Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An extremely SAD case: structure of a putative redox-enzyme maturation protein from *Archaeoglobus fulgidus* at 3.4 Å resolution

Olga Kirillova,[a] Maksymilian Chruszcz,[a] Igor A. Shumilin,[a] Tatiana Skarina,[b] Elena Gorodichtchenskaia,[b] Marcin Cymborowski,[a] Alexei Savchenko,[b] Aled Edwards[b] and Wladek Minor[a]*

[a]Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA, and [b]Banting and Best Department of Medical Research and Structural Genomics Consortium, 112 College Street, University of Toronto, Toronto, Ontario M5G 1L6, Canada

Correspondence e-mail: wladek@iwonka.med.virginia.edu

Received 18 September 2006
Accepted 18 December 2006

PDB Reference: AF0173, 2o9x, r2o9xsf.

This paper describes the crystal structure of AF0173, a putative redox-enzyme maturation protein (REMP) from *Archaeoglobus fulgidus*. The REMPs serve as chaperones in the maturation of extracytoplasmic oxidoreductases in archaea and bacteria. The all-helical subunits of AF0173 form a dimer arising from the interaction of residues located in a funnel-shaped cavity on one subunit surface with an uncut expression tag from the other subunit. This cavity is likely to represent a binding site for the twin-arginine motif that interacts with REMPs. The conservation of the overall fold in AF0173 and bacterial REMPs as well as the presence of conserved residues in their putative binding sites indicates that REMPs act in a similar manner in archaea and bacteria despite their limited sequence similarity. A model of the binding of the twin-arginine motif by AF0173 is suggested. The solution of the AF0173 structure by the single anomalous dispersion method represents an extreme case of SAD structure determination: low resolution (3.4 Å), the absence of NCS and the presence of only two anomalously scattering atoms in the asymmetric unit. An unusually high solvent content (73%) turned out to be important for the success of the density-modification procedures.

1. Introduction

The purpose of structural genomics projects is to use the power of structural annotation to provide functional clues about new proteins. In many structural genomics efforts, there is a drive to work on high-resolution diffraction data from high-quality crystals so that the crystallographic software can perform most of the structure determination and refinement in an automated manner. This strategy can be both highly productive and economical, but also has the drawback that more challenging crystallographic problems are set aside.

This paper describes the structure of AF0173, Midwest Center for Structural Genomics (MCSG) code APC5565, a protein from the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus*. The function of this protein was unknown prior to this study, but the structural, sequence and genetic analysis described here demonstrates that AF0173 is the first characterized archaeal redox-enzyme maturation protein (Turner *et al.*, 2004) and is likely to function in a manner similar to the well studied bacterial chaperones DmsD and TorD. The bacterial REMPs play a role in the maturation of periplasmic respiratory enzymes such as trimethylamine *N*-oxide (TMAO), dimethyl sulfoxide (DMSO), nitrate and sulfur reductases and formate dehydrogenase. These oxido-reductases contain at least one redox-active cofactor such as molybdopterin, an iron–sulfur center or heme. The cofactor is acquired in the cytoplasm before the enzymes are exported

from the cell in a folded conformation through the twin-arginine translocation (Tat) system (Berks *et al.*, 2000; Santini *et al.*, 1998; Sargent *et al.*, 1998; Weiner *et al.*, 1998). The activity of the oxidoreductases is either abolished or significantly decreased in the absence of REMPs (Blasco *et al.*, 1998; Ilbert *et al.*, 2003; Oresnik *et al.*, 2001; Pommier *et al.*, 1998; Ray *et al.*, 2003). Although the specific role(s) of REMPs have not been firmly established, they are thought either to act as system-specific chaperones that bind to the unfolded cofactor-binding subunit in order to facilitate introduction of the cofactor or to prevent premature export of unfolded forms of the enzyme by binding to the twin-arginine consensus motif (Turner *et al.*, 2004).

The structure of AF0173 was established using one of the less well traveled structure-determination paths. The AF0173 crystals diffracted to low resolution (3.4 Å), had no non-crystallographic symmetry and only had two anomalously scattering atoms in the 181 amino acids present in the asymmetric unit. Nevertheless, the structure was determined using the single anomalous diffraction (SAD) technique, making AF0173 one of only 38 structures in the PDB (at the time of deposition) solved by SAD using diffraction data to worse than 3 Å. Our analysis suggests that a carefully performed SAD experiment may be more widely applicable than has previously been appreciated.

Table 1
Summary of the data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell (3.52–3.40 Å)

Data collection	
Space group	$P6_522$
Unit-cell parameters (Å)	$a = b = 137.5$, $c = 64.3$
Wavelength (Å)	0.9793
Resolution range (Å)	34.4–3.4
No. of observations	131304
No. of unique reflections	5277
Completeness† (%)	99.8 (100.0)
Anomalous completeness‡ (%)	99.9 (100.0)
$R_{\rm merge}$†	0.070 (0.507)
Anomalous $R_{\rm merge}$‡	0.058 (0.500)
Anomalous average redundancy‡	14.0 (14.1)
Average $I/\sigma(I)$	37.0 (5.2)
Phasing statistics	
Anomalous phasing power	2.6
Anomalous $R_{\rm cullis}$	0.57
Figure of merit (acentric) (*MLPHARE*)	0.34
After density modification (*DM*)	0.84
Model and refinement statistics	
Resolution range (Å)	34.4–3.4
No. of reflections (working set)	5008
No. of reflections (test set)	254
Completeness (%)	100
$R_{\rm work}$	0.237
$R_{\rm free}$	0.285
Restraints (r.m.s. observed)	
Bond length (Å)	0.022
Bond angle (°)	2.51
Average isotropic B value (Å^2)	78
Protein residues/atoms	162/1130

† Friedel mates merged. ‡ Friedel mates separated.

