJ1099 and in determining whether it is a bifunctional catalyst.

2. Materials and methods

2.1. Macromolecule production

2.1.1. Production and purification of native His₆-MJ1099. MJ1099 with an N-terminal 6×His tag (His₆-MJ1099) was codon-optimized for expression in *Escherichia coli* by gene synthesis (Genscript, Piscataway, New Jersey, USA) and cloned into the T7 expression vector pE-T41a (Novagen, Darmstadt, Germany) via *Nde*I and *Hind*III sites using standard protocols (Sambrook & Russell, 2001). After sequence verification, pET-41a-His₆-MJ1099 was transformed into *E. coli* BL21 (DE3) RIL cells (Stratagene, La Jolla, California, USA). The resulting strain [*E. coli* BL21 (DE3)-RIL/pET-41a-His₆-MJ1099] was grown and harvested as described previously (McNamara *et al.*, 2014) and the cells were frozen in liquid N₂ and stored at −80°C until use.

For the purification of native His₆-MJ1099, all manipulations were carried out at 4°C or on ice. Approximately 10 g of cells were suspended 1:3(w:v) in 50 mM Tris–HCl pH 8.0 containing 300 mM NaCl, 5% glycerol, 20 mM imidazole, 1 mg ml^{−1} lysozyme, 0.01 mg ml^{−1} DNase, 4 mM dithiothreitol (DTT) and Roche cOmplete protease inhibitor (Roche, Nutley, New Jersey, USA). Cells were broken using an Emulsiflex homogenizer (Avestin Inc., Ottawa, Ontario, Canada) as described by Peti & Page (2007). The lysate was clarified by centrifugation at 32 000g for 30 min and then passed through a 0.45 µm syringe filter (Millipore). His₆-MJ1099 was purified from filtered lysate using nickel-affinity chromatography. Buffer A consisted of 50 mM Tris–HCl pH 8, 300 mM NaCl, 5% glycerol, 4 mM DTT (added just before use). Buffer B was similar to buffer A, but also included 500 mM imidazole. Nickel-affinity chromatography was performed as previously described using a 5 ml HisTrap HP column (GE Healthcare, Piscataway, New Jersey, USA; McNamara *et al.*,

2014). The best fractions from the nickel-affinity chromatography were pooled and concentrated to 15 mg ml^{−1} using a 30 kDa cutoff centrifugal concentrator (EMD Millipore). Concentrated His₆-MJ1099 was further purified using an ÄKTApurifier and an S200 16/300 column (GE Healthcare) loaded with 250 µl sample and developed at 0.4 ml min^{−1} with 20 mM Tris–HCl pH 8, 100 mM NaCl, 5% glycerol, 4 mM DTT. The best fractions were concentrated as described above, frozen in liquid nitrogen and stored at −80°C until used.

2.1.2. Production and purification of selenomethionine-labeled His₆-MJ1099. Plasmid pET-41a-His₆-MJ1099 (see above) was transformed into *E. coli* BL21-Gold (DE3) for production of selenomethionine (SeMet)-labeled protein as described in Arbing *et al.* (2010) but with cell growth at 18°C overnight after induction of protein expression. Cells were harvested by centrifugation and the cell pellet was frozen in liquid nitrogen prior to lysis. The frozen cell pellet was thawed and resuspended in buffer A [20 mM Tris–HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 0.1% *N,N*-dimethyldodecylamine *N*-oxide, 1 mM EDTA, cOmplete protease inhibitor (Roche), 1 mM PMSF, 0.5 µg ml^{−1} DNase I, 1 mg ml^{−1} lysozyme]. Cells were lysed using an Emulsiflex C-3 cell homogenizer (Peti & Page, 2007), MgCl₂ was added to a final concentration of 10 mM and the lysate was clarified by centrifugation at 15 000g for 30 min at 4°C. The clarified supernatant was filtered with a 1.0 µm syringe filter and loaded onto a HisTrap column (GE Healthcare) equilibrated in buffer A. The column was washed extensively and the target protein was eluted with buffer B (buffer A with 250 mM imidazole). Fractions containing pure protein were pooled, concentrated and then further purified by size-exclusion chromatography using a Sephacryl S-100 column (GE Healthcare) equilibrated in buffer C (20 mM Tris–HCl pH 8.0, 150 mM NaCl). The fractions containing pure protein were concentrated to 12 mg ml^{−1} for crystallization screening.

