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

Structure of SAICAR synthase from *Thermotoga maritima* at 2.2 Å reveals an unusual covalent dimer

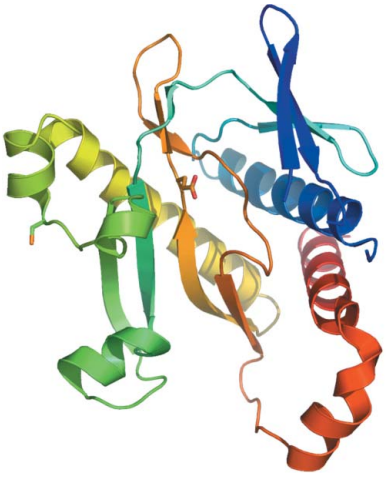
Rongguang Zhang,[a] Tatiana Skarina,[b] Elena Evdokimova,[b] Aled Edwards,[b,c] Alexei Savchenko,[c] Roman Laskowski,[d] Marianne E. Cuff[a] and Andrzej Joachimiak[a]*

[a]Biosciences Division, Midwest Center for Structural Genomics, Structural Biology Center, Argonne National Laboratory, Argonne, IL 60439, USA, [b]Ontario Centre for Structural Proteomics, University of Toronto, University Health Network, Toronto, Ontario M5G 1L7, Canada, [c]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada, and [d]Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

Correspondence e-mail: andrzejj@anl.gov

Received 28 October 2005
Accepted 14 March 2006

PDB Reference: SAICAR synthase, 1kut.



As a part of a structural genomics program, the 2.2 Å resolution crystal structure of the PurC gene product from *Thermotoga maritima* has been solved. This 26.2 kDa protein belongs to the phophoribosylaminoimidazole-succinocarboxamide or SAICAR synthase family of enzymes, the members of which are involved in *de novo* purine biosynthesis. SAICAR synthase can be divided into three subdomains: two $\alpha+\beta$ regions exhibiting structural homology with ATP-binding proteins and a carboxy-terminal subdomain of two α-helices. The asymmetric unit contains two copies of the protein which are covalently linked by a disulfide bond between Cys126(*A*) and Cys126(*B*). This 230-amino-acid protein exhibits high structural homology with SAICAR synthase from baker's yeast. The protein structure is described and compared with that of the ATP–SAICAR synthase complex from yeast.

1. Introduction

Nucleotides are central players in the metabolism of the cell. Not only do they play critical roles in energy metabolism and DNA replication, but they also serve as regulators, signal molecules and enzyme cofactors. As such, the enzymes involved in their biosynthesis may serve as important action points for antimicrobial and anticancer drugs. The *de novo* synthesis of purine and pyrimidine nucleotides is largely invariant among known organisms in biology (Mathews & Van Holde, 1996). Purine biosynthesis proceeds from 5-phopho-α-D-ribosyl-1-pyrophosphate (PRPP) to inosinic acid (IMP) in a multi-step pathway. Phosphoribosylaminoimidazole-succinocarboxamide (SAICAR) synthase catalyzes the ATP-dependent transfer of an aspartate molecule to the carboxyl group of 4-carboxy-5-amino-imidazole ribonucleotide. In most prokaryotes, plants and yeast, SAICAR synthase is a monofunctional enzyme. In animals, it is the amino-terminal domain of a larger protein that also exhibits phosphoribosylaminoimidazole carboxylase (AIRC) activity.

Previously, the structure of SAICAR synthase from yeast has been determined to 1.9 Å resolution (Levdikov *et al.*, 1998). It is a 34.5 kDa monomer of 306 amino acids folded into three subdomains: two $\alpha+\beta$ domains and an α-helical carboxy-terminus. A deep central cleft dominates the surface topology, to which all three subdomains contribute residues. By locating highly conserved amino acids and solving the structure of an ATP–SAICAR synthase complex, Ledvikov and coworkers identified the active site as being within this cleft. Most interactions between the adenine base of ATP and the protein involve main-chain atoms, with the exception of Glu219.

