

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

^aJoint Center for Structural Genomics, <http://www.jcsg.org>, USA, ^bDepartment of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA, ^cProgram on Bioinformatics and Systems Biology, Burnham Institute for Medical Research, La Jolla, CA, USA, ^dStanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA, ^eDépartement de Biochimie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke (Quebec), Canada, ^fProtein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA, ^gCenter for Research in Biological Systems, University of California, San Diego, La Jolla, CA, USA, and ^hPhoton Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Correspondence e-mail: wilson@scripps.edu

Received 25 September 2009

Accepted 23 November 2009

PDB References: SPO0140 from *Silicibacter pomeroyi* DSS-3, 2re3; Sbal_2486 from *Shewanella baltica*, 2ra9.

Structures of the first representatives of Pfam family PF06938 (DUF1285) reveal a new fold with repeated structural motifs and possible involvement in signal transduction

The crystal structures of SPO0140 and Sbal_2486 were determined using the semiautomated high-throughput pipeline of the Joint Center for Structural Genomics (JCSG) as part of the NIGMS Protein Structure Initiative (PSI). The structures revealed a conserved core with domain duplication and a superficial similarity of the C-terminal domain to pleckstrin homology-like folds. The conservation of the domain interface indicates a potential binding site that is likely to involve a nucleotide-based ligand, with genome-context and gene-fusion analyses additionally supporting a role for this family in signal transduction, possibly during oxidative stress.

1. Introduction

To extend the structural coverage of proteins with uncharacterized biological function, we targeted Pfam protein family Pfam06938 (DUF1285), for which we determined the structures of two representative members. The *SPO0140* gene of *Silicibacter pomeroyi* DSS-3, a marine α -proteobacterium, encodes a protein with a molecular weight of 21.2 kDa (residues 1–193) and a calculated isoelectric point of 5.7. The *Sbal_2486* gene of *Shewanella baltica* OS155, a psychrotrophic marine γ -proteobacterium, encodes a protein with a molecular weight of 17.7 kDa (residues 1–157) and a calculated isoelectric point of 4.8.

2. Materials and methods

2.1. Protein production and crystallization

The clones for SPO0140 and Sbal_2486 were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method (Klock *et al.*, 2008). The gene encoding SPO0140 (GenBank YP_165412, gi:56695065, UniProt Q5LWU5) was amplified by polymerase chain reaction (PCR) from *Silicibacter pomeroyi* DSS-3 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5'-ctgtacttccaggcATGAGCGG-ACAAAAGCCTGTGAAACC-3'; reverse primer, 5'-aattaagtgcggttaGCCGCGCTCCAGCTCCTCGACCGTCATC-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 (MGSDKIHSHHHHENLYFQ/G), was PCR-amplified with V-PIPE (Vector) primers (forward primer, 5'-taacgcgacttaattaactcgtttaaacggtctcagc-3'; reverse primer, 5'-gccctggaagtaggttttcgtgatgatgatg-3'). The 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 a selenomethionine-containing medium at 310 K with suppression of normal methionine synthesis. 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 the lysate was clarified by centrifugation at 32 500g for 30 min. The soluble fraction was passed over 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 flow-through and wash fractions were combined and concentrated to 13.9 mg ml⁻¹ by centrifugal ultrafiltration (Millipore) for crystallization trials. SPO0140 was crystallized by mixing 200 nl protein solution with 200 nl crystallization solution and using a 50 µl reservoir using the nanodroplet vapor-diffusion method (Santarsiero *et al.*, 2002) with standard Joint Center for Structural Genomics (JCSG; <http://www.jcsg.org>) crystallization protocols (Lesley *et al.*, 2002). The crystallization reagent consisted of 20% (v/v) ethanol and 100 mM Tris pH 8.5. Glycerol was added to a final concentration of 17% (v/v) as a cryoprotectant. A rod-shaped crystal of approximate dimensions 0.08 × 0.02 × 0.02 mm was harvested after 29 d at 277 K for data collection. Initial screening for diffraction was carried out using the Stanford Automated Mounting (SAM) system (Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, California, USA). The diffraction data were indexed in the tetragonal space group *P*₄₃₂₁.

The gene encoding Sbal_2486 (GenBank YP_001050848.1, gi:126174699, UniProt A3D5G6) was amplified from *Shewanella baltica* OS155 genomic DNA. Using the PIPE method (Klock *et al.*, 2008), the initial clone was generated by using I-PIPE (Insert) primers (forward primer, 5'-ctgtactccagggcATGGAAAAGATGACTGACAGTATTCAAC-3'; reverse primer, 5'-aattaagtcgcttaCTGCTCATTAGATCAGATAAATTG-3'; target sequence in upper case) that included sequences for the predicted 5' and 3' ends. Cloning, expression and purification were performed as described above for SPO0140. Crystals obtained from the full-length construct were not suitable for data collection. Bioinformatic predictions suggested that a 12-residue N-terminal truncation might produce better diffracting crystals than the full-length (157 residues) wild-type protein. Cloning attempts were initiated to generate nested truncations in steps of four residues around this prediction for a truncated sequence. Additionally, C-terminal truncations (NB these were not part of the prediction to improve crystallization) were attempted. Truncation clones were successfully generated for the construct boundaries 9–157, 17–157, 21–157, 1–153, 1–149, 1–145 and 1–141. These constructs were screened in parallel for solubility, crystallization and diffraction (Table 1). No clone was obtained for the initially proposed truncation (13–157) construct. The other three N-terminal truncations produced soluble protein that led to harvestable crystals that were of higher quality and diffracted better than the crystals of the full-length protein. The only C-terminal truncation that produced soluble protein was that for residues 1–153. This construct also led to harvestable crystals, which again were of higher quality and diffracted better than the full length construct. The improvement was less than that observed for the N-terminal truncation constructs. By making several truncation constructs and empirically testing all of them, we found a construct that was better suited to crystallization

Table 1

Summary of expression and diffraction screening results for Sbal_2486 constructs.

