

Abhinav Kumar,<sup>a,b</sup> Andrei Lomize,<sup>c</sup>  
 Kevin K. Jin,<sup>a,b</sup> Dennis Carlton,<sup>a,d</sup>  
 Mitchell D. Miller,<sup>a,b</sup> Lukasz  
 Jaroszewski,<sup>a,e,f</sup> Polat Abdubek,<sup>a,g</sup>  
 Tamara Astakhova,<sup>a,e</sup> Herbert L.  
 Axelrod,<sup>a,b</sup> Hsiu-Ju Chiu,<sup>a,b</sup> Thomas  
 Clayton,<sup>a,d</sup> Debanu Das,<sup>a,b</sup> Marc C.  
 Deller,<sup>a,d</sup> Lian Duan,<sup>a,e</sup> Julie  
 Feuerhelm,<sup>a,g</sup> Joanna C. Grant,<sup>a,g</sup> Anna  
 Grzechnik,<sup>a,d</sup> Gye Won Han,<sup>a,d</sup>  
 Heath E. Klock,<sup>a,g</sup> Mark W. Knuth,<sup>a,g</sup>  
 Piotr Kozbial,<sup>a,f</sup> S. Sri Krishna,<sup>a,e,f</sup>  
 David Marciano,<sup>a,d</sup> Daniel  
 McMullan,<sup>a,g</sup> Andrew T. Morse,<sup>a,e</sup>  
 Edward Nigoghossian,<sup>a,g</sup> Linda  
 Okach,<sup>a,g</sup> Ron Reyes,<sup>a,b</sup> Christopher L.  
 Rife,<sup>a,b</sup> Natasha Sefcovic,<sup>a,f</sup> Henry J.  
 Tien,<sup>a,d</sup> Christine B. Trame,<sup>a,b</sup> Henry  
 van den Bedem,<sup>a,b</sup> Dana Weekes,<sup>a,f</sup>  
 Qingping Xu,<sup>a,b</sup> Keith O. Hodgson,<sup>a,h</sup>  
 John Wooley,<sup>a,e</sup> Marc-André  
 Elslinger,<sup>a,d</sup> Ashley M. Deacon,<sup>a,b</sup> Adam  
 Godzik,<sup>a,e,f</sup> Scott A. Lesley,<sup>a,d,g</sup> and  
 Ian A. Wilson<sup>a,d,\*</sup>

<sup>a</sup>Joint Center for Structural Genomics, <http://www.jcsg.org>, USA, <sup>b</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA, <sup>c</sup>Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, MI, USA, <sup>d</sup>Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA, <sup>e</sup>Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA, USA, <sup>f</sup>Program on Bioinformatics and Systems Biology, Burnham Institute for Medical Research, La Jolla, CA, USA, <sup>g</sup>Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA, and <sup>h</sup>Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Correspondence e-mail: [wilson@scripps.edu](mailto:wilson@scripps.edu)

Received 17 July 2009

Accepted 15 October 2009

**PDB References:** YP\_749275.1, 2ook, r2ooksf;  
 YP\_001095227.1, 2q3l, r2q3lsf.

# Open and closed conformations of two SpoIIAA-like proteins (YP\_749275.1 and YP\_001095227.1) provide insights into membrane association and ligand binding

The crystal structures of the proteins encoded by the YP\_749275.1 and YP\_001095227.1 genes from *Shewanella frigidimarina* and *S. loihica*, respectively, have been determined at 1.8 and 2.25 Å resolution, respectively. These proteins are members of a novel family of bacterial proteins that adopt the  $\alpha/\beta$  SpoIIAA-like fold found in STAS and CRAL-TRIO domains. Despite sharing 54% sequence identity, these two proteins adopt distinct conformations arising from different dispositions of their  $\alpha 2$  and  $\alpha 3$  helices. In the 'open' conformation (YP\_001095227.1), these helices are 15 Å apart, leading to the creation of a deep nonpolar cavity. In the 'closed' structure (YP\_749275.1), the helices partially unfold and rearrange, occluding the cavity and decreasing the solvent-exposed hydrophobic surface. These two complementary structures are reminiscent of the conformational switch in CRAL-TRIO carriers of hydrophobic compounds. It is suggested that both proteins may associate with the lipid bilayer in their 'open' monomeric state by inserting their amphiphilic helices,  $\alpha 2$  and  $\alpha 3$ , into the lipid bilayer. These bacterial proteins may function as carriers of nonpolar substances or as interfacially activated enzymes.

## 1. Introduction

The YP\_749275.1 gene from *Shewanella frigidimarina* encodes a protein of unknown function with a molecular weight of 14 502 Da (residues 1–126) and a calculated isoelectric point of 4.9. An ortholog with 54% sequence identity from *S. loihica* (YP\_001095227.1) is also of unknown function, with a molecular weight of 14 105 Da (residues 1–125) and a calculated isoelectric point of 4.9.

Both sequences have been assigned to a family of 119 bacterial and archaeal proteins (PB000640) in the automatically generated Pfam-B entries (Finn *et al.*, 2008). The proteins in the PB000640 family are composed of a single domain, with one exception which is fused to a universal stress protein (UspA) domain. Profile–profile sequence-comparison methods (Jaroszewski *et al.*, 2005) detected distant homology to proteins which adopt the SpoIIAA-like fold (Kovacs *et al.*, 1998). These NTP-binding proteins are involved in regulating the sporulation sigma factor F in *Bacillus subtilis*. However, since this relationship is relatively distant (<15% sequence identity), it does not allow a direct functional inference.

Here, we report the structures of these two orthologs determined using the semi-automated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG; Lesley *et al.*, 2002) as part of the Protein Structure Initiative of the National Institute of General Medical Sciences (<http://www.nigms.nih.gov/Initiatives/PSI/>).

Both proteins have now been classified by *SCOP* (Hubbard *et al.*, 1999) as being members of a novel Sfri0576-like family which belongs to the SpoIIAA superfamily. However, despite sharing the same fold, their structures differ significantly in the relative disposition of two surface  $\alpha$ -helices and in their mode of dimerization. The arrangement of the  $\alpha$ -helices suggests that the proteins may associate with the

membrane and possibly function as carriers of nonpolar compounds similar to CRAL-TRIO domains.