As a part of our structural genomics program, we have solved the 2.2 Å resolution crystal structure of another member of the SAICAR synthase family, that from *Thermotoga maritima*. This 26.2 kDa protein is considerably smaller than its relative from yeast (34.5 kDa), with which it shares 26% primary sequence identity and an *E* value of 0.02 (FASTA/SAS). Nonetheless, the proteins share a common topology and fold. However, in the *T. maritima* enzyme the crystallographic asymmetric unit uniquely contains two copies of SAICAR synthase that are covalently linked by a disulfide bond between Cys126(*A*) and Cys126(*B*). The ATP-binding site can be located based upon high structural similarity to the yeast enzyme and other members of the ATP-binding superfamily.

Table 1
Summary of experimental crystallographic data.

(*a*) Crystallographic data.

Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 63.51, b = 43.07, c = 80.22, β = 92.30
Molecular weight (kDa)	26.2
No. of residues	230
Molecules per ASU	2
SeMet residues per ASU	10 (8 located)

(*b*) MAD data.

	Edge	Peak	Remote
Wavelength (Å)	0.9793	0.9791	0.9520
Resolution limit (Å)	2.2	2.2	2.2
No. of unique reflections	22291	21602	21094
Completeness (%)	93.3	93.9	93.7
$R_{\rm merge}$† (%)	8.6	9.2	8.4

(*c*) Phasing statistics.

	Centric		Acentric		All	
Resolution (Å)	FOM‡	Phasing power§	FOM‡	Phasing power§	FOM‡	Phasing power§
50–2.5 (26531 reflections)	0.799	2.261	0.613	2.099	0.620	2.106

† $R_{\rm merge} = \sum\sum|I_j - I_m|/\sum\sum I_j$, where I_j is the intensity of the measured reflection and I_m is the mean intensity of all symmetry-related reflections. ‡ Figure of merit from MAD phasing. § Phasing power = $F_{\rm H}/E_{\rm r.m.s.}$, where $F_{\rm H}$ is the heavy-atom structure factor and $E_{\rm r.m.s.}$ is the residual lack of closure.

2. Materials and methods

2.1. Protein cloning, expression and purification

The TM1243 gene was subcloned, expressed and its product purified and screened for crystallization as described previously (Zhang *et al.*, 2001).

2.2. Protein crystallization

Crystals for X-ray diffraction data collection were obtained from hanging-drop vapor-diffusion conditions containing 2 µl of the SeMet derivative of the protein plus 2 µl 30% PEG 4000, 0.15 *M* $MgCl_2$, 0.1 *M* Tris pH 8.5 over 2–5 d at 294 K. The crystals were flash-frozen with crystallization buffer complemented with 25% ethylene glycol. Diffraction intensity data (Table 1) were collected using the Advanced Photon Source (APS) beamline 19ID at Argonne National Laboratory.

2.3. Determination of SAICAR synthase oligomeric state using size-exclusion chromatography

FPLC size-exclusion chromatography was performed on a Superdex-200 column (10/300 mm) pre-equilibrated with 10 m*M* HEPES pH 7.5, 0.5 *M* NaCl. The column was calibrated with cytochrome *c* (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (150 kDa), β-amylase (200 kDa) and blue dextran (2000 kDa). A 25 µl SAICAR synthase protein sample at 2 mg ml^{-1} concentration or premixed with standard proteins was centrifuged at 14 000 rev min^{-1} for 10 min before being injected into the column through a 20 µl injection loop. Filtration was carried out at 277 K at a flow rate of 1 ml min^{-1}. The eluted proteins were detected by measuring the absorbance at 280 nm. SAICAR synthase elutes as a protein of approximately 53 kDa. This is nearly twice its calculated molecular weight and indicates a dimeric state in solution.

