Structural characterization of a novel subfamily of leucine-rich repeat proteins from the human pathogen Leptospira interrogans.

Isabelle Miras, Frederick Saul, Mireille Nowakowski, Patrick Weber, Ahmed Haouz, William Shepard, Mathieu Picardeau

electronic reprint

BIOLOGICAL
CRYSTALLOGRAPHY

ISSN: 1399-0047
journals.iucr.org/d

Structural characterization of a novel subfamily of leucine-rich repeat proteins from the human pathogen *Leptospira interrogans*

Isabelle Miras, Frederick Saul, Mireille Nowakowski, Patrick Weber, Ahmed Haouz, William Shepard and Mathieu Picardeau

Acta Cryst. (2015). D**71**, 1351–1359

BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Structural characterization of a novel subfamily of leucine-rich repeat proteins from the human pathogen *Leptospira interrogans*

Isabelle Miras,[a] Frederick Saul,[a] Mireille Nowakowski,[b] Patrick Weber,[a] Ahmed Haouz,[a] William Shepard[c]* and Mathieu Picardeau[d]*

[a]Institut Pasteur, Plate-forme de Cristallographie, CNRS–UMR 3528, Paris, France, [b]Institut Pasteur, Plate-forme Protéines Recombinantes, CNRS–UMR 3528, Paris, France, [c]Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin, BP48, Gif-sur-Yvette, France, and [d]Institut Pasteur, Unité de Biologie des Spirochètes, Paris, France. *Correspondence e-mail: william.shepard@synchrotron-soleil.fr, mpicard@pasteur.fr

Received 19 November 2014
Accepted 8 April 2015

Edited by K. Miki, Kyoto University, Japan

Keywords: spirochaetes; *Leptospira*; LRR motif; crystal structure.

PDB references: LIC12234, 4tzh; LIC10831, 4u06; LIC11098, 4u08; LIC12759, 4u09

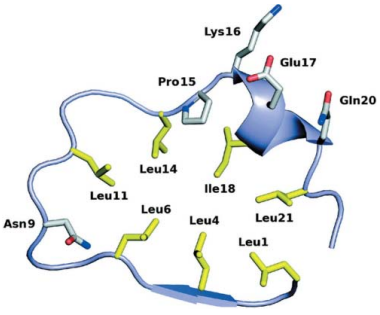
Supporting information: this article has supporting information at journals.iucr.org/d

Pathogenic *Leptospira* spp. are the agents of leptospirosis, an emerging zoonotic disease. Analyses of *Leptospira* genomes have shown that the pathogenic leptospires (but not the saprophytes) possess a large number of genes encoding proteins containing leucine-rich repeat (LRR) domains. In other pathogenic bacteria, proteins with LRR domains have been shown to be involved in mediating host-cell attachment and invasion, but their functions remain unknown in *Leptospira*. To gain insight into the potential function of leptospiral LRR proteins, the crystal structures of four LRR proteins that represent a novel subfamily with consecutive stretches of a 23-amino-acid LRR repeat motif have been solved. The four proteins analyzed adopt the characteristic α/β-solenoid horseshoe fold. The exposed residues of the inner concave surfaces of the solenoid, which constitute a putative functional binding site, are not conserved. The various leptospiral LRR proteins could therefore recognize distinct structural motifs of different host proteins and thus serve separate and complementary functions in the physiology of these bacteria.

1. Introduction

Leptospira are zoonotic pathogens with a worldwide distribution and are the aetiological agents of leptospirosis. The World Health Organization (WHO) estimates that more than one million severe cases of leptospirosis occur annually, with a case fatality rate of approximately 10% (Abela-Ridder *et al.*, 2010). Thus, leptospirosis has emerged as a major public health concern, especially in the developing world (Lau *et al.*, 2010). However, the virulence mechanisms and, more generally, the biology of pathogenic *Leptospira* remain largely unknown, partly owing to a lack of efficient genetic tools for pathogenic *Leptospira* spp. (Ko *et al.*, 2009). Pathogenic *Leptospira* spp. are not considered as intracellular pathogens but they can reside, at least transiently, within host cells (Barocchi *et al.*, 2002; Toma *et al.*, 2011). To remain virulent in the animal infection model, *Leptospira* must be motile (Lambert *et al.*, 2012), adhere to host cells (Adler, 2014), acquire exogenous sources of iron (Murray *et al.*, 2009), resist stress conditions (Eshghi *et al.*, 2012; Lourdault *et al.*, 2011; King *et al.*, 2014) and evade the host immune system (Nally *et al.*, 2005).

Genome analyses have revealed that pathogenic *Leptospira* spp. have an expanded repertoire of genes encoding proteins with leucine-rich repeat (LRR) domains, but the roles of these proteins in *Leptospira* remain unknown. The LRR domain is present in over 6000 proteins currently available in sequence



databases, and has been identified in viral, bacterial, archaeal and eukaryotic organisms (Kobe & Kajava, 2001). Most LRR domains are involved in protein–ligand or protein–protein interactions, and proteins containing these domains are involved in diverse biological processes including signal transduction, cell adhesion and the innate immune response.

Molecular mimicry, including structural mimicry of eukaryotic LRRs, is a common strategy used by pathogens to interfere with the functions of the host to the advantage of the pathogen (Doxey & McConkey, 2013). Various pathogenic bacteria express virulence-associated proteins which contain LRR domains, with the internalins from the intracellular pathogen *Listeria monocytogenes* being the most well studied examples (Bierne *et al.*, 2007). The internalins are outer membrane proteins involved in adhesion and invasion of the host cell (Ireton, 2007). Other bacterial LRR-containing proteins include YopM in *Yersinia pestis* (Leung *et al.*, 1990), IpaH in *Shigella flexneri* (Fernandez-Prada *et al.*, 2000) and SspH and SlrP in *Salmonella* spp. (Miao *et al.*, 1999; Bernal-Bayard *et al.*, 2010), all of which are effector proteins that are directly translocated into eukaryotic cells to alter host signal transduction. The membrane-bound LRR-containing lipoproteins Slr from *Streptococcus* (Bober *et al.*, 2011; Reid *et al.*, 2003) and LrrA from the spirochete *Treponema denticola* have also been shown to play a role in cell adhesion (Ikegami *et al.*, 2004). Structurally, LRR domains consist of tandems of two or more repeat units forming a curved α/β-solenoid structure (Kobe & Kajava, 2001; Bella *et al.*, 2008). The overall topology of LRR domains depends on the sequence and the number of repeat units. To our knowledge, the length of the LRR repeat can vary from a short 20-residue motif in *Yersinia* YopM (Leung *et al.*, 1990) to the longer 28–29-residue repeat of the eukaryotic ribonuclease inhibitor (Kobe & Kajava, 2001).

