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

Overexpression, purification, crystallization and preliminary X-ray analysis of putative molybdenum cofactor biosynthesis protein C (MoaC2) from *Mycobacterium tuberculosis* H37Rv

Shubhra Srivastava, Vijay Kumar Srivastava, Ashish Arora* and J. Venkatesh Pratap*

Molecular and Structural Biology Division, CSIR – Central Drug Research Institute, Chattar Manzil Palace, Lucknow 226 001, India

Correspondence e-mail: ashish_arora@cdri.res.in, jvpratap@cdri.res.in

Received 21 February 2012
Accepted 16 April 2012

Rv0864 (MoaC2) from *Mycobacterium tuberculosis* is one of the enzymes in the molybdenum cofactor (Moco) biosynthesis pathway. Together with MoaA, MoaC is involved in the conversion of guanosine triphosphate (GTP) to precursor Z, the first step in Moco synthesis. Full-length MoaC2 (17.5 kDa, 167 residues) was cloned in *Escherichia coli* and purified to homogeneity. Crystals of recombinant *M. tuberculosis* MoaC2 were grown by vapour diffusion using a hanging-drop setup. Diffracting crystals grew in a condition in which 3 µl protein solution at 10.5 mg ml^{-1} was mixed with 1.5 µl reservoir solution (0.025 *M* potassium sodium tartrate tetrahydrate pH 8.0) and equilibrated against 1000 µl reservoir solution. Diffraction data extending to 2.5 Å resolution were collected at 100 K. The crystal belonged to the cubic space group $P2_13$, with unit-cell parameter 94.5 Å. Matthews coefficient (V_M) calculations suggested the presence of two molecules in the asymmetric unit, corresponding to a solvent content of about 39%. Molecular-replacement calculations using the *E. coli* homologue as the search model gave an unambiguous solution.

1. Introduction

Molybdenum cofactor (Moco) is synthesized by a conserved biosynthetic pathway in organisms spanning archaea, eubacteria and higher eukaryotes (Chan *et al.*, 1995). Moco contains mononuclear molybdenum, which is an essential nutrient for growth in micro-organisms, plants and animals. Molybdenum is found in the form of its divalent anion in soil and oceans, with the latter being the more abundant source. Moco-containing enzymes occupy key positions between living organisms and the physical environment (Stiefel, 2002); in humans, genetic deficiencies in Moco-biosynthesis enzymes have been shown to cause pleiotropic loss of sulfite oxidase, aldehyde oxidase and xanthine dehydrogenase, usually resulting in autosomal recessive and fatal diseases (Leimkühler *et al.*, 2005; Reiss, 2000; Reiss & Johnson, 2003). Based on the cofactors utilized, Moco-containing enzymes have been grouped into two categories: those that use the iron–sulfur cluster-based iron–molybdenum cofactor (FeMo-co) and those that use a pterin-based cofactor (Hille, 1994, 1996). Although some details of Moco biosynthesis remain unclear, the pathway can broadly be divided into three steps: a guanosine derivative, most likely to be GTP, is converted to precursor Z, which is then converted to molybdopterin, followed by incorporation of the metal (Wuebbens *et al.*, 2000). In *Escherichia coli*, seven proteins have been reported to be involved in the Moco biosynthesis pathway, as shown in Fig. 1(*a*) (Sanishvili *et al.*, 2004). Together with MoaA, MoaC is involved in the first step of Moco biosynthesis, *i.e.* the conversion of a guanosine derivative to precursor Z. This process is different from other pterin-biosynthetic pathways as C8 of the purine is inserted between the second and third ribose C atoms rather than being eliminated (Wuebbens & Rajagopalan, 1995). Overall, the first step of Moco biosynthesis appears to involve a rather sophisticated rearrangement reaction. At present, it is unclear whether MoaA and

