Crystal structure of a glycerate kinase (TM1585) from Thermotoga maritima at 2.70 Å resolution reveals a new fold

28 AUTHORS, INCLUDING:

Marc A Elsliger
The Scripps Research Institute
155 PUBLICATIONS **3,332** CITATIONS

SEE PROFILE

Timothy M McPhillips
AbsoluteFlow
86 PUBLICATIONS **3,067** CITATIONS

SEE PROFILE

John Wooley
University of California, San Diego
163 PUBLICATIONS **3,250** CITATIONS

SEE PROFILE

STRUCTURE NOTE

Crystal Structure of a Glycerate Kinase (TM1585) from *Thermotoga maritima* at 2.70 Å Resolution Reveals a New Fold

Robert Schwarzenbacher,[1,3] Daniel McMullan,[1,4] S. Sri Krishna,[1,2,3] Qingping Xu,[1,5] Mitchell D. Miller,[1,5] Jaume M. Canaves,[1,3] Marc-André Elsliger,[1,6] Ross Floyd,[1,5] Slawomir K. Grzechnik,[1,3] Lukasz Jaroszewski,[1,2,3] Heath E. Klock,[1,4] Eric Koesema,[1,4] John S. Kovarik,[1,5] Andreas Kreusch,[1,4] Peter Kuhn,[1,6] Timothy M. McPhillips,[1,5] Andrew T. Morse,[1,3] Kevin Quijano,[1,4] Glen Spraggon,[1,4] Raymond C. Stevens,[1,6] Henry van den Bedem,[1,5] Guenter Wolf,[1,5] Keith O. Hodgson,[1,5] John Wooley,[1,3] Ashley M. Deacon,[1,5] Adam Godzik,[1,2,3] Scott A. Lesley,[1,4,6] and Ian A. Wilson[1,6]*

[1]*Joint Center for Structural Genomics*
[2]*Burnham Institute for Medical Research, La Jolla, California*
[3]*Center for Research in Biological Systems, University of California, San Diego, La Jolla, California*
[4]*Genomics Institute of the Novartis Research Foundation, San Diego, California*
[5]*Stanford Synchrotron Radiation Laboratory, Stanford University, Menlo Park, California*
[6]*The Scripps Research Institute, La Jolla, California*

Introduction. The TM1585 gene of *Thermotoga maritima* encodes a putative glycerate kinase (EC 2.7.1.-)[1,2] with a molecular weight of 44,589 Da (residues 1–417) and a calculated isoelectric point of 5.73. This enzyme is part of the Entner-Doudoroff pathway II (non-phosphorylative), where it catalyzes the ATP-dependent conversion of (R)-glycerate to 2-phospho-(R)-glycerate and ADP [Fig. 1(A)]. Two groups of glycerate kinases are known, and they share no significant sequence similarity despite their presumably identical biochemical activities. The first group consists of glycerate kinases from bacterial species, primarily of the Firmicutes group and the gamma subdivision of Proteobacteria. The second group includes TM1585 and glycerate kinases from eukaryotes and archaea in addition to several bacterial species.[3] The initial annotation for TM1585 as a glycerate kinase was based on its homology to a gene responsible for complementation in *Methylobacterium extorquens* AM1 mutants lacking glycerate kinase activity.[4] However, other family members related to TM1585 are annotated as putative glycerate dehydrogenases/hydroxypyruvate reductases based on genetic analysis of the tartrate utilization pathway in *Agrobacterium vitis*.[5] Glycerate kinase and glycerate dehydrogenase/hydroxypyruvate reductase catalyze successive steps in the serine metabolism pathway.[4] Recent biochemical studies have shown a glycerate-2-kinase activity for TM1585 (A. Osterman et al., personal communication). In humans, deficiency of glycerate kinase leads to D-glyceric aciduria (OMIM 220120), which is characterized by massive amounts of D-glyceric acid in the urine.[6] Clinical features include delayed psychomotor development, mental retardation, and seizures[7]. Here, we report the crystal structure of TM1585, the first structural representative of the glycerate-2-kinase family, which was determined using the semiau-

tomated, high-throughput pipeline of the Joint Center for Structural Genomics (JCSG).[8]