2.2. Crystallization

Protein crystals were obtained using the hanging-drop vapor-diffusion method. Crystallization trials were carried out at the UCLA Macromolecular Crystallization Facility using 96-well plates and commercially available sparse-matrix screens. For sparse-matrix screens, a Mosquito liquid-handling device (TTP LabTech, Cambridge, Massachusetts, USA) was used to set up 210 nl drops with three drops per well with protein:reservoir ratios of 1:1, 1:2 and 2:1. Optimizations were performed manually using 24-well Linbro plates. The crystals used for the native structure determination were obtained by optimization of well F1 of The MPD Suite (Qiagen). Reservoirs contained 500 µl 0.1–0.25 M sodium citrate pH 4.0, 10–35% 2-methyl-2,4-pentanediol (MPD). Hanging drops (3 µl) contained His₆-MJ1099 (in purification buffer at 15 mg ml^{−1}) and reservoir in the following ratios: 1:2, 1.5:1.5 and 2:1(v:v). Crystals took about 1–2 d to develop at room temperature. The best diffracting crystal tested formed in 0.25 M sodium citrate pH 4.0, 26% MPD. No additional cryoprotectant was used. Similar well conditions were used for co-crystallization studies with folate and phosphoribosylpyrophosphate (PRPP), which were both used at 0.5 mM.

For SeMet-labeled MJ1099, crystallization leads were optimized in 24-well plates. Strongly diffracting crystals grew to full size in 3–4 weeks using a 1:1 ratio of protein to reservoir solution (230 mM ammonium citrate, 20% PEG 3350). Crystals were cryoprotected by briefly transferring them into a cryoprotectant solution (5 mM Tris–HCl pH 8.0, 40 mM NaCl, 200 mM ammonium citrate, 20% PEG 3350, 7% PEG 400) followed by flash-cooling in liquid nitrogen.

Table 1
Data collection and processing for MJ1099.
Values in parentheses are for the outer shell.

	Native	Phosphate-bound	SeMet-labeled protein		
			Peak	Inflection	High remote
Diffraction source	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C
Wavelength (Å)	1.0393	1.0393	0.9792	0.9795	0.9717
Temperature (°C)	−173	−173	−173	−173	−173
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Crystal-to-detector distance (mm)	250.0	300.0	370.0	370.0	370.0
Rotation range per image (°)	0.5	1.0	0.5	0.5	0.5
Total rotation range (°)	180.0	360.0	360.0	270.0	180.0
Exposure time per image (s)	0.5	1.0	0.5	0.5	0.5
Space group	C222 ₁	C2	P2 ₁	P2 ₁	P2 ₁
a, b, c (Å)	64.76, 152.60, 153.11	144.87, 153.03, 132.84	61.38, 173.87, 152.34	61.44, 173.85, 152.45	61.48, 174.03, 152.44
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 111.9, 90.0	90.0, 96.7, 90.0	90.0, 96.7, 90.0	90.0, 96.8, 90.0
Mosaicity (°)	0.336	0.223	0.18	0.18	0.18
Resolution range (Å)	76.3–1.7 (1.74–1.70)	69.14–2.40 (2.47–2.40)	86.94–2.23 (2.29–2.23)	75.70–2.38 (2.44–2.38)	87.02–2.27 (2.32–2.27)
Total No. of reflections	533070	357726	1028353	607734	470186
No. of unique reflections	159732	25747	296605	240639	255407
Completeness (%)	98.2 (89.8)	98.3(96.5)	97.4 (81.7)	96.1 (84.2)	87.5 (47.7)
Multiplicity	3.33	13.89	3.47	2.53	1.84
⟨I/σ(I)⟩	9.48 (2.03)	9.17 (2.07)	11.71 (1.81)	8.88 (1.87)	9.62 (1.62)
CC _{1/2}	99.4 (75.6)	99.5 (89.1)	99.7 (73.5)	99.5 (74.8)	99.6 (75.7)
R _{merge} [†] (%)	7.6 (55.2)	7.3 (58.8)	7.1 (72.3)	8.3 (55.5)	5.9 (44.8)
R _{i,m} (%)	9.1 (66.0)	8.7 (69.3)	8.4 (66.0)	10.5 (64.2)	8.1 (61.3)
Overall B factor from Wilson plot (Å ²)	23.04	57.99	37.27	37.97	34.83

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

2.3. Data collection and processing

Diffraction data were collected at −173°C on NE-CAT beamline 24-ID-C using a DECTRIS PILATUS 6M detector at the Advanced Photon Source at Argonne National Laboratory (Table 1). For SeMet-labelled protein, a three-wavelength multi-wavelength anomalous dispersion (MAD) data set was collected from a single crystal and the data were processed with *XDS* (Kabsch, 2010).

