



BIOLOGICAL  
CRYSTALLOGRAPHY

**Volume 71 (2015)**

**Supporting information for article:**

**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**

## SUPPORTING INFORMATION

**Table S1: LRR proteins in *Leptospira* spp.**

Proteins for which the structures are reported in the present study are shaded in grey

								orthologues in	
<i>L. interrogans</i> str.				LRR	other			<i>L. borgpetersenii</i>	<i>L. biflexa</i>
Fiocruz L1-130	TrEMBL	aa	kDa	number <sup>a</sup>	domain <sup>b</sup>	SignalP <sup>c</sup>	PSORTB <sup>d</sup>	str. L550 (id%)	str. Patoc (id%)
LIC10828	Q72U36	378	43.99	14	X	NO	Extracellular	X	X
LIC10829 <sup>e</sup>	Q72U35	402	46.29	17	X	NO	Extracellular	X	X
LIC10830	Q72U34	521	60.16	12	X	YES	Extracellular	X	X
LIC10831	Q72U33	377	44.20	13	X	YES	Extracellular	X	X
LIC11051	Q72TH0	685	78.36	8	WGR	NO	Extracellular	LBL_1017 (61%)	X
LIC11097	Q72TC4	413	48.21	13	X	YES	Extracellular	X	X
LIC11098	Q72TC3	426	48.97	17	X	YES	Extracellular	X	X
LIC11180	Q72T41	288	33.43	10	X	NO	Extracellular	X	X
LIC11504	Q72S80	266	31.06	7	X	NO	Extracellular	X	X
LIC11505	Q72S79	572	66.28	20	X	YES	Extracellular	X	X
LIC11507	Q72S77	500	58.44	8	X	YES	Extracellular	X	X
LIC12234 <sup>f</sup>	Q72Q78	215	24.71	8	X	NO	Extracellular	X	X
LIC12375	Q72PU2	301	35.20	8	X	YES	Extracellular	X	X
LIC12401	Q72PR6	217	24.58	5	X	YES	Extracellular	LBL_0836 (74%)	LEPBl2289 (40%)
LIC12512	Q72PF9	122	14.16	3	X	YES	Cytoplasmic	X	X
LIC12676	Q72P01	657	76.25	3	WGR	NO	Unknown	X	X
LIC12759	Q72NS0	423	48.76	17	X	YES	Extracellular	X	X
LIC12899	Q72ND5	272	31.03	9	X	NO	Extracellular	X	X
LIC12901	Q72ND3	1616	184.69	7	WGR	NO	Extracellular	LBL_2267 (72%)	X
LIC20055	Q75FX6	291	38.31	2	X	NO	Extracellular		X
LIC20154	Q75FM8	241	28.26	4	X	YES	Cytoplasmic	LBL_4175 (79%)	X

<sup>a</sup> LRR finder: <http://www.lrrfinder.com/>

<sup>b</sup> InterProScan (<http://www.ebi.ac.uk/interpro/scan.html>). WGR: This domain is found in a variety of polyA polymerases as well as the *E. coli* molybdate metabolism regulator P33345 and other proteins of unknown function.

<sup>c</sup> Hidden Markov models using SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP3.0/>).