Diffraction screening was carried out by collecting two diffraction images 90° apart and evaluating the resulting images for resolution, diffraction strength, ice rings and spot quality.

Construct	Soluble	No. of crystals screened	Best screening resolution† (Å)	Median screening resolution (Å)	Best crystal-quality score‡	Best spot-quality score§	Average spot-quality score
1–157	+	213	3.0	6.5	6	8	6.7
1–153	+	19	2.8	5.8	7	8	8.0
1–149	–						
1–145	–						
1–141	–						
9–157	+	68	2.2	4.2	10	9	8.1
17–157	+	31	2.2	3.9	10	10	8.4
21–157	+	34	2.5	4.5	8	9	7.9

† Note that crystals of the full-length construct (1–157) were screened using the synchrotron beam at SSRL, while crystals of the truncated constructs were screened using the X-ray microsource. The resolution from synchrotron data collection is typically 0.6–1.2 Å better than the screening resolution obtained using the microfocussed sealed-tube system. The resolution from synchrotron-screened crystals is more comparable to the final resolution after data collection. ‡ Crystal quality is an overall integer score of 0–10 that is assigned to assess the suitability of the crystal for data collection, with 10 being the best-quality crystals and 0 corresponding to no diffraction. The score is based on resolution, spot quality, diffraction strength, single *versus* multiple lattices, ice-ring pathology and other factors. Only crystals with scores of 5 or better are saved for further evaluation and data collection. § Spot quality is assigned as an integer score of 0–10, with 10 corresponding to nice clean spots, 5–6 corresponding to mostly elongated or anisotropic spots, 1–3 corresponding to split spots and 0 corresponding to extreme streaks or powder-like patterns.

and structure determination. A crystal of the 9–157 construct was used for structure solution. The primers used to generate the 9–157 truncation clone by PIPE mutagenesis were I-PIPE (Insert) forward primer 5'-ctgtactccagggcCAACACACTCAAACAATTCCGCCG-CCG-3' and reverse primer 5'-gccctggaagtagcaggtttctgctgatgatgatgatg-3' (Klock *et al.*, 2008). Purified Sbal_2486 was concentrated to 20 mg ml⁻¹ by centrifugal ultrafiltration (Millipore) for crystallization trials. Sbal_2486 was crystallized by mixing 200 nl protein solution and 200 nl crystallization solution and using a 50 µl reservoir volume using the nanodroplet vapor-diffusion method with standard JCSG crystallization protocols. The crystallization reagent consisted of 20% (v/v) 2-propanol, 20% (w/v) PEG 4000 and 0.1 M sodium citrate pH 5.6. Ethylene glycol (1,2-ethanediol) was added to a final concentration of 10% (v/v) as a cryoprotectant. A rhombohedral crystal of approximate size 0.1 × 0.1 × 0.1 mm was harvested after 20 d at 277 K for data collection. Initial screening for diffraction was carried out using the SAM system and an X-ray microsource (Miller & Deacon, 2007) installed at SSRL. The diffraction data were indexed in the orthorhombic space group *P*₂₁₂₁.

2.2. Data collection, structure solution and refinement

For SPO0140, multiple-wavelength anomalous diffraction (MAD) data were collected on beamline BL11-1 at SSRL at wavelengths corresponding to the inflection (λ_3 -2re3), high-energy remote (λ_1 -2re3) and peak (λ_2 -2re3) of a selenium MAD experiment. The data sets were collected at 100 K with a MAR Mosaic 325 mm CCD detector (Rayonix) using the *Blu-Ice* (McPhillips *et al.*, 2002) data-collection environment. The MAD data were integrated and reduced using *MOSFLM* (Leslie, 1992) and were scaled using the program *SCALA* (Collaborative Computational Project, Number 4, 1994). Selenium-substructure solution and phasing were performed with *SHELXD* (Sheldrick, 2008) and *autoSHARP* (Bricogne *et al.*, 2003) with a mean figure of merit of 0.49 for 12 selenium sites (NB there are five unique Se sites per chain, but SeMet131 adopts two different

Table 2

Summary of crystal parameters, data collection and refinement statistics for SPO0140 (PDB code 2re3).

Values in parentheses are for the highest resolution shell.

	λ_1 -2re3	λ_2 -2re3	λ_3 -2re3
Space group	$P4_22_12$		
Unit-cell parameters	$a = b = 75.37, c = 182.69$		
Data collection			
Wavelength (Å)	0.9184	0.9791	0.9794
Resolution range (Å)	29.5–2.50 (2.56–2.50)	29.5–2.50 (2.56–2.50)	29.5–2.50 (2.56–2.50)
No. of observations	141130	141301	141452
No. of unique reflections	19038	19111	19093
Completeness (%)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
Mean $I/\sigma(I)$	12.3 (2.9)	12.0 (2.6)	12.3 (2.9)
R_{merge} on I † (%)	12.9 (76.0)	13.4 (80.9)	12.5 (71.8)
Model and refinement statistics			
Resolution range (Å)	29.5–2.50		
No. of reflections (total)	18959‡		
No. of reflections (test)	972		
Completeness (%)	99.7		
Data set used in refinement	λ_1 -2re3		
Cutoff criterion	$ F > 0$		
$R_{\text{cryst}}§$	0.215		
$R_{\text{free}}¶$	0.258		
Stereochemical parameters			
Restraints (r.m.s.d, observed)			
Bond lengths (Å)	0.014		
Bond angles (°)	1.56		
Average isotropic B value (Å ²)	25.4		
ESU based on R_{free} value†† (Å)	0.28		
Protein residues/atoms	367/2949		
Water/other solvent molecules	186/1		

