Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure of a fatty acid-binding protein from *Bacillus subtilis* determined by sulfur-SAD phasing using in-house chromium radiation

**Jie Nan,[a] Yanfeng Zhou,[a] Cheng Yang,[a,b] Erik Brostromer,[a] Ole Kristensen[c] and Xiao-Dong Su[a,d]***

[a]National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, People's Republic of China, [b]Rigaku/MSC Inc., 9009 New Trails Drive, The Woodlands, TX 77381, USA, [c]University of Copenhagen, Faculty of Pharmaceutical Sciences, 2100 Copenhagen, Denmark, and [d]Shenzhen Graduate School of Peking University, Shenzhen 518055, People's Republic of China

Correspondence e-mail: xdsu@pku.edu.cn

Received 26 January 2009
Accepted 3 March 2009

**PDB Reference:** *Bs*DegV, 3fys, r3fyssf.

Sulfur single-wavelength anomalous dispersion (S-SAD) and halide-soaking methods are increasingly being used for *ab initio* phasing. With the introduction of in-house Cr X-ray sources, these methods benefit from the enhanced anomalous scattering of S and halide atoms, respectively. Here, these methods were combined to determine the crystal structure of *Bs*DegV, a DegV protein-family member from *Bacillus subtilis*. The protein was cocrystallized with bromide and low-redundancy data were collected to 2.5 Å resolution using Cr $K\alpha$ radiation. 17 heavy-atom sites (ten sulfurs and seven bromides) were located using standard methods. The anomalous scattering of some of the *Bs*DegV S atoms and Br atoms was weak, thus neither sulfurs nor bromides could be used alone for structure determination using the collected data. When all 17 heavy-atom sites were used for SAD phasing, an easily interpretable electron-density map was obtained after density modification. The model of *Bs*DegV was built automatically and a palmitate was found tightly bound in the active site. Sequence alignment and comparisons with other known DegV structures provided further insight into the specificity of fatty-acid selection and recognition within this protein family.

## 1. Introduction

In recent years, structural genomics centres have been geared towards the protein structure determination of complete genomes, mostly the human genome and pathogen genomes (Manjasetty *et al.*, 2008). Novel and automated technologies and methods are therefore required and have been developed to speed up structure determination. Synchrotron radiation has made great progress (Dauter, 2005) and to date most protein crystal structures have been solved based on synchrotron diffraction data. However, owing to the limited beamtime availability, synchrotron sources are still not always accessible to many protein crystallographers, particularly when the synchrotron source is far away. Thus, in-house facilities and methods for data collection and phasing are important and necessary for structure determination when a constant flow of diffractable crystals become available.

For *ab initio* phasing using in-house data, several methods are available such as single-wavelength anomalous dispersion (SAD), single or multiple isomorphous replacement (SIR, MIR) and SIR/MIR supplemented with anomalous scattering (SIRAS, MIRAS). All of these approaches utilize the isomorphous or anomalous scattering differences between the native and so-called derivative crystals (the native and deri-

vative crystals may be the same in the case of SAD), which are mostly contributed from artificially introduced heavy atoms or from those naturally present in the crystal. In addition to the conventionally used heavy atoms such as Pt, Au, Hg *etc.*, halide salts are becoming more and more popular (Yogavel *et al.*, 2007) and are often used in 'quick cryosoaking' owing to the simplicity of crystal derivative preparation (Dauter *et al.*, 2000). Being small monatomic ions, halides usually substitute the solvent waters at the protein surface without much preference in coordination geometry. The limitations of this method are the low occupancy and unpredictable number of the halide sites, which can lead to difficulties in substructure determination (Usón *et al.*, 2003). One of the approaches to overcoming this problem that has been investigated is the use of $KI/I_2$ solutions to introduce triiodide and pentaiodide anions, which have a wider range of binding sites; however, this approach is not suitable for PEG-based solutions owing to phase separation on the addition of $KI/I_2$ (Evans & Bricogne, 2003).

Being an atom that is naturally found in the cysteine and methionine residues of native proteins, sulfur can be used as the sole anomalous scatterer in SAD phasing (S-SAD). The first case of S-SAD structural determination can be traced back to 1981 (Hendrickson & Teeter, 1981) and was followed by the proposal of computational methods (Wang, 1985). Advances in cryocrystallography and detector technology have made it possible to measure weak anomalous scattering differences and in the meantime phasing quality has been greatly improved by the development of new methods and algorithms. After S-SAD phasing had been reported to be facilitated by synchrotron data collected at a high wavelength (>1.54 Å; Dauter *et al.*, 1999), the successful use of in-house data from Cu X-rays in S-SAD structure determination was also reported (Yang & Pflugrath, 2001; Ren *et al.*, 2005; Sarma & Karplus, 2006). S-SAD phasing is not always successful, but some general guidelines have emerged. Firstly, based on a minimum anomalous signal of 0.6% (Wang, 1985), at least three sulfurs per 100 residues are required for the SAD method when using a Cu X-ray source. Secondly, highly redundant data of sufficiently high resolution should be measured for the precise estimation of anomalous differences. Thirdly, it has been shown that the solvent content also affects the phasing quality (Debreczeni *et al.*, 2003). There are also other uncertain factors such as radiation damage *etc.* that can affect the phasing procedure and the final phase quality. Sometimes high-resolution native synchrotron data are needed as a supplement in order to obtain interpretable maps by phase expansion (Debreczeni *et al.*, 2003).

