

Crystal structure of an LRR protein with two solenoids

Cell Research (2013) 23:303-305. doi:10.1038/cr.2012.159; published online 13 November 2012

Dear Editor,

The leucine rich repeat (LRR) is a versatile protein motif (~20-30 amino acids) found in a variety of protein in all life forms. Tandem arrays of two or more repeats in an LRR domain form a structure similar to a horseshoe-shaped solenoid. Largely by acting as scaffolds for protein-protein interactions, LRR domains participate in diverse biological processes [1, 2], as exemplified by the large family of LRR-containing receptor like kinases (LRR-RLKs) in plants [3]. Nearly all the known typical LRR structures comprise an N-terminal cap (LRRNT) and some of them have a C-terminal cap (LRRCT) as well, shielding the two ends of the conserved central LRR segment [1, 2]. An LRR domain is characterized by a repetitive sequence pattern rich in leucine residues. However, in some LRR proteins including some plant LRR-RLKs and mammalian Toll-like receptors (TLR7-9), the LRR repeats are interrupted by sequences of 30-70 residues termed island domain or non-LRR region. A recent genome-wide study identified more than 600 proteins belonging to this family of LRR proteins [4]. Compared with the canonical ones, these atypical LRR proteins are much less well-characterized structurally. BRI1 is the only atypical LRR protein whose LRR domain structure has been reported [5, 6]. Despite that its superhelical structure is markedly different from those of the typical LRR proteins, the interrupted LRRs in BRI1 still pack continuously, forming a single solenoid structure as observed in all the other known LRR structures. However, atypical LRRs like TLR7-9 have been predicted to form two solenoids [7], although experimental evidence for this is lacking. Here, we describe the crystal structure of the LRR domain of TMK1 (receptor-like Trans-Membrane Kinase 1) [8], an atypical LRR protein from *Arabidopsis* that contains two solenoids and discuss its biological implications.

TMK1 is an LRR-RLK [8] and contains a non-LRR region in its extracellular LRR domain (TMK1-LRR) (Supplementary information, Figure S1), although its functional information is limited. We expressed TMK1-LRR (residues 1~469) in insect cells, purified the protein as previously described [5] and solved its crystal

structure with molecular replacement. The structure was refined to a resolution of 1.55 Å with R_{factor} of 16.9% and R_{free} of 19.0% (Supplementary information, Table S1). Except for 23 residues from the N-terminus and 22 from the C-terminus, all the remaining residues were well-defined and included in the finally refined model.

Strikingly different from the horseshoe structures of the canonical LRR proteins, the overall structure of the TMK1-LRR is shaped like the Arabic number “7” (Figure 1A) and can be clearly divided into two LRR domains. The C-terminal LRR (C-LRR, see below for further evidence) domain packs nearly perpendicularly against the spine of the N-terminal LRR (N-LRR). Its uncapped C-terminus might in part result from the exposed carbonyl oxygen atoms from the last repeat (Supplementary information, Figure S2), which can increase protein solubility. The structure showed the existence of 13 LRRs in the TMK1-LRR, with 10 from the N-LRR and 3 from the C-LRR (Figure 1A). Continuous packing of the 10 LRRs with lengths ranging from 21 to 26 residues (Supplementary information, Figure S3) forms a regular LRR domain, resembling the N-terminal portion of BRI1 as indicated by database search using the DALI server (Supplementary information, Figure S4A). Except for LRR1, LRR3, LRR4 and LRR7 (Supplementary information, Figure S3), all the other LRRs from the N-LRR contain the sequence GxL/i/vP (x stands for any amino acid) specific for the plant LRR proteins [9]. Like BRI1 and other LRR proteins, the N-LRR is stabilized by an LRRNT comprising a disulfide bond between Cys53 and Cys60 (Figure 1A). In contrast, the other end of the N-LRR domain is only partially shielded by a loop from the C-LRR. The convex outer surface of the N-LRR contains 3₁₀ α-helices and loops of various lengths, but no β-sheet observed in BRI1 and PGIP1 [10].

