Cell. Mol. Life Sci. 65 (2008) 2307 – 2333
1420-682X/08/152307-27
DOI 10.1007/s00018-008-8019-0

Cellular and Molecular Life Sciences

Review

The leucine-rich repeat structure

J. Bella[a,*], K. L. Hindle[a], P. A. McEwan[b] and S. C. Lovell[a]

[a] Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT (UK), Fax: +44-161-275-1505, e-mail: jordi.bella@manchester.ac.uk
[b] School of Pharmacy, Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham NG7 2RD (UK)

Received 11 January 2008; received after revision 7 March 2008; accepted 10 March 2008
Online First 14 April 2008

Abstract. The leucine-rich repeat is a widespread structural motif of 20–30 amino acids with a characteristic repetitive sequence pattern rich in leucines. Leucine-rich repeat domains are built from tandems of two or more repeats and form curved solenoid structures that are particularly suitable for protein-protein interactions. Thousands of protein sequences containing leucine-rich repeats have been identified by automatic annotation methods. Three-dimensional structures of leucine-rich repeat domains determined to date reveal a degree of structural variability that translates into the considerable functional versatility of this protein superfamily. As the essential structural principles become well established, the leucine-rich repeat architecture is emerging as an attractive framework for structural prediction and protein engineering. This review presents an update of the current understanding of leucine-rich repeat structure at the primary, secondary, tertiary and quaternary levels and discusses specific examples from recently determined three-dimensional structures.

Keywords. Leucine-rich repeat, protein structure, protein conformation, protein-protein interactions, molecular sequence data, molecular models, protein engineering, repetitive sequences.

Introduction

Several protein architectures are built from tandems of internal repetitive motifs [1, 2]. Perhaps the best-known example is the leucine-rich repeat (LRR), a widespread structural motif that has been identified in thousands of protein sequences in all life forms, from viruses to eukaryotes. Each LRR is typically 20–30 amino acids long and unusually rich in the hydrophobic amino acid leucine. The defining feature of the LRR motif is an 11-residue hallmark sequence L*xx*L*x*L*xx*N*x*L (*x* being any amino acid) [3], although other hydrophobic residues can substitute the leucine and asparagine residues at the consensus positions. Proteins and domains containing tandems of two or more LRRs form the continuously expanding LRR superfamily [4]. Members include intracellular, extracellular and membrane-attached proteins with such varied functions as cell adhesion and signalling [5–7], extracellular matrix assembly [8], platelet aggregation [9], neuronal development [10, 11], RNA processing [12, 13], adhesion and invasion of pathogenic bacteria to host cells [14, 15], disease resistance and pathogen recognition in plants [16–18] and immune response [19–23]. Despite their apparently unrelated functions, LRR proteins and domains share a common

* Corresponding author.

structural framework that makes them suitable for protein-protein interactions.

Here we review recent findings on the structure and function of LRR proteins and summarise the current understanding of LRR architecture at the different structural levels, from amino acid sequence to quaternary arrangements. Several excellent analyses and reviews of LRR structure have been published through the years [3, 24–29]. We refer the reader to those when appropriate to avoid unnecessary duplication.

Structural characteristics of LRR proteins

The first crystal structure of an LRR protein, ribonuclease inhibitor (RNI) established the structural nature of the repeats [30]. Since then nearly 90 structures of LRR proteins or domains have been determined (Table 1) and several others await release from the Protein Data Bank [31]. Protein domains with LRR architecture form curved solenoid structures where each repeat is a turn of the solenoid (Fig. 1). The LRR superfamily thus belongs to a larger structural clan that includes many different types of solenoid folds [32]. The concave side of the LRR domains is defined by a parallel β-sheet to which each LRR contributes one strand. The strands are interwoven with a variety of structural elements in the convex side: α-helices, 3_{10} helices, polyproline II helices, β-turns and even short β-strands. The curved LRR architecture appears to be well suited for building protein-protein interaction modules and it is generally thought that the concave surfaces of the LRR structures contain their ligand-binding sites. Such mode of binding has been demonstrated in several crystal structures of LRR protein domains in complex with their ligands (Fig. 1C, D) [12, 33–45], but some LRR proteins use alternative surfaces for ligand binding [46–50]. The recent discovery that some LRR proteins use their concave side to form very stable dimers (Fig. 1B) [51, 52] suggests that they may also use alternative modes of binding [53].

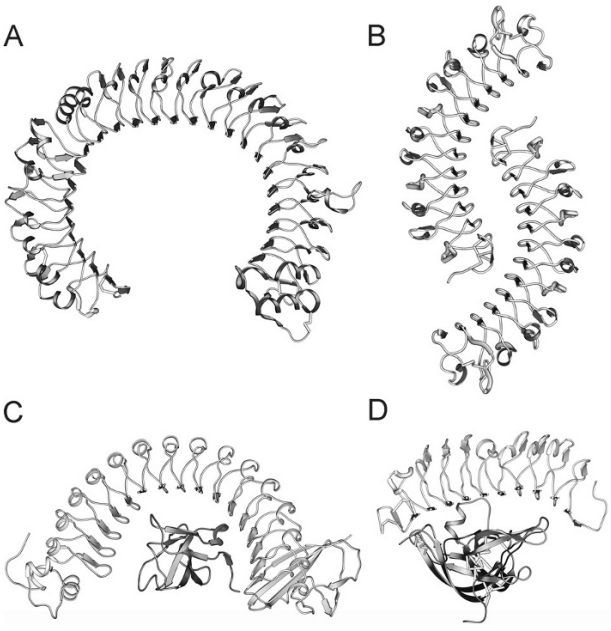

Figure 1. Tertiary and quaternary structures of leucine-rich repeat (LRR) domains viewed from the sides of the respective solenoids. All drawings are shown at scale. (*A*) The TLR3 structure (PDB i.d. 2A0Z) illustrates the characteristic solenoid fold where each LRR corresponds to one turn of the solenoid. Its overall shape is three-quarters of a torus and is made of 25 LRRs of different lengths (see Table 1). The concave side is made of a continuous β-sheet, whereas the convex side shows a variety of secondary helical structures (α-helix, 3_{10}, polyproline II) as well as stretches of consecutive β-turns and short β-strands. The inner space appears well suited for protein-protein interactions. (*B*) The DCN structure (PDB i.d. 1XKU) is an example of homodimeric arrangement of LRR domains, where the surface on the concave side is used as dimerization interface. (*C*) An example of an LRR protein bound to its ligand: INLA in complex with the N-terminal domain of human E-cadherin (PDB i.d. 1O6S). The cadherin domain is largely surrounded by the half-torus LRR structure. A special capping domain (LRRCI, described in the text) closes the C-terminal end of the INLA LRR domain. (*D*) Even short LRR domains can engage their ligands through their concave surface, as shown by the ternary complex between the extracellular LRR domain of FSHR and its ligand follicle-stimulating hormone (FSH), which is made of two chains α and β (PDB i.d. 1XWD). The LRR structure is much more open and only makes contact with one side of the ligand, but still uses its concave surface to that purpose. This figure was prepared with SETOR [160].

Secondary structure and repeat length

The chain conformation in the convex side of an LRR domain is partially related to the length of the individual component repeats. For the purposes of structural discussion, we have adopted the standard orientation shown in Figure 2. The RNI structure contains 16 repeats: 14 alternating in lengths between 28 and 29 amino acids and two shorter ones flanking them (Table 1). Each LRR in the RNI structure has a continuous α-helix in its convex side that is connected to the strands forming the β-sheet in its concave side by two loops (Fig. 2A). The "ascending" loops link the C-terminal ends of the β-strands to the N termini of the α-helices; the "descending" loops link the C-terminal ends of the helices to the N termini of the following β-strands. Since RNI was the first LRR structure to be determined, all LRR proteins and domains are classified as α/β folds (α-helices alternating with β-strands) or β-α superhelices in structural databases such as SCOP [54] or CATH [55]. Paradoxically, many of the LRR structures determined to date have very little or no α-helical conformation. The

Table 1. Structures of leucine-rich repeat (LRR) proteins or domains determined to date (standalone or in complex with ligands), and abbreviations used in the text. The numbers and lengths of LRRs correspond to those observed in the actual structures, which may differ from the predictions based on sequence only or from full sequences when only a fragment has been crystallized. Repeats involved in capping motifs LRRNT, LRRCT, LRRCE and LRRCap (see text) are underlined in the list of LRRs. For the sake of completeness some very long terminal repeats with irregular structure are included as well in the LRR sequences.

LRR protein	PDB ids	LRRs	LRR sequence	References
BGN: Biglycan core protein	2FT3	12	21 $(24)_2$ 21 24 25 21 $(24)_2$ 23 30 27	[52]
CD14: Monocyte differentiation antigen CD14	1WWL	10	37 27 26 28 24 28 27 25 21 22	[49]
DCN: Decorin core protein	1XCD 1XEC 1XKU	12	21 $(24)_2$ 21 24 26 21 $(24)_2$ 23 30 27	[51]
DLC1: Dynein light chain 1	1DS9 1M9L	6	25 22 23 22 25 34	[72, 73]
FSHR: Follicle-stimulating hormone receptor	1XWD	9	21 24 $(25)_2$ 24 26 23 25 21	[38]
GGTA: Rab geranylgeranyl transferase α	1DCE 1LTX	5	22 23 22 25 27	[142, 143]
GPIBA: Glycoprotein Ibα	1GWB 1M0Z 1M10 1OOK 1P8V 1P9A 1QYY 1SQ0 1U0N	9	21 24 22 23 $(24)_3$ 23 28	[36, 37, 39, 47, 48, 56, 144]
INLA: Internalin A	1O6S 1O6T 1O6V 2OMV 2OMY 2OMW	15	$(22)_5$ 21 $(22)_9$	[35, 45]
INLB: Internalin B	1D0B 1H6T 1M9S 1OTM 1OTN 1OTO 2UZX, 2UZY	7	$(22)_7$	[66, 98, 100, 145, 146]
INLC: Internalin C	1XEU	6	$(22)_2$ 21 $(22)_3$	[147]
INLH: Internalin H	1H6U	8	$(22)_8$	[98]
LINGO1: Lingo-1 receptor ectodomain	2ID5	14	21 $(24)_{12}$ 26	[88]
NOGO: Nogo receptor ectodomain	1OZN 1P8T	10	21 $(24)_2$ 25 $(24)_5$ 28	[69, 70]
NTRK1: Neurotrophic tyrosine kinase receptor type 1 ectodomain	2IFG	5	23 25 24 23 26	[148]
PGIP2: Polygalacturonase inhibitor 2	1OGQ	10	26 25 $(24)_2$ 25 23 24 23 24 26	[65]
PP32: Acidic nuclear phosphoprotein PP32	2JE0 2JE1	5	25 22 24 25 27	[149]
RNA1P: RAN GTPase-activating protein 1	1K5D 1K5G 1YRG 2CA6	11	33 28 34 28 37 28 29 28 30 29 30	[46, 150]
RNI: Ribonuclease inhibitor	1A4Y 1DFJ 1Z7X 2BEX 2BNH 2Q4G	16	25 (28 29)$_7$ 28/27	[30, 33, 34, 41, 42, 151, 152]
SKP2: S-phase kinase- associated protein 2	1FQV 1FS2 2ASS 2AST	10	$(23)_2$ 25 24 26 $(27)_2$ $(25)_2$ 24	[40, 153]
SLITD3: Slit protein 3rd LRR domain	1W8A	6	21 25 $(24)_3$ 29	[83]
SLIT2D2: Slit2 protein 2nd LRR domain	2V9S 2V9T	7	21 $(24)_5$ 26	[44]
SLIT2D3: Slit2 protein 3rd LRR domain	2V70	7	21 25 $(24)_4$ 27	[154]
TAP: Tip-associating protein (nuclear RNA export factor 1)	1FO1 1FT8 1KOH 1KOO	4	43 26 24 31	[13, 155]
TIR1: Transport inhibitor response 1 protein	2P1M 2P1N 2P1O 2P1P 2P1Q	18	23 39 25 26 28 27 24 30 24 25 24 34 25 35 24 $(25)_2$ 25	[74]
TLR1H: Toll-like receptor 1 ectodomain (human, hybrid with VLR hagfish protein)	2Z7X	20	21 $(24)_2$ 21 25 23 25 24 33 27 26 29 22 24 26 25 22 23 24 28	[50]
TLR2H: Toll-like receptor 2 ectodomain (human and mouse, hybrids with VLR hagfish protein)	2Z81 2Z82 2Z7X	21	21 $(24)_3$ $(25)_2$ $(24)_2$ 27 28 30 29 24 27 26 23 21 20 22 24 28	[50]
TLR2S: Toll-like receptor 2 ectodomain (human, shorter hybrid with VLR hagfish protein)	2Z80	12	21 $(24)_3$ $(25)_2$ $(24)_2$ 27 28 24 28	[50]
TLR3: Toll-like receptor 3 ectodomain (human)	1ZIW 2A0Z	25	21 $(24)_5$ 26 24 27 26 $(24)_2$ 33 24 28 24 25 24 26 24 32 $(24)_2$ 25 28	[57, 84]
TLR4: Toll-like receptor 4 ectodomain (mouse)	2Z64	23	21 $(24)_4$ 25 28 23 27 30 24 22 21 22 26 23 25 24 25 $(24)_3$ 27	[43]
TLR4H: Toll-like receptor 4 ectodomain (human, hybrid with VLR hagfish protein)	2Z63	22	21 $(24)_4$ 25 28 23 27 30 25 22 21 22 26 23 (25 24)$_2$ 24 28	[43]
TLR4S: Toll-like receptor 4 ectodomain (human, shorter hybrid with VLR hagfish protein)	2Z62 2Z65	10	21 $(24)_4$ 25 28 23 24 28	[43]
TMOC: Tropomodulin C-terminal domain	1IO0 1PGV	4	29 28 28 30/31	[156, 157]
U2A: Small nuclear ribonucleoprotein A'	1A9N	5	23 22 24 25 27	[12]

Table 1 *(Continued)*

LRR protein	PDB ids	LRRs	LRR sequence	References
VLRA29: Hagfish variable lymphocyte receptor A29	2O6Q	9	21 $(24)_7$ 32	[71]
VLRB59: Hagfish variable lymphocyte receptor B59	2O6S	7	21 $(24)_5$ 30	[71]
VLRB61: Hagfish variable lymphocyte receptor B61	2O6R	6	21 $(24)_4$ 30	[71]
VLRBT4: Hybrid VLRB61 (hagfish)-Toll-like receptor 4 ectodomain (human)	2Z66	11	21 24 26 23 25 24 25 $(24)_2$ 25 27	[43]
YOPM: *Yersinia* outer membrane protein M	1G9U 1JL5	15	$(20)_3$ (22 20)$_3$ $(20)_6$	[67]

reason is that shorter LRRs are more common than those in the RNI structure, and repeats with 20–24 amino acids show other secondary structures in their convex sides. This variability is illustrated in Figure 2. Repeats with 27–29 residues, like those in the RNI structure, have an α-helix running through the convex side between two consecutive β-strands (Fig. 2A). Very short repeats (20- or 21-residues long) use a more extended main chain conformation with segments of polyproline II helix instead (Fig. 2B, C). The high conservation of one or two proline residues in the convex sides of these short LRRs is a consequence of the polyproline II conformation. Repeats with intermediate length, 22–26, can adopt a variety of secondary structure combinations to build their convex sides, with a prevalence of the C=O(i)···H–N(i+3) hydrogen bonding topology (3_{10} helices and β-turns). Thus, the predominant secondary structures in the convex sides of the 22-residue LRRs from the internalins are short 3_{10} helices and β-turns (Fig. 2D), sometimes combined with individual γ-turns. However, repeats with 22–23 amino acids from other proteins may use a combination of polyproline II and β-turns or 3_{10} helical segments (Fig. 2E). The typical length of an LRR is 24 amino acids and several of the structures shown in Table 1 contain one or more tandems of 24-residue LRRs. These repeats often have short β-strands at the beginning of their convex sides followed by stretches with β-turns, short segments of 3_{10} helices or a combination of both (Fig. 2F). Of particular interest is the tandem β-turn motif where two or three consecutive β-turns form a continuous hydrogen-bonded structure (Fig. 3; see also [56] and Fig. 2 therein). These tandems of β-turns, ubiquitous in 24-residue LRRs, result in an amphipatic, flattened structure where hydrophobic residues are inserted into the LRR hydrophobic core and hydrophilic residues are pointing outside (Fig. 3) [56]. LRRs with 25 or 26 amino acids are intermediate in length and their convex sides can show a variety of secondary structures: α-helices like those in the RNI inhibitors (but with less turns), 3_{10} helices, or strand/tandem β-turn combinations similar to those of 24-residue LRRs.