In addition to the sulfurs from cysteine and methionine residues and sulfate ions, there are often other atoms that can potentially be used in SAD phasing, such as the P atom in phosphate, Cl atoms from buffers and (for example) the calcium, magnesium, iron and zinc metal ions present in many metalloproteins. Also, with the introduction of in-house chromium radiation sources, the anomalous scattering factor $f$ of sulfur is doubled at the Cr $K\alpha$ wavelength of 2.29 Å compared with the Cu $K\alpha$ wavelength (1.54 Å; Yang *et al.*, 2003; Xu *et al.*, 2005; Watanabe *et al.*, 2005; Li *et al.*, 2008). Indeed, most of these atoms have a significant increase (twofold) in anomalous scattering using Cr radiation. The anomalous scattering factors $f$ of some common atoms with Cu $K\alpha$ and Cr $K\alpha$ radiation are summarized in Supplementary Table 1[1].

The Pfam DegV protein family received its name based on sequence similarity to the DegV protein from *Bacillus subtilis*. There are 710 members of the DegV family and all of them are from bacteria, mostly Firmicutes, according to the Pfam database (Finn *et al.*, 2008). The function of DegV proteins is not yet completely understood, but structural evidence indicates that they can bind different fatty acids. In order to study their function and to understand how selectivity is achieved by the various DegV-family members, we determined the structure of DegV from *B. subtilis* (*Bs*DegV) in complex with an innate tightly bound palmitate. The native *Bs*DegV protein was cocrystallized with bromide. X-ray diffraction data were collected using an in-house Cr $K\alpha$ X-ray source and the structure was solved by the SAD method using the anomalous dispersion from both S and Br atoms. The anomalous signals from different atoms as well as their phasing power were calculated and compared. As in the other known structures of DegV proteins, a fatty acid was observed in the crystal structure. Combined with structure-based multiple sequence alignment, analysis of the binding pocket will help us to better understand this family of proteins with regard to the specificity of fatty-acid selection and recognition.

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The *Bs*DegV gene was amplified from the genomic DNA of *B. subtilis* by PCR using the primers 5′-CGC**GGATCC**ATGAATATTGCAGTCGTAACAGA-3′ and 5′-CCG**CTCGAG**TTATTTAAAACACCAGCAAATTCC-3′. The primers contain *Bam*HI and *Xho*I restriction sites, respectively, which are shown in bold in the sequences above. After digestion with *Bam*HI and *Xho*I, the PCR-amplified fragment was ligated into the kanamycin-resistant expression vector pET-28a(+) (Novagen, USA) which was linearized with the same two restriction endonucleases. The final construct contained the *Bs*DegV-coding sequence fused to a poly-His (6×His) tag at the N-terminus.

The plasmid was transformed into *Escherichia coli* strain BL21 (DE3) for protein expression. Cells were grown in Luria–Bertani (LB) broth supplemented with 50 µg ml$^{-1}$ kanamycin at 310 K until the optical density at 600 nm reached 0.6. Recombinant protein expression was induced by adding isopropyl $\beta$-D-1-thiogalactopyranoside to a final concentration of 0.5 m*M*. The culture was transferred to 303 K and incubated for 4 h. *E. coli* cells were harvested by centrifugation (6700*g*, 10 min), resuspended in lysis buffer (20 m*M* Tris–HCl pH 7.5

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5154). Services for accessing this material are described at the back of the journal.

**Table 1**
Statistics of data collection and structure refinement.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Wavelength (Å) | 2.2909 |
| Space group | $P2_1$ |
| Unit-cell parameters (Å, °) | $a$ = 42.1, $b$ = 84.7, $c$ = 42.2, $\alpha$ = 90.0, $\beta$ = 94.1, $\gamma$ = 90.0 |
| Resolution (Å) | 25.0–2.5 (2.6–2.5) |
| No. of unique reflections | 9588 (721) |
| No. of reflections in the $R_{free}$ data set | 479 (36) |
| Completeness | 93.5 (65.3) |
| $R_{merge}$† (%) | 3.7 (8.0) |
| $\langle I\rangle/\langle\sigma(I)\rangle$ | 60.3 (31.6) |
| Redundancy | 3.7 (2.2) |
| Refinement | |
| $R_{work}/R_{free}$‡ (%) | 14.8/22.5 |
| No. of protein atoms | 2229 |
| No. of ligand atoms | 29 |
| No. of waters | 132 |
| R.m.s.d. bond length (Å) | 0.006 |
| R.m.s.d. bond angle (°) | 0.998 |
| Ramachandran plot (%) | |
| Favoured | 91.7 |
| Allowed | 7.9 |
| Disallowed | 0.4 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{work}$ and $R_{free} = \sum_{hkl}[|F(h)_o| - |F(h)_c|]/\sum_{hkl}|F(h)_o|$ for reflections in the working and test sets, respectively. $R_{free}$ was calculated using 5% of the data.

and 500 m*M* NaCl) and then lysed by sonication. After centrifugation (34 700*g*, 30 min) at 277 K twice, the supernatant was purified using a HiTrap nickel column (GE Healthcare, USA) at room temperature. The His-tagged protein was eluted with a buffer containing 20 m*M* Tris–HCl pH 7.5, 0.5 *M* NaCl and 200 m*M* imidazole and further purified by size-exclusion chromatography (Superdex 75, GE Healthcare, USA) in the final crystallization buffer (20 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl). The purified protein was concentrated to about 10 mg ml$^{-1}$ by centrifugation using an Amicon Ultra-15 concentrator (Millipore, USA) with a 10 kDa cutoff.

### 2.2. Crystallization

Initial crystallization screening was performed using the Hampton Research Crystal Screen, Crystal Screen II and Index Screen kits and the hanging-drop vapour-diffusion method. Crystallization drops prepared by mixing 1 µl protein solution with 1 µl reservoir solution were incubated at 289 K. Crystals suitable for data collection were obtained using a condition containing 0.15 *M* potassium bromide, 30% polyethylene glycol monomethyl ether (PEG MME) 2000.