† $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. ‡ The number of unique reflections that were used in refinement is typically slightly less than the total number that were integrated and scaled. Reflections are excluded owing to systematic absences, negative intensities and rounding errors in the resolution limits and unit-cell parameters. § $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. ¶ R_{free} is the same as R_{cryst} but for 5.1% of the total reflections chosen at random and omitted from refinement. †† Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

conformations resulting in two partial occupancy 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 remote (λ_1 -2re3) data set. The refinement included experimental phase restraints in the form of Hendrickson–Lattman coefficients from *SHARP*, NCS restraints (positional weight 0.5 and thermal weight 2.0) and TLS refinement with one TLS group per chain. Data-reduction and refinement statistics for SPO0140 are summarized in Table 2.

For Sbal_2486, MAD data were collected on beamline 8.2.2 at the ALS at wavelengths corresponding to the inflection (λ_2 -2ra9), low-energy remote (λ_3 -2ra9) and peak (λ_1 -2ra9) of a selenium MAD experiment. The inflection and remote data were collected first using an interleaved protocol with a 10° wedge size over a total sweep of 100° and the peak data were then collected in a 130° sweep. The data sets were collected at 100 K with a Quantum 315 CCD detector (ADSC). The MAD data were integrated and reduced using *MOSFLM* and were scaled with the program *SCALA*. Selenium-substructure solution and phasing were performed with *SHELXD* and *autoSHARP* with a mean figure of merit of 0.48 for a single selenium site. Automatic model building was performed with *ARP/wARP*. Model completion and refinement were performed with *Coot* (Emsley & Cowtan, 2004) and *REFMAC5.2* (Murshudov *et al.*, 1999) using the peak (λ_1 -2ra9) data set. The refinement included experimental phase restraints in the form of Hendrickson–Lattman coefficients from *SHARP* and restrained anisotropic ADP refinement.

Table 3

Summary of crystal parameters, data collection and refinement statistics for Sbal_2486 (PDB code 2ra9).

Values in parentheses are for the highest resolution shell.

	λ_1 -2ra9	λ_2 -2ra9	λ_3 -2ra9
Space group	$P2_12_12_1$		
Unit-cell parameters (Å)	$a = 38.41, b = 62.29, c = 73.25$		
Data collection			
Wavelength (Å)	0.9795	0.9798	1.0000
Resolution range (Å)	29.9–1.40 (1.44–1.40)	29.9–1.40 (1.44–1.40)	29.8–1.40 (1.44–1.40)
No. of observations	172763	130260	115916
No. of unique reflections	35222	34664	32638
Completeness (%)	99.6 (98.5)	98.2 (91.5)	92.6 (67.8)
Mean $I/\sigma(I)$	16.3 (2.7)	14.1 (2.1)	16.7 (1.9)
R_{merge} on I † (%)	5.8 (61.3)	6.0 (62.9)	4.6 (58.4)
Model and refinement statistics			
Resolution range (Å)	29.9–1.40		
No. of reflections (total)	35168‡		
No. of reflections (test)	1764		
Completeness (%)	99.4		
Data set used in refinement	λ_1 -2ra9		
Cutoff criterion	$ F > 0$		
$R_{\text{cryst}}§$	0.162		
$R_{\text{free}}¶$	0.196		
Stereochemical parameters			
Restraints (r.m.s.d, observed)			
Bond lengths (Å)	0.014		
Bond angles (°)	1.52		
Average isotropic B value (Å ²)	20.1		
ESU based on R_{free} value†† (Å)	0.06		
Protein residues/atoms	127/1032		
Water/other solvent molecules	231/8		

† $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. ‡ The number of unique reflections that were used in refinement is typically slightly less than the total number that were integrated and scaled. Reflections are excluded owing to systematic absences, negative intensities and rounding errors in the resolution limits and unit-cell parameters. § $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. ¶ R_{free} is the same as R_{cryst} but for 5.0% of the total reflections chosen at random and omitted from refinement. †† Estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Cruickshank, 1999).

Data-reduction and refinement statistics for Sbal_2486 are summarized in Table 3.

2.3. Validation and deposition

The quality of the crystal structures was analyzed using the JCSG Quality Control server (<http://smb.slac.stanford.edu/jcsg/QC>). This server processes the coordinates and data through a variety of validation tools including *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Davis *et al.*, 2007), *WHAT IF* v.5.0 (Vriend, 1990), *RESOLVE* (Terwilliger, 2003) and *MOLEMAN2* (Kleywegt, 2000), as well as several in-house scripts, and summarizes the output. Protein quaternary-structure analysis used the *PISA* server (Krissinel & Henrick, 2007). Fig. 1(b) was adapted from an analysis using *PDBsum* (Laskowski *et al.*, 2005) and all others were prepared with *PyMOL* (DeLano Scientific). Atomic coordinates and experimental structure factors for SPO0140 at 2.5 Å resolution and Sbal_2486 at 1.4 Å resolution have been deposited in the PDB and are accessible under codes 2re3 and 2ra9, respectively.

