# Expression, limited proteolysis and preliminary crystallographic analysis of IpaD, a component of the *Shigella flexneri* type III secretion system

**Steven Johnson,[a,b] Pietro Roversi,[a] Marianela Espina,[c] Janet E. Deane,[a] Susan Birket,[c] William D. Picking,[c] Ariel Blocker,[b] Wendy L. Picking[c] and Susan M. Lea[a,b]\***

[a]Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, England, [b]Sir William Dunn School of Pathology, University of Oxford, England, and [c]Department of Molecular Biosciences, University of Kansas, USA

Correspondence e-mail: susan.lea@path.ox.ac.uk

IpaD, the putative needle-tip protein of the *Shigella flexneri* type III secretion system, has been overexpressed and purified. Crystals were grown of the native protein in space group $P2_12_12_1$, with unit-cell parameters $a = 55.9$, $b = 100.7$, $c = 112.0$ Å, and data were collected to 2.9 Å resolution. Analysis of the native Patterson map revealed a peak at 50% of the origin on the Harker section $v = 0.5$, suggesting twofold non-crystallographic symmetry parallel to the $b$ crystallographic axis. As attempts to derivatize or grow selenomethionine-labelled protein crystals failed, in-drop proteolysis was used to produce new crystal forms. A trace amount of subtilisin Carlsberg was added to IpaD before sparse-matrix screening, resulting in the production of several new crystal forms. This approach produced SeMet-labelled crystals and diffraction data were collected to 3.2 Å resolution. The SeMet crystals belong to space group $C2$, with unit-cell parameters $a = 139.4$, $b = 45.0$, $c = 99.5$ Å, $\beta = 107.9°$. An anomalous difference Patterson map revealed peaks on the Harker section $v = 0$, while the self-rotation function indicates the presence of a twofold noncrystallographic symmetry axis, which is consistent with two molecules per asymmetric unit.

## 1. Introduction

*Shigella flexneri*, the causative agent of shigellosis, is a Gram-negative bacterial pathogen that initiates infection by invading cells of the colonic epithelium in a process which is dependent on a type III secretion system (T3SS; Phalipon & Sansonetti, 2003). The T3SS is composed of a basal body, which traverses both bacterial membranes, and an external needle through which effector proteins are secreted (Blocker *et al.*, 2001). During *S. flexneri* infection, secretion is activated by contact of the tip of the needle complex with host cells, resulting in formation of a pore in the host-cell membrane that is contiguous with the needle. Other effector proteins are injected *via* this apparatus directly into the host-cell cytoplasm (for a review, see Johnson *et al.*, 2005).

Of the T3SS effectors, IpaB, IpaC and IpaD are essential for invasion (Menard *et al.*, 1993). IpaB and IpaC have been demonstrated to insert into the host-cell membrane and are therefore suggested to form the translocation pore (Blocker *et al.*, 1999). Cells lacking IpaD are not only incapable of pore insertion and invasion, but also demonstrate impaired effector secretion control (Picking *et al.*, 2005). IpaD was originally proposed to form a plug in the T3SS with IpaB (Menard *et al.*, 1994), but more recent data has demonstrated that its role is more complex (Picking *et al.*, 2005). Deletions within the N-terminal third of the molecule do not affect invasion of the bacteria into host cells, although the insertion of IpaB/IpaC into membranes is slightly impaired. However, even short deletions (five residues) in the C-terminus of IpaD completely abolish the invasive phenotype and pore insertion. These observations, combined with the functional homology between IpaD and LcrV, a protein recently shown to be assembled at the tip of the *Yersinia pestis* T3SS needle (Mueller *et al.*, 2005), suggest that IpaD may play a role in regulating insertion of the IpaB/IpaC translocon pore from the tip of the *S. flexneri* needle.

Our laboratory has recently determined the crystal structure of MxiH, the needle subunit, and assembled a molecular model of the

T3SS needle by docking it into a 16 Å EM reconstruction of the *S. flexneri* needle (Cordes *et al.*, 2003; Deane, Cordes *et al.*, 2006; Deane, Roversi *et al.*, 2006). As part of an ongoing study into the interactions of the T3SS with host cells, we have now expressed, purified and crystallized IpaD. The three-dimensional structure of IpaD, in conjunction with the structure of the MxiH needle and other tip proteins such as LcrV (Derewenda *et al.*, 2004), will hopefully allow us to delineate the mechanism of translocon insertion.

