

Debanu Das,<sup>a,b</sup> Piotr Kozbial,<sup>a,c</sup>  
Gye Won Han,<sup>a,d</sup> Dennis Carlton,<sup>a,d</sup>  
Lukasz Jaroszewski,<sup>a,c,e</sup> Polat  
Abdubek,<sup>a,f</sup> Tamara Astakhova,<sup>a,e</sup>  
Herbert L. Axelrod,<sup>a,b</sup> Constantina  
Bakolitsa,<sup>a,c</sup> Connie Chen,<sup>a,f</sup> Hsiu-Ju  
Chiu,<sup>b</sup> Michelle Chiu,<sup>a,f</sup> Thomas  
Clayton,<sup>a,d</sup> Marc C. Deller,<sup>a,d</sup> Lian  
Duan,<sup>a,e</sup> Kyle Ellrott,<sup>a,e</sup> Marc-André  
Elslinger,<sup>a,d</sup> Dustin Ernst,<sup>a,f</sup> Carol L.  
Farr,<sup>a,d</sup> Julie Feuerhelm,<sup>a,f</sup> Anna  
Grzechnik,<sup>a,d</sup> Joanna C. Grant,<sup>a,f</sup>  
Kevin K. Jin,<sup>a,b</sup> Hope A. Johnson,<sup>a,d</sup>  
Heath E. Klock,<sup>a,f</sup> Mark W. Knuth,<sup>a,f</sup>  
S. Sri Krishna,<sup>a,c,e</sup> Abhinav Kumar,<sup>a,b</sup>  
David Marciano,<sup>a,d</sup> Daniel  
McMullan,<sup>a,f</sup> Mitchell D. Miller,<sup>a,b</sup>  
Andrew T. Morse,<sup>a,e</sup> Edward  
Nigoghossian,<sup>a,f</sup> Amanda Nopakun,<sup>a,d</sup>  
Linda Okach,<sup>a,f</sup> Silvy Oommachen,<sup>a,b</sup>  
Jessica Paulsen,<sup>a,f</sup> Christina Puckett,<sup>a,f</sup>  
Ron Reyes,<sup>a,b</sup> Christopher L. Rife,<sup>a,b</sup>  
Natasha Sefcovic,<sup>a,c</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,c</sup> Tiffany  
Wooten,<sup>a,f</sup> Qingping Xu,<sup>a,b</sup> Keith O.  
Hodgson,<sup>a,g</sup> John Wooley,<sup>a,e</sup> Ashley M.  
Deacon,<sup>a,b</sup> Adam Godzik,<sup>a,c,e</sup> Scott A.  
Lesley,<sup>a,d,f</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,  
California, USA, <sup>c</sup>Program on Bioinformatics  
and Systems Biology, Burnham Institute for  
Medical Research, La Jolla, California, USA,  
<sup>d</sup>Department of Molecular Biology, The Scripps  
Research Institute, La Jolla, California, USA,  
<sup>e</sup>Center for Research in Biological Systems,  
University of California, San Diego, La Jolla,  
California, USA, <sup>f</sup>Protein Sciences Department,  
Genomics Institute of the Novartis Research  
Foundation, San Diego, California, USA, and  
<sup>g</sup>Photon Science, SLAC National Accelerator  
Laboratory, Menlo Park, California, USA

Correspondence e-mail: wilson@scripps.edu

Received 16 April 2009  
Accepted 13 May 2009

PDB Reference: KPN03535, 3flz, r3f1zsf.

## The structure of KPN03535 (gi|152972051), a novel putative lipoprotein from *Klebsiella pneumoniae*, reveals an OB-fold

KPN03535 (gi|152972051) is a putative lipoprotein of unknown function that is secreted by *Klebsiella pneumoniae* MGH 78578. The crystal structure reveals that despite a lack of any detectable sequence similarity to known structures, it is a novel variant of the OB-fold and structurally similar to the bacterial Cpx-pathway protein NlpE, single-stranded DNA-binding (SSB) proteins and toxins. *K. pneumoniae* MGH 78578 forms part of the normal human skin, mouth and gut flora and is an opportunistic pathogen that is linked to about 8% of all hospital-acquired infections in the USA. This structure provides the foundation for further investigations into this divergent member of the OB-fold family.

### 1. Introduction

KPN03535 (gi|152972051) is an orphan protein that is exclusively found in *Klebsiella pneumoniae* MGH 78578 (an opportunistic human pathogen belonging to enterbacteriales of gammaproteobacteria; Galperin *et al.*, 2007; Gill *et al.*, 2006; Frank & Pace, 2008; Ley *et al.*, 2008) and *K. pneumoniae* 342 (three-residue substitution). It consists of 132 residues with a calculated pI of 9.40 and a predicted signal peptide. The N-terminus of KPN03535 has a lipoprotein signature, indicated by the presence of an LSGC motif (von Heijne, 1989), as well as predictions from *LipoP* 1.0 (Juncker *et al.*, 2003). It is a singleton protein that has not been assigned to any Pfam family, but sequence-based fold-prediction methods (Ginalski *et al.*, 2003) suggest similarity to members of the PF01336 family (OB-fold nucleic acid-binding domain). We determined the crystal structure of KPN03535 in order to explore this extremely divergent member of the commonly occurring OB-fold. Structural comparisons show similarities to the OB-fold-containing Cpx-pathway protein NlpE, single-stranded DNA-binding (SSB) proteins, bacterial OB-fold (BOF) and toxin proteins, which enable inferences about function that may now be tested biochemically. This structure should serve as a basis for understanding structure–function relationships in any newly discovered proteins with a similar sequence, such as those identified by whole microbial genome sequencing and metagenomic surveys of the human microbiome.