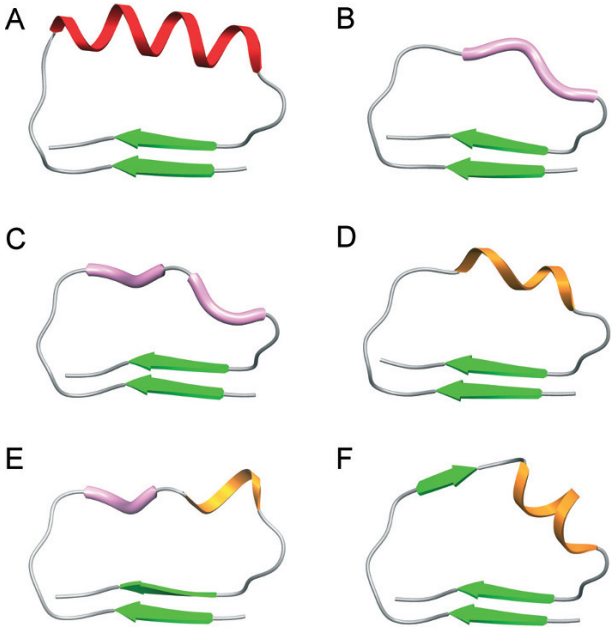

Figure 2. Diversity of secondary structures in the convex side of LRRs. In each panel, the β-strands from two consecutive LRRs are shown. (*A*) A 28-residue repeat from the RNI structure, showing a continuous α-helix in the convex side. (*B*) A 20-residue repeat from the YOPM structure, showing one segment of polyproline II conformation. (*C*) A 21-residue repeat from the DCN structure, with two polyproline II segments. (*D*) A 22-residue repeat from the INLH structure, showing a β-turn in tandem with a 3_{10} helical turn. (*E*) A 23-residue repeat from the DCN structure, where a polyproline II segment is followed by a single turn of 3_{10} helix. (*F*) A 24-residue repeat from the TLR3 structure, showing a short β-strand in the convex side followed by two turns of 3_{10} helix. In all figures green arrows represent β-strands, red ribbons α-helices, orange ribbons 3_{10}-helix/β-turn tandem structures, and pink tubes segments of polyproline II helix. 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". Figure prepared with SETOR [160].

The structural preferences observed indicate a gradual transition in the convex sides of LRRs as their length increases, from polyproline II conformation to

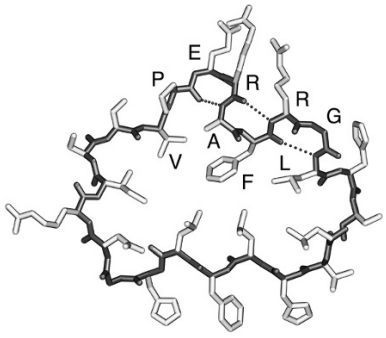

Figure 3. Three consecutive β-turns in the convex side of LRR-VII from the NOGO structure (PDB i.d. 1OZN, residues ^{180}LTHLFLHGNRISSVPERAFRGLHS^{203}). The sequence VPERAFRGL is indicated, where underlined residues have their side chains in the hydrophobic core of the repeat. Figure prepared with SETOR [160].

3_{10} helix and β-turns, and then to α-helices. Such a trend only applies to regular LRRs, i.e. those that do not have any extension on their convex, ascending or descending, sides. The structures in Table 1 show several examples of extended LRRs (Fig. 4; see also Fig. 5 from [57]), where additional residues loop out from the expected path of a regular LRR before rejoining it some residues later. These extensions adopt a variety of secondary structures. Although to date there is no known example of an LRR extension forming a separate folded domain, such a situation has been observed for a related fold, the β-helix solenoid, in the structure of a haemoglobin protease from pathogenic *Escherichia coli* [58, 59].

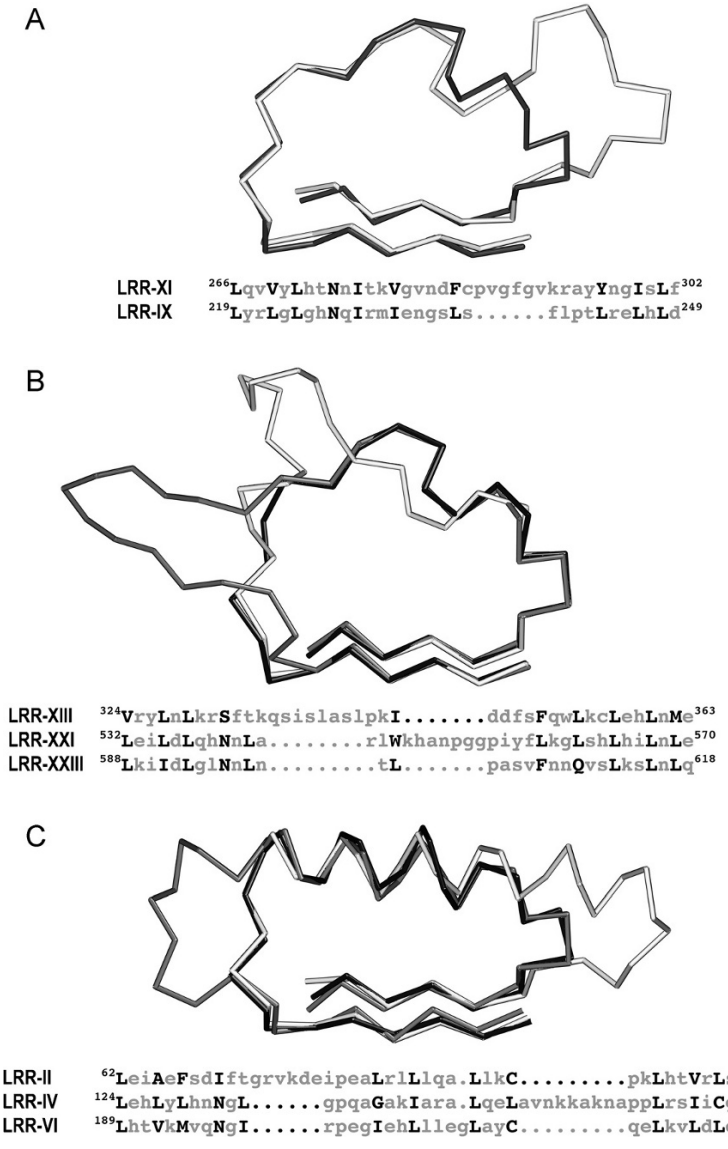

Figure 4. Examples of extended LRRs, shown as Cα traces, and their sequence alignments with regular LRRs. Residues in the consensus hydrophobic positions are showed in upper case. (*A*) Repeat XI in the BGN structure (PDB i.d. 2FT3) is significantly longer than the rest (30 amino acids) and shows a lateral extension (the "ear") that is characteristic of the SLRP family. The regular LRR-IX (24 amino acids) is superimposed for comparison. The ear extension occurs in the descending loop and the polypeptide chain regains its normal path for the following repeat. (*B*) Extended repeats XIII (33 amino acids) and XXI (32 amino acids) from the TLR3 structure, (PDB i.d. 2A0Z) together with regular LRR-XXIII (24 amino acids). The extension to LRR-XIII comes out from the ascending loop whereas the one to LRR-XXI occurs in the convex region. (*C*) The extended repeats LRR-II (34 amino acids) and LRR-IV (37 amino acids) from RNA1P (PDB i.d. 2CA6) extend the convex side α-helix towards the ascending and descending sides respectively. Figure prepared with SETOR [160].

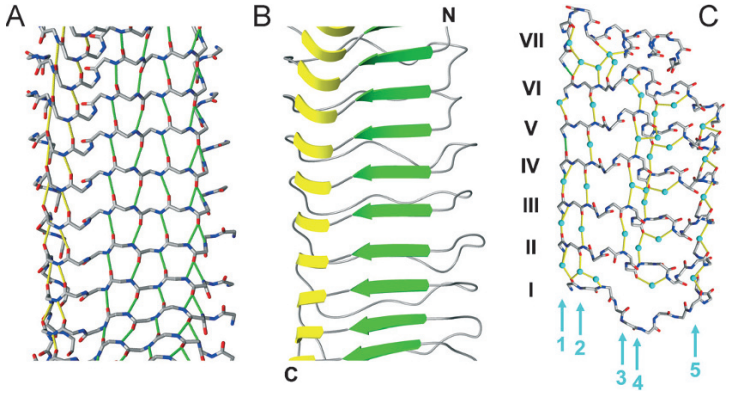

Figure 5. Hydrogen bonding and secondary structure motifs between LRRs (*A*) Main chain hydrogen bonding in the concave side of the YOPM structure (PDB i.d. 1JL5). Hydrogen bonds from the parallel β-sheet are shown as green dotted lines. Inter-repeat, main chain hydrogen bonds at either side of the α_L residue from each LRR are shown as yellow dotted lines. These hydrogen bonds form a continuous structure running the entire length of the LRR domain, shown schematically in (*B*) by the yellow ribbons C-terminal to the parallel β-strands. The combination of repetitive α_L turns and parallel β-strands can be described as an α_L bounded β-sheet (see text), and is observed in all LRR structures to date. The N and C labels in (B) indicate schematically the direction of the polypeptide chain. (*C*) Hydration spines in the VLRA29 structure (PDB i.d. 2O6Q). Water molecules (shown as isolated spheres) connect the convex sides of consecutive LRRs by hydrogen bonding the main chain peptide groups. Hydrogen bonds are shown as yellow dotted lines. Roman numerals I to VII indicate the seven LRRs of VLRA29, of which only the convex side is shown. Up to five continuous spines can be identified running perpendicularly to the polypeptide main chain (numbered 1–5). Each LRR in the VLRA29 structure has an isolated β-strand at the beginning of its convex side. These β-strands are too distant from their neighbouring strands for direct hydrogen bonding and therefore do not form a second β-sheet on the convex side; they are connected by the water spines 1 and 2 instead. Figure prepared with SETOR [160].

Hydrogen bonding connectivity across repeats

Typical ribbon-diagram representations of LRR domains can be deceptive about the amount of secondary structure and hydrogen bonding present in the main chain conformation. The common structural element to all LRR domains is the parallel β-sheet lining the concave side of the solenoid. Each β-strand is usually connected by five hydrogen bonds to the strand from the following repeat (the first C=O accepts hydrogen bonds from two N–H groups) (Fig. 5A). Four residues per strand participate in characteristic backbone-backbone hydrogen bonds. Additional hydrogen bonding connects contiguous repeats to greater or lesser extent in all the LRR structures. For instance, the first turn of the ascending loop (immediately following the parallel β-strand in each LRR), is connected by one or two hydrogen bonds to the corresponding turn in the next repeat. Taking the consensus sequence motif L*xx*L*x*L*xx*N*x*L as reference, the amino acid in position 8 (N-terminal to the consensus asparagine) is always in a left-handed α-helical (α_L) conformation. The successive alignment of residues in that position creates, due to the repetitive solenoid structure, a continuous left-handed corner that runs for the length of the entire LRR domain (Fig. 5B). Such a corner could be described as one quarter of a left-handed α-helix (or π-helix in the case of LRRs made of shorter repeats), with typical conformational angles (φ,ψ) in the α_L helical region (around +60° and +50°, respectively). This structure is usually supported by conserved side-chain interactions, such as the "asparagine ladder" already described in the first structural determination of an LRR molecule [25]. Not surprisingly similar structural motifs are observed in other solenoid structures such as β-helices [25, 60]. Pickersgill and colleagues [60–62] have noticed the unusual nature of these α_L-based turns on parallel β-sheets and identified them as a new protein structural motif, coining the term "α_L bounded β-strand" to describe it. More recently, these turns have been classified in the broader category of β-arcs, and the term "β-arcades" used to describe the stacking in register and through hydrogen bonding of β-arcs in β-solenoid proteins [63, 64].

Repetitive patterns of main chain hydrogen bonding can also occur between the ascending loops of consecutive repeats. Di Matteo and colleagues [65] have proposed that these hydrogen bonds give rise to a second, distorted parallel β-sheet in the PGIP2 structure. Although such description is debatable in terms of conformational angles, the presence of additional inter-repeat hydrogen bonding on the flanks of the PGIP2 structure (and to lesser extent in other LRR structures), emphasises the similarities between the LRR fold and the β-helical architecture seen in pectate lyases and other solenoid structures. Notwithstanding its structural classification, the ascending side of LRR domains forms another contiguous surface that can potentially provide additional or alternative recognition sites for protein-protein interactions [65].

Repeats with 24–26 amino acids typically show a short β-strand at the beginning of their convex side (Fig. 2F). Consecutive LRRs containing these short strands could in principle form continuous β-sheets on the convex side of the LRR domain. Pairs of such hydrogen-bonded β-strands occur in the two known structures of members of the small LRR protein and proteoglycan (SLRP) family, DCN and BGN, where the LRR lengths follow an imperfect 21–24–24 pattern (Table 1), and each pair of consecutive longer LRRs forms small, isolated β-sheets in their convex side (see Fig. 5 from [53]). However, it is also observed in other LRR structures that the hydrogen bonding connectivity between similar β-strands in the convex side extends only to two, exceptionally three, adjacent LRRs. In fact, the majority of β-strands from the convex side of the LRR structures are isolated and are only connected to β-strands in adjacent repeats through water bridges (see below). For example, in the Toll-like receptor 3 (TLR3) structure only a few pairs of consecutive LRRs form small two-stranded β-sheets in the convex side (LRRs VII with VIII, X with XI, XII with XIII, XV with XVI) and there is one instance of a short three-stranded β-sheet (LRRs XVII, XVIII and XIX). These groups are not connected to each other by direct hydrogen bonding. In all other LRRs with β-strands in the convex side, these are isolated and do not form extended secondary structures. The recently determined LRR structures of variable lymphocyte receptors from hagfish (Table 1) have tandems of several consecutive LRRs with 24 amino acids each, but no continuous β-sheet is formed in the convex side, as all the β-strands are isolated from each other.

Nevertheless, ordered water molecules are observed in LRR crystal structures (when the resolution of the data allows), connecting these isolated β-strands as well as other secondary structure elements on the convex side of the LRR domains (Fig. 5C). Repetitive patterns of inter-repeat water bridges were reported in the early high-resolution structural determinations of INLB and YOPM proteins [66, 67]. The water molecules in these bridges are tightly bound, as indicated by their low temperature factors, and therefore have a clear structural role. They form networks that compensate for the absence of direct hydrogen bonding interactions between repeats in the convex side of the LRR domains, possibly reaching a compromise between flexibility and structural stability [67]. These water networks are quite regular (Fig. 5C; see also Fig. 4 from [66] and Fig. 8 from [67]). Several structures determined at 1.6 Å resolution or better (PDB entries 1O6V, 1H6T, 1OZN, 2O6S) have confirmed the previous observations that water molecules are organised in columnar hydrogen-bonded structures, running the entire length of the convex side of the LRR domains and connecting adjacent LRRs by hydrogen bonding. The term "water spines" has been used to describe such structures [66]. Up to five spines can be seen in the convex side of the VLRA29 structure, connecting consecutive LRRs repeats perpendicularly to the path of the polypeptide chain (Fig. 5C). The same patterns of hydration are observed in other LRR structures with similar secondary structure elements in their convex sides, although due to the lower resolution of these structures not all the water molecules can be located confidently.

The hydration patterns in LRR structures are prime examples of the often overlooked role of water molecules as building elements of protein secondary structures. These patterns become more obvious in repetitive structures such as the LRR proteins or the triple helix of collagen [68], and reinforce the notion that ordered water molecules are integral components of protein structures and important contributors to the conformations of these proteins.

Hydrophobic core and LRR stability

It is normally assumed that the main contribution to the stability of LRR domains and proteins comes from the hydrophobic inner core of the solenoid, where the conserved leucines and other aliphatic residues are packed. The side chains of these residues are isolated from the solvent and are arranged with the characteristic close-packed structure seen in the hydrophobic core of globular proteins. Such packed structure optimises van der Waals interactions between side chains and confers stability to the entire LRR domain. Given the repeating structure of LRR domains, individual repeats largely follow the same principles and their side chains often completely fill the inner space (Fig. 6A). Phenylalanine residues are especially suitable for that task due to the large hydrophobic surface of their side chains; they are highly conserved in the medium-length LRRs (24–26 residues) and can form continuous phenylalanine spines when tandems of such repeats are present [69–71]. Nevertheless, some LRRs can show internal "holes" when analysed in isolation (Fig. 6B). The side chains of the neighbouring LRRs fill these holes in an interlocking manner (Fig. 6C), thus ensuring that the inner hydrophobic core of the solenoid remains compact and free of destabilising cavities. Repeats with polyproline II conformation in their convex side are appropriate for such purpose as their residue side chains point away from the helix axis in three directions roughly 120° from each other; the direction of one side chain is approximately perpendicular to the main plane of the

LRR and fills the hole in the following repeat. The structural locking of neighbouring LRRs prevents the LRR structure from opening and exposing its hydrophobic core to the solvent. It also has implications for the curvature path of the LRR domains, which may not be constant between different repeats because of the existence of these interlocking repeats.