3. Results and discussion

3.1. Overall structure

The crystal structure of SPO0140 (Fig. 1) was determined to 2.5 Å resolution using the MAD method. Data collection, model and refinement statistics are summarized in Table 2. The final model included two protomers (residues 10–192 for chain A, residues 10–193

for chain *B*), one glycerol molecule and 186 water molecules in the asymmetric unit. No electron density was observed for residues Gly0 (which remained at the N-terminus after cleavage of the purification tag), SeMet1–Pro9 in chains *A* and *B* and Gly193 in chain *A*. Side-chain atoms of Lys108, Thr137 and Glu139 in chain *A* and Ser10, Lys108, Gln133, Thr137 and Glu139 in chain *B* had poorly defined electron density and were omitted from the model. The Matthews coefficient (V_M ; Matthews, 1968) was $3.0 \text{ \AA}^3 \text{ Da}^{-1}$ and the estimated solvent content was 59.0%. The Ramachandran plot produced by *MolProbity* (Davis *et al.*, 2007) showed that 97.0% of the residues were in favored regions, with no outliers.

The crystal structure of Sba1_2486 was determined to 1.4 Å resolution using the MAD method. Data collection, model and refinement statistics are summarized in Table 3. The final model included a monomer of 127 residues, seven ethylene glycol molecules, one sodium ion and 231 water molecules in the asymmetric unit. No electron density was observed for residues Gly0 (which remained at

the N-terminus after cleavage of the purification tag), Gln9–Cys28 and Glu156–Gln157. The Matthews coefficient (V_M ; Matthews, 1968) was $2.6 \text{ \AA}^3 \text{ Da}^{-1}$ and the estimated solvent content was 52.5%. The Ramachandran plot produced by *MolProbity* showed that 97.6% of the residues were in favored regions, with no outliers.

SPO0140 is an α/β protein comprising two domains (Fig. 1). The N-terminal domain (residues 10–93) consists of a β -meander– α -helix– β -meander core (residues 24–92), with the two β -meanders hydrogen bonding along the first and sixth strands to form a twisted mixed six-stranded β -sheet. Two N-terminal helices (H1 and H2) pack against the sheet and complete this domain. The same $\beta_3\alpha\beta_3$ unit (β_1 –3, H3 and β_4 –6; residues 36–88) is encountered again in the C-terminal domain (β_{10} –12, H5 and β_{13} –15; residues 129–83), with an additional three-stranded meander (β_7 –9; residues 95–122) packing perpendicularly against the second meander to form a β -sandwich (Fig. 1*a*). Both repeats share the same overall fold and topology and their structural comparison is considered to be signifi-