2.4. Structure solution and refinement

The structure was solved with *AutoSol* and an initial model was built with *AutoBuild*, both of which are part of the *PHENIX* suite of crystallography programs (Adams *et al.*, 2010). The asymmetric unit of the SeMet-labeled protein crystal contained 12 monomers. One model was used to solve a higher resolution native data set by molecular replacement using the program *Phaser* (McCoy *et al.*, 2007) and the higher resolution model was refined at 1.7 Å resolution. The simulated annealing omit map was prepared with *PHENIX* (Adams *et al.*, 2010).

2.5. Media and chemicals

The rich medium used was lysogeny broth (LB), also known as Luria–Bertani medium (EMD Millipore, Gibbstown, New Jersey, USA; Bertani, 1951). Antibiotics, DNase I and lysozyme were from Sigma Chemical Company (St Louis, Missouri, USA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) and DTT were from Gold Biotechnology Inc. (St Louis, Missouri, USA).

2.6. Protein methods

SDS–PAGE was performed using Bio-Rad Ready gels and Bio-Rad Mini-Protean Tetra electrophoresis cells according to the manufacturer’s instructions (Bio-Rad, Hercules, California, USA). Coomassie Brilliant Blue R-250 was used to stain proteins. Bio-Rad protein-assay reagent (Bio-Rad), which is based on the method of Bradford, was used to determine the concentration of protein samples with bovine serum albumin as the standard (Bradford, 1976).

2.7. Bioinformatic analyses

SiteHound-web (Hernandez *et al.*, 2009) was used to search for possible MJ1099 active sites. The average clustering algorithm was used, and a multi-probe characterization of the putative ligand-binding site was performed using multiple energy cutoffs. To construct multiple sequence alignments, homologs of MJ1099 were identified by *BLAST* searches of the DOE Joint Genome Institute Integrated Microbial Genomes database. Alignments were prepared with *MUSCLE* (Edgar, 2004) and analyzed with *Jalview* (Waterhouse *et al.*, 2009). All images of protein structures were made using *PyMOL* (v1.5; Schrödinger).

2.8. Accession number

The coordinates of the final models and the merged structure factors have been deposited in the Protein Data Bank with PDB codes 4u9p (apoenzyme) and 4rc1 (MJ1099 with phosphate bound).

3. Results

3.1. His₆-MJ1099 purification and molecular-mass estimation

N-terminally His-tagged MJ1099 was produced at high levels in *E. coli* and purified by Ni–NTA and gel-filtration chromatography. By Ni–NTA chromatography, His₆-MJ1099 eluted as a symmetrical peak between 125 and 250 mM imidazole. In the gel-filtration step, a minor void peak was observed, but the major His₆-MJ1099 peak was centered at 12.54 ml with a small upstream shoulder (Supplementary Fig. S1[†]). The best fractions were concentrated using a 30 kDa centrifugal filter to give a protein concentration of 15 mg ml^{−1}. SDS–PAGE indicated that concentrated MJ1099 was >95% pure (Supplementary Fig. S1). This sample was frozen in liquid nitrogen and later used for crystallographic studies. We also estimated the molecular mass of native His₆-MJ1099 based on gel-filtration chro-

[†] Supporting information has been deposited in the IUCr electronic archive (Reference: SW5076).

Table 2

Structure solution and refinement for MJ1099.

Values in parentheses are for the outer shell.

