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

Structure of a putative molybdenum-cofactor biosynthesis protein C (MoaC) from *Sulfolobus tokodaii* (ST0472)

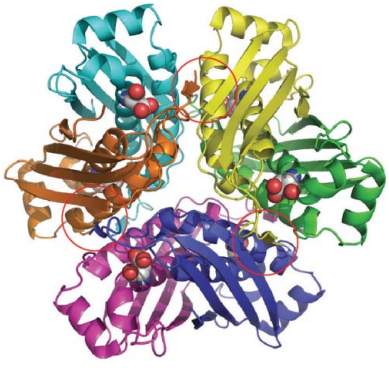
Hiromi Yoshida,[a] Mitsugu Yamada,[a] Seiki Kuramitsu[b] and Shigehiro Kamitori[a]*

[a]Division of Structural Biology, Life Science Research Center and Faculty of Medicine, Kagawa University, 1750-1 Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0793, Japan, and [b]Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan

Correspondence e-mail: kamitori@med.kagawa-u.ac.jp

Received 18 March 2008
Accepted 27 May 2008

PDB Reference: MoaC, 2ohd, r2ohdsf.



The crystal structure of a putative molybdenum-cofactor (Moco) biosynthesis protein C (MoaC) from *Sulfolobus tokodaii* (ST0472) was determined at 2.2 Å resolution. The crystal belongs to the monoclinic space group *C*2, with unit-cell parameters a = 123.31, b = 78.58, c = 112.67 Å, β = 118.1°. The structure was solved by molecular replacement using the structure of *Escherichia coli* MoaC as the probe model. The asymmetric unit is composed of a hexamer arranged as a trimer of dimers with noncrystallographic 32 symmetry. The structure of ST0472 is very similar to that of *E. coli* MoaC; however, in the ST0472 protein an additional loop formed by the insertion of seven residues participates in intermonomer interactions and the new structure also reveals the formation of an interdimer β-sheet. These features may contribute to the stability of the oligomeric state.

1. Introduction

The molybdenum cofactor (Moco) is an essential component for enzymes such as xanthine oxidase, sulfite oxidase and nitrate reductase and therefore plays an important role in biological systems (Rajagopalan, 1991). Moco has a tricyclic pyranopterin structure with the dithiolene moiety coordinating a Mo atom and is synthesized by an evolutionarily conserved biosynthetic pathway which is found in archaea, eubacteria and eukaryotes (Mendel & Bittner, 2006). In *Escherichia coli*, seven proteins have been reported to be involved in the Moco-biosynthetic pathway, as shown in Fig. 1(*a*) (Sanishvili *et al.*, 2004). Genomic analysis of a thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7, shows that the ST0472 protein (151 amino-acid residues, 17 125 Da) could be a putative Moco-biosynthesis protein C (MoaC), which is involved in the first step of Moco biosynthesis. ST0472 has an amino-acid sequence identity of 48% to *E. coli* MoaC as shown in Fig. 1(*b*) (Wuebbens *et al.*, 2000). Here, we report the crystal structure of ST0472 and its structural comparison with *E. coli* MoaC.