© 2012 International Union of Crystallography
All rights reserved

MoaC act independently of each other or form a protein–protein complex. MoaA belongs to the *S*-adenosylmethionine (SAM)-dependent radical enzyme superfamily, members of which catalyse the formation of protein or substrate radicals by reductive cleavage of SAM by a [4Fe–4S] cluster (Frey & Magnusson, 2003; Jarrett, 2003; Sofia *et al.*, 2001). Moco biosynthesis also plays an important role in *Mycobacterium tuberculosis* as Moco is an essential cofactor of a diverse group of redox enzymes as annotated by the TB Structural Genomics Consortium (http://genolist.pasteur.fr/TubercuList/). In *M. tuberculosis* H37Rv three orthologous genes are involved in the Moco-biosynthesis pathway: *moaC1* (*Rv3111*), *moaC2* (*Rv0864*) and *moaC3* (*Rv3324c*). *moaC2* has been reported to be downregulated approximately 2.8-fold in the nutrient-starvation model of *M. tuberculosis* (Betts *et al.*, 2002). Here, we report the cloning, overexpression, purification, crystallization and preliminary X-ray analysis of Rv0864 (MoaC2) from *M. tuberculosis* H37Rv.

2. Experimental methods

2.1. Cloning

Genomic DNA of *M. tuberculosis* H37Rv was prepared as described by Kremer *et al.* (2005). The primers used for the isolation and amplification of the genes by polymerase chain reaction (PCR) from the genomic DNA of *M. tuberculosis* H37Rv were as follows: forward primer, 5′-GGAT**CCATGG**CCAGGGCTTCTGGGGC-3′; reverse primer, 5′-CTAG**CTCGAG**TCGCCTCGTCCAGGTCCCG-C-3′. The forward primer contains an *Nco*I restriction-enzyme site and the reverse primer contains an *Xho*I restriction-enzyme site (shown in bold). 10 ng *M. tuberculosis* H37Rv genomic DNA was used in a typical 50 µl PCR reaction. The PCR reactions were carried out using an MJ Mini Personal Thermo Cycler (Bio-Rad). Each reaction consisted of an initial denaturation step of 2 min at 368 K which facilitated melting of the genomic DNA, followed by 30 cycles of denaturation at 367 K for 1 min, primer annealing at 338 K for 1 min and primer extension at 345 K for 1.5 min. The PCR reaction was terminated with a final extension step at 345 K for 20 min and cooled to 277 K. The PCR products were extracted from the gel using a gel-extraction kit (Amersham Biosciences).

The PCR product of Rv0864 was digested with *Nco*I and *Xho*I in 1× NEBuffer 4 for 2 h at 310 K in a water bath and the digested products were run on a 1.5% agarose gel and purified using a DNA gel-extraction kit. The *Nco*I–*Xho*I-digested PCR product was ligated into the vector pET-28b digested with the same restriction enzymes. The ligation mixture was then directly transformed into chemically competent *E. coli* DH5α cells. Positive clones were identified by restriction digestion with *Nco*I and *Xho*I restriction enzymes. Clones with the required insert and vector sizes were selected as positive clones and were verified by sequencing using T7 promoter primers at the DNA Sequencing Facility at the University of Delhi South Campus, New Delhi, India. The construct, which contained an additional eight residues (LEHHHHHH) at the C-terminus of the Rv0864 sequence, was overexpressed in pET-28b.

2.2. Overexpression

The expression vector was introduced into *E. coli* BL21 (DE3) strain and the recombinant strain was cultured in LB medium supplemented with 50 µg ml^{-1} kanamycin at 310 K and was induced with 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside on reaching an $OD_{600\ nm}$ of 0.6. After induction, the culture was grown for a further 5–6 h at 310 K. The cells were then collected by centrifugation and the cell pellet from 1 l culture was suspended in 20 ml 50 m*M* Tris–HCl buffer containing 300 m*M* NaCl and 10 m*M* imidazole pH 8.0 (buffer *A*). The cells were disrupted using a Constant Cell Disruption System [Labmate (Asia) Pvt. Ltd] at 152 MPa at room temperature and the sample was centrifuged at 10 000*g* for 1 h at 277 K.