### 2. Materials and methods

#### 2.1. KPN03535 expression, purification and crystallization

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE; Klock *et al.*, 2008) cloning method. The gene encoding KPN03535 (gi|152972051; Swiss-Prot A6TEE6) was amplified by polymerase chain reaction (PCR) from *K. pneumoniae* MGH 78578 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5'-ctgtacttcaggcGCTTCTAAAGCCTTTATTCCGCGGGAG-3'; reverse primer, 5'-aattaagtcgcttaTTTAACCACCTTGGGATTCTGTAGC-GTC-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 (MGSDKIHSHHHHEN-

LYFQG), was PCR-amplified with V-PIPE (Vector) primers (forward primer, 5'-taacgcgacttaattaactgtttaaaccggtctccagc-3'; reverse primer, 5'-gccctggaagtacaggtttctgtgatgatgatgatg-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 on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Using the PIPE method, the gene segment encoding residues Met1–Leu22 was deleted for expression of soluble protein as these residues were initially predicted to correspond to either a signal peptide using *SignalP* (Bendtsen *et al.*, 2004) or transmembrane helices using *TMHMM-2.0* (Krogh *et al.*, 2001). Expression was performed in selenomethionine-containing medium. 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 passed 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

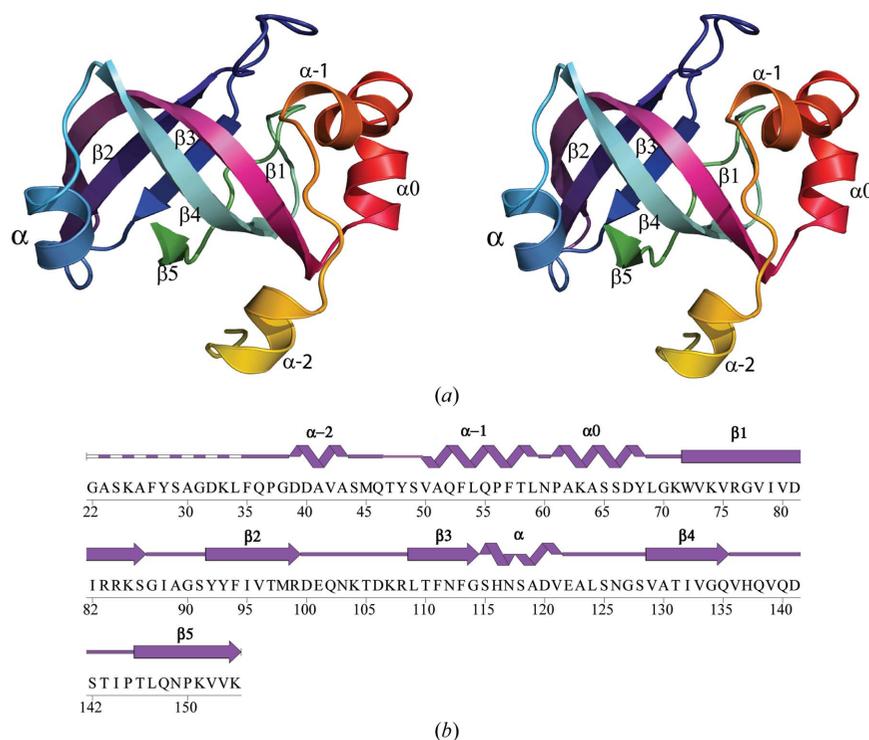
**Table 1**

Crystallographic data and refinement statistics for KPN03535 (PDB code 3flz).

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 97.42, b = 105.51, c = 181.25$
Data collection	
Wavelength (Å)	0.9792 [Se peak ( $\lambda_1$ )]
Resolution range (Å)	29.9–2.46 (2.52–2.46)
No. of observations	509996
No. of unique reflections	68362
Completeness (%)	99.8 (99.7)
Mean $I/\sigma(I)$	15.4 (2.5)
$R_{\text{merge}}$ on $I^\dagger$ (%)	11.1 (69.6)
Model and refinement statistics	
Resolution range (Å)	29.9–2.46
No. of reflections (total)	68310 $\ddagger$
No. of reflections (test)	3458
Completeness (%)	99.7
Data set used in refinement	$\lambda_1$
Cutoff criterion	$ F  > 0$
$R_{\text{cryst}}^\S$	0.192
$R_{\text{free}}^\S$	0.228
Stereochemical parameters	
Restraints (r.m.s.d. observed)	
Bond angle ( $^\circ$ )	1.70
Bond length (Å)	0.017
Average isotropic $B$ value ( $\text{Å}^2$ )	38.2 $\P$
ESU $\ddagger\ddagger$ based on $R_{\text{free}}$ (Å)	0.22
Protein residues/atoms	1182/9162
Water/PEG molecules	323/2

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ .  $\ddagger$  Typically, the number of unique reflections used in refinement is 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.  $\S R_{\text{cryst}} = \sum_{hkl} | |F_{\text{obs}}| - |F_{\text{calc}}| | / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure-factor amplitudes, respectively.  $R_{\text{free}}$  is as for  $R_{\text{cryst}}$  but for 5.1% of the total reflections chosen at random and omitted from refinement.  $\P$  This value represents the total  $B$  that includes TLS and residual  $B$  components.  $\ddagger\ddagger$  ESU, estimated overall coordinate error (Collaborative Computational Project, Number 4, 1994; Tickle *et al.*, 1998).