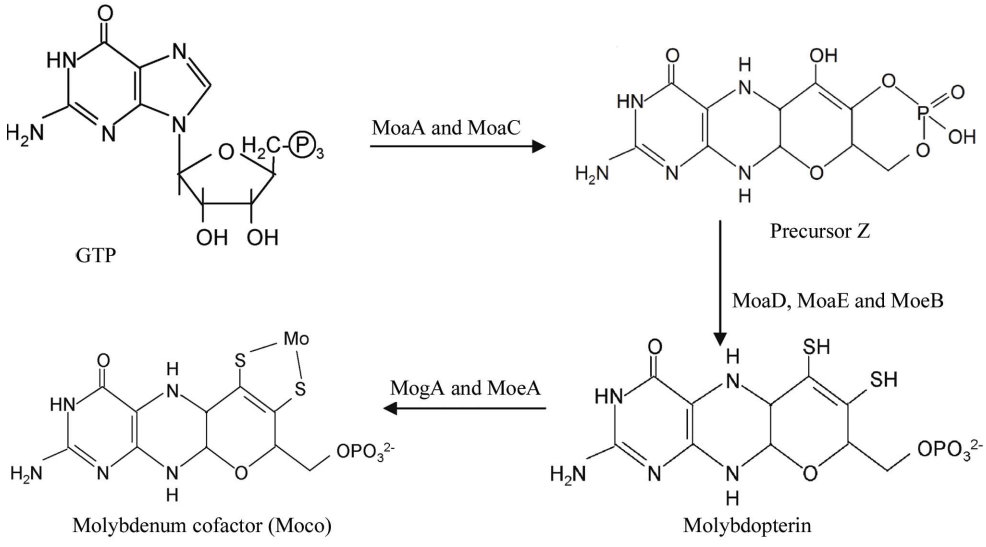
2. Materials and methods

The ST0472 gene (NCBI Gene ID 1458415) from *S. tokodaii* strain 7 genomic DNA was cloned by the PCR method using the designed primers forward (5′-ATAT**CATATG**ACCGAAGCTAAAATCGTAGATATTTCATCT-3′) and reverse (5′-ATAT**GGATCC**GCGGCCGCTTATTACTTCATATCGTCGTAAGT-3′), which include *Nde*I and *Bam*HI sites, respectively (shown in bold). The digested PCR product with *Nde*I and *Bam*HI was inserted into the expression vector pET11a (Novagen, San Diego, California, USA) linearized with *Nde*I and *Bam*HI. The recombinant plasmid was used for protein expression in *E. coli* Rosetta-gami (DE3) as the host strain. Cells harbouring the plasmid were cultured at 310 K overnight in 4× LB medium containing 50 µg ml^{-1} ampicillin. After cell disruption by sonication, the lysate was incubated at 343 K for 10 min and centrifuged to remove denatured protein. The supernatant was applied onto a Resource Iso column (GE Healthcare Biosciences Corp., Piscataway, New Jersey, USA) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 5 m*M* β-mercaptoethanol and 1.35 *M*

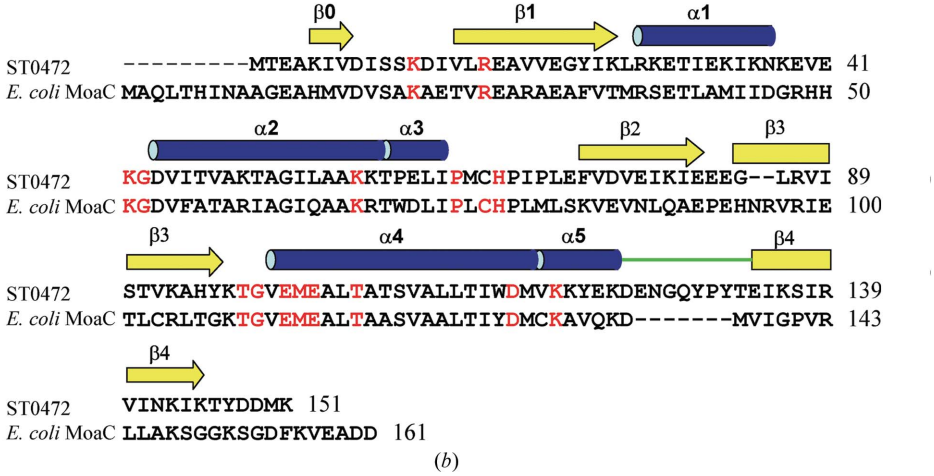
ammonium sulfate. ST0472 was found in the flowthrough fraction. After desalting the flowthrough fraction using a HiPrep 26/10 desalting column (GE Healthcare Biosciences Corp.), the protein solution was applied onto a Resource Q column (GE Healthcare Biosciences Corp.) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 5 m*M* β-mercaptoethanol. ST0472 was eluted with a linear gradient of 0–1.3 *M* NaCl and dialyzed against 10 m*M* potassium phosphate buffer pH 7.0 containing 5 m*M* β-mercaptoethanol. Dialyzed protein solution was loaded onto a Bio-Scale CHT5-I column (Bio-Rad Laboratories, Hercules, California, USA) equilibrated with the same buffer for dialysis and eluted with a linear gradient of 10–500 m*M* potassium phosphate. After gel filtration on a HiLoad 16/60 Superdex75 column (GE Healthcare Biosciences Corp.) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 5 m*M* β-mercaptoethanol and 150 m*M* NaCl, the protein was again applied onto the Resource Q column under the same conditions. The eluted ST0472 was equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 5 m*M* β-mercaptoethanol and 130 m*M* NaCl. The purified protein was concentrated to a final concentration of 15.5 mg ml^{-1}, using Vivaspin 20-5k (Sartorius Stedim Japan KK, Tokyo, Japan). Protein concentration was determined by measuring the absorbance at 280 nm using an absorption coefficient of 14 440 M^{-1} cm^{-1} calculated on the basis of the numbers of tryptophan and tyrosine residues (Pace *et al.*, 1995).