2.3. Protein purification

The clear supernatant thus obtained was applied onto a 5 ml Ni–NTA Superflow column pre-equilibrated with buffer *A* at room temperature. Unbound proteins were removed by washing with 30 m*M* imidazole in lysis buffer (buffer *A*). The bound MoaC2 was finally eluted with five column volumes of buffer *A* containing 250 m*M* imidazole. The column fractions were pooled, dialyzed to

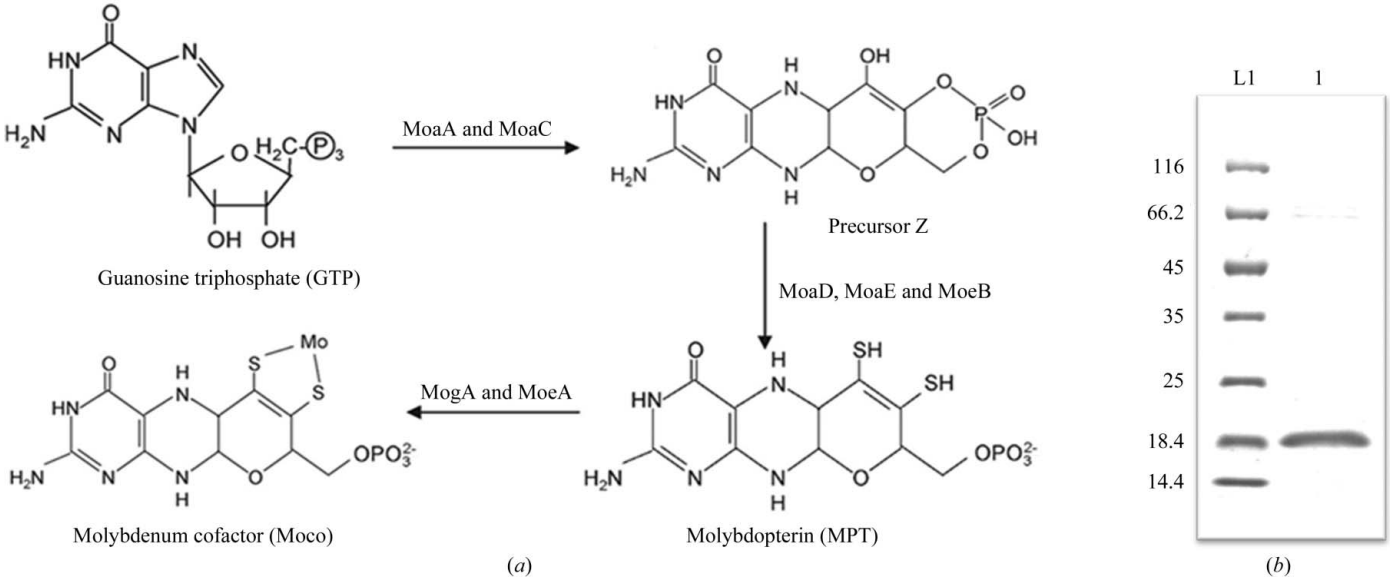

Figure 1
(*a*) Schematic representation of the molybdenum cofactor (Moco) biosynthesis pathway (adapted from Sanishvili *et al.*, 2004). (*b*) 12% SDS–PAGE of the expressed MoaC2. Lane L1, molecular-mass marker (labelled in kDa); lane 1, 12 µg purified Rv0864 protein.

remove imidazole and concentrated using an Amicon Centricon membrane filter (5 kDa cutoff). The concentrated samples were subsequently purified by FPLC (Bio-Rad) using a prepacked Superdex 75 10/300 column (GE Healthcare, USA) in a buffer consisting of 20 m*M* Tris–HCl, 50 m*M* NaCl, 1 m*M* EDTA pH 8.0 and were more than 99% pure as seen on 12% SDS–PAGE (Fig. 1*b*).