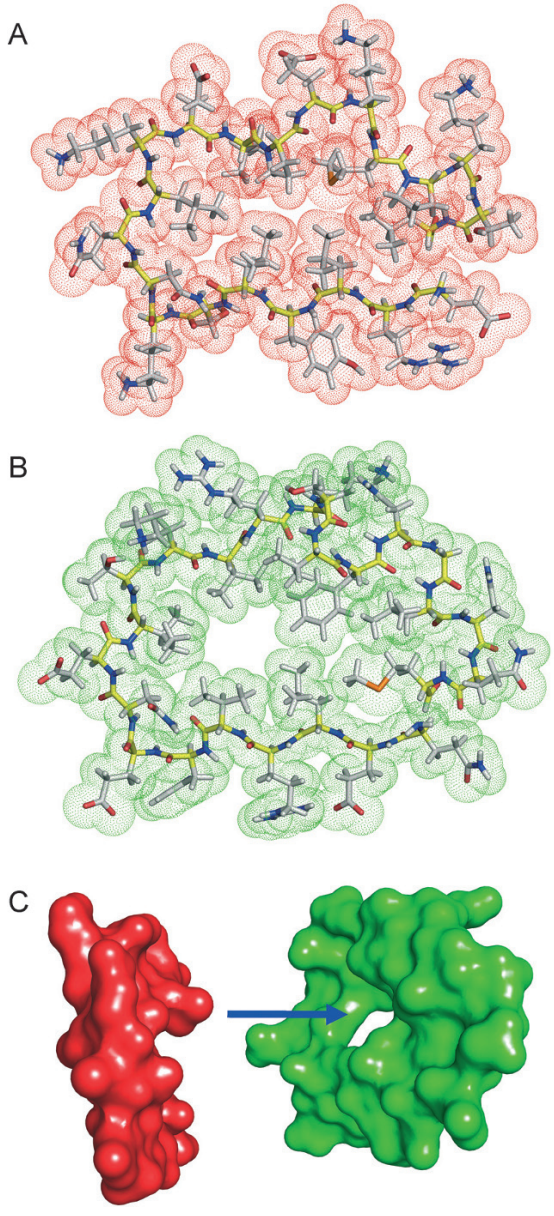

Figure 6. Close packing of side chains in the hydrophobic core of LRRs. (*A*) Dot surface representation of LRR-IV (21 residues) from the DCN structure (PDB i.d. 1XKU). Hydrogen atoms have been built in ideal positions. Main chain carbon atoms are shown in yellow; side chain carbon atoms are in white. (*B*) Dot surface representation of the 24-residue LRR-V from the same DCN structure. (*C*) Accessible surface representation of the two repeats LRR-IV (red) and LRR-V (green); one side chain from LRR-IV fills the gap on the following repeat. This figure was prepared with PyMOL [161].

Capping structures (discussed below) play similar roles by protecting the hydrophobic residues of the first and last LRRs. Consequently, the surfaces of the individual LRRs will show a significant degree of complementarity with those of their immediate neighbours or any capping structures covering them. Such complementarity should be taken into consideration in studies where truncated or chimeric LRR domains are engineered, as truncations or poorly complementary LRR surfaces might leave a cavity unprotected and destabilise the domain. It is always advisable to test the structural integrity of these engineered forms to ensure that the LRR domains have folded correctly.

There are two exceptions to a compact inner hydrophobic core in the LRR structures determined to date (Table 1). The CD14 structure [49] has an unusual architecture that exposes its hydrophobic core behind a collapsed N-terminal capping structure, creating a pocket that extends through the first two LRRs. This pocket is lined by the typical hydrophobic residues seen in other LRR proteins, yet inclusion of unusual tryptophan and arginine side chains contribute to maintain its opening. The maintenance of its hydrophobic character is thought to be essential for the recognition of the lipid chains of lipopolysaccharide, a known CD14 ligand [49]. The entire LRR structure is unusual in other aspects: it lacks the conserved asparagine residues from the consensus motif sequence and instead replaces them with small hydrophobic residues, and it uses several tryptophan residues in its inner hydrophobic core; these large groups extend the backbone chains from LRRs III, V and VIII on their convex sides while maintaining a compact hydrophobic core. The second exception is presented by the DLC1 structures [72, 73]. These structures show extended cavities running across the hydrophobic inner cores of several consecutive LRRs. They show very poor packing of internal residue side chains and the stacking of LRRs does not eliminate the resulting holes. The DLC1 structures are the only currently available example of LRR structural determination by NMR and present large, significant differences with structures of similar sequences determined crystallographically (already noted in [27]). We think that rather than representing an unusual LRR architecture, the DLC1 structures may show locally poor definition of side chain conformation and packing due to limited experimental information.

Not many studies have been reported on the stability of LRR proteins and their susceptibility to site-directed mutagenesis, if compared for example with the more extensive information available for globular proteins. The net stabilities of two SLRPs, biglycan and decorin, have been estimated to be lower than

2 kcal/mol based on their denaturation curves in guanidinium chloride and their melting temperatures measured by circular dichroism and microcalorimetry [52]. Scott and co-workers [51, 52] conclude that the LRR domains of these proteoglycans are relatively unstable and that dimerisation is essential for their stability at physiological temperature, presumably through the extensive side chain interactions seen in the dimeric crystal structures of decorin and biglycan. Indirect information on LRR stability can be deduced from an analysis of mutations in LRR proteins that result in diseases [29]. The authors of this study summarise the frequency of such mutations at each position of the consensus LRR sequences. Some mutations are expected to disrupt the ligand-binding surfaces on these LRR proteins (for example on their concave or convex sides), but others are more likely to have a structural impact resulting in loss of function. Obvious examples in this later category are mutations in the N-terminal and C-terminal caps, mutations to amino acids in the hydrophobic cores, changes to the consensus asparagine residues, deletions of entire LRRs, insertions of a few amino acid residues, and truncated proteins resulting from nonsense mutations [29]. Changes of similar nature in any LRR protein can be expected to have consequences for its thermal stability or structural integrity.

Curvature and twist in LRR structures

The structural principles of LRR protein domains have been the subject of analysis in the past [3, 26–29]. Given their curved (horseshoe) shape and the importance of the inner space in ligand binding, attempts have been made to define and predict the curvature of LRR domains. Several geometrical parameters have been defined by Enkhbayar and co-workers [28] (Fig. 7): the radius of curvature R, the overall angle of rotation about the central axis φ and the average angle per repeat $\langle\varphi\rangle$, the twist angle τ between adjacent repeats, and the tilt angle θ between the direction of the β-sheet and the axis of the best fitting circle. Analysis of these parameters for 14 LRR proteins available at the time indicated that structures with α-helices in their convex sides have more pronounced curvature (smaller R and larger $\langle\varphi\rangle$) than structures with 3_{10} or polyproline II helices (Fig. 8). This difference in curvature can be explained by the differences in diameter of the different secondary structure elements on the convex side, α-helices being wider than 3_{10} helices, polyproline II helices or tandem β-turns. Therefore, it is a mistake to model shorter repeats from longer ones by simply removing helical turns and assume that the curvature will remain intact. On the contrary, the spacing of adjacent β-strands, equivalent to the parameter $D = 2\,R \sin(\langle\varphi\rangle/2)$, is quite conserved (4.9–5.2 Å) across different LRR structures independently of the number of repeats or the radius or curvature [28]. Despite the general trend, the overall angle of rotation is not distributed evenly between repeats. Plots of step angles between consecutive LRRs against step number result in largely oscillating patterns (see Fig. 8 in [53]) that reflect only in part underlying alternances between lengths of consecutive repeats. Some structures even present individual kinks or bends where their general curvature trend is locally interrupted [38, 43, 50, 74].

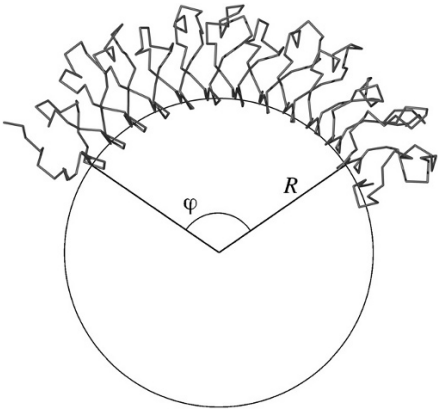

Figure 7. Overall curvature parameters for the concave side of LRR domains, as defined by Enkhbayar et al. [28]. A circle is fitted to the coordinates of equivalent atoms from each repeat (in this example the Cα atoms from the central residues in the parallel β-sheet). The method is illustrated with the DCN structure (PDB i.d. 1XKU), which gives a radius of curvature $R = 28$ Å and an overall angle of rotation $\varphi = 109°$.

The curvature of LRR structures results from the differences in cross-section between the β-strands in the concave side and the different helical structures of the convex side [28, 53]. This difference distinguishes them from the related β-helical structures, where the presence of continuous β-sheets on the different sides of the solenoid results in straight, tubular structures [64]. As mentioned earlier, several LRR structures contain small β-strands on the convex sides of their repeats. Due to the curvature of the LRR domains these strands cannot form continuous β-sheets on the convex side and appear either isolated or in small clusters of two, at most three hydrogen-bonded strands, with water molecules filling the gaps between strands too separated for hydrogen bonding (Fig. 5C). Alternatively, pairs of LRRs with 24–26 residues are intercalated with shorter repeats lacking β-strands. Such is the situation observed in SLRPs [53].

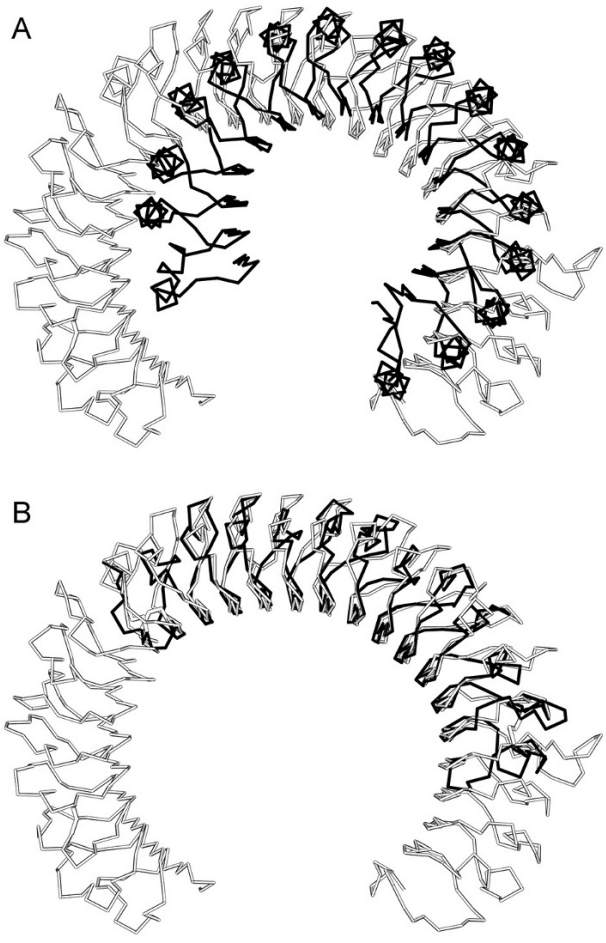

Figure 8. Different degrees of curvature in LRR domains. (*A*) Superposition of the RNI (black) and TLR3 (grey) structures (PDB i.d. 2BNH and 2A0Z, respectively) shown as Cα traces. The RNI structure has a smaller radius of curvature in its inner concave side and a larger average angle per repeat. (*B*) Superposition of the DCN (black) and TLR3 (grey) structures (PDB i.d. 1XKU and 2A0Z, respectively) shown as Cα traces. Both structures show similar radius of curvature in their inner concave sides and similar average angles per repeat. Figure prepared with SETOR [160].

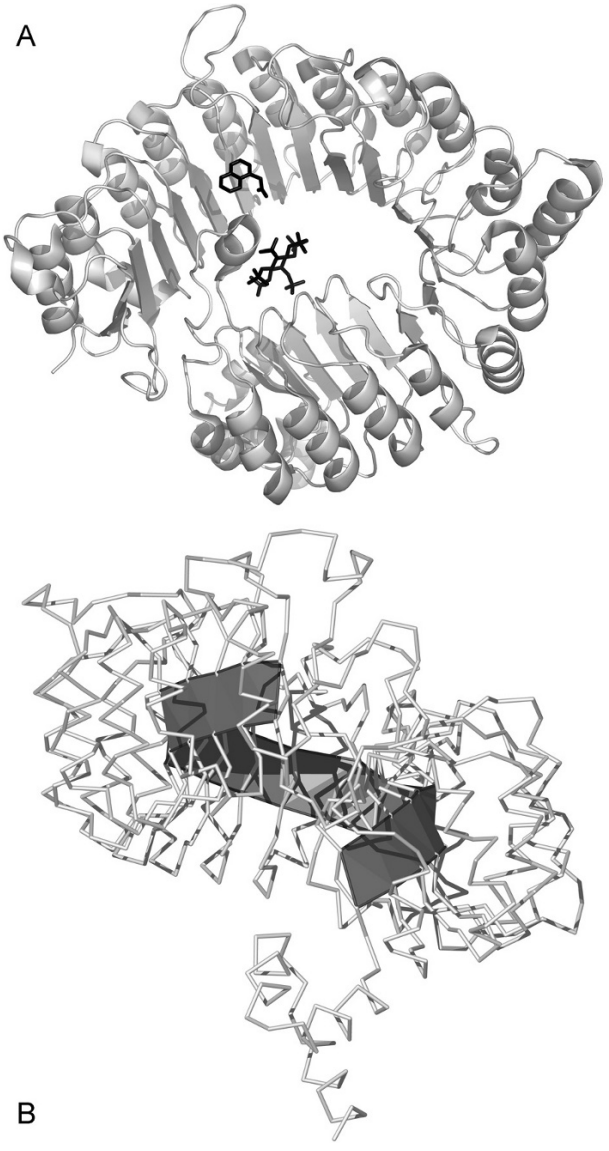

Figure 9. The TIR1 structure has enough LRRs to form a closed ring. The resulting enclosed inner space is surrounded by the parallel β-sheet. (*A*) Ribbon diagram of TIR1 (PDB i.d. 2P1O). The inner space provides the binding sites for two small ligands: inositolhexakisphosphate and the synthetic auxin analog 1-naphthalene acetic acid. (*B*) Cα trace representation of TIR1 with a ribbon connecting the consecutive strands of the parallel β-sheet from the LRR domain. Notice the helical path of the parallel β-sheet. Figure prepared with PyMOL [161].

Structures with a large value of the overall rotation angle φ such as RNI or TLR3 (Fig. 1A) suggest that LRR structures with sufficient repeats could form closed circles (or more properly toruses). The recently determined TIR1 structure forms one such closed ring where the continuously running β-sheet in the concave side lines an internal cavity (Fig. 9) that is used for binding small plant hormones (auxins) and inositol hexakisphosphate as a cofactor. In fact, the TIR1 structure can be better described as a whole turn of a helix [74], and the internal β-sheet follows a helical path as illustrated in Figure 9B.

The known tendency of β-sheets to adopt a right-handed twist is reflected in the shape adopted by the concave side of the LRR structures. The parallel β-sheet can be regarded as a surface analogous to a Möbius strip [28, 29], and most LRR structures show a right-handed twist, although there is a significant degree of variability in the average degree of twist per LRR and there are even some structures with left-handed overall twists [28].

Ligand binding and quaternary structure of LRR proteins and domains

The concave space defined by the curved LRR architecture presents a suitable environment for macromolecular ligand binding. Structures with large overall rotation angles ($\varphi > 180^\circ$) such as RNI, TLR3, INLA or YOPM, define inner circular regions with *R* values between 18 and 25 Å [28]. The term "horseshoe" is usually applied to describe the shape of these molecules. Globular protein ligands with radii in that range would be largely encased in these inner spaces, thus maximising the number of protein-protein contacts. Several three-dimensional structures of complexes between LRR proteins or domains and their ligands have been determined (Table 2). Many show the mode of binding just described, with the ligand surrounded to large or lesser extent by the concave surface of the parallel β-sheet from the LRR domain (Figs. 1C, D; 10A, B). In general, ligand binding does not introduce major rearrangements in the structure of the LRR domains and the free and bound forms superimpose quite well, which may present an advantage in terms of reduction of entropic penalty upon binding [69]. In some cases the radius of curvature is seen to increase slightly in the bound form [28], suggesting some degree of structural flexibility to accommodate larger ligands.