2.4. Data collection

Diffraction data were collected at 100 K at beamline 19ID of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. Three-wavelength inverse-beam MAD data (peak, 12.6610 keV, 0.9793 Å; inflection point, 12.6640 keV, 0.9791 Å; high-energy remote, 13.0240 keV, 0.9520 Å) were collected from an SeMet-labeled protein crystal to a Bragg spacing of 2.2 Å. One crystal (0.5 × 0.2 × 0.6 mm) was used for data collection at 100 K using 3 s exposures and 1° oscillations at a 200 mm crystal-to-detector distance. The total oscillation range was 160° as predicted using the strategy module within the *HKL*2000 package (Otwinowski & Minor, 1997). The crystal belonged to space group $P2_1$, with unit-cell parameters a = 63.51, b = 43.07, c = 80.22 Å, β = 92.30°. All data were processed and scaled within *HKL*2000 to $R_{\rm merge}$ values of 8.6, 9.2 and 8.4% for the inflection point, peak and remote data, respectively (Table 1).

2.5. Structure determination and refinement

The structure was determined by MAD phasing and refined to 2.2 Å against the averaged peak data using the *Crystallography & NMR System* (*CNS*; Brünger *et al.*, 1998). The initial model was built manually using *QUANTA* (Oldfield, 1996). The asymmetric unit dimer was refined to an R factor of 24.5% and an $R_{\rm free}$ of 28.1% using peak data from 10 to 2.2 Å Bragg spacing. The final model contains 3502 non-H protein atoms and 81 water molecules, with an average overall B value of 30 $Å^2$. Further refinement details can be found in Table 2. Both monomers have a missing surface loop and unmodeled

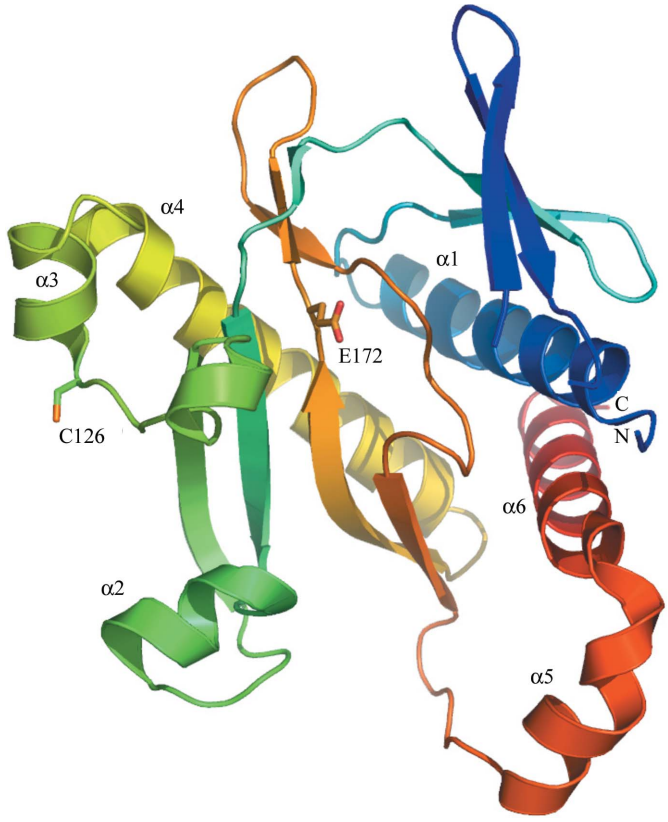


Figure 1
Ribbon schematic of the *T. maritima* SAICAR synthase monomer. The peptide chain is colored blue to red from the N- to the C-terminus. Conserved residue Glu172 is labeled E172; Cys126 is labeled C126 and is cross-linked to its counterpart in the second molecule of the asymmetric unit. All ribbon figures were generated using *PyMOL* (DeLano, 2002).

terminal residues, as they were unobserved in the electron-density maps. These include the N-terminus (*A*1–*A*6) and *A*25–*A*35 of the first monomer and *B*1, *B*27–*B*31 and *B*199–*B*201 of the second. Chain *A* has 213 modelled residues and chain *B* has 222. The main-chain torsion angles for 89.7% of the residues fall within the most favoured regions of a Ramachandran plot, 8.7% fall in additional allowed areas and the remaining five residues fall in generously allowed areas. No unusual geometries were detected by *PROCHECK* (Laskowski, 2001).