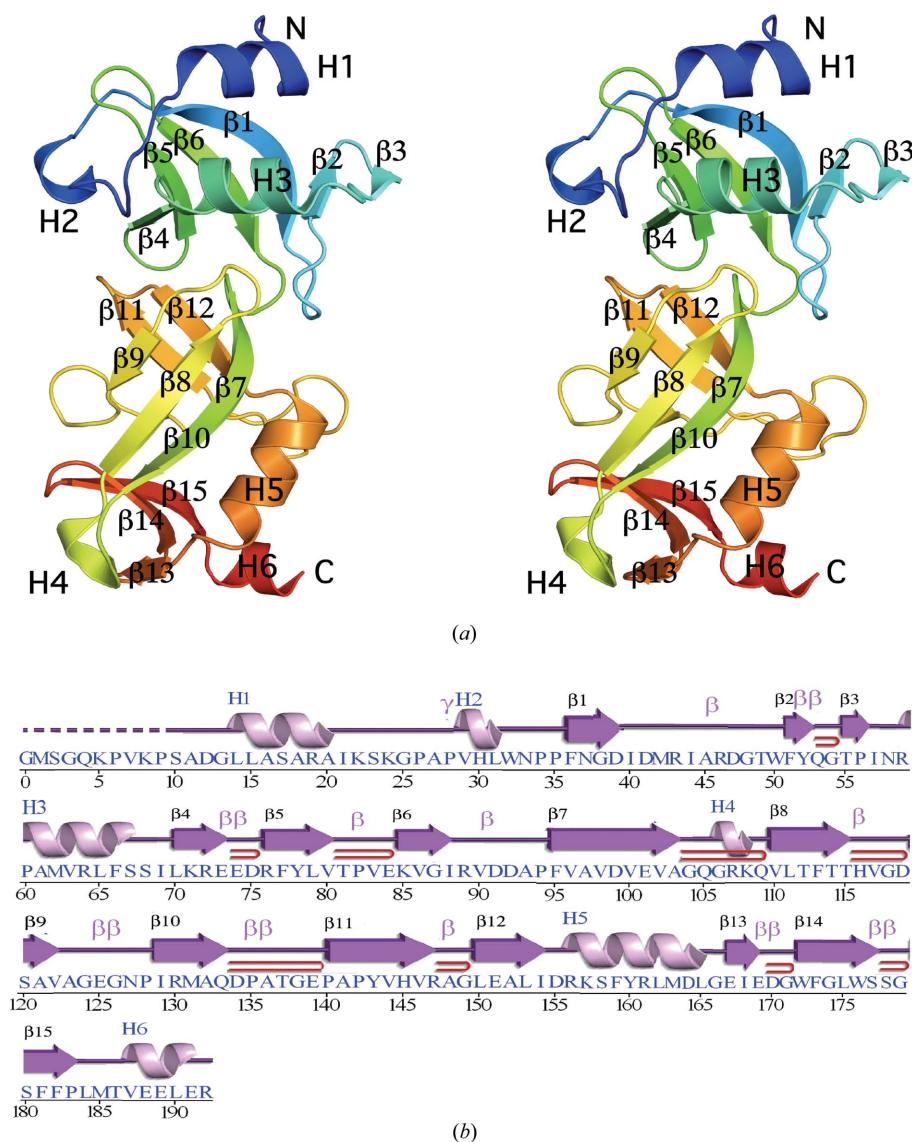


Figure 1 Crystal structure of SPO0140 from *Silicibacter pomeroyi*. (a) Stereo ribbon diagram of the SPO0140 monomer (chain *A*) color-coded from the N-terminus (blue) to the C-terminus (red). Helices (H1–H6) and β -strands (β_1 – β_{15}) are indicated. (b) Diagram showing the secondary-structure elements of SPO0140 superimposed on its sequence in accordance with *PDBsum* (<http://www.ebi.ac.uk/pdbsum>). For SPO0140, the α -helices (H1, H3, H5 and H6), 3_{10} -helices (H2 and H4), β -strands (β_1 – β_{15}), β -turns (β) and γ -turns (γ) are indicated. The β -hairpins are indicated by red loops. Residues not included in the final model are indicated with a dashed line.

cant by different alignment methods. Both *FATCAT* (Ye & Godzik, 2003) and *DALI* (Holm *et al.*, 2008) show a statistically significant similarity, with the *FATCAT* alignment yielding a C^α r.m.s.d. of 3.0 Å over 51 residues (Figs. 2a and 2b) and a sequence identity of 8% (*DALI* Z score 4.0, r.m.s.d. of 3.6 Å over 48 residues, 15% sequence identity). The main difference involves the orientation of the β_3 meanders with respect to each other, resulting in different hydrogen-bonding patterns (*i.e.* the two meanders are connected in the N-terminal domain but form two separate sheets in the C-terminal domain).

Three smaller β_3 - α -helix ($\beta_3\alpha$) repeats (residues 36–67, 129–164 and 167–192) were also identified in the structure and can be generated from the first repeat by an approximate 90° anticlockwise rotation along the helix axis between the first and second motif and a clockwise 90° rotation between the second and third (Fig. 2c). Although the C-terminal helix is both shorter and differently oriented with respect to the meander when compared with the first two $\beta_3\alpha$ repeats (Fig. 2d), this unit may constitute a minimal supersecondary structure motif for this protein.

The two domains of Sbal_2486 (residues 29–84 and 85–155) show close structural similarity to SPO0140 (C^α r.m.s.d. of 2.8 Å over 127 residues with a sequence identity of 25%). Structural elements missing from Sbal_2486 involve SPO0140 helices H1 and H2 and part of strand β_1 , as well as the C-terminal β -meander (strands β_{13} – β_{15}) and helix H6 (Fig. 3a). Thus, although the $\beta_3\alpha\beta_3$ repeat described for SPO0140 is present in the N-terminal domain of Sbal_2486, the loss of the C-terminal meander results in a truncated version of this repeat in the second domain. The hypothesis of the fold having originated *via* duplication of the $\beta_3\alpha\beta_3$ repeat is supported by homolog sequence analysis, which shows the N-terminal domain to be more strongly

conserved than the C-terminal domain. Helix H1 and the β_7 – β_9 meander can then be viewed as additions to the core repeat that help to stabilize each domain. Alternatively, the fold can also be viewed as consisting of a repetition of β_3 units followed either by an α -helix or a turn. The β_3 -turn repeats (strands β_4 – β_6 and β_7 – β_9) are conserved in both structures, with the conservation of the β_7 – β_9 meander arguing in favor of this possibility. In the case of the $\beta_3\alpha$ motifs, three repeats are encountered in longer homologs, such as SPO0140, and two are found in shorter versions, such as Sbal_2486.

Searches with *FATCAT* (Ye & Godzik, 2003), *DALI* (Holm *et al.*, 2008) or *SSM* (Krissinel & Henrick, 2004) revealed no significant hits for the N-terminal domain of SPO0140 or Sbal_2486. In all three methods, the closest structural neighbor of the SPO0140 C-terminal domain is the PA2021 protein (PDB code 1ywy; Y. C. Lin, G. Liu, Y. Shen, A. Yee, C. H. Arrowsmith & T. Szyperki, unpublished work) from *Pseudomonas aeruginosa*, with a C^α r.m.s.d. of 3.4 Å over 67 residues and a sequence identity of 12% (Fig. 3b). The similarity extends over strands β_7 – β_9 and the second $\beta_3\alpha$ motif (strands β_{10} – β_{12} and helix H5). *SCOP* classifies PA2021 as a pleckstrin-homology (PH) domain-like barrel. However, both this similarity and the similarity of the same region of SPO0140 to the canonical prokaryotic PH domain (PDB code 3hsa; Joint Center for Structural Genomics, unpublished work; Fig. 3c) are not statistically significant in all of the algorithms employed (*DALI* Z score 0.4, r.m.s.d. of 4.6 Å over 42 residues, 10% sequence identity; *SSM* Z score 1.4, *P* score 0, r.m.s.d. of 3.2 Å over 53 residues, 4% sequence identity; Laskowski *et al.*, 2005) and the overall topologies of the two domains are different, with the PH domain containing an additional N-terminal strand and a longer helix in a different orientation with respect to the DUF1285 β_3 -helix meander (β_{10} – β_{12}). These considerations led us to classify