Typical values of the total buried surfaces in complexes between LRR domains and their ligands are in the 2500–3400-Å^2 range for RNI, FSHR and INLA bound to different ligands, to more moderate values as the size of the ligand and the concave surface decrease (Table 2). In principle, large buried surfaces can be indicative of very tight complexes. Those of RNI with different members of the ribonuclease superfamily show equilibrium dissociation constants in the femtomolar range [42, 75], amongst the tightest on record for protein-protein complexes. Conversely, the interaction between INLA and the N-terminal domain of E-cadherin (EC1) is surprisingly weak, in the micromolar range [35]. In this case, just two rationally chosen amino acid substitutions can increase the binding affinity by four orders of magnitude [45, 76]. Therefore, the amount of buried surface alone is not a reliable indicator of the strength of the interaction between LRR domains and their ligands, and specific interactions need to be taken into account. Ligands are more irregular in shape than the LRR concave surface to which they bind; thus, these specific interactions will occur across discrete groups of residues at both sides of the interface. Consequently, it is generally observed that LRR ligands tend to make few non-contiguous contacts with the curved LRR β-sheet (Fig. 10). This imperfect surface complementarity and the need for water molecules to fill the resulting gaps are throught to be responsible for the weak binding affinity observed in the INLA-EC1 complex [45]. Electrostatic interactions can make an important contribution to these non-contiguous contacts, as illustrated by the extremely large electrostatic energies of interaction between RNI and angiogenin [77] and the effects of mutations in charged residues on the stability of RNI complexes with ribonuclease 1 [42]. Overall, the studies on RNI and INLA complexes with different ligands indicate that the formation of high-affinity associations between LRR proteins or domains and their ligands relies on the existence of sufficient short-range interactions across the interface, normally clustered in non-contiguous regions, and that charged residues may be particularly efficient in increasing the binding affinity by forming tight hydrogen bonds and increasing the association rates of the complexes [42]. Interestingly, changes in one or two residues may be sufficient to increase or decrease the binding affinity by several orders of magnitude.

The same principles apply to LRR domains and proteins with shorter overall rotation angles ($\varphi < 180^\circ$), which can been described as "banana"-shaped [70]. Banana-shaped LRR proteins and domains do not wrap around their ligands so extensively as the horseshoe-shaped ones, yet they form tight associations with equilibrium dissociation constants in the nanomolar range. Several structures of banana-shaped LRR molecules in complex with ligands have been determined (Table 2). An interesting example in this category is that of GPIBA bound to the A1 domain of von Willebrand factor (VWF-A1) [36, 37]. The interacting surface on the concave side of GPIBA is extended but discontinuous (Fig. 10B), with two main sites of interaction: a β-hairpin that is part of the N-terminal capping motif (LRRNT) of GPIBA, and a flexible loop at the C-terminal capping region (LRRCT), that is disordered in the uncomplexed GPIBA structure but adopts a β-hairpin structure upon complex formation. These two sites have been named β-finger and β-switch, respectively, and extend as grasping arms from the concave side of the GPIBA LRR domain to embrace the VWF-A1 domain. This arm-assisted binding represents a variation to the general theme of the interaction between LRR domains and their ligands. The total buried surface in this complex is around 2100 Å^2, with nanomolar-range dissociation constants [36]. Large patches of opposite charge on the interacting surfaces of GPIBA and VWF-A1 have been proposed to promote long-range electrostatic interactions rather than participating directly in complex stabilisation [36]. A similar binding strategy is observed in the crystal structure of

Table 2. Representative quaternary assemblies seen in LRR crystal structures. Values of total buried surface calculated using the Protein Interfaces, Surfaces and Assemblies (PISA) server [158, 159].

	PDB i.d.	Buried surface (Å^2)	References
Ligand binding mainly to the concave side			
RNI complex with eosinophil-derived neurotoxin	2BEX	3430	[41]
RNI complex with ribonuclease 1	1Z7X	2880	[42]
RNI complex with angiogenin	1A4Y	2680	[34]
RNI complex with ribonuclease A	1DFJ	2590	[33]
FSHR complexed with its ligand FSH	1XWD	2580	[38]
INLA in complex with the N-terminal domain of E-cadherin	1O6S 2OMY	2650, 2620	[35, 45]
GPIBA in complex with von Willebrand factor A1 domain	1M10 1SQ0	2090, 2110	[36, 37]
U2B"-U2A' complex bound to a fragment of small nuclear RNA	1A9N	1980	[12]
SLIT2D2 in complex with the 1st Ig domain from Robo1	2V9T	1380	[44]
TLR4S in complex with lymphocyte antigen 96 (MD-2) bound to endotoxin antagonist Eritoran (concave side binding)	2Z65	1230	[43]
TIR1 bound to inositol hexakisphosphate	2P1M	740	[74]
TIR1 bound to different auxins	2P1N 2P1O 2P1P 2P1Q	460–400	[74]
Ligand binding to the C-terminal region and concave side			
Complex between INLB and receptor tyrosine kinase Met	2UZX 2UZY	2850	[100]
Ternary complex between GPIBA, von Willebrand factor A1 domain and borocetin	1U0N	2830	[39]
SKP2–SKP1 complex (C-terminal binding)	1FQV	2980	[153]
Ternary complex SKP2–Skp1–Cks1 (Skp1 to C-terminal region, Cks1 to concave side)	2ASS	2990, 1760	[40]
Quaternary complex SKP2–Skp1–Cks1–P27 phosphopeptide (Skp2 to C-termina region, Cks1/P27 to concave side)	2AST	2950, 2261	[40]
TLR4S in complex with lymphocyte antigen 96 (MD-2) bound to endotoxin antagonist Eritoran (C-terminal binding)	2Z65	1680	[43]
GPIBA in complex with thrombin, site II	1OOK	1100	[47]
Ligand binding to the ascending and concave sides			
Ternary complex RNA1P–Ran (GNP-Mg)–RanBP1	1K5D	2630	[46]
Mouse TLR4 in complex with mouse MD2	2Z64	2210	[43]
Ligand binding to the descending and concave sides			
GPIBA in complex with thrombin, site I	1OOK 1P8V	2120, 1410	[47, 48]
Ligand binding to both terminal regions			
TIR1 in complex with Skp1-like protein 1A (ASK1)	2P1M	3150	[74]
Ligand binding mainly to the convex side			
GPIBA in complex with thrombin, site III	1P8V	1890	[48]
Dimerization through the concave side			
BGN antiparallel dimer (concave to concave side)	2FT3	2630	[52]
DCN antiparallel dimer	1XKU	2320	[51]
INLA antiparallel crystallographic dimer 1	1O6V	3740	[35]
TLR2 antiparallel crystallographic dimer	2Z80	2970	[50]
INLH antiparallel crystallographic dimer	1H6U	2190	[98]
VLRA29 antiparallel crystallographic dimer	2O6Q	1670	[71]
SLIT22 antiparallel crystallographic dimer	2V9S	1490	[44]
GPIBA crystallographic dimer 1 (convex side to concave side of another monomer)	1GWB	3330	[56]
TLR4 crystallographic dimer 1 (LRRCT to concave side of another monomer)	2Z62	2100	[43]
NOGO crystallographic dimer (LRRCT to concave side of another monomer)	1P8T	1570	[70]
GPIBA crystallographic dimer 2 (LRRCT to concave side of another monomer)	1P9A 1QYY	1550	[47, 144]
RNA1P crystallographic dimer (N-terminus and descending side to concave side of another monomer)	1YRG	1330	[150]
Dimerization through the ascending side			
TLR1H–TLR2H heterodimer	2Z7X	1620	[50]
TLR3 crystallographic dimer	2A0Z 1ZIW	1530, 1260	[57, 84]
TLR4 crystallographic dimer 2 (LRRCT to ascending side of the other monomer)	2Z63	1770	[43]
INLA antiparallel crystallographic dimer 2, ascending side to ascending side	1O6T	2040	[35]

Table 2 *(Continued)*

	PDB i.d.	Buried surface (Å²)	References
C-terminus to C-terminus dimers			
TAP crystallographic dimer	1FT8 1KOO	1880, 1630	[13, 155]
CD14 tail-to-tail crystallographic dimer	1WWL	1220	[49]
Larger assemblies			
LINGO1 crystallographic tetramer	2ID5	9330	[88]
YOPM crystallographic tetramer	1JL5	9170	[67]

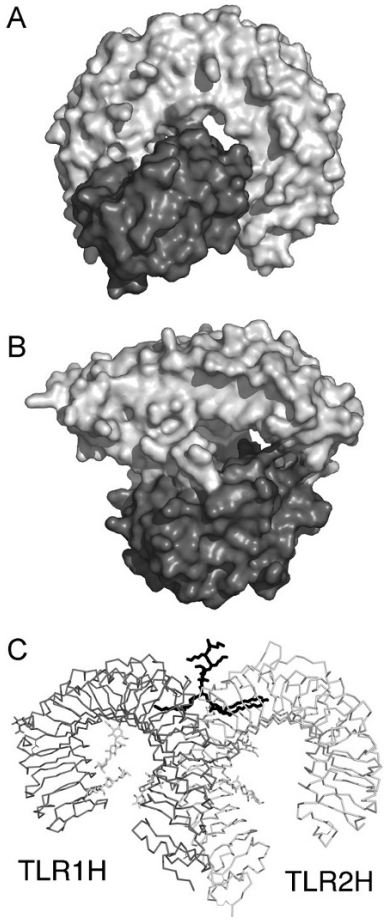

Figure 10. Examples of quaternary arrangements involving LRR domains. (*A*) Surface representation of the complex between RNI (light grey) and ribonuclease 1 (dark grey) (PDB i.d. 1Z7X). The concave inner side of RNI is almost completelly filled by the bound ligand. (*B*) Surface representation of the complex between GPIBA (light grey) and VWF-A1 (dark grey) (PDB i.d. 1M10). (*C*) Cα trace representation of the heterodimer of TLR1H and TLR2H induced by shared binding to a lipopeptide ligand (shown in black) (PDB i.d. 2Z7X). Notice the lipid chains of the ligand penetrating deeply the hydrophobic cores of both LRR domains. A very similar dimer is observed in the TLR3 structures (PDB i.d. 1ZIW, 2A0Z), between two symmetry-related monomers. Figure prepared with PyMOL [161].

the spliceosomal U2B"-U2A' protein complex bound to an RNA fragment [12]. The U2A' protein contains an LRR domain whose concave side binds the ribonucleoprotein domain of U2B". This LRR domain shows a β-hairpin N-terminal extension and a loop extension in its C-terminal region. These structures form two arms that reach out from the concave side at both its terminal ends and embrace the ribonucleoprotein domain of U2B" [12], in a manner completely equivalent to the β-finger and β-switch from GPIBA interacting with VWF-A1.

The more recent structures of the complex between SLIT2D2 and the N-terminal domain of Robo1 [44] and the complex between FSHR and FSH [38] show that both lack terminal arm structures. The interface between SLIT2D2 and Robo1 is non-contiguous and can be divided in two regions: one is predominantly electrostatic in nature with direct salt bridges and hydrogen bonds between the two proteins; the other is predominantly hydrophobic with extensive contacts between hydrophobic side chains reinforced by hydrogen bonding interactions. The total buried surface in this complex is relatively small (1380 Å^2, Table 2) but the dissociation constant remains in the low nanomolar values [44]. The FSHR/FSH complex (Fig. 1D) is unusual in that the interface is contiguous and all β-strands from the concave side of FSHR are in contact with the bound FSH ligand. As a consequence the total buried surface in this complex is very large (Table 2) and comparable to that seen for the complexes of horseshoe-shaped RNI or INLA with their ligands. Electrostatic interactions also seem to be very important in the FSHR/FSH complex, with an exceptionally high buried-charge density, several direct salt-bridge interactions across the interface, and complementarity of the predominant charge on the interacting surfaces of ligand and receptor [38, 78].

The completely enclosed inner space of TIR1 in complex with different auxins (plant hormones) (Fig. 9) show how closed ring LRR domains can be

used to bind small molecules. The bound auxins are proposed to act as molecular glues that enhance TIR1-substrate interactions [74].

Although ligand-binding through concave surfaces is the most common in LRR domains and proteins, it is not exclusive. Several structures of complexes show additional surfaces engaged in ligand binding. For example, binding of Ran(GNP)–RanBP1 to RNA1P occurs at one side (ascending) of RNA1P [46]. Some LRR proteins and domains are glycosylated in their concave sides, which makes them partially or completely unsuitable for ligand binding. Thus, the complex between the mouse versions of TLR4 and MD-2 involves both the concave and ascending sides of TLR4 as part of its concave surface is blocked by N-linked oligosaccharide structures [43]. A few structures show the C-terminal regions engaged in ligand binding (Table 2) and in the case of TIR1 its closed ring structure allows the partner protein ASK1 to bind simultaneously to both terminal regions [74]. From the structures available to date, it seems that the convex and descending sides of LRR domains are not preferred binding surfaces (Table 2). Three different thrombin interaction sites with GPIBA have been reported from two crystal structures [47, 48, 79]. Of those, site I is common to both structures and on the descending side of the LRR domain of GPIBA near the C terminus, whereas binding of thrombin to the convex side of GPIBA (site III) is only observed in one of the structures [48]. It seems, however, that the thrombin-GPIBA interaction is dominated by GPIBA residues Leu275–Tyr279, which are C-terminal to the LRR domain, and specifically by two sulphated tyrosine residues in that sequence [80]. Additional interactions with different sites on the surface of the LRR domain would then be secondary.

Many of the LRR proteins and domains crystallised to date show putative quaternary structures with extensive dimerisation or multimerisation interfaces. It is not easy to conclude, from the arrangement of monomers alone, if these oligomerisation states represent stable structures in solution or are simply unstable associations brought together by the crystalline lattice. The total buried surfaces of some of these dimers and oligomers are comparable to those observed for very stable complexes (Table 2), but as we have seen earlier these measurements on their own are not reliable indicators of high affinity macromolecular association. The BGN and DCN structures show dimeric arrangements where the concave surfaces provide the dimerisation interface (Fig. 1B). Several close interactions occur between monomers across the interface and the residues involved are conserved in related SLRPs [51–53]. There is ample experimental evidence for the existence of BGN and DCN dimers in solution even at subnanomolar concentration [51, 52, 81], and the observed small angle X-ray scattering of DCN in solution is consistent with the dimeric arrangement seen in the crystal structures [53, 81]. The total buried surface in these dimers is 2300–2500 $Å^2$, in the same range as the complexes of RNI or INLA with their ligands. It is still unknown if the DCN and BGN dimers are the functional units or if they are able to dissociate upon ligand binding [53]. However, the LRR domains of these proteoglycans appear to be relatively unstable and it has been shown that dimerisation is essential for their stability at physiological temperature [52].

Dimerisation through the concave sides has been predicted for other LRR proteins known to form dimers such as opticin [53, 82] or the fourth LRR domain from Slit [83], and antiparallel dimers similar to those of BGN or DCN are actually observed in the crystal structures of quite a number of structures (Table 2), some burying large surfaces at their interfaces. Other crystallographic dimers are observed in which alternative sides provide the interface. Particularly intriguing is the tail-to-tail dimeric arrangement observed in the crystal structure of CD14, which almost perfectly mimics a larger LRR domain with twice the number of repeats and a large, continuous β-sheet in its concave surface ([49] and Fig. 1 therein) The existence of CD14 dimers in solution is reported by the same authors, although to date it is unclear if these dimers have any biological role (their similarity in shape and dimensions with the TLR3 structure and possible interactions with TLR4 are discussed in [84]). However, for most of the crystallographic dimers noted in Table 2 there is no biophysical evidence of stability in solution or functional evidence of dimerisation (with the important exceptions discussed below), and thus it should be assumed that they are a consequence of the respective crystalline lattices and that the normal quaternary state of these proteins is monomeric.

The important exceptions mentioned relate to the dimeric structures of TLRs. TLRs are pattern recognition receptors from the immune innate system that recognise conserved molecular structures in pathogens. They all share an ectodomain with LRR architecture [23, 85]. The structure of TLR3 [57, 84] (Fig. 1A) shows that most of its molecular surface is masked by glycosylation, including the concave, convex and descending sides; a similar situation was predicted for other TLRs based on sequence analysis [84]. The glycosylation of the concave side of TLR3 and other TLRs was initially unexpected, as the concave surface had been predicted to contain the ligand binding sites of these receptors [86]. Analysis of charge distribution on the TLR3 surface identified a

positive charge region on the glycosylation-free ascending side, thus suggesting a binding site for dsRNA, the ligand for TLR3 [57, 84]. Interestingly the ascending side was also seen to provide the interface for a TLR3 crystallographic dimer. The total buried surface of that dimer is relatively small (Table 2) and TLR3 as a soluble ectodomain is monomeric in solution. However, it was noted that full-length TLR3 is a membrane protein and could form a stable or transient dimer when embedded in the membrane. Such dimer could provide an extended binding surface where two TLR3 molecules could bind simultaneously to two sides of the dsRNA molecule. A model for dsRNA binding and signalling involving dimer formation has been discussed [84]. Further understanding of the interaction of TLRs with their ligands has come with the recent structural determinations by Jie-Oh Lee and colleagues. Two landmark papers [43, 50] reveal how TLRs recognise their ligands and show evidence of ligand-induced dimerisation of TLR LRR domains. The structure of the lipopeptide induced TLR1H–TLR2H heterodimer (Fig. 10C) [50] shows the same mode of dimerisation observed between identical monomers in the TLR3 structure, through the ascending sides of both monomers. A single molecule of lipopeptide is shared by both TLR1H and TLR2H by inserting two of its lipid chains into a pocket on TLR2H and a third lipid chain into a narrow channel in TLR1H [50]. Both pocket and channel are on the ascending sides of their respective TLRH molecules and run deeply inside the hydrophobic core of the LRR domains (Fig. 10C). These interactions illustrate yet another mode of ligand binding for the versatile LRR architecture. On a separate paper [43], the same group presents two structures of TLR4 and TLR4S bound to MD-2, a soluble protein required for TLR4 signalling, plus evidence of lipopolysaccharide-induced dimerisation of the TLR4/MD-2 complex by gel filtration chromatography, native PAGE analysis and cross-linking. Together these studies show the existence of at least two different modes of ligand recognition by TLRs and provide model scenarios for ligand-induced TLR activation *via* ectodomain dimerisation [43, 50, 87].