## 2. Materials and methods

### 2.1. Protein production and crystallization

The clones for YP\_749275.1 and YP\_001095227.1 were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock *et al.*, 2008). The gene encoding YP\_749275.1 (GenBank YP\_749275; gi:114561762; Swiss-Prot Q087X8) was amplified from *S. frigidimarina* NCIMB 400 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5'-ctgtactccagggcATGGATATGAAGAAACATGGTTTATCG-3'; reverse primer, 5'-aattaagtcgcttaATATCGAAGCCATTTCAAGGCGTCATC-3'; target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSD-KIHSHHHHHENLYFQ/G), was PCR-amplified with V-PIPE (Vector) primers (forward primer, 5'-taacgcgacttaattaactcgtttaacgctccagc-3'; reverse primer, 5'-gccctggaagtagcaggtttctgatgatgatgatg-3'). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the V-PIPE/I-PIPE mixture and dispensed onto selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in selenomethionine-containing medium at 310 K with suppression of normal methionine synthesis (Van Duyne *et al.*, 1993). At the end of fermentation, lysozyme was added to the culture to a final concentration of 250  $\mu\text{g ml}^{-1}$  and the cells were harvested and frozen. After one freeze-thaw cycle, the cells were homogenized in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine-HCl (TCEP)] and passed through a Microfluidizer (Microfluidics). The lysate was clarified by centrifugation at 32 500g for 30 min and loaded onto nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer; the resin was washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP] and the protein was eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was buffer-exchanged with TEV buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) using a PD-10 column (GE Healthcare) and incubated with 1 mg TEV protease per 15 mg of eluted protein. The protease-treated eluate was run over nickel-chelating resin (GE Healthcare) pre-equilibrated with HEPES crystallization buffer (20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP) and the resin was washed with the same buffer. The flowthrough and wash fractions were combined and concentrated to 17.7 mg ml<sup>-1</sup> by centrifugal ultrafiltration (Millipore) for crystallization trials and crystallized by mixing 200 nl protein with 200 nl crystallization solution in sitting drops above a 50  $\mu\text{l}$  reservoir volume using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard JCSG crystallization protocols (Lesley *et al.*, 2002). The crystallization reagent for YP\_749275.1 consisted of 0.2 M calcium acetate and 20.0% PEG 3350. A diamond-shaped crystal of approximate dimensions 50  $\times$  50  $\times$  50  $\mu\text{m}$  was harvested after 10 d at 277 K. Ethylene glycol was added to the crystal as a cryoprotectant to a final concentration of 8% (v/v). Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Light-source (SSRL, Menlo Park, California, USA). The diffraction data

were indexed in monoclinic space group C2. The oligomeric state of YP\_749275.1 in solution was determined using a 1  $\times$  30 cm Superdex 200 column (GE Healthcare) coupled with miniDAWN static light-scattering and Optilab differential refractive-index detectors (SEC/SLS; Wyatt Technology). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM sodium chloride and 0.02% (w/v) sodium azide. The molecular weight was calculated using the *ASTRA* v.5.1.5 software (Wyatt Technology).

The gene encoding YP\_001095227.1 (GenBank YP\_001095227; gi:127514030; Swiss-Prot A3QHM0) was amplified by polymerase chain reaction (PCR) from *Shewanella* sp. PV-4 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5'-ctgtactccagggcATGAGCAGCAATCTACATGTATCGCC-3'; reverse primer, 5'-aattaagtcgcttaACACAGCCAGTCGAGTGCATCGCGT-3'; target sequence in upper case) that included sequences for the predicted 5' and 3' ends. Cloning, expression and purification were performed as described for YP\_749275.1. Purified YP\_001095227.1 was concentrated to 17.4 mg ml<sup>-1</sup> by centrifugal ultrafiltration (Millipore) for crystallization trials. YP\_001095227.1 was crystallized using the nanodroplet vapor-diffusion method as described for YP\_749275.1. The crystallization reagent for YP\_001095227.1 consisted of 0.2 M NaCl, 37% 2-methyl-2,4-pentanediol and 0.1 M Tris pH 7.33. A rod-shaped crystal of approximate dimensions 70  $\times$  30  $\times$  300  $\mu\text{m}$  was harvested after 36 days at 277 K. Initial screening for diffraction was carried out as described for YP\_749275.1. The diffraction data were indexed in orthorhombic space group C222<sub>1</sub>. The oligomeric state of YP\_001095227.1 in solution was determined using a 0.8  $\times$  30 cm Shodex Protein KW-803 column (Thomson Instruments) pre-calibrated with gel-filtration standards (Bio-Rad).

### 2.2. Data collection, structure solution and refinement

For YP\_749275.1 and YP\_001095227.1, selenium multiwavelength anomalous diffraction (MAD) data were collected on beamline 11-1 at SSRL at wavelengths corresponding to the inflection, high-energy remote and peak. The data sets were collected at 100 K using a MAR 325 CCD detector and the *BLU-ICE* data-collection environment (McPhillips *et al.*, 2002). The MAD data were integrated using *MOSFLM* (Leslie, 1992) and scaled with the program *SCALA* from the *CCP4* suite (Collaborative Computational Project, Number 4, 1994). Phasing was performed with *SHELXD* (Schneider & Sheldrick, 2002) and *autoSHARP* (Bricogne *et al.*, 2003), which resulted in a mean figure of merit of 0.35 to 1.8 Å resolution for YP\_749275.1 with five selenium sites and of 0.43 to 2.25 Å resolution for YP\_001095227.1 with six selenium sites. Automatic model building was performed with *ARP/wARP* (Cohen *et al.*, 2004). Model completion and refinement were performed with *Coot* (Emsley & Cowtan, 2004) and *REFMAC5.2* (Winn *et al.*, 2003) using the inflection-wavelength data. The refinement of YP\_749275.1 included experimental phase restraints in the form of Hendrickson-Lattman coefficients from *SHARP* and TLS refinement with one TLS group per chain. The refinement of YP\_001095227.1 included experimental phase restraints, NCS restraints (positional weight 0.5 and thermal weight 2.0) and TLS refinement with one TLS group per chain. Data-collection and refinement statistics are summarized in Table 1.

### 2.3. Validation and deposition

The quality of the crystal structure was analyzed using the JCSG quality-control server. This server verifies the stereochemical quality of the model using *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Davis *et al.*, 2007), *WHATIF* (Vriend, 1990) and *RESOLVE*

**Table 1**

Summary of crystal parameters, data collection and refinement statistics for YP\_749275.1 and YP\_001095227.1.

Values in parentheses are for the highest resolution shell.