The expression, localization, structure and function of leptospiral LRR-containing proteins have not been explored experimentally. We have initiated the characterization of these pathogen-specific *Leptospira* spp. proteins by solving the crystal structures of four representative LRR proteins from *L. interrogans*. Our results offer insight into a family of pathogen-specific proteins that may play key roles in host–pathogen interactions.

2. Materials and methods

2.1. Cloning, expression and purification of LRR proteins

Constructs deprived of the N-terminal signal peptide were designed to express the recombinant LRR protein as a soluble secreted protein. The LRR ORF portions coding for residues 24–191 in LIC12234, 30–377 in LIC10831, 30–426 in LIC11098 and 24–423 in LIC12759 were amplified by PCR from genomic DNA of strain Fiocruz L1-130 and cloned into the *Escherichia coli* expression vector pDEST17 (Gateway technology, Invitrogen) with the insertion of an N-terminal TEV cleavage site. For LIC10831, which lacks methionine, we designed a single-mutant variant LIC10831-T266M using a structural model generated with the *I-TASSER* server (Zhang, 2008; Roy *et al.*, 2010, 2012). The substitution of the Thr266 codon with that for methionine was performed using the QuikChange multisite-directed mutagenesis kit (Agilent Technologies) according to the manufacturer's instructions.

Protein expression was performed in *E. coli* BL21 (DE3) pLysS (Novagen) cells transformed with the plasmid pDEST-rLRR and grown at 30°C in high-density medium (HDM) with a microfermentor unit according to established protocols (Frachon *et al.*, 2006). Overexpression was achieved by induction with 1 m*M* IPTG at 14°C for 15 h; the cells were then harvested by centrifugation, washed with PBS (140 m*M* NaCl, 2.7 m*M* KCl, 10 m*M* Na_2HPO_4, 1.8 m*M* KH_2PO_4 pH 7.3) and frozen at −80°C. Selenomethionine (SeMet)-labelled proteins were produced in *E. coli* B834 (DE3) cells (Novagen) grown in flasks with DLM medium (Bernard & Payton, 2001) supplemented with 200 mg l^{-1} L-selenomethionine (Calbiochem), and overexpression was induced as above. The same purification protocol was followed for both the native and the selenomethionine-labelled proteins. Frozen cells were thawed, resuspended in lysis buffer (100 m*M* Tris–HCl, 200 m*M* NaCl pH 8.0) and lysed using a French press. The lysate obtained from 3 g (wet weight) of frozen cells was centrifuged at 26 800*g* for 1 h, filtered on a 0.45 µm membrane and loaded onto a 1 ml HisTrap Ni^{2+}–IMAC column (GE Healthcare). The recombinant LRR proteins were eluted by applying a 0–500 m*M* imidazole gradient in the same buffer. Fractions containing the recombinant protein, as confirmed by 4–12% SDS–PAGE, were pooled, centrifuged at 6000*g* for 10 min to eliminate precipitated material and dialysed overnight at 18°C against a buffer consisting of 100 m*M* Tris–HCl pH 7.5, 200 m*M* NaCl, 1 m*M* DTT. The His_6 tag was removed *via* proteolysis by incubating the sample overnight at 4°C with recombinant His_6-tagged TEV protease in a 30:1(*w*:*w*) ratio. The mixture was passed by gravity flow through 0.7 ml Ni–NTA resin (Qiagen) in order to eliminate the TEV protease and the cleaved His_6 expression tag. The resulting elute was concentrated and injected onto a HiLoad 16/60 Superdex 75 size-exclusion column (GE Healthcare) equilibrated in the same buffer at a flow rate of 1 ml min^{-1}. Peak fractions containing the LRR protein were pooled and concentrated with a 5 kDa cutoff V concentrator (Corning).

2.2. Crystallization

Preliminary crystallization screens were carried out at 18°C by the sitting-drop vapour-diffusion method using a Mosquito (TTP Labtech) nanolitre dispensing system. Manual optimization was performed by the hanging-drop method in 24-well plates by mixing 2 µl protein solution and 2 µl reservoir solution and equilibrating against 1 ml reservoir solution. The optimized conditions for crystal growth were as follows: native LIC12234 (10.5 mg ml^{-1}), 17.5%(*w*/*v*) PEG 8K, 200 m*M* zinc acetate, 100 m*M* cacodylate pH 6.5; SeMet-labelled LIC10831-T226M variant (26 mg ml^{-1}), 20%(*w*/*v*) PEG 4K, 10%(*v*/*v*) 2-propanol, 10 m*M* zinc acetate, 100 m*M* Na HEPES pH 7.5; SeMet-labelled LIC11098 (25 mg ml^{-1}), 14%(*w*/*v*) PEG 8K,

Table 1
Crystallographic parameters, data and refinement statistics.

Values in parentheses are for the highest resolution shell.

	LIC12234	LIC10831	LIC11098	LIC12759
Crystal parameters				
Space group	$P2_1$	$P4_22_12$	$P3_221$	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 59.5, b = 65.6, c = 59.9, β = 116.86	a = b = 122.1, c = 58.2	a = b = 105.7, c = 161.7	a = 75.2, b = 104.8, c = 116.6
Data statistics				
Resolution range (Å)	41.4–1.39 (1.46–1.39)	48.2–1.90 (2.00–1.90)	46.5–1.95 (2.06–1.95)	47.8–1.95 (2.06–1.95)
No. of unique reflections	78102 (9640)	35305 (4989)	76748 (11079)	67851 (9753)
Multiplicity	3.1 (3.1)	6.5 (5.3)	8.3 (8.5)	10.9 (11.0)
R_{merge}	0.066 (0.604)	0.053 (0.627)	0.096 (0.962)	0.097 (0.974)
$R_{p.i.m.}$	0.045 (0.407)	0.032 (0.431)	0.036 (0.351)	0.045 (0.449)
Completeness (%)	94.1 (80.0)	99.7 (98.3)	100.0 (100.0)	100.0 (100.0)
$\langle I/\sigma(I)\rangle$	9.1 (2.0)	16.1 (2.1)	11.8 (2.3)	13.6 (2.4)
Wilson plot B factor (Å^2)	20.9	34.2	37.7	32.1
Refinement				
Resolution (Å)	32.8–1.39 (1.43–1.39)	48.3–1.90 (1.96–1.90)	46.45–1.95 (2.00–1.95)	47.8–1.95 (2.00–1.95)
No. of reflections	78067 (4347)	34851 (2778)	76690 (5608)	67774 (4919)
R factor, working set	0.181 (0.272)	0.201 (0.415)	0.208 (0.275)	0.187 (0.252)
R_{free}	0.206 (0.296)	0.236 (0.419)	0.238 (0.272)	0.213 (0.270)
Non-H atoms				
Protein	3157	2843	6326	6379
Waters	453	364	488	609
Zn^{2+} ions	8	4	9	6
Ca^{2+} ions	—	—	1	—
R.m.s. deviations from ideal				
Bond lengths (Å)	0.010	0.010	0.010	0.010
Bond angles (°)	1.16	1.12	1.14	1.15
Ramachandran plot (%)				
Preferred regions	94.5	93.5	93.6	94.0
Allowed regions	5.2	5.9	5.9	5.6
Outliers	0.3	0.6	0.5	0.4

200 m*M* calcium acetate, 10 m*M* zinc acetate, 100 m*M* cacodylate pH 6.5; SeMet-labelled LIC12759 (20.8 mg ml^{-1}), 25%(*w*/*v*) PEG 4K, 125 m*M* zinc acetate, 200 m*M* magnesium chloride, 100 m*M* MES pH 6.5. For X-ray data collection, the crystals were flash-cooled in liquid nitrogen using Paratone/paraffin oil (50/50%) as a cryoprotectant, except for LIC11098 where the cryoprotectant was made up of the crystallization solution mixed with 25%(*v*/*v*) glycerol.