Higher crystallographic oligomers present additional difficulties in terms of biological relevance. Two interesting cases are presented here (Table 2). The YOPM structure forms the same tetrameric structure in three different crystal forms [67]. The YOPM tetramer is a double helix that defines a hollow cylinder with an inner diameter of about 35 Å. Each strand is a continuous end-to-end dimer of YOPM monomers connected through their C-terminal repeats, in a manner similar to the CD14 dimer discussed above. The interaction between the four monomers buries a considerable amount of hydrophobic surface and most of the exposed hydrophobic surfaces on the monomers are involved in protein-protein contacts in the tetramer. Still, the biological significance of the YOPM tetramer is unclear at present time. It is possible that the tetramer is a crystallisation artifact resulting from high concentration of calcium ions [67], which are seen mediating interactions between monomers in the tetramer. Gel filtration chromatography coupled to light scattering shows a monomeric state either with or without calcium, whereas cross-linking indicates oligomerisation upon addition of calcium [67]. The second example of LRR domain tetramerisation is the LINGO1 structure. Four LINGO1 molecules form a ring-like, closed tetramer with approximate four-fold symmetry and an internal hole of approximately 45 Å in diameter [88]. LINGO1 tetramers are seen in two different crystal forms and have been detected in solution by gel filtration chromatography, dynamic light scattering, analytical ultracentrifugation and chemical cross-linking at micromolar concentrations, where the protein is predominantly monodisperse in its tetrameric form. Thus, LINGO1 oligomerisation is not an artifact of crystallisation and may have some yet uncharacterised functional role, perhaps facilitating the formation of multimeric signalling complexes on the cell surface [88].

Prediction of ligand binding sites

Research groups investigating the structure and function of LRR proteins worldwide have been remarkably successful in elucidating the molecular basis of ligand binding by direct structural determination, as the large number of complexes in Table 2 attests. However, it is still necessary in some cases to deduce possible regions of ligand binding from the structures of the unbound LRR proteins and domains, as obtaining the desired complexes may prove difficult and time consuming. Often, recently determined structures of LRR proteins or domains can be used to generate reliable *in silico* models for homologous LRR proteins and these models can be investigated for clues on ligand binding that may guide further mutagenesis and functional studies. In favourable cases, the existence of significant hydrophobic pockets or extended LRRs may be combined with mutagenesis studies to identify specific functional sites [49]. In less favourable cases several strategies of increasing levels of sophistication have been used to identify putative ligand binding sites on LRR structures. Typically, the structure of the unbound LRR domain will be analysed for clustering of

charged residues, hydrophobic residues, exposed aromatic residues, exposed histidine residues, or any other category of amino acids that may seem appropriate to the researcher. Different categories are typically displayed on molecular surface representations of the LRR domains using appropriate colour schemes (see for example Fig. 6 from [70] or Fig. 4 from [69]). This procedure, sometimes referred to as "hotspot" analysis, relies on the accumulated knowledge of the molecular characteristics of surface interactions in macromolecular complexes determined by X-ray crystallography. Conspicuous surface clusters (patches) are suggestive of possible regions involved in ligand recognition and binding. Analysis of conservation in homologous sequences is normally performed in parallel and different levels of conservation are mapped on molecular surface representations using some colour scheme (see for example Fig. 5 from [71]). Any information available on members of a family known to share binding to a common ligand or to show differential specificity is checked against residue conservation trends.

The relatively simple topology of LRR proteins and domains allows performing some of these analyses on schematic surface projections of the different surfaces. Buried structural residues can be omitted for clarity, and individual residues are shown as "square pixels" in a two-dimensional representation. This procedure was pioneered in the analysis of the YOPM structure (Fig. 10 in [67]), and was used to illustrate the conservation of residues involved in dimerisation of members of the SLRP family (Fig. 3 from [51]; Fig. 9 from [53]).

Automatic methods for the analysis of sequence variability and hotspot identification are currently available [89–94]. These methods identify conserved residues and map them on the known protein structures, aiming to locate clusters of conservation on the protein surface and to predict protein-protein binding sites. They can be applied to the investigation of protein binding sites on LRR structures, shown for example in the prediction of the hypervariable concave surfaces of hagfish variable lymphocyte receptors (VLRs) as the most probable antigen binding sites [71].

LRR primary structure and automatic sequence annotation

It is not uncommon in the LRR literature to find the terms "domain" and "repeat" being applied indistinctly to the same structural or sequence entity. Such an approach can often lead to confusion and should be avoided. Thus, individual repeats should not be described as "domains" but rather the term "LRR domain" should only be applied to clearly separate sets of two or more LRRs in tandem. Similarly, capping structures that integrate into the LRR fold and do not exist as independently folded units should not be classified as domains on their own (the terms "motif", "module" or "subdomain" seem more appropriate in this case). This recommended nomenclature is illustrated for the *Drosophila* Slit (Table 3), a multidomain protein with four LRR domains in tandem, each of them containing its own tandem of 6–8 LRRs capped both at the N terminus and C terminus with LRRNT and LRRCT motifs (discussed below). The current annotation for the Slit sequence in the Swiss-Prot database (accession code P24014, SLIT_DROME), which assigns 24 LRRs numbered 1–24, can convey the wrong notion of a continuous, single LRR domain with 24 repeats similar to the TLR3 structure.

An additional level of confusion arises from the intrinsic repetitive nature of LRR sequences. Given a protein sequence containing several LRRs, there are many possible ways of defining the origin of each repeat. Depending on what convention is used in automatic sequence annotation, the apparent number of repeats of a given LRR protein may change slightly and some repeats may not be identified at all. Furthermore, the presence of extended LRRs adds its own difficulty to the correct identification of the complete repeat structure of a given LRR sequence. Consequently, in the absence of any three-dimensional structural information, automatic sequence annotation methods are very reliable in the identification of LRR-containing protein sequences but very inefficient in predicting correctly the number of LRRs, their lengths and their boundaries.

The most straightforward way of defining the repeat structure of an LRR protein is to place the consensus sequence at the beginning of each repeat. Thus, individual LRRs adopt the consensus form $\mathrm{L}xx\mathrm{L}x\mathrm{L}xx\mathrm{N}x\mathrm{L}x_n$ where n depends on the length of the repeat and the leucine residues are often replaced by other hydrophobic amino acids. Consecutive LRRs of different lengths can then be aligned so that the mainly conserved region is on the left, corresponding to the parallel β-sheet on the concave side, and the variable region is on the right, corresponding to the convex side and the ascending and descending loops. This convention, illustrated for the *Drosophila* Slit protein in Table 3, also allows for easy incorporation of extended LRRs as repeats with unusually long variable regions (Fig. 4), since the extensions are unlikely to occur on the consensus region.

The LRR motif is described in several sequence databases for protein domain identification and

Table 3. Domain and repeat organization in the LRR region of the *Drosophila* SLIT protein (SwissProt i.d. P24014). Positions of residue conservation are shown in bold type. Roman numerals indicate LRR number on each domain.

First LRR domain		
LRRNT		R**C**PRV**C**S**C**T
	I	GLN**V**D**C**SHRGLTS**V**PRKISAD
LRR	II	**V**ER**L**EL**L**QG**N**N**L**TV**I**YETDFQRLTK
	III	**L**RM**L**QL**L**TD**N**QI**H**TI**E**RNSFQDLVS
	IV	**L**ER**L**RL**L**NN**N**R**L**KA**I**PENFVTSSAS
	V	**L**LR**L**DI**S**NN**V**I**T**TV**G**RRVFKGAQS
	VI	**L**RS**L**QL**D**NN**Q**I**T**C**L**DEHAFKGLVE
	VII	**L**EI**L**TL**N**NN**N**L**T**S**L**PHNIFGGLGR
LRRCT	VIII	**L**RA**L**RL**S**D**N**P**F**A**C**D**C**HLSWLSRFLRSAT RLAPYTR**C**QSPSQLKGQNVADLHDQEFK**C**SGLTEHAPMECGAEN
Second LRR domain		
LRRNT		S**C**PHP**C**R**C**A
	I	DGI**V**D**C**REKS**L**TS**V**PVTLPDD
LRR	II	TTE**L**R**L**EQ**N**F**I**TE**L**PPKSFSSFRR
	III	**L**RR**I**DL**S**NN**N**I**S**R**I**AHDALSGLKQ
	IV	**L**TT**L**V**L**YG**N**K**I**KD**L**PSGVFKGLGS
	V	**L**QL**L**LL**N**A**N**E**I**SC**I**RKDAFRDLHS
	VI	**L**SL**L**S**L**YD**N**N**I**QS**L**ANGTFDAMKS
LRRCT	VII	IKT**V**H**L**AK**N**P**F**I**C**D**C**NLRWLADYLHKNP IETSGAR**C**ESPKRMHRRRIESLREEKFK**C**SWDELRMKLSGECRMDS
Third LRR domain		
LRRNT		D**C**PAM**C**H**C**E
	I	GTT**V**D**C**TGRG**L**KE**I**PRDIPLH
LRR	II	TTE**L**L**L**ND**N**E**L**GR**I**SSDGLFGRLPH
	III	**L**VK**L**E**L**KR**N**Q**L**TG**I**EPNAFEGASH
	IV	**I**QE**L**Q**L**GE**N**K**I**KE**I**SNKMFLGLHQ
	V	**L**KT**L**N**L**YD**N**Q**I**SC**V**MPGSFEHLNS
LRRCT	VI	**L**TS**L**N**L**AS**N**P**F**NCNCHLAWFAEWLRKKS LNGGAAR**C**GAPSKVRDVQIKDLPHSEFK**C**SSENSEGCLGDG
Fourth LRR domain		
LRRNT		Y**C**PPS**C**T**C**T
	I	GTV**V**R**C**SR**N**Q**L**KE**I**PRGIPAE
LRR	II	TSE**L**Y**L**ES**N**E**I**EQ**I**HYERIRHLRS
	III	**L**TR**L**D**L**SN**N**Q**I**TI**L**SNYTFANLTK
	IV	**L**ST**L**I**I**SY**N**K**L**QC**L**QRHALSGLNN
	V	**L**RV**L**S**L**HG**N**R**I**SM**L**PEGSFEDLKS
LRRCT	VI	**L**THIA**L**GS**N**P**L**Y**C**D**C**GLKWFSDWIKLDY VEPGIAR**C**AEPEQMKDKLILSTPSSSFV**C**RGRVRNDILAKCN

analysis such as SMART [95], Pfam [96] and InterPro [97] (Table 4). Different databases may use slightly different definitions of the LRR motifs, especially in relation to the position of the consensus sequence with respect to the repeat boundaries. Tools for automatic sequence annotation built in these databases will identify most protein sequences that have an LRR structure, but will almost invariably fail to identify all the LRRs from a given protein sequence, as individual LRRs can easily tolerate one or more substitutions of amino acids in the consensus sequence to non-hydrophobic or charged ones. However, subsequent manual annotation should easily detect all the LRRs in a protein sequence once it has been identified as a member of the LRR superfamily.

A classification system for LRR subfamilies was proposed years ago by Kajava and co-workers [3, 26, 27] on the basis of patterns of sequence conservation. Seven subfamilies were defined (Table 5): (1) RI-like (*e.g.* RNI, RNA1P), with typical LRRs 28–29 residues long, α-helical conformation in their convex sides (Fig. 2A), very little twist, and a high degree of curvature; (2) SD22-like (*e.g.*, INLA, INLB, INLC, INLH), with typical LRRs 22 residues long and 3_{10} conformation in the convex side (Fig. 2D); (3) cysteine-containing (SKP2), with LRRs usually 26 residues long and a distinctive consensus sequence; (4) bacterial (YOPM), with short LRRs typically 20 residues long and polyproline II conformation in their convex sides (Fig. 2B, C); (5) typical (*e.g.*

Table 4. Current database descriptors for LRR and related structural motifs. The third column for each database shows number of sequences containing the motifs (InterPro), number of motifs (Pfam) and both (SMART), as of 1 January 2008. Associated motifs, capping structures and some related domains are also included. See text for a description of the different entries.

	InterPro			SMART			Pfam		
LRR	IPR001611	LRR	15555				PF00560	LRR_1	59894
	IPR013101	LRR_2	720				PF07723	LRR_2	582
	IPR011713	LRR_3	371				PF07725	LRR_3	336
				SM00370	LRR	107629 / 18380			
LRR_typ	IPR003591	LRR_sub-typ	3715	SM00369	LRR_TYP	18512 / 7830			
LRR_ri	IPR003590	LRR_RNase_inh _sub-typ	22	SM00368	LRR_RI	127 / 97			
LRR_cc	IPR006553	LRR_cys_ sub-typ	147	SM00367	LRR_CC	744 / 367			
LRR_bac				SM00364	LRR_BAC	0 / 0			
LRR_sd22				SM00365	LRR_SD22	0 / 0			
Capping									
LRRNT	IPR000372	LRR_cys_N	2484	SM00013	LRRNT	5596 / 5033	PF01462	LRRNT	1572
LRRNP	IPR013210	LRRNT_2	2082				PF08263	LRRNT_2	1326
LRRCT	IPR000483	LRR_C	1341	SM00082	LRRCT	4715 / 3903	PF01463	LRRCT	399
LRRCap	IPR003603	U2A'_phospho protein32A_C	338	SM00446	LRRcap	871 / 795			
LRRCI	IPR012569	LRR_adjacent	489				PF08191	LRR_adjacent	375
Variants									
L domain	IPR000494	EGF_rcpt_L	486				PF01030	Recep_L_domain	894
LRRV	IPR004830	LRV	54				PF01816	LRV	321

Table 5. Consensus sequences of the LRR subfamilies according to Kajava and Kobe [3, 26]. '–', possible insertion sites; 'ϕ', non-polar residue; 'x', any residue. Residue conservation positions are shown in uppercase (more than 50 % of sequences) or lowercase (more than 30 % of sequences).

Subfamily	Conserved region											Variable region																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
RI-like	L	x	x	L	x	L	x	x	$^{N}/_{C}$	x	L	x	x	x	g	ϕ	x	x	L	x	x	o	L	x–	x	x	x	x
SD22-like	L	x	x	L	x	L	x	x	N	x	I	x	x	I	x	x	L	x–	x	L	x	x						
Cys-containing	L	x	x	L	x	L	x	x	c	x–	x	I	T	D	x	x	ϕ	x	x	L	a	x–	x	c	x	x		
Bacterial	L	x	x	L	x	V	x	x	N	x	L	x	x	L	P	$^{e}/_{d}$	L	–	P	x	x							
Typical	L	x	x	L	x	L	x	x	N	x	L	x	x	L	p	x	x	Φ	F	x–	x	L	x	x				
Plant specific	L	x	x	L	x	L	x	x	N	x	L	$^{t}/_{s}$	g–	x	I	P	x	x	L	G	x	L	x–	x				
TpLRR	L	x	x	I	x	L	x–	x	x	L	x	x	I	g	x	x	A	F	x	x	$^{C}/_{N}$	x–	x					

GPIBA, NOGO, DCN, LINGO1, VLRA29), the most populated subfamily with LRRs typically 24–26 residues long and 3_{10} helical or tandem β-turn conformations in their convex sides (Fig. 2F); (6) plant specific (PGIP2), with LRRs similar in length to the 'typical' subfamily but with different consensus sequence; and (7) TpLRR (no structure available yet), with 23-residue LRRs lacking the conserved cysteine or asparagine residues. Descriptions for some of these subfamilies were initially incorporated in several of the sequence databases (Table 4) but they are slowly becoming abandoned in favour of more generic descriptors that can be applied automatically to as many LRR sequences as possible. Hence, the depopulation observed in different databases for the SD22 or 'bacterial' LRR subfamilies.

What is the shortest sequence for an LRR?

Traditionally it has been accepted that the lengths of the LRR motifs range between 20 and 30 amino acids. An upper limit of 30 amino acids is probably correct for regular LRRs, but extended or terminal LRRs may have significantly higher numbers of residues (see Table 1 and Fig. 4 for some examples of very long LRRs). The absolute lower limit for an LRR is normally considered 20 residues, and such is the length of the shortest LRRs of known structure (as seen in the YOPM structure). Length annotation for individual repeats in sequence databases such as SwissProt can sometimes show very short values for LRR lengths, but these are often the result of partial identification of repeats, ignoring the irregular regions of extended repeats, or use of different conventions on the beginning and end of LRR sequences.