Protein (PDB code)	YP_749275.1 (2ook)			YP_001095227.1 (2q3l)		
Space group	C2			C222 <sub>1</sub>		
Unit-cell parameters (Å, °)	$a = 80.58, b = 40.418, c = 78.05, \beta = 92.2$			$a = 40.39, b = 113.96, c = 130.69$		
Data collection	$\lambda_1$ Se	$\lambda_2$ Se	$\lambda_3$ Se	$\lambda_1$ Se	$\lambda_2$ Se	$\lambda_3$ Se
Wavelength (Å)	0.9184	0.9791	0.9788	0.9184	0.9793	0.9790
Resolution range (Å)	28.6–1.80	28.6–1.80	28.6–1.85	28.7–2.25	28.7–2.25	28.7–2.25
	(1.85–1.80)	(1.85–1.80)	(1.90–1.85)	(2.31–2.25)	(2.31–2.25)	(2.31–2.25)
No. of observations	85996 (6294)	85818 (6269)	79298 (5823)	104546 (7445)	104441 (7352)	105302 (7527)
No. of unique reflections	22625 (1640)	22618 (1639)	20914 (1518)	14575 (1060)	14580 (1043)	14628 (1053)
Completeness (%)	96.5 (95.3)	96.5 (95.0)	96.5 (95.7)	99.0 (100.0)	98.9 (100.0)	98.9 (100.0)
Mean $I/\sigma(I)$	13.6 (2.1)	13.3 (1.7)	13.1 (1.9)	17.8 (3.6)	17.2 (3.1)	16.4 (2.8)
$R_{\text{merge}}$ on $I^\dagger$	0.056 (0.653)	0.056 (0.763)	0.059 (0.672)	0.067 (0.467)	0.071 (0.546)	0.075 (0.629)
$R_{\text{meas}}$ on $I^\ddagger$	0.065 (0.758)	0.066 (0.886)	0.069 (0.781)	0.072 (0.504)	0.077 (0.588)	0.081 (0.678)
Model and refinement statistics						
Data set used in refinement	$\lambda_1$ MAD Se			$\lambda_1$ MAD Se		
Cutoff criterion	$ F  > 0$			$ F  > 0$		
$R_{\text{cryst}}^\S$	0.183			0.181		
$R_{\text{free}}^\P$	0.233			0.239		
Resolution range (Å)	28.6–1.80			28.7–2.25		
No. of reflections (total)	22625			14570		
No. of reflections (test set)	1162			754		
Completeness (%)	96.5			98.8		
Stereochemical parameters						
Restraints (r.m.s.d. observed)						
Bond lengths (Å)	0.012			0.017		
Bond angles (°)	1.53			1.76		
Average isotropic $B$ value (Å <sup>2</sup> )	35.4			51.3		
ESU <sup>††</sup> based on $R_{\text{free}}$ value (Å)	0.141			0.216		
Protein residues/atoms	247/2026			248/1962		
Water molecules/other solvent molecules	224/6			95/13		

<sup>†</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ . <sup>‡</sup>  $R_{\text{meas}} = \frac{\sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$  (Diederichs & Karplus, 1997). <sup>§</sup>  $R_{\text{cryst}} = \frac{\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{hkl} |F_{\text{obs}}|}$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure-factor amplitudes, respectively. <sup>¶</sup>  $R_{\text{free}}$  is the same as  $R_{\text{cryst}}$ , but for 5.1% (2ook) and 5.2% (2q3l) of the total number of reflections that were chosen at random and omitted from refinement. <sup>††</sup> Estimated overall coordinate error (Cruickshank, 1999; Collaborative Computational Project, Number 4, 1994).

(Terwilliger, 2003), as well as several in-house scripts, and summarizes the outputs. Protein quaternary-structure analysis was carried out using the PISA server (Krissinel & Henrick, 2007). Fig. 1(c) was adapted from ESPript (Gouet *et al.*, 1999) and all other figures were prepared with PyMOL (DeLano Scientific). The atomic coordinates and experimental structure factors for YP\_749275.1 and YP\_001095227.1 have been deposited in the PDB under codes 2ook and 2q3l, respectively.

#### 2.4. Orientation of proteins in membranes

The spatial orientations of membrane-associated proteins with respect to the lipid bilayer, including the maximal penetration depths ( $D$ ) of protein residues in the hydrocarbon core and the free energy of transfer of the protein from water to the membrane ( $\Delta G_{\text{transf}}$ ), were calculated using the PPM program, as previously described by Lomize *et al.* (2006). Two major contributions are considered in the current version of PPM: (i) the water–lipid transfer energy calculated with atomic solvation parameters (favorable for nonpolar C and S atoms and unfavorable for polar N and O atoms) and (ii) the de-ionization penalty for charged residues. The smoothing function with a decay parameter of 1 Å was used to describe the gradual polarity changes at the lipid head group–hydrocarbon core boundary.

### 3. Results

#### 3.1. Overall structure

The crystal structures of YP\_749275.1 and YP\_001095227.1 (Fig. 1) were determined independently by the MAD method to 1.80 and 2.25 Å resolution, respectively. Data-collection and refinement

statistics are summarized in Table 1. The final model of YP\_749275.1 consists of a protein dimer (residues 2–126 for chain *A* and residues 3–79 and 82–126 for chain *B*), six ethylene glycols and 224 water molecules in the asymmetric unit. Similarly, the structure of YP\_001095227.1 consists of a protein dimer (residues 0–125 for chain *A* and residues 0–48 and 53–125 for chain *B*), two sodium ions, two chloride ions, nine 2-methyl-2,4-pentanediol molecules and 95 water molecules in the asymmetric unit.

The Matthews coefficient ( $V_M$ ; Matthews, 1968) and estimated solvent content are 2.16 Å<sup>3</sup> Da<sup>−1</sup> and 43.0% for YP\_749275.1, and 2.63 Å<sup>3</sup> Da<sup>−1</sup> and 53.3% for YP\_001095227.1. The Ramachandran plot produced by MolProbity shows that 98.8% (YP\_749275.1) and 97.5% (YP\_001095227.1) of the residues are in favored regions, with no outliers.

YP\_749275.1 and YP\_001095227.1 both adopt the SpoIIAA-like fold according to SCOP and CATH (Cuff *et al.*, 2009). This fold consists of four turns of  $\beta/\alpha$  superhelix with an additional N-terminal  $\beta$ -strand. The five strands and four helices are arranged in the order  $\beta_1$ – $\beta_2$ – $\alpha_1$ – $\beta_3$ – $\alpha_2$ – $\beta_4$ – $\alpha_3$ – $\beta_5$ – $\alpha_4$ . The two proteins share the same topology (Fig. 1), although some noticeable differences are apparent in the lengths of the  $\beta$ -strands and  $\alpha$ -helices.