2.4. Crystallization and data collection

Preliminary crystallization trials were initiated by the hanging-drop vapour-diffusion method in 24-well crystallization plates using Crystal Screen and Crystal Screen 2 from Hampton Research at 295 K. 2 µl protein solution (6.5 mg ml^{-1}) was mixed with 1 µl reservoir solution and equilibrated against 500 µl reservoir solution. A crystalline precipitate appeared in Crystal Screen condition No. 2 (0.4 *M* potassium sodium tartrate tetrahydrate) after two weeks. This condition was further optimized by varying the protein concentration (6–12.5 mg ml^{-1}), the precipitant concentration (0.025–0.1 *M*) and the reservoir volume (500–1000 µl). Good diffraction-quality crystals appeared in about 15 d from a drop consisting of 3 µl protein solution (10.5 mg ml^{-1}) and 1.5 µl reservoir solution (0.025 *M* potassium sodium tartrate tetrahydrate) equilibrated against 1000 µl reservoir solution.

For X-ray data collection, crystals were mounted on CryoLoops (Hampton Research), rinsed with cryoprotectant solution [30%(*v*/*v*) paraffin oil in the reservoir solution] and flash-cooled directly in a nitrogen stream at 100 K. The crystal diffracted to 2.5 Å resolution (Fig. 2*a*) and diffraction data were collected in-house on a Rigaku MicroMax-007 HF X-ray generator (Cu $K\alpha$) using a MAR345dtb detector. The reflections were indexed using *iMOSFLM* (Battye *et al.*, 2011) and scaled with *SCALA* (Evans, 2006). The *CTRUNCATE* program (French & Wilson, 1978) was used to convert intensities to structure factors. The diffraction data statistics are presented in Table 1.

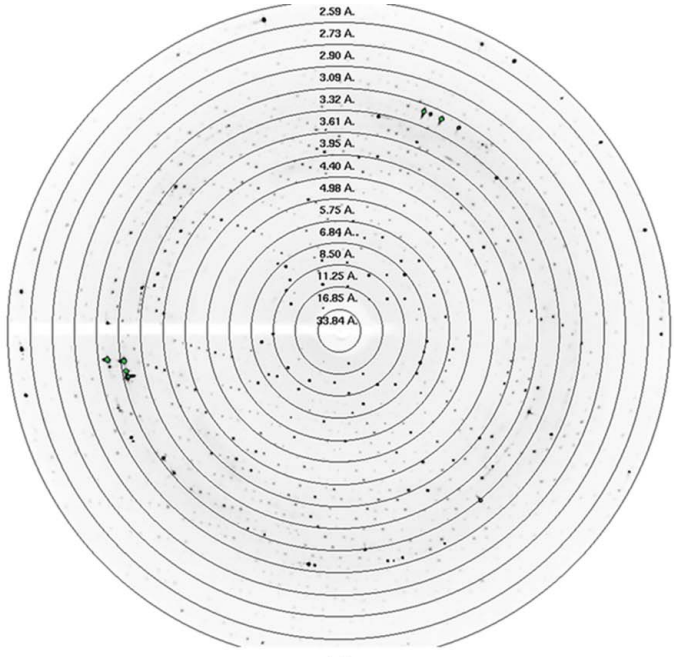

Figure 2
(*a*) Snapshot of the diffraction pattern of the MoaC2 crystal. The image corresponds to a 1° oscillation with 3 min exposure time and a crystal-to-detector distance of 250 mm. The outermost circle corresponds to 2.5 Å resolution. (*b*) Sequence alignment of *M. tuberculosis* MoaC2 and its homologue from *E. coli* (PDB entry 1ekr). The secondary-structural elements corresponding to the *E. coli* protein are shown at the top. Helices and strands are represented by coils and arrows, respectively. Conserved residues are highlighted in red boxes. The sequence alignment was produced using the program *ClustalW* (Thompson *et al.*, 1994) and the figure was generated using the program *ESPript* (Gouet *et al.*, 1999).