	Native	Phosphate-bound	SeMet-labeled protein		
			Peak	Inflection	High remote
Resolution range (Å)	76.3–1.70 (1.74–1.70)	69.14–2.40 (2.47–2.40)	86.94–2.23 (2.29–2.23)	75.70–2.38 (2.44–2.38)	87.02–2.27 (2.32–2.27)
Completeness (%)	98.2 (89.8)	98.3 (96.5)	97.4 (81.7)	96.1 (84.2)	87.5 (47.7)
σ Cutoff	0.0	0.0			
No. of reflections, working set	82236	102608			
No. of reflections, test set	4145	5118			
$R_{\text{work}}^{\dagger}$ (%)	20.17 (24.61)	19.29 (23.36)			
$R_{\text{free}}^{\ddagger}$ (%)	22.29 (23.76)	21.05 (28.24)			
Cruickshank DPI	0.100	0.305			
No. of non-H atoms					
Protein	5146	16051			
Ion	0	0			
Ligand	6	30			
Water	122	150			
Total	5274	16231			
R.m.s. deviations					
Bonds (Å)	0.01	0.01			
Angles (°)	1.08	1.19			
Average B factors (Å ²)					
Protein	32.0	103.6 §			
Ion	—	—			
Ligand	45.25	129.1			
Water	29.19	67.57			
Ramachandran plot¶ (%)					
Most favoured	92.5	92.0			
Allowed	7.0	7.9			
Generously allowed	0.5	0.1			
Disallowed	0.0	0.0			

[†] $R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$. [‡] $R_{\text{free}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ calculated using a random set containing 5% reflections that were not included throughout refinement. § The MJ1099 phosphate model has a high overall B factor owing to anisotropy in the diffraction data. ¶ The percentages of residues in Ramachandran plot regions were determined using *PROCHECK*.

matography. By comparison with protein standards (log molecular mass *versus* retention volume) the molecular mass of native His₆-MJ1099 was estimated at 171 000 Da. The calculated mass of the His₆-MJ1099 monomer is 26 053 Da. This suggests 6.6 monomers in the biological unit, which in conjunction with analysis of crystallographic data (below) suggests MJ1099 is a homo-hexamers.

3.2. Crystallography

Crystals of native MJ1099 and SeMet-labeled MJ1099 were prepared as described in §2. The native crystals were mostly rod-shaped and up to 1 mm in length (Supplementary Fig. S1). Crystals of SeMet MJ1099 were rod-shaped but somewhat smaller in size. The best native crystal diffracted to 1.7 Å resolution. The SeMet crystal used for structure determination diffracted to 2.3 Å resolution. We were unable to solve the structure by molecular replacement using the native data set and homology-based models obtained *via Phyre*² (Kelley & Sternberg, 2009); the best model was 16% identical in amino-acid sequence with 97% coverage. The structure was then solved by the MAD method and the solution was used as a molecular-replacement model for a higher resolution (1.7 Å) native data set. The final model was built with native data and refined to 1.7 Å resolution with residual errors of $R_{\text{work}} = 0.20$ and $R_{\text{free}} = 0.22$ (Table 2). All amino-acid residues were fitted to the electron density except residues 213–218 (KGGDRN), which are presumed to be in a disordered loop, and the last amino acid at the C-terminus (Lys235).

3.3. Subunit and quaternary structure

The tertiary structure of MJ1099 is a TIM barrel (Fig. 1). Although variations are seen in the TIM superfamily, MJ1099 has the classic TIM (β/α)₈-fold and forms a barrel composed of eight parallel β -strands surrounded by eight α -helices (Vega *et al.*, 2003). The TIM

superfamily is a large and diverse group of proteins (Gerlt & Raushel, 2003; Henn-Sax *et al.*, 2001; Nagano *et al.*, 2002; Vega *et al.*, 2003; Wierenga, 2001). The Structural Classification of Proteins (SCOP) database (v.1.73) currently lists 33 TIM β/α -barrel fold families. TIM-barrel proteins have diverse functions and the great majority are enzymes (Gerlt & Raushel, 2003; Henn-Sax *et al.*, 2001; Nagano *et al.*, 2002; Vega *et al.*, 2003; Wierenga, 2001). Because of their wide distribution and sequence variation, there is not yet a consensus that a TIM-barrel fold indicates an evolutionary relationship (Vega *et al.*, 2003).