3. Results and discussion

3.1. Structure description

SAICAR synthase from *T. maritima* is a 26.2 kDa globular protein composed of segregated α-helical and antiparallel β-sheet regions with approximate dimensions 55 × 45 × 35 Å (Fig. 1). The secondary-structural elements are labeled and numbered as they appear along the polypeptide in Fig. 2, which is a structure-based primary sequence alignment with SAICAR synthase from baker's

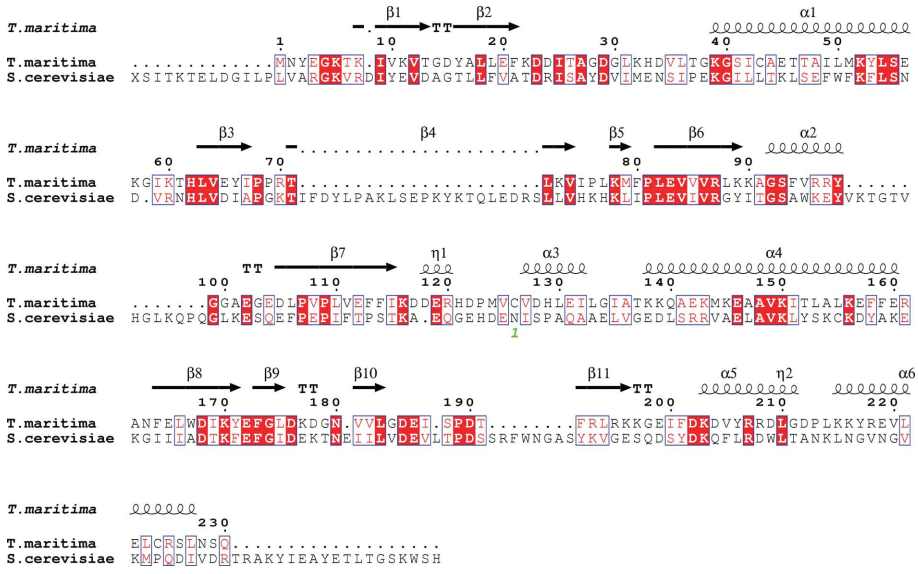


Figure 2
Structure-based primary sequence alignment of SAICAR synthases from *T. maritima* and baker's yeast (*S. cerevisia*). Identical residues are highlighted and secondary-structure motifs are labeled above: β-strands are depicted as arrows and α-helices as coils. The numbering is that for the *T. maritima* enzyme. This figure was generated using the *ESPript* utility (Gouet *et al.*, 1999).