<sup>d</sup> PSORTb version 3.0 <http://www.psort.org/psortb/>

<sup>e</sup> probable pseudogene with a premature stop codon

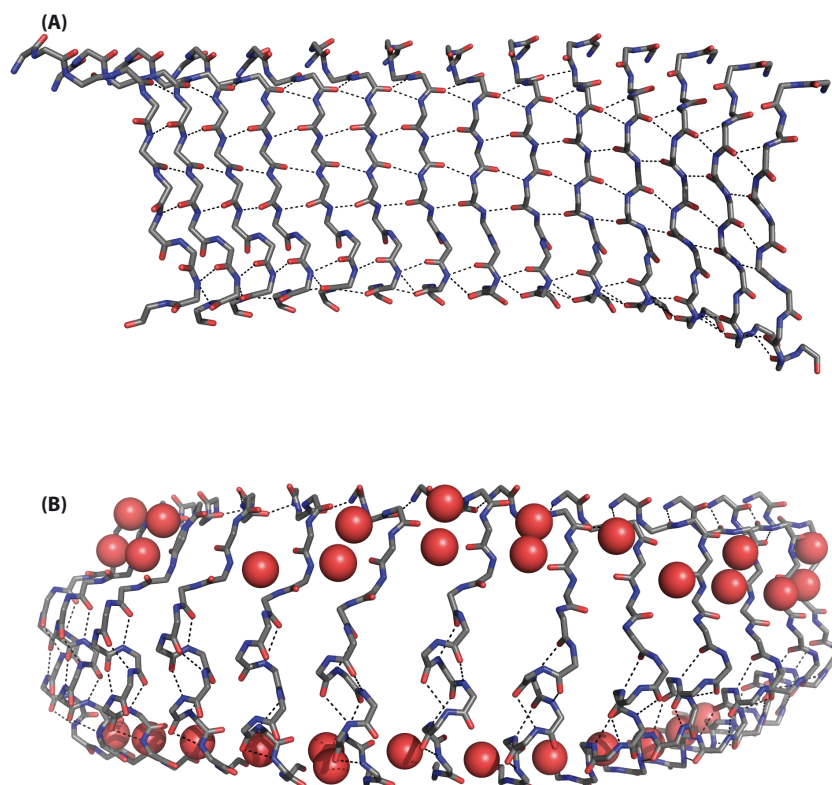
<sup>f</sup> correction of incorrect ORF start site

**Table S2: Dynamic light scattering (DLS) data**

Protein	Hydrodynamic radius (nm)	Apparent Molecular weight estimated by DLS (kDa)	Theoretical molecular weight of monomer (kDa)
LIC10831	3	44	40.86
LIC11098	3.1	48	45.61
LIC12759	3	45	45.95
LIC12234	2.2	22	22.05
LIC12234 + 50mM ZnAc	2.9	40	22.05

DLS measurements were performed with a DynaPro-MS800 molecular-sizing instrument (Protein solutions, Lakewood, NJ, USA). Protein samples in presence or in absence of 50mM zinc acetate were loaded into a 45 $\mu$ l quartz cuvette. The hydrodynamic radius and molecular mass were determined from 50 measurements at 18°C. Data were analyzed using DYNAMICS 6.9.2.11 software. The results show that all the recombinant LRR proteins are monomeric in solution in the absence of zinc, and the presence of zinc induces the dimerization of LIC12234.

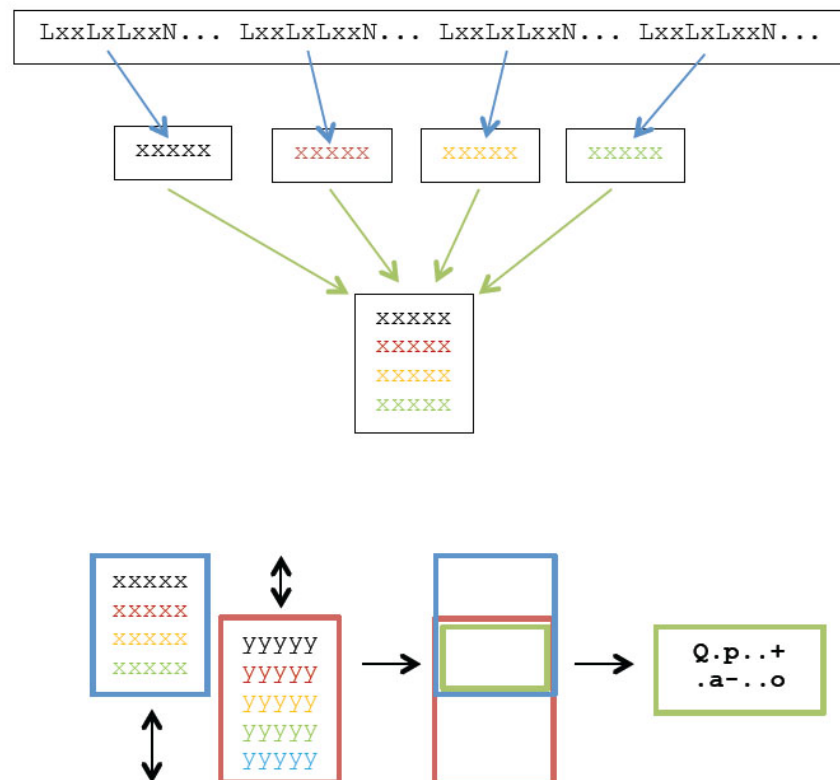
**Figure S1: Conserved hydrogen bond network and water-mediated interactions in LIC10831.** (A) Extended network of hydrogen bonds in the inner concave surface. (B) Hydrogen bonds in the helical coils on the outer convex surface, and conserved network of water molecules between repeat units.