2. Materials and methods

2.1. Protein expression and purification

The ORF of AF0173 was amplified by PCR from *A. fulgidus* genomic DNA (ATCC) and cloned into the *Nde*I and *Bam*HI sites of a modified pET15b cloning vector (Novagen). In the modified vector, the thrombin cleavage site is substituted by the TEV protease cleavage site and a double stop codon is introduced downstream from the *Bam*HI site (Zhang *et al.*, 2001). The overexpressed protein contains an additional N-terminal sequence MGSSHHHHHHSSGRENLYFQ↓GH that includes the TEV protease recognition site. Selenomethionine-substituted AF0173 was overexpressed in the methionine-auxotroph *Escherichia coli* strain B834 (DE3) (Novagen). The cells were grown in supplemented M9 medium at 310 K. When the OD_{600} reached 0.6, protein expression was induced with 0.4 m*M* IPTG and continued overnight at 288 K. The cells were lysed by sonication in binding buffer (500 m*M* NaCl, 5% glycerol, 0.5 m*M* TCEP, 5 m*M* imidazole, 50 m*M* HEPES pH 7.5) with the addition of 0.5% nonionic detergent Igepal CA-630 (Sigma) and 1 m*M* each of PMSF and benzamidine. The lysate was clarified by centrifugation and passed through a DE52 (Whatman) column pre-equilibrated in binding buffer. The flowthrough fraction was applied onto a Ni–NTA agarose column (Qiagen) and washed extensively. The His-tagged AF0173 was eluted from the column in elution buffer (500 m*M* NaCl, 5% glycerol, 0.5 m*M* TCEP, 250 m*M* imidazole, 50 m*M* HEPES pH 7.5). Attempts to cleave the tag from the protein by treatment with recombinant His-tagged TEV protease were unsuccessful. The tag-containing AF0173 was dialyzed in 10 m*M* HEPES pH 7.5, 500 m*M* NaCl, concentrated and used for crystallization.

2.2. Crystallization and data collection

Crystals suitable for X-ray diffraction experiments were obtained by the hanging-drop vapor-diffusion method at 294 K. The drop was composed of 2 µl of a 22 mg ml^{-1} solution of SeMet AF0173 and 2 µl of well buffer containing 3 *M* ammonium acetate and 0.1 *M* sodium acetate at pH 4.2and 2% MPD. A crystal of dimensions 0.4 × 0.4 × 0.3 mm was cryoprotected by transferring it to crystallization buffer containing 2% glycerol, 2% ethylene glycol and 2% sucrose. Low-temperature (100 K) data were collected from an SeMet-substituted sample at the Structural Biology Center 19ID beamline at Argonne National Laboratory (Rosenbaum *et al.*, 2006) and were processed with *HKL*-2000 (Otwinowski & Minor, 1997). The protein crystallized in space group $P6_522$, with one protein subunit per asymmetric unit. The Matthews coefficient ($V_{\rm M}$) of the crystal was 4.6 Å^3 Da^{-1} and its estimated solvent content was 73%. Data sets were only collected at one wavelength (0.9793 Å). The choice of SAD experiment was a calculated one. All available crystals had reasonable size but diffracted very poorly. One crystal was selected for an experiment to characterize radiation damage. The exposure time chosen for the first data collection turned out to be too long and resulted in fast crystal decay, but did not improve the diffraction limit. The exposure time for the final SAD

experiment was chosen to minimize the influence of radiation damage by limiting the increase in the overall *B* factor during scaling and to maximize the amount of anomalous signal that could be obtained from a single data set by collecting a complete highly redundant data set. In addition, the 'solvability' of the structure was checked during data collection with *HKL*-3000 (a.k.a. *HKL*-2000_ph; Minor *et al.*, 2006). During the test for 'solvability', data from a not fully complete low-redundancy data set were used in *HKL*-3000 to obtain an initial map and model. The initial map and model were obtained before data collection ended and their quality allowed us to decide that collection of an additional data set and change of the wavelength were not necessary. Especially important for the success of the experiment were complete low-resolution data and high solvent content, both of which strongly affect the density-modification procedure. As noticed previously (Rice *et al.*, 2000), density modification is critical for the success of SAD experiments. Data-collection, phasing and refinement statistics are summarized in Table 1.

2.3. Structure solution and refinement

HKL-3000, a new software package that interacts with *SHELXD* (Schneider & Sheldrick, 2002), *SHELXE* (Sheldrick, 2002), *MLPHARE* (Otwinowski, 1991), *DM* (Cowtan & Main, 1993; Cowtan & Zhang, 1999), *CCP*4 (Collaborative Computational Project, Number 4, 1994), *SOLVE* (Terwilliger & Berendzen, 1999) and *RESOLVE* (Terwilliger, 2002, 2004), was used to solve the structure and build the initial model. As shown in Table 1, the data collected for AF0173 are complete and highly redundant. Both Se atoms present in the subunit were found. The N-terminal selenomethionine was well ordered and its Se atom was pronounced in the substructure solution. Three sites were used for phasing as the second SeMet had a double conformation. Initial phases obtained from *MLPHARE* were improved by density-modification procedures as implemented in *DM*. The unusually high solvent content (73%) was critical for the success of the density-modification procedures and was also instrumental in resolving the SAD phase ambiguity. Owing to the low resolution, the initial model obtained with *RESOLVE* contained fewer than half of the main-chain atoms. Substantial manual effort was then required in order to build the rest of the model. The model was rebuilt with a multi-cycle iterative procedure combining manual building with the program *O* (Jones, 2004), refinement with *REFMAC*5 (Murshudov *et al.*, 1997) and density modification with *DM*. The final model consisted of a single polypeptide chain of 162 amino acids that included nine residues of the uncut N-terminal tag and residues 1–153 of AF0173. No electron density was observed for the C-terminal segment (residues 154–163). The refinement of the model converged to an *R* factor of 23.7% ($R_{\rm free}$ = 28.5%). For $R_{\rm free}$ calculations, a subset of 4.8% of the reflections was used (Brünger, 1992). The refinement was performed with overall and individual isotropic *B* factors. The normal procedure for such low-resolution data would suggest only overall *B*-factor refinement, but individual *B*-factor refinement allowed the description of molecular mobility as presented in Fig. 1(*a*). In the final stage, TLS refinement was applied (Schomaker & Trueblood, 1998). For TLS refinement only one group was used, with the exclusion of the mobile loop (residues 56–64). The model validation was performed with *SFCHECK* (Vaguine *et al.*, 1999), *Map2mod* (Kirillova & Minor, 2006), *ADIT* (Yang *et al.*, 2004), *MOLPROBITY* and *KING* (Lovell *et al.*, 2003). 76% of the residues were positioned in the most favored regions of the Ramachandran plot and 24% were in additional allowed regions. There were no residues in the disallowed regions. Coordinates and structure factors were