A more convincing case for LRRs shorter than 20 residues is presented by the sequence of the Mimivirus protein R380 (Table 6). Analysis of the repeat structure of that protein sequence suggests at least 14 short LRRs, of which 8 have 19 residues, including a tandem of 6 consecutive 19-residue LRRs. The conservation of Pro residues in the variable region and sequence similarity with the 20-residue-long LRRs from YOPM suggests a polyproline II conformation for the convex side of these repeats (as in Fig. 2B, C). Nevertheless, no further LRR sequences have been identified to date containing consecutive 19-residue LRRs; this suggests that these repeats are exceptionally short and probably represent the minimal length compatible with an LRR structure.

Table 6. Predicted repeat structure of the putative mimivirus LRR protein R380 (accession nos. Q5UQX3, YR380_MIMIV, the first 30 residues are ommited as they do not conform to a LRR structure). A fragment of the sequence of the YOPM protein (accession number P17778, YOPM_YERPE) encompassing a tandem of six contiguous 20-residue LRRs is also shown for comparison. Residue conservation is shown in bold type.

YR380_MIMIV

LRR	*N*	Sequence
I	19	**I**GFQM**L**KS**I**N**I**KR**YP**EFSY
II	22	**L**EK**L**F**I**NN**N**N**L**KQ**LP**DPQYLPK
III	19	**I**KE**L**V**C**SY**N**I**L**TH**IP**F**YP**N
IV	21	**L**IK**L**D**I**SH**N**Q**V**QN**I**NV**Y**NQSK
V	20	**L**LY**L**D**C**SF**N**K**N**IETRI**F**LPE
VI	21	**C**RE**L**Y**V**ND**A**N**I**SK**L**EINYFPN
VII	19	**L**RI**L**D**C**SN**N**N**I**SRI**S**S**L**SS
VIII	19	**L**IE**L**N**I**QN**N**H**I**TE**LP**S**YP**Q
IX	19	**L**VRI**M**A**D**NN**K**L**C**Y**VP**T**FP**N
X	19	**L**LS**L**S**V**SY**N**N**I**VK**I**TDQ**P**L
XI	19	**L**KK**L**V**A**NN**N**S**V**IE**L**GN**LP**K
XII	19	**L**KF**F**D**L**SF**N**K**L**NS**V**TIPSS
XIII	22	**A**KY**I**F**L**QF**N**N**F**VS**V**DIDNCIGC
XIV	21	**V**KE**L**Q**V**DF**N**I**Y**SR**I**YSKYFDN
		IYA**I**N**I**QT**N**RDKLHYYLQQYSQLSNEHI\u

YOPM_YERPE

LRR	*N*	Sequence
IX	20	**L**TT**I**Y**A**DN**N**L**L**KT**LP**D**LPP**S
X	20	**L**EA**L**N**V**RD**N**Y**L**TD**LP**E**LP**QS
XI	20	**L**TF**L**D**V**SE**N**I**F**SG**L**SE**LPP**N
XII	20	**L**YY**L**N**A**SS**N**E**I**RS**L**CD**LPP**S
XIII	20	**L**EE**L**N**V**SN**N**K**L**IE**LP**A**LPP**R
XIV	20	**L**ER**L**I**A**SF**N**H**L**AE**VP**E**LP**QN
XV	20	**L**KQ**L**H**V**EY**N**P**L**RE**FP**D**IP**ES

Associated motifs, cappings and related domains

Several motifs and domains defined in protein sequence databases are associated with or appear often related to the LRR fold (Table 4). Capping structures are the most common of these associated motifs. Extracellular LRR proteins and domains are often flanked at the N and C termini by disulphide-bonded caps that protect the hydrophobic core of the first and last LRRs. Both N-terminal (LRRNT) and C-terminal (LRRCT) capping motifs are described in databases for protein domain identification and analysis such as SMART [95], Pfam [96] and InterPro [97] (Table 4).

In the typical LRRNT capping motif a single β-strand runs antiparallel to the main β-sheet and is followed by a short LRR of 20 or 21 residues (Fig. 11A). The consensus sequence contains four cysteines in a Cx_nCxCx_mC pattern, with n and m being variable numbers. The cysteines form a disulphide knot that connects the antiparallel β-strand to the first LRR. The main structural elements of this motif appear to be maintained irrespective of the number and spacing of cysteines [53]. The LRRCT capping motif contains normally four cysteines that stabilise the local structure with two disulphide bonds. An α helix covers the hydrophobic core of the last LRR (Fig. 11B). This capping motif seems to occur slightly less often than LRRNT (Table 1). Both LRRNT and LRRCT capping structures are common in extracellular and membrane-associated LRR proteins. Several structures have been determined containing both of these caps. Representative examples are GPIBA, LINGO1, NOGO, NTRK1, the different SLIT domains, the TLR different structures and their engineered chimeras, and the VLR structures (Table 1). At the primary sequence level, many cysteine-capped LRR proteins have been automatically annotated as having only one of the two capping structures, although close inspection of their sequences shows that in many cases both capping motifs are present. A variant of LRRNT specific to plants (LRRNP in Table 4) has been defined in the Pfam database. The PGIP2 structure [65] is the only current example for this motif.

In an early sequence analysis of disulphide-bonded C-terminal capping motifs Kajava and colleagues [3, 26] proposed their classification in four different subfamilies named CF1–4. The most common flanking motif, CF1, contains four cysteine residues and is typically detected by LRRCT database descriptors. The second motif, CF2, contains only two cysteines and is specific to SLRP family; the third, CF3, contains three cysteines and is specific to G-protein coupled receptors, and CF4 is a plant specific motif with two cysteines. As seen previously for the LRR subfamilies, protein sequence databases do not include descriptors for the different subfamilies of C-terminal capping motifs and only a general LRRCT motif is described. With the determination of two structures from the SLRP family, DCN and BGN, the structure of the CF2 subtype of C-terminal capping motifs was elucidated. The capping motifs in SLRPs are structurally quite different from LRRCT (Fig. 11C). They show a laterally extended repeat that always precedes the last, C-terminal repeat, which is connected by disulphide bonds [51–53]. This particular lateral extension, which was referred as the "ear", can be identified on sequences of most members of the SLRP family. The ear-containing capping structures have a sequence that is sufficiently different from other LRRCT motifs [53] so as not to be recognised by current automatic annotation tools. A sequence analysis using expression patterns specifically designed to identify sequences containing ear repeats has been carried out in our laboratory; the results of this analysis will be published shortly (Park et al., manuscript in preparation). Since the structure of these capping motifs is very different from the LRRCT one, we have coined the term LRRCE (for LRR C-terminal capping structure containing ear repeats) to designate it.

Two additional capping structures, not based on disulphide bonding, are currently defined in sequence databases (Table 4). The LRRCI motif, specific of the internalins, is in fact a proper domain with an all-β strand topology [35, 98, 99]. Examples of this domain are found in the different internalin structures (INLA, INLB, INLC, and INLH); it is also detected at the sequence level in other members of the internalin family. The LRRCI domain belongs to the family of immunoglobin-like domains with two sandwiched β-sheets; three strands in one of these sheets cap the parallel β-sheet of the LRR domain. LRRCI domains were initially thought to have a mainly structural function, increasing the stability of the internalin LRR domains [99]. However, the recent crystal structure of INLB in complex with a fragment of the human tyrosine kinase receptor Met shows that LRRCI domains can also engage in host cell receptor interactions, providing a secondary interaction interface that in the case of INLB-Met is critical for receptor activation [100]. The other capping motif, LRRCap, occurs at the C terminus of many intracellular LRR domains [101]. It includes the last, irregular LRR, 27–34 residues long, and it protects the hydrophobic core of the LRR domain. Representative structures containing LRRCap motifs are DLC1, GGTA, PP32, TAP and U2A.

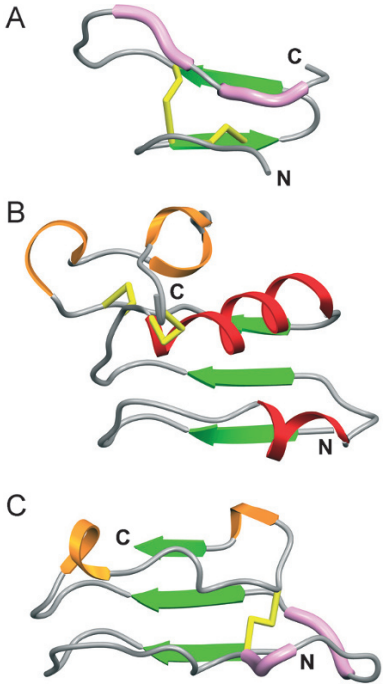

Figure 11. Ribbon diagrams of cysteine-capping motifs in LRR structures, viewed from the convex side of the LRR domains: (*A*) LRRNT capping motif in the DCN structure (PDB i.d. 1XKU); (*B*) LRRCT capping motif in the NOGO structure (PDB i.d. 1OZN); (*C*) LRRCE capping motif in the DCN structure (1XKU). Different secondary structure elements are identified as in Figure 1, with yellow sticks indicating disulphide bonds. The N- and C-terminal ends in each panel are indicated. Figure prepared with SETOR [160].

Two more structural motifs are related to the LRR fold: the leucine-rich repeat variant (LRRV) and the L-domains. The bacteria-specific LRRV structure also shows a repetitive sequence rich in leucines but it is completely different from the LRR domains and proteins that are the subject of this review. Rather than a parallel β-sheet interwoven with different helical topologies, the LRV repetitive structural motif alternates α- and 3_{10} helices arranged in a right-handed superhelix. One crystal structure has been determined with the LRRV fold [102] (PDB i.d. 1LRV). The second related motif is the L domain, a solenoid domain observed in the structures of the extracellular regions of tyrosine-kinase receptors such as the insulin receptor (IR), insulin-like growth factor

receptors (IGFR) and epidermal growth factor receptors (EGFR). The N-terminal regions of these receptor families contain two homologous LRR-like domains named L1 and L2, separated by a cysteine-rich (CR) region that is also specific to these receptors. The three-domain tandem L1-CR-L2 surrounds a central space of sufficient size to accommodate macromolecular ligands. Several structures of these receptors have been determined [103–106], and a recent review summarises the latest findings and current understanding on the structure and function of these receptors [107]. The domains L1 and L2 themselves have a three-dimensional structure remarkably similar to that of LRR domains: they form a right-handed, single-stranded coil, with a parallel β-sheet in one half side of the coil and short 3_{10} helices, β-turns and α-helices on the more irregular opposite half. The resulting slight curvature makes them similar to the bacterial LRR domains represented by the YOPM structure [67]. L domains pack large numbers of leucine residues inside the hydrophobic core arranged in a way clearly reminiscent of that seen in LRR structures. The individual "repeats" building up these L domains have similar lengths to the LRRs (20–30 residues), but their repetitive sequences are less regular and do not show the characteristic pattern of the LRR sequences, *i.e.* the reason why they are not normally classified as members of the LRR superfamily (with the exception of the SCOP database [54]). L domains should not be described as β-solenoids either, as they do not show β-sheet connectivity on the more irregular half of their coil structures [64].

Sequence-based prediction and engineering of LRR structures

The number of determined three-dimensional structures of LRR domains (Table 1) pales in comparison with the large number of sequences of proteins containing domains with LRR architecture. Nevertheless, due to the intrinsic repetitive nature of these LRR domains, it is possible to make useful structural predictions from sequences of proteins containing LRRs (a statement that really applies to most protein architectures with repetitive substructures, from collagen to β-solenoids [64, 108]). These structural predictions coupled with sequence conservation analysis may be of sufficient accuracy to investigate biological function and certainly of value in suggesting mutagenesis experiments, chimeric designs, and other biochemical studies [109–128], or in understanding the effect of disease-causing mutations in LRR-containing proteins [29, 129–133]. In the near future, it will be possible to obtain realistic and accurate three-dimensional representations of naturally occurring LRR proteins or domains on the basis of a sufficient set of known three-dimensional LRR structures. Furthermore, understanding the structural principles of LRR structure will allow designing *de novo* LRR structures with novel, designed sequences. Given the level of confidence that can be achieved in predicting the overall fold of LRR proteins, the LRR scaffold offers excellent possibilities for direct protein engineering (see below).

Kajava, Kobe and collaborators [3, 27, 32, 108, 134] have specifically addressed the structural details of the LRR fold and evaluated the success of modelling procedures by comparing predicted models of LRR domains with the experimentally determined structures. They have identified several levels of structural prediction for an LRR-containing protein sequence, each with increasing complexity. Sequences of regular LRRs can be dissected into the conserved 11-residue motifs located in the concave side of the LRR domain and forming the parallel β-sheet, and the remaining variable regions that connect consecutive repeats in the convex side (Table 5). Thus, the first level of structural prediction is the positioning of residues with respect to the path of the polypeptide main chain. Due to the high residue conservation across regular LRRs it is normally easy to predict which residues are involved in the hydrophobic core of the LRR domain and which are pointing outside, amenable for protein-protein interactions. These predictions start losing validity in irregular LRRs such as those with insertions (extended LRRs) or those at the terminal regions. The next level in complexity is the prediction of the main chain conformation in individual LRRs. Evident trends in the possible conformations of the convex zone and their correlation with number and conservation of residues have been described earlier. As the number of experimentally determined structures increases, the level of accuracy in predicting the conformation of the convex zone of individual LRRs will improve. The final level of complexity is the integration of the individual repeats in the complete curved solenoid structure.

The early modelling strategy adopted by Kajava and co-workers for LRR structures relied on a restrained conformational search to produce models of repeats, where four types of restraints were considered: (a) the assignment of individual residue side chains as internal or external; (b) the general course of the polypeptide main chain accounting for the different twists and turns of the repeat; (c) the preferred conformations observed in other protein structures and the geometry of hydrogen bond formation; and

(d) the ability of local conformations to accept proline residues when these have been observed in at least one LRR structure [134]. The first generation of models obtained using this constrained approach was compared with subsequently determined crystal structures [27]. The models could correctly predict general architecture, curvature, backbone conformation and assignment of internal or external side chains. Models performed better in predicting the backbones of LRRs containing α helices or 3_{10} helices but were less satisfactory for those repeats containing segments of polyproline II helices. Not surprisingly, finer details such as the correct choice of rotamer side chain conformation were less accurate, and in the case of the YOPM structure the models failed to predict correctly the overall molecular twist.

The increase in number of experimentally determined LRR structures will undoubtedly improve the accuracy of newly produced models, at least for the regular LRRs with relatively high conservation of the consensus residues. An approach currently being developed in our laboratory consists in modelling individual repeats based on the most homologous repeat with known three-dimensional structure and then assembling these individual model repeats using overall domain structures as templates. While the backbone conformation of the assembled models can be relatively accurate for regular LRRs, prediction of overall parameters such as domain curvature or twist presumably will need to be improved.

An immediate benefit of the reliability in sequence-based structural prediction of LRR proteins and domains is the ability to engineer modified, chimeric, or even *de novo* structures. A few studies have adopted the comparatively simple approach of generating LRR chimeras by swapping repeats between related sequences or between different species [135–139]. These repeat-swap chimeras (often referred as "domain swaps" in the original publications) were used to ask functional questions such as ligand recognition and binding specificity. Decisions about repeat boundaries were taken on the basis of homology modelling [140], as the corresponding structures were not known at the time of the chimera designs. Adopting a different approach, Plückthun and co-workers [141] designed a protein LRR library based on the mammalian ribonuclease family. Starting from the RNI structure, they produced proteins with 4–14 LRR modules containing randomised amino acids at the concave surface. The clones, expressed in *E. coli*, produced monomeric, stable and highly soluble proteins with the expected α-helical circular dichroism [141]. Interestingly, the same design principles of this artificial library apply to the naturally evolved "library" of variable lymphocyte receptors (VLRs) from hagfish [71]: structural residues are conserved, insertions and deletions in the LRR pattern are not allowed, sequence variability concentrates on the concave surfaces, and the conservation of capping LRRs follows different rules than the rest. It can be concluded that modularity is inherent to the LRR scaffold and that, at least for these two systems, there is a large degree of flexibility in the amino acid composition of the concave surface without affecting the overall stability of the LRR fold [71].

Very recently, an elegant application of the structural prediction of LRR structure has been devised by Lee and co-workers [43, 50] in their structural studies of TLRs dimerisation. They developed hybrids in which truncated fragments of different TLRs were combined with capping motifs LRRNT and LRRCT from hagfish VLR proteins. The strategy, described in detail in [43] as the "hybrid LRR technique", was very successful in producing soluble hybrids that were key to the structural determinations of TLRs. Possible applications of this technique for future studies involving LRR proteins were also discussed in the same paper [43].

