Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the cytoplasmic domain of *Yersinia pestis* YscD, an essential component of the type III secretion system

George T. Lountos,[a,b] Joseph E. Tropea[b] and David S. Waugh[b]*

[a]Basic Science Program, SAIC-Frederick Inc., National Cancer Institute at Frederick, Frederick, MD 21702-1201, USA, and [b]Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, MD 21702-1201, USA

Correspondence e-mail: waughd@mail.nih.gov

Received 28 October 2011
Accepted 16 December 2011

PDB Reference: cytoplasmic domain of YscD, 4a0e.

The *Yersinia pestis* YscD protein is an essential component of the type III secretion system. YscD consists of an N-terminal cytoplasmic domain (residues 1–121), a transmembrane linker (122–142) and a large periplasmic domain (143–419). Both the cytoplasmic and the periplasmic domains are required for the assembly of the type III secretion system. Here, the structure of the YscD cytoplasmic domain solved by SAD phasing is presented. Although the three-dimensional structure is similar to those of forkhead-associated (FHA) domains, comparison with the structures of canonical FHA domains revealed that the cytoplasmic domain of YscD lacks the conserved residues that are required for binding phosphothreonine and is therefore unlikely to function as a true FHA domain.

1. Introduction

The Gram-negative enterobacterium *Yersinia pestis* is the causative agent of bubonic plague or 'Black Death', one of the most deadly diseases in recorded history (Titball & Leary, 1998; Raoult *et al.*, 2000). Although confirmed cases of plague are relatively infrequent at the present time, it remains a significant public health concern in the 21st century, as shown by recent outbreaks in India, Southeast Asia, Africa and even North America (Stenseth *et al.*, 2008; Schrag & Wiener, 1995). Furthermore, given the course of current world events, there is also concern about the potential use of *Y. pestis* as an agent of bioterrorism (Inglesby *et al.*, 2000; Ligon, 2006).

Like many other Gram-negative bacterial pathogens, *Y. pestis* transports a number of effector proteins across both bacterial membranes and the eukaryotic plasma membrane into the cytosol of eukaryotic cells *via* a type III secretion system (T3SS). The Yop (*Yersinia* outer proteins) effectors, which are thought to be delivered through a hollow needle conduit, allow the bacteria to overpower the innate immune response of the host by interfering with signal transduction pathways that regulate the actin cytoskeleton, phagocytosis, apoptosis and the inflammatory response (Navarro *et al.*, 2005; Trosky *et al.*, 2008). The membrane-spanning structure associated with the T3SS, which has been termed the injectisome (Cornelis, 2006; Galán & Wolf-Watz, 2006; Blocker *et al.*, 2008), is composed of at least 21 *Yersinia* secretion (Ysc) proteins (Cornelis, 2002; Plano *et al.*, 2001).

YscD is a member of the EscD/PrgH/YscD family of proteins, which are conserved components of many T3SSs. In concert with other Ysc proteins, YscD forms a ring-shaped assembly that spans the inner membrane of the bacterium at the base of the injectisome (Spreter *et al.*, 2009; Hueck, 1998;

Marlovits *et al.*, 2004). YscD is a 419-residue transmembrane protein that contains a predicted cytoplasmic forkhead-associated (FHA) domain (residues 1–121; Pallen *et al.*, 2002), a single-pass transmembrane segment (residues 122–142) and a larger periplasmic domain (residues 143–419; Ross & Plano, 2011). Studies of YscD and its homologs have shown that it is an essential component of the T3SS injectisome (Plano & Straley, 1995). Deletion analysis of YscD revealed that both the cytoplasmic and periplasmic domains are necessary for its function (Ross & Plano, 2011). Removal of the cytoplasmic domain completely abolishes type III secretion, yet its role in the secretion process has not yet been elucidated. Deletion analysis of the periplasmic domain of YscD has shown that its C-terminal region (residues 354–419) is also essential for secretion and that it mediates binding interactions between YscD and the outer-membrane YscC secretion complex (Ross & Plano, 2011). Although the structures of some putative YscD homologs from other T3SSs are known (one cytoplasmic and one periplasmic domain; McDowell *et al.*, 2011; Spreter *et al.*, 2009), no structures of YscD itself have been reported. Therefore, to gain some insight into the function of the essential cytoplasmic domain of YscD, we solved its crystal structure at a resolution of 2.04 Å and compared it with those of canonical FHA domains.

2. Materials and methods

2.1. Cloning and protein expression

The N-terminal domain of *Y. pestis* YscD (Ser2–Arg121) was amplified from the full-length gene by the polymerase chain reaction (PCR) using the following oligodeoxyribonucleotide primers: 5′-GAG AAC CTG TAC TTC CAG AGT TGG GTC TGT CGT TTT TAT CAA GGG-3′ and 5′-G GGG ACC ACT TTG TAC AAG AAA GCT GGG TTA TTA TCG TGA ACG AGG TAA CCT GTC GGT TGG-3′ (primer R). The resulting PCR amplicon was used as the template for a second PCR with the following primers: 5′-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC GGA GAA CCT GTA CTT CCA G-3′ and primer R. The amplicon from the second PCR was inserted by recombinational cloning into the entry vector pDONR221 (Invitrogen, Carlsbad, California, USA) and the nucleotide sequence was confirmed. The open reading frame of the N-terminal domain of YscD (Ser2–Arg121), now with a recognition site (ENLYFQ/S) for tobacco etch virus (TEV) protease added to its N-terminus, was moved by recombinational cloning into the destination vector pDEST527 (Protein Expression Laboratory, SAIC-Frederick, Frederick, Maryland, USA) to produce pJT170. This plasmid directs the expression of the N-terminal domain of YscD with a hexahistidine tag preceding the TEV protease recognition site. The fusion protein was expressed in *Escherichia coli* strain BL21-CodonPlus (DE3)-RIPL (Stratagene, La Jolla, California, USA). Cells were grown to mid-log phase ($OD_{600} \simeq 0.5$) at 310 K in LB broth containing 100 µg ml^{-1} ampicillin, 30 µg ml^{-1} chloramphenicol and 0.2% glucose. Overproduction of the fusion protein was induced with isopropyl β-D-1-thiogalactopyranoside at a final concentration of 1 m*M* for 4 h at 303 K. The cells were pelleted by centrifugation and stored at 193 K. Pilot studies indicated poor cleavage of the fusion protein by TEV protease. To rectify this situation, DNA encoding three glycine residues was inserted between the TEV protease recognition site and Ser2 of YscD using the QuikChange mutagenesis kit (Stratagene, La Jolla, California, USA). The resulting construct, pJT173, produced a fusion protein that was completely cleaved by TEV protease.