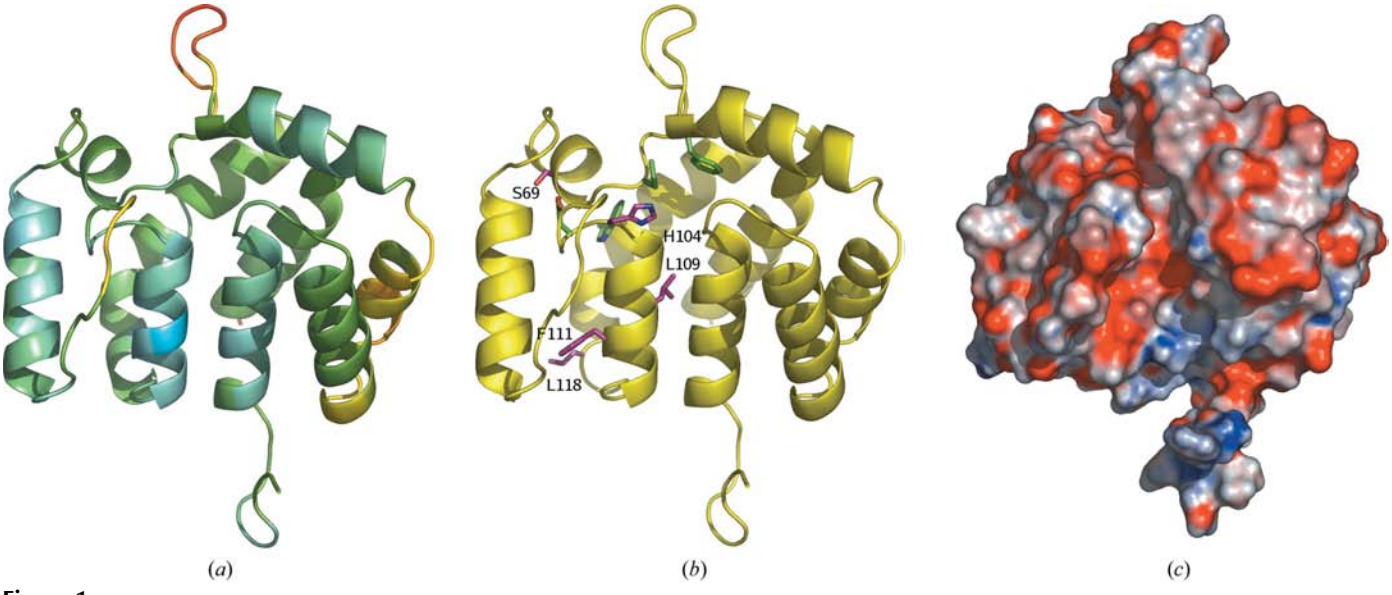

Figure 1
Structure of the AF0173 subunit. (*a*) Overall structure of AF0173 with *B*-factor distribution. Red corresponds to the highest *B* values and blue to the lowest. (*b*) Ribbon diagram with side chains belonging to invariant (green) and highly conserved (purple) residues. (*c*) Electrostatic surface of AF0173 (red corresponds to negative and blue to positive charges).

deposited in the PDB with accession code 1ze0 and were recently replaced by 2o9x.

3. Results and discussion

3.1. Overall structure

The subunit of AF0173 has an all-helical architecture (Fig. 1*b*). It contains ten α-helices that account for 60% of the protein residues. A disulfide bridge is formed between Cys64 and Cys132. The asymmetric unit of the crystal contains a monomer of AF0173, but a dimer is generated by applying a twofold crystallographic symmetry operation (Fig. 2). The existence of the dimer is consistent with the results of size-exclusion chromatography, which demonstrate a reversible equilibrium between the dimeric form of AF0173 and its monomeric form present only at low concentration.

A ten-residue segment of an uncut N-terminal tag that includes the TEV protease recognition site is visible in the electron density. This segment protrudes out of the globular subunit of AF0173 and binds in a cavity on the surface of the other subunit in the dimer (Fig. 2). The tag binding is responsible for dimer formation, providing 2685 Å^2 of the 3050 Å^2 total contact area between subunits (approximately 88%). The inaccessibility of the TEV protease recognition site in the dimer explains the failure to remove the His tag by proteolytic cleavage.

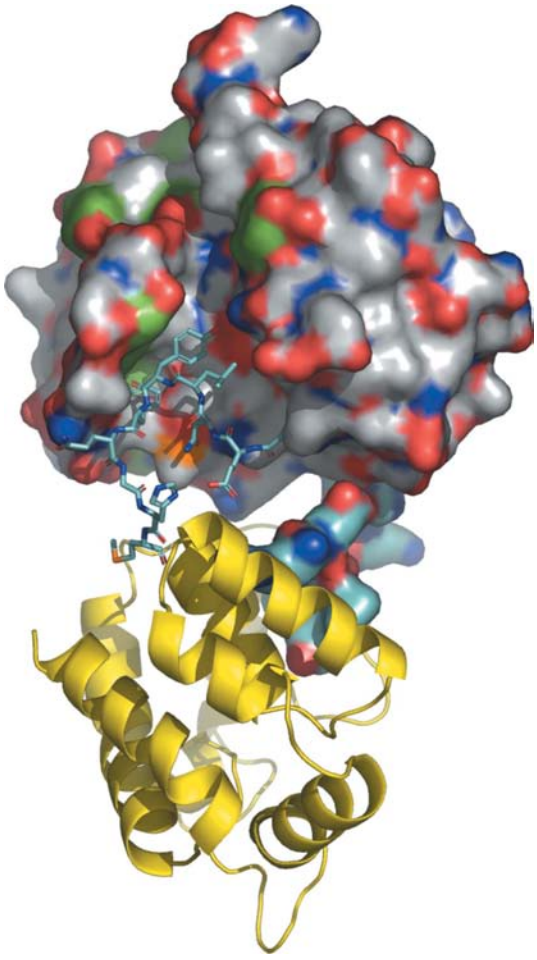

Figure 2
The dimer of AF0173. In the subunit represented as a surface, tyrosine residues are green and other residues are shown using standard color code (C, gray; O, red; N, blue; S, orange). Dimer formation is caused by the interaction of the His tags (cyan) and the funnel-like peptide-binding grooves.