Members of the (β/α)₈-barrel superfamily have diverse quaternary structures including monomers, dimers, tetramers and decamers (Gerlt & Raushel, 2003; Henn-Sax *et al.*, 2001; Nagano *et al.*, 2002; Vega *et al.*, 2003; Wierenga, 2001). Analysis of the MJ1099 structural model with the *PISA* software and gel-filtration chromatography of purified enzyme (see above) indicated that MJ1099 is a homo-hexamer arranged as a trimer of dimers. The dimer interface involves interactions between helices $\alpha 3/\alpha 3$ on different subunits and reciprocal $\alpha 2$ – $\alpha 3$ interactions. These interactions bury 2194 Å² of surface area according to *PISA* (Krissinel & Henrick, 2007). The surfaces that hold the dimers together in the hexameric structure involve several helices and loops and the buried surface area is 1488 Å² at each of the three interfaces (Krissinel & Henrick, 2007).

3.4. Bioinformatic prediction of the MJ1099 active site

The *SiteHound-web* software was used to examine MJ1099 for potential substrate-binding sites (Hernandez *et al.*, 2009) (Fig. 2). This software uses chemical probes to identify small-molecule protein-binding sites. Four probes were used: a phosphate oxygen probe (OP) designed to find binding sites for phosphorylated ligands, a methyl carbon probe (CMET) which is optimized to identify binding sites

dominated by van der Waals interactions, an aromatic carbon probe (CR1) and a hydroxyl oxygen probe (OA). The best binding site was ranked first with three probes (OP, CR1 and OA) and second with the fourth probe (CMET). Importantly, this site is located near the C-terminal face of the β -barrel as expected for TIM-barrel active sites (Vega *et al.*, 2003). The highest interaction energy was obtained with the OP probe (−3369 arbitrary units) and this value was substantially higher than the second ranked site found (−1478 arbitrary units), supporting a genuine binding site (Gherzi & Sanchez, 2009). This pocket is relatively large at 40 Å³ and is primarily formed from 18 amino acids that are highly conserved among MJ1099 and homologs

(the percentage conservation of each amino acid is indicated in parentheses): Leu4 (97%), Ser6 (98%), Asp25 (99%), Lys27 (99%), Leu34 (98%), Lys85 (99%), Asp151 (100%), Thr152 (100%), Lys155 (100%), Ala186 (98%), Gly187 (99%), Ser188 (82%), Gly206 (96%), Arg208 (99%), Gly209 (78%) and Ala210 (99%). In addition, two of 18 amino acids that form the putative binding pocket show conservation of chemically similar amino acids as indicated parenthetically: Ile189 (100% Ile or Leu or Val) and Val207 (98% Val or Phe) (Fig. 2). Thus, the site described above represents a reasonable active site for MJ1099.

3.5. An MJ1099 phosphate-binding site

DELTA-BLAST (Domain Enhanced Lookup Time Accelerated-Basic Local Alignment Search Tool; Boratyn *et al.*, 2012) and *Phyre*² (Kelley & Sternberg, 2009) searches were performed using MJ1099 and various homologs as query sequences. When the *M. extorquens* MJ1099 homolog (ORF22) was used as the query, a number of hits (expect = 5×10^{-24} to 3×10^{-21}) were found to quinolinate phosphoribosyl transferase/nicotinate-nucleotide pyrophosphorylase, which uses quinolinic acid and PRPP as substrates. In addition, a *Phyre*² search aligned MJ1099 with phosphoribosyl anthranilate isomerase (98.9% confidence and 89% coverage). This tentatively suggested that a phosphoribose might be a substrate for MJ1099 and prior studies had proposed that the fifth step in H₄MPT side-chain biosynthesis involves the transfer of a phosphoribosyl group from PRPP to a pterin-containing intermediate (White, 2001). Therefore, attempts were made to co-crystallize MJ1099 in the presence of a pterin-containing analog (folate) and PRPP. A good-quality crystal was obtained when MJ1099 was crystallized in the presence of folate and PRPP (1 mol:1 mol ligand:protein) and a 2.4 Å data set was collected. Molecular replacement was performed using the structure determined above and a model was refined to an R_{work} of 0.19 and R_{free} of 0.21 (Tables 1 and 2). An $F_o - F_c$ map revealed a region of density indicative of a phosphate or a sulfate ion and this density was also observed in an $F_o - F_c$ simulated-annealing map (Fig. 3). No added sulfate or phosphate was present in the crystallization condition, but phosphate may have been produced by hydrolysis of PRPP which is unstable in aqueous solution, particularly at acid and alkaline pH values (the pH of the well solution was 4.0; Peel, 1969). Therefore, the observed density is likely to correspond to inorganic phosphate. Interestingly, this phosphate was bound primarily by four highly conserved amino acids (**Lys155**, **Ser188**, **Arg208** and Arg217) three of which (indicated in bold) were part of the binding site predicted by *SiteHound-web* (Fig. 3). It was also observed that a previously unstructured loop (213–218, RGGDRN) became ordered upon phosphate binding and that the arginine in this loop is involved in electrostatic interactions with the ligand (phosphate).