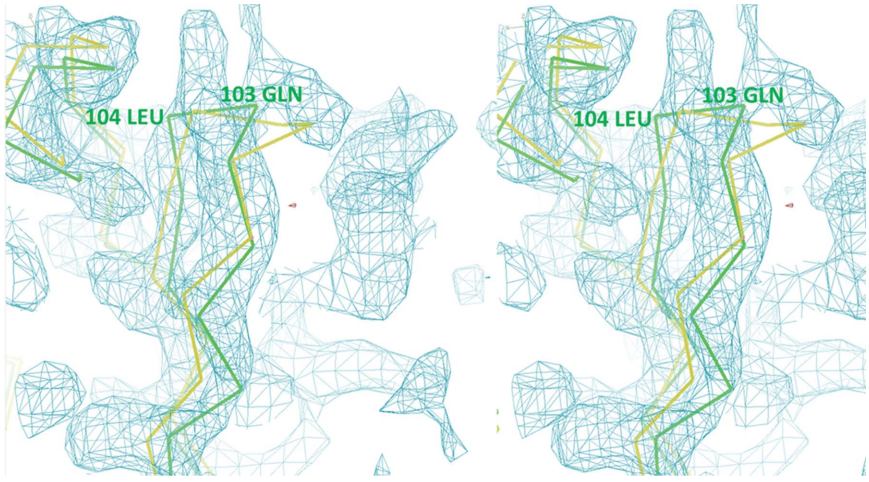

Figure 3
Stereoview of the $2|F_o| - |F_c|$ map contoured at 1.0σ corresponding to a region involving a two-residue deletion in the *M. tuberculosis* sequence with respect to the *E. coli* homologue (residues 92 and 93). C^{α} traces of *E. coli* MoaC (yellow) and Rv0864 (green) are shown; the residue numbers are those of the *M. tuberculosis* sequence. This figure was generated using *PyMOL* (v.1.2r3pre; Schrödinger LLC).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.54
Space group	$P2_13$
Unit-cell parameter (Å)	94.5
Matthews coefficient (Å^3 Da^{-1})	1.98
Solvent content (%)	39
Resolution range (Å)	33–2.5 (2.64–2.50)
No. of observations	90770 (12994)
No. of unique reflections	7563 (1097)
Multiplicity	12.0 (11.8)
$\langle I/\sigma(I)\rangle$	55.3 (34.7)
Completeness (%)	99.8 (100)
R_{merge}†	0.03 (0.05)
Refinement statistics	
R factor (%)	27.7
R_{free} (%)	37.7
Protein atoms	1951
Water molecules	0
R.m.s.d. bond lengths (Å)	0.006
R.m.s.d. bond angles (°)	1.209
Ramachandran plot statistics: residues in (%)	
Favoured regions	91.0
Allowed regions	8.2
Disallowed regions	0.8

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of the *i* observations.