The cavity on the AF0173 surface that interacts with the tag has a funnel-like shape, with a 15 Å wide and 9 Å deep mouth extending into a 7 Å wide neck that runs to the opposite pole of the subunit (Fig. 1*c*). The tag binds at the mouth of the cavity (Fig. 2). The binding is predominantly hydrophobic, with only three polar contacts and a major contribution from the tag Leu(−5)-Tyr(−4)-Phe(−3) segment.

3.2. Genetic, sequence and structure-similarity analysis

The function of AF0173 has not been experimentally determined. In genetic databases, AF0173 is annotated as a reductase-assembly protein, but its sequence similarity to biochemically characterized oxidoreductase components from other organisms is not sufficient for confident functional assignment. A *BLAST* (Altschul *et al.*, 1997) search resulted in more than 80 archaeal and bacterial sequence homologs, many of which are described as hypothetical proteins, components of anaerobic dehydrogenases or TorD cytoplasmic chaperones.

Structure-similarity searches with *FFAS* (Rychlewski *et al.*, 2000) and *DALI* (Holm & Sander, 1993) revealed the three-dimensional structure of AF0173 to be closely related to the structures of DmsD (formerly YnfI), a putative anaerobic dehydrogenase component from *Salmonella typhimurium* (PDB code 1s9u), and the TorD chaperone from *Shewanella massilia* (PDB code 1n1c; Tranier *et al.*, 2003).

The sequence identity between AF0173 and DmsD from *S. typhimurium* is 19% and superposition of 124 C$^\alpha$ atoms of these proteins results in a root-mean-square deviation (r.m.s.d.) of 1.6 Å. The sequence identity between AF0173 and TorD from *S. massilia* is 16%. TorD exists in both monomeric and dimeric forms (Tranier *et al.*, 2002); however, the three-dimensional structure has only been elucidated for the dimer. The dimeric form exhibits domain swapping between the two monomers (Tranier *et al.*, 2003). Each of the two globular modules of the dimer results from the association of the N-terminal moiety of one subunit with the C-teminal moiety of the other. Structure analysis and small-angle X-ray scattering data suggest that the globular module of the dimer is structurally similar to the monomer of TorD (Tranier *et al.*, 2003). Superposition of 109 C$^\alpha$ atoms of this globular module and AF0173 gives an r.m.s.d. of 1.8 Å.

The chaperones DmsD and TorD are involved in the biogenesis of the multisubunit membrane-bound enzymes DMSO reductase and TMAO reductase, respectively. These enzymes, which are found in bacteria and archaea, allow anaerobic respiration using *S*-oxides and *N*-oxides as terminal electron acceptors. Only bacterial enzymes have been characterized biochemically, but their structural similarity to

AF0173 and their similar organization of operons suggest that the bacterial and archaeal homologs share common features.

In the *A. fulgidus* genome, AF0173 belongs to the operon that also encodes the membrane, FeS-binding and molybdopterin-binding subunits of the molybdopterin oxidoreductase. The gene content of this operon is similar to that of the *dms* operon of DMSO reductase, but differs from that of the *tor* operon of TMAO reductase, which encodes the molybdopterin-binding subunit TorA, the membrane-bound *c*-type cytochrome TorC and the TorD chaperone. The *A. fulgidus* reductase has not been studied biochemically and its terminal electron acceptor is unknown. At the present state of knowledge, the AF0173 chaperone is most likely to be involved in the biogenesis of an archaeal DMSO reductase. However, further biochemical studies are needed to verify the specificity of this enzyme as *BLAST* searches reveal that its different subunits also show homology with nitrate, perchlorate, selenate, chlorate and polysulfide reductases.