4. Discussion

Here, we present the structure of MJ1099 solved by anomalous scattering methods to 1.7 Å resolution. The results indicated that MJ1099 is a member of the TIM-barrel superfamily and that the biological unit is a homohexamer. A putative active site was found by bioinformatics methods and was shown to be highly conserved among MJ1099 homologs. Importantly, this site was located on the C-terminal face of the β -barrel, which is the location of the active sites of virtually all TIM-barrel proteins (Vega *et al.*, 2003). Thus, a good active-site candidate was found. The results presented here also suggest a phosphorylated substrate or cofactor. About two thirds of TIM-barrel enzymes use substrates or cofactors that contain one or

Figure 1
Subunit and quaternary structure. (a) MJ1099 folds into a classic TIM-barrel (β/α)₈ fold composed of eight parallel β -strands surrounded by eight α -helices. (b) The biological unit of MJ1099 is a homohexamer that assembles as a trimer of dimers.

more phosphate groups (Vega *et al.*, 2003). A *SiteHound-web* probe optimized for phosphorylated substrates identified a high-scoring binding site and an inorganic phosphate ion was bound to this site in one crystal form (PDB entry 4rc1). However, the biochemical role of MJ1099 in H₄MPT synthesis was not investigated in this study because the required substrates are unavailable commercially and are difficult to synthesize (Chistoserdova *et al.*, 1998, 2005).

Prior genetic studies indicated that a homolog of MJ1099 found in *Methylobacterium* (ORF22) was used for the biosynthesis of H₄MPT (Chistoserdova *et al.*, 2005). An ORF22 gene knockout was unable to synthesize dephospho-H₄MPT and was complemented by an ORF22 minimal clone (Chistoserdova *et al.*, 2005). This is very strong evidence that ORF22 from *Methylobacterium* has a role in dephos-

pho-H₄MPT biosynthesis. Bioinformatic analyses found that ORF22 and MJ1099 are 32% identical in amino-acid sequence with 92% coverage (expect = 4×10^{-24}) and that 16/18 amino acids predicted to form the MJ1099 binding pocket are conserved in *Methylobacterium* ORF22 (the two changes observed were conservative, Ile to Leu and Ala to Gly). Analyses also showed that MJ1099 is the only homolog of ORF22 present in *Methanocaldococcus* which is known to require H₄MPT for growth. Thus, bioinformatic analyses suggest MJ1099 may be used for H₄MPT biosynthesis. Surprisingly, however, a recent study showed that MJ1099 can catalyze the condensation of glyceraldehyde-3-phosphate and dihydroxyacetone phosphate to form 4-(hydroxymethyl)-2-furancarboxaldehyde phosphate (4-HFC-P) and that this reaction is needed for the biosynthesis of methanofuran

Figure 2
The predicted active site of MJ1099. (a) The small-molecule-binding site predicted by *SiteHound-web* is shown as orange spheres. (b) Interacting amino acids found by *SiteHound-web* (not all interactions are visible in this view; a complete list is given in the text). (c) Surface view of the binding pocket. (d) Surface view created with highly conserved residues colored magenta and variable residues colored cyan. The conservation level was determined and coloring was performed with the *ConSurf* server (Ashkenazy *et al.*, 2010; Landau *et al.*, 2005).

(Miller *et al.*, 2014). Like H₄MPT, methanofuran is a required C1 carrier in methanogenesis (DiMarco *et al.*, 1990). This tentatively suggests that MJ1099 and homologs are bifunctional enzymes used for the synthesis of two key C1 carriers: H₄MPT and methanofuran. However, given the somewhat low similarity between OFR22 from *Methylobacterium* and MJ1099 from *Methanocaldococcus* (32% identity and 42% homology) further work will be needed to determine whether these enzymes are indeed bifunctional. The structure presented here should help unravel the physiological roles of OFR22 and MJ1099. In addition, it should help with mechanistic studies of MJ1099 and with the identification of small molecules useful for the inhibition of methanogens for biotechnological applications.