### 2.3. Data collection and processing

X-ray diffraction data were collected using an in-house X-ray source consisting of a Rigaku MicroMax-007 HF generator with VariMax Cr optics and a Rigaku R-AXIS IV$^{++}$ imaging-plate detector equipped with a helium cone. The crystal was cryoprotected by quickly swiping it through reservoir solution supplemented with 15% glycerol, quickly frozen in liquid nitrogen and then mounted in a random orientation. SAD data were collected from a single crystal using an oscillation range of 0.5° and 4 min exposure time per image with the crystal-to-detector distance set to 125 mm. A total of 720 images were collected. The data were indexed, integrated, corrected for absorption and scaled using the program *HKL*-2000 (Otwinowski & Minor, 1997). Diffraction data were limited by an improper detector placement and thus reflections beyond a resolution of 2.5 Å were excluded. The statistics of the collected data are summarized in Table 1.

### 2.4. Structure determination and refinement

Initial analysis of the data quality was performed using the programs *XPREP* (Sheldrick, 2001) and *XTRIAGE* (Zwart *et al.*, 2005) from the *PHENIX* software package (Adams *et al.*, 2002). The crystal of *Bs*DegV belonged to the monoclinic space group $P2_1$, with unit-cell parameters $a$ = 42.1, $b$ = 84.7, $c$ = 42.2 Å. From the crystal parameters, it was suggested that there was one molecule per asymmetric unit, with a Matthews coefficient of 2.2 Å$^3$ Da$^{-1}$ and 44.7% solvent content.

*SHELXD* (Sheldrick, 2008) was used to find the locations of the heavy-atom sites, which were then refined using the software *MLPHARE* (Otwinowski, 1991) with data to 2.8 Å resolution. The anomalous difference Fourier map obtained from the *MLPHARE* phases was used to search for additional heavy-atom sites. 17 heavy atoms were found in total and the anomalous Cullis *R* factor (defined as the anomalous lack of closure over the anomalous difference) given by *MLPHARE* was 0.69. All of the located sites were treated as sulfurs and input into the program *SOLVE* (Terwilliger, 2003) to calculate the initial phases. The phases were improved by density modification using the program *DM* (Cowtan, 1994) and the final real-space free *R* factor was 0.116. The model was automatically traced using the program *RESOLVE* (Terwilliger, 2003) and was completed manually using *Coot* (Emsley & Cowtan, 2004). The structure was refined with *REFMAC*5 (Murshudov *et al.*, 1997) and the *phenix.refine* module of the *PHENIX* software package (Afonine *et al.*, 2005). The geometry of the model was checked using the program *PROCHECK* (Laskowski *et al.*, 1993). Refinement statistics are listed in Table 1. The graphics program *PyMOL* (DeLano, 2002) was used in the production of the figures as well as for structural analysis.

## 3. Results and discussion

### 3.1. Anomalous diffraction

Cell-content analysis suggested that there was one molecule per asymmetric unit in the crystal. The *Bs*DegV protein contains 281 amino acids, including seven methionine residues and three cysteine residues, which means that there are ten sulfurs per asymmetric unit. The expected value of the anomalous signal at the Cr $K\alpha$ wavelength was about 1.15% as calculated using the Hendrickson formula (Hendrickson & Teeter, 1981) modified by introducing angular dependence, $\langle\Delta F^{\pm}\rangle/\langle F\rangle_{calc} = 2^{1/2}(f''_A N_A^{1/2})/[f_{eff}(\theta)N_P^{1/2}]$ (Dauter *et al.*, 2002; Shen *et al.*, 2003). The observed $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ ratio of the diffraction data is shown in Fig. 1. Generally, the experimental

data agree well with the theoretical Bijvoet ratio at low resolution (Dauter *et al.*, 2002), but the deviation increases dramatically with increasing resolution owing to the decrease in measurement accuracy. In the case of *Bs*DegV (Fig. 1), the observed Bijvoet ratio is much higher than the calculated value even in the low-resolution bins, which suggested that there were some other heavy atoms with anomalous scattering in the crystal. Considering that the protein was cocrystallized with bromide, we concluded that the additional anomalous signal was from Br atoms. The measurability of the anomalous diffraction as defined in the *PHENIX* software package was analyzed using the *phenix.reflection_statistics* utility (Fig. 1). In contrast to the anomalous signal, the measurability fell as the resolution increased. A more sudden increase in the Bijvoet ratio value and a measurability that was still in the encouraging range was observed at 2.8 Å resolution, suggesting a good high-resolution cutoff for substructure determination.

### 3.2. Substructure solution and SAD phasing

Apart from the ten sulfurs from the protein, the total number of anomalous scatterers in the crystal was unknown. Thus, 20 sulfurs were set to be searched for in *SHEXLD* with data truncated to 2.8 Å resolution as discussed above. The correlation coefficients output by *SHEXLD* were 46.95 for all data and 22.2 for weak data and the substructure solution is listed in Table 2 (the atom sites were identified after the structure solution of *Bs*DegV). The coordinates of the highest nine peaks were input to *MLPHARE* for refinement and further heavy-atom search. *MLPHARE* was run for seven cycles until the search converged with 17 anomalous scattering sites, of which 16 corresponded to the 16 highest peaks in the