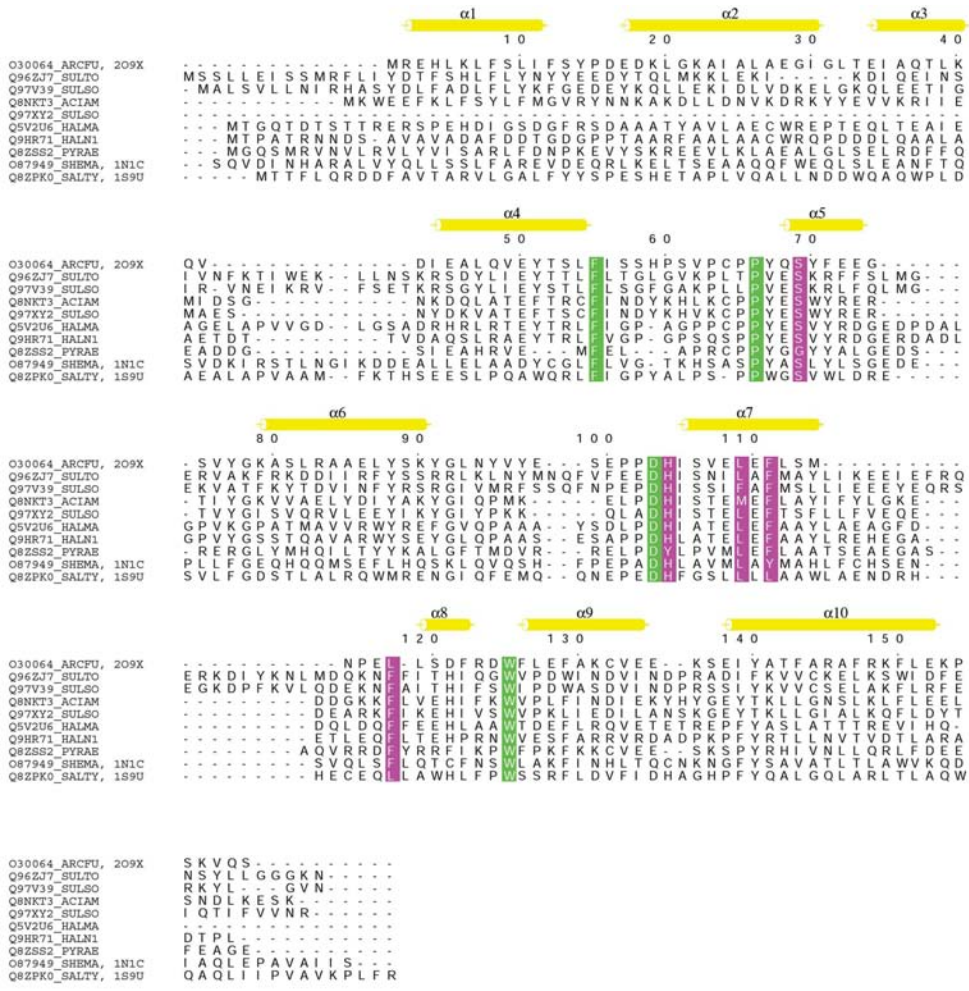

Figure 3
Structure-based sequence alignment of eight putative archaeal redox-enzyme maturation proteins and two bacterial REMPs present in the PDB. The proteins are identified by UniProtKB/TrEMBL accession Nos. and PDB codes in the cases where the three-dimensional structure is known. AF0173 is on the first line of the alignment. The residue numbering and secondary-structure elements for AF0173 are represented at the top of the alignment. Invariant (green) and highly conserved (purple) residues are highlighted. This figure was created using *ALSCRIPT* (Barton, 1993).

3.3. Localization of the twin-arginine binding site

The DmsD and TorD chaperones are the most extensively studied representatives of redox-enzyme maturation proteins (REMPs; Turner *et al.*, 2004). The REMPs are essential for the maturation of extracytoplasmic oxidoreductases, which contain redox-active cofactors acquired in the cytoplasm. After cofactor insertion, these oxidoreductases are exported through the membrane in the folded state by the twin-arginine translocation (Tat) system. The exported protein is expressed with an N-terminal signal sequence (S/T)-RR*X*FLK recognized by the Tat system. The REMP does not constitute part of the active assembled reductase, but is essential for its maturation. It binds to the unfolded reductase subunit in the absence of cofactor, facilitating cofactor insertion and preventing premature export. Upon cofactor insertion, REMP is released from the folded subunit that is exported by the Tat translocon (Sargent *et al.*, 2002; Turner *et al.*, 2004).

While none of the archaeal REMPs has been studied biochemically, the structural similarity of AF0173 to bacterial proteins and the presence of the twin-arginine signal sequence SRRDFIK in the coexpressed molybdopterin-binding reductase subunit AF0176 suggest that AF0173 and the well studied bacterial DmsD and TorD chaperones function in a similar manner. DmsD has been shown to specifically interact with the precursor but not with the mature forms of both DmsA and TorA (Oresnik *et al.*, 2001). In addition, TorD has been shown to interact with the unfolded but not the native mature portion of TorA (Pommier *et al.*, 1998). The TorD/DmsD chaperones have been suggested to contain binding sites for both the signal peptide and the unfolded mature portion of the molybdopterin-binding subunit (Sargent *et al.*, 2002). The signal peptides of DmsA and TorA do not possess obvious sequence similarity apart from the presence of the twin-arginine motif (Voordouw, 2000). Therefore, the twin-arginine recognition sequence appears to be solely responsible for the interaction between the signal peptide and TorD/DmsD. Since the twin-arginine motif is conserved in proteins exported by

the Tat system, the chaperone residues that interact with this motif are also likely to be conserved. The unfolding of the protein exposes the buried nonpolar amino acids and the chaperone residues interacting with the unfolded mature molybdopterin-binding subunit are likely to be hydrophobic.

A search with *BLAST* and *PROFUNC* (Laskowski *et al.*, 2005) found eight AF0173 homologues in archaeal genomes. The genes encoding seven of them belong to operons that encode a putative oxidoreductase with a twin-arginine signal sequence on the molybdopterin-binding subunit. An alignment of the sequences of these seven archaeal and two bacterial REMPs with known structures is presented in Fig. 3. Despite the low sequence similarity between most of these proteins, four residues, Phe55, Pro66, Asp103 and Trp125 (AF0173 numbering), are invariant and five more residues are highly conserved (Fig. 1*b*).

Several lines of evidence indicate that the twin-arginine motif of the molybdopterin-binding subunit binds in the funnel-shaped cavity that interacts with the N-terminal tag of the other subunit in the AF0173 dimer. All nine invariant and conserved residues in the AF0173 structure are located within or in close vicinity of this cavity (Fig. 2). Two of the three surface-accessible invariant residues, Phe55 and Pro66, are located in the neck of the cavity. The invariant Asp103 and Trp125 are positioned behind Pro66 and the conserved His104 and Ile105. All four conserved residues are located on the helix that forms the bottom of the cavity and the side chain of the surface-accessible Phe111 interacts with the side chain of Phe(−3) of the tag.

The funnel-shaped cavity contains several patches of hydrophobic residues and is particularly favored by tyrosine residues. Eight of the ten tyrosines present in AF0173 are located in or near the cavity (Fig. 2). Tyrosine side chains are very well suited to mediating recognition at protein–protein interfaces (Fellouse *et al.*, 2004). They are capable of providing multiple types of contacts, including hydrogen bonds, hydrophobic interactions and electrostatic interactions. In addition, tyrosine side chains avoid electrostatic repulsion and can adjust their position by changing a few torsion angles. Tyrosine residues account for about 10% of the total complementarity-determining regions and about 25% of the antigen contacts in antigen-binding fragments (Fabs; Mian *et al.*, 1991). Three of the AF0173 tyrosine residues located in the funnel-shaped cavity, Tyr88, Tyr95 and Tyr97, form hydrophobic contacts with the side chains of Tyr(−4) and Phe(−3) of the tag.

