PROTEIN STRUCTURE REPORT

Crystal structures of possible lysine decarboxylases from *Thermus thermophilus* HB8

MUTSUKO KUKIMOTO-NIINO,[1] KAZUTAKA MURAYAMA,[1]
MIYUKI KATO-MURAYAMA,[1] MIKI IDAKA,[1] YOSHITAKA BESSHO,[1]
AYAKO TATSUGUCHI,[1] RYOKO USHIKOSHI-NAKAYAMA,[1] TAKAHO TERADA,[1,2]
SEIKI KURAMITSU,[2,3] MIKAKO SHIROUZU,[1,2] AND SHIGEYUKI YOKOYAMA[1,2,4]

[1]RIKEN Genomic Sciences Center, Tsurumi, Yokohama 230-0045, Japan
[2]RIKEN Harima Institute at SPring-8, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan
[3]Graduate School of Science, Osaka University, Toyonaka, Osaka 560–0043, Japan
[4]Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

(RECEIVED August 3, 2004; FINAL REVISION August 3, 2004; ACCEPTED August 6, 2004)

Abstract

TT1887 and TT1465 from *Thermus thermophilus* HB8 are conserved hypothetical proteins, and are annotated as possible lysine decarboxylases in the Pfam database. Here we report the crystal structures of TT1887 and TT1465 at 1.8 Å and 2.2 Å resolutions, respectively, as determined by the multiwavelength anomalous dispersion (MAD) method. TT1887 is a homotetramer, while TT1465 is a homohexamer in the crystal and in solution. The structures of the TT1887 and TT1465 monomers contain single domains with the Rossmann fold, comprising six α helices and seven β strands, and are quite similar to each other. The major structural differences exist in the N terminus of TT1465, where there are two additional α helices. A comparison of the structures revealed the elements that are responsible for the different oligomerization modes. The distributions of the electrostatic potential on the solvent-accessible surfaces suggested putative active sites.

Keywords: structural genomics; *Thermus thermophilus*; hypothetical protein; lysine decarboxylase; Rossmann fold

TT1887 from *Thermus thermophilus* HB8 is a conserved hypothetical protein, which consists of 171 amino acid residues (18.5 kDa). It is annotated as a possible lysine decarboxylase in the Pfam database (PF03641) (Bateman et al. 2002), and shows high sequence identity (35%) to TT1465 from *T. thermophilus* HB8, which consists of 217 amino acid residues (24.3 kDa) (Fig. 1A). They share the sequence motif PGGxGTxxE, which is highly conserved among 140 predicted bacterial and yeast proteins with unknown function, including proteins annotated as lysine decarboxylases. The crystal structures of two members of this family, the hypothetical proteins TM1055 from *Thermotoga maritima* (PDB: 1RCU) and Yvdd from *Bacillus subtilis* (PDB: 1T35), were recently solved at 2.5 Å and 2.72 Å resolutions, respectively. In the *T. maritima* TM1055 structure, there are four monomers in the asymmetric unit, while there are eight in the *B. subtilis* Yvdd structure. The monomeric structures of the two proteins share the similar α/β topology of the Rossmann fold.

We now report the crystal structures of two putative lysine decarboxylases, TT1887 and TT1465 from *T. thermophilus* HB8, at 1.8 Å and 2.2 Å resolutions, respectively. The structures were determined by the multiwavelength anomalous dispersion (MAD) method. The structures of the TT1887 and TT1465 monomers share the Rossmann fold, and are quite similar to those of the *T. maritima* TM1055

Reprint requests to: Shigeyuki Yokoyama, RIKEN Genomic Sciences Center, 1–7–22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan; e-mail: yokoyama@biochem.s.u-tokyo.ac.jp; fax: +81-45-503-9195.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.041012404.