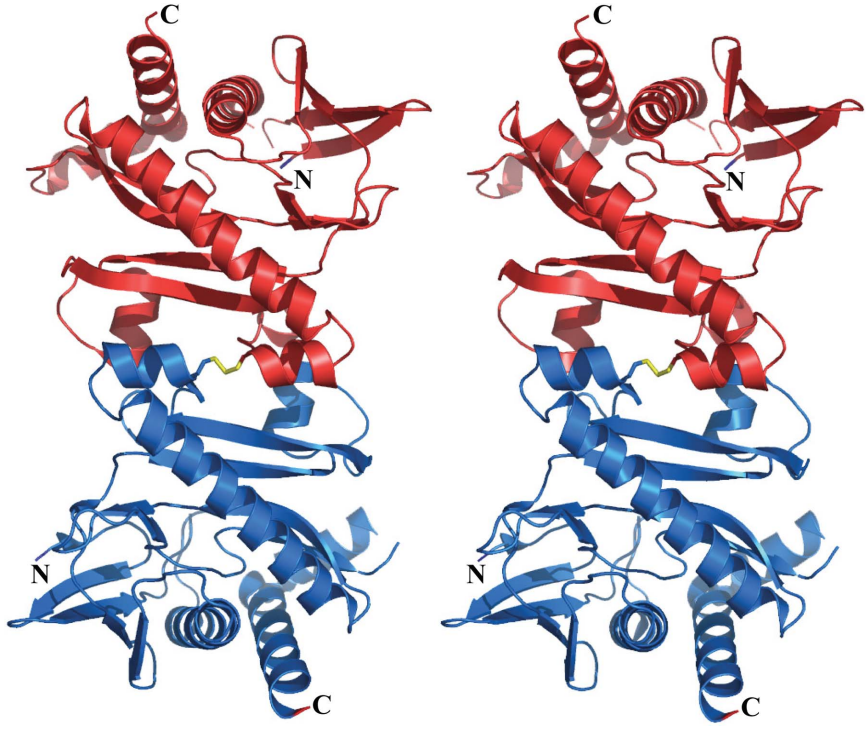


Figure 3
Dimeric SAICAR synthase. The covalent dimer is pictured here in stereo with one peptide chain in red and the other in blue. The amino-termini are labeled N and the carboxy-termini are labeled C. The disulfide bridge is located across a (noncrystallographic) twofold axis and is shown as yellow sticks.

yeast (Gouet *et al.*, 1999). The monomer can be divided into the same three subdomains as the yeast enzyme: two $\alpha+\beta$ regions exhibiting structural homology with ATP-grasp fold proteins and the protein kinase superfamily and a final carboxy-terminal subdomain of two α-helices. Each subdomain can be identified in relation to the three longest α-helices in the molecule: $\alpha1$ (19 residues), $\alpha4$ (24 residues) and $\alpha6$ (~15 residues). The two $\alpha+\beta$ subdomains each contains a major helix ($\alpha1$ or $\alpha4$) and an antiparallel four- to five-stranded β-sheet. As they occur in the molecule, the planes defined by the two β-sheets are roughly perpendicular to each other. A deep central cleft occurs between the two subdomains, with both contributing amino-acid residues to the ATP-binding site. The final subdomain is an α-helical 'L' at the C-terminus, with $\alpha6$ stacking alongside $\alpha1$ to complete the protein.

3.2. *T. maritima* SAICAR synthase is a covalent dimer

Two copies of *T. maritima* SAICAR synthase reside in the crystallographic asymmetric unit. Pictured in Fig. 3, they form an elongated dimer of roughly 85 × 40 × 30 Å. The amino- and carboxy-termini occur at the far extremes of the dimer and on opposite sides to each other. The central $\alpha+\beta$ subdomain forms the dimer core and helix $\alpha4$ lies obliquely across the C-terminal side of the dimer. Short helix $\alpha3$ straddles the noncrystallographic twofold symmetry axis, with Cys126(A) forming a disulfide bond with Cys126(B) which spans the rotation axis.

One monomer possesses a solvent-accessible surface area of approximately 11 913 Å^2. Upon dimer formation, a total of 758 Å^2 is lost, as calculated by *The Quaternary Structure File Server* (*PQS*) and a solvation free energy of −108 kJ mol^{-1} is gained (Henrick & Thornton, 1998). 8–9 Å to either side of the disulfide bridge, salt bridges occur between Asp128(A) and Lys145(B) and between Asp128(B) and Lys145(A). The dimer interface is a mixture of hydrophobic and hydrophilic interactions, with a few bound water molecules directly or indirectly hydrogen bonding to each monomer. We observe that a dimer occurs both in solution at 277 K and in the crystal grown at room temperature.