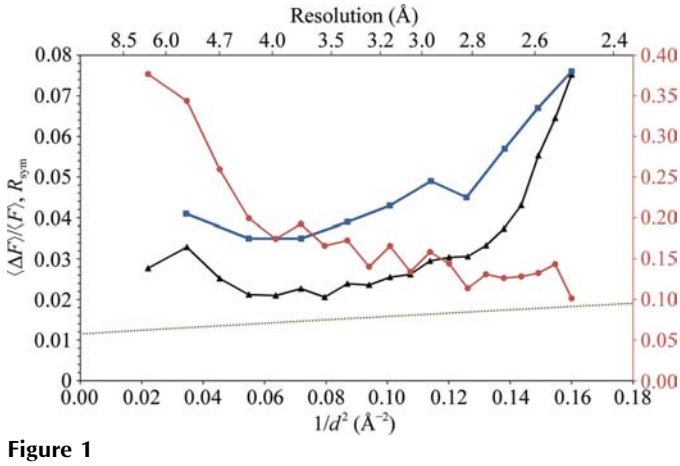


**Figure 1**
The Bijvoet ratio as a function of resolution. The dotted line in green is the theoretical value of the Bijvoet ratio calculated from the Hendrickson formula modified by introducing the factor of the diffraction angle, $2^{1/2}(f''_A N_A^{1/2})/[f_{eff}(\theta)N_P^{1/2}]$. The observed Bijvoet ratio $\langle\Delta F^{\pm}\rangle/\langle F\rangle$ is plotted as black triangles. $R_{sym}$ is plotted as blue squares. The red line represents the measurability of the anomalous signals at different resolutions and the values are given on the right *y* axis. The measurability was calculated using the *phenix.reflection_statistics* module from the *PHENIX* software package. The measurability given by *phenix.reflection_statistics* is defined as the fraction of Bijvoet-related intensity differences for which $|\Delta I|/\sigma(\Delta I) > 3.0$, $\min[I(+)/\sigma I(+), I(-)/\sigma I(-)] > 3.0$ holds (Zwart *et al.*, 2005).

**Table 2**
Substructure solution from *SHELXD*, with each peak annotated with the corresponding atom.

| Site | Height | Atom |
|---|---|---|
| 1 | 1.000 | Br1 |
| 2 | 0.702 | Cys277 |
| 3 | 0.663 | Met151 |
| 4 | 0.652 | Br3 |
| 5 | 0.645 | Cys279 |
| 6 | 0.636 | Br2 |
| 7 | 0.610 | Cys120 |
| 8 | 0.551 | Met227 |
| 9 | 0.538 | Met25 |
| 10 | 0.476 | Met17 |
| 11 | 0.440 | Met105 |
| 12 | 0.329 | Met160 |
| 13 | 0.316 | Met1 |
| 14 | 0.299 | Br7 |
| 16 | 0.229 | Br4 |

*SHEXLD* output. Not knowing the type of the atoms, all located sites were treated as sulfurs for phasing using *SOLVE*. The initial phases were improved by solvent flattening and histogram matching using *DM* and the figure of merit (FOM) increased from 0.410 to 0.908. With an easily interpretable map, as illustrated in Fig. 4(*c*), automatic model building by *RESOLVE* traced 195 of 281 residues. The remaining part of the structure was manually built and refined.

### 3.3. Comparison of the anomalous scattering of Met $S^{\delta}$, Cys $S^{\gamma}$ and bromide

After structure completion and refinement, it was possible to identify the atom types of the substructure and to correlate the atoms to the peaks from the *SHELXD* result (Table 2). Of the 17 sites located per asymmetric unit, ten originated from *Bs*DegV S atoms and seven were bromides from the crystal-

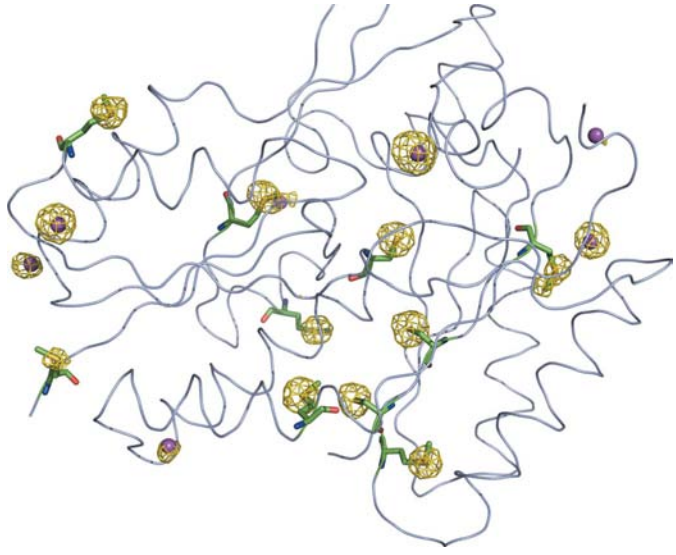

**Figure 2**
Anomalous Fourier map calculated with final phases at 3.5$\sigma$. The *Bs*DegV structure is shown as a backbone representation; methionines and cysteines are shown in green sticks and bromides are highlighted as purple balls.

lization buffer binding to the protein surface (Fig. 2). To verify that all the heavy-atom sites had been located, an anomalous difference map was calculated from the refined model and no strong peaks were found. The anomalous peak heights of the located heavy atoms agreed well with the rank from the *SHEXLD* output (Fig. 3). Since the crystallization buffer contained chlorine, the seven identified bromides could also be chlorides or both types of atoms bound in the same sites. However, the anomalous signal of the first three sites was too strong to arise from chloride alone, which confirmed the presence of bromide in the crystal. For convenience of discussion and analysis, we assumed that all the nonsulfur heavy atoms were bromides.

All three cysteine $S^{\gamma}$ atoms displayed similar anomalous scattering, but the methionine $S^{\delta}$ atoms showed varying scattering properties. The anomalous signals of the $S^{\delta}$ atoms in Met151, Met25 and Met227 were similar to those of the $S^{\gamma}$ atoms in the cysteine residues and higher than those in Met105, Met17 and Met160. The peak height of the $S^{\delta}$ atom in Met1 was $5\sigma$, one third of the highest peak for the sulfurs and just above the default detection threshold used in many programs. Previous studies on radiation damage by soft X-rays, including that from in-house chromium radiation, have revealed that the anomalous signal of sulfurs can be reduced in different ways depending on the residue type and the surrounding environment (Burmeister, 2000; Evans *et al.*, 2003; Yang *et al.*, 2003).