#### 3.2. Comparison of the YP\_749275.1 and YP\_001095227.1 structures

Despite sharing high sequence identity (54%), the two structures align with an overall r.m.s.d. of 4.2 Å (123 aligned C $\alpha$  atoms). However, most of the deviations occur around the  $\alpha_2$  and  $\alpha_3$  helices (Fig. 2a). The r.m.s.d. decreases to 1.6 Å over 92 C $\alpha$  atoms when these two helices are excluded from the alignment.



region' where the two orthologs adopt distinct conformations. This cluster includes residues from  $\beta 1$  (His6, Gly7),  $\alpha 1$  (Gly27, Leu29, Thr30, His31 and Tyr34) and  $\alpha 2$  (Ala69, Ala70, Trp71, Asp72, Asp73 and Gly77). It is noteworthy that some conserved residues appear to stabilize the 'closed' conformation (Fig. 4c), whereas others stabilize the 'open' conformation. For example, in the 'closed' state, Trp71 is buried from the solvent and participates in multiple van der Waals interactions with surrounding aromatic and aliphatic residues (Leu8, Leu29, Tyr34, Leu74 and Trp65), while His6 and His31 face the solvent and are not involved in any stabilizing hydrogen bonds. Conversely, in the 'open' state, Trp71 is solvent-exposed and may participate in protein-membrane interactions, whereas His6 and His31 form stabilizing hydrogen bonds with Asp33 and Asp72, respectively. In the open conformation, two other conserved residues, Tyr34 and Asp73, are located inside the hydrophobic cavity and are hydrogen bonded to each other (Fig. 4b). The conservation of these residues indicates their functional importance and suggests that they may be involved in hydrogen bonding and/or ionic interactions with a bound ligand.

This 'switch-region' cluster is supplemented by two residues from the adjacent subunit in the dimer (Ile12 and Arg14 in the  $\beta 1$  strand). Arg14 of one subunit hydrogen bonds to Asp33 of the other subunit, which is likely to provide some stability and specificity to the dimer formation (YP\_749275.1). Ile12 engages in hydrophobic interactions

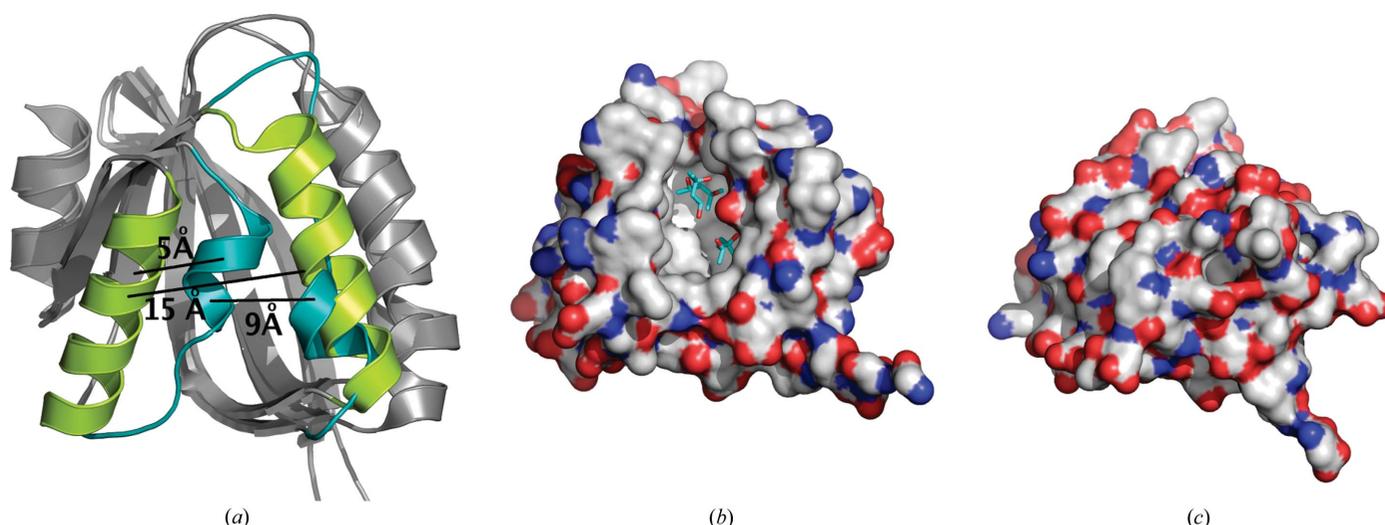
with Met40 (Val40 in YP\_001095227.1). However, Arg14 does not form any contacts in the 'open' monomeric structure (YP\_001095227.1). Thus, the conserved Ile12 and Arg14 may contribute to stabilization of the dimeric state of the protein in solution.

The second cluster consists of residues from  $\beta 4$  (Ala88 and Gly91),  $\beta 5$  (Phe114) and  $\alpha 4$  (Ala120, Trp123 and Leu124) that aid the interaction of  $\alpha 4$  with the  $\beta$ -sheet.

### 3.5. Comparison with other structures

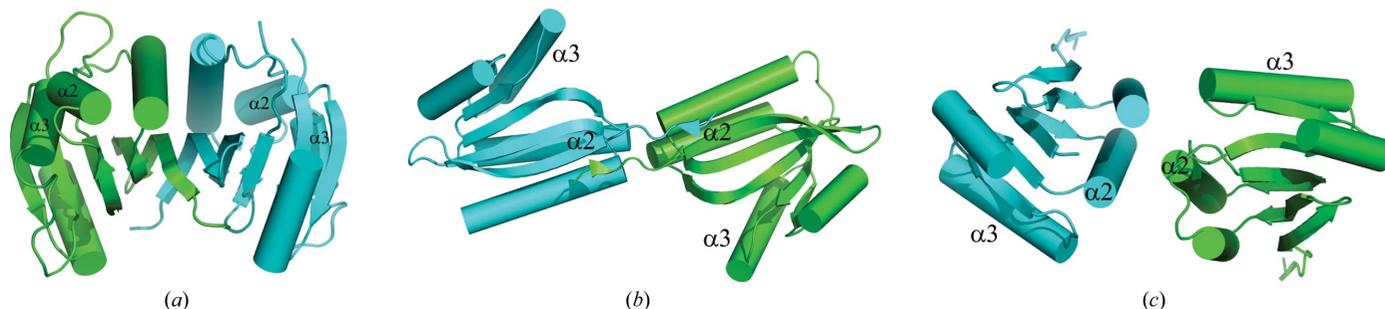
Based on the SCOP classification, the SpoIIAA-like fold consists of two structural superfamilies: (i) the bacterial sporulation anti-sigma factor antagonist SpoIIAA superfamily that contains the STAS domain (PF01740 in Pfam) and (ii) the CRAL-TRIO superfamily of eukaryotic carriers of nonpolar substances (PF03765 and PF00650 in Pfam).