**Figure 1**

Crystal structure of KPN03535. (a) Stereo ribbon representation of the KPN03535 monomer color-coded from the N-terminus (yellow) to the C-terminus (green). The nomenclature for helix  $\alpha$  and strands  $\beta 1$ – $\beta 5$  follows that used for the OB-fold (Murzin, 1993). Helices  $\alpha-2$ ,  $\alpha-1$  and  $\alpha 0$  are unique to KPN03535. (b) Diagram showing the secondary-structure elements of KPN03535 superimposed on the primary amino-acid sequence. Helices and  $\beta$ -strands are indicated. The protein was expressed with a purification tag that was removed, leaving a residual Gly residue at the N-terminus followed by the KPN03535 sequence.

the same buffer. The flowthrough and wash fractions were combined and concentrated for crystallization trials to 16.1 mg ml<sup>-1</sup> by centrifugal ultrafiltration (Millipore). KPN03535 was crystallized by mixing 100 nl protein solution with 100 nl crystallization solution in a sitting drop over a 50 µl reservoir volume 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 contained 31% polyethylene glycol 600 and 0.1 M CHES pH 9.5. No further cryoprotectant was added to the crystal. A cube-shaped crystal with approximate dimensions 80 × 80 × 80 µm was harvested after 42 d at 293 K for data collection. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM; Cohen *et al.*, 2002) at the Stanford Synchrotron Radiation Lightsource (SSRL; Menlo Park, California, USA). The diffraction data were indexed in the orthorhombic space group *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub>. The molecular weight and oligomeric state were determined using a 1 × 30 cm Superdex 200 column (GE Healthcare) in combination with static light scattering (Wyatt Technology). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM NaCl and 0.02% (*w/v*) sodium azide.

2.2. X-ray data collection and structure determination

Single-wavelength anomalous diffraction (SAD) data were collected to 2.46 Å resolution on beamline 9-2 at SSRL at the

wavelength corresponding to the peak ( $\lambda_1$ ) of a selenium absorption edge using the *Blu-Ice* data-collection environment (McPhillips *et al.*, 2002). A data set was collected at 100 K using a MAR Mosaic 325 CCD detector (Rayonix USA). The SAD data were integrated and reduced using *MOSFLM* (Leslie, 1992) and scaled with the program *SCALA* (Collaborative Computational Project, Number 4, 1994). Phasing was performed with *SHELXD* (Sheldrick, 2008) and *auto-SHARP* (Vonrhein *et al.*, 2007) [20 selenium sites per asymmetric unit, overall FOM (acentric/centric) = 0.34/0.12, overall phasing power (anomalous differences) = 1.2] and automated iterative model building was performed with *RESOLVE* (Terwilliger, 2003). Model completion and crystallographic refinement were performed with *Coot* (Emsley & Cowtan, 2004) and *REFMAC5* (Collaborative Computational Project, Number 4, 1994) with TLS (one group per monomer) refinement (Winn *et al.*, 2003) and medium NCS restraints for all chains. Data and refinement statistics are summarized in Table 1.

The quality of the crystal structure was analyzed using the JCSG Quality Control server, which verifies the stereochemical quality of the model using *AutoDepInputTool* (Yang *et al.*, 2004), *MolProbity* (Davis *et al.*, 2004) and *WHATIF* 5.0 (Vriend, 1990), the agreement between the atomic model and the data using *SFCHECK* 4.0 (Vaguine *et al.*, 1999) and *RESOLVE* (Terwilliger, 2000), the protein sequence using *ClustalW* (Thompson *et al.*, 1994), atom occupancies using *MOLEMAN2* (Kleywegt, 2000) and the consistency of NCS