2.2. Protein purification

All procedures were performed at 277–281 K. 5–10 g *E. coli* cell paste was suspended in 150 ml ice-cold 50 m*M* sodium phosphate pH 7.5, 200 m*M* NaCl, 25 m*M* imidazole (buffer *A*) containing 1 m*M* benzamidine–HCl (Sigma Chemical Co., St Louis, Missouri, USA) and Complete EDTA-free protease-inhibitor cocktail tablets (Roche Molecular Biochemicals, Indianapolis, Indiana, USA). The cells were lysed using an APV-1000 homogenizer (Invensys APV Products, Albertslund, Denmark) at 69 MPa and centrifuged at 30 000*g* for 30 min. The supernatant was filtered through a 0.22 µm polyethersulfone membrane and applied onto a 12 ml Ni–NTA Superflow column (Qiagen, Valencia, California, USA) equilibrated in buffer *A*. The column was washed to baseline with buffer *A* and eluted with a linear gradient of imidazole to 250 m*M*. Fractions containing recombinant His_6-GGG-YscD (Ser2–Arg121) were pooled, concentrated using an Amicon YM10 membrane (Millipore, Bedford, Massachusetts, USA), diluted with 50 m*M* sodium phosphate pH 7.5, 200 m*M* NaCl buffer to reduce the imidazole concentration to about 25 m*M* and digested overnight at room temperature with His_6-tagged TEV protease (Kapust *et al.*, 2001). The digest was applied onto a 12 ml Ni–NTA Superflow column equilibrated in buffer *A* and recombinant GGG-YscD (Ser2–Arg121) emerged in the column effluent. The effluent was incubated overnight at 277 K with 10 m*M* dithiothreitol, concentrated as above and applied onto a HiPrep 26/60 Sephacryl S-200 HR column (GE Healthcare Biosciences, Piscataway, New Jersey, USA) equilibrated in 25 m*M* Tris pH 7.2, 150 m*M* NaCl and 2 m*M* TCEP buffer. The peak fractions containing YscD were pooled and concentrated to 30–35 mg ml^{-1} (estimated at 280 nm using a molar extinction coefficient of 12 490 M^{-1} cm^{-1}). Aliquots were flash-cooled in liquid nitrogen and stored at 193 K. The final product was judged to be >95% pure by SDS–PAGE. The molecular weight was confirmed by electrospray ionization mass spectroscopy.

2.3. Protein crystallization

Initial crystallization screens were conducted by the sitting-drop vapor-diffusion method using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, California, USA) and 96-well Intelli-Plates (Art Robbins Instruments). The N-terminal domain (GGG-Ser2–Arg121) of YscD was screened against a variety of sparse-matrix kits from Hampton Research (Aliso Viejo, California, USA) and Qiagen using a protein concentration of 33.2 mg ml^{-1}. Crystals were obtained

Table 1
X-ray data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	YscD-N, KI derivative	YscD-N
X-ray source	MicroMax-007 HF	22-ID, SER-CAT
Wavelength (Å)	1.5418	1.2
Resolution (Å)	50–2.56 (2.65–2.56)	50–2.04 (2.08–2.04)
Space group	*P*432	*P*432
Unit-cell parameters (Å)	$a = b = c = 117.8$	$a = b = c = 118.2$
Total reflections/unique reflections	391535/17144	93479/18490
Completeness (%)	100 (100)	98.9 (99.3)
$R_{\rm merge}$† (%)	5.8 (32.3)	6.1 (63.6)
Mean $I/\sigma(I)$	50.9 (12.6)	41.6 (2.1)
Multiplicity	22.8 (22.5)	5.1 (4.0)
No. of heavy-atom sites found	11	
Figure of merit	0.40	
Refinement statistics		
Resolution (Å)		39.4–2.04
No. of reflections		17908
$R_{\rm work}$‡ (%)		21.1
$R_{\rm free}$‡ (%)		25.6
No. of atoms/mean *B* factor (Å^2)		
Protein chain *A*		825/44.0
Protein chain *B*		825/52.2
Water		88/52.9
R.m.s. deviations from ideal geometry		
Bond lengths (Å)		0.007
Bond angles (°)		1.1
Ramachandran plot		
Most favored (%)		94.8
Additionally allowed (%)		5.2
MolProbity analysis		
Clash score		4.54 [98th percentile]
Protein-geometry score		1.25 [99th percentile]
C^{β} deviations > 0.25 Å		0
Residues with bad bonds		0
Residues with bad angles		0
PDB code		4a0e

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of multiply recorded reflections. ‡ $R = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$. $R_{\rm free}$ is the *R* value calculated for 5% of the data set not included in the refinement.

from Qiagen JCSG Core IV screen condition No. 27 [0.085 *M* Tris–HCl pH 8.5, 25.5%(*w*/*v*) PEG 4000, 0.17 *M* sodium acetate, 15%(*v*/*v*) glycerol] at 291 K. These crystals initially diffracted to approximately 3.0 Å resolution using the home X-ray source. Optimization of the crystals was performed by the hanging-drop vapor-diffusion method and examining the effects of a variety of additives from the Hampton Research Additive Screen. Crystals for data collection were obtained by mixing 5 µl protein solution (33.2 mg ml^{-1}), 4 µl reservoir solution [0.085 *M* Tris–HCl pH 8.5, 25.5%(*w*/*v*) PEG 4000, 0.17 *M* sodium acetate, 15%(*v*/*v*) glycerol] and 1 µl TCEP (100 m*M*). The drops were sealed over 1 ml crystallization well solution and stored at 291 K. Block-shaped crystals grew within one week to approximate dimensions of 0.2 × 0.2 × 0.2 mm.