Of the seven bromides, the first three sites show strong anomalous scattering with high occupancies. The anomalous signal of the other four bromides is in the same range as that of the methionine Met1 $S^{\delta}$ atom with the lowest peak height (Fig. 3 and Supplementary Fig. 1). This can be explained by the weak nonspecific binding of the halide ions. They tend to bind electrostatically either to protein hydrogen-donor groups, particularly arginine and lysine residues, or to water, or to bind through hydrophobic interactions at nonpolar protein surfaces (Dauter *et al.*, 2000). Without much preference for specific coordination geometry, halides often provide only a low anomalous signal owing to their poor occupancy, as is the case for the four Br atoms Br4–7 in the structure of *Bs*DegV.

### 3.4. Comparison of the phasing power of sulfur and bromide

In order to compare the phasing power of the sulfurs and bromides in *Bs*DegV, SAD phases were calculated using only sulfurs, only bromides or all sites. To eliminate the bias of resolution cutoff, data at different resolutions were used for phase calculation. The coordinates of the selected atoms were input into *SOLVE* and the initial phases were improved in a consistent manner by solvent flattening and histogram matching with *DM*. The improved phases were compared with the phases calculated from the final model and phase errors were estimated. The phase-error plot clearly showed that the quality of the phases calculated using both atom types was much better than that obtained when either of the types of atoms were used alone (Fig. 4*d*). In agreement with this, as

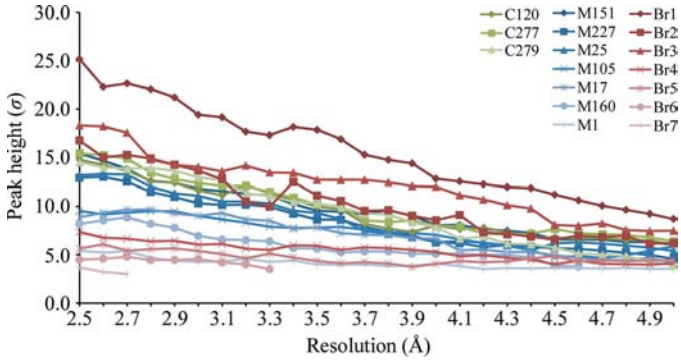


**Figure 3**
The anomalous signal reduction as a function of resolution. The peak heights of the heavy atoms were derived from the anomalous Fourier map, which was calculated using the final phases.

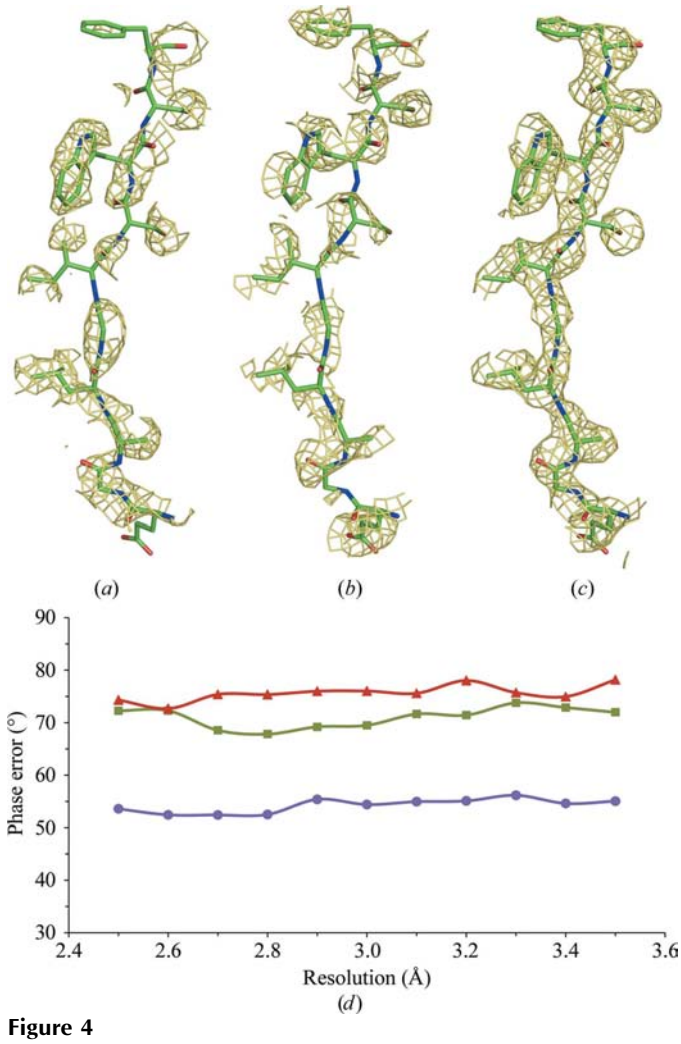


**Figure 4**
Experimental electron-density map of *Bs*DegV calculated using diffraction data to 2.5 Å resolution and SAD phases after *DM*. (*a*) Electron-density map when only sulfurs were used for SAD phasing. (*b*) Electron-density map when only bromides were used for SAD phasing. (*c*) Electron-density map when all heavy atoms were used for SAD phasing. (*d*) SAD phase errors (after *DM*) against the final model as a function of resolution cutoff. Different colours represent the phases calculated from different anomalous scatterers: phases calculated from sulfurs are shown in red, those from bromides are shown in green and those from both are shown in purple.

illustrated by the electron-density maps, neither sulfurs nor bromides alone were sufficient to give an easily interpretable map from SAD phasing (Figs. 4*a*–4*c*).