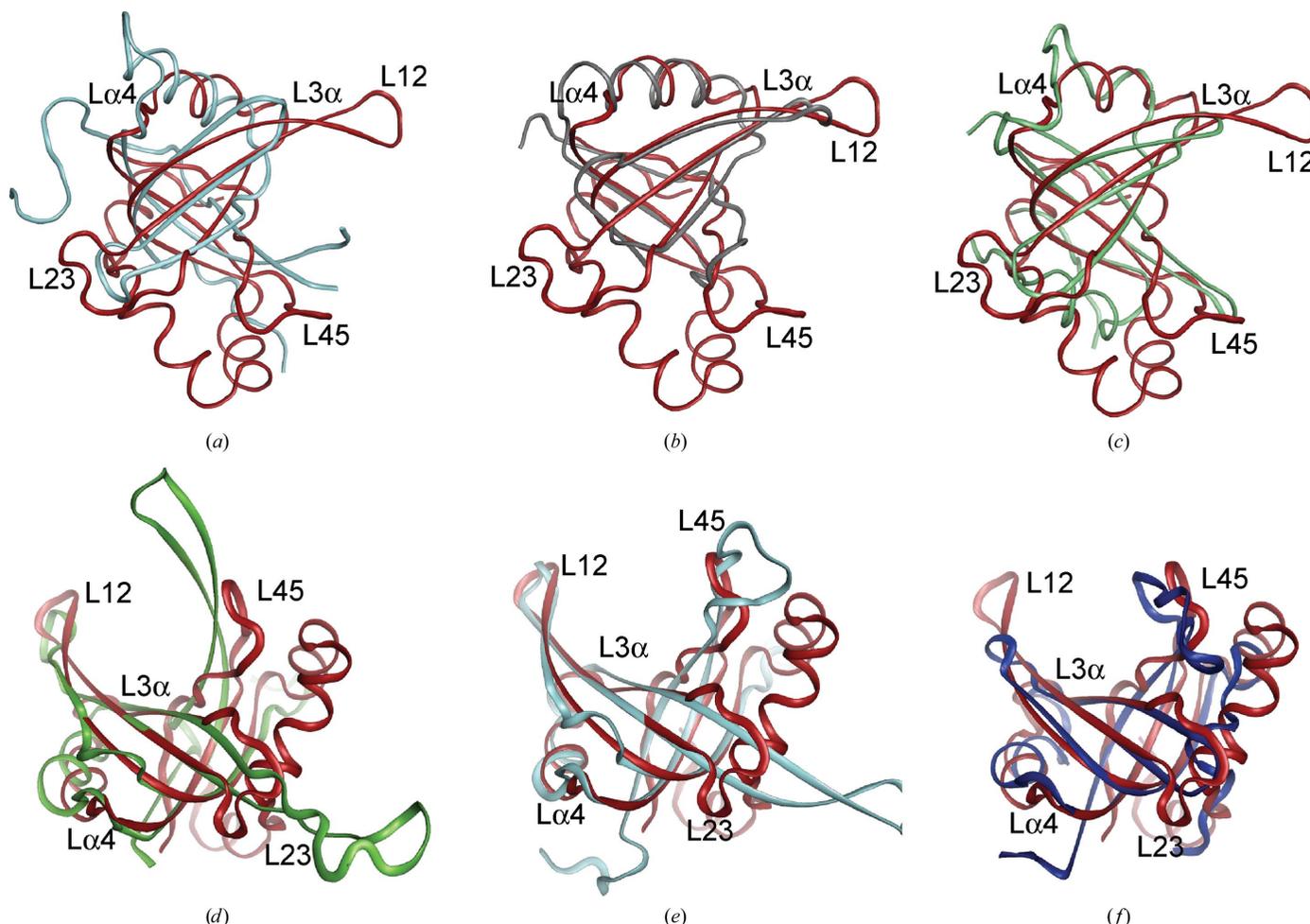


Figure 2 Superimposition of the crystal structure of KPN03535 (red) on OB-fold proteins that have N-terminal lipoprotein sequence, such as (a) NlpE, (b) shiga toxin and (c) BOF, and single-stranded DNA-binding proteins (SSBs), such as (d) *E. coli* SSB, (e) *E. coli* PriB and (f) *T. thermophilus* aspartyl-tRNA synthetase.

pairs. This analysis also evaluates difference in  $R_{\text{cryst}}/R_{\text{free}}$ , expected  $R_{\text{free}}/R_{\text{cryst}}$  and maximum/minimum  $B$  values by parsing the refinement log-file and PDB header. Protein quaternary structure analysis was performed using the PISA server (Krissinel & Henrick, 2005). Fig. 1(b) was adapted from an analysis using PDBsum (Laskowski *et al.*, 2005) and all other figures were prepared with PyMOL (DeLano, 2002). Atomic coordinates and experimental structure factors for KPN03535 have been deposited in the PDB under accession code 3flz. A systematic search for other proteins of similar structure was conducted using several different methods including the DALI server (Holm *et al.*, 2008), the protein structure comparison service SSM at the European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm>; Krissinel & Henrick, 2004) and the flexible structure-alignment method FATCAT (Ye & Godzik, 2003).

### 3. Results and discussion

#### 3.1. Overall structure

Residues 1–22 of the full-length protein (1–154) were initially predicted to represent a signal peptide and were removed during cloning. The crystallized protein is comprised of a glycine left after cleavage of the expression and purification tag followed by KPN03535 residues 23–154. The final model contains ten monomers (chains A–J), two PEG molecules (PEG 600 fragments from the crystallization solution) and 323 water molecules in the asymmetric unit. The ten monomers are almost identical in structure and completeness and superimpose extremely well, with pairwise r.m.s.d. values ranging from only 0.2 to 0.4 Å. Residues 23–35 in chains A, B, C, E and J, 23–36 and 154 in chain D, 23–35 and 154 in chain F, 23–36 in chains G and H, and 23–38 in chain I and the N-terminal glycine in all chains are disordered and have not been modeled. The Matthews

coefficient (Matthews, 1968) is  $\sim 3.2 \text{ \AA}^3 \text{ Da}^{-1}$ , with an estimated solvent content of  $\sim 62\%$ . The Ramachandran plot produced by MolProbity (Davis *et al.*, 2004) shows that 98.5% and 100% of amino acids are in the favored and allowed regions, respectively.