2.4. Data collection, phasing and model refinement

Crystals of the YscD N-terminal domain were retrieved from the crystallization drop using a LithoLoop (Molecular Dimensions, Apopka, Florida, USA) and flash-cooled by plunging them into liquid nitrogen without the use of an additional cryoprotectant. Native X-ray diffraction data were collected from a single crystal maintained at 100 K using a MAR 300 CCD detector on beamline 22-ID of the SER-CAT facilities at the Advanced Photon Source, Argonne National Laboratory using remote data-collection procedures. 180 frames of data were collected using a wavelength of 1.2 Å, an oscillation angle of 0.5°, a crystal-to-detector distance of 200 mm and an exposure time of 2 s. The data were integrated and scaled with *HKL*-3000 (Minor *et al.*, 2006). Data-collection statistics are outlined in Table 1.

Since at the time of data collection there was no suitable structural homolog available for phasing by molecular replacement, we sought to solve the structure of the YscD N-terminal domain by single anomalous dispersion (SAD) phasing using the quick-soak halide method (Dauter *et al.*, 2000; Dauter & Dauter, 2007). A derivative for phasing was prepared by soaking a crystal in 0.085 *M* Tris–HCl pH 8.5, 25.5%(*w*/*v*) PEG 4000, 0.17 *M* sodium acetate, 15%(*v*/*v*) glycerol and 0.5 *M* potassium iodide for 2 min, after which the crystal was retrieved from the drop using a LithoLoop and flash-cooled by plunging it into liquid nitrogen. X-ray diffraction data were collected using a MAR345 detector mounted on a Rigaku MicroMax-007 HF high-intensity microfocus X-ray generator equipped with VariMax HF optics (Rigaku, The Woodlands, Texas, USA) and operated at 40 kV and 30 mA (λ = 1.5418 Å). 360 frames of data were collected using an exposure time of 5 min, an oscillation angle of 0.5° and a crystal-to-detector distance of 175 mm. The data were integrated and scaled with *HKL*-3000 without merging the Bijvoet pairs. Although the iodide-soaked crystals diffracted X-rays to approximately 2.2 Å resolution, useful anomalous signal was only detected to 2.56 Å resolution (Dauter, 2006). The structure of the N-terminal YscD domain was solved using the automated experimental phasing with model building option in the *PHENIX* software suite (Adams *et al.*, 2010, 2011). Nine iodide sites were located in the initial substructure by the *HySS* submodule (Grosse-Kunstleve & Adams, 2003). Two additional iodide sites were located by *Phaser* and experimental phases were calculated based on 11 refined iodide sites (figure of merit 0.40; Read & McCoy, 2011). Density modification was performed with *RESOLVE* (Terwilliger, 2000). Automated model building using *RESOLVE* built 210 residues of the two molecules in the asymmetric unit of the crystal, resulting in an initial model with an *R* value of 0.27 and an $R_{\rm free}$ of 0.32 (Terwilliger, 2003).

The model derived from SAD phasing was then used as a search model for molecular replacement using the native data set to 2.04 Å resolution and the automated molecular replacement with *Phaser* option in *PHENIX* (McCoy *et al.*, 2007; Adams *et al.*, 2011). Iterative rounds of manual model rebuilding and refinement were performed with *Coot* (Emsley *et al.*, 2010) followed by refinement with *phenix.refine* (Afonine *et al.*, 2005). The refinement was monitored by the $R_{\rm free}$ value, which was calculated using a randomly selected 5% of the reflections (Brünger, 1997). Water molecules were located using *Coot* and manually inspected, and their positions were refined with *phenix.refine*. In the final stages of refine-

ment, TLS refinement was performed (Painter & Merritt, 2006). The quality of the refined model was validated with *MolProbity* (Chen *et al.*, 2010). Secondary-structure elements were assigned with *DSSP* (Kabsch & Sander, 1983) and sequence alignments were performed with *ClustalW* (Goujon *et al.*, 2010). Topology plots were generated using *PDBSum* (Laskowski, 2009). All figures were prepared with *PyMOL* (http://www.pymol.org). The atomic coordinates and structure-factor files have been deposited in the Protein Data Bank under accession code 4a0e.

3. Results and discussion

3.1. Structure determination

X-ray data-collection and model-refinement statistics are summarized in Table 1. The structure of the YscD N-terminal domain (hereafter referred to as YscD-N) was solved by SAD phasing using crystals derivatized with iodide ions. After identifying the positions of 11 iodide ions, the experimentally determined maps at 2.56 Å resolution allowed the automated building of 210 out of 248 residues in the asymmetric unit, which contained two molecules as suggested by the Matthews coefficient (2.56 Å^3 Da^{-1}) and solvent content (51.9%) (Matthews, 1968; Kantardjieff & Rupp, 2003). The native YscD-N crystals diffracted to a resolution of 2.04 Å in space group *P*432, with unit-cell parameters $a = b = c = 118.2$ Å. Using the model derived from SAD phasing, the two molecules of YscD-N in the asymmetric unit were located and refined to a final *R* value of 21.1% and $R_{\rm free}$ of 25.6%. The final model consisted of a single non-native glycine at the N-terminus and residues Ser2–Thr108 in each monomer. There was no electron density present for residues 109–121 and therefore they were not included in the final model. An example of the quality of the electron-density map is shown in Fig. 1 for residues Trp3–Lys11. Model validation using *MolProbity* and *PROCHECK* (Laskowski *et al.*, 1993) indicated that 94.8% of the residues are located in the most favored region and none are found in disallowed regions of the Ramachandran plot. The all-atom contact clash score and *MolProbity* protein-geometry scores were in the 98th and 99th percentiles, respectively, indicating good geometry.