2.3. Data collection, structure determination and refinement

X-ray diffraction data were collected on beamline PROXIMA1 at Synchrotron SOLEIL (Saint Aubin, France). The data were processed with *XDS* (Kabsch, 2010) and scaled with either *XSCALE* from the *XDS* package or with *SCALA* from the *CCP*4 program suite (Winn *et al.*, 2011). The structures of LIC12234, LIC10831 and LIC12759 were solved by single- or multi-wavelength anomalous diffraction (SAD or MAD) experiments with data collected at an energy just above the Zn *K* edge (9.6586 keV, λ = 1.2837 Å) as determined by X-ray energy scans. Zinc sites in the crystal structure were located with *SHELXD* (Sheldrick, 2008), and SAD phasing was performed with *SHELXE* and *phenix.autosol* (Terwilliger *et al.*, 2009). This procedure led to interpretable electron-density maps that could be automatically traced with *ARP*/*wARP* (Langer *et al.*, 2008). The structure of LIC11098 was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the structure of LIC12759 as a search template. The models were improved through iterative cycles of manual adjustment and model building with *Coot* (Emsley *et al.*, 2010) and refinement with *BUSTER* (Smart *et al.*, 2012). The final models were validated using the *MolProbity* server (http://molprobity.biochem.duke.edu; Chen *et al.*, 2010). All structural figures were generated with *PyMOL* (v.1.5; Schrödinger). The crystallographic parameters, data statistics and refinement statistics are shown in Table 1. Coordinates and structure factors of LIC12234, LIC10831, LIC11098 and LIC12759 have been deposited in the Protein Data Bank with accession codes 4tzh, 4u06, 4u08, and 4u09, respectively.

3. Results and discussion

3.1. Genomic and proteomic analysis of LRR-containing proteins in *Leptospira* spp.

Genomic analysis revealed that the pathogen *L. interrogans* strain Fiocruz L1-130 contains 19 LRR-containing proteins, while the strict pathogen *L. borgpetersenii* strain L550 has five and the nonpathogenic *L. biflexa* strain Patoc contains only one LRR protein-encoding gene (Supplementary Table S1). Half of the LRR protein-encoding genes in *L. interrogans* are clustered. For example, the LRR protein-encoding genes LIC10828, LIC10829, LIC10830 and LIC10831 are co-localized, suggesting DNA rearrangement and/or recombination events probably favoured by the repetition of homologous sequences. The number of LRR protein-encoding genes in

pathogenic *Leptospira* spp. greatly exceeds the number of such genes in almost all other pathogenic bacteria, with the exception of the family of internalins in *Listeria monocytogenes* (Bierne *et al.*, 2007). A bioinformatics analysis of the cellular localization (*PSORTb* server: http:/www.psort.org/psortb/) suggests that 17 out of 19 LRR-containing proteins from *L. interrogans* strain Fiocruz L1-130 are extracellular. Their corresponding molecular weight varies from 14 to 185 kDa depending on the number of LRR repeats and the

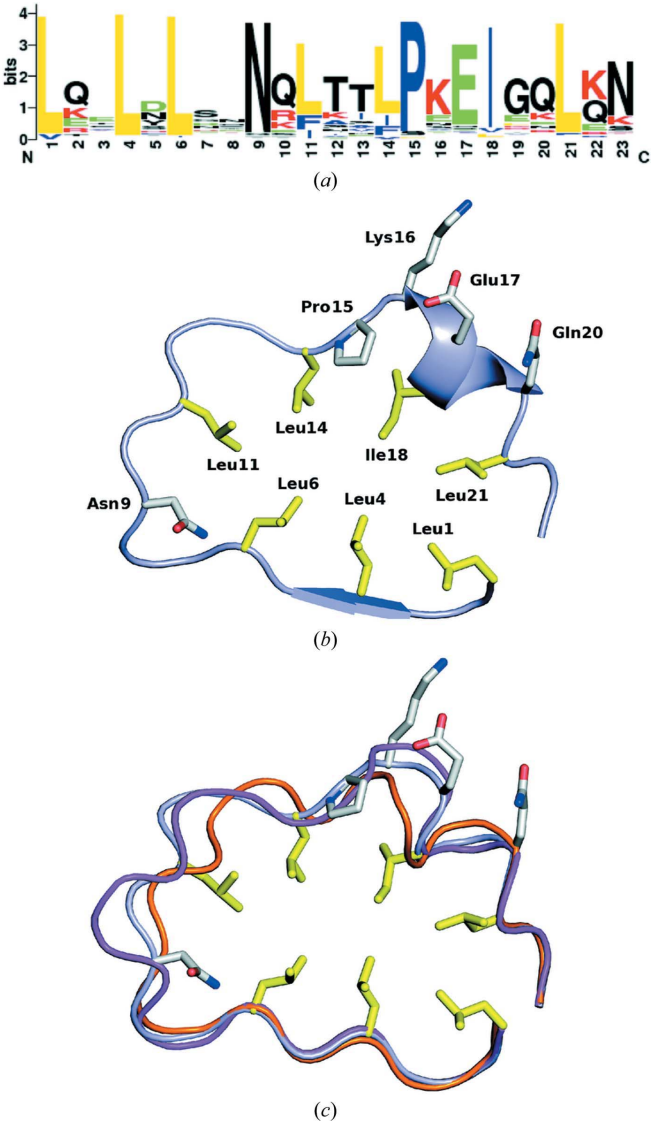


Figure 1
Primary- and secondary-structural analysis of the consensus sequence of the 23-amino-acid LRR repeat units in *L. interrogans*. (*a*) A graphical sequence diagram, generated with *WebLogo* (Crooks *et al.*, 2004), representing 148 distinct LRR units from *L. interrogans* serovar Copenhageni strain Fiocruz L1-130. The degree of sequence conservation at each position is indicated by the height of the letters. (*b*) Secondary-structure topology of the canonical 23-amino-acid repeat of leptospiral LRR proteins. Side chains of the buried hydrophobic residues are shown in yellow. Other significantly conserved residues in the consensus sequences are shown in grey. (*c*) Superposition of the leptospiral LRR repeat unit (23 amino acids, cyan) with InlB (22 amino acids, orange) and XopL (24 amino acids, magenta).

