

# Helical Parameters and Correlations of Tandem Leucine Rich Repeats in Proteins

Purejav Enkhbayar<sup>1#</sup>, Hiroki Miyashita<sup>2#</sup>, Robert H Kretsinger<sup>3\*</sup> and Norio Matsushima<sup>2\*</sup>

<sup>1</sup>Department of Biophysics and Bioinformatics, School of Biology and Biotechnology, National University of Mongolia, Ulaanbaatar 210646/377, Mongolia

<sup>2</sup>Institute of Tandem Repeats, Sapporo 004-0882, Japan

<sup>3</sup>Department of Biology, University of Virginia, Charlottesville 22904, USA

<sup>#</sup>These two authors contribute equally to this study

## Abstract

Leucine rich repeats (LRRs) are present in over 20,000 proteins from viruses to eukaryotes. Most LRR units are 20-30 residues long and can be divided into a highly conserved segment and a variable segment. Eight classes have been recognized. Two to sixty-two units occur in tandem to form an LRR structure. The tertiary structures of these LRRs are helical, in which the  $\beta$ -strands of the highly conserved segments stack in parallel. This helix consists of a super helical arrangement of repeating structural units. We call it a coil of solenoids. We have used our program HELFIT to assign helical parameters to 642 LRRs of known structures of 114 proteins. We report these parameters and their correlations with eight classes of LRR, with the number of repeat units in the LRR, with oligomerization, and with ligand state of the LRR. The helical parameters of the eight LRR classes frequently overlap one another. However, the constant distance between parallel  $\beta$ -strands is the primary determinant of the helical parameters of the LRRs. When the repeat number,  $n$ , in LRRs is small, the LRR structures are more variable and, by inference, more flexible. In the LRRs with  $n \geq 10$ ,  $\Delta z$  (the rise per repeat unit) of the "RI-like" and "Cysteine-containing" classes is smaller than those of "SDS22-like", and "Plant-specific" classes. This difference is ascribed mainly to the difference in the structural units. The helical parameters of the LRRs unambiguously describe both right handed and left handed helices, helical dimers, and subdomains if they exist. Moreover, the helical parameters sensitively detect structural changes induced by protein, protein interactions, glycosylation, and/or mutation.

**Keywords:** HELFIT; LRR; Solenoid; Parallel  $\beta$ -sheet; BRI1; Oligomerization; Protein-protein interactions

**Abbreviations:** 3D: Three dimensional; AMIGO-1: Amphotericin-induced Proteins 1; APLIC: *Anopheles* Plasmodium-responsive Leucine-rich Repeat Protein 1; AtCOI1: Coronatine-insensitive Protein 1 from *Arabidopsis thaliana*; AtRPK2: LRR receptor-like serine/threonine-protein Kinase RPK2 from *Arabidopsis thaliana*; AtTIR1: Transport Inhibitor Response 1 from *Arabidopsis thaliana*; AtTMK1: Probable Receptor Protein Kinase TMK1 from *Arabidopsis thaliana*; BAK1: BRASSINOSTEROID INSENSITIVE 1-associated Receptor Kinase 1; BRI1: Protein BRASSINOSTEROID INSENSITIVE 1; BRL1: Serine/threonine-protein Kinase BRI1-like 1; CD14: Myeloid Cell-Specific Leucine-Rich Glycoprotein; C $\alpha$ : alpha-carbon; DLC1: Flagellar Outer Arm Dynein Light Chain 1; dsRNA: Double-stranded RNA; FSHR: Follicle Stimulating Hormone Receptor; FSL2: LRR Receptor-Like Serine/Threonine-Protein Kinase FLS2; Glib $\beta$ : Glycoprotein Ib beta; GPIba: Glycoprotein Ib alpha; GPIX: Glycoprotein IX; HCS: Highly Conserved Segment; HEL: Anti-Hen Egg White Lysozyme; InIA: Internalin A; InIB: Internalin B; InIC: Internalin C; InIF: Internalin F; InIH: Internalin H; InIJ: Internalin J; InIK: Internalin K; IR: Island region; LEGL7: *Legionella* Leucine-Rich Repeat-Containing Protein; LGR: Leucine-Rich Repeat-Containing G Protein-Coupled Receptor; LINGO-1: Leucine-Rich Repeat and Immunoglobulin-Like Domain-Containing Nogo Receptor-Interacting Protein 1; LRIM1: Leucine-rich Immune Molecule 1; LRR: Leucine Rich Repeat; NGL-2: Netrin-G2 Ligand/ Leucine-Rich Repeat-Containing Protein 4; NGL-3: Netrin-G3 Ligand/ Leucine-Rich Repeat Containing Protein 4b; NgR-4: Reticulon-4 receptor/Nogo receptor/Nogo-66 receptor; NLR4: NLR Family CARD Domain Containing Protein 4; NLRX1: NLR Family Member X1; PP32/ANP32A/LANP: Acidic Leucine-rich Nuclear Phospho-protein 32 Family Member A; PGIP: Polygalacturonase Inhibitor; RabGGT $\alpha$ : Geranylgeranyl Transferase type-2 Subunit Alpha; RanGAP: Ran GTPase Activating Protein 1; RI:

Ribonuclease Inhibitor; RMSD: The Root Mean Square Deviation from the best fit helix; SERK1: Somatic Embryogenesis Receptor Kinase 1; Skp2: S-phase Kinase-Associated Protein 2; SspH: *Salmonella* Secreted Protein H; TAP: mRNA export factor TAP; TLR: Toll-like Receptor; Trk-A: Tyrosine Kinase Receptor A; TSHR: Thyroid-Stimulating Hormone Receptor; U2 snRNP A': U2 Small Nuclear Ribonucleoprotein A'; VLR: Variable Lymphocyte Receptor; VS: Variable Segment; YopM: Leucine-rich Effector Protein from *Yersinia pestis*

## Introduction

LRRs are present in over 20,000 proteins from viruses to eukaryotes, as listed in the data bases of PFAM, SMART, PROSITE, and InterPro [1-4]. All organisms for which sequence data are available have at least one LRR protein. Most LRR proteins are involved in protein, protein interactions, as observed in the plant immune response and the mammalian innate immune response [5-9]. Beyond immunity, extensive functional diversity occurs among LRR proteins, including their involvement in apoptosis, autophagy, ubiquitin related processes, nuclear mRNA transport, neuronal development, and the type III secretion system in pathogenic bacteria [10-16]. Furthermore, plant LRR proteins including LRR containing receptor-like kinases and LRR

**\*Corresponding authors:** Norio Matsushima, Institute of Tandem Repeats, Sapporo 060-8556, Japan, Tel/Fax: +81 11 886 0087; E-mail: [norio\\_irreko@outlook.jp](mailto:norio_irreko@outlook.jp)

Robert H Kretsinger, Department of Biology, University of Virginia, Charlottesville 22904, USA, Tel: 434-982-5764; Fax: 434-982-0019; E-mail: [rhk5i@virginia.edu](mailto:rhk5i@virginia.edu)

Received June 03, 2014; Accepted June 20, 2014; Published June 24, 2014

**Citation:** Enkhbayar P, Miyashita H, Kretsinger RH, Matsushima N (2014) Helical Parameters and Correlations of Tandem Leucine Rich Repeats in Proteins. J Proteomics Bioinform 7: 139-150. doi:10.4172/0974-276X.1000314

**Copyright:** © 2014 Enkhbayar P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

containing receptor like proteins act as signal amplifiers in the cases of tissue damage, of establishing symbiotic relationships, and of affecting developmental processes [17-19].

Two to sixty-two LRR units occur in tandem. Each unit is typically 20-30 residues long and can be divided into a highly conserved segment (HCS) and a variable segment (VS) [6,20,21]. The HCS consists of an eleven residue stretch, LxxLxLxxNxL, or a twelve residue stretch, LxxLxLxxCxxL, in which “L” is Leu, Ile, Val, or Phe; “N” is Asn, Thr, Ser, or Cys; and “C” is Cys, Ser or Asn. Eight classes of LRRs-“RI-like”, “Cysteine-containing”, “SDS22-like”, “IRREKO”, “Bacterial”, “Plant-specific”, “Typical”, and “TpLRR”-have been recognized [6,20-22]. They are characterized by different lengths and consensus sequences of the VS of their repeat units. Most of the known LRRs consist of only one class of units; however a few have a mixture of two classes. A super-motif of LRRs occurs in a group of LRR proteins including the subfamily of small LRR proteoglycan including biglycan, and decorin, and TLRs 7, 8, and 9 [23-25]. Moreover, the first, N-terminal LRR unit in the “Typical” or “IRREKO” LRR domains is frequently occupied by a “Bacterial” motif [21]. Crystal and/or solution structures of representatives of all the eight classes of LRR are available.

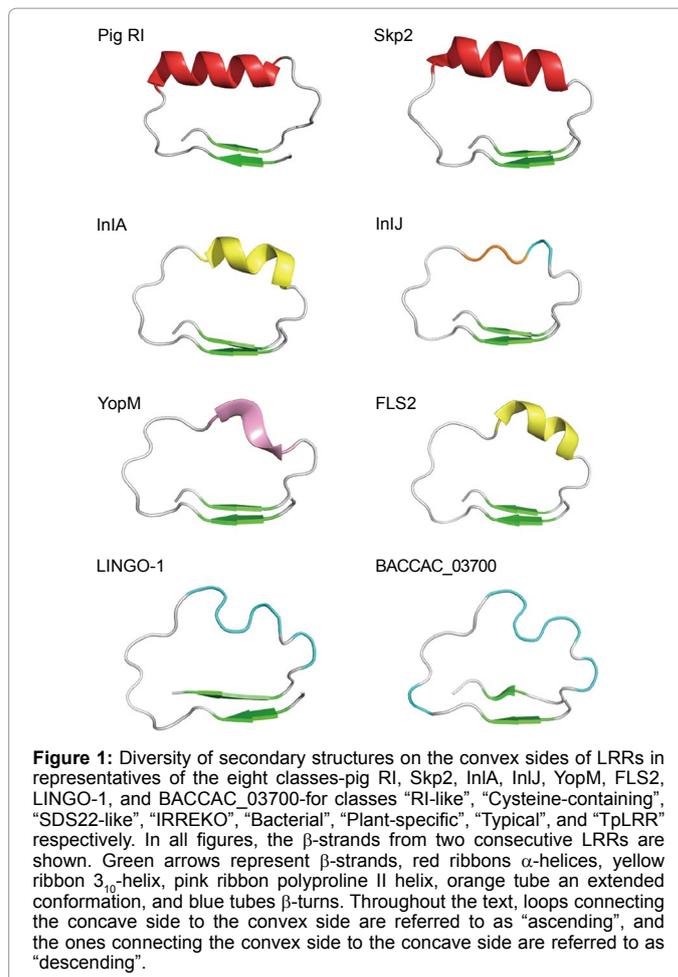
Three residues at positions 3 to 5, xLx, in the HCS form a short  $\beta$ -strand [5-9]. These  $\beta$ -strands from tandem LRR units stack parallel and then the LRRs form an arc or a helix. This helix is not simple but consists of a super helical arrangement of repeating structural units. We call it a coil of solenoids.

The concave face, consisting of the HCSs of each LRR unit, consists of a parallel  $\beta$ -sheet, one strand from each HCS unit. The convex face, consisting of the VS of each LRR unit, is made of a variety of secondary structures including the  $\alpha$ -helix,  $3_{10}$ -helix, polyproline II helix, and an extended conformation or a tandem arrangement of  $\beta$ -turns (Figure 1). The various secondary structures on the convex side are connected to the strands forming the  $\beta$ -sheet in its concave side by two loops (Figure 1). The “ascending loop” links the C-terminal end of the  $\beta$ -strand in the HCS to the N-terminus of the characteristic secondary structure in the VS [9] (Figure 1). The “descending loop” links the C-terminal end of the characteristic secondary structure in the VS to the N-terminus of the  $\beta$ -strand in the HCS of the following unit. Most of the known LRR structures have an N- and/or C-cap that shields the hydrophobic core of the first LRR unit at the N-terminus and/or the last unit at the C-terminus [5-9]. In extracellular proteins or extracellular regions, these caps frequently consist of Cys clusters consisting of two or four Cys residues on the N- and C-terminal sides of the LRRs.

Non-LRR, Island Regions, IRs, interrupting LRRs are widely distributed; they are referred to as “islands” or “loop outs” [26-28]. A large number of plant LRR proteins including LRR containing receptor like kinases and LRR containing receptor like proteins have IRs [24,28]. In our analyses we treat the LRR regions on either side of its IR as distinct helices, as well as analyzing the entire LRR as a single helix.

Enkhbayar et al. [29-31] developed a program, HELFIT, that determines the helical parameters-helix axis, pitch (with handedness), radius, and number of points or units per turn-as well as RMSD. Only four points are required to define a helix. The trace of a helix with pitch=0.0Å is a circle; with N=exactly 1.0 the trace is a straight line.

Here we have determined the helical parameters for each LRR and evaluated correlation with class of LRR, with the number of repeat units in the LRR, with oligomerization, and with ligand state of the LRR to reveal some structural features of LRR structures. We demonstrate that the constant distance between parallel  $\beta$ -strands is the primary



**Figure 1:** Diversity of secondary structures on the convex sides of LRRs in representatives of the eight classes-pig RI, Skp2, InIA, InIJ, YopM, FLS2, LINGO-1, and BACCAC\_03700-for classes “RI-like”, “Cysteine-containing”, “SDS22-like”, “IRREKO”, “Bacterial”, “Plant-specific”, “Typical”, and “TpLRR” respectively. In all figures, the  $\beta$ -strands from two consecutive LRRs are shown. Green arrows represent  $\beta$ -strands, red ribbons  $\alpha$ -helices, yellow ribbon  $3_{10}$ -helix, pink ribbon polyproline II helix, orange tube an extended conformation, and blue tubes  $\beta$ -turns. Throughout the text, loops connecting the concave side to the convex side are referred to as “ascending”, and the ones connecting the convex side to the concave side are referred to as “descending”.

determinant of the helical parameters of the LRRs. When the repeat number in LRRs is small, the LRR structures are more variable and, by inference, more flexible.

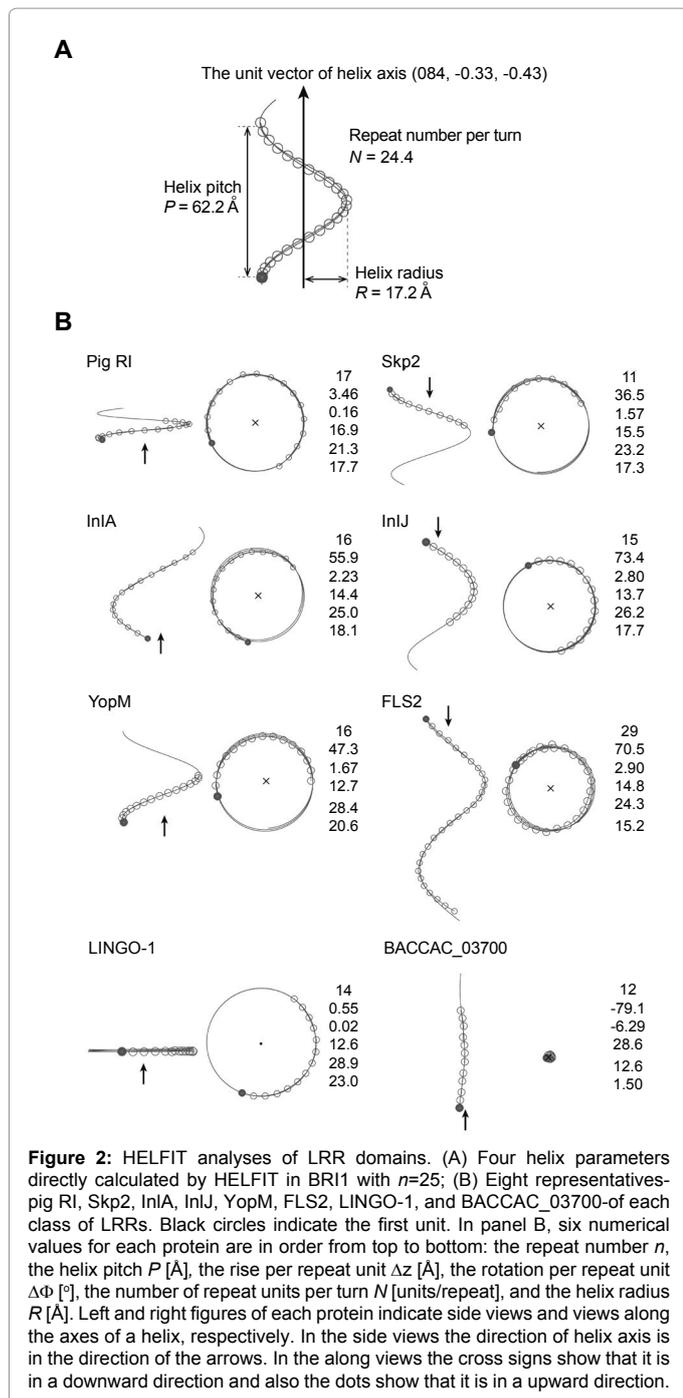
## Methods

A helix consisting of  $n$  repeat units may be characterized by helix axis, pitch ( $P$ ), helix radius ( $R$ ), and number of repeat units per turn ( $N$ ). HELFIT [29-31] computes these parameters in which the helix axis is represented by the unit vector (Figure 2A). These parameters also yield the rise per repeat unit ( $\Delta z=P/N$ ) and the rotation per repeat unit ( $\Delta\Phi=360^\circ/N$ ). Moreover, HELFIT gives rmsd:

$$\text{rmsd} = \left[ \left( \text{the minimum of } \frac{\sum (d_i)^2}{n} \right)^{1/2} \right]$$

where  $d_i$  is the closest distances from data point to the trace of the helix. Here  $p = \left( \frac{\text{rmsd}}{(n-1)^{1/2}} \right)$  gives the regularity of helix independent of its length. If  $p$  is relatively large, the LRR is not well described by a helix and might be more easily visualized as another shape, e.g. an ellipse.

In most LRRs the three residues at positions 3-5 in the HCS form a short  $\beta$ -strand, which is almost completely conserved in all LRR structures [5-9]. These  $\beta$ -strands from adjacent repeat units form a parallel, pleated sheet on the inside, or concave, surface of the helix. As reference points for HELFIT we use the coordinates of the  $\alpha$ -carbon ( $C\alpha$ ) of the consensus leucine residue at position 4 (corresponding to



the middle of each  $\beta$ -strand) in individual LRR repeat units.

Three structures of LRRs with  $n=3$  have been determined. In these cases we performed 3D circle fitting that directly fits a circle to a set of 3D data points, because only three points define a circle [32]. The 3D circle fitting, meaning  $P=0.0$  and  $\Delta z=0.0$  as an a priori assumption determines the radius of the circle, the number per circle, the center of the circle, and the root mean square deviation from the best fit circle. The 3D circle fitting was also utilized for the estimation of the average distance between adjacent repeats by the helical parameters. Furthermore, the center of the circle was used for the calculation of the distance between the two monomers in TLR dimers.

For grouping of eight LRR classes of known LRR domains we performed similarity searches by FASTA and by BLAST using representative LRR domains clearly belonging to individual classes as query sequences. The representative LRR domains were selected from those for which the 3D structures have been solved. “RI-like” LRRs are from RI, “Cysteine-containing” from Skp2, “SDS22-like” from InIA, “IRREKO” from InIJ, “Bacterial” from YopM, “Plant-specific” from FLS2, “Typical” from LINGO-1, and “TpLRR” from BACCAC\_03700.

Secondary structure assignments based on 3D coordinates were made using the DSSP program ([http://crdd.osdd.net/raghava/ccpdb/beta2\\_up.php](http://crdd.osdd.net/raghava/ccpdb/beta2_up.php)) [33].

## Results

### Helical parameters

Figure 2B shows HELFIT computed helices in representatives—RI, Skp2, InIA, InIJ, YopM, FLS2, LINGO-1, and BACCAC\_03700—of each of the eight classes of LRRs. The fits are very good. Supplementary Table 1 shows the helical parameters— $P$ ,  $\Delta z$ ,  $\Delta\Phi$ ,  $N$ , and  $R$  of the 642 LRRs of known structures of 114 different proteins (Supplementary Table 1) [34-145].

### The correlation of $\Delta z$ , $\Delta\Phi$ , and $R$

$D$  is the average  $C\alpha(i)-C\alpha(i+1)$  distance between adjacent repeats, in which the  $C\alpha$  atoms are at position 4 in the  $\beta$ -strand [32].  $D$  is a function of  $\Delta z$ ,  $\Delta\Phi$ , and  $R$ .

$$D = \left[ \left\{ 2R \sin\left(\frac{\Delta\Phi}{2}\right) \right\}^2 + (\Delta z)^2 \right]^{\frac{1}{2}} \quad (1)$$

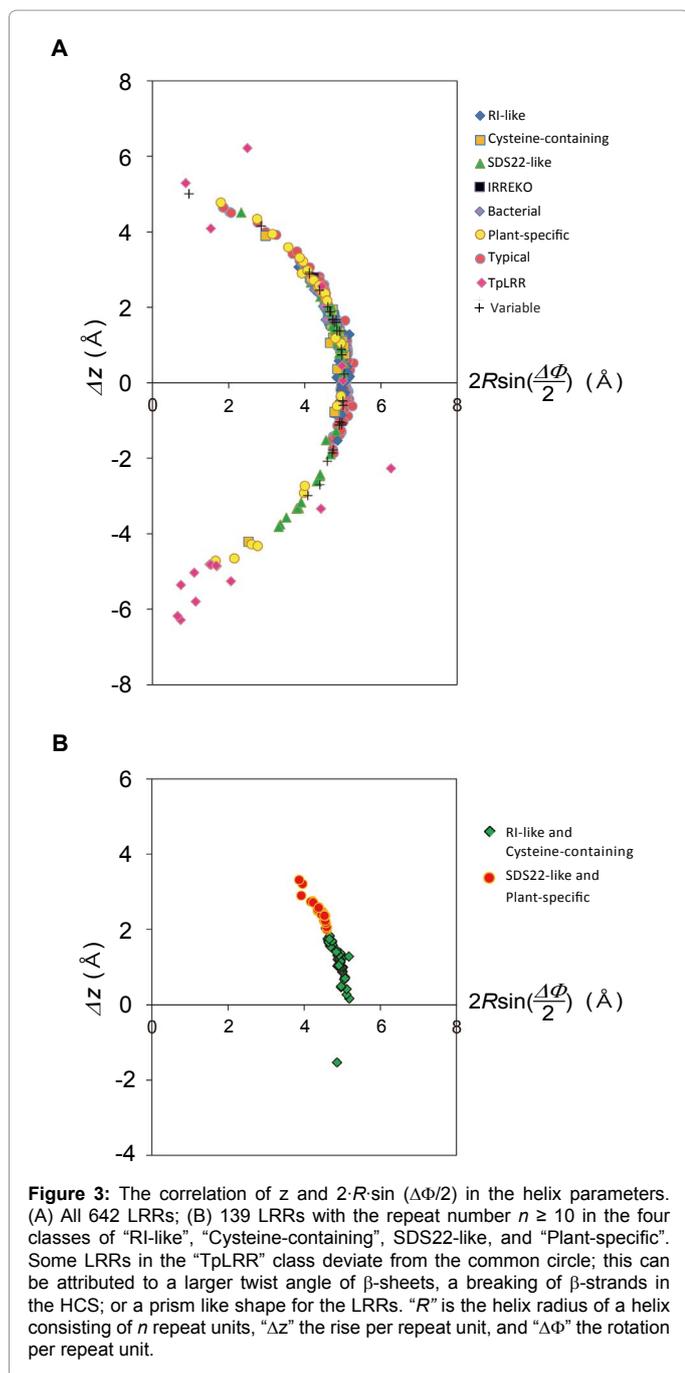
The data points of  $2 \cdot R \cdot \sin(\Delta\Phi/2)$  versus  $\Delta z$  fall on a circle with radius  $D$ ; although, in the “TpLRR” class the data deviate from this circle (Figure 3). The circle fittings using in total 627 data points (except for those of the “TpLRR” proteins) gives  $D=5.02 \pm 0.00 \text{ \AA}$  [32]. This  $D$  corresponds to the inter-strand distance that is in the range of 4.5 to 5.5  $\text{\AA}$  [146]; this distance allows the formation of hydrogen bonds between parallel strands. In a canonical, parallel  $\beta$ -sheet the strands are aligned at the same level parallel to the axes of the individual strands. This occurs in LRR structures; the observed  $D$  is consistent with the distance between strands in the canonical parallel  $\beta$ -sheet.

The  $D$  value in the “TpLRR” class is  $5.52 \pm 0.14 \text{ \AA}$  and thus is larger and more variable than those in other classes. The calculation of the  $C\alpha(i)-C\alpha(i+1)$  distance revealed  $D>6.9 \text{ \AA}$  between neighboring  $\beta$ -strands, as observed in FAEPRAA2165\_0102, EUBVEN\_01088, BACCAC\_03700, and BACOVA\_01565. This greater distance arises from water bridges in the hydrogen bonds between the strands.

Most of the LRRs have  $R=15 \rightarrow 30 \text{ \AA}$  (Supplementary Table 1). However, five LRRs in human Trk-A shows very small  $R$  ( $=0.89 \text{ \AA}$ ) [143], while LRRs 7-13 in BT\_1240 has very large  $R$  ( $=277 \text{ \AA}$ ). Even in the extreme cases the  $2 \cdot R \cdot \sin(\Delta\Phi/2)$  and  $\Delta z$  values fall on the circle in figure 3. In the former the helix axis is nearly perpendicular to the  $\beta$ -strand axis. In contrast, in the latter the helix axis is nearly parallel to the  $\beta$ -strand axis.

The helical parameters of the eight LRR classes frequently overlap one another (Figure 2B). The  $\Delta z$  and  $2 \cdot R \cdot \sin(\Delta\Phi/2)$  values in the “Typical” class are distributed broadly (Figure 3).

Structures of LRRs in GPIB $\beta$ , a GPIB $\beta$ /GPIX chimera, and ZP\_02034617.1 with  $n=3$  are available (Supplementary Table 1)



[144]. The circle fitting analysis yields  $D = 4.90 \text{ \AA} \pm 0.01 \text{ \AA}$  that seems to be smaller than in LRRs with  $n \geq 4$ . This smaller  $D$  might reflect interactions of side chains bridging N- and C-caps [144].

### Left handed helices

Most LRRs form right handed helices, that is  $\Delta z > 0.0 \text{ \AA}$ . However, left handed helices are adopted by LEGL7, CD14, Lmof2365\_1307, FSHR, TLR1, and TLR2, in which the repeat number  $n \geq 7$ ;  $\Delta z = -0.78 \rightarrow -3.17 \text{ \AA}$  (Figure 4 and supplementary Table 1).

### Helical dimers in CD14, RI, and LGR4

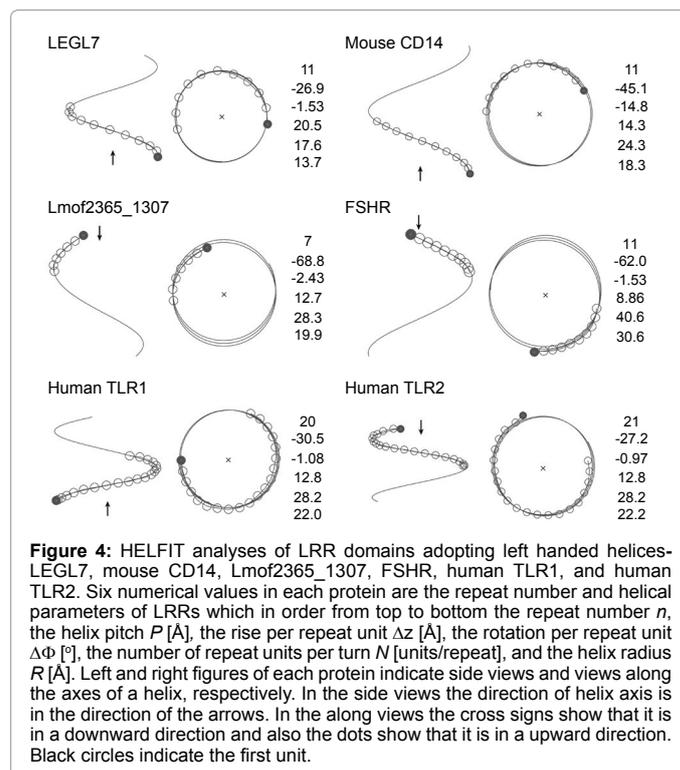
In the homo-dimers of some LRRs not only the individual monomer

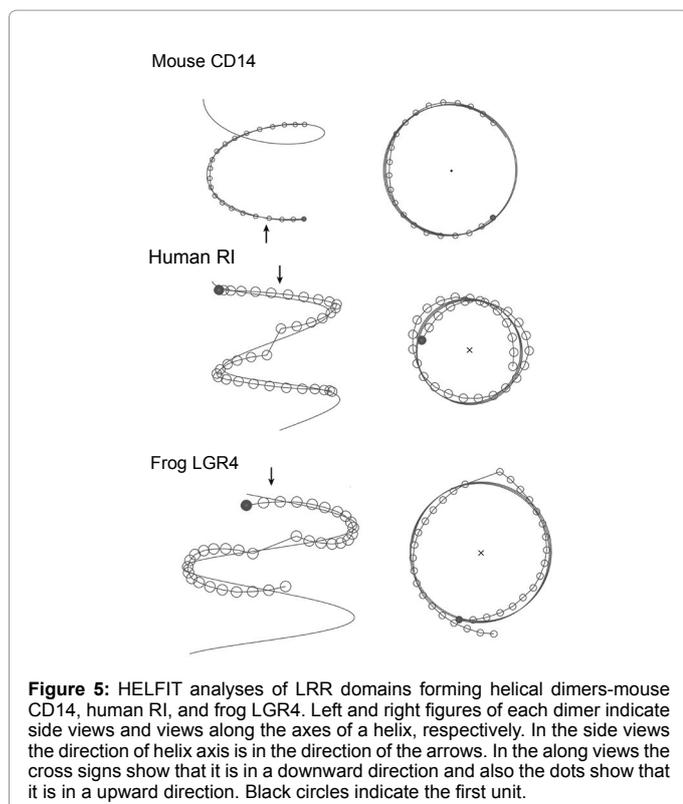
but also the entire dimer forms a single helix. We found four examples. Mouse CD14, in which the LRR does not conform to any of the eight classes, forms a dimer in the crystallographic asymmetric unit as well as in solution [139]. Dimerization in the crystal is mediated by LRR residues in the loop between  $\beta$ -strands in the HCS of repeats 12 and 13. The C-terminal  $\beta$ -strands of the two  $\beta$ -sheets from the two monomers interact in an antiparallel fashion and form a large and continuous  $\beta$ -sheet encompassing the entire CD14 dimer. The two monomers are related by a two fold axis of rotation, perpendicular to their common helix axis. The CD14 monomer forms a left handed helix, as does the CD14 dimer (Figure 5 and supplementary Table 1).

Human RI complexed with human angiogenin or ribonuclease I forms a dimer with human RI-angiogenin/ribonuclease I [36,37]. The whole molecule of human RI consists of only LRRs; all belong to "RI-like" class. The two complexes are held together by many hydrogen bonds formed between the N-terminal  $\beta$ -strands of the two human RI molecules, leading to the formation of an antiparallel  $\beta$ -sheet. The entire RI dimer adopts a right handed helix whose helical parameters are similar to those of its individual monomers. These two hRI molecules are related by an approximate two fold axis of rotation, nearly perpendicular to the common helical axis.

Mouse RI bound to mouse ribonuclease I forms a tetramer in the crystallographic asymmetric unit. The tetramer consists of two dimers that are related by a glide plane. The monomers of both dimers adopt a right handed helix whose parameters are quite similar to those in the dimer of human RI- angiogenin/ribonuclease I (Supplementary Table 1).

Frog LGR4, whose LRRs belong to the "RI-like" class, also forms a homo-dimer in the crystal and in solution [104]. The two monomers in the LGR4 dimer are in close proximity at their C-terminal sides and are related by a two fold axis of rotation-TLRs 1-6, TLR8, RP105, LGR4, LGR5, and Drosophila Toll, "TpLRR" class-FAEPRAA2165\_01021 and





EUBVEN\_01088, and “Variable” class-BT\_1240 and BAGEGG\_03329. Most of the  $p$  (error) values are larger than 0.20 (Supplementary Table 1). The angle between the two helix axes ( $\Omega$ ) best characterizes the discontinuity (Supplementary Table 2). The LRRs with two subdomains may be grouped into four categories.

The overall shape of the entire LRR in TLRs 1-6, TLR8, RP105, and BAGEGG\_03329, is well approximated as half or three quarters of an ellipse. For example, the LRRs in human TLR3 with  $n=25$  consist of two subdomains, LRRs 1-12 and LRRs 13-25 (Figure 6). The 25 units of TLR3 forms a nearly flat ellipse with  $p=0.23$  Å. This  $p$  value is larger than  $p=0.09$  Å of units 1-12 in subdomain 1 and of units 13-25 in subdomain 2. Similar characteristics are observed in other LRRs that consist of two subdomains. The angle,  $\Omega$ , between the two helix axes of the subdomains ranges from  $13.4 \rightarrow 19.2^\circ$  (Supplementary Table 2). This tilt generates a larger  $p$  for the entire LRRs.

AtTMK1, BRI1, BRL1, AtRKP2 and Drosophila Toll contain LRRs interrupted by a non-LRR, IR. For example, AtTMK1 contains 15 units interrupted by a non-LRR IR;  $n_1=11$  and  $n_2=4$  [75]. The two helix axes are nearly perpendicular to one another;  $\Omega=74^\circ$  (Figure 6 and Supplementary Table 2). The non-LRR IR has a cluster of four Cys residues with the pattern of  $Cx_6Cx_{29}Cx_7C$ . The formation of a disulfide bridge in  $Cx_7C$  suggests that the IR acts as an N-cap, further supporting the formation of the second subdomain. The first LRR domain in Drosophila Toll with  $n=19$  forms a half ellipse [95]. Therefore, the LRR domain is divided into two subdomains.

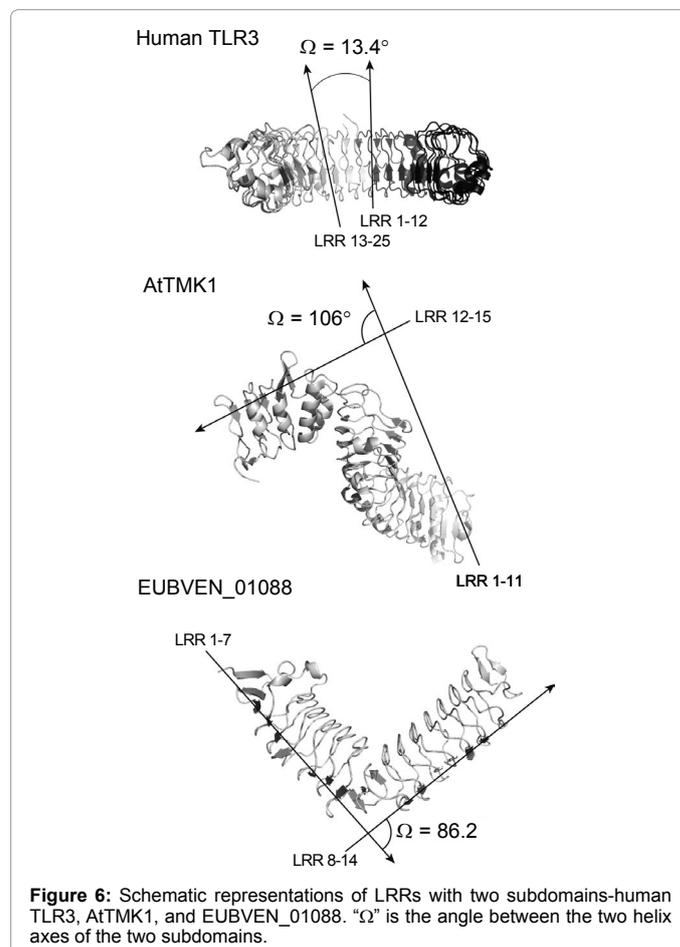
EUBVEN\_01088, FAEPRAA2165\_01021, and BT\_1240 contain LRRs in which the continuity of the parallel  $\beta$ -sheet is disrupted. The “TpLRR” domains in EUBVEN\_01088 and FAEPRAA2165\_01021 with  $n=14$  are kinked at the central seventh and eighth repeats with 35 and 37 residues, respectively;  $\Omega=71 \rightarrow 88^\circ$  (Figure 6 and supplementary Table 2). BT\_1240 contains 13 LRRs of which four repeats are similar to

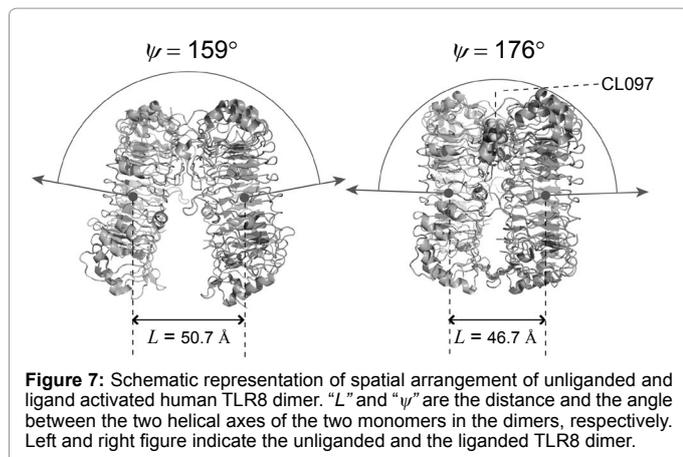
the “TpLRR” consensus. BT\_1240 forms a homo-dimer in the crystal, in which the two molecules are in close proximity at the C-terminal sides. The LRR unit 7 is 30 residues long and disrupts the continuity of the parallel  $\beta$ -sheet;  $\Omega=64 \rightarrow 70^\circ$  (Supplementary Table 2).

The LRR domains in human LGR4 and LGR5 with  $n=19$ , and frog LGR4 with  $n=17$  are kinked between the central eleventh and twelfth repeats [104-107];  $\Omega=19.1 \rightarrow 40.1^\circ$ . The units of LRR11 and of LRR12 do not strictly obey the  $LxxLxLxxNxL$  rule. Instead, the Asn residues are replaced by A309 and T332, respectively. As a result, the asparagine ladder (in which the Asn side chains from different turns are stacked and form hydrogen bond connecting turns) breaks in this region, thereby generating two longer  $\beta$ -strands [104-107].

### Homo-oligomers

Proteins that contain LRRs form homo-dimers, -trimers, -tetramers, -pentamers, -hexamers, and -octamers in crystals. The LRRs of these homo-oligomers are related by non-crystallographic symmetry and hence experience slightly different packing environments. There are two patterns. The first case is that the helical parameters of individual monomers constituting the homo-oligomers are similar to each other. In this case, their repeat numbers are large,  $n \geq 11$ . Examples include, homo-octamers-NLRC4 ( $n=16$ ), Skp1 ( $n=11$ ), RP105 ( $n=23$ ), AtTIR1 ( $n=18$ ), and AtCOI1 ( $n=18$ ); homo-hexamers-SspH1 ( $n=10$ ) and biglycan ( $n=12$ ); homo-pentamer-EUBVEN\_01088 14; homo-tetramers-RanGAP1 ( $n=11$ ), LINGO-1 ( $n=14$ ), Drosophila Toll ( $n=8$ ), and CARMIL ( $n=16$ ), and homo-dimers-NGL-3 ( $n=11$ ) and InlA





( $n=16$ ). In the second pattern the helical parameters differ significantly among the individual monomers. In this case, their repeat numbers are small;  $n=5$  or  $7$ . They include homo-tetramers-TAP,  $n=5$ , VLR2913,  $n=5$ , bovine coupling factor B,  $n=5$ , and human pp32,  $n=5$ ; the homo-hexamer-human pp32,  $n=5$ ; and the homo-octamer-InlK,  $n=7$ .

### Eight classes of LRR

When the repeat number,  $n$ , is  $\leq 8$ , the helical parameters for individual monomers of the homo-oligomers are sometimes variable. The helix parameters were not determined in these cases.

The consensus sequence of the “RI-like” class is  $LxxLxLxxNx(L/C)xxxgoxxLxxoLxxxxx$ . The repeat length is 28-29. Most of their VSs adopt an  $\alpha$ -helical conformation ( $\beta$ - $\alpha$  structural units) (Figure 1). “RI-like” LRR proteins include RI ( $n=17$ ), NLR4 ( $n=16$ ), CARMIL ( $n=16$ ), and NLRX1 ( $n=8$ ) [34-40]. The helix parameters range over:  $P=-6.3 \rightarrow 35.6 \text{ \AA}$ ,  $\Delta z=-0.21 \rightarrow 1.76 \text{ \AA}$ ,  $\Delta\Phi=11.4 \rightarrow 18.5^\circ$ ,  $N=19.5 \rightarrow 31.6$  units/repeat, and  $R=14.7 \rightarrow 24.8 \text{ \AA}$  (Supplementary Table 1).

The LRR domains of RanGAP from human and fungi,  $n=11$ , and tropomodulin from chicken and *Caenorhabditis elegans*,  $n=5$ , show little similarity to the RI LRRs [41-45]. However, their repeat units are in  $\beta$ - $\alpha$  conformations that are the same as those of the “RI-like” class. Correspondingly, their helical parameters are similar to those of the four “RI-like” LRR proteins.

The LRR domain in LEGL7 is similar to that in CARMIL (E-value= $6.2 \cdot 10^{-4}$  in FASTA). We tentatively assign the LEGL7 LRR to the “RI-like” class. The LRR domain consists of eleven units. The consensus is  $VxxLxLxxNxLxxxSxxELxxxLxIPxx$  with 29 residues. DSSP analysis indicates that the “ascending loops” in the eight repeats adopt a  $3_{10}$ -helix of one turn at the underlined residues and the VS adopts an  $\alpha$ -helix with 6  $\rightarrow$  10 residues. The LRR domain adopts a left-handed helix;  $P=-26.9 \text{ \AA}$ ,  $\Delta z=-1.53 \text{ \AA}$ ,  $\Delta\Phi=20.5^\circ$ ,  $N=17.6$  units/repeat, and  $R=13.7 \text{ \AA}$  (Supplementary Table 1).

“Cysteine-containing” LRR proteins include F-box/LRR-repeat protein 20 from *Rattus norvegicus*, grr1 from *Saccharomyces cerevisiae* and grrA from *Emericella nidulans*. The consensus sequence is  $LxxLxLxxCxxITDxxoxL(a/g)xx(C/L)xx$ . Tertiary structures are available for Skp2 with  $n=8$  or  $11$  [46,47]. The units of Skp2 are relatively variable; they are in  $\beta$ - $\alpha$  conformations (Figure 1). The helix parameters range over:  $P=28.1 \rightarrow 40.8 \text{ \AA}$ ,  $\Delta z=1.17 \rightarrow 1.75 \text{ \AA}$ ,  $\Delta\Phi=15.0 \rightarrow 16.2^\circ$ ,  $N=22.3 \rightarrow 23.4$  units/repeat, and  $R=16.6 \rightarrow 18.9 \text{ \AA}$  (Supplementary Table 1).

The “Cysteine-containing” LRR domain in Skp2 is similar to those

in AtTIR1 and AtCOI1 [48-50]. For both,  $n=18$ . The HCS of the LRR domains in AtTIR1 and in AtCOI1 are highly variable because the CxxI sequence is not conserved in most of these LRRs. In contrast, the VSs are conserved. We, therefore, assume that both the LRR domains belong to the “Cysteine-containing” class. The variable HCSs form relatively long  $\beta$ -strands with three to six residues and the VSs form  $\alpha$ -helices with twelve residues. The  $p$  values are large:  $p=0.33 \rightarrow 0.42 \text{ \AA}$  and thus the right handed helix is distorted. The helix parameters range over:  $P=19.3 \rightarrow 20.4 \text{ \AA}$ ,  $\Delta z=1.02 \rightarrow 1.05 \text{ \AA}$ ,  $\Delta\Phi=18.5 \rightarrow 19.3^\circ$ ,  $N=18.7 \rightarrow 19.1$  units/repeat, and  $R=14.9 \rightarrow 15.3 \text{ \AA}$ .

The consensus sequence of the “SDS22-like” class is  $LxxLxLxxNxIxxIxxLxxLxx$ . The repeat length is 21-23. The VSs strongly prefer a  $3_{10}$ -helix at the underlined residues. The individual units are in  $\beta$ - $3_{10}$  conformation (Figure 1). “SDS22-like” LRR proteins include five distinct internalins (except for InI) with  $n=6 \rightarrow 16$  [51-68]. The “SDS22-like” LRR domains adopt a right-handed helix:  $P=6.27 \rightarrow 73.7 \text{ \AA}$ ,  $\Delta z=0.23 \rightarrow 2.91 \text{ \AA}$ ,  $\Delta\Phi=11.8 \rightarrow 15.1^\circ$ ,  $N=23.9 \rightarrow 30.7$  units/repeat, and  $R=17.4 \rightarrow 22.8 \text{ \AA}$  (Supplementary Table 1).

The LRR domain in InlK shows high similarity to that in Lmof2365\_1307. These two LRR domains show significant similarity to those in InlA and internalin C2 (with E-values  $<10^{-4}$  in FASTA). We tentatively assume that the LRR domain in InlK and Lmof2365\_1307 belong to the “SDS22-like” LRR class. Three variable HCSs at the N-terminal side adopt long  $\beta$ -strands with five or eleven residues. These putative “SDS22-like” domains adopt a left-handed helix;  $P=-62.9 \rightarrow -97.4 \text{ \AA}$ ,  $\Delta z=-1.90 \rightarrow -3.32 \text{ \AA}$ ,  $\Delta\Phi=10.9 \rightarrow 14.8^\circ$ ,  $N=25.8 \rightarrow 33.1$  units/repeat, and  $R=12.9 \rightarrow 24.7 \text{ \AA}$ .

The consensus sequence of the “IRREKO” class is  $LxxLxLxxNxLxxLDLxx(N/L/Q/x)xx$  or  $LxxLxCxxNxLxxLDLxx(N/L/x)xx$  [21]. The only available 3D structure is that of InI [69]. InI with  $n=15$  belongs to this class; although, the first and the last units are “SDS22-like”. The VS of TxLDL(T/S)x(N/L)Tx adopt an extended  $\beta$  conformation and  $\beta$ -turns (Figure 1). Its helical parameters are  $P=73.3 \text{ \AA}$ ,  $\Delta z=2.80 \text{ \AA}$ ,  $\Delta\Phi=13.7^\circ$ ,  $N=26.2$  units/repeat, and  $R=17.7 \text{ \AA}$  (Supplementary Table 1).

The consensus sequence of the “Bacterial” class is  $LxxLxVxxNxLxxLP(D/E)LPxx$ . The repeat length is 20-22. The structural units are in  $\beta$ -polyproline II helix conformation (Figure 1). “Bacterial” LRR proteins include YopM ( $n=16$ ), SspH2 ( $n=13$ ), SspH1 ( $n=10$ ), and ipaH3 ( $n=9$ ) [147-150]. The helix parameters range over:  $P=46.7 \rightarrow 103 \text{ \AA}$ ,  $z=1.64 \rightarrow 2.61 \text{ \AA}$ ,  $\Delta\Phi=8.71 \rightarrow 12.7^\circ$ ,  $N=28.4 \rightarrow 36.0$  units/repeat, and  $R=20.6 \rightarrow 24.6 \text{ \AA}$  (Supplementary Table 1).

“Plant-specific” LRR proteins include protein FLS2, BRI1, BRL1, BAK1, SERK1, AtTMK1, AtRPK2, and PGIP [70-77]. BRI1 and BRL1 are homologs. BAK1 and SERK1 are also homologs. The consensus of the “plant-specific” class is  $xxLxLxxNxLxGxLPxxLxxLxx$  with 24 residues [22,151]. The AtTMK LRR domain with  $n=15$  is kinked at the central eleventh and twelfth repeats [75]. The consensus sequence, LxGxLP, at positions 11 to 16 forms a second parallel  $\beta$ -strand in the “ascending loop”. Thus, the structural unit is  $\beta$ - $\beta$ - $3_{10}$  (Figure 1). The helix parameters with  $n=9 \rightarrow 29$  range over:  $P=57.2 \rightarrow 87.0 \text{ \AA}$ ,  $\Delta z=2.39 \rightarrow 3.37 \text{ \AA}$ ,  $\Delta\Phi=13.7 \rightarrow 15.0^\circ$ ,  $N=24.0 \rightarrow 26.2$  units/repeat, and  $R=15.2 \rightarrow 17.5 \text{ \AA}$  (Supplementary Table 1).

“Typical” LRRs are the most abundant LRR class [6,20]. The consensus sequence is  $LxxLxLxxNxLxxLpxxxFxxLxx$  in which uppercase indicates more than 50% occurrence of a given residue in a certain position; lowercase indicates 30-50% occurrence; “L” is Leu, “N” is Asn, “p” is Pro, “o” indicates a non-polar residue, “F” are Phe, and “x”

is a non-conserved residue. "Typical" LRR proteins include 53 proteins (Supplementary Table 1) [71, 78-137].

Most of their VSs adopt a tandem arrangement of three  $\beta$ -turns (which we call consecutive -turns) on the convex faces (Figure 1). The consecutive -turns form a flattish amphipathic structure with main chain hydrogen bonds in a linear arrangement. Adjacent  $\beta$ -turn repeat motifs form a regular parallel packing arrangement with the core hydrophobic residues pointing alternately in and out of the plane of the structure interlocking with the corresponding hydrophobic residues from the adjacent repeat. The consecutive -turns have already been reported in the underlined eight residues of the VS consensus of xxLPxGLLxGLxx seen in three LRR units in GPIb $\alpha$  with  $n=9$  [125]. Thus, the structural units may be represented by consecutive -turns.

Phenylalanine at position 19 is internally buried and is surrounded by four Leu's, which form the hydrophobic core of the LRR structure. These Phe's of consecutive units form the spine of the LRR [89]. This Phe spine involves aromatic, aromatic interactions; the benzyl groups form stacks of Phe side chains. The first, N-terminal LRR unit in most of the "Typical" LRR domains is "Bacterial" LRR. The VS adopts the polyproline II helix instead of consecutive  $\beta$ -turns.

The helix parameters with  $n=7 \rightarrow 27$  range broadly:  $P=62.0 \rightarrow 93.2$  Å,  $\Delta z=1.53 \rightarrow 3.24$  Å,  $\Delta\Phi=7.41 \rightarrow 15.0^\circ$ ,  $N=23.9 \rightarrow 51.2$  units/repeat, and  $R=15.5 \rightarrow 38.9$  Å (Supplementary Table 1). The LRR domains may be grouped into a right handed helix, a left-handed helix, and a near flat.

Human Slit2 contains four tandem arrays (D1 to D4) of LRRs that consist of six to eight units. The structures of D2,  $n=7$ , D3,  $n=7$ , and D4,  $n=6$ , have been determined in the free state [91-93]. The helical parameters differ significantly from each other;  $D2>D4>D3$  in  $P$ ;  $D4>D2>D3$  in both  $\Delta z$  and  $\Delta\Phi$ , and  $D3>D2>D4$  in both  $N$  and  $R$  (Supplementary Table 1).

The LRR domains in biglycan, decorin, and TLR8 consist of three or four tandem repeats of a super motif of **STT** in which "S" is a "Bacterial" LRR unit and "T" is a "Typical" LRR unit [23,24]. The LRR domains of biglycan and decorin with  $n=12$  consist of four repeats of **STT**. Also the N-terminal LRR domain in TLR8,  $n=27$ , contains three repeats of **STT**. The helix parameters range over:  $P=36.7 \rightarrow 51.8$  Å,  $\Delta z=0.82 \rightarrow 1.84$  Å,  $\Delta\Phi=10.5 \rightarrow 14.3^\circ$ ,  $N=25.3 \rightarrow 34.3$  units/repeat, and  $R=19.0 \rightarrow 26.3$  Å.

The LRR domain in Xcv3220,  $n=9$ , has high sequence similarity to those in both human LGR4 with E-value= $7.7 \cdot 10^{-6}$  in FASTA and human LGR5 with E-value= $7.0 \cdot 10^{-6}$  in BLAST. We assume that this Xcv3220 LRR domain belongs to the "Typical" class. Seven of the nine repeats are represented by the consensus of LxxLxLxxCxx(x/-)LxxLPxxLxxLxx with 23-24 residues. The structural unit is  $\beta$ -3<sub>10</sub>. The LRR domain adopts a left-handed helix;  $P=22.3$  Å,  $z=0.59$  Å,  $\Delta\Phi=-9.6^\circ$ ,  $N=37.6$  units/repeat, and  $R=29.7$  Å.

The consensus sequence of the "TpLRR" class is LxxLxLxxxLxxIgxxAFxx(C/N)xx. The repeat length is 23-25. Bacterial "TpLRR" proteins include BACOVA\_04585, bacterial group 3 Ig-like proteins (BACCAC\_03700 and BACOVA\_01565), FAEPRAA2165\_01021, EUBVEN\_01088, and BT\_1240. The repeat numbers range from 12 to 14. The tertiary structures of the six "TpLRR" proteins have been determined. The ascending loops consist mainly of one  $\beta$ -turn and short  $\beta$ -strand of two residues, while the most of the descending loops consist of a single  $\beta$ -turn (Figure 1). Most of their VSs adopt a tandem arrangement of two or three  $\beta$ -turns.

The helix parameters are variable among the six proteins and

deviate significantly from those of the other classes (Supplementary Table 1). This can be attributed to a larger twist angle of  $\beta$ -sheets, to a breaking of  $\beta$ -strands in the HCS, or to a prism like shape for the LRRs. The mean twist angle of  $\beta$ -sheets in LRR proteins belonging to the other classes was  $3 \rightarrow 8^\circ$  [32]. LRRs in the "TpLRR" class sometimes do not form  $\beta$ -strands in the HCS.

Most of LRRs in BT\_0210,  $n=18$ , BAGEGG\_03329,  $n=18$ , and LRIM1,  $n=11$ , are highly variable and could be assigned to none of the eight classes [134, 138-144]. Their helix parameters range over:  $P=66.2 \rightarrow 96.2$  Å,  $\Delta z=1.66 \rightarrow 2.88$  Å and  $\Delta\Phi=8.5 \rightarrow 10.8^\circ$ ,  $N=33.4 \rightarrow 42.1$  units/repeat, and  $R=22.0 \rightarrow 31.8$  Å (Supplementary Table 1). These parameters are most similar to those in the "Bacterial" domains.

### Structural change induced by protein, protein interactions, glycosylation and/or mutation

In the LRRs the concave surface, the convex surface, N-terminal domain including N-cap, C-terminal domain including C-cap, or their combinations are involved in protein, protein interactions (Supplementary Table 3).

The LRRs of VLR and Slit2 belong to the "Typical" class, as noted. VLRB.2D,  $n=5$ , bound to a protein antigen, HEL, as well as non-bound VLR has been determined [83]. VLR uses nearly its entire concave surface to bind HEL. This binding causes large changes of the helical parameters; a decrease of  $P$ ,  $\Delta z$ , and  $\Delta\Phi$ , and an increase in  $R$  (Supplementary Table 3). The structures of D3 in human Slit2 have been determined in the free state and in the complex with the Ig domain from Robo1 [91]. As seen in the VLRB.2D-HEL complex, the Slit2 D3 domain uses its concave surface to bind Ig. The formation of the complex causes large changes of the helical parameters. Conversely, an increase of  $P$ ,  $\Delta z$ , and  $\Delta\Phi$ , and a decrease in  $R$  is observed (Supplementary Table 3). The same structural change occurs in the In1A-cadherin complex; although, it is minor (Supplementary Tables 1 and 3) [56].

The crystal structures of free LRIM1 and of free APL1 and of their hetero-dimer complex have been determined [134]. The LRIM1/APL1 complex has a single intermolecular disulfide bond. LRIM1 and APL1 contain 11 and 15 repeat units, respectively. The HELFIT analysis detects structural change in both LRR domains due to the complex formation. The LRIM1 LRR domain in the complex shows a significant increase in  $P$  and  $\Delta z$  and decrease in  $R$  (Supplementary Table 3). In contrast, the helical parameters in the APL1 LRR domain are reversed and have small changes in  $\Delta z$  and  $R$ .

The structures of frog LGR4 have been determined in the free state and in complex with R-spondin [104]. The formation of the complex induces small structural changes and increases in  $P$ ,  $\Delta z$ , and  $R$  (Supplementary Table 3).

The structures of free RabGGT $\alpha$  and its complex bound to escort protein 1 have been determined [53,54]. The helical handedness of the LRR domain with  $n=6$  reverses upon binding to REP (Supplementary Table 1).

The crystal structures of the complexes of Skp1-Skp2 and Skp1-Skp2-Cks1 are available [46,47]. The binding of Cks1 causes slight structural changes; an increase of  $P$ ,  $\Delta z$ , and  $\Delta\Phi$ , and a decrease in  $R$ . Moreover, the crystal structures of the complexes with Ran-GPPNHP-RanBP1-RanGAP and Ran-GDP-AIFx-RanBP1-RanGAP have been determined [43]. The helical parameters of these two complexes differ slightly from one another.

Many structures of GPIb $\alpha$  have been reported in the free state and

in complexes with one or two  $\alpha$ -thrombins, von Willebrand Factor A1, and PEP inhibitor [124-131]. We do not find a clear pattern of structural changes. However, it appears that the binding of two  $\alpha$ -thrombins causes a structural change.

The crystal structures of NgR-4,  $n=10$ , in both the non-glycosylated state and N-glycosylated at Asn-82 and Asn-179, are available [88,89]. N-glycosylation induces a decrease of  $P$ ,  $\Delta z$ , and  $\Delta\Phi$  and an increase of  $N$  and of  $R$  (Supplementary Table 3). The G28E mutant of bovine coupling factor B,  $n=5$ , has a smaller  $R$  than does the wild type (Supplementary Table 1) [152].

### Spatial arrangement of TLR dimers

The structures of the unliganded and ligand activated human TLR8 dimer were determined [119]. Helix parameters of the two monomers in the unliganded dimer are generally similar to those in the liganded dimer. However, upon ligand binding the spatial arrangement of the two monomers in the dimer changes. The distance between the two monomers at their C-terminal regions changes from 53 Å to 30 Å [119]. HELFIT calculates the distance and the angle between the two helical axes of the two monomers in the dimers,  $L$  and  $\psi$ . Correspondingly, the ligand binding drastically changes these two parameters;  $\Delta L=-2.6$  Å and  $\Delta\psi=17^\circ$  (Figure 7).

The structures of TLR3 complexed with dsRNA and with six different Fabs (Fab15 light and heavy chains, Fab12 light and heavy chains, and Fab1068 light and heavy chains) have been determined, as well as the structure of free TLR3 [112,113]. The formations of the complexes induce decreases in the helix radii. The dsRNA TLR3 dimer has  $L=74.5$  Å and  $\psi=148^\circ$ .

Furthermore, the hetero-dimers of human TLR1 and TLR2, and of mouse TLR2 and TLR6 have been determined [108,109]. The values of  $L$  and  $\psi$  are 44.0 Å and 168° in the TLR1-TLR2 dimer and 53.0 Å and 169° in the TLR2-TLR6 dimer, respectively.

### Discussion

#### The constant inter-strand distance is the primary determinant of the helical parameters of the LRRs

The HELFIT analysis demonstrates that the inter-strand distance ( $D$ ) is constant in the seven classes except for the “TpLRR” class;  $D=5.02 \pm 0.00$  Å (Figure 2B). In contrast,  $D$  in the “TpLRR” class is significantly larger;  $D=5.52 \pm 0.14$  Å. This can be attributed to non-uniformity of the inter-strand distance. The helix radius  $R$  in some LRRs belonging to the “TpLRR” class is extremely small or very large (Supplementary Table 1). Thus, the overall shape of these LRRs of the “TpLRR” class is best described as a prism; while those of the other classes are helices or semi-ellipses.

#### The “ascending loops” influence the helical parameters of the LRRs

The LRRs in RI and LEGL7 belong to the “RI-like” class. The RI LRR adopts a right handed helix or a horse shoe shape. In contrast, the LEGL7 LRR adopts a left handed helix (Figure 4). The LRRs form  $\beta$ -strands on the concave face (in the HCS) and  $\alpha$ -helix of 9  $\rightarrow$  15 residues on the convex face (in the VS). Both of the “descending loops” also consist of one  $\beta$ -turn. On the other hand, the ascending loops of the LEGL7 LRR include a 310-helix of one turn; while the RI LRR does not adopt this helix in the loops.

Some LRRs in NLRC4, InlK, Lmof2365\_1307, AtTIR1, AtCOI1,

TLRs 1-6, and TLR8 do not obey the LxxLxLxxNxL consensus, especially at the underlined residues. In addition, some LRRs in these proteins frequently form longer  $\beta$ -strands consisting of four to seven residues in the HCS. This affects not only the ascending loops but also the descending loops. Indeed, the right handed helices of NLRC4, AtTIR1, AtCOI1, TLRs 1-6 and TLR8 are distorted. Also the LRRs in InlK and Lmof2365\_1307, which belong to the “SDS22-like” class, adopt left handed helices; while those in the normal “SDS22-like” class form right handed helices.

The HCS of the “TpLRR” class with ten residues is shorter than eleven or twelve residues in the “Typical” class. On the other hand, the VS of the “TpLRR” class is quite similar to that of the “Typical” class. Both VSs consist mainly of consecutive  $\beta$ -turns (Figure 1). However, the helical parameters indicate that the “Typical” class adopts a helix, while the “TpLRR” class resembles a prism. These observations can be attributed to a structural difference of the ascending loops and the inter-strand distance.

The helix parameters vary among the “Typical” LRRs. Seven of the ten repeats in NgR-4 and eleven of the fourteen repeats LINGO-1 conform to the “Typical” consensus. However, their structures differ from one another. The NgR-4 LRR domain adopts a right handed helix with  $\Delta z=2.09$  Å; while the LINGO-1 LRR arc is nearly flat;  $|\Delta z|=0.06$  Å. The LINGO-1 LRRs show a structural periodicity that consists of four tandem repeats of three consecutive LRR units. The first and second VSs in the three LRR units adopt a  $\beta$ -strand plus consecutive  $\beta$ -turns, while in the last VS the corresponding  $\beta$ -strand breaks. In contrast, this periodicity is not seen in the corresponding residues in NgR-4. The different helical parameters may be due to a difference in structure of the ascending loops and the conformations on the convex face.

In conclusion, the above observations suggest that the helical parameters more strongly depend on the structures of the “ascending loops” than of the “descending loops”. Moreover, the helical parameters of the LRRs are influenced by helical elements on the convex face and the uniformity of parallel strand stacking on the concave face.

#### LRR domains having small repeat number are highly flexible

The values  $R$  and  $\Delta z$  versus  $n$  for all 642 LRRs show a lot of variation, if  $n$  is small (Figure 3). Here we describe some examples. The first is the family of VLR. The lysozyme bound VLRB.2D forms a tetramer in the crystal [83]. The two monomers adopt a right handed helix, while the other two monomers adopt a left handed helix. Sea lamprey VLRB,  $n=5$ , bound to the BcIA protein forms a trimer [84]. One monomer forms a right handed and the other monomer forms a left handed helix. The third monomer is not visible in the electron density map of the crystal structure. Tetrameric VLR 2913 has a range of helical parameters in its four monomers (Supplementary Table 1).

The crystal structures of InlB,  $n=8$ , in wild type, mutant, and complexes have been determined [60-64,85]. The InlB molecules form a monomer, dimer, and hexamer in the crystals. The helical parameters show a range of variation. InlB with an additional LRR unit inserted,  $n=9$ , forms a trimer. All the monomers have similar helical parameters.

Right handed and left handed helices exist in the dimer of SERK1,  $n=5$ , bound to BRI1 [71] and in the tetramer of bovine Coupling Factor B,  $n=5$  [152]. InlK,  $n=7$ , forms a homo-octamer in the crystal [68]. The helix parameters of these eight monomers are highly variable (Supplementary Table 1). Even with  $n=17$  in RI, the helix parameters are significantly different among four vertebrate species: both  $P$  and  $\Delta z$  decrease in order of pig, bovine, human, and mouse [34-37].

These observations strongly suggest that LRR domains with  $n \leq 7$  are more variable and are more flexible in solution. NMR data of LC1 with  $n=6$  support this inference; within the  $\beta$ -sheets the consensus leucine residues at position 4 have relatively low order parameters [52].

### Structural features of individual LRR classes

As noted, the helical parameters of the eight LRR classes frequently overlap one another. Moreover, when the repeat number in LRRs is small, the LRR structures are more variable. Taken together, the helical parameters of the LRRs with  $n \geq 10$  were compared to find the structural features (Figure 3B and supplementary Figure 1). The “TpLRR” class deviates significantly from other classes. The “Typical” class has highly variable values. The rise per subunit,  $\Delta z$ , of the “RI-like” and “Cysteine-containing” classes is clearly smaller than those of the “SDS22-like”, and “Plant-specific” classes. This is mainly ascribed to the difference in the structural units.

### Structures of non-LRR, IRs

The structures of non-LRR, IRs may be considered in two groups. In the first group, IRs allow the continuity of the parallel  $\beta$ -sheet and the LRR units to form a single domain of a regular LRR structure. Examples include the IRs of BRI1, AtRPK2, BT\_0210, and TLR8.

The BRI1 LRR domain has 25 units interrupted by a 70 residue IR between unit 21 and unit 22 [28]. B. thetaitoamicro BT\_0210 contains 18 units interrupted by a 76-residue IR between units 2 and 3. In the two LRR domains the helical handedness reverses (Supplementary Table 1). The IR forms a small domain those folds back into the interior of the helix, where it has extensive polar and hydrophobic interactions. The IR between repeat units 15 and 16 of TLR8 is 39 residues long [24]. The additional residues loop out from the expected helical path of a regular LRR before rejoining it some residues later [119]. Consequently, this IR is in the exterior of the helix, in contrast to the IRs in BRI1 and BT\_0210.

In the second case IRs disrupt the continuity of the parallel  $\beta$ -sheet and form two distinct subdomains. This case is observed in AtTMK1 [75] and *Drosophila* Toll [95]. The AtTMK1 LRR domain contains 15 repeats interrupted by a 41 residue IR between LRR11 and LRR12 [28], while the Toll LRR domain with 24 repeats has a 68 residue IR between LRR10 and LRR11 [95].

### Conclusions

We have assigned helical parameters to the 642 LRRs of known structure using the program HELFIT. These parameters and their correlations with class of LRR, with the number of repeat units in the LRR, and with oligomerization and ligand state of the LRR are described. The helical parameters of the eight LRR classes frequently overlap one another. However, the constant distance between parallel  $\beta$ -strands is the primary determinant of the helical parameters of the LRRs. When the repeat number,  $n$ , is small, the LRR structures are more variable and, by inference, more flexible. In the LRRs with  $n \geq 10$ , rise per repeat unit,  $\Delta z$ , of the “RI-like” and “Cysteine-containing” classes are smaller than those of “SDS22-like”, and “Plant-specific” classes. This is ascribed to the difference in the structural units. The helical parameters of the LRRs unambiguously define both right handed and left handed helices, and helical dimers and subdomains if they exist. Moreover, the helical parameters sensitively detect structural changes induced by protein, protein interactions, glycosylation, and/or mutation.

### Acknowledgements

We are grateful to Tuvdendorj Nomin-Erdene for her help of the HELFIT

analysis. We also thank Chai JiJie of Tsinghua University for his providing of the PDB file of AtRPK2.

### References

1. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290-301.
2. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40: D302-305.
3. Sigrist CJ, de Castro E, Cerutti L, Cucho BA, Hulo N, et al. (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* 41: D344-347.
4. Burge S, Kelly E, Lonsdale D, Mutowo-Mueller P, McAnulla C, et al. (2012) Manual GO annotation of predictive protein signatures: the InterPro approach to GO curation. *Database (Oxford)* 2012: bar068.
5. Kobe B, Deisenhofer J (1994) The leucine-rich repeat: a versatile binding motif. *Trends Biochem Sci* 19: 415-421.
6. Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* 11: 725-732.
7. Matsushima N, Enkhbayar P, Kamiya M, Osaki M, and Kretsinger R (2005) Leucine-Rich Repeats (LRRs): Structure, Function, Evolution and Interaction with Ligands. *Drug Design Rev* 2: 305-322.
8. Matsushima N, Tachi N, Kuroki Y, Enkhbayar P, Osaki M, et al. (2005) Structural analysis of leucine-rich-repeat variants in proteins associated with human diseases. *Cell Mol Life Sci* 62: 2771-2791.
9. Bella J, Hindle KL, McEwan PA, Lovell SC (2008) The leucine-rich repeat structure. *Cell Mol Life Sci* 65: 2307-2333.
10. Kofoed EM, Vance RE (2012) NAIPs: building an innate immune barrier against bacterial pathogens. NAIPs function as sensors that initiate innate immunity by detection of bacterial proteins in the host cell cytosol. *Bioessays* 34: 589-598.
11. Ng A, Xavier RJ (2011) Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity. *Autophagy* 7: 1082-1084.
12. de Wit J, Hong W, Luo L, Ghosh A (2011) Role of leucine-rich repeat proteins in the development and function of neural circuits. *Annu Rev Cell Dev Biol* 27: 697-729.
13. McPhee JB, Mena P, Zhang Y, Bliska JB (2012) Interleukin-10 induction is an important virulence function of the *Yersinia pseudotuberculosis* type III effector YopM. *Infect Immun* 80: 2519-2527.
14. Furlan G, Klinkenberg J, Trujillo M (2012) Regulation of plant immune receptors by ubiquitination. *Front Plant Sci* 3: 238.
15. Ho DN, Coburn GA, Kang Y, Cullen BR, Georgiadis MM (2002) The crystal structure and mutational analysis of a novel RNA-binding domain found in the human Tap nuclear mRNA export factor. *Proc Natl Acad Sci U S A* 99: 1888-1893.
16. Jiang J, Zhang C, Wang X (2013) Ligand perception, activation, and early signaling of plant steroid receptor brassinosteroid insensitive 1. *J Integr Plant Biol* 55: 1198-1211.
17. Tör M, Lotze MT, Holton N (2009) Receptor-mediated signalling in plants: molecular patterns and programmes. *J Exp Bot* 60: 3645-3654.
18. Steinwand BJ, Kieber JJ (2010) The role of receptor-like kinases in regulating cell wall function. *Plant Physiol* 153: 479-484.
19. Bernoux M, Ellis JG, Dodds PN (2011) New insights in plant immunity signaling activation. *Curr Opin Plant Biol* 14: 512-518.
20. Kajava AV (1998) Structural diversity of leucine-rich repeat proteins. *J Mol Biol* 277: 519-527.
21. Matsushima N, Miyashita H, Mikami T, Kuroki Y (2010) A nested leucine rich repeat (LRR) domain: the precursor of LRRs is a ten or eleven residue motif. *BMC Microbiol* 10: 235.
22. Kajava AV, Anisimova M, Peeters N (2008) Origin and evolution of GALA-LRR, a new member of the CC-LRR subfamily: from plants to bacteria? *PLoS One* 3: e1694.
23. Matsushima N, Ohyanagi T, Tanaka T, Kretsinger RH (2000) Super-motifs and evolution of tandem leucine-rich repeats within the small proteoglycans-biglycan, decorin, lumican, fibromodulin, PRELP, keratan, osteoadherin, epiphygan, and osteoglycin. *Proteins* 38: 210-225.

24. Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, et al. (2007) Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* 8: 124.
25. Miyashita H, Kuroki Y, Matsushima N (2014) Novel leucine rich repeat domains in proteins from unicellular eukaryotes and bacteria. *Protein Pept Lett* 21: 292-305.
26. Fritz-Laylin LK, Krishnamurthy N, Tör M, Sjölander KV, Jones JD (2005) Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. *Plant Physiol* 138: 611-623.
27. Matsushima N, Mikami T, Tanaka T, Miyashita H, Yamada K, et al. (2009) Analyses of non-leucine-rich repeat (non-LRR) regions intervening between LRRs in proteins. *Biochim Biophys Acta* 1790: 1217-1237.
28. Matsushima N, Miyashita H (2012) Leucine-rich Repeat (LRR) domains containing intervening motifs in plants. *Biomeolecules* 2: 288-311.
29. Enkhbayar P, Hikichi K, Osaki M, Kretsinger RH, Matsushima N (2006) 3(10)-helices in proteins are parahelices. *Proteins* 64: 691-699.
30. Enkhbayar P, Damdinsuren S, Osaki M, Matsushima N (2008) HELFIT: Helix fitting by a total least squares method. *Comput Biol Chem* 32: 307-310.
31. Enkhbayar P, Matsushima N (2012) HELFIT: Algorithm and applications. *AIP Conf. Proc* 1479: 1942-1947.
32. Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, Matsushima N (2004) Structural principles of leucine-rich repeat (LRR) proteins. *Proteins* 54: 394-403.
33. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577-2637.
34. Kobe B, Deisenhofer J (1995) A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* 374: 183-186.
35. Kobe B, Deisenhofer J (1996) Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. *J Mol Biol* 264: 1028-1043.
36. Johnson RJ, McCoy JG, Bingman CA, Phillips GN Jr, Raines RT (2007) Inhibition of human pancreatic ribonuclease by the human ribonuclease inhibitor protein. *J Mol Biol* 368: 434-449.
37. Papageorgiou AC, Shapiro R, Acharya KR (1997) Molecular recognition of human angiogenin by placental ribonuclease inhibitor--an X-ray crystallographic study at 2.0 Å resolution. *EMBO J* 16: 5162-5177.
38. Hu Z, Yan C, Liu P, Huang Z, Ma R, et al. (2013) Crystal structure of NLR4 reveals its autoinhibition mechanism. *Science* 341: 172-175.
39. Zwolak A, Yang C, Feeser EA, Ostap EM, Svitkina T, et al. (2013) CARMIL leading edge localization depends on a non-canonical PH domain and dimerization. *Nat Commun* 4: 2523.
40. Hong M, Yoon SI, Wilson IA (2012) Structure and functional characterization of the RNA-binding element of the NLRX1 innate immune modulator. *Immunity* 36: 337-347.
41. Hillig RC, Renault L, Vetter IR, Drell T 4<sup>th</sup>, Wittinghofer A, et al. (1999) The crystal structure of rna1p: a new fold for a GTPase-activating protein. *Mol Cell* 3: 781-791.
42. Hillig RC, Renault L (2006) Detecting and overcoming hemihedral twinning during the MIR structure determination of Rna1p. *Acta Crystallogr D Biol Crystallogr* 62: 750-765.
43. Seewald MJ, Körner C, Wittinghofer A, Vetter IR (2002) RanGAP mediates GTP hydrolysis without an arginine finger. *Nature* 415: 662-666.
44. Krieger I, Kostyukova A, Yamashita A, Nitani Y, Maéda Y (2002) Crystal structure of the C-terminal half of tropomodulin and structural basis of actin filament pointed-end capping. *Biophys J* 83: 2716-2725.
45. Lu S, Symersky J, Li S, Carson M, Chen L, et al. (2004) Structural genomics of *Caenorhabditis elegans*: crystal structure of the tropomodulin C-terminal domain. *Proteins* 56: 384-386.
46. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, et al. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* 408: 381-386.
47. Hao B, Zheng N, Schulman BA, Wu G, Miller JJ, et al. (2005) Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. *Mol Cell* 20: 9-19.
48. Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, et al. (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* 446: 640-645.
49. Hayashi K, Tan X, Zheng N, Hatate T, Kimura Y, et al. (2008) Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling. *Proc Natl Acad Sci U S A* 105: 5632-5637.
50. Sheard LB, Tan X, Mao H, Withers J, Ben-Nissan G, et al. (2010) Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* 468: 400-405.
51. Wu H, Maciejewski MW, Marintchev A, Benashski SE, Mullen GP, et al. (2000) Solution structure of a dynein motor domain associated light chain. *Nat Struct Biol* 7: 575-579.
52. Wu H, Blackledge M, Maciejewski MW, Mullen GP, King SM (2003) Relaxation-based structure refinement and backbone molecular dynamics of the dynein motor domain-associated light chain. *Biochemistry* 42: 57-71.
53. Zhang H, Seabra MC, Deisenhofer J (2000) Crystal structure of Rab geranylgeranyltransferase at 2.0 Å resolution. *Structure* 8: 241-251.
54. Pylypenko O, Rak A, Reents R, Niculae A, Sidorovitch V, et al. (2003) Structure of Rab escort protein-1 in complex with Rab geranylgeranyltransferase. *Mol Cell* 11: 483-494.
55. Price SR, Evans PR, Nagai K (1998) Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 394: 645-650.
56. Schubert WD, Urbanke C, Ziehm T, Beier V, Machner MP, et al. (2002) Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* 111: 825-836.
57. Wollert T, Heinz DW, Schubert WD (2007) Thermodynamically reengineering the listerial invasion complex InlA/E-cadherin. *Proc Natl Acad Sci U S A* 104: 13960-13965.
58. Marino M, Braun L, Cossart P, Ghosh P (1999) Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. *Mol Cell* 4: 1063-1072.
59. Schubert WD, Göbel G, Diepholz M, Darji A, Kloer D, et al. (2001) Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain. *J Mol Biol* 312: 783-794.
60. Ferraris DM, Gherardi E, Di Y, Heinz DW, Niemann HH (2010) Ligand-mediated dimerization of the Met receptor tyrosine kinase by the bacterial invasion protein InlB. *J Mol Biol* 395: 522-532.
61. Marino M, Banerjee M, Jonquières R, Cossart P, Ghosh P (2002) GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands. *EMBO J* 21: 5623-5634.
62. Marino M, Banerjee M, Copp J, Dramsi S, Chapman T, et al. (2004) Characterization of the calcium-binding sites of *Listeria monocytogenes* InlB. *Biochem Biophys Res Commun* 316: 379-386.
63. Niemann HH, Jäger V, Butler PJ, van den Heuvel J, Schmidt S, et al. (2007) Structure of the human receptor tyrosine kinase met in complex with the *Listeria* invasion protein InlB. *Cell* 130: 235-246.
64. Niemann HH, Gherardi E, Bley Müller WM, Heinz DW (2012) Engineered variants of InlB with an additional leucine-rich repeat discriminate between physiologically relevant and packing contacts in crystal structures of the InlB:MET complex. *Protein Sci* 21: 1528-1539.
65. Breitsprecher D, Gherardi E, Bley Müller WM, Niemann HH (2014) Crystal structure of an engineered YopM-InlB hybrid protein. *BMC Struct Biol* 14: 12.
66. Ooi A, Hussain S, Seyedarabi A, Pickersgill RW (2006) Structure of internalin C from *Listeria monocytogenes*. *Acta Crystallogr D Biol Crystallogr* 62: 1287-1293.
67. Polle L, Rigano LA, Julian R, Ireton K, Schubert WD (2014) Structural details of human tuba recruitment by InlC of *Listeria monocytogenes* elucidate bacterial cell-cell spreading. *Structure* 22: 304-314.
68. Neves D, Job V, Dortet L, Cossart P, Dessen A (2013) Structure of internalin InlK from the human pathogen *Listeria monocytogenes*. *J Mol Biol* 425: 4520-4529.

69. Bublitz M, Holland C, Sabet C, Reichelt J, Cossart P, et al. (2008) Crystal structure and standardized geometric analysis of InlJ, a listerial virulence factor and leucine-rich repeat protein with a novel cysteine ladder. *J Mol Biol* 378: 87-96.
70. Sun Y, Li L, Macho AP, Han Z, Hu Z, et al. (2013) Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science* 342: 624-628.
71. Santiago J, Henzler C, Hothorn M (2013) Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. *Science* 341: 889-892.
72. Hothorn M, Belkhadir Y, Dreux M, Dabi T, Noel JP, et al. (2011) Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* 474: 467-471.
73. She J, Han Z, Kim TW, Wang J, Cheng W, et al. (2011) Structural insight into brassinosteroid perception by BRI1. *Nature* 474: 472-476.
74. She J, Han Z, Zhou B, Chai J (2013) Structural basis for differential recognition of brassinolide by its receptors. *Protein Cell* 4: 475-482.
75. Liu P, Hu Z, Zhou B, Liu S, Chai J (2013) Crystal structure of an LRR protein with two solenoids. *Cell Res* 23: 303-305.
76. Song W, Han Z, Sun Y, Chai J (2014) Crystal structure of a plant leucine rich repeat protein with two island domains. *Sci China Life Sci* 57: 137-144.
77. Di Matteo A, Federici L, Mattei B, Salvi G, Johnson KA, et al. (2003) The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc Natl Acad Sci U S A* 100: 10124-10128.
78. Kim HM, Oh SC, Lim KJ, Kasamatsu J, Heo JY, et al. (2007) Structural diversity of the hagfish variable lymphocyte receptors. *J Biol Chem* 282: 6726-6732.
79. Han BW, Herrin BR, Cooper MD, Wilson IA (2008) Antigen recognition by variable lymphocyte receptors. *Science* 321: 1834-1837.
80. Deng L, Velikovskiy CA, Xu G, Iyer LM, Tasumi S, et al. (2010) A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey. *Proc Natl Acad Sci U S A* 107: 13408-13413.
81. Luo M, Velikovskiy CA, Yang X, Siddiqui MA, Hong X, et al. (2013) Recognition of the Thomsen-Friedenreich pancarcinoma carbohydrate antigen by a lamprey variable lymphocyte receptor. *J Biol Chem* 288: 23597-23606.
82. Kanda R, Sutoh Y, Kasamatsu J, Maenaka K, Kasahara M, et al. (2014) Crystal structure of the lamprey variable lymphocyte receptor C reveals an unusual feature in its N-terminal capping module. *PLoS One* 9: e85875.
83. Velikovskiy CA, Deng L, Tasumi S, Iyer LM, Kerzic MC, et al. (2009) Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. *Nat Struct Mol Biol* 16: 725-730.
84. Kirchdoerfer RN, Herrin BR, Han BW, Turnbough CL Jr, Cooper MD, et al. (2012) Variable lymphocyte receptor recognition of the immunodominant glycoprotein of *Bacillus anthracis* spores. *Structure* 20: 479-486.
85. Lee SC, Park K, Han J, Lee JJ, Kim HJ, et al. (2012) Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc Natl Acad Sci U S A* 109: 3299-3304.
86. Seiradake E, Coles CH, Perestenko PV, Harlos K, McIlhinney RA, et al. (2011) Structural basis for cell surface patterning through NetrinG-NGL interactions. *EMBO J* 30: 4479-4488.
87. Mosyak L, Wood A, Dwyer B, Buddha M, Johnson M, et al. (2006) The structure of the Lingo-1 ectodomain, a module implicated in central nervous system repair inhibition. *J Biol Chem* 281: 36378-36390.
88. Barton WA, Liu BP, Tzvetkova D, Jeffrey PD, Fournier AE, et al. (2003) Structure and axon outgrowth inhibitor binding of the Nogo-66 receptor and related proteins. *EMBO J* 22: 3291-3302.
89. He XL, Bazan JF, McDermott G, Park JB, Wang K, et al. (2003) Structure of the Nogo receptor ectodomain: a recognition module implicated in myelin inhibition. *Neuron* 38: 177-185.
90. Weinreb PH, Wen D, Qian F, Wildes CP, Garber EA, et al. (2010) Resolution of disulfide heterogeneity in Nogo receptor I fusion proteins by molecular engineering. *Biotechnol Appl Biochem* 57: 31-45.
91. Morlot C, Thielens NM, Ravelli RB, Hemrika W, Romijn RA, et al. (2007) Structural insights into the Slit-Robo complex. *Proc Natl Acad Sci U S A* 104: 14923-14928.
92. Morlot C, Hemrika W, Romijn RA, Gros P, Cusack S, et al. (2007) Production of Slit2 LRR domains in mammalian cells for structural studies and the structure of human Slit2 domain 3. *Acta Crystallogr D Biol Crystallogr* 63: 961-968.
93. Seiradake E, von Philipsborn AC, Henry M, Fritz M, Lortat-Jacob H, et al. (2009) Structure and functional relevance of the Slit2 homodimerization domain. *EMBO Rep* 10: 736-741.
94. Howitt JA, Clout NJ, Hohenester E (2004) Binding site for Robo receptors revealed by dissection of the leucine-rich repeat region of Slit. *EMBO J* 23: 4406-4412.
95. Parthier C, Stelter M, Ursel C, Fandrich U, Lilie H, et al. (2014) Structure of the Toll-Spatzle complex, a molecular hub in *Drosophila* development and innate immunity. *Proc Natl Acad Sci U S A* 111: 6281-6286.
96. Gangloff M, Arnot CJ, Lewis M, Gay NJ (2013) Functional insights from the crystal structure of the N-terminal domain of the prototypical toll receptor. *Structure* 21: 143-153.
97. Gangloff M, Moreno A, Gay NJ (2013) Liesegang-like patterns of Toll crystals grown in gel. *J Appl Crystallogr* 46: 337-345.
98. Lewis M, Arnot CJ, Beeston H, McCoy A, Ashcroft AE, et al. (2013) Cytokine Spatzle binds to the *Drosophila* immunoreceptor Toll with a neurotrophin-like specificity and couples receptor activation. *Proc Natl Acad Sci U S A* 110: 20461-20466.
99. Jiang X, Liu H, Chen X, Chen PH, Fischer D, et al. (2012) Structure of follicle-stimulating hormone in complex with the entire ectodomain of its receptor. *Proc Natl Acad Sci U S A* 109: 12491-12496.
100. Fan QR, Hendrickson WA (2005) Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* 433: 269-277.
101. Sanders P, Young S, Sanders J, Kabelis K, Baker S, et al. (2011) Crystal structure of the TSH receptor (TSHR) bound to a blocking-type TSHR autoantibody. *J Mol Endocrinol* 46: 81-99.
102. Sanders J, Chirgadze DY, Sanders P, Baker S, Sullivan A, et al. (2007) Crystal structure of the TSH receptor in complex with a thyroid-stimulating autoantibody. *Thyroid* 17: 395-410.
103. Kajander T, Kuja-Panula J, Rauvala H, Goldman A (2011) Crystal structure and role of glycans and dimerization in folding of neuronal leucine-rich repeat protein AMIGO-1. *J Mol Biol* 413: 1001-1015.
104. Xu K, Xu Y, Rajashankar KR, Robev D, Nikolov DB (2013) Crystal structures of Lgr4 and its complex with R-spondin1. *Structure* 21: 1683-1689.
105. Wang D, Huang B, Zhang S, Yu X, Wu W, et al. (2013) Structural basis for R-spondin recognition by LGR4/5/6 receptors. *Genes Dev* 27: 1339-1344.
106. Chen PH, Chen X, Lin Z, Fang D, He X (2013) The structural basis of R-spondin recognition by LGR5 and RNF43. *Genes Dev* 27: 1345-1350.
107. Peng WC, de Lau W, Fomeris F, Granneman JC, Huch M, et al. (2013) Structure of stem cell growth factor R-spondin 1 in complex with the ectodomain of its receptor LGR5. *Cell Rep* 3: 1885-1892.
108. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, et al. (2007) Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acetylated lipopeptide. *Cell* 130: 1071-1082.
109. Kang JY, Nan X, Jin MS, Youn SJ, Ryu YH, et al. (2009) Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* 31: 873-884.
110. Choe J, Kelker MS, Wilson IA (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 309: 581-585.
111. Bell JK, Botos I, Hall PR, Askins J, Shiloach J, et al. (2005) The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proc Natl Acad Sci U S A* 102: 10976-10980.
112. Luo J, Obmolova G, Malia TJ, Wu SJ, Duffy KE, et al. (2012) Lateral clustering of TLR3:dsRNA signaling units revealed by TLR3ecd:3Fabs quaternary structure. *J Mol Biol* 421: 112-124.
113. Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, et al. (2008) Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* 320: 379-381.
114. Ohto U, Fukase K, Miyake K, Shimizu T (2012) Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc Natl Acad Sci U S A* 109: 7421-7426.

115. Park BS, Song DH, Kim HM, Choi BS, Lee H, et al. (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 458: 1191-1195.
116. Ohto U, Yamakawa N, Akashi-Takamura S, Miyake K, Shimizu T (2012) Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. *J Biol Chem* 287: 40611-40617.
117. Zhou K, Kanai R, Lee P, Wang HW, Modis Y (2012) Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. *J Struct Biol* 177: 402-409.
118. Yoon SI, Kumarov O, Natarajan V, Hong M, Gudkov AV, et al. (2012) Structural basis of TLR5-flagellin recognition and signaling. *Science* 335: 859-864.
119. Tanji H, Ohto U, Shibata T, Miyake K, Shimizu T (2013) Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science* 339: 1426-1429.
120. Kokatla HP, Sil D, Tanji H, Ohto U, Malladi SS, et al. (2014) Structure-based design of novel human Toll-like receptor 8 agonists. *ChemMedChem* 9: 719-723.
121. Yoon SI, Hong M, Wilson IA (2011) An unusual dimeric structure and assembly for TLR4 regulator RP105-MD-1. *Nat Struct Mol Biol* 18: 1028-1035.
122. Ohto U, Miyake K, Shimizu T (2011) Crystal structures of mouse and human RP105/MD-1 complexes reveal unique dimer organization of the toll-like receptor family. *J Mol Biol* 413: 815-825.
123. Zhao Y, Malinauskas T, Harlos K, Jones EY2 (2014) Structural insights into the inhibition of Wnt signaling by cancer antigen 5T4/Wnt-activated inhibitory factor 1. *Structure* 22: 612-620.
124. Celikel R, McClintock RA, Roberts JR, Mendolicchio GL, Ware J, et al. (2003) Modulation of alpha-thrombin function by distinct interactions with platelet glycoprotein Iba1. *Science* 301: 218-221.
125. Uff S, Clemetson JM, Harrison T, Clemetson KJ, Emsley J (2002) Crystal structure of the platelet glycoprotein Ib(alpha) N-terminal domain reveals an unmasking mechanism for receptor activation. *J Biol Chem* 277: 35657-35663.
126. Varughese KI, Ruggeri ZM, Celikel R (2004) Platinum-induced space-group transformation in crystals of the platelet glycoprotein Ib alpha N-terminal domain. *Acta Crystallogr D Biol Crystallogr* 60: 405-411.
127. Huizinga EG, Tsuji S, Romijn RA, Schiphorst ME, de Groot PG, et al. (2002) Structures of glycoprotein Iba1 and its complex with von Willebrand factor A1 domain. *Science* 297: 1176-1179.
128. Zarpellon A, Celikel R, Roberts JR, McClintock RA, Mendolicchio GL, et al. (2011) Binding of alpha-thrombin to surface-anchored platelet glycoprotein Ib alpha sulfotyrosines through a two-site mechanism involving exosite I. *Proc Natl Acad Sci U S A* 108: 8628-8633.
129. Dumas JJ, Kumar R, Seehra J, Somers WS, Mosyak L (2003) Crystal structure of the GpIba1-thrombin complex essential for platelet aggregation. *Science* 301: 222-226.
130. Dumas JJ, Kumar R, McDonagh T, Sullivan F, Stahl ML, et al. (2004) Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Iba1 complex reveals conformation differences with a complex bearing von Willebrand disease mutations. *J Biol Chem* 279: 23327-23334.
131. Fukuda K, Doggett T, Laurenzi JJ, Liddington RC, Diacovo TG (2005) The snake venom protein botrocetin acts as a biological brace to promote dysfunctional platelet aggregation. *Nat Struct Mol Biol* 12: 152-159.
132. McEwan PA, Andrews RK, Emsley J (2009) Glycoprotein Iba1 inhibitor complex structure reveals a combined steric and allosteric mechanism of von Willebrand factor antagonism. *Blood* 114: 4883-4885.
133. Blenner MA, Dong X, Springer TA (2014) Structural basis of regulation of von Willebrand factor binding to glycoprotein Ib. *J Biol Chem* 289: 5565-5579.
134. Baxter RH, Steinert S, Chelliah Y, Volohonsky G, Levashina EA, et al. (2010) A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 107: 16817-16822.
135. Basquin J, Roudko VV, Rode M, Basquin C, Séraphin B, et al. (2012) Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. *Mol Cell* 48: 207-218.
136. Scott PG, McEwan PA, Dodd CM, Bergmann EM, Bishop PN, et al. (2004) Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. *Proc Natl Acad Sci U S A* 101: 15633-15638.
137. Scott PG, Dodd CM, Bergmann EM, Sheehan JK, Bishop PN (2006) Crystal structure of the biglycan dimer and evidence that dimerization is essential for folding and stability of class I small leucine-rich repeat proteoglycans. *J Biol Chem* 281: 13324-13332.
138. Kelley SL, Lukk T, Nair SK, Tapping RI (2013) The crystal structure of human soluble CD14 reveals a bent solenoid with a hydrophobic amino-terminal pocket. *J Immunol* 190: 1304-1311.
139. Kim JI, Lee CJ, Jin MS, Lee CH, Paik SG, et al. (2005) Crystal structure of CD14 and its implications for lipopolysaccharide signaling. *J Biol Chem* 280: 11347-11351.
140. de Chiara C, Menon RP, Pastore A (2008) Structural bases for recognition of Anp32/LANP proteins. *FEBS J* 275: 2548-2560.
141. Huyton T, Wolberger C (2007) The crystal structure of the tumor suppressor protein pp32 (Anp32a): structural insights into Anp32 family of proteins. *Protein Sci* 16: 1308-1315.
142. Tochio N, Umehara T, Munemasa Y, Suzuki T, Sato S, et al. (2010) Solution structure of histone chaperone ANP32B: interaction with core histones H3-H4 through its acidic concave domain. *J Mol Biol* 401: 97-114.
143. Wehrman T, He XL, Raab B, Dukipatti A, Blau H, et al. (2007) Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. *Neuron* 53: 25-38.
144. McEwan PA, Yang W, Carr KH, Mo X, Zheng X, et al. (2011) Quaternary organization of GPIb-IX complex and insights into Bernard-Soulier syndrome revealed by the structures of GPIbβ and a GPIbβ/GPIX chimera. *Blood* 118: 5292-5301.
145. Trastoy B, Lomino JV, Pierce BG, Carter LG, Günther S, et al. (2014) Crystal structure of *Streptococcus pyogenes* EndoS, an immunomodulatory endoglycosidase specific for human IgG antibodies. *Proc Natl Acad Sci U S A* 111: 6714-6719.
146. Nagarajaram HA, Reddy BV, Blundell TL (1999) Analysis and prediction of inter-strand packing distances between beta-sheets of globular proteins. *Protein Eng* 12: 1055-1062.
147. Evdokimov AG, Anderson DE, Rutzahn KM, Waugh DS (2001) Unusual molecular architecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. *J Mol Biol* 312: 807-821.
148. Quezada CM, Hicks SW, Galán JE, Stebbins CE (2009) A family of *Salmonella* virulence factors functions as a distinct class of autoregulated E3 ubiquitin ligases. *Proc Natl Acad Sci U S A* 106: 4864-4869.
149. Keszei AF, Tang X, McCormick C, Zeqiraj E, Rohde JR, et al. (2014) Structure of an SspH1-PKN1 complex reveals the basis for host substrate recognition and mechanism of activation for a bacterial E3 ubiquitin ligase. *Mol Cell Biol* 34: 362-373.
150. Zhu Y, Li H, Hu L, Wang J, Zhou Y, et al. (2008) Structure of a *Shigella* effector reveals a new class of ubiquitin ligases. *Nat Struct Mol Biol* 15: 1302-1308.
151. Miyashita H, Kretsinger R, Kuroki Y, Matsushima N (2013) Horizontal gene transfer of plant-specific leucine-rich repeats between plants and bacteria. *Nat Sci* 5: 580-598.
152. Lee JK, Belogrudov GI, Stroud RM (2008) Crystal structure of bovine mitochondrial factor B at 0.96-Å resolution. *Proc Natl Acad Sci U S A* 105: 13379-13384.