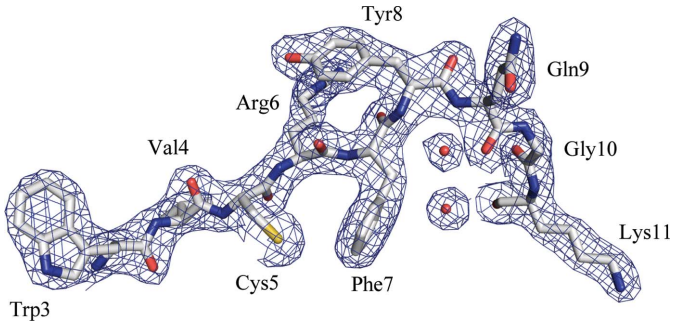
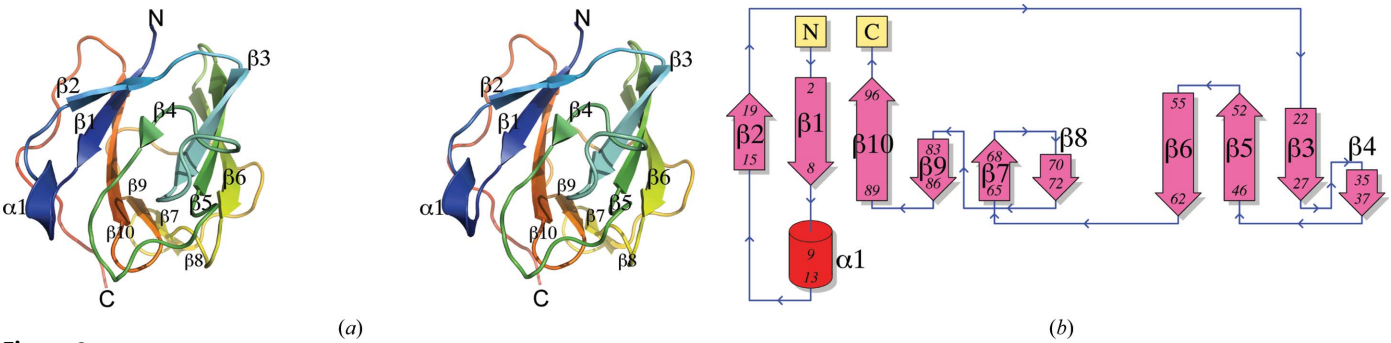


Figure 1
A representative example of the quality of the final $2F_o - F_c$ electron-density map (blue, 2.04 Å resolution, contoured at 1.0σ), highlighting the fit of residues Trp3–Lys11 to the maps.

3.2. Overall structure

YscD-N folds into a structure with ten β-strands and one α-helix (Fig. 2). The core scaffold consists of a β-sandwich composed of two β-sheets: a six-stranded antiparallel β-sheet on one side and a four-stranded mixed β-sheet on the other. A 3_{10}-helix is located between strands $\beta1$ and $\beta2$ (Fig. 2*a*). The electrostatic surface of YscD-N is characterized by large negatively charged patches on both sides of the molecule with additional small positively charged patches and dispersed hydrophobic regions (not shown).

YscD-N crystallized with a dimer in the asymmetric unit. The elution profile during size-exclusion chromatography suggested that YscD-N is a monomer in solution (data not shown), although Ross and Plano have reported that isolated YscD-N may dimerize or further oligomerize (Ross & Plano, 2011). However, the dimer of YscD-N that is found in the asymmetric unit of the crystal suggests that it is weakly associated, with monomer–monomer contacts facilitated by only five interface residues: Arg23, Thr59, Asp60, Leu74 and Gly75. The interface includes only one hydrogen-bonding pair, between Leu74 (chain *A*) and Arg23 (chain *B*), and 18 nonbonded contacts. Curiously, all but one of the residues (Arg23) that form the putative dimer interface of YscD-N are not conserved in homologs from closely related T3SSs (not shown). The total buried surface area between the two chains is approximately 285 Å^2 of each monomer as calculated using *AREAIMOL* from the *CCP*4 suite of programs (Winn *et al.*,



(*a*) (*b*)

Figure 2
(*a*) Stereo image of the YscD-N monomer with labeled topological elements. The N- and C-termini are labeled. (*b*) Topology plot of YscD-N generated with *PDBSum* (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum).

2011). This is a rather small interface for a stable protein dimer (Yan *et al.*, 2008). Lattice contacts revealed another possible dimer in the crystal with a larger buried surface area (498 Å^2; not shown), but the C-termini of the molecules in this dimer point in opposite directions so that it is exceedingly unlikely that they could simultaneously cross the inner membrane. On the other hand, the C-termini of the molecules in the dimer that is observed in the asymmetric unit of the crystal point in the same direction and both could very readily cross the membrane in this orientation. Analysis of the coordinates using the *PDBePISA* server did not reveal any specific interactions that could result in the formation of stable quaternary structures and therefore suggests that dimerization does not occur in solution (Krissinel & Henrick, 2007). The structural analog of YscD-N in the *Shigella flexneri* T3SS, MxiG, is a monomer in solution, although it was portrayed as an oligomer in a model of the 24-subunit ring structure assembled from electron-microscopy images (McDowell *et al.*, 2011). Taken together, we believe these observations suggest that the N-terminal domain of YscD (and its analogs in other T3SSs) interacts only weakly with itself, if at all, in solution and that oligomerization must be facilitated and/or stabilized by interactions with other components of the injectisome or by concomitant oligomerization of its periplasmic (C-terminal) domains.

3.3. Comparison with structural homologs

Structural homologs of YscD-N were identified by submitting the coordinates to the *PDBeFold* server (Krissinel & Henrick, 2004; http://www.ebi.ac.uk/msd-srv/ssm). The structure is similar to those of several forkhead-associated domains. The two closest structural homologs identified were the FHA domain of the CT664 type III secretion protein from *Chlamydia trachomatis* (PDB entry 3gqs; Midwest Center for Structural Genomics, unpublished work) and the *Mycobacterium tuberculosis* Rv0020c FHA domain (PDB entry 3po8; Pennell *et al.*, 2010). Structural alignment of the YscD-N (Fig. 3*a*, green) coordinates with those of the CT664 protein (Fig. 3*a*, blue) reveals that the two proteins have an r.m.s.d. value of 1.80 Å over 93 aligned residues and a sequence identity of 21%. There is a strong similarity between the core topological elements of the protein (Fig. 3*a*, arrow *a*). While the topology of the YscD-N domain can be described as $\beta\alpha\beta_{10}$, the topology of the CT664 protein is β_8, the 3_{10}-helix between strands $\beta1$ and $\beta2$ in YscD-N is not found in CT664 (Fig. 3*a*, arrow *b*) and YscD-N contains two additional small β-strands ($\beta7$ and $\beta8$) that are not present in CT664 (Fig. 3*a*, arrow *c*). Furthermore, many of the loops that connect the β-strands do not align well between the two structures.

Structural alignment of YscD-N with the Rv0020c FHA domain (Fig. 3*b*, magenta) reveals an r.m.s.d. value of 1.82 Å over 88 aligned residues and 18% sequence identity. Here again the two structures share strong topological similiarity in the core β-sheets of the proteins (Fig. 3*b*, arrow *a*). Differences between the two structures include the lack of the 3_{10}-helix between strands $\beta1$ and $\beta2$ (Fig. 3*b*, arrow *b*) in the Rv0020c FHA domain and several structural differences in the conformations of the loops connecting the β-strands. The most significant structural difference is found in the region between strands $\beta6$ and $\beta9$ in YscD-N. While there are two additional small β-strands $\beta7$ and $\beta8$ in YscD-N, only one β-strand is found in this

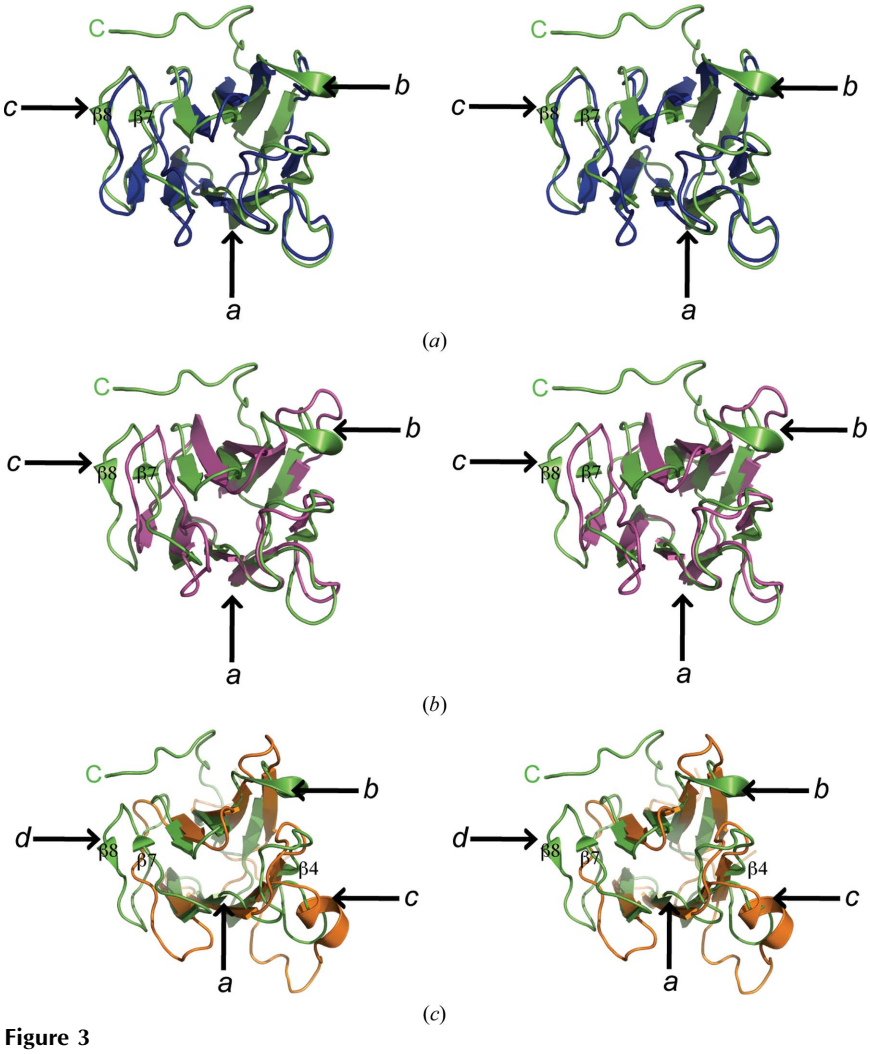


Figure 3
(*a*) Stereo image of the coordinates of YscD-N (green ribbons) superimposed with those of the *C. trachomatis* CT664 FHA domain (blue ribbons; PDB entry 3gqs). (*b*) Stereo image of the coordinates of YscD-N (green ribbons) superimposed with those of the *M. tuberculosis* Rv0020c FHA domain (magenta ribbons; PDB entry 3po8). (*c*) Stereo image of the coordinates of YscD-N (green ribbons) superimposed with those of the *S. flexneri* MxiG FHA domain (orange ribbons; PDB entry 2xxs; McDowell *et al.*, 2011). Arrows are used to highlight the structural differences described in the text.

region in Rv0020c which does not align with any β-strand in the YscD structure (Fig. 3*b*, arrow *c*). The conformation of the loops in this region also differ significantly, as none of the backbone Cα atoms in this region superimpose at all.

A recent publication by McDowell and coworkers reported the three-dimensional NMR structure of the N-terminal domain of the *S. flexneri* type III secretion protein MxiG, a putative functional counterpart of YscD-N (McDowell *et al.*, 2011). The N-terminal domain of MxiG (MxiG-N), which protrudes into the cytosol like YscD-N, also shows strong structural similarity to other FHA domains. Curiously, the structure of YscD-N deviates more from the structure of its putative functional homolog MxiG-N than from that of the non-functionally homologous Rv0020c FHA domain (Fig. 3*c*). The superimposed YscD and MxiG N-terminal domain structures exhibit an r.m.s.d. value of 2.39 Å over 77 aligned residues with only 9% sequence identity based on structure-based sequence alignment as analyzed by *PDBeFold*. As observed for other FHA domains, the two proteins share a highly similar core structure (Fig. 3*c*, arrow *a*). Significant structural differences are found in several regions of the two structures. The 3_{10}-helix located between strands β1 and β2 in YscD-N is not found in MxiG-N (Fig. 3*c*, arrow *b*). Additionally, there is a significant difference between the two proteins in the region between strands β3 and β5. In MxiG-N this region contains an α-helix that is not found in YscD-N, a longer loop with a significantly different conformation to that in YscD-N and a much longer β-strand that superimposes on strand β4 in YscD-N (Fig. 3*c*, arrow *c*). Additional conformational differences exist in the region highlighted by arrow *d* in Fig. 3(*c*). This region, which is located between strands β6 and β9 in YscD-N, reveals significant structural deviations when aligned with MxiG-N, including two small β-strands (β7 and β8) in YscD-N that are not present in MxiG-N and differences in the conformation of the loop in the same part of MxiG-N. Thus, the structural alignment of YscD-N with MxiG-N indicates that while the two proteins share strong homology to canonical FHA domains in their core structures, there are significant structural differences in their outer peripheral regions. Although both structures share the FHA fold, there is high sequence divergence between the two homologs (Fig. 4*a*). This is not unusual, however, as previous studies have shown that FHA domains exhibit a high degree of sequence diversity (Hammet *et al.*, 2003).

Sequence alignments of YscD-N with the N-terminal domains of other T3SS homologs such as PscD (*Pseudomonas aeruginosa* 152504), AscD (*Aeromonas hydrophila*) and LscD (*Photorhabus luminescens*) exhibit sequence identity ranging from 41 to 48%, but only of about 15% when any of them is aligned with MxiG-N (Fig. 4*a*). Fig. 4(*b*) illustrates the distribution of conserved surface-exposed (green) residues and buried (red) residues mapped onto the structure of YscD-N. The buried residues are primarily found within the β-sheet sandwich of the FHA domain. One striking feature that can be seen in Fig. 4(*b*) is a patch of surface residues (Asp29–Gln32 in YscD) in the loop between strands β3 and β4 that are conserved among the PscD, AscD and LscD homologs. As these residues are solvent-exposed, this may be an area of interest for targeted mutagenesis to probe possible functional roles of this loop. It is thus likely that YscD-N is a closer functional and structural homolog of PscD, AscD and LscD than it is of MxiG-N.

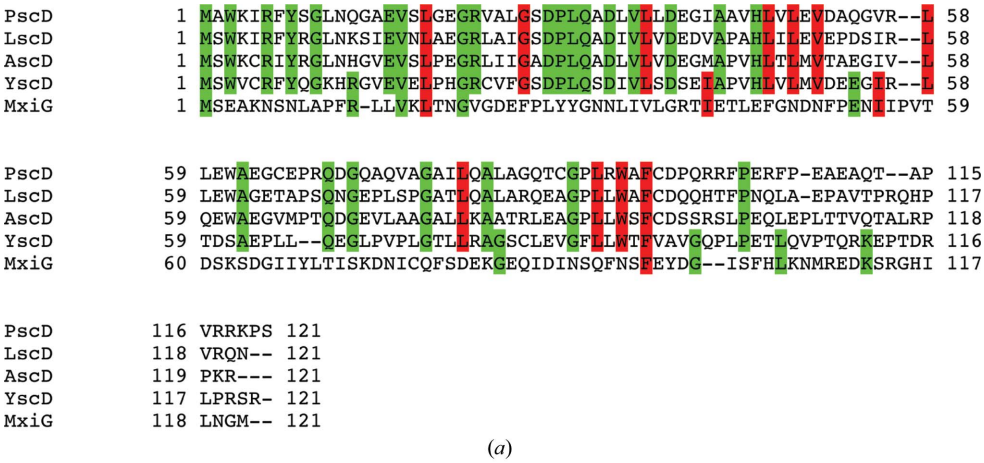


Figure 4
(*a*) Sequence alignment of the N-terminal domains of the type III secretion-system homologs PscD (*Pseudomonas aeruginosa* 152504), AscD (*A. hydrophila*), LscD (*Photorhabus luminescens*), YscD (*Y. pestis* C092) and MxiG (*S. flexneri*). Conserved surface residues are highlighted in green and buried residues are highlighted in red. (*b*) The conserved residues are mapped onto the structure of YscD-N.

3.4. Structural insights into biological function

FHA domains can be found in over 700 eukaryotic proteins such as kinases, phosphatases, kinesins, DNA-binding transcription factors and other enzymes (Hofmann & Bucher, 1995; Yaffe & Smerdon, 2004; Durocher & Jackson, 2002). Furthermore, there are more than 500 bacterial proteins containing FHA domains that have been implicated in the injection of bacterial

proteins into the host cell, transmembrane transport and cell division (Pallen *et al.*, 2002, 2005). A common and unifying characteristic of canonical FHA domains in these biological processes is their high specificity for binding phosphothreonine residues (Yaffe & Smerdon, 2004; Liang & Van Doren, 2008; Pennell *et al.*, 2010). Although the three-dimensional structure of YscD-N shows strong homology to many canonical FHA domains, functional studies conducted with the type III secretion system homolog MxiG-N suggest that it does not interact with phosphothreonine (McDowell *et al.*, 2011). We therefore compared the crystal structure of YscD-N with the crystal structures of canonical FHA domains from *M. tuberculosis* Rv0020c and human checkpoint kinase 2 (Chk2) bound to phosphothreonine-containing peptides in order to gain insight into whether YscD-N may or may not bind phosphothreonine.

In the crystal structure of the Rv0020c FHA domain complexed with the phosphopeptide KIEpTATI, direct hydrogen bonding to phosphothreonine is mediated by Arg459, Ser473, Arg474, Thr494 and Asn495 (Pennell *et al.*, 2010). The interactions with Ser473 and Arg474 are conserved among FHA domains and residues Asn495 and His519 are part of a cleft that acts as a specificity pocket for the phosphothreonine. Site-directed mutagenesis experiments in which the conserved residues Arg459, Ser473 and Asn495 were mutated to alanine resulted in large decreases in the binding affinity for the phosphothreonine peptide. In fact, the authors reported that the S473A mutation resulted in no detectable peptide binding (Pennell *et al.*, 2010). Sequence and structural alignment of the YscD-N coordinates with the FHA domains of Rv0020c and Chk2 reveal that YscD-N lacks the conserved arginine, asparagine and serine residues that are critical for the binding of phosphothreonine (Fig. 5*a*). In the structural alignment, Arg59 and Arg117 in Rv0020c (Fig. 5*b*) and Chk2 (Fig. 5*c*) (Li *et al.*, 2002), respectively, align with Pro44 in YscD-N and the conserved serine residue is replaced by alanine in YscD-N. The solution NMR structure of MxiG-N also reveals that it does not possess the appropriate arrangement of conserved residues that is required for phosphothreonine binding (McDowell *et al.*, 2011). Fig. 5(*d*) illustrates the distribution of structurally aligned conserved surface (green) and buried (red) residues between YscD-N and Rv0020c mapped onto the structure of YscD-N. The buried residues are clustered within the β-sheet sandwich and are likely to contribute to the conservation of the FHA fold, while the surface residues are distributed in various loop regions and do not

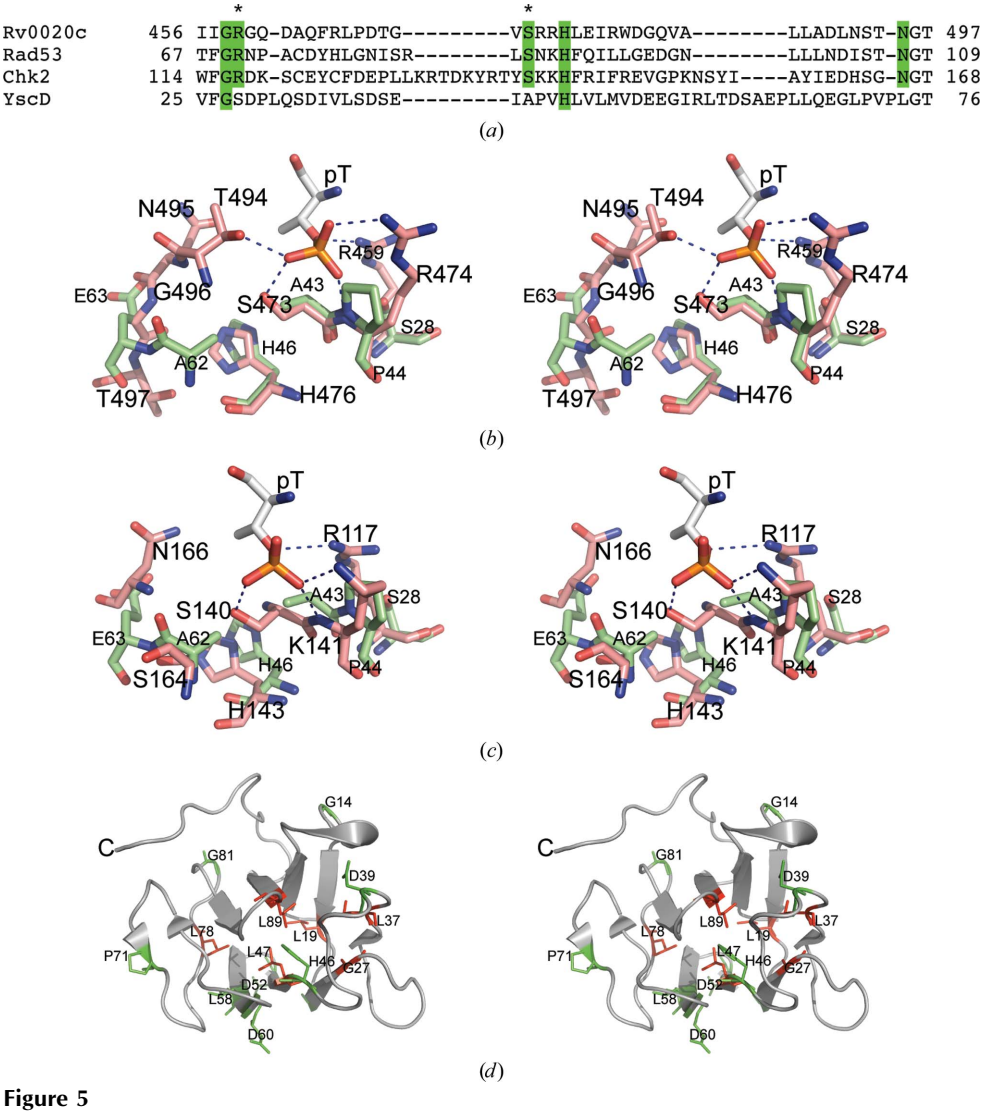


Figure 5
(*a*) Sequence alignment of residues near the pThr-binding pocket of the FHA domains in Rv0020c (*M. tuberculosis*), Rad53 (*Saccharomyces cerevisiae*), Chk2 (human checkpoint kinase 2) and YscD (*Y. pestis* C092). Residues implicated in phosphothreonine binding in FHA domains are highlighted in green. The residues highlighted with an asterisk were shown to have a significant decrease in pThr binding when mutated to alanine in Rv0020c. (*b*) Structural alignment of the coordinates of YscD-N (green) with those of the Rv0020c–phosphopeptide complex (salmon; PDB entry 3po8). O atoms are colored red, N atoms blue and phosphates orange. For clarity, only the pThr residue from the peptide and protein residues in the pThr-binding pocket are shown. (*c*) Structural alignment of the coordinates of YscD-N (green) with those of the Chk2 FHA domain in complex with a phosphopeptide (salmon; PDB entry 1gxc; Li *et al.*, 2002). (*d*) The conserved residues between the FHA domains of YscD and Rv0020c based on structural alignment are mapped onto the structure of YscD-N. Surface residues are colored green and buried residues are colored red.

immediately suggest any potential functional conservation. Therefore, although the structures of YscD-N and MxiG-N resemble those of canonical FHA domains, it is unlikely that the either of them binds phosphothreonine-containing peptides. The recent structure determination of the FHA domain of KIF13B bound to CENTA1 provides structural evidence for phosphorylation-independent binding interactions mediated by an FHA domain (Tong *et al.*, 2010). Indeed, the KIF13B FHA domain lacks the conserved phosphate-interaction residues of the pThr-binding pocket and therefore indicates that the FHA domain can perform functions other than phosphothreonine binding. Indeed, as more structures of proteins containing FHA domains are solved, additional functions of this versatile fold are likely to be discovered.

Currently, the biological function of the cytoplasmic domain of YscD remains elusive, but genetic and biochemical studies of YscD and its homologs in other T3SSs suggest that YscD-N interacts with other type III secretion-system components that are involved in the assembly of the cytoplasmic ring structure at the base of the injectisome (Johnson *et al.*, 2008; Morita-Ishihara *et al.*, 2006; Plano & Straley, 1995). The determination of the crystal structure of YscD-N is an important first step in unraveling its biological function, which could reveal potential new mechanisms for the FHA domain.

We thank Scott Cherry for technical assistance. This project was supported in part with Federal funds from the National Cancer Institute, National Institutes of Health under contract HHSN261200800001E and the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the US Government. We thank the Biophysics Resource in the Structural Biophysics Laboratory, NCI-Frederick for use of the LC/ESMS instrument. X-ray diffraction data were collected on the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22-ID of the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. W-31-109-Eng-38.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Adams, P. D. *et al.* (2011). *Methods*, **55**, 94–106.
Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.
Blocker, A. J., Deane, J. E., Veenendaal, A. K., Roversi, P., Hodgkinson, J. L., Johnson, S. & Lea, S. M. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 6507–6513.
Brünger, A. T. (1997). *Methods Enzymol.* **277**, 366–396.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cornelis, G. R. (2002). *Nature Rev. Mol. Cell Biol.* **3**, 742–752.
Cornelis, G. R. (2006). *Nature Rev.* **4**, 811–825.
Dauter, M. & Dauter, Z. (2007). *Methods Mol. Biol.* **364**, 149–158.
Dauter, Z. (2006). *Acta Cryst.* D**62**, 867–876.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Durocher, D. & Jackson, S. P. (2002). *FEBS Lett.* **513**, 58–66.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Galán, J. E. & Wolf-Watz, H. (2006). *Nature (London)*, **444**, 567–573.
Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010). *Nucleic Acids Res.* **38**, W695–W699.
Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.
Hammet, A., Pike, B. L., McNees, C. J., Conlan, L. A., Tenis, N. & Heierhorst, J. (2003). *IUBMB Life*, **55**, 23–27.
Hofmann, K. & Bucher, P. (1995). *Trends Biochem. Sci.* **20**, 347–349.
Hueck, C. J. (1998). *Microbiol. Mol. Biol. Rev.* **62**, 379–433.
Inglesby, T. V. *et al.* (2000). *JAMA*, **283**, 2281–2290.
Johnson, D. L., Stone, C. B. & Mahony, J. B. (2008). *J. Bacteriol.* **190**, 2972–2980.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). *Protein Eng.* **14**, 993–1000.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Li, J., Williams, B. L., Haire, L. F., Goldberg, M., Wilker, E., Durocher, D., Yaffe, M. B., Jackson, S. P. & Smerdon, S. J. (2002). *Mol. Cell*, **9**, 1045–1054.
Liang, X. & Van Doren, S. R. (2008). *Acc. Chem. Res.* **41**, 991–999.
Ligon, B. L. (2006). *Semin. Pediatr. Infect. Dis.* **17**, 161–170.
Marlovits, T. C., Kubori, T., Sukhan, A., Thomas, D. R., Galán, J. E. & Unger, V. M. (2004). *Science*, **306**, 1040–1042.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McDowell, M. A., Johnson, S., Deane, J. E., Cheung, M., Roehrich, A. D., Blocker, A. J., McDonnell, J. M. & Lea, S. M. (2011). *J. Biol. Chem.* **286**, 30606–30614.
Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.
Morita-Ishihara, T., Ogawa, M., Sagara, H., Yoshida, M., Katayama, E. & Sasakawa, C. (2006). *J. Biol. Chem.* **281**, 599–607.
Navarro, L., Alto, N. M. & Dixon, J. E. (2005). *Curr. Opin. Microbiol.* **8**, 21–27.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Pallen, M. J., Beatson, S. A. & Bailey, C. M. (2005). *BMC Microbiol.* **5**, 9.
Pallen, M., Chaudhuri, R. & Khan, A. (2002). *Trends Microbiol.* **10**, 556–563.
Pennell, S., Westcott, S., Ortiz-Lombardía, M., Patel, D., Li, J., Nott, T. J., Mohammed, D., Buxton, R. S., Yaffe, M. B., Verma, C. & Smerdon, S. J. (2010). *Structure*, **18**, 1587–1595.
Plano, G. V., Day, J. B. & Ferracci, F. (2001). *Mol. Microbiol.* **40**, 284–293.
Plano, G. V. & Straley, S. C. (1995). *J. Bacteriol.* **177**, 3843–3854.
Raoult, D., Aboudharam, G., Crubézy, E., Larrouy, G., Ludes, B. & Drancourt, M. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 12800–12803.
Read, R. J. & McCoy, A. J. (2011). *Acta Cryst.* D**67**, 338–344.
Ross, J. A. & Plano, G. V. (2011). *J. Bacteriol.* **193**, 2276–2289.
Schrag, S. J. & Wiener, P. (1995). *Trends Ecol. Evol.* **10**, 319–324.
Spreter, T., Yip, C. K., Sanowar, S., André, I., Kimbrough, T. G., Vuckovic, M., Pfuetzner, R. A., Deng, W., Yu, A. C., Finlay, B. B., Baker, D., Miller, S. I. & Strynadka, N. C. (2009). *Nature Struct. Mol. Biol.* **16**, 468–476.

Stenseth, N. C., Atshabar, B. B., Begon, M., Belmain, S. R., Bertherat, E., Carniel, E., Gage, K. L., Leirs, H. & Rahalison, L. (2008). *PLoS Med.* **5**, e3.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 45–49.
Titball, R. W. & Leary, S. E. (1998). *Br. Med. Bull.* **54**, 625–633.
Tong, Y., Tempel, W., Wang, H., Yamada, K., Shen, L., Senisterra, G. A., MacKenzie, F., Chishti, A. H. & Park, H.-W. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 20346–20351.
Trosky, J. E., Liverman, A. D. & Orth, K. (2008). *Cell. Microbiol.* **10**, 557–565.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yaffe, M. B. & Smerdon, S. J. (2004). *Annu. Rev. Biophys. Biomol. Struct.* **33**, 225–244.
Yan, C., Wu, F., Jernigan, R. L., Dobbs, D. & Honavar, V. (2008). *Protein J.* **27**, 59–70.