Initial crystallization screening for ST0472 was performed using the Emerald BioSystems Wizard I crystallization matrix with a 96-well Crystallization Microplate (Corning Incorporated, New York, USA) by the sitting-drop method at room temperature. Small crystals of ST0472 protein appeared on mixing 1.0 µl protein solution (15.5 mg ml^{-1}) and 1.0 µl reservoir solution [20%(*w*/*v*) PEG 3000, 0.1 *M* HEPES pH 7.5, 0.2 *M* NaCl; Wizard I-28]. After optimizing the crystallization condition using the hanging-drop vapour-diffusion method with a 24-well test plate (TPP, Switzerland), suitable-sized crystals with dimensions of 0.2 × 0.15 × 0.1 mm were obtained within 10 d using a reservoir solution containing 18%(*w*/*v*) PEG 3000, 0.1 *M* HEPES pH 7.5, 0.2 *M* NaCl.

X-ray diffraction data for ST0472 were collected at the Photon Factory (Tsukuba, Japan) on beamline BL-6A using an ADSC Quantum 4R CCD detector at 100 K, using a cryo-solution consisting of 20%(*v*/*v*) glycerol in reservoir solution. All data were processed with *HKL*-2000 (Otwinowski & Minor, 1997). Data-collection statistics are summarized in Table 1. As a search model, the structure



(*a*)



(*b*)

Figure 1
(*a*) Biosynthetic pathway of the molybdenum cofactor (Moco) from a guanosine derivative. (*b*) Sequence alignment between ST0472 and *E. coli* MoaC (PDB code 1ekr). Blue bars and yellow arrows show secondary-structure elements. The additional loop formed by the insertion of seven residues found in ST0472 is indicated by a green line. Highly conserved residues among 15 MoaC proteins from different organisms are highlighted in red.

of *E. coli* MoaC (PDB code 1ekr) was used for molecular replacement with the program *MOLREP* (Collaborative Computational Project, Number 4, 1994). After initial phase determination, the model was built with the programs *Coot* (Emsley & Cowtan, 2004) and *Xfit* (McRee, 1999) from the *XtalView* program system (McRee, 1993) and the structure was refined using the program *REFMAC*5 (Murshudov *et al.*, 1997) from the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). Finally, the structure was refined without noncrystallographic symmetry restraints by the program *CNS* (Brünger *et al.*, 1998) using conjugate-gradient minimization with Engh and Huber stereochemical parameters (Engh & Huber, 1991). The refinement statistics are presented in Table 1.