**Figure S2a. Alignment method used for the comparison of external residues on the concave surfaces of LRR containing proteins.**

A bio-informatic analysis was used to compare the residues on the exterior of the concave surface. These residues can be represented as an  $M \times N$  matrix, where  $M$  is the number of outward pointing residues along the  $\beta$ -strand and  $N$  is the number of  $\beta$ -strands in the concave surface.  $M$  is usually equal to five and can be identified from the LxxLxLxxN segment (positions 1-9) of the consensus LRR motif. The method used for comparison resembles the Repeat Conservation Mapping method (Helft *et al.*, 2011), which utilizes the consensus sequence to reduce the LRR domains into matrices as a means of predicting functional sites on the surface of LRR proteins. Similar methods have also been used in the analysis of LRR protein structures (Evdokimov *et al.*, 2001)(McEwan *et al.*, 2006, Scott *et al.*, 2004).

In the first step, the external residues of the LxxLxLxxN segments are extracted and compressed into 5-amino acid lines of 1-letter codes. The 5-residue lines are then concatenated into a  $5 \times N$  matrix. In the second step, two matrices of 1-letter codes are compared. Offsets of the matrices (blue and red boxes, below) are scanned, and the overlapping elements (green) are directly compared for each register. In the results (green box), aliphatic (**a**), aromatic (**o**), polar (**p**), negatively charged (**-**), positively charged (**+**), and variable (**.**) residues are indicated and identical residues are shown in capital letters. Percentages of identical and similar residues are determined using the concave surface with the least number of residues.



**References**

- Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. (2001). *J. Mol. Biol.* **312**, 807-821.
- Helft, L., Reddy, V., Chen, X., Koller, T., Federici, L., Fernández-Recio, J., Gupta, R. & Bent, A. (2011). *PLoS One* **6**, e21614.
- McEwan, P. A., Scott, P. G., Bishop, P. N. & Bella, J. (2006). *J. Struct. Biol.* **155**, 294-305.
- Scott, P. G., McEwan, P. A., Dodd, C. M., Bergmann, E. M., Bishop, P. N. & Bella, J. (2004). *Proc Natl Acad Sci U S A* **101**, 15633-15638.

**Figure S2b : Matrices for the external residues on the concave surfaces of LIC12234, LIC10831, LIC11098 and LIC12759.** Positively charged sidechains are shown in blue, negatively charged sidechains in red, polar sidechains in purple, aliphatic sidechains in green and aromatic sidechains in black.

<u>Lic12234</u>	<u>Lic10831</u>	<u>Lic11098</u>	<u>Lic12759</u>
QIDSR	RVDSR	TYR__	_KPK_
ESHRD	QRYHY	RINSG	RNDSF
KYDSR	QLYRS	QLNDD	QKDGG
EVFNG	QVDGS	QQHSK	QKNNN
GIYNN	QLYHS	QKKYE	QESHs
VSSSS	KSDSN	QENAH	QKNDN
RINwD	KSYSE	QTYGH	QESLS
TES__	KVFNN	ESGDH	KNDNH
	QYYSD	ESGDH	ENDRS
	QTDsY	QIHRN	KVMTG
	QTD RN	QKLNK	KTNGE
	QTFSN	QKKYE	LENYY
	LWSVY	QEDDG	KYSYH
	QTYNN	QTYGN	QEHSG
		ESDEH	EWSSN
		QTNKY	QREGN
		KKYHN	QREDS
		_IY__	_ED__