YP\_749275.1 and YP\_001095227.1 have both been assigned to a novel Sfri0576-like family in the SpoIIAA-like superfamily in SCOP. However, their structures can be aligned, without significant insertions or deletions, with structures from both superfamilies, although a higher DALI Z score was obtained for SpoIIAA proteins (Holm & Sander, 1995). Structural alignment using the SSM server (<http://www.ebi.ac.uk/msd-srv/ssm/>) of YP\_001095227.1 (PDB code 2q3l) with SpoIIAA from *Bacillus sphaericus* (PDB code 1h4z; Seavers *et*



**Figure 2**

Structural comparison between YP\_749275.1 and YP\_001095227.1. (a) Superposition of the two structures underscores their overall similarity. The major differences arise from the different relative positions and orientations of helices  $\alpha 2$  and  $\alpha 3$ . These two helices in YP\_749275.1 (cyan) are close together, while they separate in YP\_001095227.1 (green). (b) A surface representation of YP\_001095227.1 in the 'open' state shows the presence of a wide cavity. The C atoms are colored grey, O atoms red and N atoms blue. The 2-methyl-2,4-pentanediol (MPD) molecules are shown as cyan and red sticks. The opening and closing of the cavity is regulated by the movement of helix  $\alpha 3$  away from or towards  $\alpha 2$ , which involves an  $\sim 5$  Å translation and a  $90^\circ$  rotation of this helix. (c) YP\_749275.1 reveals no cavity in the 'closed' state of the protein.



**Figure 3**

Dimerization modes. (a) Monomers in YP\_749275.1 assemble with  $\beta$ -sheets lining up side by side to form a large, extended  $\beta$ -sheet. (b) YP\_001095227.1 has a different dimerization mode with a more limited buried surface area. (c) Possible alternate mode of dimer formation in YP\_001095227.1. Helices  $\alpha 2$  and  $\alpha 3$  are labeled.

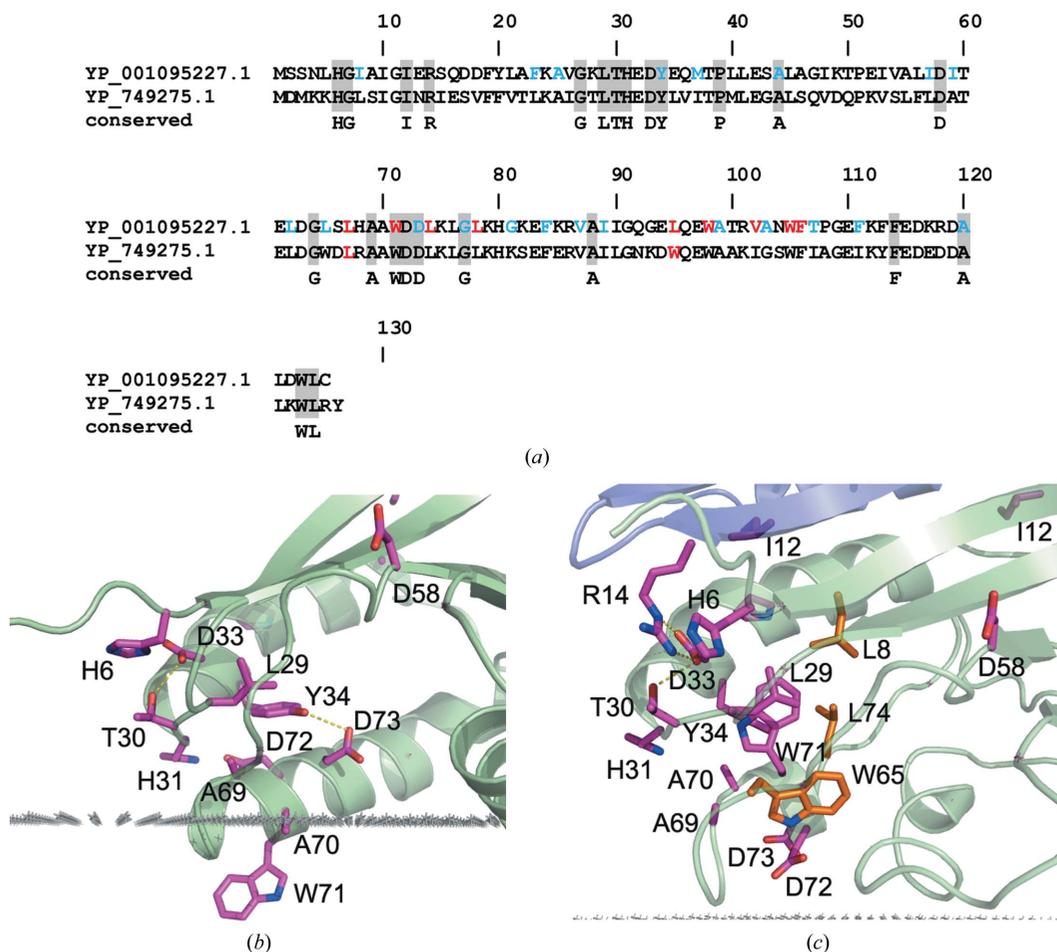
*al.*, 2001) gave an r.m.s.d. of 2.6 Å over 99 C $\alpha$  atoms. Alignment with CRAL-TRIO domains, such as  $\alpha$ -tocopherol transfer proteins ( $\alpha$ TTPs), led to almost the same r.m.s.d. Superposition of the corresponding ‘open’ structures (YP\_001095227.1 and PDB entry 1oiz; Meier *et al.*, 2003) yielded an r.m.s.d. of 3.0 Å over 104 C $\alpha$  atoms, while superposition of the ‘closed’ structures (YP\_749275.1 and PDB entry 1r5l; Min *et al.*, 2003) resulted in an r.m.s.d. of 2.9 Å over 95 C $\alpha$  atoms. The close superposition of these entire domains indicates a possible common evolutionary origin for all these proteins.