3. Results and discussion

The final model of ST0472 was refined in the resolution range 45.8–2.2 Å to a crystallographic *R* value of 20.2% (R_{free} = 25.2%) using isotropic temperature factors for all atoms. There are six protein molecules (Mol-*A*, Mol-*B*, Mol-*C*, Mol-*D*, Mol-*E* and Mol-*F*) per asymmetric unit, with r.m.s. deviations for C^{α} atoms between the six molecules of 0.73–1.36 Å. Some residues of the N-terminal regions (1–4 of Mol-*A*, 1–3 of Mol-*B*, Mol-*D*, Mol-*E* and Mol-*F* and 1–2 of Mol-*C*) and C-terminal regions (147–151 of Mol-*A*, 143–151 of Mol-*B*, 144–151 of Mol-*C*, Mol-*D* and Mol-*E* and 148–151 of Mol-*F*) are missing from the model because of poor electron density in these regions. The Ramachandran plot (Ramachandran & Sasisekharan,

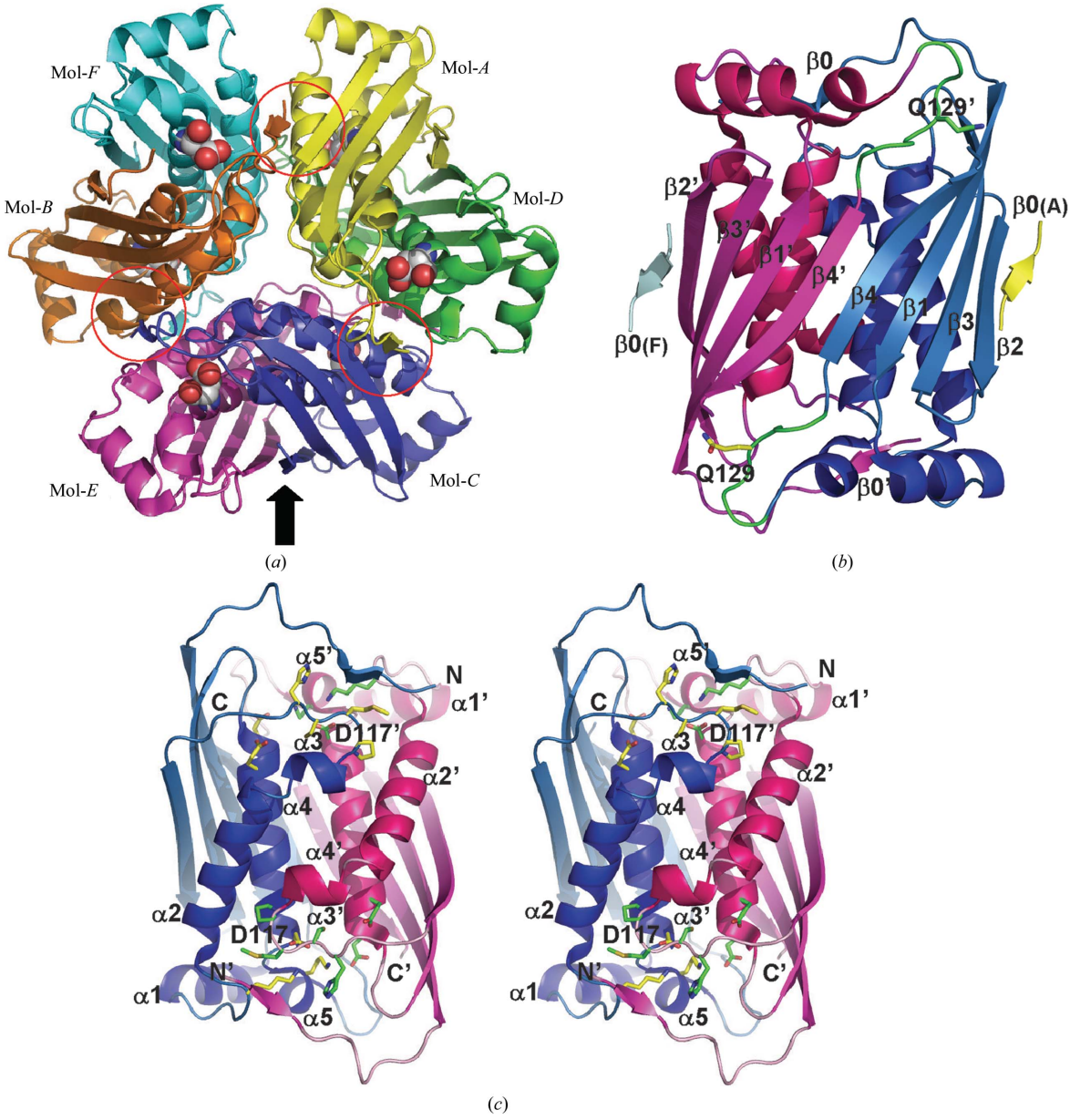


Figure 2
(*a*) The hexamer of ST0472 viewed along the threefold axis. Each colour represents a different monomer. Asp117 of each monomer, which is located at the centre of the putative active site, is shown by van der Waals spheres. Interdimer interactions forming the large β-sheet between β0 and β2 are indicated by red circles. (*b*) The dimer of ST0472 formed by Mol-*C* and Mol-*E* viewed along the twofold axis from the direction of the arrow in (*a*). The additional loop formed by the insertion of seven residues is shown in green and Gln129 residues are shown as yellow (Mol-*C*) and green sticks (Mol-*E*). The β0 strands from Mol-*A* and Mol-*F* are also illustrated. (*c*) Stereoview of the back side of (*b*). Highly conserved residues forming the putative active site are shown as yellow sticks (Mol-*C*) and green sticks (Mol-*E*). Asp117 residues are labelled.

Table 1
Data-collection and refinement statistics.

Data-collection details	
Temperature (K)	100
Wavelength (Å)	0.978
Resolution range (Å)	50.00–2.20 (2.28–2.20)
No. of measured reflections	164887
No. of unique reflections	48362 (4788)
Redundancy	3.4 (3.2)
Completeness (%)	99.9 (99.6)
Mean $I_o/\sigma(I_o)$	11.3 (4.63)
R_{merge}† (%)	5.5 (26.8)
Space group	*C*2
Unit-cell parameters (Å, °)	a = 123.31, b = 78.58, c = 112.67, β = 118.1
Refinement details	
Resolution range (Å)	45.78–2.20 (2.34–2.20)
No. of reflections	46051 (6304)
Completeness (%)	95.1 (87.4)
R factor (%)	20.2 (27.6)
R_{free}‡ (%)	25.2 (30.9)
R.m.s.d. bond length (Å)	0.006
R.m.s.d. bond angles (°)	1.30
No. of atoms	
Protein	6687
Solvent	443
Overall mean B value (Å^2)	29.4
PDB code	2ohd

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement and $\langle I(hkl)\rangle$ is the weighted mean of all measurements of $I(hkl)$. ‡ R_{free} is calculated using 10% of reflections.

1968) calculated using the program *PROCHECK* (Laskowski *et al.*, 1993) demonstrated that 90.0% of the residues lie in the most favoured regions and 10.0% of the residues are in additionally allowed regions.

The biological unit of *E. coli* MoaC was reported to be a hexamer with 32 symmetry (Wuebbens *et al.*, 2000). Six molecules of ST0472 per asymmetric unit are also arranged as a hexamer (Fig. 2*a*) in the same manner as in *E. coli* MoaC. The hexamer is formed from three dimers (Mol-*A* and Mol-*D*, Mol-*B* and Mol-*F*, Mol-*C* and Mol-*E*) with twofold symmetry, in which 22% of the total accessible surface of each monomer is used for the dimer interface as calculated using the program *AREAIMOL* in the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The three dimers are associated around a threefold axis to form the hexamer, in which 22% of the total accessible surface of each dimer is buried.