We thank the members of the Yeates laboratory for valuable discussions and general assistance. Our thanks to Michael Collazo for help with setting up crystal trays and to Jamie Lu for growth of cells for the production of selenomethionine-labeled protein. We also

thank M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov and F. Murphy at NE-CAT beamline 24-ID of the APS, Argonne National Laboratory, and Gabriela Naranjo and Alejandro Penaloza (California State University, Fullerton) for initial production and purification studies of OFR22, the MJ1099 homolog in *M. extorquens*. Selenomethionine-labeled protein was prepared at the UCLA-DOE Protein Expression Technology Center, initial crystallization trials were performed at the UCLA-DOE X-ray Crystalization facility and in-house X-ray diffraction analyses were conducted at the UCLA Crystallography Core Facility, all of which are supported by DOE grant DE-FC02-02ER63421. X-ray diffraction data were also collected at the APS, Argonne National Laboratory, which is supported by grants from the National Center for Research Resources (5P41RR015301-10), the NIH (8 P41 GM103403-10) and the DOE under Contract DE-AC02-06CH11357. This work was also supported by the BER program of the Department of Energy Office of Science (TOY), NSF grant MCB-1020200 (MER) and Senior Intramural Research Award #3539 from California State University at Fullerton (MER).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Arbing, M. A., Kaufmann, M., Phan, T., Chan, S., Cascio, D. & Eisenberg, D. (2010). *Protein Sci.* **19**, 1692–1703.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). *Nucleic Acids Res.* **38**, W529–W533.
- Basseri, R. J., Basseri, B., Pimentel, M., Chong, K., Youdim, A., Low, K., Hwang, L., Soffer, E., Chang, C. & Mathur, R. (2012). *Gastroenterol. Hepatol.* **8**, 22–28.
- Bechard, M. E., Chhatwal, S., Garcia, R. E. & Rasche, M. E. (2003). *Biol. Proced. Online*, **5**, 69–77.
- Bertani, G. (1951). *J. Bacteriol.* **62**, 293–300.
- Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J. & Madden, T. L. (2012). *Biol. Direct*, **7**, 12.
- Bradford, M. (1976). *Anal. Biochem.* **72**, 248–254.
- Caccamo, M. A., Malone, C. S. & Rasche, M. E. (2004). *J. Bacteriol.* **186**, 2068–2073.
- Chistoserdova, L., Kalyuzhnaya, M. G. & Lidstrom, M. E. (2009). *Annu. Rev. Microbiol.* **63**, 477–499.
- Chistoserdova, L., Rasche, M. E. & Lidstrom, M. E. (2005). *J. Bacteriol.* **187**, 2508–2512.
- Chistoserdova, L., Vorholt, J. A., Thauer, R. K. & Lidstrom, M. E. (1998). *Science*, **281**, 99–102.
- Crécy-Lagard, V. de, Phillips, G., Grochowski, L. L., El Yacoubi, B., Jenney, F., Adams, M. W. W., Murzin, A. G. & White, R. H. (2012). *ACS Chem. Biol.* **7**, 1807–1816.
- DiMarco, A. A., Bobik, T. A. & Wolfe, R. S. (1990). *Annu. Rev. Biochem.* **59**, 355–394.
- Dumitru, R., Palencia, H., Schroeder, S. D., DeMontigny, B. A., Takacs, J. M., Rasche, M. E., Miner, J. L. & Ragsdale, S. W. (2003). *Appl. Environ. Microbiol.* **69**, 7236–7241.
- Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.
- Gerlt, J. A. & Raushel, F. M. (2003). *Curr. Opin. Chem. Biol.* **7**, 252–264.
- Gheri, D. & Sanchez, R. (2009). *Proteins*, **74**, 417–424.
- Grochowski, L. L. & White, R. H. (2010). *Comprehensive Natural Products II: Chemistry and Biology*, edited by T. P. Begley, pp. 711–748. New York: Elsevier.
- Grochowski, L. L., Xu, H., Leung, K. & White, R. H. (2007). *Biochemistry*, **46**, 6658–6667.
- Hansen, E. E., Lozupone, C. A., Rey, F. E., Wu, M., Guruge, J. L., Narra, A., Goodfellow, J., Zaneveld, J. R., McDonald, D. T., Goodrich, J. A., Heath, A. C., Knight, R. & Gordon, J. I. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 4599–4606.
- Henn-Sax, M., Höcker, B., Wilmanns, M. & Sterner, R. (2001). *Biol. Chem.* **382**, 1315–1320.
- Hernandez, M., Gheri, D. & Sanchez, R. (2009). *Nucleic Acids Res.* **37**, W413–W416.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kalyuzhnaya, M. G., Korotkova, N., Crowther, G., Marx, C. J., Lidstrom, M. E. & Chistoserdova, L. (2005). *J. Bacteriol.* **187**, 4607–4614.