Despite their structural similarity, the sequence identity between either YP\_749275.1 or YP\_001095227.1 and proteins in the SpoIIAA-like fold families is <15%. Comparison of the sequences of different STAS domains (PDB codes 1h4z, 1til, 1tid, 1sbo, 1auz, 2vy9 and 3f43; Seavers *et al.*, 2001; Masuda *et al.*, 2004; Etezady-Esfarjani *et al.*, 2006; Kovacs *et al.*, 1998; Marles-Wright *et al.*, 2008) identified a **GxLxH** motif in some of these proteins (Seavers *et al.*, 2001). A similar motif is conserved in YP\_749275.1 and YP\_001095227.1 (<sup>27</sup>**GKLTH**). Although the Gly and Leu residues play a structural role in providing the tight turn between the  $\beta$ 1-strand and  $\alpha$ 2-helix, the conservation of His in the STAS-domain proteins cannot easily be explained. On the other hand, the phosphorylatable serine that is conserved in all SpoIIAA (Ser58 in 1auz, Ser57 in 1h4x; Seavers *et al.*, 2001) is substituted by negatively charged Glu/Asp residues in the majority of other members of the PB000640 Pfam-B family, including YP\_

749275.1 (Asp66), although Ser is present in YP\_001095227.1 (Ser66). This key serine is located at the start of helix  $\alpha$ 2 and participates in the interaction of SpoIIAA with SpoIIAB and a nucleotide ligand. In the presence of ADP, the SpoIIAA–SpoIIAB complex is stable, while in the presence of ATP, SpoIIAA becomes phosphorylated and then dissociates (Aravind & Koonin, 2000; Najafi *et al.*, 1996). The lack of conservation of this serine in YP\_749275.1 indicates a possible loss of functional similarity to SpoIIAA.

The YP\_749275.1 and YP\_001095227.1 structures suggest that these proteins can adopt open and closed conformations (Fig. 2). This situation differs from bacterial STAS proteins, the structures of which are essentially ‘closed’ with  $\alpha$ 2 and  $\alpha$ 3 tightly packed, occluding any possible cavity formation; an open state has not yet been observed in STAS-domain proteins.

The presence of a deep cavity in YP\_001095227.1 is similar to the eukaryotic Sec14-like proteins, which also adopt the same SpoIIAA-like fold. Sec14-like proteins have a lipid-binding CRAL-TRIO domain that participates in the transport of hydrophobic substances such as lipids or  $\alpha$ -tocopherol. Like YP\_749275.1 and YP\_001095227.1, yeast and human Sec14 proteins (PDB codes 1aua, 1oiz, 1r5l, 1o6u and 3b7n; Sha *et al.*, 1998; Meier *et al.*, 2003; Min *et al.*, 2003; Stocker *et al.*, 2002; Schaaf *et al.*, 2008) have been crystallized in two alternative conformations which differ in the relative disposition of two helices located at the entrance to a hydrophobic ligand-binding



**Figure 4** Conserved residues in YP\_749275.1 and YP\_001095227.1. (a) A sequence alignment with other members of the PFAM PB000640 family (not shown) reveals several conserved residues (marked in grey boxes). Residues from the binding cavity are colored blue and residues that are predicted to penetrate to the lipid bilayer are colored red. These residues are indicated on the structure in (b) for YP\_001095227.1 and (c) for YP\_749275.1. The main cluster of conserved residues from strand  $\beta$ 1 and helices  $\alpha$ 1 and  $\alpha$ 2 is shown in purple. The protein backbone is shown in a cartoon representation. The calculated membrane boundary is shown by grey dots. A few additional nonconserved residues involved in hydrophobic interactions (Leu8, Leu74 and Trp65) in YP\_749275.1 are shown in orange.

cavity. In particular, human  $\alpha$ TTP (PDB code 1oiz) was obtained in a detergent-bound 'open' structure (Meier *et al.*, 2003) with the 'lid' helices  $\alpha$ 9 and  $\alpha$ 11 moved apart (Fig. 5c) as well as in a ligand-bound 'closed' structure (PDB code 1r5l; Min *et al.*, 2003) with  $\alpha$ 11 shifted towards  $\alpha$ 9 and blocking the entrance to the cavity (Fig. 5d).

### 3.6. Predicted protein–membrane association

Calculations using the *PPM* method (Lomize *et al.*, 2006) show that the 'open' conformation of YP\_001095227.1 can associate with the lipid bilayer by immersing its exposed nonpolar residues from  $\alpha$ 2 and  $\alpha$ 3 into the lipid acyl-chain region. The predicted depth of residue penetration into the hydrophobic core of the membrane is  $6.5 \pm 0.4$  Å and the calculated water–membrane transfer energy ( $\Delta G_{\text{transf}}$ ) is  $-47.3$  kJ mol $^{-1}$ . However, the corresponding protein membrane binding energy is expected to be smaller than the transfer energy, since part of the transfer energy must be spent on protein conformational change. The lipid-interaction residues include Leu67, Trp71, Leu74 and Leu78 from  $\alpha$ 2, and Leu95, Trp98, Val102, Trp105 and Phe106 from  $\alpha$ 3 (Fig. 5a).

In contrast, the 'closed' conformation of YP\_749275.1 was predicted to form a stable dimer by the *PISA* server, as shown in Fig. 5(b). However, this dimer is visibly asymmetric as  $\alpha$ 3 is longer by one helical turn in subunit 2 compared with subunit 1. Furthermore, two hydrophobic residues (Leu67 in  $\alpha$ 2 and Trp95 in  $\alpha$ 3) are solvent-exposed in one subunit (indicated in purple) but buried from solvent in another subunit. These two solvent-exposed nonpolar residues may anchor the protein at the hydrophobic boundary of the lipid bilayer (Fig. 5b). However, the calculated depth of penetration is only