Materials and Methods. *Protein production and crystallization.* The TM1585 gene (TIGR: TM1585; Swiss-Prot: Q9X1S1) was amplified by polymerase chain reaction (PCR) from genomic DNA using *PfuTurbo* (Stratagene) and primers corresponding to the predicted 5′- and 3′-ends. The PCR product was cloned into plasmid pMH1, which encodes an expression and purification tag (MGSD-KIHHHHHH) at the amino terminus of the full-length protein. The cloning junctions were confirmed by DNA sequencing. Protein expression was performed in a selenomethionine-containing medium using the *Escherichia coli* methionine auxotrophic strain DL41. At the end of fermentation, 250 μg/mL lysozyme was added to the culture. Bacteria were lysed by sonication after a freeze/thaw procedure in Lysis Buffer [50 mM Tris (pH 7.9), 50 mM NaCl, 0.25 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)], and the cell debris was pelleted by centrifugation at 3400 × g for 60 min. The soluble fraction was applied to nickel-chelating resin (GE Healthcare) pre-equilibrated with Equilibration Buffer [50 mM potassium phosphate (pH 7.8), 300 mM NaCl, 10% (v/v) glycerol, 0.25 mM TCEP] containing 20 mM imidazole. The resin
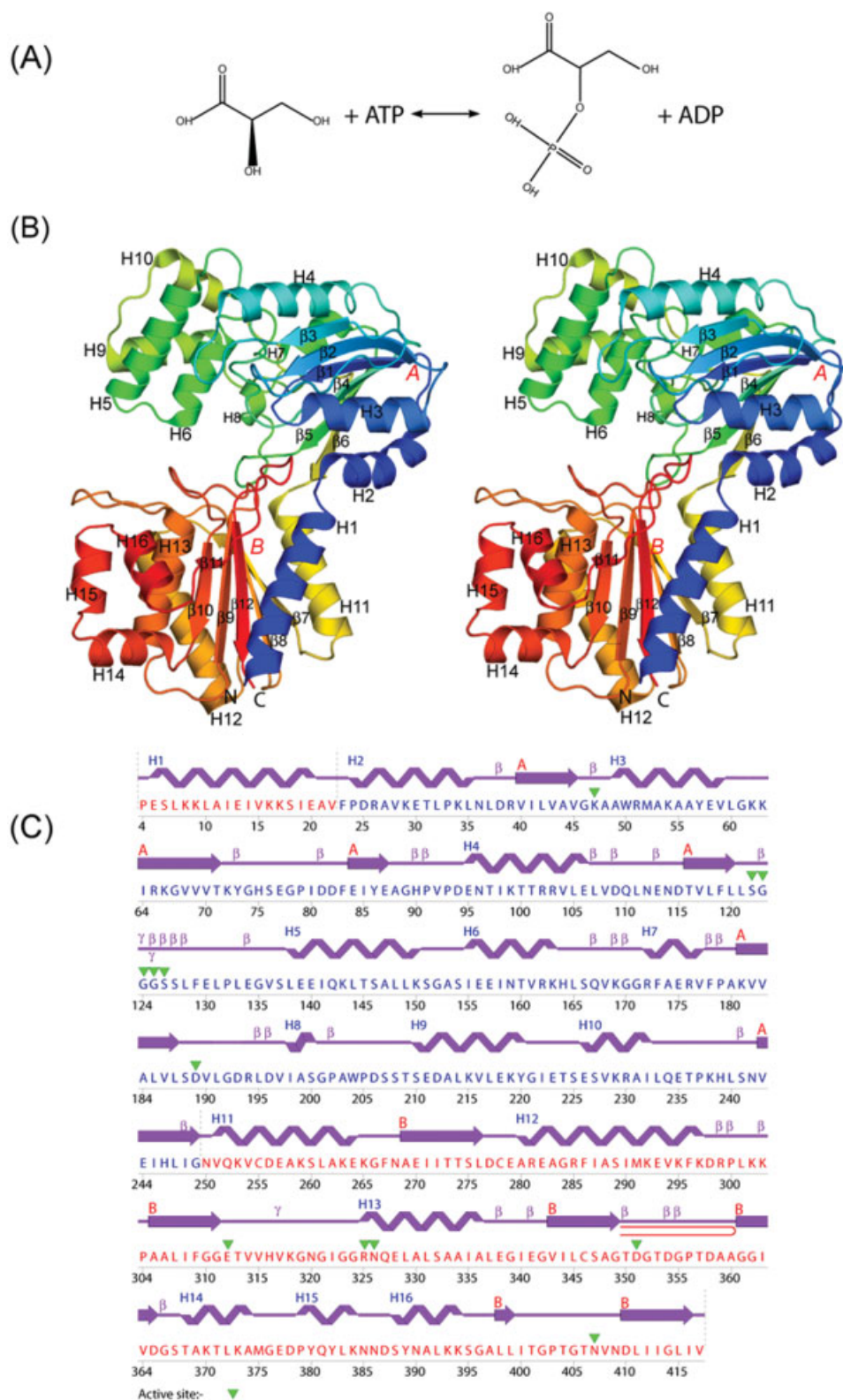
Fig. 1.   **A:** Reaction catalyzed by glycerate-2-kinase (EC 2.7.1.-). **B:** Crystal structure of TM1585. Stereo ribbon diagram of TM1585 monomer color-coded from N-terminus (blue) to C-terminus (red). Helices (H1–H16) and β-strands (β1–β12) are labeled. N- and C-terminal domains are denoted by *A* and *B*, respectively. **C:** Diagram showing the secondary structural elements in TM1585 superimposed on its primary sequence. The α-helices, 3_{10}-helix, β-strands, β-bulges, and γ-turns are indicated. The β-hairpin is depicted as a red loop. Highly conserved residues in the putative active site are marked with a green triangle. N- and C-terminal domain residues are shown in blue and red, respectively. β-strands from N- and C-terminal domains are designated in β-sheets A and B, respectively.