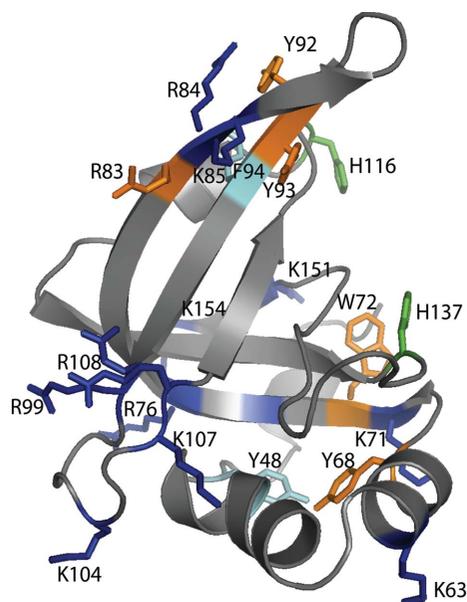
Residues 70–154 of the monomer form the OB-fold comprised of a five-stranded  $\beta$ -sheet ( $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$  and  $\beta_5$ ) capped by a short  $\alpha$ -helix ( $\alpha$ ) based on the standard OB-fold nomenclature (Murzin, 1993; Fig. 1a). The capping helix is shorter than those observed in most other OB-fold proteins (Fig. 2). Residues 36–69 constitute three additional  $\alpha$ -helices ( $\alpha-2$ ,  $\alpha-1$  and  $\alpha_0$ ) which are not observed in other structures of the same fold. The curved  $\beta$ -sheet forming the  $\beta$ -barrel core of the OB-fold is highly conserved in size and structure, while the largest variations are seen in the three loops ( $L_{12}$ ,  $L_{23}$  and  $L_{45}$ ) that extend in different directions from the core and are often functionally important.

Crystal-packing and assembly analysis using PISA (Krissinel & Henrick, 2005) supported by analytical size-exclusion chromatography and static light scattering suggest that a monomer is the likely oligomeric state. In the crystal structure, the protein assembles as two stacked pentameric rings, formed by loose interdigitation of the 'finger-like'  $\beta_1$ - $L_{12}$ - $\beta_2$  structure, with outer and inner diameters of  $\sim 80$  Å and  $\sim 40$  Å, respectively, and a thickness of  $\sim 40$  Å. The buried surface area of each monomer within each pentamer ( $\sim 540 \text{ \AA}^2$ ) and each monomer in the interface between the two pentamers ( $\sim 600 \text{ \AA}^2$ ) is low. The quaternary structure analysis does not suggest sufficiently strong and extensive interactions to enable complex formation in solution, suggesting that these pentamers could be a crystallization artifact. The N-terminus of each monomer extends into the solvent and probably does not have an impact on the oligomerization state. In the absence of any biochemical data, the functional oligomeric state of the protein remains unknown.

#### 3.2. Functional hypotheses

**3.2.1. NlpE-like.** The only other reported bacterial lipoprotein containing an OB-fold is the C-terminal domain of *E. coli* NlpE (new lipoprotein E), which is an outer membrane lipoprotein in Gram-negative bacteria involved in the envelope stress response in the Cpx pathway. It activates the Cpx, two-component, signal transduction pathway composed of the inner membrane histidine kinase CpxA and the cytoplasmic response regulator CpxR (Raivio & Silhavy, 1997). The Cpx pathway controls the production of the periplasmic protease DegP and other proteins involved in fighting cellular stress (Snyder *et al.*, 1995; Danese *et al.*, 1995; Raivio *et al.*, 1999). Other proteins are also implicated in the regulation of the Cpx pathway. For example, CpxP with an LTXXQ motif (Pfam PF07813) is involved in feedback inhibition of the Cpx pathway (Danese *et al.*, 1998; Danese & Silhavy, 1998). In *K. pneumoniae*, a periplasmic CpxP-like protein with the LTXXQ motif, KPN03534, is the neighboring gene to KPN03535. Therefore, KPN03535, like KPN03534, may play a role in the Cpx pathway, similar to NlpE. KPN03535 superimposes fairly well on *E. coli* NlpE (PDB code 2z4i; Hirano *et al.*, 2007; r.m.s.d. = 3.3 Å, 16% sequence identity, Z score 2.3; Fig. 2a). Despite extremely low sequence identity, some residues are conserved in KPN03535 (Arg76, Asp100, Thr105, Lys107, Arg108 and Asn117) from structure-based sequence alignment. However, the functional roles of these residues in NlpE are not known.

**3.2.2. Toxin and BOF-like.** Other bacterial OB-fold proteins that have an N-terminal signal sequence are toxins, such as the shiga toxin (PDB code 1r4p; Fraser *et al.*, 2004; Fig. 2b), cholera toxin (PDB code 3efx; Holmner *et al.*, 2004; r.m.s.d. = 2.8 Å, 6% sequence identity, Z score 5.8) and a bacterial OB-fold (BOF; Ginalski *et al.*, 2004) protein