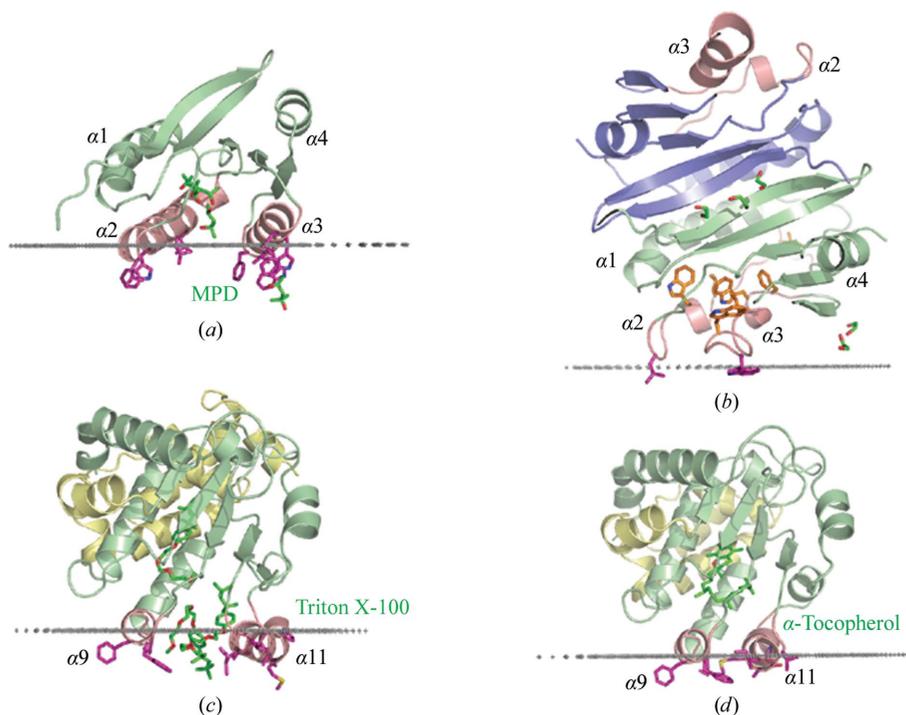
1.2 Å and  $\Delta G_{\text{transf}}$  is only  $-12.6$  kJ mol $^{-1}$ , indicating a weak association.

### 3.7. Predicted protein–protein interactions

A genomic neighborhood search performed using *STRING* (<http://string.embl.de>) relates both YP\_749275.1 and YP\_001095227.1 to a TonB-dependent receptor precursor (a co-occurrence in the same species) and a universal stress protein (UspA) domain (localization in the close genetic neighborhood). One of the homologous proteins (UniProt ID Q083D4\_SHEFN) from PfamB family PB000640 is fused to a UspA domain. A *PSI-BLAST* (Altschul *et al.*, 1997) search returns about three dozen bacterial proteins (mostly proteobacterial) using a sequence-identity cutoff of 25%. Most of these proteins have no functional annotation. A few contain a UspA domain, including Q083D4\_SHEFN, albeit with low *e*-value scores. Although the neighborhood-matching and the *BLAST* search scores are not significant enough to confer a definitive link between these proteins, they may suggest a role of these proteins in the stress-response pathway.

## 4. Discussion

YP\_749275.1 and YP\_001095227.1 can now be assigned to a new bacterial protein family which adopts the SpoIIAA-like structural fold. The structures suggest that these proteins are metamorphic, adopting two distinct conformations (open and closed) which are stabilized under different environmental conditions (Murzin, 2008). We suggest that the predicted anchoring of YP\_749275.1 and



**Figure 5**

Comparison of the putative membrane association of YP\_001095227.1 and YP\_749275.1 with that of the CRAL-TRIO domain of human  $\alpha$ -tocopherol transfer protein ( $\alpha$ -TTP). (a) YP\_001095227.1 in the 'open' conformation (PDB code 2q3l). (b) Dimeric form of YP\_749275.1 in the 'closed' conformation (PDB code 2ook). The second molecule is colored blue and the lid helices enclosing the binding cavity are colored pink. Residues that move from the surface to the protein interior during the conformational switch are colored orange. (c) Human  $\alpha$ -TTP in the 'open' conformation (PDB code 1oiz; Meier *et al.*, 2003). The N-CRAL-TRIO domain is colored yellow and the lipid-binding CRAL-TRIO domain is colored green. Molecules of detergents or bound ligands are colored in dark green here and in (d). (d)  $\alpha$ -TTP in the 'closed' conformation (PDB code 1r5l; Min *et al.*, 2003). In all figures, residues that penetrate or are proposed to penetrate the lipid bilayer are colored purple. Calculated boundaries between lipid head groups and the acyl-chain region are shown by grey dots.

YP\_001095227.1 to the lipid bilayer *via* helices  $\alpha 2$  and  $\alpha 3$  would induce a switch from the ‘closed’ to the ‘open’ conformation. Both proteins presumably exist as stable water-soluble dimers in their ‘closed’ conformation which can weakly associate with the lipid bilayer *via* nonpolar residues from the ‘lid’ helices (Fig. 5*b*). Membrane binding would promote protein activation owing to the rearrangement of the ‘lid’ helices and subsequent dimer dissociation. The membrane-associated protein in the ‘open’ conformation (Fig. 5*a*) may then bind amphiphilic ligands which have accumulated at the membrane interface.

The cellular localization of YP\_749275.1 and YP\_001095227.1 is currently unknown. We suggest that these proteins are cytoplasmic based on the following observations. Firstly, proteins from the same PfamB family are found in Gram-positive bacterial and archaeal species that lack the outer membrane and periplasmic space. Secondly, YP\_001095227.1 has a single Cys125 that is likely to remain reduced in the bacterial cytoplasm (Ritz & Beckwith, 2001). Finally, one protein from the family is fused with the UspA domain, which is a cytoplasmic protein involved in the stress-response pathway.

We further suggest that YP\_749275.1 and YP\_001095227.1 may function either as water-soluble carriers of hydrophobic compounds, similar to the related CRAL-TRIO domains, or as interfacially activated enzymes. If these proteins are ligand carriers, then they can dissociate from the membrane in their ligand-loaded state. If these proteins are interfacially activated enzymes, they would remain membrane-associated while performing their chemical reactions. The presence of conserved Tyr34 and Asp73 residues at the entrance to the hydrophobic cavity may indicate possible enzymatic activity rather than simply the formation of a hydrogen bond to a ligand. For example, similar pairs of hydrogen-bonded residues (usually a Tyr–Glu pair) are found in a glycoside hydrolase (PDB code 2fhr; Watts *et al.*, 2006) and in glycosyltransferases (PDB code 1s2g; Anand *et al.*, 2004). In these enzymes, the Tyr and Glu residues form a nucleophile that interacts with the OH group of the substrate.