Table 2
Consensus sequences of LRR classes.

The LRR repeat can be divided into a highly conserved segment (HCS), which usually consists of the 11-residue sequence L*xx*L*x*L*xx*N*x*L, followed by a variable segment (VS) (Bella *et al.*, 2008; Kobe & Kajava, 2001). This table is modified from Wei *et al.* (2008). –, possible insertion sites; ◇, nonpolar residue; *x*, any residue. Conserved residues are shown in upper case (more than 50% of sequences) or lower case (more than 30% of sequences).

Class	HCS	VS
Typical type	L*xx*L*x*L*xx*N*x*L	*xx*Lp*xx*◇F*x*–*x*L*xx*
Bacterial	L*xx*L*x*V*xx*N*x*L	*xx*LP*x*e/dL–P*xx*
Shigella IpaH	L*xx*L*x*L*xx*N*x*L	*xx*LP*xx*LP*xx*
Yersinia YopM	L*xx*L*x*A/V*xx*N*x*L	*xx*LPD/ELPP*x*
Listeria InlA	L*xx*L*x*L*xx*N*x*I	*x*DL*xx*L*xx*L*x*
Salmonella SspH2	L*xx*L*x*L*xxxx*L	*xx*LP*x*LP*xx*
Leptospira interrogans	**L*xx*L*x*L*xx*N*x*L**	***xx*LP*x*EI*xx*L*xn***
Ribonuclease inhibitor-like	L*xx*L*x*L*xx*N/C*x*L	*xxx*g◇*xx*L*xxx*L*x*–*xxxx*
SDS22-like	L*xx*L*x*L*xx*N*x*I	*xx*I*xx*L*x*–*x*L*xx*
Cysteine-containing	L*xx*L*x*L*xx*c*x*–*x*	ITD*xx*◇*xx*La*x*–*x*c*xx*
Plant-specific	L*xx*L*x*L*xx*N*x*L	t/sg–*x*IP*xx*LG*x*L*x*–*x*
Treponema pallidum	L*xx*I*x*L*x*–*xx*L*x*	*x*Ig*xx*AF*xx*C/N*x*–*x*

presence or absence of other domains (Supplementary Table S1). The leptospiral LRR domains consist of 2–20 repeat units as predicted by the *LRRfinder* server (http://www.lrrfinder.com), and a majority of the LRR proteins lack any other identifiable functional domain (with the exception of three proteins which contain a WGR domain of unknown function; Supplementary Table S1).

Primary sequence analysis of 148 repeats present in LRR proteins from *L. interrogans* strain Fiocruz L1-130 reveals a sequence identity of more than 60% between individual repeats across all LRR proteins. The individual repeat units are composed of 23 amino-acid residues, with a consensus sequence L*xx*L*x*L*xx*N*x*L*xx*LP*x*EI*xx*L*x*N (where *x* can be any amino acid; Fig. 1*a*). Each repeat unit can be divided into a highly conserved segment (HCS), which consists of an 11-amino-acid sequence, followed by a variable segment (VS) of 12 amino acids. Seven classes of LRRs have been described according to the sequence-repeat motif (Table 2) and the presence of functional non-LRR domains. The leptospiral LRR repeat sequence resembles the consensus sequence L*xx*L*x*L*xx*N*x*L*xx*LP*x*(*x*)LP*xx* characterizing the typical bacterial-type LRR (Table 2; Kobe & Kajava, 2001; Wei *et al.*, 2008) and represents a novel subfamily of LRRs with highly conserved consecutive stretches of a 23-amino-acid LRR motif. This organization is analogous to the stretches of 22-residue LRRs in internalins and 20-residue LRRs in YopM. In order to characterize the structural topology adopted by this subfamily, we have determined the crystal structures of four representative leptospiral LRR proteins.

3.2. Structural analysis of *Leptospira* LRR proteins

Our initial strategy to solve the structures was to label the LRR proteins with selenomethionine for SAD/MAD phasing experiments. However, the proteins studied here contain very few or no methionine residues. For LIC10831, which is devoid of methionines, we mutated Thr266 to methionine and then substituted in selenomethionine. For LIC12234, the X-ray

diffraction data of the native crystals displayed a significant anomalous signal. As these crystals grew in the presence of zinc acetate, we suspected that these proteins might bind zinc ions. X-ray diffraction data collected at the Se *K* edge from crystals of SeMet-labelled LIC10831-T266M revealed several additional peaks in the heavy-atom substructure. X-ray fluorescence emission spectra of the crystal confirmed the presence of zinc, which was a component of the crystallization solution. X-ray diffraction data collected below and above the Zn *K* edge established unequivocally that the additional heavy-atom sites are zinc ions bound at varying occupancies. SeMet-LIC12759, which contains two methionine residues (Met199 and Met231), was also crystallized in the presence of zinc to increase the weak anomalous signal, and this allowed us to improve the SAD-phased maps. The fourth crystal structure (SeMet-LIC11098) could be solved by molecular-replacement methods (Table 1).

The crystal structures confirm that the consensus sequence corresponds to the structural repeat. The secondary structure of each repeat coil contains a short β-strand and a short 3_{10}-helix connected by loops and turns (Fig. 1*b*). This topology closely resembles the 22-amino-acid repeat unit in the internalins (Neves *et al.*, 2013) and the 24-amino-acid repeat in the LRR domain of the type III effector (XopL) from *Xanthomonas campestris* (PDB entry 4fcg; Midwest Center for Structural Genomics, unpublished work; Fig. 1*c*). The highly conserved segment (HCS), which forms the inner concave face of the molecule, starts at the end of a β-turn from the previous repeat unit, followed by a short β-strand (L*x*L*xx*) and a β-turn where the carbonyl O atom of the sixth amino-acid residue makes a conserved hydrogen bond to the asparagine side chain of the ninth residue (L*xx*N; Fig. 1*b*). The variable segments (VS), which are located on the outer convex face of the molecule, start with a short polyproline II-like stretch of four residues (*xx*LP) preceding a short 3_{10}-helix (*x*EIx) and a final β-turn (*x*L*x*N). The side chains of the highly conserved (>50%) leucine, isoleucine and asparagine residues of the 23-amino-acid repeat are buried in the interior of the coil (Fig. 1*b*). When compared with the 22-amino-acid repeat motif of internalin B (PDB entry 1d0b; Marino *et al.*, 1999), the leptospiral 23-amino-acid repeat shows an additional proline residue at position 15 (Fig. 1*c*), and comparison with the 24-amino-acid XopL repeat unit (PDB entry 4fcg) reveals a deletion between positions 9 and 10. Despite these differences, the conserved leucine residues buried in the interior of the coil display similar orientations and packing to maintain the overall topology of the repeat coil (Fig. 1*c*).

An overall structural analysis demonstrated that these proteins fold into single domains which display the canonical horseshoe shape of curved α/β-solenoids characteristic of LRR domains (Fig. 2). The four leptospiral LRRs in this study have a similar radius of curvature, 27–30 Å as defined by Enkhbayar *et al.* (2004), despite the different number of repeat units (LIC11098 and LIC12759 with 17 tandem repeats each and LIC10831 and LIC12234 with 13 and eight repeats, respectively). Each repeat unit forms hydrogen bonds to adjacent repeats, creating a conserved hydrogen-bonding network which stabilizes the α/β-solenoid structure

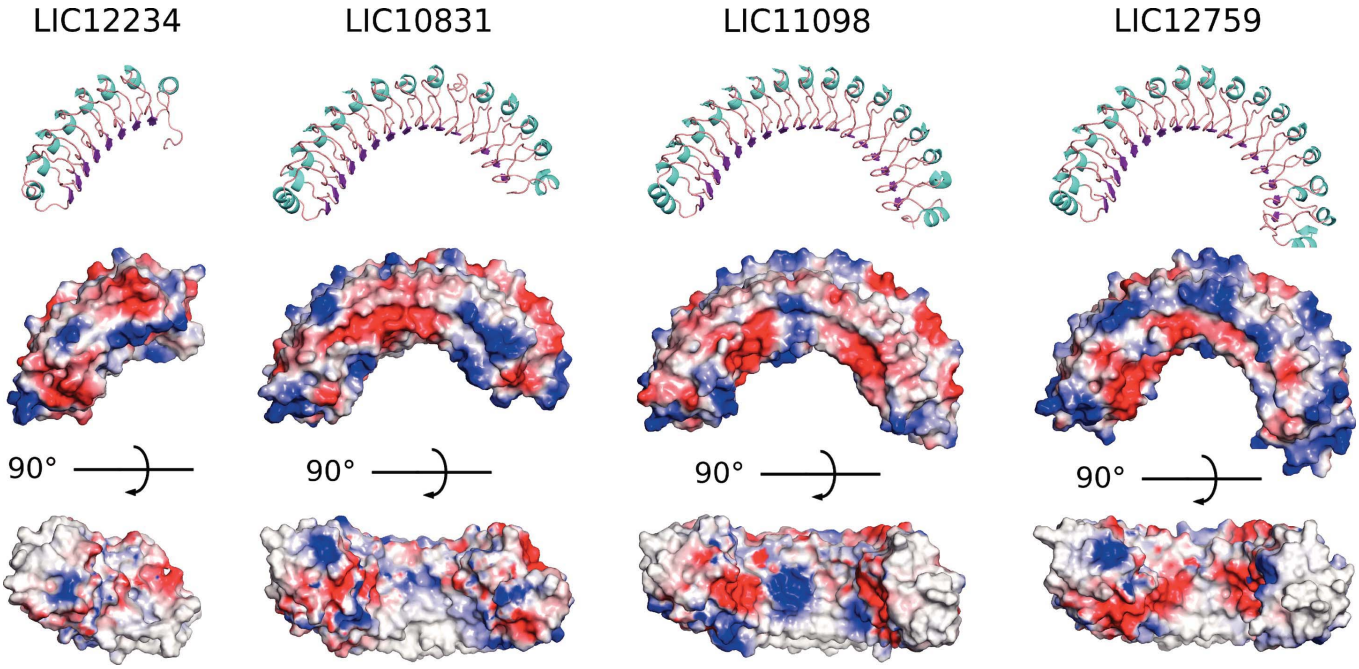


Figure 2
Structural characterization of leptospiral LRR proteins. Overall structures of the leptospiral LRR proteins LIC11098, LIC12759, LIC12234 and LIC10831. The concave inner surfaces defined by a parallel β-sheet are shown in magenta and the 3_{10}-helices forming the outer convex surface are shown in cyan. The electrostatic charge properties of the proteins are indicated (positive and negative charges are indicated in blue and red, respectively). The 90° rotation along a horizontal axis shows the distribution of charge on the inner concave surface.

Table 3
Concave surface residue identities and similarities.

Top diagonal elements are percentages of strictly identical residues. The bottom diagonal contains percentages of identical and similar residues: aliphatic (Ala, Ile, Leu, Met, Pro, Val), aromatic (Phe, Trp, Tyr), polar (Cys, Asp, Glu, Asn, Gln, Ser, Thr), negatively charged (Asp, Glu) and positively charged (His, Lys, Arg). LRR proteins: leptospiral LRR (LIC12234, LIC10831, LIC11098 and LIC12759), *L. monocytogenes* internalin A (InlA) and B (InlB), thyroid-stimulating hormone receptor (TSH-R) and *Yersinia pestis* YopM

	LIC12234	LIC10831	LIC11098	LIC12759	InlA	InlB	TSH-R	YopM
LIC12234	—	15.8	15.8	15.8	15.8	10.5	10.5	15.8
LIC10831	39.5	—	21.4	20.0	12.9	12.5	16.3	14.2
LIC11098	31.5	31.4	—	17.8	11.3	22.5	14.3	12.7
LIC12759	34.2	37.1	24.4	—	11.3	10.0	14.3	12.7
InlA	29.0	40.0	31.3	35.0	—	15.0	16.3	15.2
InlB	23.7	35.0	37.5	35.0	27.5	—	5.0	12.5
TSH-R	28.9	34.7	24.5	34.7	32.6	27.5	—	14.3
YopM	26.3	32.8	34.7	27.8	29.1	32.5	26.5	—

(Supplementary Fig. S1). Many of these bonds are formed between atoms of the polypeptide chain backbone in the parallel β-sheet lining the inner concave surface of the LRR domain. These parallel β-strands are aligned and form the framework for the curved and twisted α/β-solenoid architecture (Fig. 2). On the outer convex face of the molecule, there are practically no hydrogen bonds linking the 3_{10}-helical coils between repeat units. Instead, an extensive network of water-mediated interactions interlace the backbone atoms between the repeats (Supplementary Fig. S1).

All LRR-containing proteins have N- and C-terminal capping structures which shield the hydrophobic core of the α/β-solenoid (Bella *et al.*, 2008). In the leptospiral LRR proteins these caps are minimal, unlike the large subdomains present in the internalins (Schubert *et al.*, 2001). The leptospiral LRR caps fold into compact coils and are well conserved (Fig. 3). The N-terminal cap resembles a modified variable segment (VS) preceding the β-strand of the first consensus LRR repeat unit (Fig. 3). The cap contains a short amphiphilic α-helix with two conserved hydrophobic residues (leucine/isoleucine), which interact with the hydrophobic core of the α/β solenoid, and two exposed hydrophilic residues. The α-helix is immediately followed by a conserved proline residue which reverses the direction of the polypeptide chain, leading into the first LRR repeat. The C-terminal cap resembles a

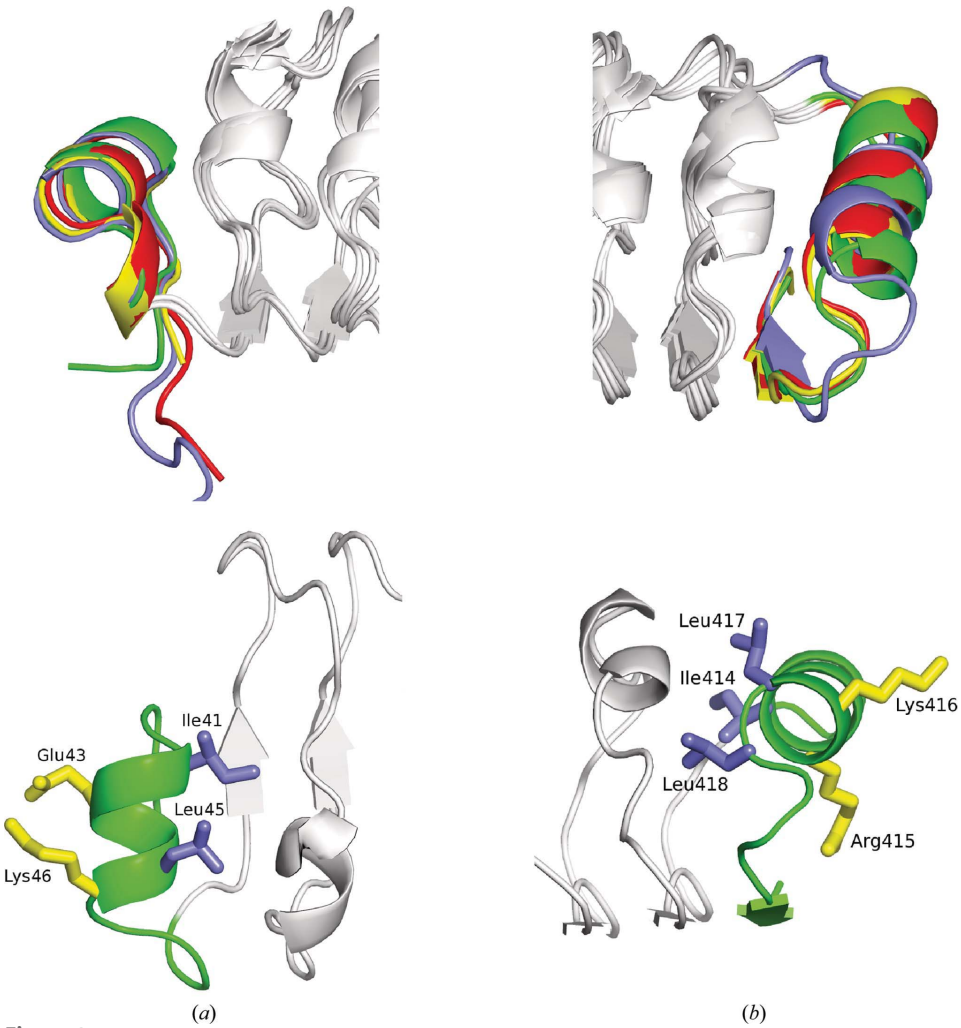


Figure 3
Structure of the (*a*) N-terminal and (*b*) C-terminal caps of the leptospiral LRR proteins. Top, superposition of the N-terminal (*a*) and C-terminal (*b*) regions of the four leptospiral LRR proteins shows the overall similarity of the capping structures (LIC11098 is shown in red, LIC12759 in yellow, LIC12234 in blue and LIC10831 in green). Bottom, selected residues of the N-terminal (*a*) and C-terminal (*b*) helices closing the α/β-solenoid in LIC11098 are shown in blue (hydrophobic) and yellow (hydrophilic); the amphiphilic character of these capping helices is strictly conserved in the four leptospiral proteins.

modified LRR containing an α-helix at the convex side and an additional strand closing the β-sheet (Fig. 3). The cap folds into a 24- or 26-amino-acid coil immediately following the last LRR repeat and contains a β-loop much like the other LRR repeats, followed by an α-helix of 2–3 turns, a β-turn and a short β-strand. The ensemble contains 5–6 well conserved hydrophobic residues which pack against the inner core of the α/β-solenoid. Three conserved hydrophobic residues are located on the buried side of the helix. The exposed face of the helix contains 4–6 charged residues (mostly arginines and lysines; Fig. 3). The C-terminal β-strand interacts with the last LRR repeat, closing the β-sheet of the inner concave surface. Overall, the C-terminal caps resemble those found in SspH1 and IpaH3 (PDB entries 4nkh and 3cvr, respectively; Keszei *et al.*, 2014; Zhu *et al.*, 2008), albeit with a shorter α-helix. The LRR capping structures have the same structural topology in the four leptospiral proteins, suggesting that they are necessary for the stability of the α/β-solenoid conformation or could act as common specific binding sites.

3.3. Clues about the binding site and protein partners

The internalins of *L. monocytogenes* are the most well documented bacterial LRR-containing proteins which exploit mammalian cell components to induce host-cell entry (Bierne *et al.*, 2007). At least three *Listeria* LRR proteins, internalin A (InlA), internalin B (InlB) and internalin C (InlC), are involved in this process by interacting with the cell-surface receptor of human E-cadherin (hEC1; Schubert *et al.*, 2002), the hepatocyte growth factor (Met; Shen *et al.*, 2000) and the sixth Src homology (SH3-6) domain of cytosolic Tuba (Polle *et al.*, 2014), respectively. Superposition of the crystal structure of LIC10831 with the InlA–hEC1 complex (PDB entry 1o6s; Schubert *et al.*, 2002) shows a remarkable similarity in the overall topology of the two LRR proteins even though the repeat unit of the internalins is smaller (22 amino acids) compared with the leptospiral LRRs (Fig. 4). The radius of the inner concave surface of LIC10831 differs only slightly from that of InlA, and consequently if the binding mode is the same we would expect LIC10831 to bind to a host protein or domain of comparable size to the binding partner of InlA (the N-terminal domain of hEC1). Of the three other leptospiral LRRs, two (LIC11098 and LIC12759) form longer arcs than InlA but maintain similar radii of curvature, while one (LIC12234) is much shorter and is of a similar length to InlB.

