Structural characterization of *Treponema pallidum* Tp0225 reveals an unexpected leucine-rich repeat architecture

Raghavendran Ramaswamy,[‡] Simon Houston,[‡] Bianca Loveless, Caroline E. Cameron and Martin J. Boulanger*

Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8P 5C2, Canada. *Correspondence e-mail: mboulanger@uvic.ca

Received 13 March 2019

Accepted 28 May 2019

Edited by R. Sankaranarayanan, Centre for Cellular and Molecular Biology, Hyderabad, India

[‡] These authors contributed equally to this work.

Keywords: syphilis; *Treponema pallidum*; X-ray crystallography; leucine rich-repeat domain.

PDB reference: leucine-rich repeat protein (TpLRR), 6mlx

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

The phylogenetically divergent spirochete bacterium *Treponema pallidum* subsp. *pallidum* is the causative agent of syphilis. Central to the capacity of *T. pallidum* to establish infection is the ability of the pathogen to attach to a diversity of host cells. Many pathogenic bacteria employ leucine-rich repeat (LRR) domain-containing proteins to mediate protein–protein interactions, including attachment to host components and establishment of infection. Intriguingly, *T. pallidum* expresses only one putative LRR domain-containing protein (Tp0225) with an unknown function. In an effort to ascribe a function to Tp0225, a comprehensive phylogenetic analysis was first performed; this investigation revealed that Tp0225 clusters with the pathogenic clade of treponemes. Its crystal structure was then determined to 2.0 Å resolution using Pt SAD phasing, which revealed a noncanonical architecture containing a hexameric LRR core with a discontinuous β -sheet bridged by solvent molecules. Furthermore, a surface-exposed, hydrophobic pocket, which was found in Tp0225 but is largely absent in canonical LRR domains from other pathogenic bacteria, may serve to coordinate a hydrophobic ligand. Overall, this study provides the first structural characterization of the sole LRR domain-containing protein from *T. pallidum* and offers insight into the unique molecular landscape of this important human pathogen.

1. Introduction

Syphilis is a chronic, multistage, sexually transmitted infection caused by the spirochete bacterium *Treponema pallidum* subsp. *pallidum* (hereafter *T. pallidum*). With an estimated 36 million cases globally and 11 million new infections per year (World Health Organization, 2011), syphilis is a major global public health threat. Despite the prevalence of this obligate human pathogen, the molecular mechanisms underlying its virulence are poorly understood. This is owing in part to the fact that *T. pallidum* is phylogenetically distinct, with approximately one-third of its predicted protein-coding genes having no known orthologs or assigned functions (Fraser *et al.*, 1998; Pětrošová *et al.*, 2013). Moreover, until recently *T. pallidum* could not be continuously cultured *in vitro* and it remains intractable to genetic manipulation (Liu *et al.*, 2010; Izard *et al.*, 2009; Edmondson *et al.*, 2018). The molecular characterization of treponemal proteins has therefore benefited from biochemical and structural studies (Parker, Houston, Pětrošová *et al.*, 2016; Parker, Houston, Wetherell *et al.*, 2016; Deka *et al.*, 2007).

Previous research investigating the protein complement of *T. pallidum* led to the identification of a protein associated with the peptidoglycan/cytoplasmic membrane complex that

harbors a putative leucine-rich repeat (LRR; Shevchenko *et al.*, 1997). Whole-genome sequencing of the *T. pallidum* Nichols strain later confirmed the presence of a corresponding protein-coding gene that was assigned the locus tag Tp0225 (Fraser *et al.*, 1998). Microarray studies further revealed that Tp0225 was expressed at more than twice the mean level of arrayed *T. pallidum* genes (Smajs *et al.*, 2005). Tp0225 was also shown to react weakly with anti-*T. pallidum* antibodies in sera from secondary-stage syphilis patients (Shevchenko *et al.*, 1997), indicating that the protein is expressed during human infection. These data provided a rationale for the further study of Tp0225, particularly when considered in the context that LRR-containing proteins often play a role in the processes of bacterial invasion, dissemination and immune evasion (Kędzierski *et al.*, 2004).