### 3.5. Structure description and homology search

The crystal structure of *Bs*DegV consists of two different $\alpha/\beta$ folds. The N-terminal domain (residues 1–155), which contains seven $\beta$-strands and six $\alpha$-helices, can be further divided into two subdomains: a three-layer $\alpha/\beta/\alpha$ core domain and a small peripheral subdomain including a three-stranded antiparallel $\beta$-sheet and one $\alpha$-helix. The C-terminal domain (residues 156–281) contains six $\beta$-strands and five $\alpha$-helices in a two-layered $\alpha/\beta$-arrangement, as shown in Fig. 5(*a*). The two domains mainly interact through hydrophobic contacts and form a ligand-binding pocket at the domain interface.

A *DALI* structural search (Holm & Sander, 1995) revealed that *Bs*DegV shares structural similarity to other DegV or DegV-like proteins as well as to the mannose transporter EIIA domain (EIIA-man) and dihydroxyacetone kinases (Dak). In a previous study, the EIIA-man, Dak and DegV proteins were classified into one superfamily named EDD (Kinch *et al.*, 2005). EIIA-man proteins are part of the bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS) mediating carbohydrate transport. Dak proteins function in phosphoryl transfer from ATP or phosphoenolpyruvate to downstream enzymes/substrates in eubacterial, plant and animal phosphorylation cascades. As they share an

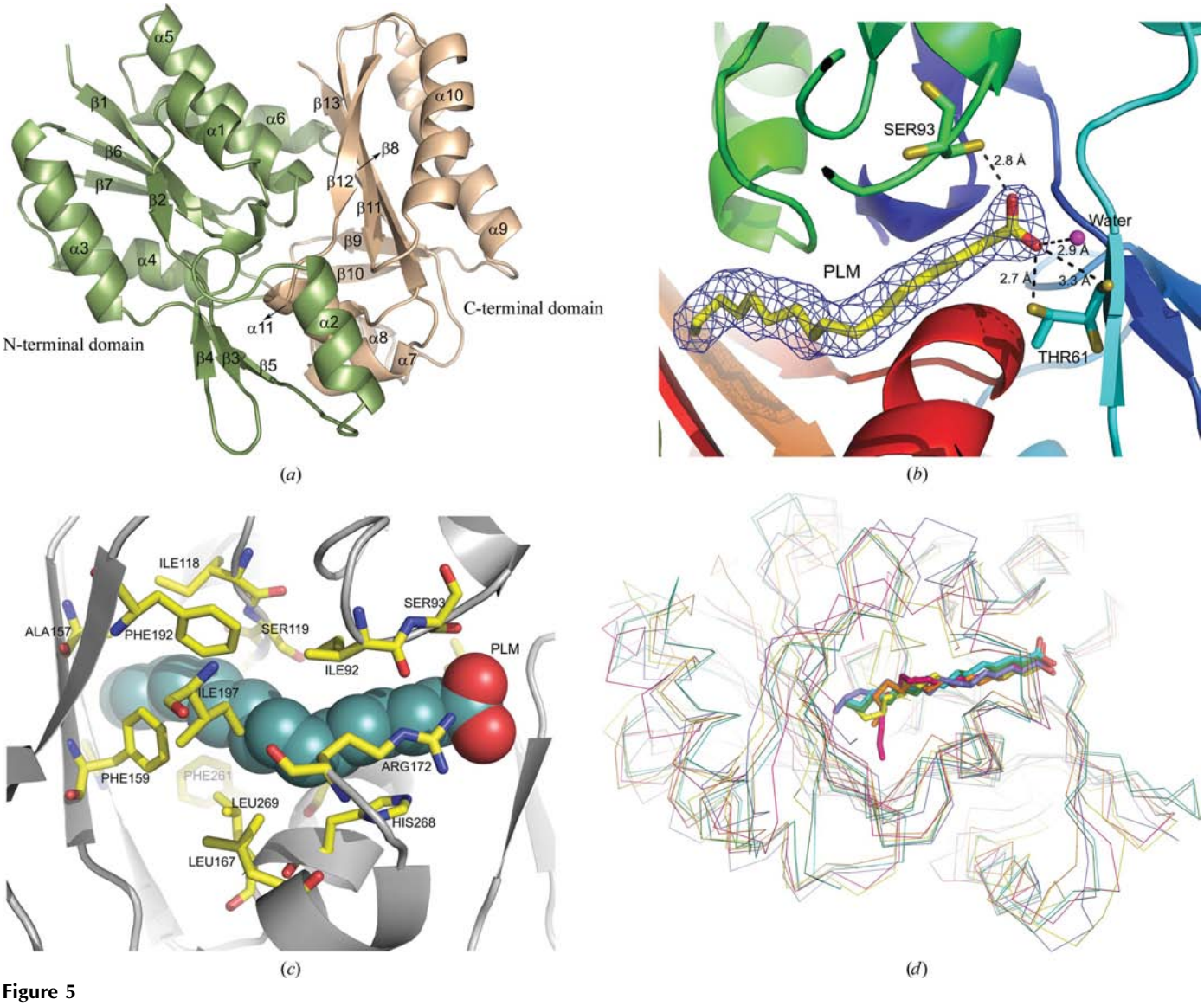


**Figure 5**
(*a*) Overall structure of *Bs*DegV in ribbon representation. The N- and C-terminal domains are represented in green and wheat, respectively. (*b*) OMIT map without palmitate. The map was calculated with ($|F_o| - |F_c|$) and contoured at 3.0$\sigma$. (*c*) Palmitate-binding site. (*d*) Superimposition of known DegV structures (shown as ribbons) and their interaction with palmitate (shown as a stick representation). Blue, *Bs*DegV; red, *Streptococcus pyogenes* DegV-like protein (PDB code 2g7z; B. Nocek, L. Volkart, S. Clancy & A. Joachimiak, unpublished work); orange, *B. stearothermophilus* APC36103 (PDB code 1pzx; R. Zhang, J. Osipiuk, M. Zhou, R. Alkire, S. Moy, F. Collart & A. Joachimiak, unpublished work); yellow, *Thermus thermophilus* DegV-family protein (PDB code 2dt8; P. H. Rehse & S. Yokoyama, unpublished work); green, *Thermotoga maritima* TM1468 (PDB code 1vpv; Joint Center for Structural Genomics, unpublished work); cyan, *Thermotoga maritima* TM841 (PDB code 1mgp; Schulze-Gahmen *et al.*, 2003).