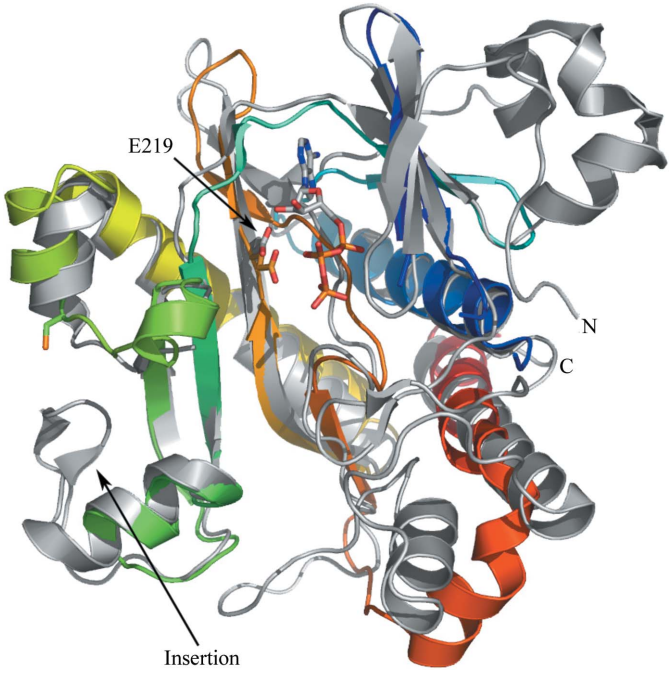


Figure 4
Structure superposition of SAICAR synthase monomers from *T. maritima* (rainbow colored from blue to red) and an ATP-bound complex (PDB code 1odb) from baker's yeast (in gray). The molecules are in an equivalent orientation to that in Fig. 1. The ATP and Glu219 (E219), one of the only side chains hydrogen bonding to the ATP, are shown in stick representation. In addition, Cys126 is also shown in stick representation and an insertion likely to obstruct dimer formation is indicated. This superposition was achieved with the *Secondary Structure Matching* tool for protein structure comparison (*SSM*; Krissinel & Henrick, 2004).

Table 2
Summary of crystallographic refinement statistics.

Resolution range (Å)	10–2.2
No. of reflections	35023
σ cutoff	0.0
R value (%)	24.5
Free R value (%)	28.1
R.m.s. deviations from ideal geometry	
Bond length (Å)	0.008
Angle (°)	1.50
Dihedral (°)	23.1
Improper (°)	0.94
No. of protein atoms	3502
No. of water atoms	81
Mean B factor, all atoms (Å^2)	30.00
Ramachandran plot statistics (%)	
Residues in most favored regions	89.7
Residues in additional allowed regions	8.7
Residues in generously allowed regions	1.6
Residues in disallowed regions	0.0

3.3. Comparison to yeast SAICAR synthase and ATP-binding sites

A *DALI* search for structurally similar proteins (Holm & Sander, 1996) produced one major homolog, with a Z score of 18.3 and a primary sequence identity of 25%: SAICAR synthase from *Saccharomyces cerevisiae* (PDB code 1a48). Approximately 188 residues of 306 superimpose with the 209-amino-acid *T. maritima* protein, with an r.m.s.d. of 2.5 Å. Fig. 4 shows a superposition of the *T. maritima* protein and an ATP-bound complex of the yeast enzyme (PDB code 1odb). The folds are quite similar, with extensive insertions in the yeast protein at the amino- and carboxy-termini as well as in a few loop regions (see Fig. 2 for details). The ATP from the complex locates the binding clefts with conserved glutamates (Glu172 for *T. maritima*, Glu219 for the yeast enzyme) in equivalent positions. Cys126, responsible for the covalent dimer, and the residues of the interface salt bridge are not conserved in the yeast protein. Furthermore, there is a 13-residue insertion between secondary-structure elements $\alpha2$ and $\beta7$ (see Figs. 2 and 4), which appears to hinder any equivalent dimer formation.

We wish to thank all members of the Structural Biology Center at Argonne National Laboratory for their help in conducting experiments and members of the Midwest Center for Structural Genomics for help in preparation of this manuscript. This work was supported by National Institutes of Health Grant GM62414-01 and by the US Department of Energy, Office of Biological and Environmental Research under contract W-31-109-Eng-38

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.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Henrick, K. & Thornton, J. M. (1998). *Trends Biochem. Sci.* **23**, 358–361.

Holm, L. & Sander, C. (1996). *Science*, **273**, 595–602.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Laskowski, R. A. (2001). *Nucleic Acids Res.* **29**, 221–222.
Levdikov, V. M., Barynin, V. V., Grebenko, A. I., Melik-Adamyan, W. R., Lamzin, V. S. & Wilson, K. S. (1998). *Structure*, **6**, 363–376.
Mathews, C. K. & Van Holde, K. E. (1996). *Biochemistry.* Menlo Park, CA, USA: Benjamin/Cummings.
Oldfield, T. J. (1996). In *Crystallographic Computing 7: Proceedings of the Macromolecular Crystallography Computing School*, edited by P. Bourne & K. Watenpaugh. Oxford University Press.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
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**, 1095–106.