Figure 3
Phosphate bound to MJ1099 as observed in one crystal form. The phosphate-binding site corresponds with the small-molecule-binding pocket predicted by SiteHound-web (Fig. 2). (a) Simulated-annealing omit map contoured at +5σ and showing key amino-acid contacts. (b) An electrostatic surface potential map shows that phosphate is bound to a positively charged region of MJ1099.

- Kelley, L. A. & Sternberg, M. J. (2009). *Nature Protoc.* **4**, 363–371.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). *Nucleic Acids Res.* **33**, W299–W302.
- Maden, B. E. (2000). *Biochem. J.* **350**, 609–629.
- Marx, C. J., Chistoserdova, L. & Lidstrom, M. E. (2003). *J. Bacteriol.* **185**, 7160–7168.
- Mashhadi, Z., Xu, H. & White, R. H. (2009). *Biochemistry*, **48**, 9384–9392.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- McNamara, D. E., Cascio, D., Jorda, J., Bustos, C., Wang, T.-C., Rasche, M. E., Yeates, T. O. & Bobik, T. A. (2014). *J. Biol. Chem.* **289**, 8852–8864.
- Miller, D., Wang, Y., Xu, H., Harich, K. & White, R. H. (2014). *Biochemistry*, **53**, 4635–4647.
- Möller-Zinkhan, D., Börner, G. & Thauer, R. K. (1989). *Arch. Microbiol.* **152**, 362–368.
- Nagano, N., Orengo, C. A. & Thornton, J. M. (2002). *J. Mol. Biol.* **321**, 741–765.
- Peel, J. L. (1969). *Data for Biochemical Research*, 2nd ed., edited by R. M. C. Dawson, D. C. Elliott, W. H. Elliott & K. M. Jones, p. 118. London: Oxford University Press.
- Peti, W. & Page, R. (2007). *Protein Expr. Purif.* **51**, 1–10.
- Rasche, M. E. & White, R. H. (1998). *Biochemistry*, **37**, 11343–11351.
- Sambrook, J. & Russell, D. (2001). *Molecular Cloning: a Laboratory Manual*, 2nd ed. New York: Cold Spring Harbor Laboratory Press.
- Samuel, B. S. & Gordon, J. I. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 10011–10016.
- Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R. K. & Gordon, J. I. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 10643–10648.
- Scott, J. W. & Rasche, M. E. (2002). *J. Bacteriol.* **184**, 4442–4448.
- Vega, M. C., Lorentzen, E., Linden, A. & Wilmanns, M. (2003). *Curr. Opin. Chem. Biol.* **7**, 694–701.
- Vorholt, J. A., Chistoserdova, L., Stolyar, S. M., Thauer, R. K. & Lidstrom, M. E. (1999). *J. Bacteriol.* **181**, 5750–5757.
- Vorholt, J. A., Marx, C. J., Lidstrom, M. E. & Thauer, R. K. (2000). *J. Bacteriol.* **182**, 6645–6650.
- Wang, S., Tionson, J. & Rasche, M. E. (2014). *J. Bacteriol.* **196**, 203–209.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
- White, R. H. (2001). *Vitam. Horm.* **61**, 299–337.
- White, R. H. (2011). *Biochemistry*, **50**, 6041–6052.
- Wierenga, R. K. (2001). *FEBS Lett.* **492**, 193–198.
- Wuebbles, D. J. & Hayhoe, K. (2000). *Non-CO₂ Greenhouse Gases: Scientific Understanding, Control and Implementation*, edited by J. van Ham, A. P. M. Baede, L. A. Meyer & R. Ybema, pp. 1–44. Dordrecht: Kluwer Academic Publishers.