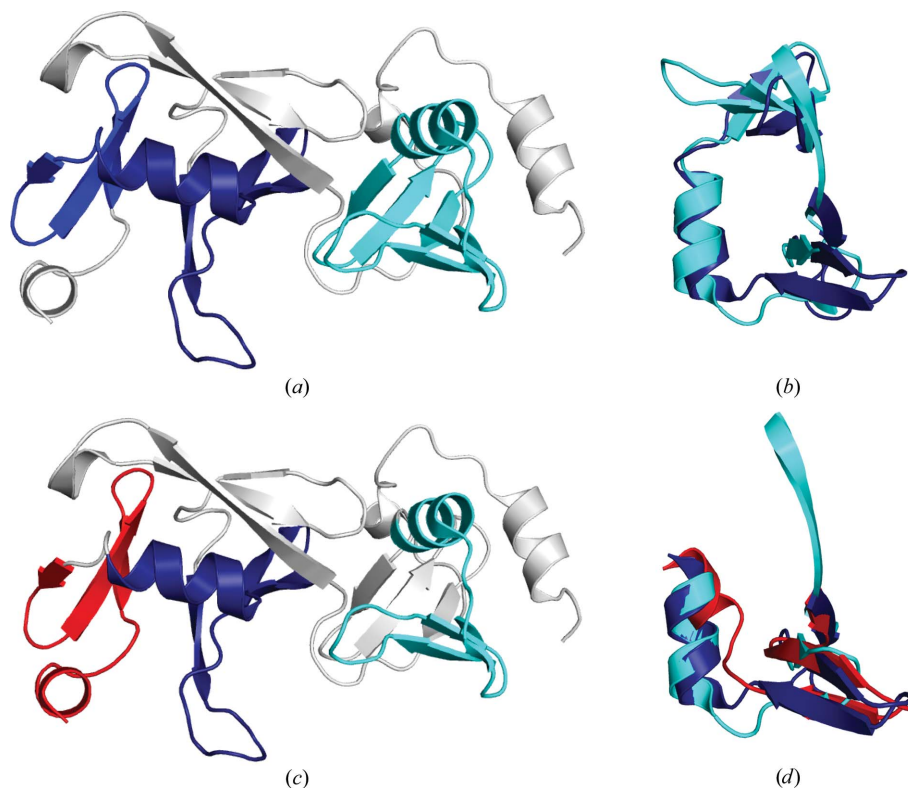


Figure 2

Structural representation of the repeated $\beta_3\alpha\beta_3$ and $\beta_3\alpha$ motifs in SPO0140. (a, b) SPO0140 contains two $\beta_3\alpha\beta_3$ motifs. (a) Ribbon diagram of SPO0140 (PDB code 2re3; residues 10–192; gray) showing the relative orientation of the two $\beta_3\alpha\beta_3$ motifs (residues 34–92, cyan; residues 126–186, blue) in the structure and (b) the same repeats superimposed. (c, d) SPO0140 contains three $\beta_3\alpha$ motifs. (c) Ribbon diagram of SPO0140 as in (a) showing the relative orientation of the three $\beta_3\alpha$ motifs (residues 36–67, cyan; residues 129–164, blue; residues 167–192, red) and (d) the three $\beta_3\alpha$ motifs superimposed.

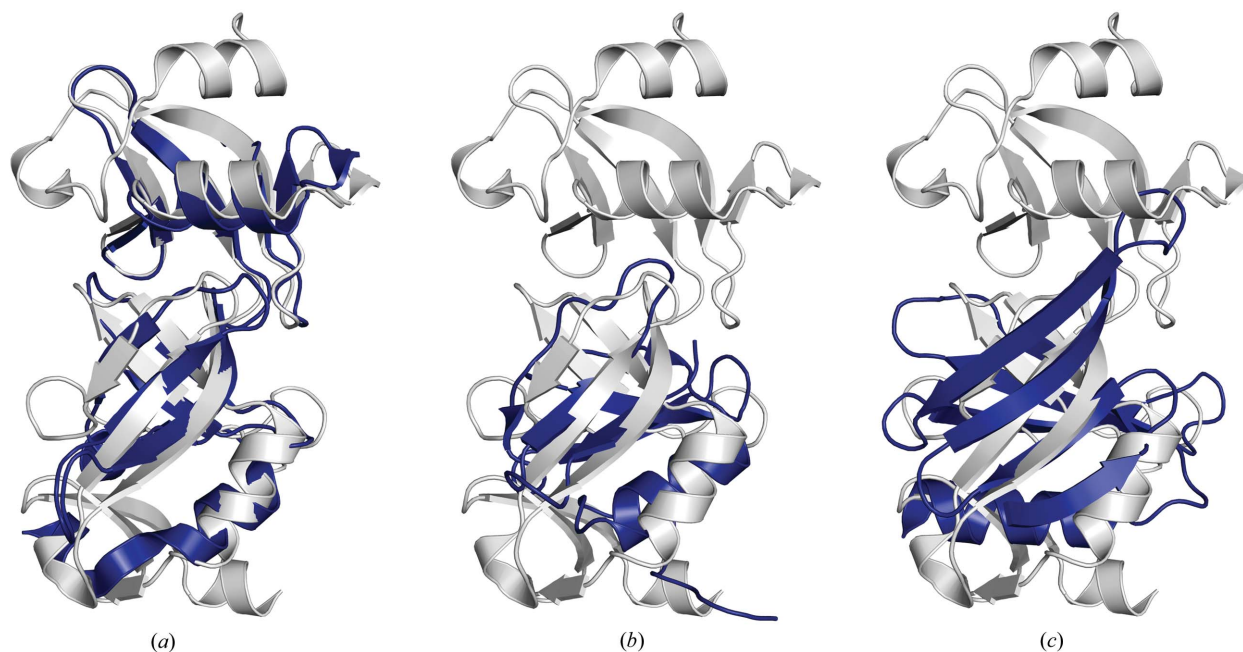


Figure 3

Structural comparisons between DUF1285 homologs and PH-like domains. Ribbon diagram showing the superposition of SPO0140 (PDB code 2re3, residues 10–192; gray) with, in blue, (a) Sbal_2486 (PDB code 2ra9, residues 29–155), another DUF1285 homolog, (b) a prokaryotic PH-like domain PA2021 (PDB code 1ywy, residues 23–96) and (c) a canonical prokaryotic PH domain (PDB code 3hsa, residues 22–179).