The structure of the DmsD chaperone from *S. typhimurium* indicates the increased mobility of the residues in the region corresponding to the funnel-shaped cavity of AF0173. The only segment disordered in DmsD is loop 117–122 at the flank of a topologically similar cavity that also contains patches of hydrophobic residues, including tyrosines. This cavity contains a bound PEG 400 molecule in the DmsD structure.

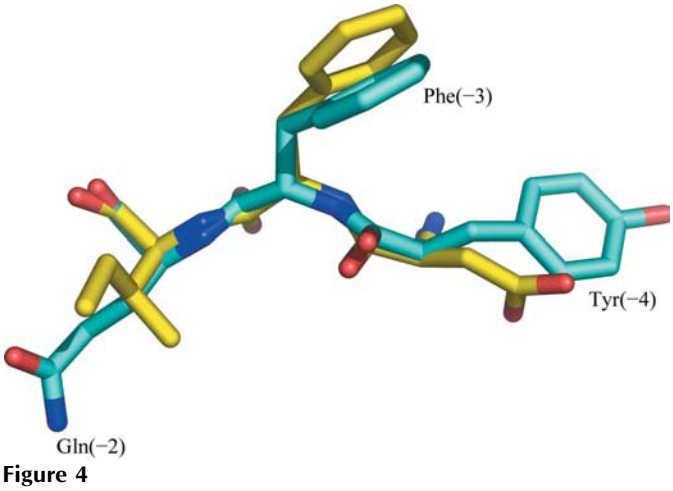

Figure 4
Comparison of the His tag [residues Tyr(−4)–Gln(−2)] of AF0173 (cyan) and model of a twin-arginine recognition-peptide conformation (yellow). High conformational similarity suggests that part of the twin-arginine motif can be bound by the residues forming the funnel-like cavity.

3.4. Structure-based modeling of twin-arginine motif binding

We used the HMMSTR/Rosetta server (Bystroff *et al.*, 2000; Bystroff & Shao, 2002; http://www.bioinfo.rpi.edu/~bystrc/hmmstr) to model the three-dimensional structure of the 20-residue signal peptide of the molybdopterin-binding subunit AF0176 that contains the twin-arginine motif recognized by AF0173. The modeling indicates that the observed coordination of the N-terminal tag segment in the funnel-shaped cavity of AF0173 partially mimics the binding of the twin-arginine motif.

The twin-arginine segment SRRDFIK is modeled as a nearly identical β-turn in all five of the best structures obtained from HMMSTR/Rosetta, although other regions of the signal peptide show some differences. The modeled conformation of the Asp-Phe-Ile fragment is nearly identical to the conformation of the Tyr(−4)-Phe(−3)-Gln(−2) tripeptide of the N-terminal tag of AF0173 that provides the major contribution to the tag binding in the funnel-shaped cavity (Fig. 4).

If the Asp-Phe-Ile tripeptide fragment is placed in the funnel-shaped cavity of AF0173 instead of the Tyr-Phe-Gln segment, it retains both main-chain hydrogen bonds and the Phe side chain fits into the same hydrophobic pocket. The third tag-coordinating hydrogen bond formed by the side chains of Tyr(−4) and Ser106 can be replaced by a bond between the side chains of Asp and Ser9, Ser13, Tyr14, Lys19 or Ser106 as well as main-chain atoms. Thus, the C-terminal part of the twin-arginine motif can readily be accommodated in the mouth of the funnel-shaped cavity of AF0173. We suggest that the bound twin-arginine segment is stretched along this cavity so that its N-terminal part including two arginines resides in the neck of the funnel, which contains a cluster of invariant residues. The N-terminal part of the twin-arginine motif can be shielded by loop 57–63 of AF0173, which protrudes out of the globular subunit at the end of the funnel neck and is present in other REMPs according to structure-based sequence alignment (Fig. 3). In addition to the residues of the twin-arginine motif, the wide mouth of the cavity may also accommodate residues exposed in the unfolded molybdopterin-binding subunit, which has been shown to bind to

Table 2
Structures determined by the SAD method at 3.0 Å resolution or lower.

Information that is not available is marked by a —. Data in parentheses are for the highest resolution shells.

PDB code	Resolution (Å)	R/R_{free} (%)	Solvent content (%)	NCS	No. of amino acids per ASU	Type of anomalous scatterer	No. of anomalous scatterers†	$I/\sigma(I)$
2o9x	3.40	24/28	73.0	No	172	Se	2, 2	37.0 (5.2)
1rhz	3.50	25/33	72.0	No	563	Se	13, 15	30.9 (2.74)
1rh5	3.20	24/29						26.9 (2.24)
1xa6	3.20	25/30	80.0	No	466	Zn, Xe	2 + 1	34.5 (3.8)
1y7e	3.20	20/26	62.0	No	458	Se	—, 7	—
1q1n	3.15	24/28	78.0	No	360	Zn	2	—
1nqh	3.10	22/27	62.0	No	594	Se	8, 8	28.2 (5.2)
1mqs	3.00	26/29	65.5	No	721	Se	10, 11	13.3 (10.9)
1nhy	3.00	24/24	67.0	No	219	Se	—, 4	50.1 (7.7)
1r8i	3.00	26/29	68.4	No	213	Se	12, 13	21.4 (3.2)
1s48	3.00	23/28	72.1	No	609	Se	12, 12	25.0 (3.3)
1z2r	4.20	28/33	74.1	Yes	2328	Hg	—	11.0 (—)
1zcd	3.45	32/32	73.6	Yes	776	Se	—, 22	30.5 (12.9)
1sru	3.30	29/31	66.4	Yes	452	Se	—, 12	—
1xri	3.30	20/25	73.8	Yes	302	Se	—, 6	20.8 (10.3)
1t8b	3.23	21/25	51.0	Yes	454	Se	58, 60	14.0 (2.1)
1q9c	3.21	26/29	55.8	Yes	1719	Se	30, 54	20.1 (4.0)
1pm3	3.15	26/29	76.5	Yes	194	Se	6, 8	8.9 (—)
1tyg	3.15	24/31	62.0	Yes	680	Se	14, 16	15.0 (3.8)
1yrh	3.11	20/23	52.9	Yes	1688	Se	—, 48	5.6 (4.5)
1xb4	3.10	33/33	49.0	Yes	808	Se	—, 20	12.3 (4.9)
1z3h	3.10	24/29	50.9	Yes	1936	Se	—, 34	21.9 (5.9)
1xiq	3.05	23/29	45.3	Yes	942	Se	36	7.1 (1.9)
1yvk	3.01	24/28	55.0	Yes	652	Se	—, 4	14.01 (2.32)
1tlw	3.10	27/30	82.0	Yes	556	Se	10, 10	12.1 (4.7)
1tlz	3.10	28/30						
1tly	3.01	26/29						
1yls	3.00	21/23	63.0	Yes	98	Se	—, 12	11.2 (2.6)
1wb3	3.20	35/37	71.0	Yes	1928	Hg	16	25.8 (1.7)
1wb2	3.10	34/35						18.3 (2.7)
1wb1	3.00	32/33						14.7 (1.5)
1yvl	3.00	24/28	59.6	Yes	1376	Au	—, 2	32.5 (5.0)
1z2c	3.00	21/29	59.9	Yes	1152	Se	—, 48	—
1ukl	3.00	24/30	62.5	Yes	1996	Se	—, 12	27.7 (3.0)
1lm7	3.00	25/30	44.7	Yes	496	Se	9, 18	32.2 (15.2)
1x3w	3.00	24/27	81.0	Yes	407	Zn, Se	5, 6	—
1ltl	3.00	25/30	64.1	Yes	1674	Zn	6, 6	—
1yq1	3.00	22/30	55.9	Yes	416	Se	—, 6	13.2 (—)