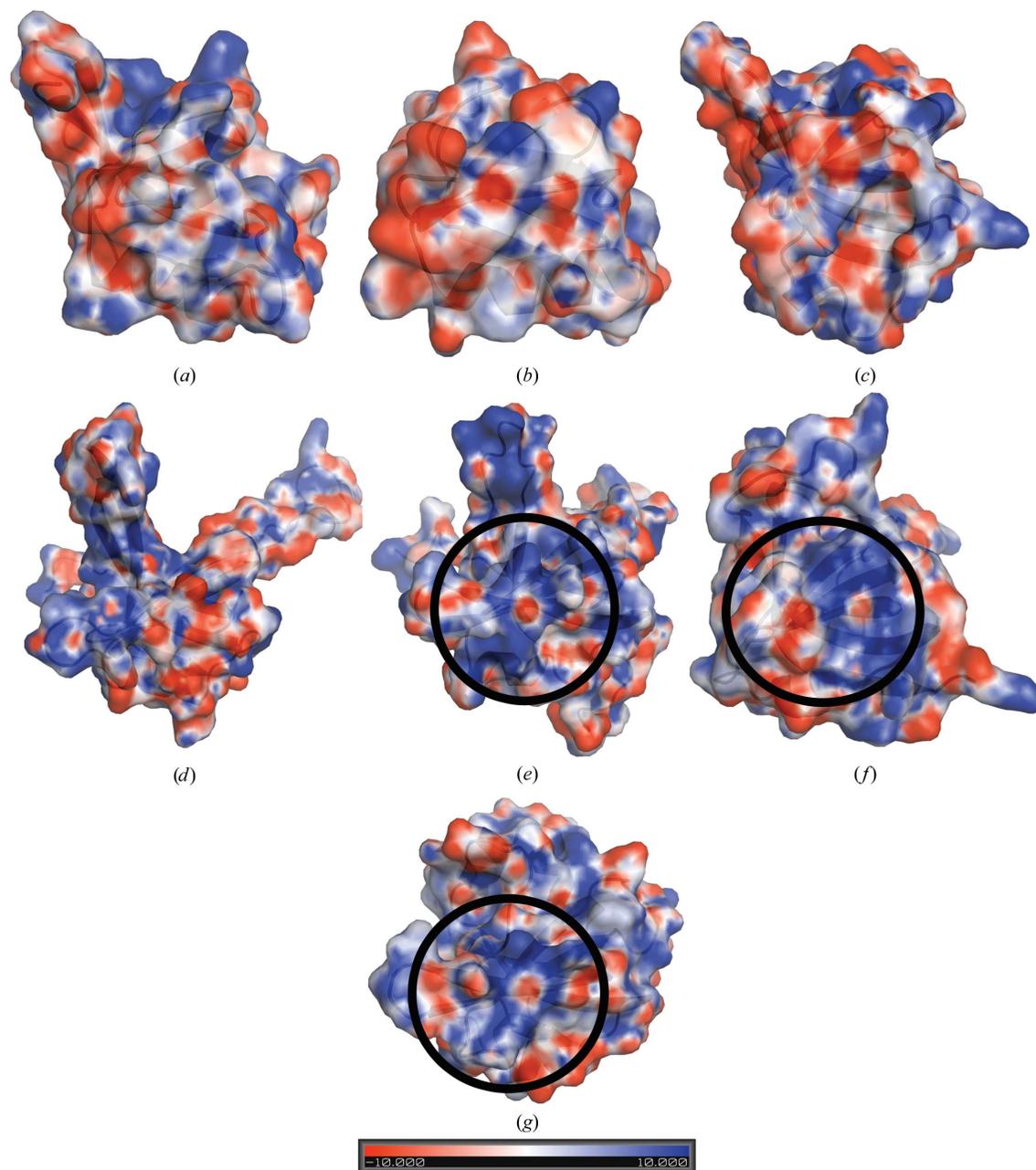
**Figure 3**

Surface-exposed charged and aromatic residues on KPN03535 that may be functionally important if this protein binds DNA or RNA (for clarity, the view of the monomer shown here is different from that shown in Fig. 4 and was obtained by a  $180^\circ$  rotation around a horizontal axis followed by a  $180^\circ$  rotation around a vertical axis). Arg83, Arg84 and Lys85 comprise the positive surface region described in Fig. 4. Arg84 and Lys85 are conserved as Arg17 and Lys18 in the *E. coli* PriB structure and as Arg29 in *T. thermophilus* aspartyl-tRNA synthetase. Phe94 is conserved in shiga toxin, but there are currently no reports of any functional role of this residue in the toxin.

(1nnx; O. Lehmann, A. Galkin, S. Pullalarevu, E. Sarikaya, W. Krajewski, K. Lim, A. Howard & O. Herzberg, unpublished work; r.m.s.d. = 2.9 Å, 14% sequence identity; Z score 7.4; Fig. 2c). Neither NlpE nor the toxins have all three of the N-terminal helices ( $\alpha 0$ ,  $\alpha-1$ ,  $\alpha-2$ ) found in KPN03535, but  $\alpha-2$  is observed in cholera toxin (3efx) and  $\alpha-1$  is observed in BOF protein (1nnx). The capping helix  $\alpha$  in KPN03535 is shorter than in the toxins and NlpE, although it is similar to that observed in BOF protein. The  $\beta$ -strands forming the curved  $\beta$ -barrel in all these structures are of similar length, but with differences in the loop sizes that connect the  $\beta$ -strands.

**3.2.3. Single-stranded DNA-binding protein, SSB-like.** Single-stranded DNA-binding proteins (SSBs) also possess OB-folds and

are involved in a multitude of cellular functions, such as DNA replication, transcription, recombination, repair, translation, cold-shock response and maintenance of telomeres (Theobald *et al.*, 2003; Chase & Williams, 1986; Wold, 1997; Meyer & Laine, 1990; Lohman & Ferrari, 1994; Lohman *et al.*, 1996). KPN03535 is structurally similar to OB-fold SSBs, including *E. coli* SSB (PDB code 1eyg; Raghunathan *et al.*, 2000; r.m.s.d. 2.7 Å; 13% sequence identity; Z score 7.0; Fig. 2d), *E. coli* PriB (PDB code 1v1q; Liu *et al.*, 2004; r.m.s.d. 2.3 Å; 13% sequence identity, Z score 8.0; Fig. 2e), *Thermus thermophilus* aspartyl-tRNA synthetase (PDB code 110w; Ng *et al.*, 2002; r.m.s.d. 2.6 Å; 11% sequence identity; Z score 9.0; Fig. 2f) and human mitochondrial SSB (PDB code 3ull; Yang *et al.*, 1997; r.m.s.d.