The overall protein fold and secondary structures are almost completely conserved between ST0472 and *E. coli* MoaC. A *DALI* search (Holm & Sander, 1993) shows that the r.m.s. deviation and *Z* score for 135 spatially corresponding C^α atoms are 0.9 Å and 22.5, respectively, between ST0472 and *E. coli* MoaC. The ST0472 monomer is globular, with an antiparallel β-sheet consisting of four β-strands (β1–β4) and five α-helices (α1–α5) which are located on the same side relative to the β-sheet; by dimerization, the two β-sheets are further connected to form a larger β-sheet, as shown in Fig. 2(*b*). Based on the analysis of sequence conservation among 18 MoaC proteins from different organisms, the putative active site is reported to be formed by Pro65–His68 (the loop between α3 and β2), Glu101–Glu103 (α4) and, from the opposing monomer of the dimer, Lys42 (loop α1 and α2), Asp117 (α4) and Lys120 (α4) (Wuebbens *et al.*, 2000). The corresponding site in ST0472 formed by these highly conserved residues is located at each end of the dimer (Fig. 2*c*), with a similar orientation of the putative active-site residues as was observed in the structure of the *E. coli* enzyme.

A striking structural difference between ST0472 and *E. coli* MoaC is found in the loop region between α5 and β4. In ST0472, seven residues (Glu126, Asn127, Gly128, Gln129, Tyr130, Pro131 and Tyr132) are inserted into this region to form the additional loop shown in green in Figs. 1(*b*) and 2(*b*). In the dimer, this additional loop of ST0472 makes intermonomer interactions with β4 of the opposing subunit which may contribute to the stability of the dimer structure. In particular, the side chain of Gln129 in the additional loop forms hydrogen-bond interactions with Asn142 and/or Lys143 of the opposing monomer, while in *E. coli* MoaC the shorter three-residue loop (Lys134, Asp135 and Met136) connects α5 and β4 without taking part in intermonomer interactions. Another structural difference is found in the interdimer surface of the hexamer. Three large β-sheets of dimers are located on the surface of the hexamer and cover the bundle of α-helices. In ST0472 three residues (Lys5, Ile6 and Val7) in the N-terminal loop region adopt a β-strand conformation (β0) and Ile6 makes a hydrogen-bond pair with Val75 in β2 of the adjacent dimer, as shown in Fig. 2(*a*). Consequently, the extended β-sheet consists of ten β-strands from four monomers, *e.g.* β0 from Mol-*A*, β2, β3, β1, β4 from Mol-*C*, β4, β1, β3, β2 from Mol-*E* and β0 from Mol-*F*, giving favourable interdimer interactions for formation of the hexamer, as shown in Fig. 2(*b*).

In this study, the putative Moco-biosynthesis protein C from *S. tokodaii*, ST0472, shows a highly similar structure, including the putative active site, to *E. coli* MoaC; however, there are significant structural differences between them. In ST0472 the insertion of seven residues to form an additional loop and the formation of an interdimer β-sheet may contribute to the stability of the oligomeric state of the enzyme. To investigate this further, additional experiments using site-directed mutant proteins will be required.

This study was supported by the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology of Japan. This research was performed with the approval of the Photon Factory Advisory Committee and the National Laboratory for High Energy Physics, Japan.

References

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.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McRee, D. E. (1993). *Practical Protein Crystallography*. New York: Academic Press.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Mendel, R. R. & Bittner, F. (2006). *Biochim. Biophys. Acta*, **1763**, 621–635.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pace, N. C., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). *Protein Sci.* **4**, 2411–2433.
Rajagopalan, K. V. (1991). *Adv. Enzymol.* **64**, 215–290.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.
Sanishvili, R., Beasley, S., Skarina, T., Glesne, D., Joachimiak, A., Edwards, A. & Savchenko, A. (2004). *J. Biol. Chem.* **40**, 42139–42146.
Wuebbens, M. M., Liu, M. T., Rajagopalan, K. & Schindelin, H. (2000). *Structure*, **8**, 709–718.