† For SeMet-substituted proteins, numbers of localized and possible anomalous scatterers are given.

REMPs. The structural studies of the complex of AF0173 with the twin-arginine motif necessary to the detailed characterization of the binding are under way in our laboratory.

3.5. SAD at low resolution

The use of anomalous diffraction has become a dominant phasing method in macromolecular crystallography for *de novo* structure determination. Most often, the MAD technique (Hendrickson, 1991; Hendrickson *et al.*, 1990) is used to obtain initial phases, but in most cases single-wavelength anomalous diffraction (SAD) can suffice (Dauter *et al.*, 2002; Rice *et al.*, 2000; Fig. 5). The SAD technique is extremely useful, for example, when diffraction data are obtained from a fixed-wavelength source or if a crystal decays quickly during data collection and it is only possible to obtain one data set of high quality. However, the use of a single wavelength is not always sufficient to determine the structure, so preliminary phasing and data analysis should be performed immediately after the first data set has been collected (Dauter, 2002). Ideally, as in the case of AF0173, the initial phasing provides sufficient information and further data collection can be omitted.

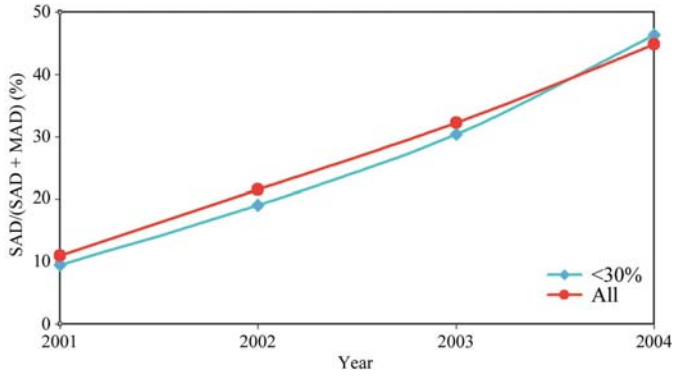

Figure 5
The fraction of structures solved by SAD in a given year. The blue line represents the fraction of structures with lower than 30% sequence identity to those deposited previously.

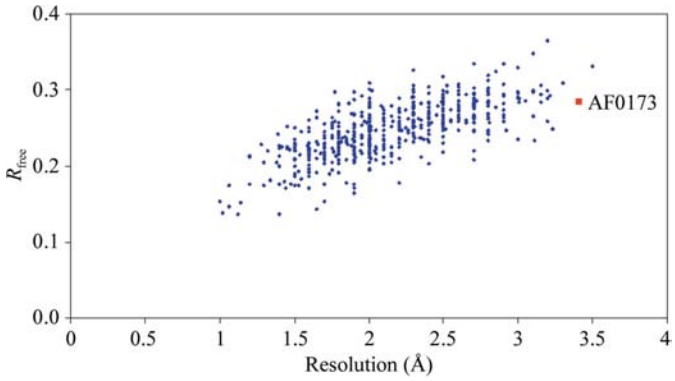

Figure 6
R_{free} plotted against resolution for all structures solved by SAD and deposited in the PDB. AF0173 is shown in red.

The SAD method is used even more rarely for protein crystals that diffract to low resolution, in particular those that do not display noncrystallographic symmetry. There are only 38 structures of the 595 solved by SAD and deposited in the PDB prior to that described in this paper had a resolution equal or lower than 3 Å (Table 2) (compared with 89 of 2180 for MAD). Excluding highly homologous structures (1rhz and 1rh5; 1wb1, 1wb2 and 1wb3; 1tly, 1tlw and 1tlz), the number of low-resolution SAD cases is reduced to 32. Of these nonredundant structures, only ten, including AF0173, do not possess NCS.

The success of both the SAD and MAD methods benefits greatly from the process of phase improvement by solvent flattening and noncrystallographic averaging, when possible. Unlike MAD, SAD requires a mechanism to resolve the phase ambiguity. The utility of these methods to improve the phases depends on the resolution, the solvent content and the order of noncrystallographic symmetry. The impact of high NCS increases for structure solution when the initial phases are very poor (Rossmann & Arnold, 2001) and even relatively modest NCS electron-density averaging can significantly improve the initial phases.