**Figure 4**

Comparison of the electrostatic surface potentials of monomers of (a) NlpE, (b) shiga toxin, (c) BOF, (d) *E. coli* SSB, (e) *E. coli* PriB, (f) *T. thermophilus* aspartyl-tRNA synthetase and (g) KPN03535. All the figures are in approximately the same orientation and reflect the surface view that would be presented for oligonucleotide binding, as in tRNA synthetase. The figure reveals that the positively charged surface patch (central blue portion in black circles) on the KPN03535 most closely resembles that of *E. coli* PriB and is also similar to that seen in aspartyl-tRNA synthetase. In KPN03535, this positively charged patch is formed by Arg83, Arg84 and Lys85. The corresponding conserved residues are Arg17 and Lys18 in PriB and Arg29 in aspartyl-tRNA synthetases, respectively.

2.7 Å; 8% sequence identity, Z score 7.1). The N-terminal  $\alpha$ -1 and  $\alpha$ 0 secondary-structure elements in KPN03535 are partially conserved in aspartyl-tRNA synthetase, but not in the other structures. Many of the loops in OB-fold ssDNA-binding proteins are either involved in interactions with DNA or in quaternary interactions that result in the various oligomeric forms. For example, loop L<sub>45</sub>, which makes the most interactions with DNA in PriB (Huang *et al.*, 2006) and aspartyl-tRNA synthetase, is similar to that of KPN03535, but is much longer in *E. coli* and in human mitochondrial SSBs. Among the surface-exposed Arg, Lys and aromatic residues that could be functionally relevant if KPN03535 were to bind DNA or RNA (Fig. 3), Arg84 and Lys85 of KPN03535 are conserved and correspond to Arg17 and Lys18 in PriB, where Lys18 is involved in ssDNA-binding (Huang *et al.*, 2006). Arg83 and Arg99 of KPN03535 are conserved in aspartyl tRNA synthetase as Arg29 (equivalent to Arg28 in the *E. coli* aspartyl-tRNA synthetase that binds to tRNA; Eiler *et al.*, 1999) and Arg39. Multiple structural alignment of various OB-fold proteins using the POSA method (Ye & Godzik, 2005) suggests that KPN03535 has a closer relationship to tRNA synthetases than to the BOF protein and is most distant from OB-fold toxins.

Analysis of the electrostatic surface potential indicates that KPN03535 most closely resembles PriB and aspartyl-tRNA synthetase (Fig. 4), with a prominent positively charged area similar to the DNA-binding region of these two proteins. Interestingly, this patch is different from that observed in the *E. coli* SSB, which reflects the known differences in ssDNA-binding modes of SSB and PriB. The basic nature of KPN03535 (pI 9.4) also hints at the possibility of oligonucleotide binding.

In conclusion, the crystal structure of KPN03535 reveals a novel divergent member of the prevalent OB-fold and suggests that it is most likely to be a nucleic acid-binding protein. As for the recently solved structure of MPN554 from *Mycoplasma pneumoniae* (Das *et al.*, 2007), another novel OB-fold with unknown cellular function but with single-stranded DNA-binding properties, the structure of KPN03535 reveals that further exploration of the functionality of the OB-fold is necessary. Bacterial lipoproteins have many important functions and are potential vaccine candidates (Steere *et al.*, 1998; Myers *et al.*, 2007). *K. pneumoniae* is an opportunistic pathogen that is prevalent in immunocompromised patients in hospitals and in patients with liver disease (Hidron *et al.*, 2008; Pope *et al.*, 2008). Functional inferences that can be drawn from this crystal structure should now allow focused structure-assisted biochemistry to establish the exact molecular and cellular role for this protein.

Additional information about KPN03535 is available from TOPSAN (Krishna *et al.*, 2010) <http://www.topsan.org/explore?PDBid=3flz>.

This work was supported by the National Institute of General Medical Sciences, Protein Structure Initiative Grant U54 GM074898. Portions of this research were performed at the Stanford Synchrotron Radiation Lightsource (SSRL). SSRL is a national user facility at SLAC National Accelerator Laboratory, operated by Stanford University on behalf of the United States 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 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. Genomic DNA from *K. pneumoniae* MGH 78578 (ATCC No. 700721D) was obtained from the American Type Culture Collection (ATCC).