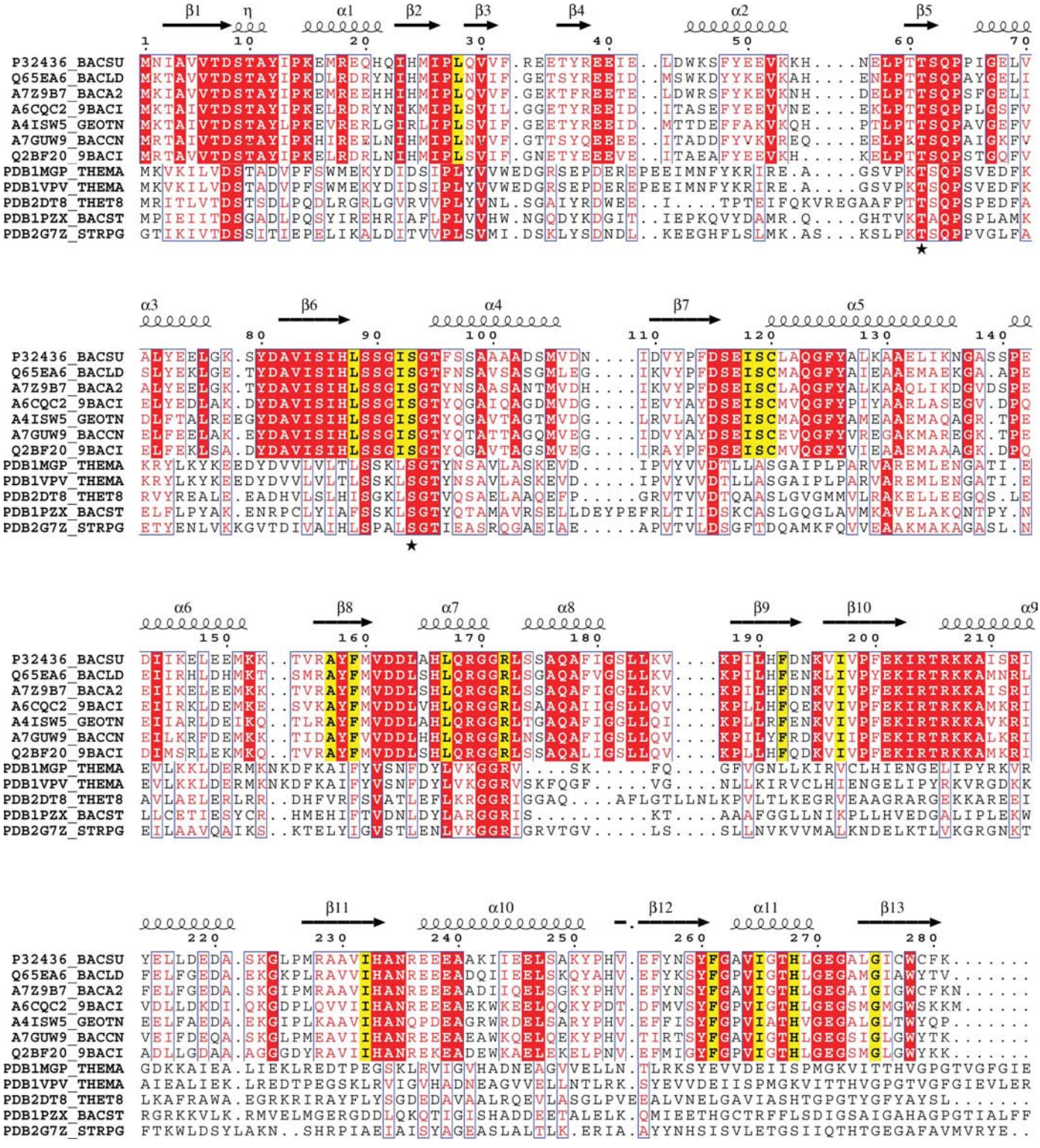


**Figure 6**
Structure-based multiple sequence alignment of DegV-family proteins. The first seven sequences are amongst the most similar proteins found using a *BLAST* search (http://www.ncbi.nlm.nih.gov/BLAST) against the nonredundant sequence database; proteins are named with SWISS-PROT ID and organism (BACSU, *B. subtilis*; BACLD, *B. licheniformis*; BACA2, *B. amyloliquefaciens*; 9BACI, *Bacillus* sp.; GEOTN, *Geobacillus thermodenitrificans*; BACCN, *B. cereus* subsp. *cytotoxis*). The last five sequences are the five most similar structures found by a *DALI* search against the PDB database (http://www.pdb.org); proteins are named with PDB code and organism as detailed for Fig. 5(*c*). Identical residues are highlighted with red boxes and similar residues are shown in red. The binding sites of the fatty-acid carboxylate group are indicated by black stars below the respective columns. Residues highlighted with yellow boxes are conserved residues for the recognition of the fatty-acid hydrocarbon tail. The alignment was prepared using *ClustalW* (Larkin *et al.*, 2007) and the figure was generated using *ESPript* (Gouet *et al.*, 1999).

evolutionarily conserved domain, it has been suggested that DegV proteins may also possess phosphotransferase activity (Erni *et al.*, 2006).