**Figure S2c: One-dimensional alignment of the external residues on the concave surfaces of LIC10831 and LIC11098.** Each row of the M x N matrix represents the five external residues of a beta strand on the concave surface. The MxN matrices derived from two LRR proteins (Seq1, Seq2) are matched by shifting the alignment of the columns and comparing the identities and homologies of the residues. A match matrix is determined for each shift, representing the homology of individual positions on the concave surface. Identical residues are indicated in capital letters. Residues which do not match but have common chemical character are represented with the following code : **+** = positively charged, **-** = negatively charged, **a** = aliphatic, **o** = aromatic, **p** = polar. Dissimilar residues are represented by a period « . », and orphan residues lacking an opposite partner are represented as an underscore « \_ » .

```
seq1 = Lic10831 ; seq2 = Lic11098
```

```
Offset1 = 2 , Offset2 = 2
```

Seq1	seq2	Matrix
_ _ _	TYR _	_ _ _
_ _ _	RINSG	_ _ _
RVDSR	QLNDD	.a...
QRYHY	QQHSK	Q....
QLYRS	QKKYE	Q....
QVDGS	QENAH	Q..a.
QLYHS	QTYGH	Q.Y..
KSDSN	ESGDH	.p...
KSYSE	ESGDH	.p...
KVFNN	QIHRN	.a..N
QYYSD	QKLNK	Q..p.
QTDSY	QKKYE	Q....
QTDRN	QEDDG	Q.D..
QTFSN	QTYGN	QTo.N
LWSVY	ESDEH	.....
QTYNN	QTNKY	QT...
_ _ _	KKYHN	_ _ _
_ _ _	_IY _	_ _ _

```
[offset1, offset2, N_, Ni, Na, No, Np, N-, N+, N., Ntotal]
```

```
[2, 2, 20, 15, 3, 1, 3, 0, 0, 48, 90]
```

The alignment results reveal 15 identical residue pairs and 3 aliphatic, 1 aromatic, 3 polar and 48 non-matching pairs.

### Figure S2d : Example of the MxN alignment of LIC10831 and InIA

The result show 9 identical residues, plus 1 aliphatic match, 17 polar matches and 1 negatively charged match. There are 42 non-matching pairs ( Seq1 from LIC10831 and Seq2 from InIA).

