PROTEINS: Structure, Function, and Bioinformatics 54:174–177 (2004)

STRUCTURE NOTE

Crystal Structure of an Iron-Containing 1,3-Propanediol Dehydrogenase (TM0920) From *Thermotoga maritima* at 1.3 Å Resolution

Robert Schwarzenbacher,[1,4] Frank von Delft,[1,6] Jaume M. Canaves,[1,4] Linda S. Brinen,[1,2] Xiaoping Dai,[1,6] Ashley M. Deacon,[1,2] Marc A. Elsliger,[1,6] Said Eshaghi,[1,3] Ross Floyd,[1,2] Adam Godzik,[1,4] Carina Grittini,[1,6] Slawomir K. Grzechnik,[1,4] Chittibabu Guda,[1,4] Lukasz Jaroszewski,[1,4] Cathy Karlak,[1,6] Heath E. Klock,[1,3] Eric Koesema,[1,3] John S. Kovarik,[1,2] Andreas Kreusch,[1,3] Peter Kuhn,[1,2] Scott A. Lesley,[1,3] Daniel McMullan,[1,3] Timothy M. McPhillips,[1,2] Mark A. Miller,[1,4] Mitchell D. Miller,[1,2] Andrew Morse,[1,4] Kin Moy,[1,6] Jie Ouyang,[1,4] Rebecca Page,[1,6] Alyssa Robb,[1,3] Kevin Rodrigues,[1,3] Thomas L. Selby,[1,6] Glen Spraggon,[1,3] Raymond C. Stevens,[1,6] Henry van den Bedem,[1,2] Jeff Velasquez,[1,6] Juli Vincent,[1,3] Xianhong Wang,[1,4] Bill West,[1,4] Guenter Wolf,[1,2] Keith O. Hodgson,[1,2] John Wooley,[1,4,5] and Ian A. Wilson[1,6]*

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

Introduction: The TM0920 gene of *Thermotoga maritima* encodes a predicted iron-containing 1,3-propanediol dehydrogenase (EC 1.1.1.202) with a molecular weight of 39,944 Da and a predicted isoelectric point of 6.1. 1,3-propanediol dehydrogenase catalyzes the oxidation of propane-1,3-diol to 3-hydroxypropanal with the simultaneous reduction of $NADP^+$ to NADPH. Another four alcohol dehydrogenase paralogues of TM0920 have been identified in the *Thermotoga* proteome (TM0111, TM0285, TM0423, and TM0820), as well as hundreds of orthologues in the three Domains of Life. Here, we report the crystal structure of TM0920 determined using the semiautomated high-throughput pipeline of the Joint Center for Structural Genomics.[1] The crystal structure is consistent with its predicted function.

The crystal structure of TM0920 (Fig. 1A) was determined to 1.3 Å resolution using the multiple-wavelength anomalous dispersion (MAD) method. Data collection and refinement statistics are summarized in Table I. The final model includes two protein molecules (residues 1–359), two nicotinamide-adenine dinucleotide phosphate molecules ($NADP^+$), two Fe^{2+} ions, and 895 water molecules. The Matthews' coefficient (V_m) for TM0920 is 2.15 $Å^3$/Da and the estimated solvent content is 42.8%. The Ramachandran plot produced by PROCHECK 3.4[2] shows that 92.8% of the residues are in the most favored regions, 7.1% in additional allowed regions and 0.2% in generously allowed regions. No residues lie in disallowed regions.

The TM0920 monomer consists of a single polypeptide chain of 359 residues, composed of 19 helices (13 α-helices, and 6 3_{10}-helices), and 8 β-strands (Fig. 1B). The total α-helix, 3_{10}-helix, and β-strand content is 48.7, 7.2, and 12.3%, respectively. Each monomer has two distinct domains separated by a deep cleft. The α/β N-terminal domain A (residues 1 to 177) consists of 8 β-strands (β1 to β8) and 9 helices (H1 to H9), arranged as a 6-stranded parallel β-sheet (β1 to β5, and β8, with a topology −4X −1X 2X 1X 1X) flanked by α-helices in a Rossmann-fold topology, and two β-strands (β6 and β7) forming a β-hairpin. The C-terminal domain B (residues 178 to 359) comprises helices H10 to H19 organized in two helical bundles. The TM0920 structure represents a dehydroquinate synthase-like fold[7] and is the third member determined for the glycerol dehydrogenase-like structural family (together with 1KQ3 and 1JQA). Analogy with other dehydrogenases and analysis of the crystal packing suggest that a monomer is the biologically-relevant form.

A BLAST search of the TM0920 protein versus the Swissprot database indicated that the highest level of sequence identity (32% identity, 47% considering conserved and partially conserved residues) corresponded to proteins from *Klebsiella pneumoniae* (Swissprot: Q59477) and *Citrobacter freundii* (Swissprot: P44513). Accordingly, TM0920 was annotated as an iron-dependent 1,3-propanediol dehydrogenase. Biochemical data for these two

Grant sponsor: National Institutes of Health, Protein Structure Initiative: Grant number: P50-GM 62411.

*Correspondence to: Dr. Ian Wilson, JCSG, The Scripps Research Institute, BCC206, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail: wilson@scripps.edu

Received 22 July 2003; Accepted 23 July 2003

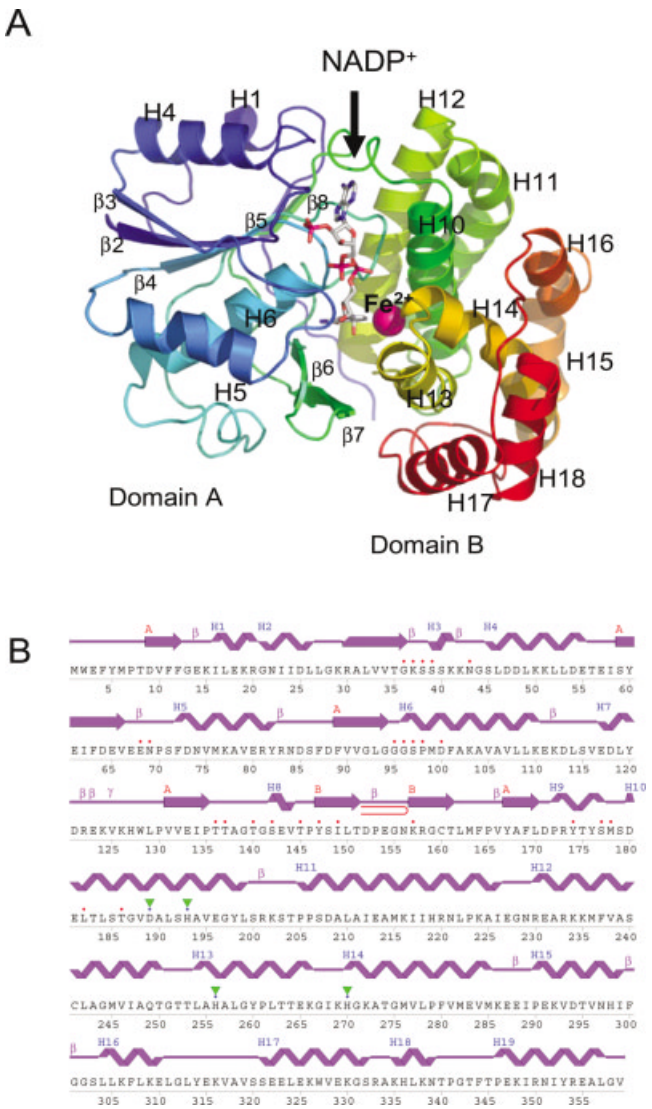

Fig. 1. Crystal structure of TM0920. **A:** Ribbon diagram of *Thermotoga maritima* TM0920 1,3-propanediol dehydrogenase color coded from N-terminus (blue) to C-terminus (red) showing the domain organization and location of the active site (arrow). α-helices (H1–H19) and β-strands (β1–β8) are indicated. The $NADP^+$ molecule is shown in ball and stick. The Fe^{2+} ion is shown as a sphere in magenta. **B:** Diagram showing the secondary structure elements in TM0920 superimposed on its primary sequence. Residues located in the active site and interacting with Fe^{2+} are indicated by blue dots and green triangles. Residues interacting with $NADP^+$ are indicated by red dots. The location of the β-hairpin formed by β-strands 6 and 7 is depicted in red.