The amino acids (322-362) from the non-LRR region form a compact structure with two 3₁₀ α-helices and a β-strand (Figure 1A). In stark contrast to the non-LRR region in BRI1-LRR that contacts the concave inner surface of BRI1-LRR [5, 6], the non-LRR region of TMK1-LRR interacts with the spine of the N-LRR (Figure 1A and 1B). Gln292 is located at the center of the interface and forms bidentate hydrogen bonds with the carbonyl

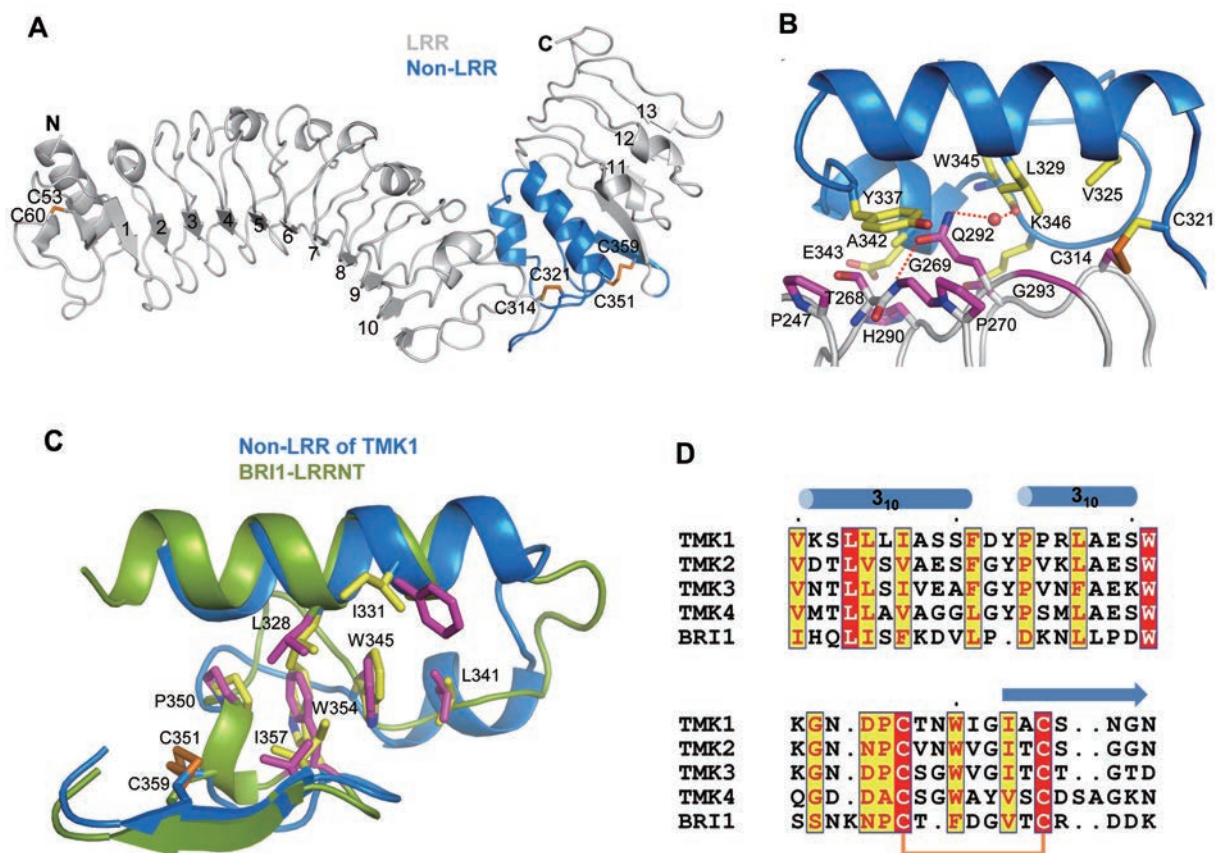


Figure 1 TMK1-LRR contains two solenoids. **(A)** Overall structure of TMK1-LRR domain. The LRRs (numbered as indicated) and the non-LRR region of TMK1-LRR are colored in gray and blue, respectively. The three disulfide bonds (Cys53-Cys60, Cys314-Cys321 and Cys351-Cys359) are shown in orange. "N" and "C" represent N- and C-terminus, respectively. **(B)** Interaction of the non-LRR region with the N-terminal LRR (N-LRR) of TMK1-LRR. The side chains from the non-LRR region and N-LRR are shown in yellow and magenta, respectively. The red-dotted lines represent hydrogen bonds. The red sphere stands for the oxygen atom of the water molecule. **(C)** Structure alignment of the non-LRR region of TMK1-LRR (residues 322-362, in blue) with LRRNT of BRI1 (residues 33-72, in light green). The side chains of the amino acids from the structural cores of both proteins are shown in yellow and magenta and only those amino acids from the non-LRR region of TMK1 are labeled. **(D)** Sequence alignment between the non-LRR region of TMK1 homologs from *Arabidopsis* and the LRRNT of BRI1. The identical amino acids are red-shaded and similar ones are yellow-shaded. The orange line linking two cysteine residues represents a disulfide bond.

oxygen of Trp345 (water-mediated) and the main-chain nitrogen of Gly269. Gln292 also forms Van der Waals interactions with its neighboring residues from the C-LRR. Many contacts, primarily hydrophobic ones, from the periphery of the interface further strengthen the interactions between the non-LRR region and the N-LRR. For example, Tyr337 of the non-LRR region tightly packs against the $C\alpha$ atom of Gly269 and the side chains of Pro247 and Pro270 from the N-LRR, and His290 of the N-LRR contacts the side chains of Ala342 and Glu343 from the non-LRR region. In addition, and perhaps more importantly, the disulfide bond between Cys314 and Cys321 appears critical for the packing of the two LRR domains. The amino acids involved in the interactions between the N-LRR and the non-LRR region are highly conserved

among TMKs from *Arabidopsis* (Supplementary information, Figure S1), suggesting their conserved roles in protein folding. Our structural analyses indicate that a unique positioning of the non-LRR region in TMK1-LRR disrupts the continuous packing of LRRs, resulting in the formation of two LRR solenoid structures.

Unexpectedly, structural comparison using DALI search revealed that the non-LRR domain (residues 322-362) of TMK1-LRR is the closest structure homolog of the LRRNT from BRI1 (residues 33-72), with an RMSD of 2.7 Å over 40 aligned $C\alpha$ atoms (Figure 1C). Consistently, structure-based sequence alignment indicated that the amino acids around the hydrophobic cores of the non-LRR domain of TMKs and BRI1-LRRNT are well conserved (Figure 1D). Additionally, as observed in

the LRRNTs of many other LRR proteins, the non-LRR domain is further stabilized by a disulfide bridge formed between Cys351 and Cys359, which are conserved in the non-LRR of TMKs and BRI1-LRRNT (Figure 1C and 1D). The β -strand from the non-LRR domain packs tightly against the first LRR from the C-LRR, forming a regular LRR structure, which resembles that of the N-terminal portion of BRI1-LRR (Supplementary information, Figure S4B). Together, these observations suggest that the non-LRR region of TMK1-LRR acts as an LRRNT, further supporting the formation of the second LRR domain in TMK1-LRR.

It is widely believed that LRR proteins adopt a horse shoe-like structure. Although LRR structures with deviations from the canonical structure have been reported [5, 6, 11, 12], the continuous packing of the LRRs in these structures still results in the formation of a single solenoid. Our study shows a unique LRR structure containing two LRR domains, providing experimental evidence for a previous prediction that a non-LRR region-containing LRR protein can contain two solenoids [7]. The unexpected finding that the non-LRR region of TMK1 resembles the LRRNT of BRI1 is consistent with the model in which the family of non-LRR region-containing LRR proteins underwent duplication and fusion during evolution [4]. The non-LRR domain makes contacts with the spine region of the N-LRR, different from its interaction with the concave inner surface of the LRR solenoid, which is observed in BRI1-LRR, thus disrupting the continuous packing of LRRs in TMK1-LRR. Supporting its capping role, the non-LRR domain packs tightly against its immediately following LRR, resulting in the formation of a complete LRR structure.

The non-LRR regions from several LRR proteins including BRI1 have been shown to be important for ligand recognition [5, 6, 13]. However, our crystal structure raises a cautionary note on functional annotations for the non-LRR region of an LRR protein. The non-LRR in TMK1-LRR appears more critical for structural integrity than for ligand recognition or protein-protein interaction, although we cannot rule out the latter possibilities. A challenging question derived from the structure is how the two LRR domains function. Paucity of functional information renders it difficult to associate the structure of TMK1 with its functions. However, it is possible that TMK1-LRR provides more than one scaffold for protein

interaction or ligand recognition. For example, the inner surfaces of the two LRR domains might interact with different TMK1 partners if they do exist. Establishment of structure-function relationship of TMK1 awaits future functional studies.

Accession numbers

The coordinates and structural factors have been deposited in PDB database under the accession code 4HQ1.

Acknowledgments

We thank F Yu and J He at Shanghai Synchrotron Radiation Facility (SSRF). JC was funded by the State Key Program of the National Natural Science Foundation of China (31130063) and the Ministry of Science and Technology of China (2010CB835300).

Peiyuan Liu^{1,2*}, Zehan Hu^{2,3*}, Bin Zhou¹, Simiao Liu^{4,5}, Jijie Chai^{4,5}

¹College of Biological Sciences, China Agricultural University, Beijing 100083, China; ²National Institute of Biological Sciences, No. 7 Science Park Road, Beijing 102206, China; ³College of Life Sciences, Beijing Normal University, Beijing 100875, China; ⁴Tsinghua-Peking Center for Life Sciences, Beijing 100084, China; ⁵School of Life Sciences, Tsinghua University, Beijing 100084, China

*These two authors contributed equally to this work.

Correspondence: Jijie Chai

E-mail: chajij@mail.tsinghua.edu.cn

References

- 1 Kobe B, Kajava AV. *Curr Opin Struct Biol* 2001; **11**:725-732.
- 2 Botos I, Segal DM, Davies DR. *Structure* 2011; **19**:447-459.
- 3 Chang C, Schaller GE, Patterson SE, *et al.* *Plant J* 2011; **66**:117-127.
- 4 Matsushima N, Mikami T, Tanaka T, *et al.* *Biochim Biophys Acta* 2009; **1790**:1217-1237.
- 5 She J, Han Z, Kim TW, *et al.* *Nature* 2011; **474**:472-476.
- 6 Hothorn M, Belkhadir Y, Dreux M, *et al.* *Nature* 2011; **474**:467-471.
- 7 Matsushima N, Tanaka T, Enkhbayar P, *et al.* *BMC Genomics* 2007; **8**:124.
- 8 Chang C, Schaller GE, Patterson SE *et al.* *Plant Cell* 1992; **4**:1263-1271.
- 9 Kajava AV. *J Mol Biol* 1998; **277**:519-527.
- 10 Di Matteo A, Federici L, Mattei B, *et al.* *Proc Natl Acad Sci USA* 2003; **100**:10124-10128.
- 11 Kim HM, Park BS, Kim JI, *et al.* *Cell* 2007; **130**:906-917.
- 12 Tan X, Calderon-Villalobos LI, Sharon M, *et al.* *Nature* 2007; **446**:640-645.
- 13 Kinoshita T, Caño-Delgado A, Seto H, *et al.* *Nature* 2005; **433**:167-171.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)