To further investigate the function of Tp0225, we first performed a comprehensive bioinformatic analysis, which showed that Tp0225 clusters more closely with the pathogenic clade of *T. pallidum* orthologs. We next determined its three-dimensional structure to 2.0 Å resolution using Pt SAD phasing. Consistent with the phylogenetically divergent nature of *T. pallidum*, structural analysis reveals several intriguing features of Tp0225, including (i) a noncanonical LRR fold with a discontinuous β -sheet structure that relies on an ordered network of solvent waters to bridge opposing β -strands and (ii) an uncommon surface-accessible hydrophobic pocket.

2. Materials and methods

2.1. Bioinformatic analysis of Tp0225 LRR-domain orthologs

Using *BLASTp*, 19 full-length orthologs (Supplementary Table S1) with homology to the Tp0225 LRR domain (AGN75438) were identified. When more than one ortholog was identified in a single *Treponema* species, the ortholog with the most significant *E* value was selected for further analyses. All treponemal orthologs had an *E*-value cutoff of 2.0×10^{-10} . The outgroup ortholog from *Bacillus galactosidilyticus* had an *E*-value cutoff of 3.0×10^{-3} . To identify regions of homology, the amino-acid sequences of each of the 19 full-length orthologs were individually aligned with the LRR domain of Tp0225 (residues Leu60–Pro230) using the pairwise sequence-alignment tool *EMBOSS Needle* (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). A multiple sequence alignment (MSA) of Tp0225 LRR-domain orthologs was generated using *Clustal Omega* (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). A phylogenetic tree was inferred from the MSA using the maximum-likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The initial tree for the heuristic search was obtained automatically by applying the neighbor-join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log-likelihood value. The tree was drawn to scale, with branch lengths corresponding to

the number of substitutions per site. All positions with less than 95% site coverage were eliminated. Evolutionary analyses were conducted using the *MEGA (Molecular Evolutionary Genetics Analysis)* software v.7 (Kumar *et al.*, 2008). The tree was rooted with the LRR domain from the Tp0225 ortholog from *B. galactosidilyticus* (WP_064468207).

2.2. Domain analysis of full-length Tp0225 orthologs from treponemes

The full-length sequences were submitted to the NCBI Conserved Domain Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>; Marchler-Bauer *et al.*, 2017) and searched against the Conserved Domain Database (v.3.16) using the default *E*-value threshold of 0.01. Specific hits, superfamily hits and multi-domain models were selected, whereas nonspecific hits were excluded from the analysis. All predicted domains had an *E*-value cutoff of 4.7×10^{-3} . The *SignalP* (<http://www.cbs.dtu.dk/services/SignalP/>; Petersen *et al.*, 2011) and *LipoP* (<http://www.cbs.dtu.dk/services/LipoP/>; Juncker *et al.*, 2003) servers were used to predict signal peptide cleavage sites and lipoprotein signal peptides, respectively.

2.3. Purification, crystallization and data collection

The sequence corresponding to *tp0225* was obtained from GenBank (Benson *et al.*, 2013; accession No. AAC45304.1). The *tp0225* sequence encoding the mature protein was codon-optimized, synthesized by GenScript and cloned into a modified pAcGP67b vector incorporating an N-terminal hexahistidine tag and a TEV cleavage site. Recombinant Tp0225 was produced using the baculovirus insect-cell system in *Trichoplusia ni* cells and was purified by nickel-affinity chromatography. The hexahistidine tag was removed by TEV cleavage and Tp0225 was further purified by size-exclusion chromatography in HBS (20 mM HEPES pH 7.5, 150 mM NaCl). Crystals of purified Tp0225 were obtained using the Index screen (Hampton Research) at 291 K. The final drops consisted of 0.3 μ l Tp0225 at 15 mg ml⁻¹ and 0.3 μ l reservoir solution and were equilibrated against 55 μ l reservoir solution (0.1 M bis-Tris-HCl pH 6.5, 25% PEG 3350). Crystals were derivatized with platinum by soaking with 2 mM K₂PtCl₄ at 291 K for 24 h. A single crystal was cryopreserved in mother liquor supplemented with 12.5% glycerol for 20 s and flash-cooled to 100 K directly in the cryostream. Diffraction data were collected on beamline 14-1 at Stanford Synchrotron Radiation Laboratory (SSRL) at a wavelength of 0.9202 Å.

2.4. Data processing, structure solution and refinement

Diffraction data were processed to resolutions of 2.80 Å (derivative) and 2.00 Å (native) using *iMosflm* (Battye *et al.*, 2011) and *AIMLESS* (Evans & Murshudov, 2013). The structure was solved by single-wavelength anomalous dispersion; 15 high-confidence Pt sites (from four molecules in the asymmetric unit) were identified and refined using *Crack2* (Ness *et al.*, 2004). The native Tp0225 structure was solved by molecular replacement using a single Tp0225 chain from the Pt-phased model in *Phaser* (McCoy, 2007). *Coot* (Emsley &

Cowan, 2004) was used for model building and the selection of solvent atoms and the model was refined in *phenix.refine* (Afonine *et al.*, 2012). Structural validation was performed with *MolProbity* (Chen *et al.*, 2010). 5% of the reflections were set aside for R_{free} calculation. Data-collection and refinement statistics are presented in Table 1. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB entry 6mlx).

3. Results and discussion

3.1. The Tp0225 LRR domain clusters in the pathogenic clade of treponemes

In contrast to other pathogenic bacteria, which contain multiple LRR proteins, Tp0225 is the only LRR protein predicted in *T. pallidum*. To place Tp0225 in the appropriate evolutionary context, we first compared Tp0225 with the most closely related orthologs from several treponemes. Intriguingly, phylogenetic analysis showed that the Tp0225 LRR domain is most closely related to orthologs from pathogens that cause systemic infections, is less related to orthologs from treponemes that cause localized infections and is even more distantly related to orthologs from nonpathogenic treponemes (Fig. 1, left). The N- and C-terminal regions flanking the LRR domains were also identified in all orthologs from pathogenic treponemes, but were present in only two of the five orthologs from nonpathogenic treponemes (Fig. 1, right). These observations suggest the potential for important functional differences between Tp0225 orthologs from treponeme pathogens and nonpathogens.

BLAST and conserved-domain searches of the Tp0225 N- and C-termini failed to identify functionally annotated domains, which is consistent with the phylogenetically distinct nature of *T. pallidum*. However, conserved domains of known function were identified in the termini of orthologs from the oral pathogens *T. socranskii* and *T. putidum* and in the ortholog from the termite gut commensal *T. azotonutricium* (Fig. 1, right). The *T. socranskii* ortholog was predicted to contain a prolipoprotein diacylglycerol transferase (LGT) domain C-terminal to its LRR domain, which is involved in lipoprotein biogenesis. Interestingly, our phylogenetic analysis also identified the LRR domain of the *T. socranskii* ortholog as the most closely related domain to the LRR domains of the orthologs from the systemic infection-causing treponemes (Tp0225 and its orthologs from the *T. pallidum* subspecies and *T. paraluisuniculi*). The *T. putidum* ortholog was predicted to contain a C-terminal domain belonging to the multifunctional Sel1-like repeat/tetratricopeptide repeat (SLR/TPR) domain family. The *T. azotonutricium* ortholog was predicted to contain an N-terminal bacterial Toll/interleukin-1 receptor (TIR 2) domain that may function as a structural mimic of the host Toll-like receptor (Chan *et al.*, 2009), and a C-terminal tetratricopeptide repeat (TPR) domain. The N-terminus of LRR proteins is suggested to impart hyperfunctionality to the proteins (Matsushima *et al.*, 2005) and, as shown here, it is

Table 1

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Tp0225, SAD	Tp0225
Data-collection statistics		
Space group	$P4_12_12$	$P4_12_12$
a, b, c (Å)	104.23, 104.23, 184.84	104.08, 104.08, 185.81
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Wavelength (Å)	0.9202	1.200
Resolution range (Å)	90.79–2.80 (2.95–2.80)	92.90–2.00 (2.11–2.00)
Measured reflections	612912 (85697)	666202 (26099)
Unique reflections	25890 (3707)	67107 (7859)
Multiplicity	23.7 (23.1)	9.9 (3.3)
Completeness (%)	100.0 (97.5)	96.2 (78.3)
$\langle I/\sigma(I) \rangle$	19.3 (6.6)	15.1 (2.4)
$R_{\text{merge}}^\dagger$	0.183 (0.997)	0.087 (0.443)
$CC_{1/2}$	0.999 (0.959)	0.999 (0.820)
No. of Pt sites	15	N/A
Refinement statistics		
Resolution (Å)		69.31–2.00
$R_{\text{cryst}}/R_{\text{free}}^\ddagger$		0.188/0.222
Molecules in asymmetric unit		4
Matthews coefficient (Å ³ Da ^{−1})		2.54
Solvent content (%)		51.63
No. of atoms		
Protein chain A		1611
Protein chain B		1616
Protein chain C		1621
Protein chain D		1616
Solvent		664
B values (Å ²)		
Protein chain A		32.8
Protein chain B		25.9
Protein chain C		32.3
Protein chain D		26.34
Solvent		37.59
R.m.s. deviation from ideality		
Bond lengths (Å)		0.007
Bond angles (°)		1.07
Ramachandran statistics (%)		
Most favored		98.8
Allowed		1.2
Disallowed		0.0

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection. $^\ddagger R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively.

likely that different orthologs use the N- and C-termini to perform species-specific functions.

3.2. Tp0225 adopts an atypical LRR architecture

T. pallidum is a phylogenetically divergent organism, hence the primary structure often reveals few details of the protein structure and function. To correlate the divergent sequence of Tp0225 with the protein fold, we determined its structure to a resolution of 2.0 Å using platinum-based SAD phasing. Overall the structure is well ordered, with the exception of the first nine residues of the N-terminal cap. Structural analysis reveals a central superhelical topology with low secondary-structure content and repeating short β -strands [Fig. 2(b), purple] framed by an N-terminal compact β -hairpin structure [Fig. 2(b), green] and a C-terminal single amphipathic α -helix [Fig. 2(b), blue]. The central LRR domain (Leu60–Asp231)

consists of six tandemly repeating units (LRR units) of varying lengths (23–30 residues). Each LRR unit adopts a β -turn with a conserved, hydrophobic motif directed towards the interior of the molecule that is involved in stacking interactions and is crucial for the formation of the hydrophobic core in the LRR domain [Fig. 2(b), bottom]. Further stabilizing the LRR core are three distinct structural ‘spines’ formed by a network of repeating phenylalanines (Phe85, Phe108, Phe131, Phe154, Phe177, Phe200 and Phe223), cysteines (Cys88, Cys111, Cys133, Cys157, Cys180, Cys203 and Cys226) and solvent waters [Fig. 2(b), right]. The Phe and Cys spines appear to be conserved in most of the pathogenic treponemes that cause

systemic and localized infections, such as *T. paraluisancuniculi*, *T. socranskii* and *T. phagedenis*. Moreover, Phe, Cys and water spines are also observed in many LRR structures across various species and help in maintaining the structural integrity of LRRs (Kobe & Deisenhofer, 1994). With respect to the oxidation state of the cysteines, a previous study indicated that Tp0225 might be localized to the periplasmic space, which is likely to be oxidizing (Shevchenko *et al.*, 1997). However, the buried nature of the cysteines within the LRR core is consistent with the reduced form observed in the crystal structure. While the N-terminal (Lys26–Tyr59) and C-terminal (Val232–Phe246) substructures provide a hydrophilic cap for

Figure 1

Phylogenetic and domain analysis of Tp0225 orthologs from pathogenic and nonpathogenic treponemes. Left: a phylogenetic tree of LRR domains from treponemal orthologs of Tp0225 was inferred using the maximum-likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992). The tree was drawn to scale, with branch lengths corresponding to the number of substitutions per site. Evolutionary analyses were conducted using *MEGA7* (Kumar *et al.*, 2008). The tree was rooted with the LRR domain from the Tp0225 ortholog from *B. galactosidilyticus* (WP_064468207). The branches for pathogenic and nonpathogenic treponemes are indicated in red and blue, respectively. Right: domain analyses of full-length Tp0225 and orthologous proteins used to construct the phylogenetic tree are shown. Tp0225 ortholog domain predictions were performed using the NCBI Conserved Domain search tool. SPI/SPII, signal peptide with signal peptidase I/II cleavage sites. LRR, leucine-rich repeat region. LGT, prolipoprotein diacylglycerol transferase domain. SLR/TPR, Sel1-like repeat/tetratricopeptide repeat domain. TIR 2, bacterial Toll/interleukin-1 receptor domain. Gray bar, signal peptide. Green bar, non-LRR domain-containing N-terminus. Purple bar, LRR domain/LRR domain-containing region. Blue bar, non-LRR domain-containing C-terminus.

the hydrophobic LRR core, they do not conform to canonical LRR caps, which frequently consist of Cys clusters containing two or four Cys residues (Matsushima & Miyashita, 2012).

A structural comparison of closely related orthologs identified by *DALI* (Holm & Rosenström, 2010) revealed that in contrast to the canonical contiguous β -structure, Tp0225 adopts a noncontiguous β -sheet with a break in the secondary structure following $\beta 5$ [Fig. 2(c)]. A network of bridging waters compensates for the loss of a single β -strand and maintains the hydrogen-bonding pattern formed by the β -strands in the contiguous regions of the structure [Fig. 2(c)]. Despite the solvent bridge, which has the potential to impart on flexibility, Tp0225 adopts a linear topology. Notably, this architecture contrasts with the curved architecture of canonical LRR domains such as *Escherichia coli* LRR (EcLRR;

PDB entry 4r6j; Park *et al.*, 2015), which relies on carbonyl O atoms being directed inwards towards the solenoid core at the point of curvature [Fig. 2(c)].

3.3. Surface-accessible hydrophobic pocket of Tp0225 LRR

The general biological function associated with LRRs is to provide a structural framework for the formation of specific protein–protein interactions or protein–lipid interactions (Kobe & Kajava, 2001). Intriguingly, structural analysis of the LRR domain of Tp0225 revealed the presence of a surface-accessible hydrophobic pocket in the LRR core [Fig. 3(a)], a rare feature in structurally characterized LRRs. To gain insight into the potential role of this pocket, we compared Tp0225 with the top ten structural orthologs identified by

Figure 2

Tp0225 adopts a noncanonical LRR architecture with a noncontiguous β -sheet. (a) Tp0225 contains an N-terminal signal peptide domain (SP; residues 1–25), an LRR N-terminal capping region (NTR; 26–60; pale green), a leucine-rich repeat domain (LRR; 60–230; purple) which harbors six LRRs, and an LRR C-terminal capping region (CTR; 231–246; blue). (b) Top left: secondary-structure depiction of Tp0225, highlighting the different domains using the same coloring scheme as in (a). Top right: schematic diagram of the topology of the Tp0225 structure. σ_A -weighted $2F_o - F_c$ electron density is contoured at 1.5σ , highlighting the Phe, Cys and solvent spine. Bottom: the organization of the LRR2 module with residues forming the conserved hydrophobic core. The structure-based sequence alignment of the six LRR modules shows the consensus motif. Hydrophobic residues are denoted ϕ . (c) Left: solvent molecules (red spheres) in Tp0225 that maintain the hydrogen-bonding pattern between β -strands. Hydrogen-bonding patterns are also shown for EvLRR (cyan) and EcLRR (green). In both Tp0225 and EvLRR the O atoms lie in the plane of the β -sheet, whereas they point inwards in EcLRR (right), which breaks the hydrogen bond, resulting in curvature.

Figure 3 Surface analysis highlighted the presence of an unusual LRR-domain hydrophobic pocket. (a) Top: surface representation of Tp0225 (purple) with a solid box indicating the surface pocket. Inset: enlarged side view of the predominantly hydrophobic surface pocket. Middle: surface representation of *F. praetensis* LRR (FpLRR; PDB entry 4fs7). Bottom: surface representation of *Homo sapiens* TLR2 (HsTLR2; light blue) with its ligand (triacylated lipopeptide; green) (PDB entry 2z7x). The inset indicates the surface pocket. HsTLR crystallized as a dimer with two monomers, TLR1 and TLR2. For the sake of clarity, we show only TLR2. (b) Mapping of conserved (burgundy) and variable (teal) residues from different Tp0225 orthologs (see Supplementary Table S1 for accession numbers) onto the Tp0225 core using ConSurf (Ashkenazy *et al.*, 2010). An area of conserved residues surrounding the pocket is indicated by the dotted circle.

DALI (Holm & Rosenström, 2010). Structural comparison revealed that of the DALI hits, only the LRR domain from the functionally uncharacterized *Faecalibacterium praetensis* (FpLRR; PDB entry 4fs7; Joint Center for Structural Genomics, unpublished work) harbored a similar pocket [Fig. 3(a)]. Notably, the LRR-containing human Toll-like receptor 2 (HsTLR2; PDB entry 2z7x; Jin *et al.*, 2007) also presents a hydrophobic surface pocket (400 Å³) that binds a 16-carbon fatty-acid chain [Fig. 3(a)]. The pockets in both FpLRR and Tp0225, however, are significantly smaller (200 and 100 Å³, respectively), consistent with a requirement to coordinate a smaller hydrophobic ligand [Fig. 3(a)]. Supporting a role for the functionality of the hydrophobic pocket in Tp0225, a ConSurf analysis (Ashkenazy *et al.*, 2010) showed a clear clustering of conserved residues among orthologs from pathogenic treponemes that form the surface pocket [Fig. 3(b)].

4. Conclusion

T. pallidum is a highly invasive pathogen, yet little is known about the detailed molecular mechanism underlying pathogenesis. In the present study, we showed that Tp0225 clusters with homologs from pathogenic treponemes. Structural analysis revealed that Tp0225 adopts an atypical LRR domain with a flattened topology and incorporates a noncontiguous β -sheet bridged by solvent waters. A surface-accessible hydrophobic pocket that is largely absent in LRR domains from other bacterial pathogens suggests the potential for coordinating a binding partner.

Acknowledgements

The authors declare that this research was conducted in the absence of any commercial or financial relationships that

could be construed as a potential conflict of interest. RR, SH, CEC and MJB designed the experiments. RR, SH, BL and MJB performed the experiments. RR, SH and MJB wrote the paper with editorial support from all authors. We thank the staff at the Stanford Synchrotron Radiation Lightsource.

Funding information

Grants in support of this work include a Canadian Institutes for Health Research Grant MOP82915 (MJB), a Public Health Service Grant (AI-051334) from the National Institutes of Health (CEC) and a Discovery Grant (327186) from the Natural Sciences and Engineering Research Council of Canada (CEC). MJB and CEC gratefully acknowledge the Canada Research Chair program for salary support.

References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). *Nucleic Acids Res.* **38**, W529–W533.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). *Nucleic Acids Res.* **41**, D36–D42.
- Chan, S. L., Low, L. Y., Hsu, S., Li, S., Liu, T., Santelli, E., Le Negrate, G., Reed, J. C., Woods, V. L. & Pascual, J. (2009). *J. Biol. Chem.* **284**, 21386–21392.
- 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.* **D66**, 12–21.
- Deka, R. K., Brautigam, C. A., Tomson, F. L., Lumpkins, S. B., Tomchick, D. R., Machius, M. & Norgard, M. V. (2007). *J. Biol. Chem.* **282**, 5944–5958.
- Edmondson, D. G., Hu, B. & Norris, S. J. (2018). *MBio*, **9**, e01153-18.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., Gwinn, M., Hickey, E. K., Clayton, R., Ketchum, K. A., Sodergren, E., Hardham, J. M., McLeod, M. P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J. K., Chidambaram, M., Utterback, T., McDonald, L., Artiach, P., Bowman, C., Cotton, M. D., Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Sandusky, M., Weidman, J., Smith, H. O. & Venter, J. C. (1998). *Science*, **281**, 375–388.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Izard, J., Renken, C., Hsieh, C.-E., Desrosiers, D. C., Dunham-Ems, S., La Vake, C., Gebhardt, L. L., Limberger, R. J., Cox, D. L., Marko, M. & Radolf, J. D. (2009). *J. Bacteriol.* **191**, 7566–7580.
- Jin, M. S., Kim, S. E., Heo, J. Y., Lee, M. E., Kim, H. M., Paik, S.-G., Lee, H. & Lee, J.-O. (2007). *Cell*, **130**, 1071–1082.
- Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). *Bioinformatics*, **8**, 275–282.
- Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2003). *Protein Sci.* **12**, 1652–1662.
- Kędzierski, Ł., Montgomery, J., Curtis, J. & Handman, E. (2004). *Arch. Immunol. Ther. Exp.* **52**, 104–112.
- Kobe, B. & Deisenhofer, J. (1994). *Trends Biochem. Sci.* **19**, 415–421.
- Kobe, B. & Kajava, A. V. (2001). *Curr. Opin. Struct. Biol.* **11**, 725–732.
- Kumar, S., Nei, M., Dudley, J. & Tamura, K. (2008). *Brief. Bioinform.* **9**, 299–306.
- Liu, J., Howell, J. K., Bradley, S. D., Zheng, Y., Zhou, Z. H. & Norris, S. J. (2010). *J. Mol. Biol.* **403**, 546–561.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwatz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y. & Bryant, S. H. (2017). *Nucleic Acids Res.* **45**, D200–D203.
- Matsushima, N. & Miyashita, H. (2012). *Biomolecules*, **2**, 288–311.
- Matsushima, N., Tachi, N., Kuroki, Y., Enkhbayar, P., Osaki, M., Kamiya, M. & Kretsinger, R. (2005). *Cell. Mol. Life Sci.* **62**, 2771–2791.
- McCoy, A. J. (2007). *Acta Cryst.* **D63**, 32–41.
- Ness, S. R., de Graaff, R. A., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
- Park, K., Shen, B. W., Parmeggiani, F., Huang, P.-S., Stoddard, B. L. & Baker, D. (2015). *Nature Struct. Mol. Biol.* **22**, 167–174.
- Parker, M. L., Houston, S., Pětrošová, H., Lithgow, K. V., Hof, R., Wetherell, C., Kao, W.-C., Lin, Y.-P., Moriarty, T. J., Ebady, R., Cameron, C. E. & Boulanger, M. J. (2016). *PLoS Pathog.* **12**, e1005919.
- Parker, M. L., Houston, S., Wetherell, C., Cameron, C. E. & Boulanger, M. J. (2016). *PLoS One*, **11**, e0166274.
- Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature Methods*, **8**, 785–786.
- Pětrošová, H., Pospíšilová, P., Strouhal, M., Čejková, D., Zbaníková, M., Mikalová, L., Sodergren, E., Weinstock, G. M. & Šmajš, D. (2013). *PLoS One*, **8**, e74319.
- Shevchenko, D. V., Akins, D. R., Robinson, E., Li, M., Popova, T. G., Cox, D. L. & Radolf, J. D. (1997). *J. Bacteriol.* **179**, 3188–3195.
- Šmajš, D., McKevitt, M., Howell, J. K., Norris, S. J., Cai, W.-W., Palzkill, T. & Weinstock, G. M. (2005). *J. Bacteriol.* **187**, 1866–1874.
- World Health Organization (2011). *Prevalence and Incidence of Selected Sexually Transmitted Infections*. Geneva: World Health Organization. <https://www.who.int/reproductivehealth/publications/rtis/9789241502450/en/>.