the SPO0140 and Sbal_2486 structures as a new fold, an assessment that is supported by a preliminary *SCOP* analysis (Alexey Murzin, personal communication).

SPO0140 crystallized with two monomers (*A* and *B*) in the asymmetric unit. The largest packing interface buries $\sim 810 \text{ \AA}^2$ of solvent-accessible surface per monomer and involves mainly helix H1, strand $\beta 1$ and the H3– $\beta 4$ loop from the N-terminal domain and helix H4 and strands $\beta 14$ – $\beta 15$ from the C-terminal domain. However, packing analysis using *PISA* (Krissinel & Henrick, 2007) suggests that this would not be a stable dimerization interface. The results of analytical size-exclusion chromatography (anSEC) support the assignment of a monomer as the quaternary state in solution for SPO0140.

Sbal_2486 crystallized with one monomer in the asymmetric unit and *PISA* again suggested a monomer as the most probable oligomeric form. The largest packing interface buries $\sim 760 \text{ \AA}^2$ of solvent-accessible surface area but is not predicted to be stable for complex formation. However, the results of anSEC coupled with static light scattering (SLS) for Sbal_2486 indicate a dimeric form in solution under the test conditions. This suggests that either the observed crystal-packing interface is stable or the oligomerization state is dependent on the buffer conditions, forming a dimer in the anSEC buffer but disassociating in the crystallization solution prior to crystal formation.

3.2. Analysis of a conserved cavity

Analysis of SPO0140 and Sbal_2486 using the *CastP* server (Binkowski *et al.*, 2003) revealed that the largest cavity occurs at the interdomain interface. Surface-conservation analysis using *ConSurf* (Landau *et al.*, 2005) showed this to be the largest and most highly conserved contiguous region in the DUF1285 family; it includes loops $\beta 1$ – $\beta 2$, $\beta 6$ – $\beta 7$ and strand $\beta 4$ from the N-terminal domain and strands $\beta 10$ – $\beta 12$ from the C-terminal domain. No strictly conserved residues

are found in this region among DUF1285 homologs, leading us to propose that this region might act as a binding site, but not one that exhibits catalytic activity.

A search against a database of cognate binding sites using *IsoCleft* (Najmanovich *et al.*, 2008), a graph-matching algorithm that searches for similarities in both geometry and chemical composition, identified shared features between the inter-domain cleft of SPO0140 and sugar, phosphate and purine-binding proteins [PDB codes 1pwh (Yamada *et al.*, 2003), 1dqa (Istvan *et al.*, 2000), 1dm3 (Modis & Wierenga, 2000), 1gpe (Wohlfahrt *et al.*, 1999), 1v0j (Beis *et al.*, 2005), 1hwy (Smith *et al.*, 2001), 2vfs (Forneris *et al.*, 2008) and 1q6p (Scapin *et al.*, 2003)]. Similar hits (adenosylcobalamin, heme, dideoxy sugars, NAD, thiamine diphosphate) were obtained for Sbal_2486 (Supplementary Table S1¹). These similarities, combined with the high conservation observed in this region of the DUF1285 structures and genome-context analysis (see below), indicate that a nucleotide-based ligand may bind along the interdomain interface.

3.3. Gene-fusion and genome-context analysis

The DUF1285 homolog from *Marinobacter* sp. ELB17 includes a phosphoglycerate mutase domain (Pfam PF00300) as part of the gene preceding DUF1285. This domain is further annotated as belonging to the subgroup of SixA phosphohistidine phosphatases (IPR004449). In prokaryotes, transcriptional profiling has shown that expression of phosphoglycerate mutase is increased under conditions of oxidative stress (Nodop *et al.*, 2008), while the SixA phosphohistidine phosphatase in *E. coli* has been implicated in signal transduction under conditions of anaerobic respiratory growth (Matsubara & Mizuno, 2000). However, as this is the only example of such a domain fusion

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5122).

encountered in the DUF1285 family, the scope for functional speculation is limited.

Several genes predicted (<http://string.embl.de>) to have functional associations with SPO0140 are located in the same genome neighborhood. Nudix hydrolases are observed in the genome neighborhood of the majority of DUF1285 homologs, including SPO0140. Nudix hydrolases are pyrophosphatases that control the cellular concentrations of a variety of nucleoside diphosphate derivatives, including nucleoside diphosphates and triphosphates and their oxidized forms, dinucleoside polyphosphates, NADH and other signaling compounds (Kraszewska, 2008). In plants, these enzymes have been implicated in oxidative signaling (Jambunathan & Mahalingam, 2006; Mahalingam *et al.*, 2006), including the maintenance of cellular redox homeostasis (Ge *et al.*, 2007) and resistance to exogenous reactive oxygen species (Tong *et al.*, 2009). In bacteria, nucleotide-based second messengers are involved in a range of signaling functions (Pesavento & Hengge, 2009), including the oxidative stress response (Johnstone & Farr, 1991). Similarly, the DUF1285 family might carry out a signaling function related to oxidative stress, possibly through binding to a small nucleotide derivative.