### 3.6. Fatty acid-binding site

Although not included in the crystallization solution, a palmitate was observed to be bound in the cavity formed at the domain interface of the *Bs*DegV structure (Fig. 5*b*). The residues contributing to the deeply buried binding pocket are Thr61, Ser93, Leu28, Leu88, Ile92, Ser93, Ile118, Ser119, Cys120, Ala157, Phe159, Leu167, Arg172, Phe192, Ile197, Ile232, Phe261, Ile265, His268, Leu269, Gly275 and Cys277 (Fig. 5*c*). The carboxylate group of the palmitate forms hydrogen bonds to Thr61, Ser93 and a water molecule. The remaining part of the ligand is recognized by the protein mostly through hydrophobic and van der Waals interactions. All of the DegV crystal structures determined to date have a palmitate bound in this conserved pocket, with the exception of the SPy1493 structure (PDB code 2g7z; B. Nocek, L. Volkart, S. Clancy & A. Joachimiak, unpublished work), which however contains a related hexaenoate molecule at the ligand-binding site. The fatty-acid orientation and the interactions of the carboxylate group with the surrounding residues are similar in all DegV proteins of known structure (Fig. 5*d*).

A structure-based multiple sequence alignment (Fig. 6), including sequences with high homology to *Bs*DegV and other DegV-family members of known structure, indicates that the main recognition sites for the carboxylate group of the ligand, Thr61 and Ser93, are rather conserved. In contrast, the residues required for binding the fatty-acid hydrocarbon tail are only conserved within a subgroup of the DegV protein family. This indicates that proteins of the DegV family tend to bind fatty acids, particularly the carboxylate group, in a conserved manner. However, the binding of specific fatty acids is length-dependent and the recognition and orientation, even for the same ligand, is only conserved to some extent.

## 4. Conclusions

In theory, many proteins contain enough sulfurs to make in-house data collection and S-SAD phasing an option in crystal structure determination. However, the observed anomalous signals are often lower than expected owing to limited crystal quality, an inadequate data-collection strategy, radiation damage and other factors. Radiation from a Cr X-ray source increases the anomalous scattering factors $f$ of sulfur and many other atoms compared with their values when a Cu X-ray source is used. Another possible method to consider is halide soaking (or cocrystallization), in which new anomalous scatterers such as bromide and iodide are introduced to replace water or chloride. Both methods have proved to be effective in previous studies (Dauter *et al.*, 2000; Evans & Bricogne, 2003; Yang *et al.*, 2003; Xu *et al.*, 2005; Watanabe *et al.*, 2005; Yogavel *et al.*, 2007; Li *et al.*, 2008). Here, in our structural study of the protein *Bs*DegV, the anomalous dispersion of sulfurs was quite weak and bromide incorporation was inefficient. With the combination of bromide co-crystallization and the Cr $K\alpha$ radiation-enhanced S-SAD method, the structure was successfully solved. Considering the simplicity of sample preparation, it is confirmed that this approach combining halide soaking and enhancement of anomalous scattering signals using a Cr $K\alpha$ X-ray source can be routinely used for in-house S-SAD phasing.

This work was supported by grants from the Ministry of Education of the People's Republic of China (104003 and a special SAD phasing grant) and from the National Natural Science Foundation of China (30325012, 30530190). OK and X-DS would like to thank the Alfred Benzon Foundation for support.

## References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **42**, contribution 8.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dauter, Z. (2005). *Prog. Biophys. Mol. Biol.* **89**, 153–172.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, USA.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Erni, B., Siebold, C., Christen, S., Srinivas, A., Oberholzer, A. & Baumann, U. (2006). *Cell. Mol. Life Sci.* **63**, 890–900.
Evans, G. & Bricogne, G. (2003). *Acta Cryst.* D**59**, 1923–1929.
Evans, G., Polentarutti, M., Djinovic Carugo, K. & Bricogne, G. (2003). *Acta Cryst.* D**59**, 1429–1434.
Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. & Bateman, A. (2008). *Nucleic Acids Res.* **36**, D281–D288.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Kinch, L. N., Cheek, S. & Grishin, N. V. (2005). *Protein Sci.* **14**, 360–367.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Li, Y., Bahti, P., Shaw, N., Song, G., Chen, S., Zhang, X., Zhang, M., Cheng, C., Yin, J., Zhu, J.-Y., Zhang, H., Che, D., Xu, H., Abbas, A., Wang, B.-C. & Liu, Z.-J. (2008). *Proteins*, **71**, 2109–2113.
Manjasetty, B. A., Turnbull, A. P., Panjikar, S., Bussow, K. & Chance, M. R. (2008). *Proteomics*, **8**, 612–625.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

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.
Ren, H., Wang, L., Bennett, M., Liang, Y., Zheng, X., Lu, F., Li, L., Nan, J., Luo, M., Eriksson, S., Zhang, C. & Su, X.-D. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 303–308.
Sarma, G. N. & Karplus, P. A. (2006). *Acta Cryst.* D**62**, 707–716.
Schulze-Gahmen, U., Pelaschier, J., Yokota, H., Kim, R. & Kim, S.-H. (2003). *Proteins*, **50**, 526–530.
Sheldrick, G. M. (2001). *XPREP*, v.6.12. Bruker–Nonius Inc., Madison, USA.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shen, Q., Wang, J. & Ealick, S. E. (2003). *Acta Cryst.* A**59**, 371–373.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Usón, I., Schmidt, B., von Bülow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 57–66.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Watanabe, N., Kitago, Y., Tanaka, I., Wang, J., Gu, Y., Zheng, C. & Fan, H. (2005). *Acta Cryst.* D**61**, 1533–1540.
Xu, H. *et al.* (2005). *Acta Cryst.* D**61**, 960–966.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.
Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* D**59**, 1943–1957.
Yogavel, M., Gill, J., Mishra, P. C. & Sharma, A. (2007). *Acta Cryst.* D**63**, 931–934.
Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, contribution 7.