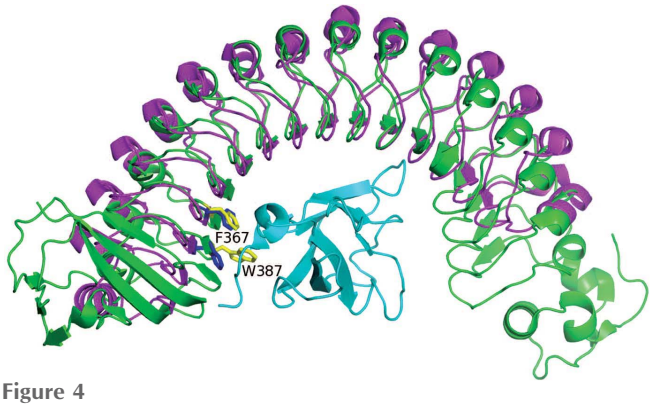


Figure 4
Superposition of LIC10831 with the internalin A–human E-cadherin complex (PDB entry 1o6s). The N-terminal domain of E-cadherin (hEC1, shown in cyan) binds to the inner concave surface of InlA (green) (Schubert *et al.*, 2002). The radius of curvature of the two LRR proteins is similar. Superposition of LIC10831 with InlA gives an r.m.s. deviation of 2.2 Å for 200 C^{α} matches, suggesting that LIC10831 (magenta) would bind to a host protein of comparable size to the E-cadherin N-terminal domain. The two residues of InlA directly involved in binding to hEC1 which are conserved in LIC10831 (Phe305 and Trp326) are indicated in yellow and blue, respectively.

The LRR-containing proteins provide a structural framework for specific protein–protein interactions, particularly the concave face, which constitutes a putative functional binding site. The electrostatic potential of the concave face has been mapped in the four leptospiral LRR proteins (Fig. 2); the distribution of negative and positive charges are well dispersed over the concave surface, suggesting that these proteins can bind different partners. To support this hypothesis, a bioinformatic analysis was used to compare amino-acid residues on the exterior of the concave surface. A description of the alignment algorithm and an example of the results are illustrated in Supplementary Fig. S2. Table 3 summarizes the results of the alignments. The range of identity and similarity matches between the eight LRR-containing proteins compared (LIC12234, LIC10831, LIC11098, LIC12759, InlA, InlB, TSH-R and YopM) are of the order of 5–23% and 22–40%, respectively. Many of these matches occur at the bottom edges of the concave surfaces, where the bacterial LRR proteins have a Q/E ladder. Other matches are well dispersed over the rest of the concave surface. Within the subset of comparisons between leptospiral LRRs (identical and similar matches of 16–22% and 24–40%, respectively), few matches occur in the central lanes of the concave surface. A search for surface matches among four nonleptospiral LRR-containing proteins (InlA, InlB, TSH-R and YopM) reveals identity and similarity matches ranging from 5–17% and 26–33%, respectively. It should be noted that these proteins have different physiological functions, and when they are compared with the leptospiral LRR proteins the level of identical and similar matches is roughly equivalent (10–23 and 24–40%, respectively; Table 3). These levels of identity and similarity suggest that the concave surfaces of this small set of proteins do not resemble each other, and thus the binding partners of these proteins may be different. For example, only nine residues of the InlA concave surface were found to be identical to those of LIC10831. Of these, some are directly involved in cadherin–hEC1 binding, including Phe367 and Trp387 (internalin numbering; Fig. 4), while most are asparagine and glutamine residues on the periphery of the surface. A certain amount of caution, however, must be employed when interpreting such comparative results, as the binding affinities of LRR domains are often weak, mediated by water molecules, and may be dependent upon the presence of certain ions. The disassociation constants (K_d) for InlA are in the micromolar range (albeit maintaining high selectivity) and the LRR–host binding interfaces contain a large number of buried solvent molecules as well as divalent cations. As a case in point,

Wollert and coworkers have shown that structure-guided site-specific mutations on the concave surface of InlA are capable of altering the K_d value by as much as four orders of magnitude (Wollert, Heinz *et al.*, 2007). Indeed, a single point mutation in E-cadherin determines the selectivity between mouse and human proteins (Wollert, Pasche *et al.*, 2007; Lecuit *et al.*, 1999). Consequently, any structure-based prediction of host protein binding would require biochemical validation.

The presence of zinc ions in the LRR structures suggests that certain divalent metal ions may be critical for binding to the host protein. In the four leptospiral LRR crystal structures the concave surfaces of the proteins interact with neighbouring molecules, forming a dimeric or tetrameric quaternary organization (Supplementary Fig. S3). Dynamic light-scattering experiments (Supplementary Table S2) indicate that these proteins are monomeric in solution, and the presence of zinc induces the dimerization of LIC12234. Inspection of the dimer interface in the crystal structures reveals numerous hydrophobic interactions and the presence of zinc- or calcium-binding sites between the protein monomers (Supplementary Figs. S3 and S4). The majority of these zinc sites are four-coordinate with tetrahedral geometry. The zinc ions bind to the side chains of histidine, aspartic acid and/or glutamic acid, as well as chlorides and/or water (Supplementary Fig. S4). Other LRR-containing proteins are also known to bind metal ions (typically calcium, but also zinc) either alone or in complex with other proteins. In the internalin A–E-cadherin complex, for example, a calcium ion is found at the binding interface between the two proteins (Schubert *et al.*, 2002), and the presence of such divalent metal ions may be a prerequisite for the formation of the complex between the LRR domain and the host target protein. However, for the structures presented here it is difficult to ascertain whether the zinc-binding sites are crystallization artifacts or whether they have proper physiological or biochemical functions.

In conclusion, the relatively large number of LRR-encoding genes in pathogenic *Leptospira*, together with their requirement for virulence in other Gram-negative bacteria, prompted us to structurally characterize these proteins, which are recognized by the sera of human leptospirosis patients (data to be published elsewhere). Our aim in the future is to study the secretome of pathogenic *Leptospira*, focusing on LRR proteins and their interactions with the host. Such studies are essential to further our understanding of spirochaete biology, which remains largely unexplored. Ultimately, structural and functional knowledge of these proteins will be useful for the development of potential vaccine candidates and diagnostic tools.

4. Related literature

The following references are cited in the Supporting Information for this article: Evdokimov *et al.* (2001), Helft *et al.* (2011), McEwan *et al.* (2006) and Scott *et al.* (2004).

Acknowledgements

We acknowledge access to the PROXIMA1 beamline at SOLEIL and the support from its staff. We thank Jacques Bellalou for assistance in protein-production optimization and Vincent Bondet for protein purification at an early stage of the work (PFPR, Institut Pasteur). This work was supported by the Institut Pasteur and the French Ministry of Research (ANR-08-MIE-018). The authors declare no conflicts of interest.