The SPO0140 protein family (DUF1285, PF06938) is encountered mainly in proteobacteria and contains around 200 sequence homologs which vary between 150 and 250 residues in length. Availability of more DUF1285 sequences and structures might shed light on the evolutionary history of this intriguing protein family. The information presented here, in combination with further biochemical and biophysical studies, should yield valuable insights into the functional role of SPO0140 and Sbal_2486. Models for SPO0140 and Sbal_2486 homologs can be accessed at http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=2re3A and http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=2ra9A, respectively.

Additional information about SPO0140 and Sbal_2486 is available from TOPSAN (Krishna *et al.*, 2010) at <http://www.topsan.org/explore?PDBid=2re3> and <http://www.topsan.org/explore?PDBid=2ra9>, respectively.

4. Conclusions

The first structural representatives of the DUF1285 family revealed a novel fold consisting of repeated motifs. Sequence-conservation and genome-context analysis suggests a signaling role possibly involving binding to a small nucleotide derivative under conditions of oxidative stress.

We thank N. Grishin (University of Texas Southwestern Medical Center), L. Aravind (NCBI, Bethesda), A. Murzin (MRC, Cambridge) and A. Andreeva (MRC, Cambridge) for stimulating discussions. This work was supported by the National Institute of General Medical Sciences, Protein Structure Initiative grant No. U54 GM074898. RJN is a member of the FRISQ-funded Centre de Recherche Clinique Étienne-Le Bel. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL) and Advanced Light Source (ALS). 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). The ALS is

supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division of the US Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. Genomic DNA from *Silicibacter pomeroyi* DSS-3 (ATCC 700808D) was obtained from the American Type Culture Collection (ATCC). *Shewanella baltica* OS155 was a gift 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

- Beis, K., Srikanthasan, V., Liu, H., Fullerton, S. W., Bamford, V. A., Sanders, D. A., Whitfield, C., McNeil, M. R. & Naismith, J. H. (2005). *J. Mol. Biol.* **348**, 971–982.
- Binkowski, T. A., Naghibzadeh, S. & Liang, J. (2003). *Nucleic Acids Res.* **31**, 3352–3355.
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Pacionek, 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.
- 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.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Forneris, F., Heuts, D. P., Delvecchio, M., Rovida, S., Fraaije, M. W. & Mattevi, A. (2008). *Biochemistry*, **47**, 978–985.
- Ge, X., Li, G. J., Wang, S. B., Zhu, H., Zhu, T., Wang, X. & Xia, Y. (2007). *Plant Physiol.* **145**, 204–215.
- Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
- Istvan, E. S., Palnitkar, M., Buchanan, S. K. & Deisenhofer, J. (2000). *EMBO J.* **19**, 819–830.
- Jambunathan, N. & Mahalingam, R. (2006). *Planta*, **224**, 1–11.
- Johnstone, D. B. & Farr, S. B. (1991). *EMBO J.* **10**, 3897–3904.
- Kleywegt, G. J. (2000). *Acta Cryst.* **D56**, 249–265.
- Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). *Proteins*, **71**, 982–994.
- Kraszewska, E. (2008). *Acta Biochim. Pol.* **55**, 663–671.
- 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. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). *Nucleic Acids Res.* **33**, W299–W302.
- Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
- Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). *J. Mol. Biol.* **351**, 614–626.
- 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**.
- Mahalingam, R., Jambunathan, N., Gunjan, S. K., Faustin, E., Weng, H. & Ayoubi, P. (2006). *Plant Cell Environ.* **29**, 1357–1371.
- Matsubara, M. & Mizuno, T. (2000). *FEBS Lett.* **470**, 118–124.
- 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.
- Miller, M. D. & Deacon, A. M. (2007). *Nucl. Instrum. Methods Phys. Res. A*, **582**, 233–235.
- Modis, Y. & Wierenga, R. K. (2000). *J. Mol. Biol.* **297**, 1171–1182.
- Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* **D55**, 247–255.
- Najmanovich, R., Kurbatova, N. & Thornton, J. (2008). *Bioinformatics*, **24**, i105–i111.

- Nodop, A., Pietsch, D., Hocker, R., Becker, A., Pistorius, E. K., Forchhammer, K. & Michel, K. P. (2008). *Plant Physiol.* **147**, 747–763.
- Pesavento, C. & Hengge, R. (2009). *Curr. Opin. Microbiol.* **12**, 170–176.
- 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.
- Scapin, G. *et al.* (2003). *Biochemistry*, **42**, 11451–11459.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Smith, T. J., Peterson, P. E., Schmidt, T., Fang, J. & Stanley, C. A. (2001). *J. Mol. Biol.* **307**, 707–720.
- Terwilliger, T. C. (2003). *Acta Cryst.* **D59**, 1174–1182.
- Tong, L., Lee, S. & Denu, J. M. (2009). *J. Biol. Chem.* **284**, 11256–11266.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Wohlfahrt, G., Witt, S., Hendle, J., Schomburg, D., Kalisz, H. M. & Hecht, H.-J. (1999). *Acta Cryst.* **D55**, 969–977.
- Yamada, T., Komoto, J., Takata, Y., Ogawa, H., Pitot, H. C. & Takusagawa, F. (2003). *Biochemistry*, **42**, 12854–12865.
- Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst.* **D60**, 1833–1839.
- Ye, Y. & Godzik, A. (2003). *Bioinformatics*, **19**, Suppl. 2, ii246–ii255.