3. Results and discussion

MoaC2 was successfully cloned in *E. coli* and purified to homogeneity with a C-terminal His_6 tag. The molecular weight of 17.5 kDa for monomeric His_6-tagged MoaC2 was confirmed by 12% SDS–PAGE. The typical yield of protein was 10–15 mg per litre of culture. After processing the reflections using *iMOSFLM* and scaling with *SCALA*, the resulting R_{merge} was 0.03, with a mosaicity of 0.7°. Matthews coefficient (V_M) calculations using the *CCP*4 suite (Winn *et al.*, 2011) suggested the presence of two monomers in the asymmetric unit, with a solvent content of 39% and a V_M of 1.98 Å^3 Da^{-1} (Matthews, 1968). A sequence-based search for MoaC2 against the sequences in the PDB using *BLAST* (http://blast.ncbi.nlm.nih.gov) showed 44% sequence identity with *E. coli* MoaC (PDB entry 1ekr; Wuebbens *et al.*, 2000), which was used as a search model for molecular replacement. *Phaser* (McCoy *et al.*, 2007) as implemented in the *CCP*4 suite was used to perform the molecular-replacement calculations, which resulted in a unique solution. Two copies of the search model were unambiguously placed in the asymmetric unit with acceptable *Z* scores and log-likelihood gain (LLG). The rotation- and translation-function *Z* scores after positioning the first molecule were 4.1 and 13.5, respectively, while the corresponding LLG was 101. These values were 3.2, 15.2 and 353, respectively, after the second molecule had been placed. The solutions were visually examined for steric clashes and none were found. The initial model from *Phaser* was converted into a polyalanine model using the program *CHAINSAW* (Stein, 2008). A total of 5% of the reflections were kept aside for the calculation of R_{free} (Brünger, 1992). An initial rigid-body refinement was followed by positional refinement using *REFMAC*5 (Murshudov *et al.*, 2011) from the *CCP*4 suite. Conventional electron-density maps ($2|F_o| - |F_c|$ and $|F_o| - |F_c|$) were then used to progressively fit residues as per the sequence using *Coot* (Emsley & Cowtan, 2004). Sequence alignment of Rv0864 (Fig. 2*b*) with its *E. coli* counterpart showed a deletion in a loop in Rv0864 with respect to *E. coli* MoaC at positions 103–104 (corresponding to *E. coli* MoaC residues 92–93). This was clearly observed in this electron-density map (Fig. 3) and the region was appropriately modelled. Iterations of model building using *Coot* (Emsley & Cowtan, 2004) followed by refinement enabled us to fit about 120 of 167 residues (excluding the C-terminal additional residues in the construct) in both chains of the asymmetric unit, with *R* and R_{free} of 27.7% and 37.7%, respectively. The Ramachandran plot (Ramachandran & Sasisekharan, 1968) calculated using the program *PROCHECK* (Laskowski *et al.*, 1993) demonstrated that 91.0% of the residues lie in the most favoured regions, 8.2% of the residues are in additionally allowed regions and 0.8% are in disallowed regions. The current refinement statistics are given in Table 1. Further model building and refinement including simulated annealing using *PHENIX* (Adams *et al.*, 2010) is currently in progress.

The in-house X-ray diffraction facility was established with financial support from an Institutional Grant and computational support was from a Department of Science and Technology, Government of India grant to JVP and Council of Scientific and Industrial Research, India (CSIR) grant CMM0017. JVP and VKS thank Ruchir Kant for maintaining the X-ray laboratory facilities in CDRI. This work was

supported by CSIR network grant NWP0038. SS and VKS acknowledge the Indian Council of Medical Research, New Delhi and the Department of Biotechnology, Government of India, respectively, for fellowships. The CDRI communication number for this manuscript is 8248.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. (2002). *Mol. Microbiol.* **43**, 717–731.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. & Rees, D. C. (1995). *Science*, **267**, 1463–1469.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Frey, P. A. & Magnusson, O. T. (2003). *Chem. Rev.* **103**, 2129–2148.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hille, R. (1994). *Biochim. Biophys. Acta*, **1184**, 143–169.
Hille, R. (1996). *Chem. Rev.* **96**, 2757–2816.
Jarrett, J. T. (2003). *Curr. Opin. Chem. Biol.* **7**, 174–182.
Kremer, K., Au, B. K., Yip, P. C., Skuce, R., Supply, P., Kam, K. M. & van Soolingen, D. (2005). *J. Clin. Microbiol.* **43**, 314–320.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leimkühler, S., Charcosset, M., Latour, P., Dorche, C., Kleppe, S., Scaglia, F., Szymczak, I., Schupp, P., Hahnewald, R. & Reiss, J. (2005). *Hum. Genet.* **117**, 565–570.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
Reiss, J. (2000). *Hum. Genet.* **106**, 157–163.
Reiss, J. & Johnson, J. L. (2003). *Hum. Mutat.* **21**, 569–576.
Sanishvili, R., Beasley, S., Skarina, T., Glesne, D., Joachimiak, A., Edwards, A. & Savchenko, A. (2004). *J. Biol. Chem.* **40**, 42139–42146.
Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. (2001). *Nucleic Acids Res.* **29**, 1097–1106.
Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
Stiefel, E. I. (2002). *Met. Ions Biol. Syst.* **39**, 1–29.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wuebbens, M. M., Liu, M. T. W., Rajagopalan, K. & Schindelin, H. (2000). *Structure*, **8**, 709–718.
Wuebbens, M. M. & Rajagopalan, K. V. (1995). *J. Biol. Chem.* **270**, 1082–1087.