The natural ligands for these bacterial proteins remain unknown. The shape and the hydrophobic character of the cavity indicate a binding site for relatively large and poorly soluble compounds such as flavins, naphthoquinones or other substituted heterocycles. It is noteworthy that riboflavin and menaquinone are important for the growth of *Shewanella* cells on poorly soluble minerals, as they participate in the electron transfer to low-potential electron acceptors (Newman & Kolter, 2000; Marsili *et al.*, 2008). *Shewanella* produces significant amounts of flavins (riboflavin and riboflavin-5'-phosphate) that mediate extracellular electron transfer, leading to reduction, chelation and uptake of ferric iron by the cells (Marsili *et al.*, 2008). The uptake of iron complexes (with riboflavin or hydroxamate) can be facilitated by the TonB-dependent transport system (Schauer *et al.*, 2008). Therefore, the genomic neighborhood link between YP\_749275.1/YP\_001095227.1 and the TonB receptor (which also has a high co-occurrence with the nicotinamide mononucleotide transporter PnuC) may be of functional significance. In addition, the connection with UspA domains suggests a possible role of these proteins in the stress-response pathway.

Additional information about YP\_749275.1 and YP\_001095227.1 is available from TOPSAN (Krishna *et al.*, 2010) at <http://www.topsan.org/explore?PDBid=2ook> and <http://www.topsan.org/explore?PDBid=2q3l>.

This work was supported by the National Institutes of Health Protein Structure Initiative grant Nos. P50 GM62411 and U54 GM074898 from the National Institute of General Medical Sciences (<http://www.nigms.nih.gov>). Portions of this research were carried out

at the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences). *Shewanella* sp. PV-4 and *S. frigidimarina* NCIMB-400 were gifts from Dr Jim K. Fredrickson, Pacific Northwest National Laboratory, Washington, USA. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

## References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
- Anand, R., Kaminski, P. A. & Ealick, S. E. (2004). *Biochemistry*, **9**, 2384–2393.
- Aravind, L. & Koonin, E. V. (2000). *Curr. Biol.* **10**, R53–R55.
- Binkowski, T. A., Naghibzadeh, S. & Liang, J. (2003). *Nucleic Acids Res.* **31**, 3352–3355.
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* **D59**, 2023–2030.
- Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
- Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* **D60**, 2222–2229.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cruickshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.
- Cuff, A. L., Sillitoe, I., Lewis, T., Redfern, O. C., Garratt, R., Thornton, J. & Orengo, C. A. (2009). *Nucleic Acids Res.* **37**, D310–D314.
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.
- Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Etezady-Esfarjani, T., Placzek, W. J., Herrmann, T. & Wüthrich, K. (2006). *Magn. Reson. Chem.* **44**, S61–S70.
- Finn, R. D., Tate, J., Mistry, J., Coghill, 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. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
- Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
- Hubbard, T. J., Ailey, B., Brenner, S. E., Murzin, A. G. & Chothia, C. (1999). *Nucleic Acids Res.* **27**, 254–256.
- Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. (2005). *Nucleic Acids Res.* **33**, W284–W288.
- Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). *Proteins*, **71**, 982–994.
- Kovacs, H., Comford, D., Lord, M., Campbell, I. D. & Yudkin, M. D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5067–5071.
- Krishna, S. S., Weekes, D., Bakolitsa, C., Elsliger, M.-A., Wilson, I. A., Godzik, A. & Wooley, J. (2010). *Acta Cryst.* **F66**, 1143–1147.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
- Lomize, A. L., Pogozheva, I. D., Lomize, M. A. & Mosberg, H. I. (2006). *Protein Sci.* **15**, 1318–1333.
- Marles-Wright, J., Grant, T., Delumeau, O., van Duinen, G., Firbank, S. J., Lewis, P. J., Murray, J. W., Newman, J. A., Quin, M. B., Race, P. R., Rohou, A., Tichelaar, W., van Heel, M. & Lewis, R. J. (2008). *Science*, **322**, 92–96.
- Marsili, E., Baron, D. B., Shikhare, I. D., Coursolle, D., Gralnick, J. A. & Bond, D. R. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 3968–3973.
- Masuda, S., Murakami, K. S., Wang, S., Olson, C. A., Donigian, J., Leon, F., Darst, S. A. & Campbell, E. A. (2004). *J. Mol. Biol.* **340**, 941–956.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
- Meier, R., Tomizaki, T., Schulze-Briese, C., Baumann, U. & Stocker, A. (2003). *J. Mol. Biol.* **331**, 725–734.
- Min, K. C., Kovall, R. A. & Hendrickson, W. A. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 14713–14718.
- Murzin, A. G. (2008). *Science*, **320**, 1725–1726.
- Najafi, S. M., Harris, D. A. & Yudkin, M. D. (1996). *J. Bacteriol.* **178**, 6632–6634.
- Newman, D. K. & Kolter, R. (2000). *Nature (London)*, **405**, 94–97.
- Ritz, D. & Beckwith, J. (2001). *Annu. Rev. Microbiol.* **55**, 21–48.
- Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G. & Stevens, R. C. (2002). *J. Appl. Cryst.* **35**, 278–281.
- Schaaf, G., Ortlund, E. A., Tyeryar, K. R., Mousley, C. J., Ile, K. E., Garrett, T. A., Ren, J., Woolls, M. J., Raetz, C. R., Redinbo, M. R. & Bankaitis, V. A. (2008). *Mol. Cell*, **29**, 191–206.
- Schauer, K., Rodionov, D. A. & de Reuse, H. (2008). *Trends Biochem. Sci.* **33**, 330–338.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst. D* **58**, 1772–1779.
- Seavers, P. R., Lewis, R. J., Brannigan, J. A., Verschueren, K. H., Murshudov, G. N. & Wilkinson, A. J. (2001). *Structure*, **9**, 605–614.
- Sha, B., Phillips, S. E., Bankaitis, V. A. & Luo, M. (1998). *Nature (London)*, **391**, 506–510.
- Stocker, A., Tomizaki, T., Schulze-Briese, C. & Baumann, U. (2002). *Structure*, **10**, 1533–1540.
- Terwilliger, T. C. (2003). *Acta Cryst. D* **59**, 1174–1182.
- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Watts, A. G., Oppezzo, P., Withers, S. G., Alzari, P. M. & Buschiazio, A. (2006). *J. Biol. Chem.* **281**, 4149–4155.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst. D* **60**, 1833–1839.