## References

- Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). *J. Mol. Biol.* **340**, 783–795.
- Chase, J. W. & Williams, K. R. (1986). *Annu. Rev. Biochem.* **55**, 103–136.
- Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Danese, P. N., Oliver, G. R., Barr, K., Bowman, G. D., Rick, P. D. & Silhavy, T. J. (1998). *J. Bacteriol.* **180**, 5875–5884.
- Danese, P. N. & Silhavy, T. J. (1998). *J. Bacteriol.* **180**, 831–839.
- Danese, P. N., Snyder, W. B., Cosma, C. L., Davis, L. J. & Silhavy, T. J. (1995). *Genes Dev.* **9**, 387–398.
- Das, D., Hyun, H., Lou, Y., Yokota, H., Kim, R. & Kim, S.-H. (2007). *Proteins*, **67**, 776–782.
- Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). *Nucleic Acids Res.* **32**, W615–W619.
- DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific LLC, Palo Alto, California, USA.
- Eiler, S., Dock-Breggeon, A., Moulinier, L., Thierry, J. C. & Moras, D. (1999). *EMBO J.* **18**, 6532–6541.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Frank, D. N. & Pace, N. R. (2008). *Curr. Opin. Gastroenterol.* **24**, 4–10.
- Fraser, M. E., Fujinaga, M., Cherney, M. M., Melton-Celsa, A. R., Twiddy, E. M., O'Brien, A. D. & James, M. N. (2004). *J. Biol. Chem.* **279**, 27511–27517.
- Galperin, M. Y. (2007). *Environ. Microbiol.* **9**, 2385–2391.
- Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. & Nelson, K. E. (2006). *Science* **312**, 1355–1359.
- Ginalski, K., Elofsson, A., Fischer, D. & Rychlewski, L. (2003). *Bioinformatics*, **19**, 1015–1018.
- Ginalski, K., Kinch, L., Rychlewski, L. & Grishin, N. V. (2004). *FEBS Lett.* **567**, 297–301.
- Heijne, G. von (1989). *Protein Eng.* **2**, 531–534.
- Hidron, A. I., Edwards, J. R., Patel, J., Horan, T. C., Sievert, D. M., Pollock, D. A. & Fridkin, S. K. (2008). *Infect. Control Hosp. Epidemiol.* **29**, 996–1011.
- Hirano, Y., Hossain, M. M., Takeda, K., Tokuda, H. & Miki, K. (2007). *Structure*, **15**, 963–976.
- Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
- Holmmer, A., Lebens, M., Teneberg, S., Angstrom, J., Okvist, M. & Krenzel, U. (2004). *Structure*, **12**, 1655–1667.
- Huang, C. Y., Hsu, C. H., Sun, Y. J., Wu, H. N. & Hsiao, C. D. (2006). *Nucleic Acids Res.* **34**, 3878–3886.
- Juncker, A. S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2003). *Protein Sci.* **12**, 1652–1662.
- 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.
- 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. (2005). *CompLife 2005*, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 163–174. Berlin/Heidelberg: Springer-Verlag.
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). *J. Mol. Biol.* **305**, 567–580.
- Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). *Nucleic Acids Res.* **33**, D266–D268.
- Lesley, S. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11664–11669.
- Leslie, A. G. W. (1992). *Int CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. (2008). *Nature Rev. Microbiol.* **6**, 776–788.
- Liu, J. H., Chang, T. W., Huang, C. Y., Chen, S. U., Wu, H. N., Chang, M. C. & Hsiao, C. D. (2004). *J. Biol. Chem.* **279**, 50465–50471.
- Lohman, T. M. & Ferrari, M. E. (1994). *Annu. Rev. Biochem.* **63**, 527–570.
- Lohman, T. M., Overman, L. B., Ferrari, M. E. & Kozlov, A. G. (1996). *Biochemistry*, **35**, 5272–5279.

- 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.
- Meyer, R. R. & Laine, P. S. (1990). *Microbiol. Rev.* **54**, 342–380.
- Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
- Myers, G. S. *et al.* (2007). *Nature Biotechnol.* **25**, 569–575.
- Ng, J. D., Sauter, C., Lorber, B., Kirkland, N., Arnez, J. & Giege, R. (2002). *Acta Cryst. D* **58**, 645–652.
- Pope, J. V., Teich, D. L., Clardy, P. & McGillicuddy, D. C. (2008). *J. Emerg. Med.* doi:10.1016/j.jemermed.2008.04.041.
- Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). *Nature Struct. Biol.* **7**, 648–652.
- Raivio, T. L., Popkin, D. L. & Silhavy, T. J. (1999). *J. Bacteriol.* **181**, 5263–5272.
- Raivio, T. L. & Silhavy, T. J. (1997). *J. Bacteriol.* **179**, 7724–7733.
- 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.
- Sheldrick, G. M. (2008). *Acta Cryst. A* **64**, 112–122.
- Snyder, W. B., Davis, L. J., Danese, P. N., Cosma, C. L. & Silhavy, T. J. (1995). *J. Bacteriol.* **177**, 4216–4223.
- Steere, A. C., Sikand, V. K., Meurice, F., Parenti, D. L., Fikrig, E., Schoen, R. T., Nowakowski, J., Schmid, C. H., Laukamp, S., Buscarino, C. & Krause, D. S. (1998). *N. Engl. J. Med.* **339**, 209–215.
- Terwilliger, T. C. (2000). *Acta Cryst. D* **56**, 965–972.
- Terwilliger, T. C. (2003). *Acta Cryst. D* **59**, 38–44.
- Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. (2003). *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
- Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998). *Acta Cryst. D* **54**, 243–252.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst. D* **55**, 191–205.
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Wold, M. S. (1997). *Annu. Rev. Biochem.* **66**, 61–92.
- Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997). *Nature Struct. Biol.* **4**, 153–157.
- Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. D. (2004). *Acta Cryst. D* **60**, 1833–1839.
- Ye, Y. & Godzik, A. (2003). *Bioinformatics*, **19**, Suppl. 2, ii246–ii255.
- Ye, Y. & Godzik, A. (2005). *Bioinformatics*, **21**, 2362–2369.