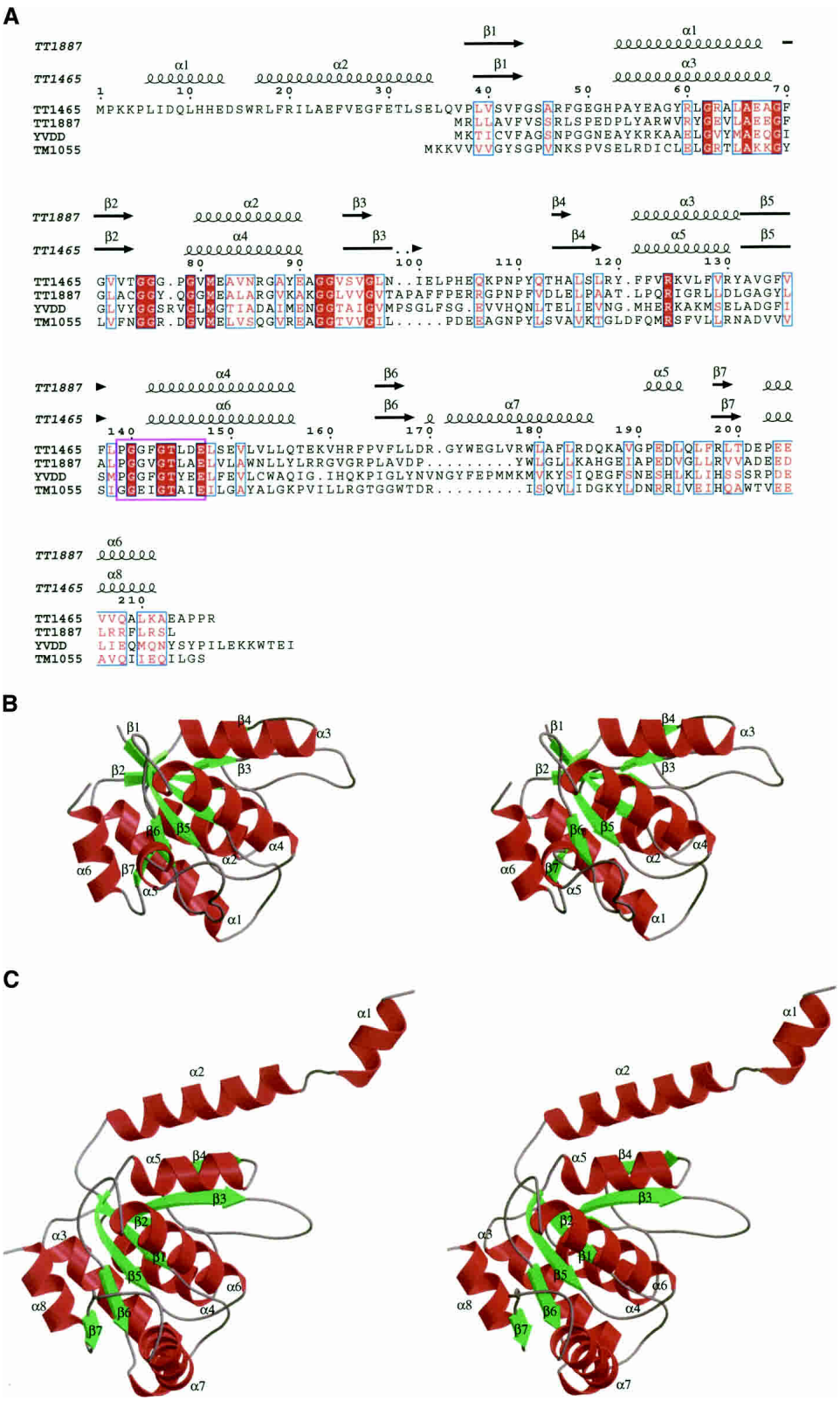

Figure 1. (*A*) Sequence alignment of lysine decarboxylase family proteins. TT1465, TT1465 from *T. thermophilus* HB8; TT1887,TT1887 from *T. thermophilus* HB8; YVDD, Yvdd from *B. subtilis*; TM1055; TM1055 from *T. maritima*. The highly conserved motifs (PGGxGTxxE) are indicated in a pink box. The secondary structures for TT1465 and TT1887 are shown *above* the sequences. (*B*) Ribbon representation of the TT1887 monomer (stereo view). The α-helices are shown in red, and the β-strands are in green. (*C*) Same view of the TT1465 monomer (stereo view).

and *B. subtilis* Yvdd monomers. However, their quaternary structures differ remarkably: TT1887 is a homotetramer, while TT1465 is a homohexamer. A structural comparison of these four members of this protein family, and the locations of putative active sites will be discussed.

Results and Discussion

The crystals of TT1887 belong to the C-centered orthorhombic space group $C222_1$, with unit cell constants of $a = 40.66$ Å, $b = 129.82$ Å, $c = 119.85$ Å, and contain two protein molecules per asymmetric unit. The structure of TT1887 was refined to 1.8 Å by the MAD method. The crystallographic data are summarized in Table 1. The final model comprises 342 amino acid residues (two protein molecules) and 403 water molecules in the asymmetric unit. The TT1887 monomer consists of a single domain composed of six α helices flanked by a seven-stranded β sheet, which is characteristic of the Rossmann fold (Fig. 1B). A DALI homology search revealed that TT1887 resembles nucleoside 2-deoxyribosyltransferase (PDB: 1F8X, root mean square deviation [RMSD] 3.5 Å over 116 Cα atoms; Armstrong et al. 1996), the negative transcriptional regulator NmrA (PDB: 1K6I, RMSD 3.2 Å over 114 Cα atoms; Stammers et al. 2001), biliverdin IX beta reductase (PDB: 1HE2, RMSD 2.9 Å over 116 Cα atoms; Pereira et al. 2001) and other Rossmann fold-like proteins.

The crystals of TT1465 belong to the primitive orthorhombic space group $P2_12_12_1$, with unit cell constants of $a = 65.65$ Å, $b = 83.80$ Å, $c = 265.14$ Å, and contain six protein molecules per asymmetric unit. The structure of TT1465 was refined to 2.2 Å by the MAD method (Table 1). The final model includes 1240 amino acid residues (5–212 in monomer A, 6–212 in monomer B, 2–102 and 108–212 in monomer C, 3–102 and 108–215 in monomer D, 4–101 and 108–215 in monomer E, and 2–100 and 108–213 in monomer F), 12 phosphate ions, and 363 water molecules in the asymmetric unit. Various numbers of amino acid residues at the N and C termini of all six monomers (A–F) and in the β3–β4 loop of four monomers (C–F) were not visible in the electron density map, and were thus omitted from the final model. The overall structure of the TT1465 monomer is basically similar to that of the TT1887 monomer (RMSD 1.9 Å over 168 Cα atoms) (Fig. 1C). The notable structural differences exist in the N terminus, which has two additional α helices (α1 and α2), and in the α7

Table 1. *X-ray data collection, phasing and refinement statistics*

	TT1887			TT1465		
	Peak	Edge	Remote	Peak	Edge	Remote
Data collection						
Wavelength (Å)	0.9792	0.9794	0.9742	0.9790	0.9793	0.9740
Resolution (Å)	50–1.8	50–1.8	50–1.8	50–2.2	50–2.2	50–2.2
Unique reflections	29,939	29,918	29,832	71,983	67,828	72,513
Redundancy	9.9	9.9	10.1	4.7	4.2	4.7
Completeness (%)	99.9 (99.2)	99.9 (99.1)	99.9 (99.7)	95.5 (88.3)	90.0 (57.9)	96.2 (92.4)
I/σ(I)	25.4 (7.1)	28.9 (7.1)	29.1 (7.7)	12.3 (2.5)	12.5 (2.4)	13.0 (2.5)
*R*sym (%)[a]	8.6 (26.8)	8.0 (26.3)	10.7 (26.2)	11.4 (46.6)	9.7 (31.6)	10.7 (46.9)
MAD analysis						
Resolution (Å)		20–1.8			20–2.2	
Number of sites		2			6	
FOM_{MIR}[b]		0.55			0.38	
$FOM_{RESOLVE}$[c]		0.67			0.62	
Refinement						
Resolution (Å)		1.8			2.2	
No. of reflections		29,544			70,481	
No. of Protein atoms		2,612			9,832	
No. of phosphate ions					12	
No. of water molecules		403			363	
*R*work (%)		18.5			19.8	
*R*free (%)[d]		22.3			25.0	
RMSD bond length (Å)		0.007			0.008	
RMSD bond angles (°)		1.4			1.5	

All numbers in parentheses represent last outer shell statistics.
[a] $R_{sym} = \sum I_{avg} - Ii/\sum Ii$, where Ii is the observed intensity and I_{avg} is the average intensity.
[b] Figure of merit after SOLVE phasing.
[c] Figure of merit after RESOLVE.
[d] R_{free} is calculated for 10% of randomly selected reflections excluded from refinement.

helix. Large sequence insertions within these regions seem to cause these structural differences (Fig. 1A). The structure of the TT1465 monomer also resembles those of *T. maritima* TM1055 (PDB: 1RCU, RMSD 2.8 Å over 156 Cα atoms) and *B. subtilis* Yvdd (PDB: 1T35, RMSD 2.6 Å over 154 Cα atoms), two members of the lysine decarboxylase family with high sequence similarities to TT1887 and TT1465 (Fig. 1A).

Two TT1887 molecules (monomers A and B) are included in the asymmetric unit, with a buried surface area of 591 Å^2 (Fig. 2A). In addition, interactions with two other symmetry-related molecules (monomers A′ and B′) are observed with larger buried surface areas of ~1750 Å^2 per monomer. Since analytical ultracentrifugation of TT1887 revealed a molecular weight value corresponding to four TT1887 molecules (data not shown), TT1887 exists as a homotetramer both in the crystal and in solution. In the structure of *T. maritima* TM1055, four protein molecules with a similar subunit arrangement are also visible in the asymmetric unit. Therefore, it is likely that *T. maritima* TM1055 also forms a homotetramer in solution.

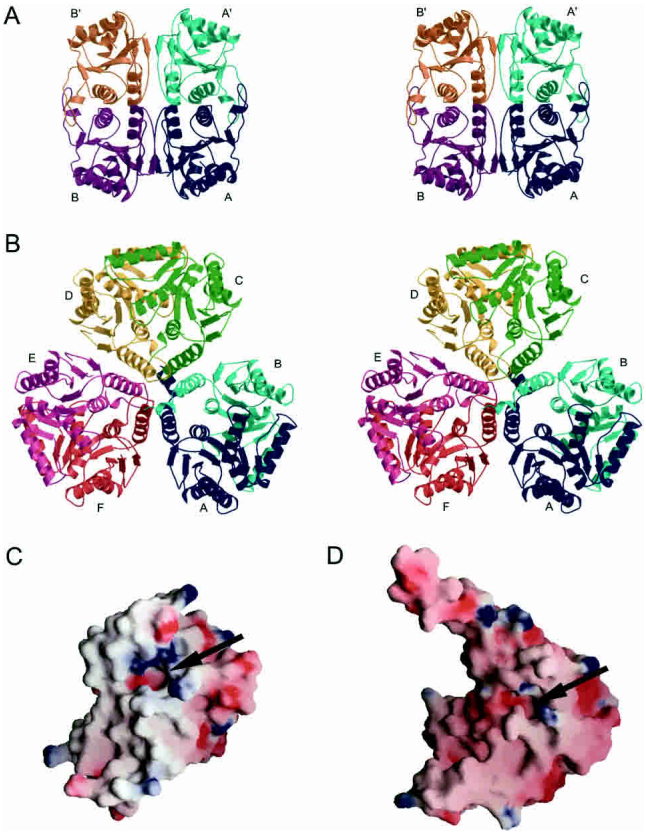

Figure 2. Ribbon representations of the TT1887 tetramer (*A*) and the TT1465 hexamer (*B*) (stereo view). Each monomer is colored differently. Electrostatic surface representations of the TT1887 monomer (*C*) and the TT1465 monomer (*D*). Blue and red surfaces represent positive and negative potentials, respectively. The locations of putative active sites are shown with arrows.

On the other hand, analytical ultracentrifugation revealed that TT1465 is a hexamer in solution (data not shown), which is consistent with the hexameric structure of TT1465 in the crystal (Fig. 2B). The two monomers (A and B) form a dimer with a buried surface area of ~2600 Å^2 per monomer. The subunit interactions between the AB dimer of TT1465 correspond to those between the AA′ dimer of TT1887 (Fig. 2A). Two other dimers (CD and EF) are related to the AB dimer by a noncrystallographic 3-fold axis. Each dimer interacts with the other two dimers with a buried surface area of 474 Å^2. The α2 helix, which is missing in TT1887, is involved in the trimer interactions. Therefore, the presence of the α2 helix seems to determine the different quaternary structures between TT1465 and TT1887. In the *B. subtilis* Yvdd structure, there are eight monomers in the asymmetric unit, which can be seen as a tetramer of the corresponding dimers (the AB dimer). The tetramer formation involves the α helix (residues 70–73) between the β3 and β4 strands in *B. subtilis* Yvdd, which is not present in either TT1887 or TT1465 (Fig. 1A).

The electrostatic potential distribution on the solvent-accessible surface of the TT1887 monomer shows the presence of a hydrophilic cavity (Fig. 2C). A similar hydrophilic cavity also exists in TT1465 (Fig. 2D). A sequence analysis revealed that three hydrophilic residues, Arg 124, Thr 144, and Glu 147 in the TT1465 sequence, are highly conserved among the protein family members (Fig. 1A). Because all three hydrophilic residues face this cavity, these residues may form the active site and act as catalytic residues.

Materials and methods

Protein expression and purification

The genes for TT1465 and TT1887 from *T. thermophilus* HB8 were cloned into pET-11b (Novagen). Selenomethionine (SeMet)-substituted TT1465 and TT1887 proteins were expressed in *Escherichia coli* B834 (DE3). The *E. coli* lysate was heated at 70°C for 15 min, and the proteins were purified by a series of HiTrap Phenyl, Resource Q and Superdex 75 column chromatography steps (Amersham Biosciences). The yields of purified TT1887 and TT1465 were 0.36 mg and 1.03 mg per 1 g wet cells, respectively.

Crystallization and data collection

The crystallization conditions were screened using a Crystal Screen kit (Hampton Research) by the hanging drop vapor diffusion method at 20°C. The crystals of TT1887 (0.84 mg/ml) were grown against a reservoir solution containing 20% PEG4000, 10% iso-propanol, and 0.1 M Na-Hepes (pH 7.5). Small crystals with a plate-like morphology (40 × 40 × 5 μm^3) were obtained after 2–3 days. They were further improved by the addition of 5 mM cadmium chloride, 5 mM sodium acetate (pH 4.6), and 1.5% PEG400, and reached a typical size of 200 × 100 × 50 μm^3 with sufficient quality for data collection. The crystals of TT1465 (1.0 mg/ml) were grown against a reservoir solution consisting of 1.0 M ammonium dihydrogen phosphate, 20 mM magnesium formate, and

0.05 M Tris-HCl (pH 8.5). Crystals with a rod-like morphology ($300 \times 100 \times 5$ μm^3) were obtained after 2 wk. Data collection was carried out at 100 K with 20% glycerol as a cryoprotectant. The MAD data were collected at three different wavelengths at BL26B1, SPring-8 (Harima), and were recorded on a MAR imaging plate. All diffraction data were processed with the HKL2000 program (Otwinowski and Minor 1997).

Structure determination and refinement

The program SOLVE (Terwilliger and Berendzen 1999) was used to locate the selenium sites and to calculate the phases, and RESOLVE was used for the density modification (Terwilliger 2001). Automatic tracing using Arp/wARP (Perrakis et al. 2001) was used to partially build the models, and the rest of the models were built and refined with the programs O (Jones et al. 1991) and CNS (Brunger et al. 1998). Refinement statistics are presented in Table 1. The quality of the model was inspected by the program PROCHECK (Laskowski et al. 1993). Structural similarities were calculated with DALI (Holm and Sander 1993). The solvent accessible surface areas were calculated with the program AREAIMOL (CCP4 1994). Graphic figures were created using the programs Molscript (Kraulis 1991) and Raster3D (Merritt and Murphy 1994). The molecular surface was created with the program GRASP (Nicholls et al. 1991). The atomic coordinates have been deposited in the Protein Data Bank, with the accession codes 1WEH for TT1887, and 1WEK for TT1465.

Acknowledgments

We thank Dr. Masaki Yamamoto for data collection at RIKEN beamline BL26B1 at SPring-8. We also thank Drs. Satoru Unzai (Yokohama City University) and David J. Scott (The University of Nottingham) for their help with analytical ultracentrifugation. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Armstrong, S.R., Cook, W.J., Short, S.A., and Ealick, S.E. 1996. Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site. *Structure* **4:** 97–107.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30:** 276–280.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., and Pannu, N.S. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Collaborative Computational Project, Number 4. 1994. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50:** 760–763.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.A* **47:** 110–118.

Kraulis, P.J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24:** 946–950.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26:** 283–291.

Merritt, E.A. and Murphy, M.E.P. 1994. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **50:** 869–873.

Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281–296.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307–326.

Pereira, P.J., Macedo-Ribeiro, S., Parraga, A., Perez-Luque, R., Cunningham, O., Darcy, K., Mantle, T.J., and Coll, M. 2001. Structure of human biliverdin IXβ reductase, an early fetal bilirubin IXβ producing enzyme. *Nat. Struct. Biol.* **8:** 215–220.

Perrakis, A., Harkiolaki, M., Wilson, K.S., and Lamzin, V.S. 2001. ARP/wARP and molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **57:** 1445–1450.

Stammers, D.K., Ren, J., Leslie, K., Nichols, C.E., Lamb, H.K., Cocklin, S., Dodds, A., and Hawkins, A.R. 2001. The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. *EMBO J.* **20:** 6619–6626.

Terwilliger T.C. 2001. Map-likelihood phasing. *Acta Crystallogr. D Biol. Crystallogr.* **57:** 1763–1775.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55:** 849–861.