Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of a conserved hypothetical protein, TTHA0849 from *Thermus thermophilus* HB8, at 2.4 Å resolution: a putative member of the StAR-related lipid-transfer (START) domain superfamily

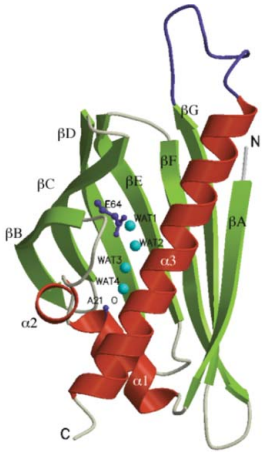
Makoto Nakabayashi,[a] Naoki Shibata,[a,b] Hirofumi Komori,[a,b] Yasufumi Ueda,[a] Hitoshi Iino,[b] Akio Ebihara,[b] Seiki Kuramitsu[b,c] and Yoshiki Higuchi[a,b]*

[a]Department of Life Science, Graduate School of Life Science, University of Hyogo and Himeji Institute of Technology, 3-2-1 Koto, Kamigori-cho, Ako-gun, Hyogo 678-1297, Japan, [b]RIKEN SPring-8 Center, 1-1-1 Koto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan, and [c]Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan

Correspondence e-mail: hig@sci.u-hyogo.ac.jp

Received 15 October 2005
Accepted 28 October 2005
Online 5 November 2005

PDB Reference: TTHA0849, 2d4r, r2d4rsf.



The crystal structure of a conserved hypothetical protein, TTHA0849 from *Thermus thermophilus* HB8, has been determined at 2.4 Å resolution as a part of a structural and functional genomics project on *T. thermophilus* HB8. The main-chain folding shows a compact $\alpha+\beta$ motif, forming a hydrophobic cavity in the molecule. A structural similarity search reveals that it resembles those steroidogenic acute regulatory proteins that contain the lipid-transfer (START) domain, even though TTHA0849 shows comparatively weak sequence identity to polyketide cyclases. However, the size of the ligand-binding cavity is distinctly smaller than other START domain-containing proteins, suggesting that it catalyses the transfer of smaller ligand molecules.

1. Introduction

TTHA0849 from *Thermus thermophilus* HB8 is a conserved hypothetical protein composed of 147 amino acids (M_r = 17 084) and exhibits weak sequence identity (about 20%) to polyketide (oligoketide) cyclases that bind multicyclic and/or aromatic compounds (Dickens *et al.*, 1995; Kendrew *et al.*, 1999; Madduri & Hutchinson, 1995; Niemi & Mäntsälä, 1995; Räty *et al.*, 2002; Torkkell *et al.*, 2000). Based on the sequence similarity between TTHA0849 and a polyketide cyclase, SnoaL from *Streptomyces nogalater*, the folding of the polypeptide backbone and the active-site environment in the two proteins are expected to be conserved (Sultana *et al.*, 2004). In order to corroborate or refute this assumption, we report the crystal structure determination at 2.4 Å resolution of TTHA0849 from *T. thermophilus* HB8. This structure determination has been carried out as part of a structural and functional genomics project on *T. thermophilus* HB8.

Unexpectedly, the crystal structure of TTHA0849 shows an overall fold resembling that of the steroidogenic acute regulatory (StAR) related lipid-transfer (START) domain primarily found in several eukaryotic proteins (Newton, 1995; Ponting & Aravind, 1999; Schultz *et al.*, 1998; Shaw, 1996; Stocco, 2000). This domain is believed to perform a diverse range of functions including signal transduction, transcriptional regulation, GTPase activation to thioester hydrolysis and allergenic induction. The START domain is a compact functional motif of the $\alpha+\beta$ type, consisting of a curved antiparallel β-sheet flanked by three or four α-helices (Gajhede *et al.*, 1996; Romanowski *et al.*, 2002; Shen *et al.*, 2005; Tsujishita & Hurley, 2000). Most START domain-containing proteins bind lipophilic molecules in a hydrophobic cavity located inside the molecule. In this paper, we describe the crystal structure of TTHA0849 and possible ligands are proposed based on the shape and the size of the cavity in comparison with other START domain-containing proteins.

2. Materials and methods

2.1. Preparation and crystallization

Escherichia coli B834(DE3) cells were transformed with expression vector pET11a carrying the TTHA0849 gene. The transformed *E. coli* was incubated for 4 h at 310 K in modified LeMaster medium (Hendrickson *et al.*, 1990), which contains 10 g l^{-1} lactose (a glucose substitute), vitamins (Kao and Michayluk Vitamin Solution; Sigma

Table 1
Summary of data-collection and refinement statistics.

Data set	Se peak	Se edge	Se remote
Data collection			
Wavelength (Å)	0.97911	0.97940	0.90000
Space group	$P2_12_12_1$		
Unit-cell parameters			
a (Å)	56.30		
b (Å)	86.49		
c (Å)	109.58		
Resolution range (Å)	46.28–2.40 (2.49–2.40)	46.34–2.40 (2.49–2.40)	43.37–2.40 (2.49–2.40)
No. of measured reflections	96534	95669	95002
No. of unique reflections	38812	38633	38701
Average redundancy	2.49 (2.55)	2.48 (2.54)	2.45 (2.53)
Completeness (%)	96.1 (99.0)	95.3 (98.8)	95.6 (99.0)
$R_{\rm merge}$	0.071 (0.182)	0.073 (0.233)	0.077 (0.230)
$\langle I/\sigma(I)\rangle$	7.7 (4.3)	7.0 (3.3)	6.5 (3.3)
Refinement statistics			
Resolution range (Å)	46.28–2.40 (2.49–2.40)		
$R_{\rm work}$	0.223 (0.333)		
$R_{\rm free}$	0.299 (0.388)		
Ramachandran plot, residues in			
Most favoured region (%)	93.1		
Additionally allowed region (%)	6.1		
Generally allowed region (%)	0.0		
Disallowed region (%)	0.8		

K3129), 25 mg l^{-1} seleno-L-methionine (SeMet) and 50 mg l^{-1} ampicillin. Cells were induced by the addition of 1 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 310 K for an additional 24 h before harvesting by centrifugation. The pellet was suspended, disrupted by ultrasonication and centrifuged. The supernatant was heated to 343 K and the denatured proteins were removed by centrifugation. The resulting supernatant was loaded onto chromatography columns containing SuperQ Toyopearl 650M (Tosoh, 30 ml), Resource Q (Amersham Biosciences, 6 ml), CHT10-1 (Bio-Rad, 10 ml) and HiLoad 16/60 Superdex 75 (Amersham Biosciences, 120 ml). Purified protein in buffer solution (20 m*M* Tris–HCl pH 9.0) was concentrated and crystallized at 293 K by the sitting-drop vapour-diffusion method using a precipitant solution containing 0.1 *M* CHES pH 9.4, 0.2 *M* ammonium sulfate and 25%(*w*/*v*) PEG 4000. Protein droplets for crystallization were prepared by mixing 1 µl protein solution (4.8 mg ml^{-1}) and 1 µl precipitant solution and droplets were equilibrated against 500 µl precipitant solution at 293 K. Approximately 20 d were required to obtain X-ray diffraction-quality crystals.

2.2. Diffraction experiment and structure analysis

Multiple-wavelength anomalous diffraction (MAD) data were collected at 100 K using a Jupiter 210 detector at beamline BL26B2 of SPring-8 (Hyogo Japan; Ueno *et al.*, 2004). The wavelengths of the X-ray beam were 0.97911 Å (Se *K* peak), 0.97940 Å (Se *K* edge) and 0.90000 Å (remote). The data were collected with a crystal-to-detector distance of 350 mm and reflections were recorded with an oscillation range per image of 0.5°. Diffraction data were indexed, integrated and scaled using *d*TREK* (Pflugrath, 1999). The calculated $V_{\rm M}$ value of 2.0 Å^3 Da^{-1} (Matthews, 1968) indicated the presence of four molecules in the asymmetric unit. During the initial stages of MAD structure determination, eight of the 12 possible Se atoms were identified with *SHARP* (de La Fortelle & Bricogne, 1997). The electron-density map was improved by solvent flattening and non-crystallographic symmetry averaging using the program *DM* (Cowtan, 1994). A coordinate model was constructed using the program *ARP*/*wARP* (Cohen *et al.*, 2004). A finalized set of atomic coordinates was obtained after iterative rounds of model building with the program *XtalView* (McRee, 1999) and refinement with *CNS* (Brünger *et al.*, 1998) by rigid-body minimization, simulated annealing, positional minimization, water-molecule identification and individual isotropic *B*-value refinement. During the refinement, non-crystallographic symmetry restraints were imposed with the exception of the following regions at the termini and the surface loops:

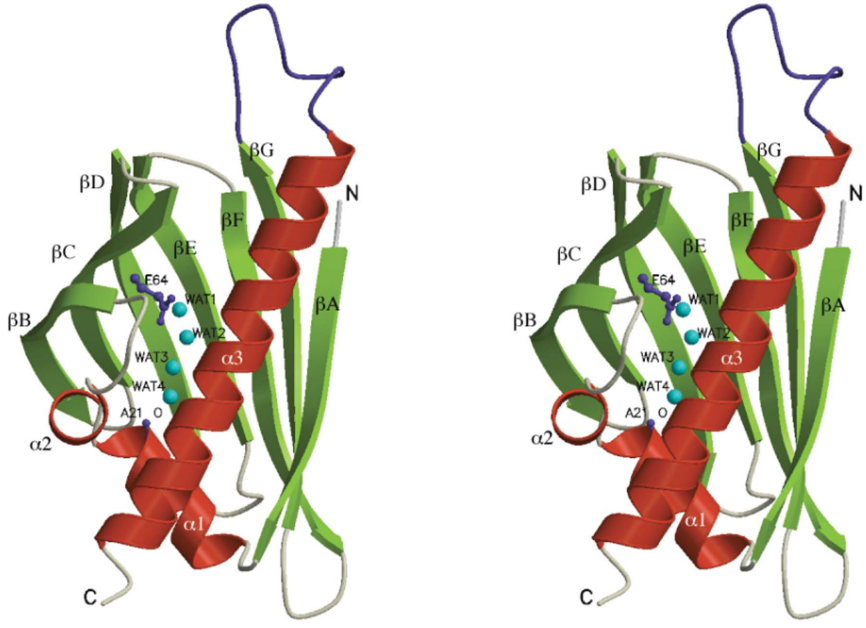


Figure 1
Stereoview of a ribbon representation of the structure of TTHA0849. Three α-helices (α1–3) and one curved up-and-down antiparallel β-sheet composed of seven β-strands (βA–βG–βF–βE–βD–βC–βB) form the hydrophobic cavity in the molecule. Four water molecules (cyan) and the atoms of Glu64 and Ala21 (blue), which comprise a hydrogen-bonding network in the cavity, are shown as a ball-and-stick model. The N- and C-terminal tails are labelled. This figure was drawn using the programs *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

Met1–Glu3, Ala41–Thr47, Pro95–Thr100, Ile114–Arg120 and Ser146–Ser147.

3. Results and discussion

The crystal belongs to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 56.30, b = 86.49, c = 109.58 Å. The R_{work} and R_{free} (Brünger, 1992) values of the refined coordinates are 0.223 and 0.299, respectively, at 2.4 Å resolution. A Ramachandran plot calculated from the refined model using *PROCHECK* (Laskowski *et al.*, 1993) showed the absence of residues in disallowed regions, with the exception of Ile114, which is on the border of the generally allowed and disallowed regions of the φ–ψ plot. Ile114 is located at the N-terminal end of a helical segment (Ile114–Leu118), showing a secondary structure intermediate between an α-helix and a 3_{10}-helix. Since its electron density is unambiguous, Ile114 is expected to be reliably positioned. X-ray data-collection and refinement statistics are given in Table 1.

The structure of TTHA0849 contains three α-helices (α1, Pro13–Lys22; α2, Leu24–Lys28; α3, Leu119–Ser146) and one curved up-and-down antiparallel β-sheet composed of seven β-strands [βA(Glu3–Ile10)–βG(Gly99–Glu109)–βF(Arg86–Glu96)–βE(Arg74–Gly82)–βD(Lys58–Asp69)–βC(Arg46–Ala55)–βB(Val34–Glu43)] (Fig. 1). A search for structurally similar domains using the program *DALI* (Holm & Sander, 1993) revealed that TTHA0849 is similar to CC1736 (function, unknown; Z score, 15.1; r.m.s.d. 2.5 Å; PDB code 1t17; Shen *et al.*, 2005), the StAR-related lipid-transport domain of MLN64 (function, lipid binding; Z score, 15.1; r.m.s.d. 2.1 Å; PDB code 1em2; Tsujishita & Hurley, 2000), birch pollen allergen Bet v 1 (function, allergen; Z score, 14.4; r.m.s.d. 3.3 Å; PDB code 1bv1; Gajhede *et al.*, 1996) and the mouse cholesterol-regulated START protein 4, StarD4 (function, lipid binding; Z score, 14.0; r.m.s.d. = 2.2 Å; PDB code 1jss; Romanowski *et al.*, 2002). In the following discussions, a comparison of the structure of TTHA0849 with those of the C-terminal segment of human MLN64, Bet v 1 and StarD4 proteins is described. The most striking structural feature of these proteins is that they contain a common tunnel-shaped feature in the α+β domain (Fig. 2), which is likely to contribute to functional specificity, even though TTHA0849 shares low sequence identity (10–14% identity) with its structural homologues. One side of the tunnel is completely closed, whereas the other end serves as an entrance through which ligands can access the interior of the molecule. Therefore, this arrangement can be described as a cavity that penetrates the molecule toward the ligand-binding site. A loop connecting α3 and βG seems to serve as a 'lid' which is involved in regulating the access of ligands into the cavity. Upon least-squares overlap of the four subunits in the asymmetric unit, an r.m.s.d. for C^{α} atoms in the region Ile114–Arg120, a linker between the loop and α3, is calculated to be 1.37 Å, a value significantly larger than the 0.41 Å value for the r.m.s.d. least-squares overlap obtained when comparing the positions of all atoms of the four subunits. This suggests that the loop has sufficient flexibility to serve as a gate. We believe that the surroundings of the loop comprise the entrance leading to the hydrophobic cavity and probably function as a gate in both conformations (open or closed) in solution. It may be necessary to suppress the flexibility of the loop in order to achieve the uniform and compact packing arrangement required to obtain diffraction-quality crystals. This is indicated from an examination of the crystal lattice packing, which shows that these loop regions on symmetry-related molecules interact.

There are four water molecules (Wat1–4) lined up along βD in the bottom portion of the cavity in TTHA0849 (Fig. 1). Two of these (Wat1 and Wat2) are within hydrogen-bonding distance of the carboxy O atoms of Glu64. The structure indicates that Wat4 is fixed by forming a hydrogen bond to the carbonyl O atom of Ala21 on α1 and Wat3 is fixed by participating in a hydrogen-bonding network with Wat4. Although water molecules are located in the cavities of other START domain-containing proteins (the numbers of water molecules in MLN64, StarD4 and Bet v 1 are 15, seven and four, respectively), these cavities are all composed of highly hydrophobic residues. The numbers of hydrophobic residues that comprise the inner wall of the cavity are 20 (TTHA0849), 12 (MLN64), 18 (Bet v 1) and 13 (StarD4). These four related proteins display a similar overall architecture for the cavity (Fig. 2); however, several detailed structural features (shape and size) of the cavity are different. In TTHA0849, since 11 residues with bulky side chains (Trp51, Trp62, Glu64, Glu66, Trp68, Phe77, Phe84, Tyr87, Trp91, Phe93 and Tyr108) are located in the cavity (Fig. 3), the size of the cavity is distinctly smaller compared with those in the other three proteins, in which the numbers of residues with bulky side-chains in the cavity are three

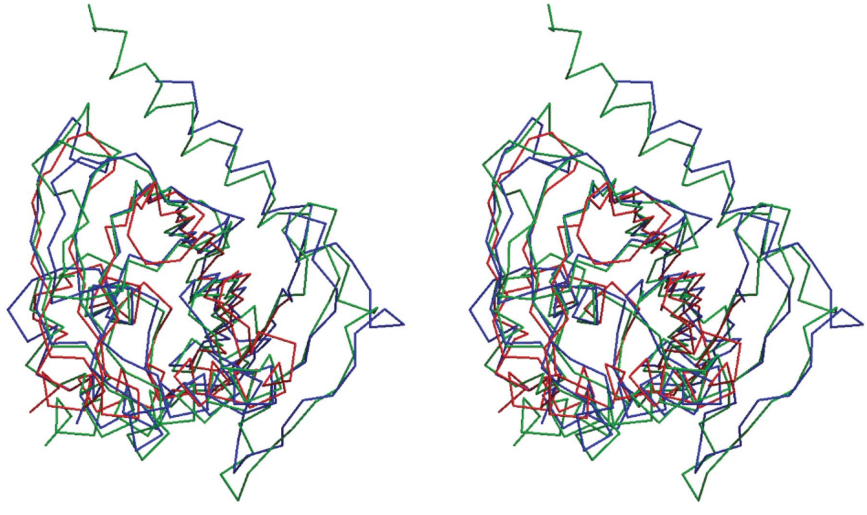


Figure 2
Stereoview of superimposed C^{α}-trace models of TTHA0849 (red), MLN64 (green) and StarD4 (blue) oriented by a 90° rotation about the horizontal axis with respect to the view in Fig. 1. This figure was drawn using the programs *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

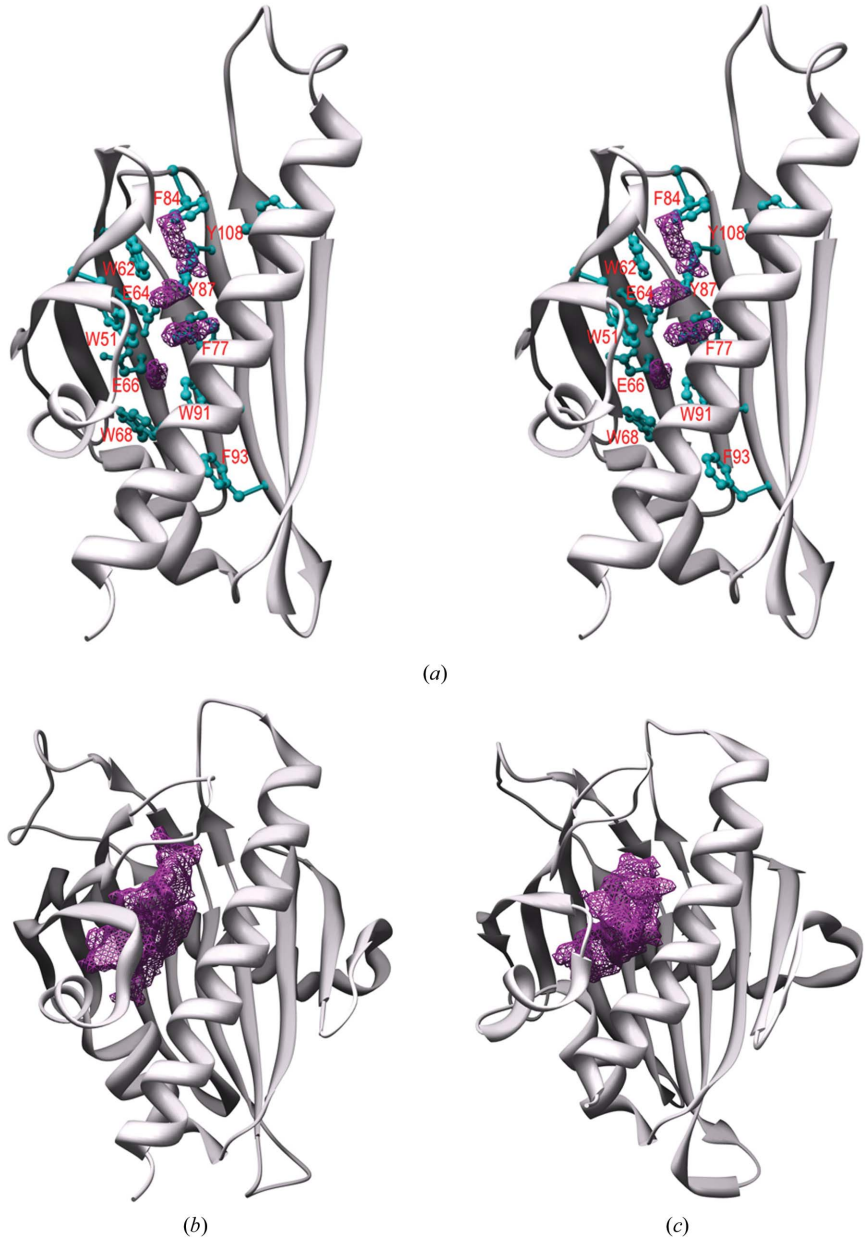


Figure 3
Ribbon diagrams with cavity-volume plots (magenta) of START domain-containing proteins in the same orientation as Fig. 1. Cavity volumes were calculated using the program *VOIDOO* with a probe radius of 1.4 Å (Kleywegt & Jones, 1994). Views of (*a*) TTHA0849 in stereo with some bulky residues (ball-and-stick model in cyan) and residue numbers in the cavity, (*b*) MLN64 and (*c*) StarD4 are shown. The cavities of StarD4 and MLN64 are composed of one connected tunnel and the volumes of the cavity are calculated to be 214 and 225 Å^3, respectively. In contrast, the space of the cavity of TTHA0849 is divided into four components and the total volume of the components is only 18 Å^3. Figures were drawn using the program *CHIMERA* (Pettersen *et al.*, 2004).

(MLN64), five (Bet v1) and four (StarD4). StarD4 and MLN64 have been reported to transfer a lipid molecule and the volumes of the cavity are 214 and 225 Å^3, respectively. In contrast, the cavity of TTHA0849 is closely packed by the side-chain atoms of aromatic residues. Therefore, the space in the cavity is divided into four components and the total volume of the components is only 18 Å^3 (Fig. 3). This distinctly smaller value suggests that the cavity of TTHA0849 does not contain sufficient space to allow molecules such as a cholesterol molecule to access the interior of the protein unless there is a drastic rearrangement of the bulky side chains, even though all the water molecules are swept out from the cavity. The outer component of the cavity (volume = 7.6 Å^3), which is the most accessible to the molecular surface, however, may play a role in the transfer of small ligands such as diatomic or triatomic molecules.

We would like to express our sincere thanks to Dr H. Axelrod of the Joint Center for Structural Genomics at the Stanford Synchrotron Radiation Laboratory for his helpful discussions. This work was supported in part by the RIKEN Structural Genomics/Proteomics Initiative and the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan, grants from the 21st Century COE Program and the JAXA project.

References

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–477.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* D**60**, 2222–2229.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dickens, M. L., Ye, J. & Strohl, W. R. (1995). *J. Bacteriol.* **177**, 536–543.
Gajhede, M., Osmark, P., Poulsen, M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., Schou, C., Lowenstein, H. & Spangfort, M. D. (1996). *Nature Struct. Biol.* **3**, 1040–1045.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Kendrew, S. G., Katayama, K., Deutsch, E., Madduri, K. & Hutchinson, C. R. (1999). *Biochemistry*, **38**, 4794–4799.
Kleywegt, G. J. & Jones, T. A. (1994). *Acta Cryst.* D**50**, 178–185.
Kraulis, J. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Madduri, K. & Hutchinson, C. R. (1995). *J. Bacteriol.* **177**, 3879–3884.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Newton, A. C. (1995) *Curr. Biol.* **5**, 973–976.
Niemi, J. & Mäntsälä, P. (1995). *J. Bacteriol.* **177**, 2942–2945.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Ponting, C. P. & Aravind, L. (1999). *Trends Biochem. Sci.* **24**, 130–132.
Räty, K., Kantola, J., Hautala, A., Hakala, J., Ylihonko, K. & Mäntsälä, P. (2002). *Gene*, **293**, 115–122.
Romanowski, M. J., Soccio, R. E., Breslow, J. L. & Burley, S. K. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 6949–6954.
Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5857–5864.
Shaw, G. (1996). *BioEssays*, **18**, 35–46.
Shen, Y., Goldsmith-Fischman, S., Atreya, H. S., Acton, T., Ma, L., Xiao, R., Honig, B., Montelione, G. T. & Szyperski, T. (2005). *Proteins*, **58**, 747–750.
Stocco, D. M. (2000). *J. Endocrinol.* **164**, 247–253.
Sultana, A., Kallio, P., Jansson, A., Wang, J.-S., Niemi, J., Mäntsälä, P. & Schneider, G. (2004). *EMBO J.* **23**, 1911–1921.
Torkkell, S., Kunnari, T., Palmu, K., Hakala, J., Mäntsälä, P. & Ylihonko, K. (2000). *Antimicrob. Agents Chemother.* **44**, 396–399.
Tsujishita, Y. & Hurley, J. H. (2000). *Nature Struct. Biol.* **7**, 408–414.
Ueno, G., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2004). *J. Appl. Cryst.* **37**, 867–873.