Almost all the low-resolution structures solved by SAD have a high solvent content; the average value of the solvent content for low-resolution SAD structures is 66%, compared with an average of 54% for all protein structures deposited in the PDB (Kantardjieff & Rupp, 2003). Interestingly, none of the ten low-resolution SAD structures lacking NCS had less than 50% solvent content. The average solvent content for low-resolution structures solved with NCS is 62%, compared with 70% when NCS was not present. In the absence of NCS, the high solvent content (~73% for AF0173) was probably the most important property of the crystal in improving the phases and the quality of the map.

Most of the structures solved by SAD at low resolution were determined using the anomalous signal from selenium incorporated into proteins as SeMet. For any phasing procedure, the greater the ratio of ordered anomalous scatterers to the number of residues, the more easily and more precisely phasing can be performed. AF0173, with only two anomalous scatterers in almost 20 kDa of protein, also appears to be an extremely SAD case.

4. Conclusions

The crystal structure reported here represents a redox-enzyme maturation protein from *A. fulgidus*. The structural similarity between archaeal and bacterial REMPs, the conservation of the residues in a putative binding site and the similar organization of the operons indicate that REMPs function in a similar fashion in both domains of prokaryotes. An unexpected structural similarity between the segment of the N-teminal tag of overexpressed AF0173 bound by the second subunit in the dimer and the twin-arginine motif allowed the localization of the putative binding site of the REMPs.

The structure turns out to be an extreme SAD case of protein crystallography, taking into account the resolution (3.4 Å), the absence of NCS and the presence of only two anomalously scattering atoms. The quality of the structural model compares favorably with other SAD low-resolution structures (Table 2; Fig. 6).

The results described in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by University of Chicago Argonne, LLC for the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. The authors would like to thank Andrzej Joachimiak and members of Structural Biology Center at Advanced Photon Source and Midwest Center for Structural Genomics for help and discussions. We would like also thank H. Zheng and H. Qiu for help with PDB statistical data analysis and F. W. Studier for providing the autoinduction protocol for the overexpression. We would like also to thank Alex Wlodawer for reading the manuscripts and making valuable comments. This work was supported by NIGMS grant P-50-GM62414.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Berks, B. C., Sargent, F. & Palmer, T. (2000). *Mol. Microbiol.* **35**, 260–274.
Blasco, F., Dos Santos, J. P., Magalon, A., Frixon, C., Guigliarelli, B., Santini, C. L. & Giordano, G. (1998). *Mol. Microbiol.* **28**, 435–447.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Bystroff, C. & Shao, Y. (2002). *Bioinformatics*, **18**, S54–S61.
Bystroff, C., Thorsson, V. & Baker, D. (2000). *J. Mol. Biol.* **301**, 173–190.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Cowtan, K. D. & Zhang, K.-Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.
Dauter, Z. (2002). *Acta Cryst.* D**58**, 1958–1967.
Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.

Fellouse, F. A., Wiesmann, C. & Sidhu, S. S. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 12467–12472.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
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.
Ilbert, M., Mejean, V., Giudici-Orticoni, M. T., Samama, J. P. & Iobbi-Nivol, C. (2003). *J. Biol. Chem.* **278**, 28787–28792.
Jones, T. A. (2004). *Acta Cryst.* D**60**, 2115–2125.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Kirillova, O. & Minor, W. (2006). *Bioinformatics*, **22**, 1660–1661.
Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Mian, I. S., Bradwell, A. R. & Olson, A. J. (1991). *J. Mol. Biol.* **217**, 133–151.
Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Oresnik, I. J., Ladner, C. L. & Turner, R. J. (2001). *Mol. Microbiol.* **40**, 323–331.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pommier, J., Mejean, V., Giordano, G. & Iobbi-Nivol, C. (1998). *J. Biol. Chem.* **273**, 16615–16620.
Ray, N., Oates, J., Turner, R. J. & Robinson, C. (2003). *FEBS Lett.* **534**, 156–160.
Rice, L. M., Earnest, T. N. & Brünger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Rosenbaum, G., Alkire, R. W., Evans, G., Rotella, F. J., Lazarski, K., Zhang, R. G., Ginell, S. L., Duke, N., Naday, I., Lazarz, J., Molitsky, M. J., Keefe, L., Gonczy, J., Rock, L., Sanishvili, R., Walsh, M. A., Westbrook, E. & Joachimiak, A. (2006). *J. Synchrotron Rad.* **13**, 30–45.
Rossmann, M. & Arnold, E. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 279–292. Dordrecht: Kluwer Academic Publishers.
Rychlewski, L., Jaroszewski, L., Li, W. & Godzik, A. (2000). *Protein Sci.* **9**, 232–241.
Santini, C. L., Ize, B., Chanal, A., Muller, M., Giordano, G. & Wu, L. F. (1998). *EMBO J.* **17**, 101–112.
Sargent, F., Berks, B. C. & Palmer, T. (2002). *Arch. Microbiol.* **178**, 77–84.
Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C. & Palmer, T. (1998). *EMBO J.* **17**, 3640–3650.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schomaker, V. & Trueblood, K. N. (1998). *Acta Cryst.* B**54**, 507–514.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 2082–2086.
Terwilliger, T. C. (2004). *J. Synchrotron Rad.* **11**, 49–52.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Tranier, S., Iobbi-Nivol, C., Birck, C., Ilbert, M., Mortier-Barriere, I., Mejean, V. & Samama, J. P. (2003). *Structure*, **11**, 165–174.
Tranier, S., Mortier-Barriere, I., Ilbert, M., Birck, C., Iobbi-Nivol, C., Mejean, V. & Samama, J. P. (2002). *Protein Sci.* **11**, 2148–2157.
Turner, R. J., Papish, A. L. & Sargent, F. (2004). *Can. J. Microbiol.* **50**, 225–238.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Voordouw, G. (2000). *Biophys. Chem.* **86**, 131–140.
Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., Cole, J. A. & Turner, R. J. (1998). *Cell*, **93**, 93–101.
Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* D**60**, 1833–1839.
Zhang, R. G., Skarina, T., Katz, J. E., Beasley, S., Khachatryan, A., Vyas, S., Arrowsmith, C. H., Clarke, S., Edwards, A., Joachimiak, A. & Savchenko, A. (2001). *Structure*, **9**, 1095–1106.