TABLE I. Summary of Crystal Parameters, Data Collection, and Refinement Statistics for TM1585 (PDB: 2b8n)

Space group	$P2_12_12_1$		
Unit cell parameters	$a = 61.95$ Å, $b = 85.17$ Å, $c = 172.51$ Å, $\alpha = \beta = \gamma = 90.00°$		
Data Collection	λ_1MADSe	λ_2MADSe	
Wavelength (Å)	0.9050	0.9791	
Resolution range (Å)	50.0–2.52	50.0–2.73	
Number of observations	77,938	70,521	
Number of reflections	26,443	22,168	
Completeness (%)	83.4 (43.5)[a]	89.0 (55.7)[a]	
Mean I/σ(I)	10.0 (2.8)[a]	11.1 (3.5)[a]	
R_{sym} on I[c]	0.079 (0.277)[a]	0.076 (0.231)[a]	
Highest resolution shell (Å)	2.61–2.52	2.83–2.73	
Model and refinement statistics			
Resolution range (Å)	48.11–2.53	Data set used in refinement	λ_1MADSe
Number of reflections (total)	26,398[f]	Cutoff criteria	$|F| > 0$
Number of reflections (test)	1,314	R_{cryst}[d]	0.182
Completeness (% total)	84.4	R_{free}[e]	0.246
Stereochemical parameters			
Restraints (RMS observed)			
Bond length	0.013 Å		
Bond angle	1.38°		
Average isotropic B-value	29.0 Å^2		
ESU based on R_{free} value[b]	0.33 Å		
Protein residues/atoms	828 / 6244		
Solvent molecules	97		

[a]Highest resolution shell.
[b]ESU = Estimated overall coordinate error[11,25]
[c]$R_{sym} = \Sigma|I_i - <I_i>|/\Sigma|I_i|$ where I_i is the scaled intensity of the ith measurement and $<I_i>$ is the mean intensity for that reflection.
[d]$R_{cryst} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[e]R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.
[f]Typically, the number of unique reflections used in refinement is slightly less than the total number that were integrated and scaled. Reflections are excluded due to systematic absences, negative intensities, and rounding errors in the resolution limits and cell parameters.

was washed with Equilibration Buffer containing 40 mM imidazole, and the target protein was eluted with Elution Buffer [20 mM Tris (pH 7.9), 300 mM imidazole, 10% (v/v) glycerol, 0.25 mM TCEP]. The eluate was buffer exchanged with Buffer Q [20 mM Tris pH 7.9, 5% (v/v) glycerol, 0.25 mM TCEP] containing 50 mM NaCl and applied to a RESOURCE Q column (GE Healthcare) pre-equilibrated with the same buffer. The protein was eluted using a linear gradient of 50 to 500 mM NaCl in Buffer Q. The appropriate fractions were pooled, buffer exchanged with Crystallization Buffer [20 mM Tris (pH 7.9), 150 mM NaCl, 0.25 mM TCEP], and concentrated for crystallization assays to 10.5 mg/mL by centrifugal ultrafiltration (Millipore). Molecular weight and oligomeric state of the target protein were determined using a 1.0×30 cm Superdex 200 column (GE Healthcare) in combination with static light scattering (Wyatt Technology). The mobile phase consisted of 20 mM Tris (pH 8.0), 150 mM NaCl, 0.02% (w/v) sodium azide. The protein was crystallized using the nanodroplet vapor diffusion method [9] with standard JCSG crystallization protocols.[8] The crystallization reagent contained 65% (v/v) 2-methyl-2,4-pentanediol (MPD), 0.1 M Bicine (pH 9.0). The crystals were indexed in orthorhombic space group $P2_12_12_1$ (Table I).

Data collection. Multi-wavelength anomalous diffraction (MAD) data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 9-2 at wavelengths corresponding to the high energy remote (λ_1) and inflection (λ_2) of a selenium MAD experiment. The datasets were collected at 100 K using an ADSC CCD detector. The MAD data were integrated and reduced using XDS and then scaled with the program XSCALE.[10] Data statistics are summarized in Table I.

Structure solution and refinement. The structure was determined with 2.53 Å selenium MAD data using the CCP4 suite[11] and SOLVE/RESOLVE.[12] Model completion and refinement were performed with dataset λ_1 using COOT[13] and REFMAC5.[14] Since the data are incomplete to 2.53 Å, we define the nominal resolution as 2.70 Å, which is the resolution of a dataset that is 100% complete and has the same number of reflections as observed in that dataset. Nevertheless, the 2890 reflections observed between 2.70 and 2.53 Å (52.8% complete for this shell) are included in the refinement. Refinement statistics are summarized in Table I.

Validation and deposition. Analysis of the stereochemical quality of the model was accomplished using AutoDepInputTool,[15] MolProbity,[16] SFcheck 4.0,[17] and WHATIF 5.0.[18] Protein quaternary structure analysis was performed using the PQS server.[19] Figures were prepared with PyMOL (DeLano Scientific). Atomic coordinates and experimental structure factors for TM1585 at 2.70 Å resolution have been deposited in the PDB and are accessible under the code 2b8n.
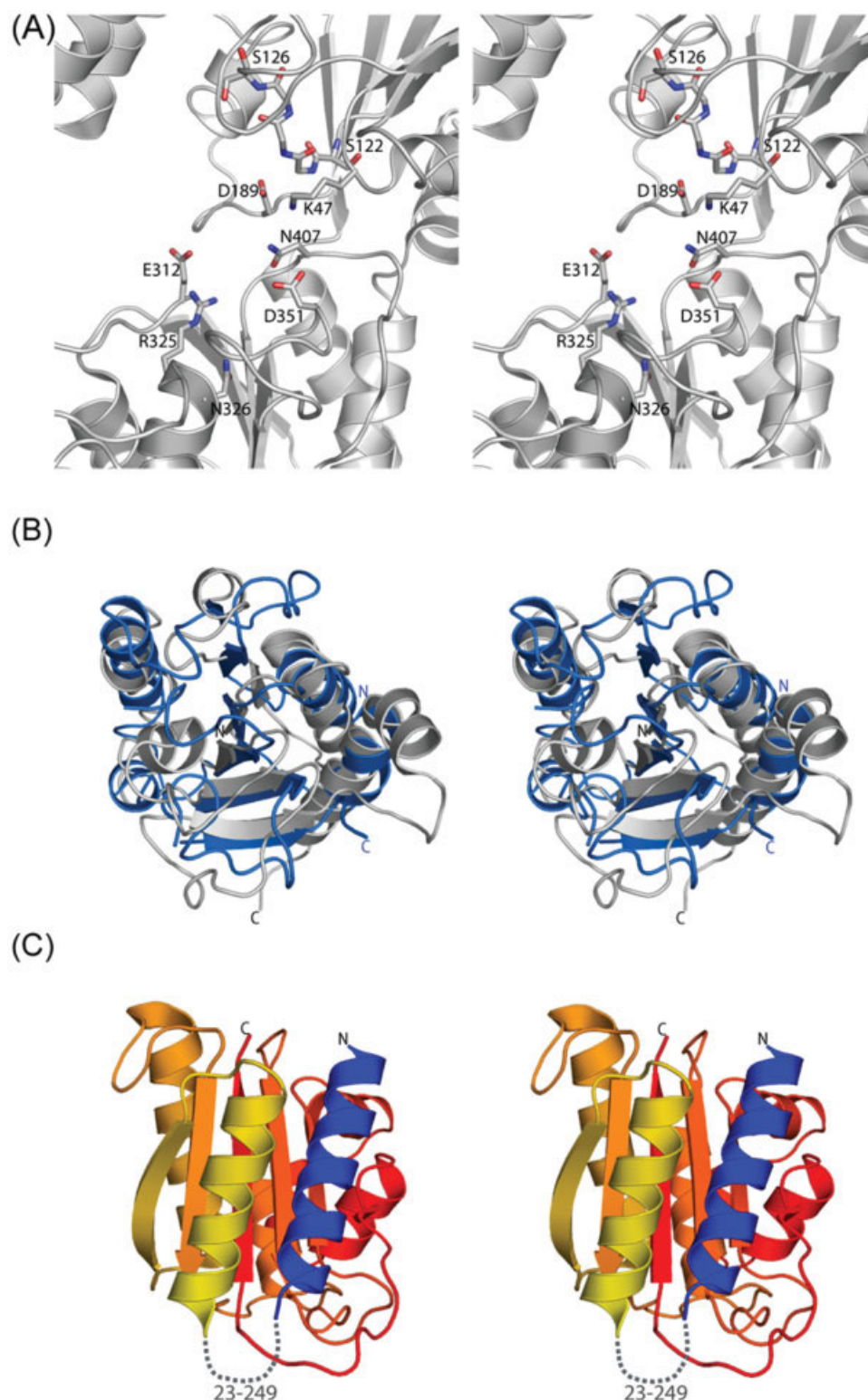
Fig. 2.   **A:** The putative active site of TM1585 is located in the cleft between the N-terminal Rossmann-like and C-terminal domains. The highly conserved residues from the Rossmann-like domain (Lys47, Asp189, and the glycine-rich loop 122-SGGGS-126) and the C-terminal domain (Glu312, Arg325, Asn326, Asp351, and Asn407) are shown in ball-and-stick representation. Atoms are colored as follows: carbon (grey), oxygen (red), and nitrogen (blue). **B:** Stereo diagram of a structural alignment of the Rossmann fold domains of TM1585 (residues 23–249, blue) and glycerate kinase from *Neisseria meningitidis* (residues 1–55 and 261–368, grey). **C:** Stereo diagram of the TM1585 C-terminal domain (residues 4–22 and 250–417), which belongs to a new fold.

Results and Discussion. The crystal structure of TM1585 [Fig. 1(B)] was determined to a nominal resolution of 2.70 Å using the MAD method. Data collection, model, and refinement statistics are summarized in Table I. The final model includes two TM1585 monomers (residues 4–417 for chains A and B) and 97 water molecules in the asymmetric unit. No electron density was observed for residues 1–3 or the expression and purification tag. The Matthews' coefficient $(V_m)^{20}$ for TM1585 is 2.6 $Å^3$/Da, and the estimated solvent content is 52.4%. The Ramachandran plot produced by MolProbity[16] shows that 98.2% and 100.0% of the residues are in favored and allowed regions, respectively. TM1585 is composed of 12 β-strands (β1–β12), 15 α-helices (H1–H7, H9–H16), and one 3_{10}-helix (H8) [Fig. 1(B and C)]. The total β-strand, α-helical, and 3_{10}-helical content is 18.6%, 39.9%, and 0.7%, respectively. The TM1585 monomer is comprised of two dissimilar α/β domains: an N- terminal, Rossmann-like domain (residues 23–249) and a C-terminal domain that adopts a new fold (residues 4–22 and 250–417). The N-terminal domain has a central, six-stranded, parallel β-sheet with a strand order of 654123, as well as an all-helical subdomain that is formed from the packing of helices H5–H6 and H9–H10. The C-terminal domain contains a six-stranded, mixed β-sheet with strand order 126345 and seven helices packed on both sides of the β-sheet.

The active site is likely to be in the cleft between the N- and C-terminal α/β domains. Highly conserved amino acids with side-chains pointing into the putative active site include seven residues from the Rossmann-like domain (Lys47, Asp189, and the glycine-rich loop 122-SGGGS-126) and five residues from the C-terminal domain (Glu312, Arg325, Asn326, Asp351, and Asn407) [Fig. 2(A)]. Each of these domains contains one highly conserved basic residue (Lys47 in the N-terminal Rossmann-like domain and Arg325 in the C-terminal domain) that could potentially interact with the triphosphate tail of ATP. However, consistent with the nucleotide-binding function of other canonical Rossmann-fold proteins, it is likely that the N-terminal domain binds ATP. The C-terminal domain is possibly involved in substrate binding with the active site located in the cleft between these domains.

Analysis of the crystallographic packing of TM1585 using the PQS server[19] indicates a monomer [Fig. 1(B)]. This finding is inconsistent with results from analytical size exclusion chromatography in combination with static light scattering, which suggest a dimer. However, the monomer contains a catalytically competent active site, which suggests that the enzyme can function as a monomer.

A DALI[21] structural similarity search using the N-terminal domain of TM1585 (residues 23–249) as a query found remote structural similarity to a number of Rossmann-fold domains. The top DALI hit is to the structure of glutamyl-tRNA reductase from *Methanopyrus kandleri* (PDB: 1gpj),[22] which aligns with a root-mean-square deviation (RMSD) of 3.2 Å and a sequence identity of 12% over 114 C_α atoms. Of particular interest is the similarity to the N-terminal domain (residues 1–55 and 261–368) of glycer-

ate kinase from *Neisseria meningitidis* (PDB: 1to6, Midwest Structural Genomics Center), which belongs to the other glycerate kinase family.[3] Both domains can be structurally aligned with an RMSD of 3.4 Å and a sequence identity of 16% over 101 C_α atoms [Fig. 2(B)]. Although these domains have a similar overall fold, the putative active site residues are not conserved between these two families of glycerate kinases. A DALI [21] search using the C-terminal domain of TM1585 as a query failed to find significant similarities to proteins in which all elements of the glycerate kinase C-terminal domain are present, indicating that this domain is a new fold. Remote structural similarity to cobalt precorrin-4-methyltransferase CbiF domain 2 (PDB: 1cbf) [23] was found by ProSMoS (Grishin NV, unpublished).[24] Both domains display a central, mixed β-sheet of five β-strands (order 12534) surrounded by α-helices. However, in TM1585, the domain includes an extra α-helix at the N-terminus and an extra α/β insertion between the last two β-strands that attaches to the edge of the five β-strands. Although the topological arrangement of the secondary structural elements of these domains are similar, the packing of the connecting helices around the β-sheet differ significantly. Hence, we classify this domain as a new fold in agreement with the SCOP database [Fig. 2(C)].

TM1585 contains homologous sequences in archaea, bacteria, and eukaryotes. Models for TM1585 homologs can be accessed at http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=TM1585. The TM1585 crystal structure reported here reveals a novel fold and is the first structural representative of the glycerate-2-kinase family. Recent biochemical studies (A. Osterman et al., personal communication) have shown that TM1585 has a glycerate-2-kinase activity for TM1585. Mutagenesis studies of active site residues and further crystallographic studies are required to elucidate the mechanism of enzyme function. The information presented here could form the starting point for the rational design of such experiments.

NOTE ADDED IN PROOF

Recent work by Reher et al. (FEMS Microbiol Lett **259** (2006) 113–119), where they propose a 2-phosophoglycerate kinase activity for a related enzyme from *Picrophilus torridus*, lends further support to our assertion that TM1585 is a glycerate-2-kinase.

REFERENCES

1. Ichihara A, Greenberg DM. Studies on the purification and properties of D-glyceric acid kinase of liver. J Biol Chem 1957;225: 949–958.
2. Doughty CC, Hayashi JA, Guenther HL. Purification and properties of D-glycerate 3-kinase from *Escherichia coli*. J Biol Chem 1966;241:568–572.
3. Cheek S, Ginalski K, Zhang H, Grishin NV. A comprehensive update of the sequence and structure classification of kinases. BMC Struct Biol 2005;5:6.
4. Chistoserdova L, Lidstrom ME. Identification and mutation of a gene required for glycerate kinase activity from a facultative methylotroph, *Methylobacterium extorquens AM1*. J Bacteriol 1997;179:4946–4948.
5. Crouzet P, Otten L. Sequence and mutational analysis of a tartrate utilization operon from *Agrobacterium vitis*. J Bacteriol 1995;177:6518–6526.
6. Bonham JR, Stephenson TJ, Carpenter KH, Rattenbury JM, Cromby CH, Pollitt RJ, Hull D. D(+)-glyceric aciduria: etiology and clinical consequences. Pediatr Res 1990;28:38–41.
7. Duran M, Beemer FA, Bruinvis L, Ketting D, Wadman SK. D-glyceric acidemia: an inborn error associated with fructose metabolism. Pediatr Res 1987;21:502–506.
8. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 2002;99:11664–11669.
9. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, et al. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr 2002;35:278–281.
10. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst 1993;26:795–800.
11. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
12. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55:849–861.
13. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132.
14. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 2003;374:300–321.
15. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2004;60:1833–1839.
16. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res 2004;32: W615–619.
17. Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr 1999;55:191–205.
18. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56, 29.
19. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23:358–361.
20. Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33:491–497.
21. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
22. Moser J, Schubert WD, Beier V, Bringemeier I, Jahn D, Heinz DW. V-shaped structure of glutamyl-tRNA reductase, the first enzyme of tRNA-dependent tetrapyrrole biosynthesis. Embo J 2001;20:6583–6590.
23. Schubert HL, Wilson KS, Raux E, Woodcock SC, Warren MJ. The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. Nat Struct Biol 1998;5:585–592.
24. Kinch LN, Qi Y, Hubbard TJ, Grishin NV. CASP5 target classification. Proteins 2003;Suppl 6:340–351.
25. Tickle IJ, Laskowski RA, Moss DS. Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of gammaB- and betaB2-crystallin. Acta Crystallogr D Biol Crystallogr 1998;54:243–252.