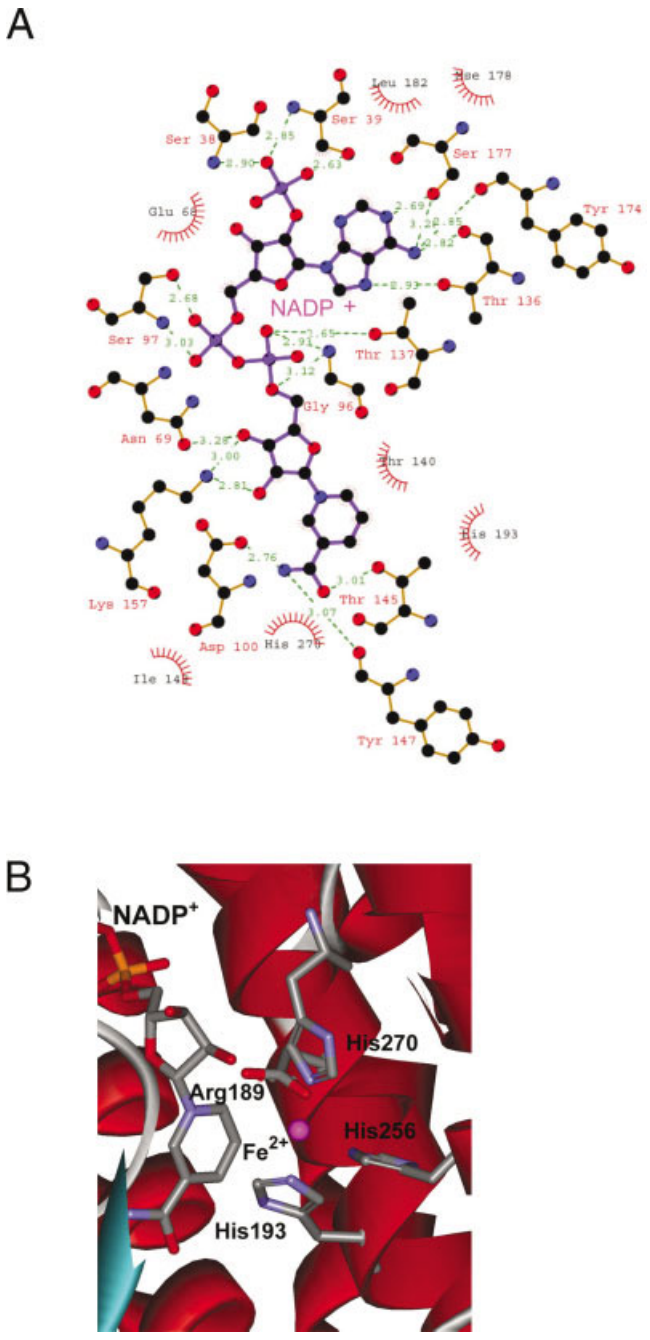

Fig. 2. **A:** Schematic representation of the interactions between $NADP^+$ and its interacting residues. The $NADP^+$ molecule is depicted in purple and the protein residues are in orange. The atoms are indicated as follows: carbon (black), oxygen (red), nitrogen (blue), and phosphorus (purple). Hydrogen bonds are represented as dashed green lines. Residues implicated in hydrophobic interaction are represented as barbed circle sections. **B:** Close-up view of the active site showing residues coordinating the metal ion.

proteins[3,4] were used as supporting evidence to model density in the active site as an Fe^{2+} ion. The Fe^{2+} ion is deeply located in the catalytic cleft and has a square pyramidal coordination with Asp189, His193, His256, and His270. All of these residues are situated in domain B (Fig. 2B). The $NADP^+$ binding site is located between domains A and B. The interactions between $NADP^+$ and the surrounding residues in the active site are depicted in Figure 2A. Interestingly, the density for the fully refined model is weak for the nicotinamide ring and shows residual difference density peaks near the nicotinamide NC5 atom. In addition, the side-chain of the metal coordinating His256 has a dual conformation and the Fe^{2+} ion, refined with SHELX,[16] appears to be only partially (60%) occupied. These findings indicate a possible superposition of states in the active site, which could not be completely accounted for with the current model.

A structural similarity search, performed using the

TABLE I. Summary of Crystal Parameters, Data Collection, and Refinement Statistics for TM0920 (PDB: 102D)

Space group	$P2_1$		
Unit cell parameters	a = 58.10 Å, b = 85.40 Å, c = 72.20 Å, $\alpha = \gamma = 90°$, $\beta = 96.2°$		
Data collection	λ_0	λ_1MADSe	λ_2MADSe
Wavelength (Å)	0.9184	0.9794	0.9184
Resolution range (Å)	72.0–1.30	30.0–2.60	30.0–2.60
Number of observations	542,582	157,092	157,648
Number of unique reflections	167,354	21,628	21,635
Completeness (%)	97.9	99.6	99.6
(In highest resolution shell, %)	99.1	99.6	99.6
Mean I/σ(I)	5.6	12.8	12.3
(In highest resolution shell)	1.2	9.4	10.0
R_{sym} on I[a]	0.076	0.042	0.043
(In highest resolution shell)	0.585	0.069	0.063
Sigma cutoff	0.0	0.0	0.0
Highest resolution shell (Å)	1.37–1.30	2.74–2.60	2.74–2.60
Model and refinement statistics			
Resolution range (Å)	72.0–1.30	Data set used in refinement	λ_0
Number of reflections (total)	158,936	Cutoff criteria	$\|F\| > 0$
Number of reflections (test)	8,371	R_{cryst}[b]	0.137
Completeness (% total)	97.8	R_{free}[c]	0.170
Stereochemical parameters			
Restraints (RMS observed)			
Bond length		0.015 Å	
Bond angle		1.53°	
Average isotropic B-value		10.9 $Å^2$	
ESU-based on R value		0.05 Å	

[a] $R_{sym} = \Sigma|I_i - \langle I_i \rangle| / \Sigma|I_i|$ where I_i is the scaled intensity of the ith measurement, and $\langle I_i \rangle$ is the mean intensity for that reflection.
[b] $R_{cryst} = \Sigma| |F_{obs}| - |F_{calc}| | / \Sigma|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[c] R_{free} as for R_{cryst}, but for 5% of the total reflections chosen at random.
[d] ESU = Estimated overall coordinate error.[12,17]

DALI server[5] indicated that TM0920 is structurally similar to a glycerol dehydrogenase from *Bacillus stearothermophilus* (PDB: 1JQA),[6] with an RMSD value of 3.0 Å for the superimposition of 321 Cα atoms with 20% sequence identity. A structural alignment between TM0920 and TM0423 (PDB:1KQ3), another alcohol dehydrogenase from *Thermotoga maritima* whose structure has also been determined by the Joint Center for Structural Genomics,[8] has an RMSD value of 3.1 Å for the superimposition of 322 Cα atoms with 21% sequence identity. According to the Structural Classification of Proteins database (SCOP),[7] the TM0920 structure has a dehydroquinate synthase-like fold, and represents the third member of the glycerol dehydrogenase-like structural family (together with 1KQ3 and 1JQA). Models for a number of TM0920 orthologues and paralogues can be accessed at http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=TM0920

The TM0920 structure represents an iron-containing 1,3-propanediol dehydrogenase in complex with $NADP^+$. We expect that the information reported here will yield valuable insights into the determinants for catalysis, substrate specificity, and thermal stability of this family of alcohol dehydrogenases.

Materials and Methods: *Protein production:* An iron-containing 1,3-propanediol dehydrogenase (TIGR: TM0920; Swissprot: Q9X022) was PCR amplified using Pfu (Stratagene) from *Thermotoga maritima* strain MSB8 genomic DNA using primer pairs encoding the predicted 5′- and 3′-ends of TM0920. The PCR product was cloned into plasmid pMH1, which encodes a purification tag consisting of the residues MGSDKIHHHHHH at the amino terminus of the full-length protein. The cloning junctions were confirmed by sequencing. Protein expression was performed in selenomethionine-containing medium using the *Escherichia coli* methionine auxotrophic strain DL41. Bacteria were lysed by sonication after a freeze-thaw procedure in Lysis Buffer (50mM Tris pH 7.9, 50mM NaCl, 1mM MgCl2, 0.25mM Tri (2-carboxyethyl) phosphine hydrochloride (TCEP), 1mg/ml lysozyme) and cell debris pelleted by centrifugation at 3600× g for 60 minutes. The soluble fraction was applied to a nickel chelate resin (Pharmacia) previously equilibrated with Equilibration Buffer (50mM KH_2PO_4 pH 7.8, 0.25mM TCEP, 10% v/v glycerol, 0.3M NaCl) containing 20 mM imidazole. The resin was washed with Equilibration Buffer containing 40 mM imidazole, and protein was eluted with Elution Buffer (20 mM Tris pH 7.9, 10% v/v glycerol, 0.25 mM TCEP, 300 mM imidazole). Buffer exchange was performed to remove imidazole from the protein eluate and the protein in Buffer Q (20 mM Tris pH 7.9, 50 mM NaCl, 5% v/v glycerol, 0.25 mM TCEP) was then applied to a Resource Q column (Pharmacia). Protein was eluted using a linear gradient to 500 mM NaCl. Appropriate fractions were buffer exchanged into size exclusion chromatography (SEC) Buffer (20 mM Tris pH 7.9, 150 mM NaCl, 0.25 mM TCEP). The protein was concentrated for crystallization assays by centrifugal ultrafiltration (Millipore). The protein was crystallized using the nanodroplet vapor diffusion method[9] with standard JCSG crystallization protocols.[1] The crystallization buffer con-

tained 20% PEG-300, 5% (w/v) PEG-8000, 10% glycerol, and 0.1 M Tris-HCl at pH 8.5. The crystals were indexed in the monoclinic space group $P2_1$ (Table I).

Data collection: Native and multi-wavelength anomalous diffraction data were collected at Stanford Synchrotron Radiation Laboratory (SSRL, Stanford, USA) on beamline 11-1 at multiple wavelengths using the BLU-ICE[10] data collection environment (Table I). All data sets were collected at 100 K using a Quantum 315 CCD detector. Data were integrated and reduced using Mosflm[11] and then scaled with the program SCALA from the CCP4 suite.[12] Data statistics are summarized in Table I.

Structure solution and refinement: The structure was determined using the software packages SnB,[13] the CCP4 suite,[12] and SOLVE.[14] An initial model was built using the ARP/wARP package.[15] The structure was refined to a resolution of 1.3 Å using REFMAC5[12] and SHELX.[16] Refinement statistics are summarized in Table I. The final model contains two protein molecules (residues 1 to 359), two $NADP^+$ molecules, two Fe^{2+} ions, two TRIS (2-amino-2-hydroxymethyl-propane-1,3-diol) molecules and 895 water molecules in the asymmetric unit. The model for chain B contains two residues, His −1 and 0, from the purification tag.

Validation and deposition: Analysis of the stereochemical quality of the models was accomplished using the JCSG Validation Central suite, which integrates seven validation tools: Procheck 3.5.4, SFcheck 4.0, Prove 2.5.1, ERRAT, WASP, DDQ 2.0, and Whatcheck. The Validation Central suite is accessible at http://www.jcsg.org. Atomic coordinates of the final model and experimental structure factors of TM920 have been deposited in the PDB and are accessible under the code 1O2D.

Acknowledgments: This work was supported by NIH Protein Structure Initiative grant P50-GM 62411 from the National Institute of General Medical Sciences (www.nigms.nih.gov). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a National user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences). The JCSG is grateful for constructive comments from Christian Cambillau.

REFERENCES

1. Lesley SA, Kuhn P, Godzik A, Deacon, AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Wooley J, Taylor SS, Wilson IA, Schultz PG, Stevens RC. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 2002;99:11664–11669.
2. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
3. Daniel R, Boenigk R, Gottschalk G. Purification of 1,3-propanediol dehydrogenase from *Citrobacter freundii* and cloning, sequencing, and overexpression of the corresponding gene in *Escherichia coli*. J Bacteriol 1995;177:2151–2156.
4. Johnson EA, Lin CC. *Klebsiella pneumoniae* 1,3 propanediol:NAD^+-oxidoreductase. J Bacteriol 1987;169:2050–2054.
5. Holm L, Sander C. Dali: A network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
6. Wilkinson KW, Baker PJ, Rice DW, Stillman TJ, Gore MG, Krauss O, Atkinson T. Crystallization of glycerol dehydrogenase from *Bacillus stearothermophilus*. Acta Crystallogr 1995;D51:830–832.
7. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
8. Brinen LS, Canaves JM, Dai X, Deacon AM, Elsliger MA, Eshaghi S, Floyd R, Godzik A, Grittini C, Grzechnik SK, Guda C, Jaroszewski L, Karlak C, Klock HE, Koesema E, Kovarik JS, Kreusch A, Kuhn P, Lesley SA, McMullan D, McPhillips TM, Miller MA, Miller MD, Morse A, Moy K, Ouyang J, Robb A, Rodrigues K, Selby TL, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, Wang X, West B, Wolf G, Taylor SS, Hodgson KO, Wooley J, Wilson IA. Crystal structure of a zinc-containing glycerol dehydrogenase (TM0423) from *Thermotoga maritima* at 1.5 Å resolution. Proteins 2003;50:371–374.
9. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens RC. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr 2002;35:278–281.
10. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat 2002;9:401–406.
11. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography 1992;26.
12. Collaborative Computational Project Number 4. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr 1994;D50:760–763.
13. Weeks CM, Miller R. The design and implementation of *SnB* v2.0. J Appl Crystallogr 1999;32:120–124.
14. Terwilliger TC, Berendzen J. Automated structure solution for MIR and MAD. Acta Crystallogr 1999;D55:849–861.
15. Perrakis A, Morris RM, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463
16. Sheldrick G, Schneider T, SHELXL: high-resolution refinement. Methods Enzymol 1997;277:319–343.
17. 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 1998;D54:243–252.