## 2. Experimental procedures

### 2.1. Expression and purification

The DNA fragment of the *ipaD* gene encoding residues 15–332 was produced using the methods described previously (Harrington *et al.*, 2003; Picking *et al.*, 2005) and subcloned into the pET-15b vector. This construct placed the gene in frame with an N-terminal hexahistidine tag. $IpaD_{N\Delta14}$ was expressed in *Escherichia coli* BL21 (DE3) cells grown in LB media containing 100 µg ml$^{-1}$ ampicillin. Cells were grown at 310 K to an $A_{600\,nm}$ of 0.6 and protein overexpression was induced by the addition of 1.0 m$M$ IPTG. After 3 h, cells were
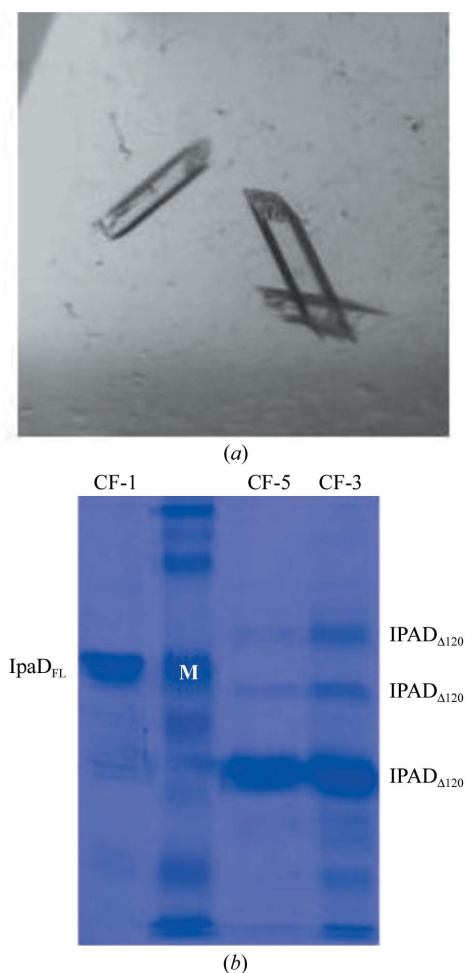


(a)



(b)

#### Figure 1

(a) Monoclinic crystals of IpaD crystal form 2 produced by in-drop proteolysis. (b) SDS–PAGE of protein derived from crystals with (CF-3 and CF-5) and without (CF-1) subtilisin treatment (see text for full description). The subscript denotes the portion of IpaD present in each band as determined by sequencing (FL, full length; Δ120, N-terminal deletion to residue 120). Molecular-weight markers are shown in lane 2; the marker identified with 'M' corresponds to a MW of 42 kDa.

harvested by centrifugation (4000g, 15 min, 277 K) and pellets were frozen at 193 K. Cell pellets were resuspended in lysis buffer [20 m$M$ Tris–HCl pH 7.5, 500 m$M$ NaCl and Complete EDTA-free protease inhibitor cocktail (Roche)] and lysed using an Emulsiflex-C5 Homogeniser (Glen Creston, UK). The resultant cell suspension was centrifuged (20 000g, 30 min, 277 K) and the soluble fraction was applied onto a pre-charged 5 ml His-Trap HP nickel-affinity column (Amersham Biosciences). Protein was eluted using a gradient of 0–1 $M$ imidazole in lysis buffer. Fractions containing $IpaD_{N\Delta14}$ were further purified by gel-filtration chromatography using a HiLoad 26/60 Superdex 75 column (Amersham Biosciences) equilibrated in 20 m$M$ Tris–HCl pH 7.5, 100 m$M$ NaCl, 10 m$M$ DTT. $IpaD_{N\Delta14}$ eluted as a single monomeric peak and was demonstrated to be pure by SDS–PAGE; mass-spectrometric analysis (data not shown) confirmed the molecular weight of the protein minus the N-terminal methionine (37 374 Da). Fractions containing purified $IpaD_{N\Delta14}$ were pooled, concentrated using 5 kDa molecular-weight VivaSpin centrifugal concentrators (VivaScience, Sartorius Group) to approximately 70 mg ml$^{-1}$ and stored at 193 K in 50 µl aliquots.

SeMet-labelled $IpaD_{N\Delta14}$ was produced by expression in the *E. coli* *met$^{-}$* auxotrophic strain B834 (DE3). Cultures were grown in LB media to an $A_{600\,nm}$ of 0.2, pelleted (4000g, 15 min, 277 K) and washed in phosphate-buffered saline three times. Washed cells were used to inoculate SelenoMet Medium Base containing SelenoMet Nutrient Mix, L-selenomethionine (40 µg ml$^{-1}$; Molecular Dimensions Ltd, UK) and ampicillin (100 µg ml$^{-1}$). Cultures were grown at 310 K until an $A_{600\,nm}$ of 0.6 was reached, whereupon the solutions were cooled to 293 K and protein overexpression was induced by the addition of 1.0 m$M$ IPTG. After 16 h, cells were harvested and SeMet-labelled protein was purified as described above. Mass-spectrometric analysis confirmed full incorporation of SeMet, including the N-terminal methionine (data not shown), in contrast to the native protein. We assume this is a consequence of the incorporation of the selenium making the N-terminal peptide bond less likely to be cleaved.

### 2.2. Crystallization and limited proteolysis

Initial crystallization conditions were obtained by sparse-matrix screening (Jancarik & Kim, 1991) using the sitting-drop vapour-diffusion technique. Drops were prepared by mixing 0.2 µl protein solution (70 mg ml$^{-1}$, 20 m$M$ Tris–HCl pH 7.5, 100 m$M$ NaCl, 10 m$M$ DTT) with 0.2 µl reservoir solution using a Tecan crystallization robot (Tecan UK, Theale, UK) and were equilibrated against 100 µl reservoir solution at 293 K. Initial crystals of $IpaD_{N\Delta14}$ (crystal form 1) grew in 3 d in condition No. 33 of Molecular Dimensions Screen 1 [30%(*w/v*) PEG 4000, 0.1 $M$ Tris–HCl pH 8.5, 0.2 $M$ MgCl$_2$]. Subsequent screening using handmade conditions yielded diffraction-quality crystals in PEG 4000 concentrations ranging from 25 to 28%(*w/v*). Mercury derivatives were prepared by adding 1 µl of 10 m$M$ HgCl$_2$ in reservoir solution to the 0.4 µl drop. Crystals were soaked for 3 h at 293 K.

One $IpaD_{N\Delta14}$ crystal form (crystal form 2) grew over a period of 12 months in 10%(*v/v*) PEG 400, 0.1 $M$ NaCl, 0.1 $M$ Tris–HCl pH 7.0 (Fig. 1a) and was discovered to be an N-terminal truncation of $IpaD_{N\Delta14}$. In an attempt to reproduce this and find new crystal forms, in-drop limited proteolysis was carried out (as in Gaur *et al.*, 2004). The broad-spectrum serine protease subtilisin Carlsberg from *Bacillus licheniformis* (Sigma–Aldrich) was added to $IpaD_{N\Delta14}$ in a 1:5000 enzyme:protein ratio by weight before sparse-matrix screening was carried out as described above. Diffraction-quality native crystals were obtained in Molecular Dimensions Screen 1 condition No. 24 [30%(*v/v*) PEG 400, 0.1 $M$ Na HEPES pH 7.5, 0.2 $M$ MgCl$_2$] (crystal

**Table 1**
IpaD X-ray diffraction data-collection statistics.

Values in parentheses are for the highest resolution shells.

| | CF-1 | | CF-2 | CF-3 | CF-4A | CF-4B SeMet MAD | | | | CF-5 |
|---|---|---|---|---|---|---|---|---|---|---|
| | Native | Mercury soak | Native | Native | Native | Peak 1 | Peak 2 | Peak 3 | Remote | Native |
| X-ray source | ESRF, ID14-3 | ESRF, ID23-1 | ESRF, ID14-1 | ESRF, ID23-2 | ESRF, ID23-1 | ESRF, ID23-1 | | | | ESRF, BM16 |
| Detector | MAR CCD | MAR CCD | ADSC scanner | MAR CCD | ADSC scanner | ADSC scanner | | | | MAR CCD |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $C2$ | $C2$ | $C2$ | $C2$ | | | | $P2_1$ |
| Unit-cell parameters | | | | | | | | | | |
| $a$ (Å) | 55.9 | 56.2 | 77.9 | 204.6 | 139.4 | 136.9 | | | | 44.9 |
| $b$ (Å) | 100.7 | 102.4 | 91.5 | 45.1 | 45.0 | 44.3 | | | | 205.9 |
| $c$ (Å) | 112.0 | 111.5 | 54.9 | 56.8 | 99.5 | 100.0 | | | | 110.9 |
| $\beta$ (°) | | | 96.4 | 105.2 | 107.9 | 107.4 | | | | 102.5 |
| Wavelength (Å) | 0.931 | 1.0055 | 0.934 | 0.873 | 0.9757 | 0.9792 | 0.9792 | 0.9792 | 0.9757 | 0.9790 |
| Resolution limits | 75–2.7 | 38–3.2 | 42–2.0 | 30–3.5 | 43–2.8 | 34–3.4 | 34–3.4 | 34–3.2 | 48–3.3 | 41–3.9 |
| (Å) | (2.8–2.7) | (3.4–3.2) | (2.1–2.0) | (3.7–3.5) | (3.0–2.8) | (3.6–3.4) | (3.6–3.4) | (3.4–3.2) | (3.5–3.3) | (4.1–3.9) |
| Completeness (%) | 90.1 (68.2) | 99.9 (100.0) | 99.2 (99.8) | 96.5 (96.5) | 99.0 (96.3) | 99.2 (96.4) | 99.4 (96.7) | 99.5 (100.0) | 97.8 (87.5) | 86.4 (86.7) |
| Unique reflections | 14594 | 11143 | 24238 | 5840 | 14425 | 7938 | 8009 | 9703 | 8518 | 7495 |
| Multiplicity | 7.0 (3.0) | 7.8 (7.9) | 3.0 (3.0) | 4.6 (4.7) | 5.2 (4.6) | 6.9 (6.6) | 7.0 (6.8) | 3.6 (3.7) | 6.8 (6.2) | 8.8 (8.2) |
| $R_{merge}$† | 14.1 (48.2) | 14.0 (37.2) | 5.3 (41.8) | 13.2 (41.6) | 11.5 (29.2) | 9.6 (33.6) | 8.4 (23.1) | 11.0 (33.5) | 8.9 (25.8) | 22.7 (42.1) |
| $I/\sigma(I)$ | 3.9 (1.6) | 3.8 (1.8) | 7.9 (1.7) | 3.0 (1.5) | 3.1 (1.9) | 5.6 (2.2) | 5.6 (3.0) | 4.5 (2.0) | 5.9 (2.7) | 2.6 (1.7) |
| $R_{anom}$‡ | — | — | — | — | — | 6.7 (15.0) | 5.5 (9.0) | 5.4 (11.4) | 4.0 (10.8) | — |

† $R_{merge} = 100 \times \sum_h [\sum_i |\langle I(h)\rangle - I(h)_i|/\sum_i I(h)_i]$, where $I(h)_i$ is the $i$th observation of reflection $h$ and $\langle I(h)\rangle$ is the mean intensity of all observations of $h$. ‡ $R_{anom} = 100 \times \sum_h |\langle I^+\rangle - \langle I^-\rangle|/\sum_h (\langle I^+\rangle + \langle I^-\rangle)$, where $\langle I^+\rangle$ and $\langle I^-\rangle$ are the mean intensities of the Bijvoet pairs for observation $h$.

form 3) and Molecular Dimensions Screen 2 condition No. 19 [10%(w/v) PEG 8000, 0.1 M Na HEPES pH 7.5, 8%(v/v) ethylene glycol] (crystal form 4A; Fig. 1b). Diffraction-quality SeMet crystals were obtained in Molecular Dimensions Screen 1 condition No. 23 [28%(v/v) PEG 400, 0.1 M Na HEPES pH 7.5, 0.2 M CaCl₂] (crystal form 5) and Molecular Dimensions Screen 1 condition No. 34 [30%(v/v) PEG 400, 0.1 M Tris–HCl pH 8.5, 0.2 M sodium citrate] (crystal form 4B). Numerous other conditions in Molecular Dimensions Screens 1 and 2 yielded crystals that were too small to collect data from.

### 2.3. Data collection and processing

Crystal forms 1, 2, 3, 4A and 5 were cryoprotected in reservoir solution containing 20%(v/v) glycerol, 20%(v/v) PEG 400, 5%(v/v) PEG 400, 25%(v/v) ethylene glycol and 5%(v/v) PEG 400, respectively, and flash-cryocooled in liquid nitrogen prior to data collection. Crystal form 4B was mounted and flash-cryocooled directly in the cryostream.

Diffraction data were recorded at 100 K (Table 1). Data were indexed and integrated in *MOSFLM* (Leslie, 1992) and scaled with *SCALA* (Evans, 1997) within the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). A fluorescence scan on a



**Figure 2**
Harker section $v = 0.5$ of the native Patterson map of IpaD crystal form 1 calculated using *PATTERSON* (Collaborative Computational Project, Number 4, 1994) at 3.5 Å. The map is drawn with a minimum contour level of 1.5σ with 0.5σ increments.

SeMet crystal (CF-4B) across the Se edge yielded values of $f' = -8.0$ and $f'' = 5.2$ using the program *CHOOCH* (Evans & Pettifer, 2001).

### 3. Results and discussion

Intact IpaD$_{N\Delta14}$ yielded plate-shaped crystals (CF-1) belonging to space group $P2_12_12_1$ (Table 1). Calculation of a native Patterson map revealed a peak at ∼50% of the origin on Harker section $v = 0.5$, consistent with two molecules per asymmetric unit related by a pure translation (Fig. 2). The value of the Matthews coefficient is 2.1 Å³ Da⁻¹ for two molecules, corresponding to a solvent content of 42% (Matthews, 1968). Attempts to grow diffraction-quality SeMet crystals in this crystal form were unsuccessful. Soaking the crystals with mercury increased the quality of diffraction of the crystals but did not yield a reliable set of Hg sites, despite using the available Patterson-solving software.

During optimization of CF-1 a second crystal form was observed (CF-2) which took 12 months to grow and belongs to space group $C2$ (Fig. 1). One crystal was used for data collection and the remainder were re-solubilized, run on SDS–PAGE and N-terminal sequenced. This revealed a major band corresponding to an N-terminal truncation beginning at residue 144 (data not shown). The value of the Matthews coefficient is 2.4 Å³ Da⁻¹ for two molecules, corresponding to 49% solvent content.

The contents of CF-2 revealed that the N-terminal third of IpaD is susceptible to degradation and suggested that limited proteolysis could trim these regions. In an attempt to reproduce CF-2 and search for new crystal forms, the broad-spectrum protease subtilisin Carlsberg was added in trace amounts to IpaD$_{N\Delta14}$ before screening. Plate-shaped crystals were grown in three new crystal forms (CF-3, CF-4 and CF-5; Table 1) and analysis by SDS–PAGE revealed various degrees of N-terminal truncation (Fig. 1b). The value of the Matthews coefficient is 2.1 and 2.5 Å³ Da⁻¹ for CF-3 and CF-4, respectively, for two molecules per asymmetric unit and is 2.6 Å³ Da⁻¹ in CF-5 for four molecules.

Anomalous difference Patterson maps for SeMet derivatives of CF-4 were calculated within *autoSHARP* (Vonrhein et al., 2005) using $(E^2 - 1)$ coefficients and strict outlier rejection. The positions of ten Se atoms were obtained using the Patterson-solving software within
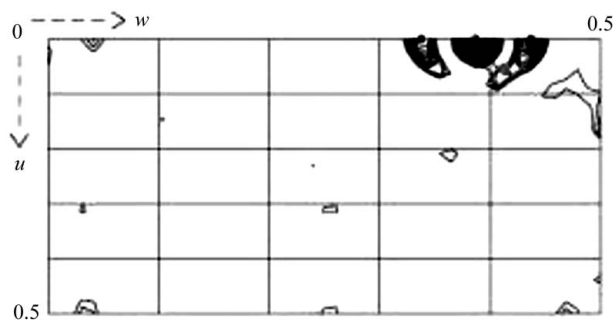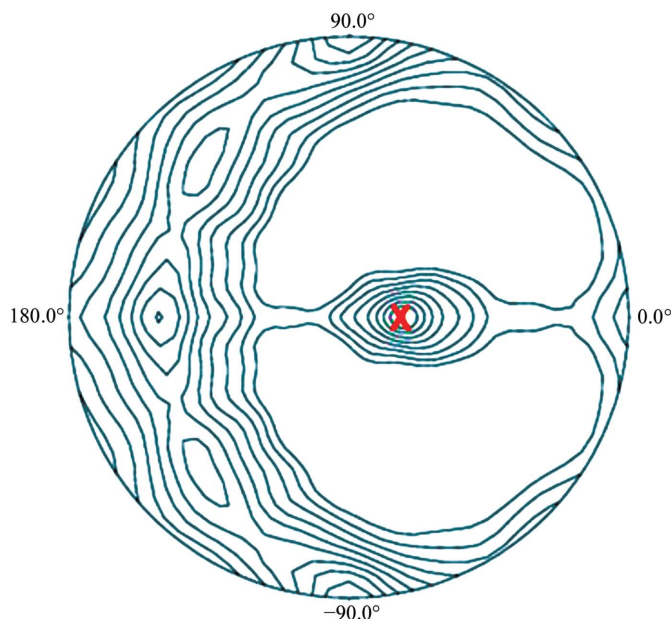
**Figure 3**
The $\kappa = 180°$ section of the self-rotation function calculated for the crystal form 4 SeMet peak 1 data set using *MOLREP* (Collaborative Computational Project, Number 4, 1994) with an integration radius of 20.0 Å and data in the resolution range 15–4 Å. The peak (marked with an X) at $(\omega, \varphi) = (22, 0°)$ represents 91% of the peak for the crystallographic twofold axis.

*autoSHARP* and were refined using *SHARP*. The ten sites clustered into two sets of five related by the twofold peak in the self-rotation function (Fig. 3). Structure determination and refinement in the various crystal forms is under way.

## References

Blocker, A., Gounon, P., Larquet, E., Niebuhr, K., Cabiaux, V., Parsot, C. & Sansonetti, P. (1999). *J. Cell Biol.* **147**, 683–693.

Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P. & Allaoui, A. (2001). *Mol. Microbiol.* **39**, 652–663.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cordes, F. S., Komoriya, K., Larquet, E., Yang, S., Egelman, E. H., Blocker, A. & Lea, S. M. (2003). *J. Biol. Chem.* **278**, 17103–17107.

Deane, J. E., Cordes, F. S., Roversi, P., Johnson, S., Kenjale, R., Picking, W. D., Picking, W. L., Lea, S. M. & Blocker, A. (2006). *Acta Cryst.* F**62**, 302–305.

Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). *Proc. Natl Acad. Sci. USA*, doi:10.1073/pnas.0602689103.

Derewenda, U., Mateja, A., Devedjiev, Y., Routzahn, K. M., Evdokimov, A. G., Derewenda, Z. S. & Waugh, D. S. (2004). *Structure*, **12**, 301–306.

Evans, G. & Pettifer, R. (2001). *J. Appl. Cryst.* **34**, 82–86.

Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.

Gaur, R. K., Kupper, M. B., Fischer, R. & Hoffmann, K. M. (2004). *Acta Cryst.* D**60**, 965–967.

Harrington, A. T., Hearn, P. D., Picking, W. L., Barker, J. R., Wessel, A. & Picking, W. D. (2003). *Infect. Immun.* **71**, 1255–1264.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Johnson, S., Deane, J. E. & Lea, S. M. (2005). *Curr. Opin. Struct. Biol.* **15**, 700–707.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Menard, R., Sansonetti, P. & Parsot, C. (1994). *EMBO J.* **13**, 5293–5302.

Menard, R., Sansonetti, P. & Parsot, C. (1993). *J. Bacteriol.* **175**, 5899–5906.

Mueller, C. A., Broz, P., Muller, S. A., Ringler, P., Erne-Brand, F., Sorg, I., Kuhn, M., Engel, A. & Cornelis, G. R. (2005). *Science*, **310**, 674–676.

Phalipon, A. & Sansonetti, P. J. (2003). *Crit. Rev. Immunol.* **23**, 371–401.

Picking, W. L., Nishioka, H., Hearn, P. D., Baxter, M. A., Harrington, A. T., Blocker, A. & Picking, W. D. (2005). *Infect. Immun.* **73**, 1432–1440.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2005). In *Crystallographic Methods*, edited by S. Doublié. Totowa, NJ, USA: Humana Press.