Concluding remarks

In this review we have given an account of the current understanding on the structure of LRR proteins and domains at all four structural levels, from amino acid sequence to macromolecular associations. The LRR fold is very versatile, and especially suited for proteins involved in macromolecular interactions. This versatility translates into an incredible variety of biological function. The pioneering elucidation of the crystal structure of the porcine ribonuclease inhibitor has been followed by the determination of almost 90 structures of LRR proteins and domains. From these structures a great deal of knowledge has been obtained, both on the structural principles of this protein architecture and on the way in which these proteins recognise and interact with their ligands. This knowledge has been extremely useful in allowing reliable structural and functional predictions for this family of proteins, and has opened the door to successful engineering attempts that hold great promise for future research. We hope that this review does justice to the work of many colleagues and that it will be useful for new researchers venturing in the multicoloured field of the LRR proteins and domains.

1 Marcotte, E. M., Pellegrini, M., Yeates, T. O. and Eisenberg, D. (1999) A census of protein repeats. J. Mol. Biol. 293, 151–160.
2 Björklund, A. K., Ekman, D. and Elofsson, A. (2006) Expansion of protein domain repeats. PLoS Comput. Biol. 2, e114.
3 Kajava, A. V. (1998) Structural diversity of leucine-rich repeat proteins. J. Mol. Biol. 277, 519–527.
4 Buchanan, S. G. and Gay, N. J. (1996) Structural and functional diversity in the leucine-rich repeat family of proteins. Prog. Biophys. Mol. Biol. 65, 1–44.
5 Kresse, H. and Schönherr, E. (2001) Proteoglycans of the extracellular matrix and growth control. J. Cell Physiol. 189, 266–274.
6 Matilla, A. and Radrizzani, M. (2005) The Anp32 family of proteins containing leucine-rich repeats. Cerebellum 4, 7–18.
7 Hohenester, E., Hussain, S. and Howitt, J. A. (2006) Interaction of the guidance molecule Slit with cellular receptors. Biochem. Soc. Trans. 34, 418–421.
8 Hocking, A. M., Shinomura, T. and McQuillan, D. J. (1998) Leucine-rich repeat glycoproteins of the extracellular matrix. Matrix Biol. 17, 1–19.
9 Andrews, R. K. and Berndt, M. C. (2004) Platelet physiology and thrombosis. Thromb. Res. 114, 447–453.
10 Chen, Y., Aulia, S., Li, L. and Tang, B. L. (2006) AMIGO and friends: An emerging family of brain-enriched, neuronal growth modulating, type I transmembrane proteins with leucine-rich repeats (LRR) and cell adhesion molecule motifs. Brain Res. Rev. 51, 265–274.
11 Ko, J. and Kim, E. (2007) Leucine-rich repeat proteins of synapses. J. Neurosci. Res. 85, 2824–2832.
12 Price, S. R., Evans, P. R. and 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.
13 Liker, E., Fernandez, E., Izaurralde, E. and Conti, E. (2000) The structure of the mRNA export factor TAP reveals a *cis* arrangement of a non-canonical RNP domain and an LRR domain. EMBO J. 19, 5587–5598.
14 Niemann, H. H., Schubert, W. D. and Heinz, D. W. (2004) Adhesins and invasins of pathogenic bacteria: A structural view. Microbes Infect. 6, 101–112.
15 Bierne, H., Sabet, C., Personnic, N. and Cossart, P. (2007) Internalins: A complex family of leucine-rich repeat-containing proteins in *Listeria monocytogenes*. Microbes Infect. 9, 1156–1166.
16 DeYoung, B. J. and Innes, R. W. (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat. Immunol. 7, 1243–1249.
17 McHale, L., Tan, X., Koehl, P. and Michelmore, R. W. (2006) Plant NBS-LRR proteins: Adaptable guards. Genome Biol. 7, 212.
18 Ryan, C. A., Huffaker, A. and Yamaguchi, Y. (2007) New insights into innate immunity in *Arabidopsis*. Cell. Microbiol. 9, 1902–1908.
19 Martinon, F. and Tschopp, J. (2005) NLRs join TLRs as innate sensors of pathogens. Trends Immunol. 26, 447–454.
20 Pancer, Z. and Cooper, M. D. (2006) The evolution of adaptive immunity. Annu. Rev. Immunol. 24, 497–518.
21 West, A. P., Koblansky, A. A. and Ghosh, S. (2006) Recognition and signaling by toll-like receptors. Annu. Rev. Cell Dev. Biol. 22, 409–437.
22 Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L. and Pancer, Z. (2007) Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat. Immunol. 8, 647–656.
23 Gay, N. J. and Gangloff, M. (2007) Structure and function of Toll receptors and their ligands. Annu. Rev. Biochem. 76, 141–165.
24 Kobe, B. and Deisenhofer, J. (1994) The leucine-rich repeat: A versatile binding motif. Trends Biochem. Sci. 19, 415–421.
25 Kobe, B. and Deisenhofer, J. (1995) Proteins with leucine-rich repeats. Curr. Opin. Struct. Biol. 5, 409–416.
26 Kobe, B. and Kajava, A. V. (2001) The leucine-rich repeat as a protein recognition motif. Curr. Opin. Struct. Biol. 11, 725–732.
27 Kajava, A. V. and Kobe, B. (2002) Assessment of the ability to model proteins with leucine-rich repeats in light of the latest structural information. Protein Sci. 11, 1082–1090.
28 Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T. and Matsushima, N. (2004) Structural principles of leucine-rich repeat (LRR) proteins. Proteins 54, 394–403.
29 Matsushima, N., Tachi, N., Kuroki, Y., Enkhbayar, P., Osaki, M., Kamiya, M. and Kretsinger, R. H. (2005) Structural analysis of leucine-rich-repeat variants in proteins associated with human diseases. Cell. Mol. Life Sci. 62, 2771–2791.
30 Kobe, B. and Deisenhofer, J. (1993) Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature 366, 751–756.
31 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242.
32 Kobe, B. and Kajava, A. V. (2000) When protein folding is simplified to protein coiling: The continuum of solenoid protein structures. Trends Biochem. Sci. 25, 509–515.
33 Kobe, B. and Deisenhofer, J. (1995) A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 374, 183–186.
34 Papageorgiou, A. C., Shapiro, R. and Acharya, K. R. (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.
35 Schubert, W. D., Urbanke, C., Ziehm, T., Beier, V., Machner, M. P., Domann, E., Wehland, J., Chakraborty, T. and Heinz, D. W. (2002) Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. Cell 111, 825–836.
36 Huizinga, E. G., Tsuji, S., Romijn, R. A., Schiphorst, M. E., de Groot, P. G., Sixma, J. J. and Gros, P. (2002) Structures of glycoprotein Ibα and its complex with von Willebrand factor A1 domain. Science 297, 1176–1179.
37 Dumas, J. J., Kumar, R., McDonagh, T., Sullivan, F., Stahl, M. L., Somers, W. S. and Mosyak, L. (2004) Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Ibα complex reveals conformation differences with a complex bearing von Willebrand disease mutations. J. Biol. Chem. 279, 23327–23334.
38 Fan, Q. R. and Hendrickson, W. A. (2005) Structure of human follicle-stimulating hormone in complex with its receptor. Nature 433, 269–277.
39 Fukuda, K., Doggett, T., Laurenzi, I. J., Liddington, R. C. and Diacovo, T. G. (2005) The snake venom protein botrocetin acts as a biological brace to promote dysfunctional platelet aggregation. Nat. Struct. Mol. Biol. 12, 152–159.
40 Hao, B., Zheng, N., Schulman, B. A., Wu, G., Miller, J. J., Pagano, M. and Pavletich, N. P. (2005) Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. Mol. Cell 20, 9–19.
41 Iyer, S., Holloway, D. E., Kumar, K., Shapiro, R. and Acharya, K. R. (2005) Molecular recognition of human eosinophil-derived neurotoxin (RNase 2) by placental ribonuclease inhibitor. J. Mol. Biol. 347, 637–655.
42 Johnson, R. J., McCoy, J. G., Bingman, C. A., Phillips, G. N. Jr. and Raines, R. T. (2007) Inhibition of human pancreatic ribonuclease by the human ribonuclease inhibitor protein. J. Mol. Biol. 368, 434–449.
43 Kim, H. M., Park, B. S., Kim, J. I., Kim, S. E., Lee, J., Oh, S. C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O. J. and Lee, J. O. (2007) Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell 130, 906–917.
44 Morlot, C., Thielens, N. M., Ravelli, R. B., Hemrika, W., Romijn, R. A., Gros, P., Cusack, S. and McCarthy, A. A.

(2007) Structural insights into the Slit-Robo complex. Proc. Natl. Acad. Sci. USA 104, 14923–14928.
45 Wollert, T., Pasche, B., Rochon, M., Deppenmeier, S., van den Heuvel, J., Gruber, A. D., Heinz, D. W., Lengeling, A. and Schubert, W. D. (2007) Extending the host range of *Listeria monocytogenes* by rational protein design. Cell 129, 891–902.
46 Seewald, M. J., Korner, C., Wittinghofer, A. and Vetter, I. R. (2002) RanGAP mediates GTP hydrolysis without an arginine finger. Nature 415, 662–666.
47 Celikel, R., McClintock, R. A., Roberts, J. R., Mendolicchio, G. L., Ware, J., Varughese, K. I. and Ruggeri, Z. M. (2003) Modulation of α-thrombin function by distinct interactions with platelet glycoprotein Ibα. Science 301, 218–221.
48 Dumas, J. J., Kumar, R., Seehra, J., Somers, W. S. and Mosyak, L. (2003) Crystal structure of the GpIbα-thrombin complex essential for platelet aggregation. Science 301, 222–226.
49 Kim, J. I., Lee, C. J., Jin, M. S., Lee, C. H., Paik, S. G., Lee, H. and Lee, J. O. (2005) Crystal structure of CD14 and its implications for lipopolysaccharide signaling. J. Biol. Chem. 280, 11347–11351.
50 Jin, M. S., Kim, S. E., Heo, J. Y., Lee, M. E., Kim, H. M., Paik, S. G., Lee, H. and Lee, J. O. (2007) Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell 130, 1071–1082.
51 Scott, P. G., McEwan, P. A., Dodd, C. M., Bergmann, E. M., Bishop, P. N. and Bella, J. (2004) Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. Proc. Natl. Acad. Sci. USA 101, 15633–15638.
52 Scott, P. G., Dodd, C. M., Bergmann, E. M., Sheehan, J. K. and Bishop, P. N. (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.
53 McEwan, P. A., Scott, P. G., Bishop, P. N. and Bella, J. (2006) Structural correlations in the family of small leucine-rich repeat proteins and proteoglycans. J. Struct. Biol. 155, 294–305.
54 Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.
55 Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thornton, J. M. (1997) CATH – A hierarchic classification of protein domain structures. Structure 5, 1093–1108.
56 Uff, S., Clemetson, J. M., Harrison, T., Clemetson, K. J. and Emsley, J. (2002) Crystal structure of the platelet glycoprotein Ibα N-terminal domain reveals an unmasking mechanism for receptor activation. J. Biol. Chem. 277, 35657–35663.
57 Bell, J. K., Botos, I., Hall, P. R., Askins, J., Shiloach, J., Segal, D. M. and Davies, D. R. (2005) The molecular structure of the toll-like receptor 3 ligand-binding domain. Proc. Natl. Acad. Sci. USA 102, 10976–10980.
58 Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., Park, S. Y. and Tame, J. R. (2005) Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. J. Biol. Chem. 280, 17339–17345.
59 Jong, W. S., ten Hagen-Jongman, C. M., den Blaauwen, T., Slotboom, D. J., Tame, J. R., Wickström, D., de Gier, J. W., Otto, B. R. and Luirink, J. (2007) Limited tolerance towards folded elements during secretion of the autotransporter Hbp. Mol. Microbiol. 63, 1524–1536.
60 Jenkins, J. and Pickersgill, R. (2001) The architecture of parallel β-helices and related folds. Prog. Biophys. Mol. Biol. 77, 111–175.
61 Pickersgill, R., Jenkins, J., Harris, G., Nasser, W. and Robert-Baudouy, J. (1994) The structure of *Bacillus subtilis* pectate lyase in complex with calcium. Nat. Struct. Biol. 1, 717–723.
62 Jenkins, J., Mayans, O. and Pickersgill, R. (1998) Structure and evolution of parallel β-helix proteins. J. Struct. Biol. 122, 236–246.
63 Hennetin, J., Jullian, B., Steven, A. C. and Kajava, A. V. (2006) Standard conformations of β-arches in β-solenoid proteins. J. Mol. Biol. 358, 1094–1105.
64 Kajava, A. V. and Steven, A. C. (2006) β-rolls, β-helices, and other β-solenoid proteins. Adv. Protein. Chem. 73, 55–96.
65 Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K. A., Savino, C., De Lorenzo, G., Tsernoglou, D. and Cervone, F. (2003) The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. Proc. Natl. Acad. Sci. USA 100, 10124–10128.
66 Marino, M., Braun, L., Cossart, P. and Ghosh, P. (1999) Structure of the lnlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. Mol. Cell 4, 1063–1072.
67 Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. and Waugh, D. S. (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.
68 Bella, J., Brodsky, B. and Berman, H. M. (1995) Hydration structure of a collagen peptide. Structure 3, 893–906.
69 He, X. L., Bazan, J. F., McDermott, G., Park, J. B., Wang, K., Tessier-Lavigne, M., He, Z. and Garcia, K. C. (2003) Structure of the Nogo receptor ectodomain: A recognition module implicated in myelin inhibition. Neuron 38, 177–185.
70 Barton, W. A., Liu, B. P., Tzvetkova, D., Jeffrey, P. D., Fournier, A. E., Sah, D., Cate, R., Strittmatter, S. M. and Nikolov, D. B. (2003) Structure and axon outgrowth inhibitor binding of the Nogo-66 receptor and related proteins. EMBO J. 22, 3291–3302.
71 Kim, H. M., Oh, S. C., Lim, K. J., Kasamatsu, J., Heo, J. Y., Park, B. S., Lee, H., Yoo, O. J., Kasahara, M. and Lee, J. O. (2007) Structural diversity of the hagfish variable lymphocyte receptors. J. Biol. Chem. 282, 6726–6732.
72 Wu, H., Maciejewski, M. W., Marintchev, A., Benashski, S. E., Mullen, G. P. and King, S. M. (2000) Solution structure of a dynein motor domain associated light chain. Nat. Struct. Biol. 7, 575–579.
73 Wu, H., Blackledge, M., Maciejewski, M. W., Mullen, G. P. and King, S. M. (2003) Relaxation-based structure refinement and backbone molecular dynamics of the dynein motor domain-associated light chain. Biochemistry 42, 57–71.
74 Tan, X., Calderon-Villalobos, L. I., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M. and Zheng, N. (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446, 640–645.
75 Lee, F. S., Shapiro, R. and Vallee, B. L. (1989) Tight-binding inhibition of angiogenin and ribonuclease A by placental ribonuclease inhibitor. Biochemistry 28, 225–230.
76 Wollert, T., Heinz, D. W. and Schubert, W. D. (2007) Thermodynamically reengineering the listerial invasion complex InlA/E-cadherin. Proc. Natl. Acad. Sci. USA 104, 13960–13965.
77 Shaul, Y. and Schreiber, G. (2005) Exploring the charge space of protein-protein association: A proteomic study. Proteins 60, 341–352.
78 Dias, J. A. (2005) Endocrinology: Fertility hormone in repose. Nature 433, 203–204.
79 Sadler, J. E. (2003) Structural biology. A ménage à trois in two configurations. Science 301, 177–179.
80 Vanhoorelbeke, K., Ulrichts, H., Romijn, R. A., Huizinga, E. G. and Deckmyn, H. (2004) The GPIbα-thrombin interaction: Far from crystal clear. Trends Mol. Med. 10, 33–39.
81 Scott, P. G., Grossmann, J. G., Dodd, C. M., Sheehan, J. K. and Bishop, P. N. (2003) Light and X-ray scattering show decorin to be a dimer in solution. J. Biol. Chem. 278, 18353–18359.
82 Le Goff, M. M., Hindson, V. J., Jowitt, T. A., Scott, P. G. and Bishop, P. N. (2003) Characterization of opticin and evidence