References

Abela-Ridder, B., Sikkema, R. & Hartskeerl, R. A. (2010). *Int. J. Antimicrob. Agents*, **36**, S5–S7.
Adler, B. (2014). *Vet. Microbiol.* **172**, 353–358.
Barocchi, M. A., Ko, A. I., Reis, M. G., McDonald, K. L. & Riley, L. W. (2002). *Infect. Immun.* **70**, 6926–6932.
Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. (2008). *Cell. Mol. Life Sci.* **65**, 2307–2333.
Bernal-Bayard, J., Cardenal-Muñoz, E. & Ramos-Morales, F. (2010). *J. Biol. Chem.* **285**, 16360–16368.
Bernard, A. & Payton, M. (2001). *Curr. Protoc. Protein Sci.*, Unit 5.3. doi:10.1002/0471140864.ps0503s00.
Bierne, H., Sabet, C., Personnic, N. & Cossart, P. (2007). *Microbes Infect.* **9**, 1156–1166.
Bober, M., Mörgelin, M., Olin, A. I., von Pawel-Rammingen, U. & Collin, M. (2011). *PLoS One*, **6**, e20345.
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.
Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). *Genome Res.* **14**, 1188–1190.
Doxey, A. C. & McConkey, B. J. (2013). *Virulence*, **4**, 453–466.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T. & Matsushima, N. (2004). *Proteins*, **54**, 394–403.
Eshghi, A., Lourdault, K., Murray, G. L., Bartpho, T., Sermswan, R. W., Picardeau, M., Adler, B., Snarr, B., Zuerner, R. L. & Cameron, C. E. (2012). *Infect. Immun.* **80**, 3892–3899.
Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. (2001). *J. Mol. Biol.* **312**, 807–821.
Fernandez-Prada, C. M., Hoover, D. L., Tall, B. D., Hartman, A. B., Kopelowitz, J. & Venkatesan, M. M. (2000). *Infect. Immun.* **68**, 3608–3619.
Frachon, E., Bondet, V., Munier-Lehmann, H. & Bellalou, J. (2006). *Appl. Environ. Microbiol.* **72**, 5225–5231.
Helft, L., Reddy, V., Chen, X., Koller, T., Federici, L., Fernández-Recio, J., Gupta, R. & Bent, A. (2011). *PLoS One*, **6**, e21614.
Ikegami, A., Honma, K., Sharma, A. & Kuramitsu, H. K. (2004). *Infect. Immun.* **72**, 4619–4627.
Ireton, K. (2007). *Cell. Microbiol.* **9**, 1365–1375.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Keszei, A. F., Tang, X., McCormick, C., Zeqiraj, E., Rohde, J. R., Tyers, M. & Sicheri, F. (2014). *Mol. Cell. Biol.* **34**, 362–373.
King, A. M., Pretre, G., Bartpho, T., Sermswan, R. W., Toma, C., Suzuki, T., Eshghi, A., Picardeau, M., Adler, B. & Murray, G. L. (2014). *Infect. Immun.* **82**, 1123–1131.
Ko, A. I., Goarant, C. & Picardeau, M. (2009). *Nature Rev. Microbiol.* **7**, 736–747.
Kobe, B. & Kajava, A. V. (2001). *Curr. Opin. Struct. Biol.* **11**, 725–732.
Lambert, A., Picardeau, M., Haake, D. A., Sermswan, R. W., Srikram, A., Adler, B. & Murray, G. A. (2012). *Infect. Immun.* **80**, 2019–2025.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lau, C. L., Smythe, L. D., Craig, S. B. & Weinstein, P. (2010). *Trans. R. Soc. Trop. Med. Hyg.* **104**, 631–638.
Lecuit, M., Dramsi, S., Gottardi, C., Fedor-Chaiken, M., Gumbiner, B. & Cossart, P. (1999). *EMBO J.* **18**, 3956–3963.

Leung, K. Y., Reisner, B. S. & Straley, S. C. (1990). *Infect. Immun.* **58**, 3262–3271.
Lourdault, K., Cerqueira, G. M., Wunder, E. A. & Picardeau, M. (2011). *Infect. Immun.* **79**, 3711–3717.
Marino, M., Braun, L., Cossart, P. & Ghosh, P. (1999). *Mol. Cell*, **4**, 1063–1072.
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.
McEwan, P. A., Scott, P. G., Bishop, P. N. & Bella, J. (2006). *J. Struct. Biol.* **155**, 294–305.
Miao, E. A., Scherer, C. A., Tsolis, R. M., Kingsley, R. A., Adams, L. G., Bäumler, A. J. & Miller, S. I. (1999). *Mol. Microbiol.* **34**, 850–864.
Murray, G. L., Srikram, A., Henry, R., Puapairoj, A., Sermswan, R. W. & Adler, B. (2009). *Microbes Infect.* **11**, 311–314.
Nally, J. E., Chow, E., Fishbein, M. C., Blanco, D. R. & Lovett, M. A. (2005). *Infect. Immun.* **73**, 3251–3260.
Neves, D., Job, V., Dortet, L., Cossart, P. & Dessen, A. (2013). *J. Mol. Biol.* **425**, 4520–4529.
Polle, L., Rigano, L. A., Julian, R., Ireton, K. & Schubert, W. D. (2014). *Structure*, **22**, 304–314.
Reid, S. D., Montgomery, A. G., Voyich, J. M., DeLeo, F. R., Lei, B., Ireland, R. M., Green, N. M., Liu, M., Lukomski, S. & Musser, J. M. (2003). *Infect. Immun.* **71**, 7043–7052.
Roy, A., Kucukural, A. & Zhang, Y. (2010). *Nature Protoc.* **5**, 725–738.
Roy, A., Yang, J. & Zhang, Y. (2012). *Nucleic Acids Res.* **40**, W471–W477.
Schubert, W. D., Göbel, G., Diepholz, M., Darji, A., Kloer, D., Hain, T., Chakraborty, T., Wehland, J., Domann, E. & Heinz, D. W. (2001). *J. Mol. Biol.* **312**, 783–794.
Schubert, W. D., Urbanke, C., Ziehm, T., Beier, V., Machner, M. P., Domann, E., Wehland, J., Chakraborty, T. & Heinz, D. W. (2002). *Cell*, **111**, 825–836.
Scott, P. G., McEwan, P. A., Dodd, C. M., Bergmann, E. M., Bishop, P. N. & Bella, J. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 15633–15638.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shen, Y., Naujokas, M., Park, M. & Ireton, K. (2000). *Cell*, **103**, 501–510.
Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. & Bricogne, G. (2012). *Acta Cryst.* D**68**, 368–380.
Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* D**65**, 582–601.
Toma, C., Okura, N., Takayama, C. & Suzuki, T. (2011). *Cell. Microbiol.* **13**, 1783–1792.
Wei, T., Gong, J., Jamitzky, F., Heckl, W. M., Stark, R. W. & Roessle, S. C. (2008). *BMC Struct. Biol.* **8**, 47.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wollert, T., Heinz, D. W. & Schubert, W. D. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 13960–13965.
Wollert, T., Pasche, B., Rochon, M., Deppenmeier, S., van den Heuvel, J., Gruber, A. D., Heinz, D. W., Lengeling, A. & Schubert, W. D. (2007). *Cell*, **129**, 891–902.
Zhang, Y. (2008). *BMC Bioinformatics*, **9**, 40.
Zhu, Y., Li, H., Hu, L., Wang, J., Zhou, Y., Pang, Z., Liu, L. & Shao, F. (2008). *Nature Struct. Mol. Biol.* **15**, 1302–1308.