Offset1 = 2 , Offset2 = 0

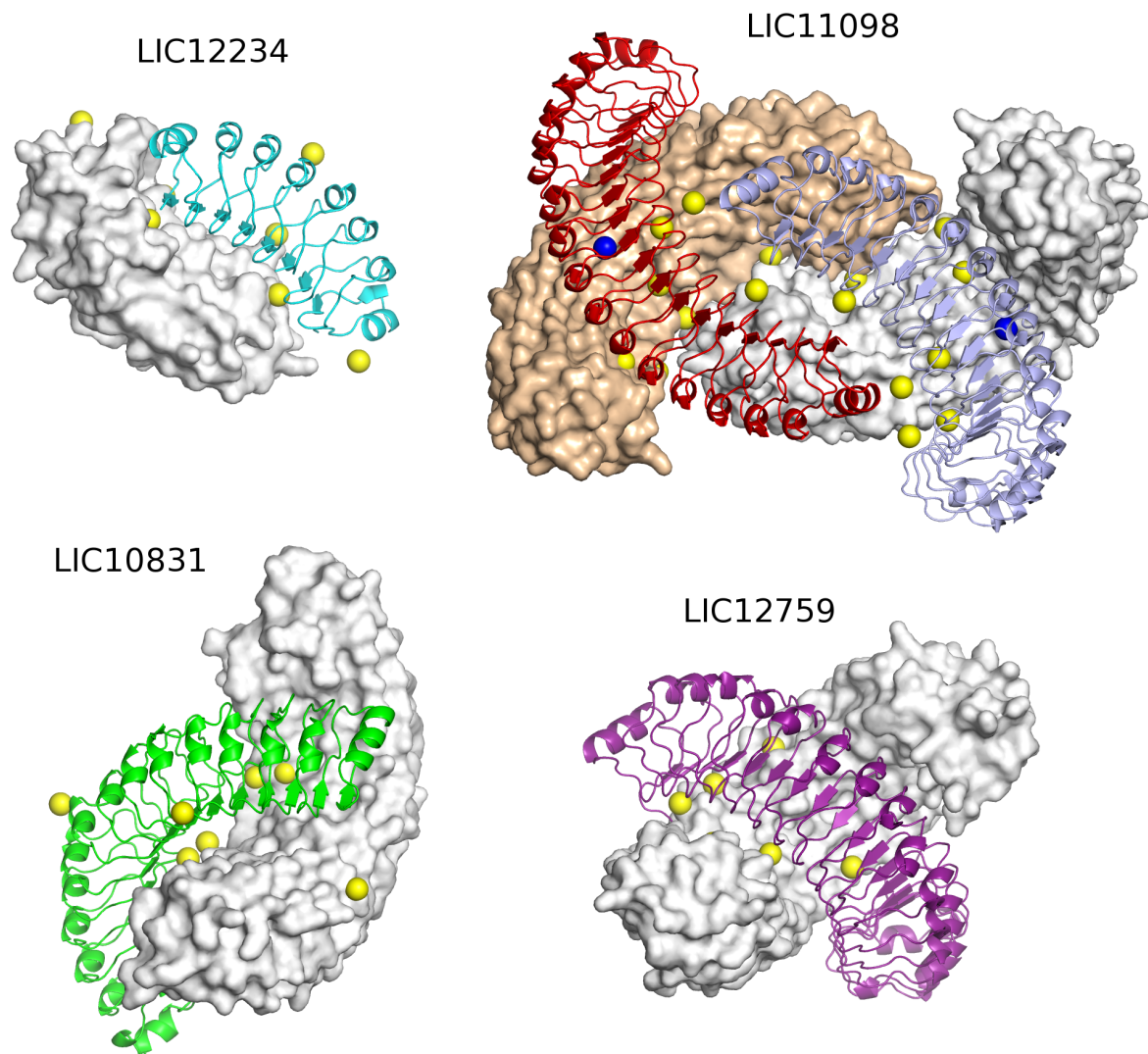
Seq1	Seq2	Matrix
- - -	TTQDR	- - -
- - -	TQNSN	- - -
RVDSR	VDLNN	...p.
QRYHY	TGTFN	p....
QLYRS	NRESS	p...p
QVDGS	QSGN	Q..Gp
QLYHS	ERDSS	....p
KSDSN	ESITN	.p.pN
KSYSE	DESNG	...p.
KVFNN	TDDAN	....N
QYYSD	TEKGA	p....
QTDSY	TNENE	pp-p.
QTDRN	TYTYF	p....
QTFSN	QRFYN	Q.F.N
LWSVY	NWSGH	.Wpa.
QTYNN	TQGND	pp.N.

[offset1, offset2, N\_, Ni, Na, No, Np, N-, N+, N., Ntotal]

[2, 0, 10, 9, 1, 0, 17, 1, 0, 42, 80]

**Figure S3. Quaternary organization in the crystal structures of the four leptospiral LRR proteins.**

Analysis of the protein interfaces in the crystal structures, using the EBI Pisa server (<http://www.ebi.ac.uk/pdbe/pisa/>), suggests a stable dimeric quaternary organization for LIC12234, LIC10831, and LIC12759, and a stable tetrameric organization for LIC11098. Zinc ions (shown as yellow spheres), and calcium in LIC11098 (blue spheres) mediate interactions between the protein monomers.



**Figure S4: Examples of zinc binding sites in the leptospiral LRR proteins.**

Independent molecules in the crystallographic asymmetric unit are colored in green and blue, respectively.