of stable dimerization in solution. J. Biol. Chem. 278, 45280–45287.
83 Howitt, J. A., Clout, N. J. and Hohenester, E. (2004) Binding site for Robo receptors revealed by dissection of the leucine-rich repeat region of Slit. EMBO J. 23, 4406–4412.
84 Choe, J., Kelker, M. S. and Wilson, I. A. (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science 309, 581–585.
85 Gay, N. J., Gangloff, M. and Weber, A. N. (2006) Toll-like receptors as molecular switches. Nat. Rev. Immunol. 6, 693–698.
86 Bell, J. K., Mullen, G. E., Leifer, C. A., Mazzoni, A., Davies, D. R. and Segal, D. M. (2003) Leucine-rich repeats and pathogen recognition in toll-like receptors. Trends Immunol. 24, 528–533.
87 Brodsky, I. and Medzhitov, R. (2007) Two modes of ligand recognition by TLRs. Cell 130, 979–981.
88 Mosyak, L., Wood, A., Dwyer, B., Buddha, M., Johnson, M., Aulabaugh, A., Zhong, X., Presman, E., Benard, S., Kelleher, K., Wilhelm, J., Stahl, M. L., Kriz, R., Gao, Y., Cao, Z., Ling, H. P., Pangalos, M. N., Walsh, F. S. and Somers, W. S. (2006) The structure of the Lingo-1 ectodomain, a module implicated in central nervous system repair inhibition. J. Biol. Chem. 281, 36378–36390.
89 Lichtarge, O., Bourne, H. R. and Cohen, F. E. (1996) An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. 257, 342–358.
90 Lichtarge, O., Yao, H., Kristensen, D. M., Madabushi, S. and Mihalek, I. (2003) Accurate and scalable identification of functional sites by evolutionary tracing. J. Struct. Funct. Genomics 4, 159–166.
91 Armon, A., Graur, D. and Ben-Tal, N. (2001) ConSurf: An algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J. Mol. Biol. 307, 447–463.
92 Chelliah, V., Chen, L., Blundell, T. L. and Lovell, S. C. (2004) Distinguishing structural and functional restraints in evolution in order to identify interaction sites. J. Mol. Biol. 342, 1487–1504.
93 Neuvirth, H., Raz, R. and Schreiber, G. (2004) ProMate: A structure based prediction program to identify the location of protein-protein binding sites. J. Mol. Biol. 338, 181–199.
94 Bradford, J. R. and Westhead, D. R. (2005) Improved prediction of protein-protein binding sites using a support vector machines approach. Bioinformatics 21, 1487–1494.
95 Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J. and Bork, P. (2006) SMART 5: Domains in the context of genomes and networks. Nucleic Acids Res. 34, D257–D260.
96 Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. and Bateman, A. (2006) Pfam: Clans, web tools and services. Nucleic Acids Res. 34, D247–D251.
97 Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silventoinen, V., Studholme, D. J., Vaughan, R. and Wu, C. H. (2005) InterPro, progress and status in 2005. Nucleic Acids Res. 33, D201–D205.
98 Schubert, W. D., Göbel, G., Diepholz, M., Darji, A., Kloer, D., Hain, T., Chakraborty, T., Wehland, J., Domann, E. and Heinz, D. W. (2001) Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain. J. Mol. Biol. 312, 783–794.
99 Freiberg, A., Machner, M. P., Pfeil, W., Schubert, W. D., Heinz, D. W. and Seckler, R. (2004) Folding and stability of the leucine-rich repeat domain of internalin B from *Listeria monocytogenes*. J. Mol. Biol. 337, 453–461.
100 Niemann, H. H., Jäger, V., Butler, P. J., van den Heuvel, J., Schmidt, S., Ferraris, D., Gherardi, E. and Heinz, D. W. (2007) Structure of the human receptor tyrosine kinase Met in complex with the *Listeria* invasion protein InlB. Cell 130, 235–246.
101 Ceulemans, H., De Maeyer, M., Stalmans, W. and Bollen, M. (1999) A capping domain for LRR protein interaction modules. FEBS Lett. 456, 349–351.
102 Peters, J. W., Stowell, M. H. and Rees, D. C. (1996) A leucine-rich repeat variant with a novel repetitive protein structural motif. Nat. Struct. Biol. 3, 991–994.
103 Garrett, T. P., McKern, N. M., Lou, M., Frenkel, M. J., Bentley, J. D., Lovrecz, G. O., Elleman, T. C., Cosgrove, L. J. and Ward, C. W. (1998) Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. Nature 394, 395–399.
104 Cho, H. S. and Leahy, D. J. (2002) Structure of the extracellular region of HER3 reveals an interdomain tether. Science 297, 1330–1333.
105 Garrett, T. P., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H. J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W. and Ward, C. W. (2002) Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. Cell 110, 763–773.
106 Cho, H. S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W. Jr. and Leahy, D. J. (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature 421, 756–760.
107 Ward, C. W., Lawrence, M. C., Streltsov, V. A., Adams, T. E. and McKern, N. M. (2007) The insulin and EGF receptor structures: New insights into ligand-induced receptor activation. Trends Biochem. Sci. 32, 129–137.
108 Kajava, A. V. (2001) Review: Proteins with repeated sequence – Structural prediction and modeling. J. Struct. Biol. 134, 132–144.
109 Jiang, X., Dreano, M., Buckler, D. R., Cheng, S., Ythier, A., Wu, H., Hendrickson, W. A. and el Tayar, N. (1995) Structural predictions for the ligand-binding region of glycoprotein hormone receptors and the nature of hormone-receptor interactions. Structure 3, 1341–1353.
110 Bhowmick, N., Huang, J., Puett, D., Isaacs, N. W. and Lapthorn, A. J. (1996) Determination of residues important in hormone binding to the extracellular domain of the luteinizing hormone/chorionic gonadotropin receptor by site-directed mutagenesis and modeling. Mol. Endocrinol. 10, 1147–1159.
111 el Tayar, N. (1996) Advances in the molecular understanding of gonadotropins-receptors interactions. Mol. Cell Endocrinol. 125, 65–70.
112 Janosi, J. B., Ramsland, P. A., Mott, M. R., Firth, S. M., Baxter, R. C. and Delhanty, P. J. (1999) The acid-labile subunit of the serum insulin-like growth factor-binding protein complexes. Structural determination by molecular modeling and electron microscopy. J. Biol. Chem. 274, 23328–23332.
113 Hines, J., Skrzypek, E., Kajava, A. V. and Straley, S. C. (2001) Structure-function analysis of *Yersinia pestis* YopM's interaction with α-thrombin to rule on its significance in systemic plague and to model YopM's mechanism of binding host proteins. Microb. Pathog. 30, 193–209.
114 Song, Y. S., Ji, I., Beauchamp, J., Isaacs, N. W. and Ji, T. H. (2001) Hormone interactions to Leu-rich repeats in the gonadotropin receptors. I. Analysis of Leu-rich repeats of human luteinizing hormone/chorionic gonadotropin receptor and follicle-stimulating hormone receptor. J. Biol. Chem. 276, 3426–3435.
115 Ceulemans, H., Vulsteke, V., De Maeyer, M., Tatchell, K., Stalmans, W. and Bollen, M. (2002) Binding of the concave

surface of the Sds22 superhelix to the α4/α5/α6-triangle of protein phosphatase-1. J. Biol. Chem. 277, 47331–47337.
116 Harton, J. A., O'Connor, W., Jr., Conti, B. J., Linhoff, M. W. and Ting, J. P. (2002) Leucine-rich repeats of the class II transactivator control its rate of nuclear accumulation. Hum. Immunol. 63, 588–601.
117 Dievart, A. and Clark, S. E. (2003) Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. Curr. Opin. Plant Biol. 6, 507–516.
118 Legouis, R., Jaulin-Bastard, F., Schott, S., Navarro, C., Borg, J. P. and Labouesse, M. (2003) Basolateral targeting by leucine-rich repeat domains in epithelial cells. EMBO Rep. 4, 1096–1102.
119 Smits, G., Campillo, M., Govaerts, C., Janssens, V., Richter, C., Vassart, G., Pardo, L. and Costagliola, S. (2003) Glycoprotein hormone receptors: Determinants in leucine-rich repeats responsible for ligand specificity. EMBO J. 22, 2692–2703.
120 Vourc'h, P., Moreau, T., Arbion, F., Marouillat-Védrine, S., Müh, J. P. and Andres, C. (2003) Oligodendrocyte myelin glycoprotein growth inhibition function requires its conserved leucine-rich repeat domain, not its glycosylphosphatidylinositol anchor. J. Neurochem. 85, 889–897.
121 Halterman, D. A. and Wise, R. P. (2004) A single-amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signaling. Plant J. 38, 215–226.
122 Tanabe, T., Chamaillard, M., Ogura, Y., Zhu, L., Qiu, S., Masumoto, J., Ghosh, P., Moran, A., Predergast, M. M., Tromp, G., Williams, C. J., Inohara, N. and Núñez, G. (2004) Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. EMBO J. 23, 1587–1597.
123 van der Hoorn, R. A., Wulff, B. B., Rivas, S., Durrant, M. C., van der Ploeg, A., de Wit, P. J. and Jones, J. D. (2005) Structure-function analysis of Cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. Plant Cell 17, 1000–1015.
124 Kolade, O. O., Bamford, V. A., Ancillo Anton, G., Jones, J. D., Vera, P. and Hemmings, A. M. (2006) *In vitro* characterization of the cysteine-rich capping domains in a plant leucine rich repeat protein. Biochim. Biophys. Acta 1764, 1043–1053.
125 Gadkari, R. A., Sandhya, S., Sowdhamini, R. and Dighe, R. R. (2007) The antigen binding sites of various hCG monoclonal antibodies show homology to different domains of LH receptor. Mol. Cell Endocrinol. 260–262, 23–32.
126 Puett, D., Li, Y., DeMars, G., Angelova, K. and Fanelli, F. (2007) A functional transmembrane complex: The luteinizing hormone receptor with bound ligand and G protein. Mol. Cell. Endocrinol. 260–262, 126–136.
127 Scott, D. J., Wilkinson, T. N., Zhang, S., Ferraro, T., Wade, J. D., Tregear, G. W. and Bathgate, R. A. (2007) Defining the LGR8 residues involved in binding insulin-like peptide 3. Mol. Endocrinol. 21, 1699–1712.
128 Stange, C., Matus, J. T., Dominguez, C., Perez-Acle, T. and Arce-Johnson, P. (2008) The N-homologue LRR domain adopts a folding which explains the TMV-Cg-induced HR-like response in sensitive tobacco plants. J. Mol. Graph. Model 26, 850–860.
129 Pusch, C. M., Zeitz, C., Brandau, O., Pesch, K., Achatz, H., Feil, S., Scharfe, C., Maurer, J., Jacobi, F. K., Pinckers, A., Andreasson, S., Hardcastle, A., Wissinger, B., Berger, W. and Meindl, A. (2000) The complete form of X-linked congenital stationary night blindness is caused by mutations in a gene encoding a leucine-rich repeat protein. Nat. Genet. 26, 324–327.
130 Albrecht, M., Domingues, F. S., Schreiber, S. and Lengauer, T. (2003) Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2. FEBS Lett. 554, 520–528.
131 Montanelli, L., Van Durme, J. J., Smits, G., Bonomi, M., Rodien, P., Devor, E. J., Moffat-Wilson, K., Pardo, L., Vassart, G. and Costagliola, S. (2004) Modulation of ligand selectivity associated with activation of the transmembrane region of the human follitropin receptor. Mol. Endocrinol. 18, 2061–2073.
132 Bogatcheva, N. V., Ferlin, A., Feng, S., Truong, A., Gianesello, L., Foresta, C. and Agoulnik, A. I. (2007) T222P mutation of the insulin-like 3 hormone receptor LGR8 is associated with testicular maldescent and hinders receptor expression on the cell surface membrane. Am. J. Physiol. Endocrinol. Metab. 292, E138–E144.
133 Majava, M., Bishop, P. N., Hagg, P., Scott, P. G., Rice, A., Inglehearn, C., Hammond, C. J., Spector, T. D., Ala-Kokko, L. and Mannikko, M. (2007) Novel mutations in the small leucine-rich repeat protein/proteoglycan (SLRP) genes in high myopia. Hum. Mutat. 28, 336–344.
134 Kajava, A. V., Vassart, G. and Wodak, S. J. (1995) Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats. Structure 3, 867–877.
135 Svensson, L., Heinegård, D. and Oldberg, A. (1995) Decorin-binding sites for collagen type I are mainly located in leucine-rich repeats 4–5. J. Biol. Chem. 270, 20712–20716.
136 Shen, Y., Romo, G. M., Dong, J. F., Schade, A., McIntire, L. V., Kenny, D., Whisstock, J. C., Berndt, M. C., López, J. A. and Andrews, R. K. (2000) Requirement of leucine-rich repeats of glycoprotein (GP) Ibα for shear-dependent and static binding of von Willebrand factor to the platelet membrane GP Ib-IX-V complex. Blood 95, 903–910.
137 Shen, Y., Dong, J. F., Romo, G. M., Arceneaux, W., Aprico, A., Gardiner, E. E., López, J. A., Berndt, M. C. and Andrews, R. K. (2002) Functional analysis of the C-terminal flanking sequence of platelet glycoprotein Ibα using canine-human chimeras. Blood 99, 145–150.
138 Arya, M., López, J. A., Romo, G. M., Dong, J. F., McIntire, L. V., Moake, J. L. and Anvari, B. (2002) Measurement of the binding forces between von Willebrand factor and variants of platelet glycoprotein Ibα using optical tweezers. Lasers Surg. Med. 30, 306–312.
139 Shen, Y., Cranmer, S. L., Aprico, A., Whisstock, J. C., Jackson, S. P., Berndt, M. C. and Andrews, R. K. (2006) Leucine-rich repeats 2–4 (Leu^{60}-Glu^{128}) of platelet glycoprotein Ibα regulate shear-dependent cell adhesion to von Willebrand factor. J. Biol. Chem. 281, 26419–26423.
140 Whisstock, J. C., Shen, Y., López, J. A., Andrews, R. K. and Berndt, M. C. (2002) Molecular modeling of the seven tandem leucine-rich repeats within the ligand-binding region of platelet glycoprotein Ibα. Thromb. Haemost. 87, 329–333.
141 Stumpp, M. T., Forrer, P., Binz, H. K. and Plückthun, A. (2003) Designing repeat proteins: Modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. J. Mol. Biol. 332, 471–487.
142 Zhang, H., Seabra, M. C. and Deisenhofer, J. (2000) Crystal structure of Rab geranylgeranyltransferase at 2.0 Å resolution. Structure 8, 241–251.
143 Pylypenko, O., Rak, A., Reents, R., Niculae, A., Sidorovitch, V., Cioaca, M. D., Bessolitsyna, E., Thoma, N. H., Waldmann, H., Schlichting, I., Goody, R. S. and Alexandrov, K. (2003) Structure of Rab escort protein-1 in complex with Rab geranylgeranyltransferase. Mol. Cell 11, 483–494.
144 Varughese, K. I., Ruggeri, Z. M. and Celikel, R. (2004) Platinum-induced space-group transformation in crystals of the platelet glycoprotein Ibα N-terminal domain. Acta Crystallogr. D Biol. Crystallogr. 60, 405–411.
145 Marino, M., Banerjee, M., Jonquières, R., Cossart, P. and 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.
146 Marino, M., Banerjee, M., Copp, J., Dramsi, S., Chapman, T., van der Geer, P., Cossart, P. and Ghosh, P. (2004) Characterization of the calcium-binding sites of *Listeria monocytogenes* InlB. Biochem. Biophys. Res. Commun. 316, 379–386.
147 Ooi, A., Hussain, S., Seyedarabi, A. and Pickersgill, R. W. (2006) Structure of internalin C from *Listeria monocytogenes*. Acta Crystallogr. D Biol. Crystallogr. 62, 1287–1293.

148 Wehrman, T., He, X., Raab, B., Dukipatti, A., Blau, H. and Garcia, K. C. (2007) Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. Neuron 53, 25–38.
149 Huyton, T. and 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.
150 Hillig, R. C., Renault, L., Vetter, I. R., Drell, T. 4th, Wittinghofer, A. and Becker, J. (1999) The crystal structure of rna1p: A new fold for a GTPase-activating protein. Mol. Cell 3, 781–791.
151 Kobe, B. and 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.
152 Levin, E. J., Kondrashov, D. A., Wesenberg, G. E. and Phillips, G. N. Jr. (2007) Ensemble refinement of protein crystal structures: Validation and application. Structure 15, 1040–1052.
153 Schulman, B. A., Carrano, A. C., Jeffrey, P. D., Bowen, Z., Kinnucan, E. R., Finnin, M. S., Elledge, S. J., Harper, J. W., Pagano, M. and Pavletich, N. P. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature 408, 381–386.
154 Morlot, C., Hemrika, W., Romijn, R. A., Gros, P., Cusack, S. and McCarthy, A. A. (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.
155 Ho, D. N., Coburn, G. A., Kang, Y., Cullen, B. R. and Georgiadis, M. M. (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. USA 99, 1888–1893.
156 Krieger, I., Kostyukova, A., Yamashita, A., Nitanai, Y. and Maeda, 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.
157 Lu, S., Symersky, J., Li, S., Carson, M., Chen, L., Meehan, E. and Luo, M. (2004) Structural genomics of *Caenorhabditis elegans:* Crystal structure of the tropomodulin C-terminal domain. Proteins 56, 384–386.
158 Krissinel, E. and Henrick, K. (2005) Detection of protein assemblies in crystals. In: Computational Life Sciences, vol. 3695, pp. 163–174, Berthold, M. R., Glen, R., Diederichs, K., Kohlbacher, O., Fischer, I. (eds.), Springer, Berlin.
159 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.
160 Evans, S. V. (1993) SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules. J. Mol. Graph. 11, 134–138.
161 DeLano, W. L. (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto.