BMC Genomics

BioMed Central

Research article

Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors

Norio Matsushima*[1], Takanori Tanaka[2], Purevjav Enkhbayar[3], Tomoko Mikami[1,4], Masae Taga[1,4], Keiko Yamada[1] and Yoshio Kuroki[5]

Address: [1]School of Health Sciences, Sapporo Medical University, Hokkaido 060-8556, Japan, [2]RIKEN Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan, [3]Faculty of Biology, National University of Mongolia, Ulaanbaatar-210646/377, Mongolia, [4]Department of Nursing, Sapporo City University, Sapporo, Hokkaido 060-0011, Japan and [5]Department of Biochemistry, School of Medicine, Sapporo Medical University, Hokkaido 060-8556, Japan

Email: Norio Matsushima* - matusima@sapmed.ac.jp; Takanori Tanaka - gouhan@gsc.riken.jp; Purevjav Enkhbayar - enkhbayar@biology.num.edu.mn; Tomoko Mikami - t.mikami@scu.ac.jp; Masae Taga - m.taga@scu.ac.jp; Keiko Yamada - oyama@sapmed.ac.jp; Yoshio Kuroki - kurokiy@sapmed.ac.jp

* Corresponding author

Published: 21 May 2007

BMC Genomics 2007, **8**:124 doi:10.1186/1471-2164-8-124

Received: 24 October 2006
Accepted: 21 May 2007

This article is available from: http://www.biomedcentral.com/1471-2164/8/124

Abstract

Background: Toll-like receptors (TLRs) play a central role in innate immunity. TLRs are membrane glycoproteins and contain leucine rich repeat (LRR) motif in the ectodomain. TLRs recognize and respond to molecules such as lipopolysaccharide, peptidoglycan, flagellin, and RNA from bacteria or viruses. The LRR domains in TLRs have been inferred to be responsible for molecular recognition. All LRRs include the highly conserved segment, LxxLxLxxNxL, in which "L" is Leu, Ile, Val, or Phe and "N" is Asn, Thr, Ser, or Cys and "x" is any amino acid. There are seven classes of LRRs including "typical" ("***T***") and "bacterial" ("***S***"). All known domain structures adopt an arc or horseshoe shape. Vertebrate TLRs form six major families. The repeat numbers of LRRs and their "phasing" in TLRs differ with isoforms and species; they are aligned differently in various databases. We identified and aligned LRRs in TLRs by a new method described here.

Results: The new method utilizes known LRR structures to recognize and align new LRR motifs in TLRs and incorporates multiple sequence alignments and secondary structure predictions. TLRs from thirty-four vertebrate were analyzed. The repeat numbers of the LRRs ranges from 16 to 28. The LRRs found in TLRs frequently consists of LxxLxLxxNxLxxLxxxxF/LxxLxx ("***T***") and sometimes short motifs including LxxLxLxxNxLxxLPx(x)LPxx ("**S**"). The *TLR7* family (TLR7, TLR8, and TLR9) contain 27 LRRs. The LRRs at the N-terminal part have a super-motif of ***STT*** with about 80 residues. The super-repeat is represented by ***STTSTTSTT*** or ***_TTSTTSTT***. The LRRs in TLRs form one or two horseshoe domains and are mostly flanked by two cysteine clusters including two or four cysteine residue.

Conclusion: Each of the six major TLR families is characterized by their constituent LRR motifs, their repeat numbers, and their patterns of cysteine clusters. The central parts of the *TLR1* and *TLR7* families and of TLR4 have more irregular or longer LRR motifs. These central parts are inferred to play a key role in the structure and/or function of their TLRs. Furthermore, the super-repeat in the *TLR7* family suggests strongly that "bacterial" and "typical" LRRs evolved from a common precursor.

Background

Toll-like receptors (TLRs) play a central role in innate immunity [1-3]. TLRs are type I integral membrane glycoproteins consisting of leucine rich repeat (LRR) motif in the ectodomain (ECD), and cytoplamic signaling domains known as Toll IL-receptor (TIR) domains, joined by a single trans membrane helix (Figure 1). They recognize and respond to a variety of components derived from pathogenic or commensal microorganisms principally bacteria and viruses. These molecules include lipids such as lipopolysaccharide (LPS) from Gram-negative bacteria and peptidoglycan fragments from bacterial cell walls, proteins such as flagellin and nucleic acids such as single-stranded and double-stranded RNA and unmethylated CpG DNA from bacteria or viruses. The ECDs including LRRs have been inferred to recognize directly various ligands. The TLR family counts 10 members in human and 12 in mice and Takifugu rubripes. Six major families of vertebrate TLRs have been proposed in a molecular dendrogram [4].

Leucine-rich repeat (LRR)-containing domains are present in over 6000 proteins listed in PFAM, PRINTS, SMART, and InterPro data bases [5-8]. All LRR repeats can be divided into a highly conserved segment (HCS) and a variable segment (VS). The HCS part consists of an eleven residue stretch, LxxLxLxxNxL, or a tweleve residue stretch, LxxLxLxxCxxL, in which "L" is Leu, Ile, Val, or Phe, "N" is Asn, Thr, Ser, or Cys, and "C" is Cys, Ser or Asn [7,9]. Seven classes of LRRs have been proposed, characterized by different lengths and consensus sequences of the VS part of repeats [9,10]. They are "RI-like", "CC", "bacterial", "SDS22-like", "plant specific", "typical", and "TpLRR". Each subfamily of small leucine-rich repeat proteoglycan (SLRP) has LRRs from more than one of the seven classes [8,11]. The structures of twenty-two different proteins that contain LRRs are available [12-51]. They include TLR3 and CD14 [48-50]. The LRR domains in all known structures adopt an arc shape. Most of the known LRR structures have a cap, which shields the hydrophobic core of the first LRR at the N-terminus or the last LRR at the C-terminus. In extracellular proteins or extracellular domains, these caps frequently consist of cysteine clusters including two or four cysteine residues [8,9].

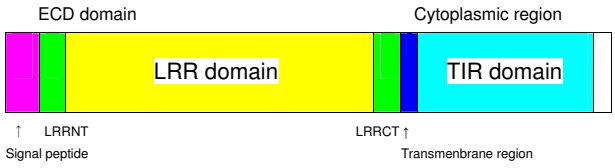

Figure 1
Structural organization of vertebrate TLRs. Mangenta is signal peptide sequence. Green is LRRNT (the cysteine clusters on the N-terminal side of LRRs) and LRRCT (the cysteine clusters on the C-terminal side of LRRs). Yellow is LRR domain. Blue is transmembrane region. Light blue is TIR domain.

The indicated repeat number of LRRs and its "phasing" (that is, what segment or residue corresponds to the beginning of a repeating unit) in individual TLRs are different among the databases (or researchers) and species. This difference reflects the irregularity of LRR motifs in TLRs. Over one hundred complete TLRs are available. Several methods of protein secondary structure predictions such as Proteus and SSPro4.0 show a correspondence of about 75% [52-54]. For the identification of LRRs we propose a new method, which uses the known structures of several LRRs, multiple sequence alignments and secondary structure predictions of TLRs. This new method indicates that each of the six recognized TLR families can be characterized by its LRR motifs, their repeat number and the motifs of two cysteine clusters flanking the LRRs. The actual repeat number of LRRs is generally larger than those reported in the databases. The present analysis leads to the hypothesis that all the LRRs in TLRs form one or two horseshoe domains.

Results

A new method for the identification of LRRs within TLRs

LRR known structures

All of the LRR domains in one protein form a single continuous structure and adopt an arc or horseshoe shape. On the inner, concave face there is a stack of parallel β-strands and on the outer, convex face there are a variety of secondary structures such as α-helix, 3_{10}-helix, polyproline II helix, or a tandem arrangement of β-turns [8,55]. The HCS part of all the LRRs consists of LxxLxLxxNxL or LxxLxLxxCxxL,, as noted, in which "L" is Leu, Ile, Val, or Phe, "N" is Asn, Thr, Ser, or Cys, and "C" is Cys, Ser or Asn [7,9]. The short β-strands are mostly formed by three residues at positions 3 through 5 in the HCS part. In most LRR proteins the β-strands on the concave face and (mostly) helical elements on the convex face are connected by short loops or β-turns. Four leucine residues at positions 1, 4, 6 and 11 participate in the hydrophobic core in LRR arcs. Similarly, conserved hydrophobic residues in the VS parts of the seven LRR classes participate in the hydrophobic core. The side chains of asparagine at position 9 form hydrogen bonds in the loop structure [6].

Structural alignments of the known LRR structures reveal that the LRR motif is surprisingly variable (Table 1). The lengths of LRRs range from 20 to 43 residues. Leucines at positions 1, 4, 6 and 11 of the HCS part are sometimes replaced by Met, Ala, or Cys, as seen in TLR3 [49,50], Internalin A (Inl-A) [26], and Internalin B (Inl-B) [22-24]. Leucines at positions 1 and 11 are also occupied by relatively hydrophilic residues such as Gly, Thr, Asn and Tyr.

Furthermore, asparagine at position 9 is occupied by hydrophobic residues such as Val, Leu and Ile. It is clear that many LRRs do not keep the complete HCS pattern and are irregular. Eighteen of the 22 known structures contain irregular LRRs. Most of the irregular HCSs can be classified into four groups; LxxLxLxx(**L/I/M**)xL, LxxLxLxx(**R/K/E**)xL, LxxLxLxx(N/S/T/**A**)xx, and LxxLxL**xxxx**x in which residues in boldface are irregular. Also there are rare examples, **xV**xxLxLxxNxL and **P**xxLxLxxNxL in follicle-stimulating hormone receptor (FSHr), and LxxLx**G**xxS/PxI in Inl-C and DLC1 (Table 1). Furthermore, an irregular LRR with (L/x)xx(L/A)xCxx(**L/R**)xLxxVPxxIPxx, which belongs to the "bacterial" motif, is frequently observed at the first LRR (LRR1) at the N-terminus of the LRR domain (Table 1).

Multiple sequence alignment

Mammalian TLR2 contains 20 LRRs, as described later. The PFAM program detects only 5–7 of the 20 LRRs, while the InterPro database (20 August, 2006) counts 13 in chicken, 14 in human, Cynomolgus monkey, dog and Chinese hamster, and 18 in bovine (Table 2). Figures 2 and 3 show the multiple sequence alignment of the LRR domain in mammalian TLR2 from 14 species. The sixth LRR (LRR6) shows canonical and irregular LRRs whose HCS parts consist of LxxLxIxx(S/T/N)xL and LxxLxIxx(**Q/D**)xL or LxxLxIxx**L**xL, respectively. The VS part is "typical". Both canonical and irregular LRR are also seen in LRR9 and LRR10. Furthermore, the HCS part of LRR4 shows LxxLxLxxNxY in which position 11 is tyrosine. This pattern was recognized in the known structures of TLR3 and lingo-1 (Table 1). The pairwise sequence identities are >35%. Thus, all LRRs in TLR2s from the 14 species can be reasonably regarded as an LRR motif.

Protein secondary structure prediction

The result of the protein secondary structure prediction of human TLR2 having 20 LRRs is shown in Figure 4. Both SSpro4.0 and Proteus predict that 15 of the 20 LRRs prefer β-strands at positions 3 through 5 and/or its neighboring positions in the HCS part. They include all five irregular motifs, LRR4, LRR5, LRR7, LRR9, and LRR11. The occurrence of β-strands in LRR6 is predicted only by Proteus. However, LRR6 with the HCS part of LEELEIDASDL is clearly a canonical LRR. All twenty including LRR6 can be reasonably identified as LRR motifs.

The identification of LRRs within TLRs

These analyses of the known LRR structures, the multiple sequence alignments and the secondary structure predictions of TLR2 provide strong evidence that allow us to identify LRRs over an extended range of sequences and inferred structures. Taken together four steps for the identification of LRRs in each member of TLRs were used.

Step 1. Detection of LRRs by the PFAM program

Step 2. Identification of a candidate LRR that can not be recognized by PFAM.

Step 3: Evaluation of protein secondary structure predictions by Proteus and SSpro4.0.

Step 4. Determination of all LRRs in each member based on the results obtained by *Steps 1–3*.

In *Step 2*, the LRR candidates are selected using the criterion that they are longer than 18 residues and that the HCS part consisting of LxxLxLxxNxL occupies at least hydrophobic residues at positions 4 and 6. The candidate includes irregular motifs that are similar to LRRs recognized by the known structures. In case there are TLRs from many species, multiple sequence alignments are also considered for identification. In *Step 3*, the preference of β-strand in the HCS part of the LRR candidate selected by *Step 1* and *Step 2* was investigated. In S*tep 4*, when the candidate prefers β-strand by either Proteus or SSpro4.0 (at least in one species), it is identified as an LRR. In some cases such as LRR12 in TLR14, the initial LRR candidate was changed into another LRR based on the results of the secondary structure prediction. The crystal structure of human TLR3 [52,53] contains 25 LRRs. The present method confirmed this. In contrast, the PFAM and SMART programs predicted only 16–17 LRRs and the databases have reported 22 LRRs (Table 2).

There are two exceptions. In five mammalian TLR6s with 20 LRRs, LRR9, **P**TLLN(F/V/L)TL(N/Q)H(I/V), that contains Pro at position 1 is not predicted to have a β-strand by both prediction methods (Figure 4). However, this pattern is seen in FSHr (Table 1). Similarly, in human and pig TLR10 with 20 LRRs LRR10, GGK(A/V)YLDHNSF, is not predicted to have a β-strand by both programs (Figure 4). However, this pattern shows remarkable similarity with the sixteenth LRR (LRR16) in TLR7 and TLR8 with 27 LRRs.

LRRs within TLRs

LRR motifs

The repeat number and "phasing" of LRRs in TLRs are summarized in Table 2 and Figures 5, 6, 7, 8, 9, and 10. The number of LRRs identified within TLRs range from 16 to 28; these numbers are larger than those reported in most databases. The "typical"; "***T***", LRR, LxxLxLxxNxLxx-LxxxxFxxLxx, occurs most frequently followed by shorter motifs including LxxLxLxxNxLxxLPx(x)LPxx ("bacterial"; "**S**") with 19–21 residues. Moreover, all of the C-terminal LRR consists of LxxLxLxxNP(F/L)xCxCxxxx(F/L)xxxx. The TLRs contain a variety of irregular LRRs (Figures 5, 6, 7, 8, 9, and 10). The first LRR at the N-terminus (LRR1) is fre-

Table 1: Irregular LRR motifs observed in the known structures of LRR-containing proteins

Protein[a]	N[b]	Irregular LRRs			PDB
		Position	Amino acid sequence[c] **LxxLxLxxNxL**		
TLR3	25	LRR1	**H**EVADCSH**L**KL	TQVPDDLPTN	1ZIW
		LRR9	LFGLFLNN**V**QL	GPSLTEKLCLELANTS	/2A0Z
		LRR13	VRYLNLKRSF**T**	KQSISLASLPKIDDFSFQWLKC	
		LRR15	LKYLSLSNSF**T**	SLRTLTNETFVSLAHSP	
		LRR18	IFEIYLSYNK**Y**	LQLTRNSFALVPS	
		LRR19	LQRLMLRR**V**AL	KNVDSSPSPFQPLRN	
CD14	11	LRR1	AADVELYG**G**G**R**	SLEYLLKRVDTEADLGQFTDIIKSLS	1WWL
		LRR2	LKRLTVRA**A**RI	PSRILFGALRVLGISG	
		LRR3	LQELTLEN**L**EV	TGTAPPPLLEATGPD	
		LRR4	LNILNLRN**V**SW	ATRDAWLAELQQWLKPG	
		LRR5	LKVLSIAQ**A**H**S**	LNFSCEQVRVFPA	
		LRR7	LQVLALRN**A**GM	ETPSGVCSALAAARVQ	
Slit	6	LRR1	**G**TTVDCTG**R**GL	KEIPRDIPLH	1W8A
Lingo-1	14	LRR8	LIVLRLRH**L**NI	NAIRDYSFKRLYR	2ID5
		LRR9	LKVLEISH**W**P**Y**	LDTMTPNCLYGLN	
Decorin	12	LRR1	LRVVQCSD**L**GL	EKVPKDLPPD	1XCD
Biglycan	12	LRR1	LRVVQCSD**L**GL	KAVPKEISPD	2FT3
FSHr	10	LRR6	LQKVLLDI**Q**D**N**	INIHTIERNSFVGLSFE	1XWD
		LRR7	**S**VILWLNKNGI	QEIHNCAFNGTQ	
		LRR9	**P**VILDISRTRI	HSLPSYGLEN	
Inl-A	16	LRR1	VTTLQADR**L**GI	KSIDGLEYLNN	1O6V
Inl-C	7	LRR1	VQNFN**G**DNSNI	QSLAGMQFFTN	1XEU
		LRR7	VNWIDLTG**Q**KC	VNEPVKYQPEL	
U2snRNPA'	5	LRR1	DRELDLRG**Y**KI	PVIRNLGATLDQ	1A9N
RabGGTα	5	LRR1	VRVLHLAH**K**DL	TVLCHLEQLLL	1DCE
DLC1	6	LRR1	**K**VELH**G**MI**P**PI	EKMDATLSTLKA	1DS9
TAP	4	LRR1	**Q**QALDLKG**L**R**S**	DPDLVAQNIDVVLNRRSCMAATLRII	1FT8
				EENIPE	
PGIP	10	LRR1	VNNLDLSG**L**NL	PKPYPIPSSLANLPYL	1OGQ
YoPM	16	LRR1	AHELELNN**L**GL	SSLPELPPH	1G9U
		LRR16	VEDLRMNS**E**RV	DPYEFAHETTDKLEDDVFE	
hRI	16	LRR15	LEQLVLYD**I**YW	SEEMEDRLQALEKDKP	1Z7X
pRI	17	LRR1	W_NLDIHC**E**QL	SDARWTELPLQQ	2BNH
RanGAP	10	LRR2	LEIAEFSD**I**F**T**	GRVKDEIPEALRLLLQALLKCPK	1YRG
cTmod	5	LRR1	LEEVNLNN**I**M**N**	IPVPTLKACAEALKTNTY	1IO0
		LRR2	VKKFSIVGTR**S**	NDPVAFALAEMLKVNNT	
		LRR4	LIELRIDN**Q**S**Q**	PLGNNVEMEIANMLEKNTT	
		LRR5	LLKFG**Y**HFTQ**Q**	PRLRASNAMMNNNDLVRK	
ceTmod	5	LRR1	LKEVNINN**M**KR	VSKERIRSLIEAACNSKH	1PGV
		LRR4	IVEFKADN**Q**R**Q**	SVLGNQVEMDMMMAIEENES	
		LRR5	LLRVGISF**A**S**M**	EARHRVSEALERNYERVRL	
Skp2	11	LRR1	WQTLDLTG**K**NL	HPDVTGRLLSQG	1FQV
		LRR4	LQNLSLWF**L**RL	SDPIVNTLAKNSN	/2ASS
		LRR7	ITQLNLSG**Y**R**K**	NLQKSDLSTLVRRCPN	
		LRR10	LKTLQVFG**I**V**P**	DGTLQLLKEA	

[a]FSHr, follicle-stimulating hormone receptor; Inl-A, internalin A; Inl -C, internalin C; U2snRNPA', spliceosomal U2A' protein; RabGGTα, rab geranylgeranyltransferase α-subunit; DLC-1, *Chlamydomonas* outer arm dynein light chain 1; TAP, mRNA export factor; PGIP, polygalacturonase-inhibiting protein; RI, ribonuclease inhibitor, RanGAP, GTPase-activating protein; Tmod, tropomodulin; Skp2, S-phase kinase-associated protein 2; h, human, p, pig; ce, Caenorhabditis elegans: [b] N; the repeat number of LRRs identified in individual known LRR structures.[c] HCS, highly conserved segment of LRR; VS, variable segment of LRR. Residues in boldface cause irregular motif.

Table 2: The repeat number of LRRs and its flanking cysteine clusters in vertebrate TLRs

Famliy[a]	Protein	Species[b]	N_L[c]	LRRNT[d]	LRRCT[d]	N_B[e]
TLR1	TLR1	h, m, p	20(8–9)	No	$CxCx_{25}Cx_{20}C$	1
	"	t	21	$CxCx_{29}Cx_{20}C$	$CxCx_{29}Cx_{20}C$	1
	TLR2	h, m, p, b, r, d, ra, g, ho, dwb, ha, cm, n	20(14–18)	Cx_5C	$CxCx_{24}Cx_{20}C$	1
	TLR2.1	c	20(13)	Cx_2Cx_5CxC	$CxCx_{24}Cx_{20}C$	1
	TLR2.2	c	20(13)	Cx_2Cx_5CxC	$CxCx_{24}Cx_{20}C$	1
	TLR2	t	20	Cx_2Cx_5CxC	$CxCx_{25}Cx_{20}C$	1
	TLR2	jf	19	Cx_2Cx_5CxC	$CxCx_{25}Cx_{20}C$	1
	TLR2	z	21	Cx_3CxCx_5CxC	$CxCx_{24}Cx_{20}C$	1
	TLR6	h, m, r, p, b	20(13–14)	No	$CxCx_{25}Cx_{20}C$	1
	TLR10	h, p	20(12)	No	$CxCx_{25}Cx_{20}C$	1
	TLR14	t, z	21	$Cx_{10}C$	$CxCx_{25}Cx_{20}C$	1
TLR3	TLR3	h, m, b, r, bu, rm, z	25(22–24)	Cx_8C	$CxCx_{25}Cx_{18}C$	1
	"	t	25	Cx_8C	$CxCx_{26}Cx_{18}C$	1
	"	jf	27	Cx_8C	$CxCx_{24}Cx_{18}C$	1
	TLR	as	27	$Cx_{15}C$	$CxCx_{24}Cx_{18}C$	1
	"	rt	27	$Cx_{15}C$	$CxCx_{24}Cx_{18}C$	1
	"	go	27	$Cx_{14}C$	$CxCx_{24}Cx_{18}C$	1
	TLRII	rt	27	$Cx_6Cx_{17}C$	$CxCx_{24}Cx_{18}C$	1
TLR4	TLR4	h, lg, pc, ob, or	23(21)	$Cx_{10}C$	$CxCx_{23}Cx_{17}C$	1
	"	m, p, r, ch, ho, b, ab, n,	23(18–19)	$Cx_{10}C$	$CxCx_{23}Cx_{18}C$	1
	"	ha, ra	23	$Cx_{10}C$	$CxCx_{23}Cx_{15}C$	1
	TLR4b	z	23	$Cx_{10}C$	$CxCx_{23}Cx_{18}C$	1
	TLR4	h [Q5VZ17]	16	$Cx_{10}C$	$CxCx_{23}Cx_{18}C$	1
	"	d	16	No	$CxCx_{23}C$	1
TLR5	TLR5	h, m, r, p, b, jhm	22(15–16)	$Cx_{11}C$	$CxCx_{24}Cx_{18}C$	1
	"	t	22	$Cx_{10}C$	$CxCx_{24}Cx_{22}C$	1
	"	rt	22	Cx_8C	$Cx_{27}C$	1
	TLRS5	t	23	Cx_8C	$Cx_{26}C$	1
TLR11	TLR11	M	25(11)	$Cx_{17}Cx_{11}Cx_{20}C$	$CxCx_{24}Cx_{19}C$	1
	TLR12	m	24(17)	$Cx_{17}Cx_{11}Cx_{20}C$	$CxCx_{24}Cx_{19}C$	1
	TLR13	m	27(21)	$Cx_{11}C$	$CxCx_{24}Cx_{16}C$	1
	TLR21	t	27	$Cx_{10}C$	$CxCx_{24}Cx_{15}C$	1
	TLR22	t	27	Cx_8C	$CxCx_{24}Cx_{18}C$	1
	TLR23	t	27	$Cx_{14}C$	$CxCx_{24}Cx_{18}C$	1
TLR7	TLR7	H, m, d	27(27–28)	$Cx_{14}C$	$CxCx_{24}Cx_{18}C$	2
	"	t	27	$Cx_{12}C$	$CxCx_{24}Cx_{18}C$	2
	TLR8	h, m, p	27(24–26)	$Cx_{12}C$	$CxCx_{24}Cx_{18}C$	2
	"	t	27	$Cx_{13}C$	$CxCx_{24}Cx_{18}C$	2
	TLR9	h, m, p, b, d, ca, ho, s, mnm	27(26)	Cx_9C	$CxCx_{23}Cx_{18}C$	2
	"	t, jf, gsb	27	$Cx_{10}C$	$CxCx_{23}Cx_{18}C$	2
	TLR	gp	28, 26	?	$CxCx_{24}Cx_{18}C$	2
Others	TLRa	Jl	21	Cx_8C	$CxCx_{22}C_{12}C$	1
	TLRb	jl	21	Cx_8C	$CxCx_{22}Cx_{20}C$	1
	TLR15	c	21	No	$CxCx_{24}Cx_{20}C$	2

[a]*The six families and others*. [b]Abbreviations: ab, American bison; as, Atlantic salmon; b, bovine; c, chicken; ca, cat; ch, Chinese hamster; cm, Cynomolgus monkey; d, dog; dwd, Domestic water buffalo; g, goat; go, Goldfish; gp, green puffer; gsb, Gilthead sea bream; h, human; ha, hamster; ho, horse; jf, Japanese flounder; jhm, Japanese house mouse; jl, Japanese lamprey; lg, lowland gorilla; m, mouse; mnm, Ma's night monkey; n, Nilgai; ob, Olive baboon; or, Orangutan; p, pig; pc, Pygmy chimpanzee; r, rat; ra, rabbit; rm, Rhesus macaque; rt, Rainbow trout; s, sheep; t, Takifugu rubripes; z, Zebrafish. [c]N_L; the repeat number of LRRs. The number of parenthesis is N_L reported in the Interpro datbase.[d] LRRNT, the cysteine clusters on the N-terminal side of LRRs; LRRCT, the cysteine clusters on the C-terminal side of LRRs. [e]N_B; the number of horseshoe domains of the LRRs.

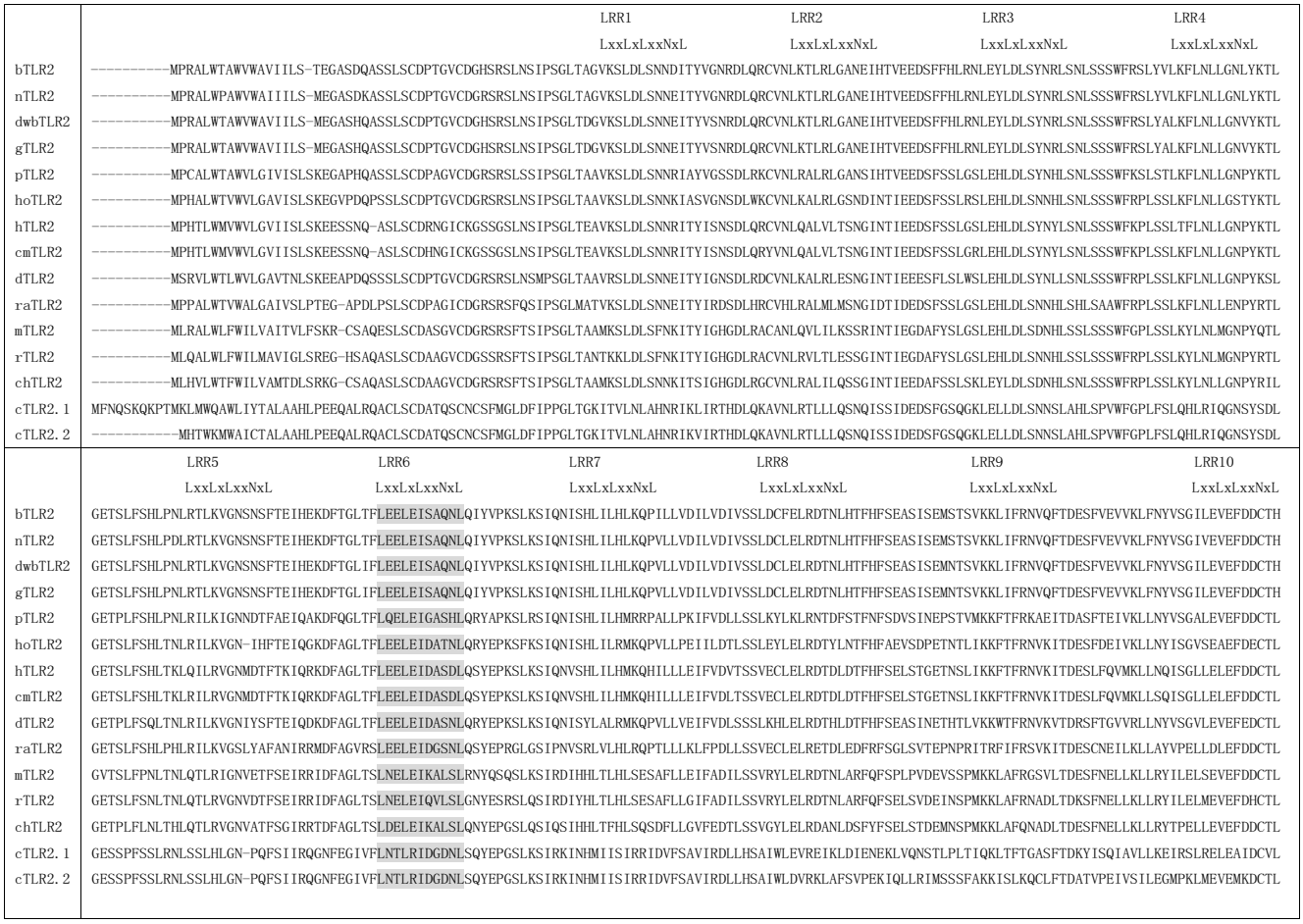

Figure 2

The multiple sequence alignment of LRRs within mammalian TLR2 from 14 species. bTLR2 [Q95LA9], nTLR2 [Q2V897], dwbTLR2 [Q2PZH4], gTLR2 [ABI31733], pTLR2 [Q59HI8], hoTLR2 [AAR08196], hTLR2 [O60603], cmTLR [Q95M53], dTLR2 [Q689D1], raTLR2 [AAM50059], mTLR2 [Q9QUN7], rTLR2 [Q6YGU2], chTLR2 [Q9R1F8], cTLR2.1 [Q9DD78], cTLR2.2 [Q9DGB6]. Abbreviations: b, Bovine; n, nilgai; dwb, domestic water buffalo; g, goat; p, pig; ho, horse; h, human; cm, Cynomolgus monkey; d, dog ; ra, rabbit; m, mouse; r, rat; ch, Chinese hamster; c, chicken. This panel shows the sequences from the N-termini to LRR10.

quently irregular, e.g. (L/x)xx(L/A)xCxx(**L**/**R**)xLxxVPxx-IPxx. This motif has been seen in the structures of TLR3, Slit, decorin, and biglycan, as noted. (Table 1). Methionine and tryptophan sometimes occupy positions 1, 4, 6 and 11 in the LxxLxLxxNxL motif, which are strongly hydrophobic. Moreover, as recognized in the known LRR structures, there are rare examples, **x**V**x**xLxLxxNxL, **P**xxLx-LxxNxL and LxxLx**G**xxS/PxI. The first motif is sometimes observed in LRR7 in human TLR10, LRR8 in Takifugu rubripes TLR14, LRR4 in chicken TLR15, LRR8 in human TLR4, and LRR16 in human TLR9 (Figure 5, 6, 7, 8, 9, and 10). Furthermore, the HCS parts of a twelve residue stretch, LxxLx(L/V/M/F)xx(S/N)xx(F/M), are sometimes observed; they include LRR5 in TLR2 from pig, bovine, nilgai, and domestic water buffalo with 20 LRRs (Figures 2 and 3), LRR11 in mouse TLR4 with 23 LRRs, and LRR14 in TLR4 from pig, bovine, rabbit, and nilgai with 23 LRRs.

LRRs in the six major families of TLRs

There are six major families of vertebarate TLRs [4]. The *TLR1* family consists of TLR1, TLR2, TLR6 and TLR10. This family contains 19–21 LRRs and has fewer numbers than do the other families except for Dog TLR4 [Q8SQH3] [56] and human TLR4 variant [Q5VZ17] in the *TLR4* family. Mammalian TLR1 contains 20 LRRs (Table 2). In contrast, Takifugu rubripes TLR1 [4] has one additional LRR at the N-terminus whose sequence is **R**NYIDLSS-RNLSSVPGDLPKE, that is a "bacterial" type. Mammalian and Takifugu rubripes TLR2 contains 20 LRRs. Japanese flounder TLR2 lacks one LRR that corresponds to LRR7 in

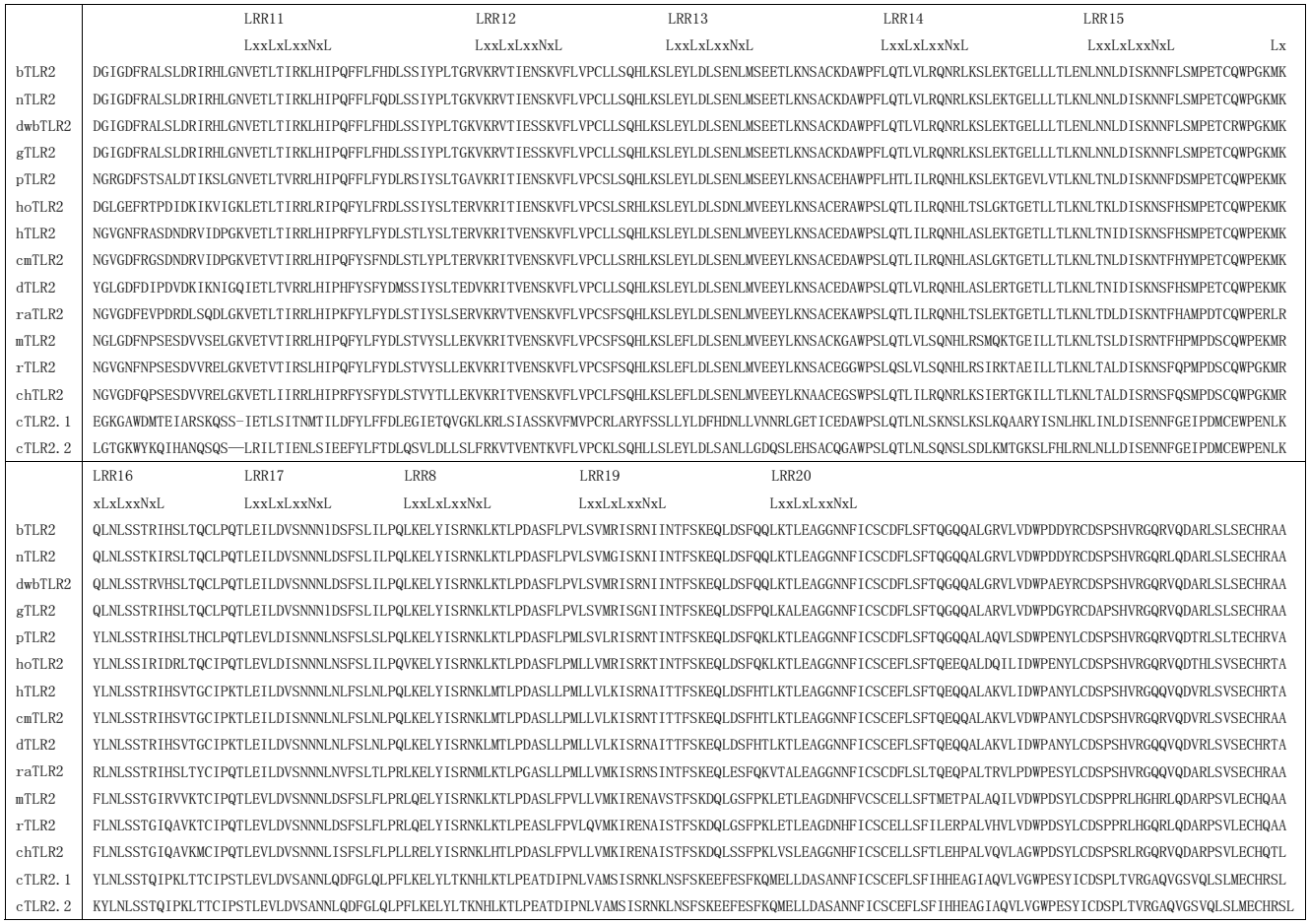

Figure 3
The multiple sequence alignment of LRRs within mammalian TLR2 from 14 species. This panel continued from Figure 2 shows the sequences from LRR11 to the C-termini.

the 20 LRRs [57]. Conversely, zebrafish TLR2 has one additional LRR at the N-terminus, as does Takifugu rubripes TLR1. This *TLR1* family shows a feature that irregular LRRs mainly concentrate at the central part of the LRR domain. In the *TLR3* family, mammalian, Takifugu rubripes and Zebrafish TLR3 contain 25 LRRs as was confirmed by the crystal structure of human TLR3 [52,53]. However, Japanese flounder TLR3 [57] contains two additional LRRs. Similarity sequence search indicates that TLRs from rainbow trout, Atlantic salmon, and goldfish are very similar to Japanese flounder TLR3. TLR4 that constitute the *TLR4* family contains 23 LRRs. Fourteen of the 23 LRRs are similar to "typical". Seven LRRs are irregular. As seen in the *TLR1* family, 5 of the 7 irregular LRRs are in the central part of the LRR domain. Dog TLR4 [Q8SQH3] [56] and human TLR4 variant [Q5VZ17] are shorter by about 200 amino acids at the N-terminus. These two TLRs contain only 16 LRRs. It is also predicted that dog TLR4 has no transmembrane region (Figure 7). TLR5 contains 22 LRRs. Ten of these 22 LRRs are clearly "typical". LRR15 in mammalian TLR5 is only 19 residues (Figure 7); the homolog in Takifugu rubripes and rainbow trout is 24 residues long. The *TLR11* family contains 24–27 LRRs. Most of LRRs in mouse TLR11, TLR12, and TLR13 are "typical". The same feature is observed in Takifugu rubripes TLR21, TLR22 and TLR23. Two Japanese lamprey TLRs appear to belong to the *TLR1* family.

The *TLR7* family consists of TLR7, TLR8 and TLR9 and contains 27 LRRs. Cross dot plots were computed for all of TLR7, TLR8, and TLR9 from human and mouse, and green puffer TLR. More important the super-motif is about 80 residues. Superposition of 21 ((7 × 6)/2) cross dotplots for the seven proteins emphasize the super-repeat of LRRs at the N-terminal part of the LRR domain (Figure 11) [11]. This super-motif comes from nine LRRs from LRR1 to LRR9 in TLR7 and TLR8, and from eight LRRs from LRR2 to LRR9 in TLR9 (Figure 12). The sequence align-

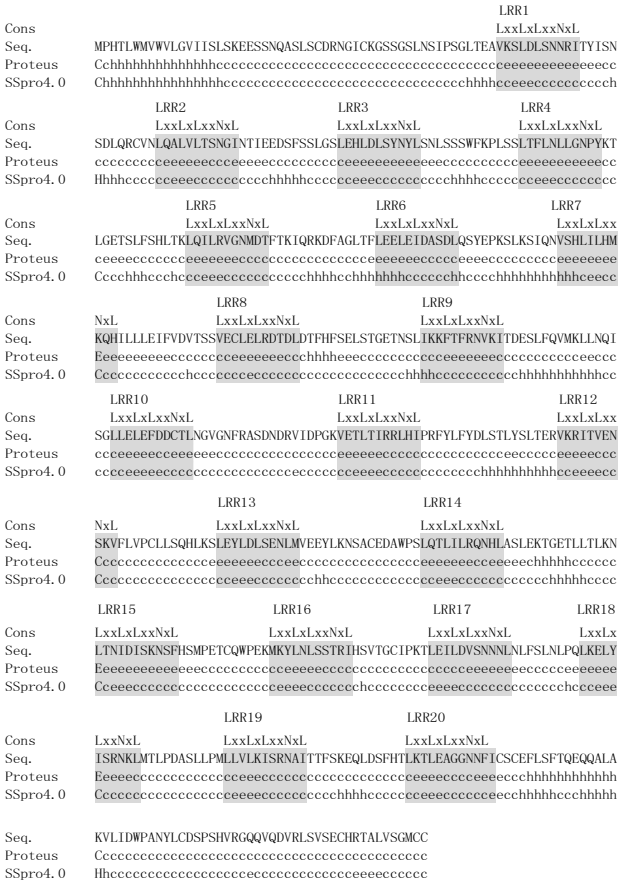

Figure 4
The secondary structure prediction of human TLR2 by SSpro4.0 and Proteus. The signal peptide and extracellular domain of hTLR2 [O60603] with 784 residues is shown; residues 1–588. The highly conserved segment of individual LRRs is highlighted by a shadow. Abbreviations: h, helix; c, coil; e, β-strand.

ment reveals two types of LRR, **S** and ***T***. The type *S* LRR is observed in the first, fourth, and seventh of the 9 tandem LRRs. All other LRRs are type ***T***. Although the third, the sixth and the seventh LRRs is longer than the second, the fifth and the eighth LRRs, their C-terminal VS parts keep the pattern of LxxxxFxxLxx that is seen in "typical" motif. Consequently their LRRs are type ***T***. Thus, there are three super-repeats, ***STTSTTSTT***, in TLR7 and TLR8, and two and two-third super-repeats, ***_TTSTTSTT***, in TLR9. Green puffer TLR forms two horseshoe domains of LRRs. The first domain is homologous to the TLR7 family and thus contains also the super-repeat of ***STTSTTSTT*** (Figure 12). LRR15 located at the central part of the 27 LRRs consists of long amino acid sequence with 73 residues in TLR7, 64 in TLR8, and 58 in TLR9, as seen in TLR15. This long LRR motif is observed in chicken TLR15. In all the case the next LRR, LRR16, is an irregular LRR that is described by (G/N)xLxLxxNx(I/L)xxVxxxxFxxLxx is similar to "typical" motif, although position 1 in the HCS part is not occupied by leucine.

Two cysteine clusters flanking the LRR domain

The LRRs within most of TLRs are flanked by two cysteine clusters, each of which contains two to five cysteine residues (Table 2 and Figures 5, 6, 7, 8, 9, and 10). Here the cysteine clusters on the N- and C-terminal sides of LRRs are termed LRRNT and LRRCT, respectively [58]. The N-terminal cluster usually consists of two cysteines, $Cx_{5-14}C$, but sometimes 3, 4 or 5 cysteines. With high frequency, as noted, the last cysteine of the clusters occupies a structurally equivalent position to those of leucines in the HCS part of LRR1. The Cx_8C motif in TLR3 and the $Cx_{10}C$ motif of TLR4 form a disulfide bond [52,59], as does the $Cx_{12}C$ motif in GPIbα [41]. The $Cx_{5-14}C$ motifs presumably form disulfide bonds. The C-terminal clusters, excepting those in three TLRs (Table 2), contain four cysteines consisting of $CxCx_{22-25}Cx_{15-20}C$. The spacing between the first and the second cysteine that are contained in the last LRR is the same for all the families. The other spacing appears to characterize each family. The $CxCx_{25}Cx_{18}C$ motif in TLR3 forms two disulfide bonds between the first and the third cysteines, and between the second and the fourth cysteines [52]. Such pairs of disulfide bonds have been observed for the $CxCx_{20}Cx_{21}C$ motif in Nogo receptor [38,39] and the $CxCx_{20}Cx_{19}C$ motif in Slit [49]. The disulfide bond connectivity can be inferred for TLRs. The C-terminal cluster for primate TLR4 ($CxCx_{23}Cx_{17}C$) is different from that of other mammalian TLR4 ($CxCx_{23}Cx_{18}C$). Only in rainbow trout TLR5 and Takifugu rubripes TLRS5 having no TIR domain, does the C-terminal cluster consists of two cysteines. There are no N-terminal cysteine clusters in TLR1, TLR6, TLR10, TLR15, and dog TLR4. However, the N-terminal amino acid sequence flanking the LRR domain might form a capping structure.

Discussion

LRRs within human TLRs

The present analyses of LRRs within vertebrate TLRs indicate that there are at least two types of LRR motifs; "typical"; "***T***", LRR, LxxLxLxxNxLxxLxxxxFxxLxx and "bacterial"; "**S**", LRR, LxxLxLxxNxLxxLPx(x)LPxx. Vertebrate TLRs contain 16–28 LRRs (Table 2 and Figures 5, 6, 7, 8, 9, and 10). Bell *et al.*, [60] have proposed that the ECDs of human TLRs comprised 19–25 LRRs including both "***T***" and "**S**" LRRs. Each member of human TLRs contain 1–2 times less LRRs than those identified here. Furthermore, in the *TLR1* family (TLR1, TLR2, TLR6 and TLR10) the LRRs at the central parts are aligned differently to each other. Such a difference is also seen in TLR4, TLR5, and TLR7. The alignments of TLR3, TLR8 and TLR9 are nearly identical except the first LRR at the N-terminus of the LRR domain and the last LRR at its C-terminus.

Figure 5

Sequence alignment of LRR domains within the six families of TLRs. (1) hTLR1 [Q15399]; hTLR2 [O60603]; hTLR6 [Q9Y2C9]; hTLR10 [Q9BXR5]; tTLR14 [Q5H726]. (2) hTLR3 [O15455]; jfTLR3 [Q76CT7]. (3) hTLR4 [O00204]; dTLR4 [Q8SQH3]. (4) hTLR5 [O60602]. (5) mTLR11 [Q6R5P0]; mTLR12 [Q6QNU9]; mTLR13[Q6R5N8]: tTLR21 [NP_001027751]; tTLR22 [Q5H723]; tTLR23 [AAW70378]; (6) hTLR7 [Q9NYK1]; hTLR8 [Q9NR97]; hTLR9 [Q9NR96]. (7) jlTLRa [Q33E93]; cTLR15 [ABB71177]. The complete amino acid sequences are shown for hTLR1 with 786 residues (res.), hTLR2 with 784 res, hTLR6 with 796 res., hTLR10 with 811 res., tTLR14 with 871 res., hTLR3 with 904 res., jfTLR3 with 961 res., hTLR4 with 839 res., dTLR4 with 636 res., hTLR5 with 858 res., hTLR7 with 1049 res., hTLR8 with 1041 res., hTLR9 with 1032 res., mTLR11 with 926 res., mTLR12 with 906 res., mTLR13 with 991 res., tTLR21 with 965 res., tTLR22 with 950 res., tTLR23 with 941res., jlTLRa with 813res., and cTLR15 with 868 res., Cysteine is highlighted in magenta. Its boldface indicates cysteines in LRRNT or LRRCT. Residues of missense mutation are highlighted in blue boldface. SIGNAL, signal peptide sequence; LRRNT, the cysteine clusters on the N-terminal side of LRRs; LRRCT, the cysteine clusters on the C-terminal side of LRRs; TRANS, transmembrane region; CYTOP, cytoplasmic region. Abbreviations: h, human; m, mouse; t, Takifugu rubripes; c, chicken; d, dog; jf, Japanese flounder. This panel shows hTLR1, hTLR2, jfTLR2 and TLR6 in the TLR1 family.

hTLR10

```
SIGNAL  MRLIRNIYIFCSIVMTAEGDAPELPEERELMTNCSNMSLRKV
        PADLTPA
        LxxLxLxxNxL
LRR1    TTTLDLSYNLL FQLQSSDFHSVSK
LRR2    LRVLILCHNRI QQLDLKTFEFNKE
LRR3    LRYLDLSNNRL KSVTWYLLAG
LRR4    LRYLDLSFNDF DTMPICEEAGNMSH
LRR5    LEILGLSGAKI QKSDFQKIAHLH
LRR6    LNTVFLGFRTL PHYEEGSLPILNTT
LRR7    KLHIVLPMDTN FWVLLRDGIKTSKILEMTNIDGKSQFVSYEM
LRR8    QRNLSLENAKT SVLLLNKVDLL
LRR9    WDDLFLILQFV WHTSVEHFQIRNVTF
LRR10   GGKAYLDHNSF DYSNTVMRTIK
LRR11   LEHVHFRVFYI QQDKIYLLLTKMD
LRR12   IENLTISNAQM PHMLFPNYPTK
LRR13   FQYLNFANNIL TDELFKRTIQLPH
LRR14   LKTLILNGNKL ETLSLVSCFANNTP
LRR15   LEHLDLSQNLL QHKNDENCSWPET
LRR16   VVNMNLSYNKL SDSVFRCLPKS
LRR17   IQILDLNNNQI QTVPKETIHLMA
LRR18   LRELNIAFNFL TDLPGCSHFSR
LRR19   LSVLNIEMNFI LSPSLDFVQSCQE
LRR20   VKTLNAGRNPF RCTCELKNFIQLE
LRRCT   TYSEVMMVGWSDSYTCEYPLNLRGIRLKDVHLHELSCNTA
TRANS   LLIVTIVVIMLVLGLAVAF
CYTOP   CCLHFDLPWYLRMLGQCTQTWHRVRKTTQEQLKRNVRF
        HAFISYSEHDSLWVKNELIPNLEKEDGSILICLYESYF
        DPGKSISENIVSFIEKSYKSIFVLSPNFVQNEWCHYEF
        YFAHHNLFHENSDHIILILLEPIPFYCIPTRYHKLKAL
        LEKKAYLEWPKDRRKCGLFWANLRAAINVNVLATREMY
        ELQTFTELNEESRGSTISLMRTDCL
```

tTLR14

```
SIGNAL  MIWTFVHFIALAVGVLA
LRRNT   STTPSPPSTTKTVTGFCRVFNS
        LxxLxLxxNxL
LRR1    GRSADCLGMQL SSVPWRQFPPS
LRR2    LEDIDLSYNRL QVINAEDFALFPR
LRR3    LRSLNLKYNNI SRIDSDAFKNNPL
LRR4    LEILDIFNNSL GEIPVAALSPLLN
LRR5    LKKLYMSNNLY KRAALAETFSTFVR
LRR6    LQTLSMGGPLV EGLKKGDFQP
LRR7    LRKLRLQEFAI KCSSNLRYYEAGSLEVVQTQKL
LRR8    GFDMAIDQRPS ALVDMLRDIANKT
LRR9    FIAIQFRNLFE FRYYTRVQDIFQG
LRR10   LKHVAAYQLIF HRGKFNENLLRMALLNLEA
LRR11   IKRLRFQYIDF ARSPTFVDNRAGSSITDLV
LRR12   LDKLDLWYISN PDVLRFDWRFTWFNN
LRR13   IRSLSIQYVYF NSVPCDSWAEMKQ
LRR14   VKVLDVSNNRL TDTYIFNQLCNYKGAAPN
LRR15   LRLFNMSNNEL TSLKDLSLLTKEFQQ
LRR16   LQELDLSRNKL GSAAESRNCIWQKS
LRR17   ITRFIVHHNNF ESSALHCLPTS
LRR18   VEFLDLSFCDL DQLDMNYFSKTSN
LRR19   LQELHLSGNKI KFIPSKWASPS
LRR20   LQSLSLDGNSF GLIGTESFQDMPR
LRR21   LSHLSAGNNPY HCTCELHAFVQET
LRRCT   ITEGKVNLTDWPWNYKCYHPEPLLNTVISQYLPGKVACDIR
TRANS   LVIVICVAATTFVVLILVLI
CYTOP   CYIFDLPWYTKATFQIIRAKYRAHKEKAAGEEGPFTYH
        AFISYSHSDADWVRDQLLPCLENNNNPYRLCIHERDFT
        PGRWIIDNIIENIENSRKVIFVLSRHFVNSEWCNYELY
        FAQQRAMGKTFSDVILVVKEP IDPNSLPSKYCKLKKM
        LSTKTYLEWPQQVNQQAFFWAQLRSVLGRPTAVTRGRQ
        SVRSRTSSASISVIGPLVDERNPEMDEDRGTEPNYEVI
        ENSLEVSHQRQIPMVAV
```

(2) The *TLR3* family

hTLR3

```
SIGNAL  MRQTLPCIYFWGGLLPFGMLCAS
LRRNT   STTKCTVS
        LxxLxLxxNxL
LRR1    HEVADCSHLKL TQVPDDLPTN
LRR2    ITVLNLTHNQL RRLPAANFTRYSQ
LRR3    LTSLDVGFNTI SKLEPELCQKLPM
LRR4    LKVLNLQHNEL SQLSDKTFAFCTN
LRR5    LTELHLMSNSI QKIKNNPFVKQKN
LRR6    LITLDLSHNGL SSTKLGTQVQLEN
LRR7    LQELLLSNNKI QALKSEELDIFANSS
LRR8    LKKLELSSNQI KEFSPGCFHAIGR
LRR9    LFGLFLNNVQL GPSLTEKLCLELANTS
LRR10   IRNLSLSNSQL STTSNTTFLGLKWTN
LRR11   LTMLDLSYNNL NVVGNDSFAWLPQ
LRR12   LEYFFLEYNNI QHLFSHSLHGLFN        F303S
LRR13   VRYLNLKRSFT KQSISLASLPKIDDFSFQWLKC
LRR14   LEHLNMEDNDI PGIKSNMFTGLIN
LRR15   LKYLSLSNSFT SLRTLTNETFVSLAHSP
LRR16   LHILNLTKNKI SKIESDAFSWLGH
LRR17   LEVLDLGLNEI GQELTGQEWRGLEN
LRR18   IFEIYLSYNKY LQLTRNSFALVPS
LRR19   LQRLMLRRVAL KNVDSSPSPFQPLRN
LRR20   LTILDLSNNNI ANINDDMLEGLEK
LRR21   LEILDLQHNNL ARLWKHANPGGPIY        H539E
                    FLKGLSH               N541A
LRR22   LHILNLESNGF DEIPVEVFKDLFE
LRR23   LKIIDLGLNNL NTLPASVFNNQVS
LRR24   LKSLNLQKNLI TSVEKKVFGPAFRN
LRR25   LTELDMRFNPF DCTCESIAWFVNWIN
LRRCT   ETHTNIPELSSHYLCNTPPHYHGFPVRLFDTSSCKDSAPFE
TRANS   LFFMINTSILLIFIFIVLLI
CYTOP   HFEGWRISFYWNVSVHRVLGFKEIDRQTEQFEYAAYIIH
        AYKDKDWVWEHFSSMEKEDQSLKFCLEERDFEAGVFELE
        AIVNSIKRSRKIIFVITHHLLKDPLCKRFKVHHAVQQAI
        EQNLDSIILVFLEEIPDYKLNHALCLRRGMFKSHCILNW
        PVQKERIGAFRHKLQVALGSKNSVH
```

jfTLR3

```
SIGNAL  MGPGGKEDERETGGRKHFQLVTFFFLLTISSLLAPISGFA
LRRNT   LKTCRIS
        LxxLxLxxNxL
LRR1    YNIAKCDKMGL TAAPRDIPSA
LRR2    VKGFDLSENKI LRVLVSDFENLPG
LRR3    LTQLDLNRNLI SQIDDGAFANLIF
LRR4    LKKLNLNNNKL VTLGENLFHGLSN
LRR5    LTELRIMSNGI KAVTLTSFKPMNS
LRR6    LKFLDFSHNKL KHITKVRSIIQHLPN
LRR7    LRELFLKKNDF TTFHSEELTNSSLQ
LRR8    LKALDLSQNPI TDFQITANVFPN
LRR9    LTWLNIGGAPG KTPVILGVRNKTFFSR
LRR10   VSTLDITGLRM TLVDNRTLLGTVNSS
LRR11   LSSLRMNAMKN NITALTHISCTIPT
LRR12   LSTLHVRHNKL TYVSSDLFKLCFN
LRR13   IREIDLTDNKI KKIRDDAFSSLQS
LRR14   LKTLSLSRNKL SSVPYATRTLPS
LRR15   LGELDLSFNNI TKLGCDDFANQTK
LRR16   LRRLRLYHNSI ASLAECVFKDLVQ
LRR17   LQVLKLQNNHL SNLNGAFRDCLPN
LRR18   LRQLLLNGNQL TALKHGEFRGLQS
LRR19   LQNLSLHENKI FNLDKGCFVGLTN
LRR20   LTDILLQNNQI RETEISKGVFNDLIN
LRR21   LRRLELRDNHI KYVNNSSLPSAPFSRLSR
LRR22   LETLAIPSQHG KGKSQLPRNLLEGLTN
LRR23   LLVFNIRNIQL ASLHKDMFNGTPQ
LRR24   LTTLDISSNEL MDLSPDLFSPIPN
LRR25   LKSLYVSRTSL RSLDYLTGANLTK
```

1

Figure 6

Sequence alignment of LRR domains within the six families of TLRs. This panel continued from Figure 5 shows hTLR10 and hTLR14 in the TLR1 family, and hTLR3 and jfTLR3 in the TLR3 family. .

```
LRR25   LKSLYVSRTSL RSLDYLTGANLTK
LRR26   LEFLQARKNEF SIISEEIIKSVPS
LRR27   LVYADFQGNSF TCKCDNAWFIKWV
LRRCT   EYYNQTQVFDAYNFECNYPLNLKGTKLLHFDIRSCSVD
TRANS   AGFLMFLSTTCTTLLFML
CYTOP   TSFTYHFLRWHLAYAYYFFLALLFDTKHKNKQPPNQYD
        AFISYNTHDEPWVVRELLPKLEGEQGWRLCLHHRDLMP
        GKPIVENIVDAIYGSRKTICVISRRYLESEWCSREMQV
        ASFRLFDEQKDVLILVFLEDIPTDELSPYYRMKKLLNK
        MSYLSWPRAAEHTELFWEKLRQALRTREDQADESFRLT
        VVDNQW
```

(3) The *TLR4* family

hTLR4

```
SIGNAL  MMSASRLAGTLIPAMAFLSCVRP
LRRNT   ESWEPCVEVVP
        LxxLxLxxNxL
LRR1    NITYQCMELNF YKIPDNLPFS
LRR2    TKNLDLSFNPL RHLGSYSFFSFPE
LRR3    LQVLDLSRCEI QTIEDGAYQSLSH
LRR4    LSTLILTGNPI QSLALGAFSGLSS
LRR5    LQKLVAVETNL ASLENFPIGHLKT       T135A
LRR6    LKELNVAHNLI QSFKLPEYFSNLTN
LRR7    LEHLDLSSNKI QSIYCTDLRVLHQMPLL
LRR8    NLSLDLSLNPM NFIQPGAFKEIR
LRR9    LHKLTLRNNFD SLNVMKTCIQGLAGLE
LRR10   VHRLVLGEFRN EGNLEKFDKSALEGLCNLT
LRR11   IEEFRLAYLDY YLDDIIDLFNCLTN       D299G
LRR12   VSSFSLVSVTI ERVKDFSYNFG
LRR13   WQHLELVNCKF GQFPTLKLKS
LRR14   LKRLTFTSNKG GNAFSEVDLPS
LRR15   LEFLDLSRNGL SFKGCCSQSDFGTTS      T399I
LRR16   LKYLDLSFNGV ITMSSNFLGLEQ
LRR17   LEHLDFQHSNL KQMSEFSVFLSLRN
LRR18   LIYLDISHTHT RVAFNGIFNGLSS
LRR19   LEVLKMAGNSF QENFLPDIFTELRN
LRR20   LTFLDLSQCQL EQLSPTAFNSLSS
LRR21   LQVLNMSHNNF FSLDTFPYKCLNS
LRR22   LQVLDYSLNHI MTSKKQELQHFPSS
LRR23   LAFLNLTQNDF ACTCEHQSFLQWIK
LRRCT   DQRQLLVEVERMECATPSDKQGMPVLSLNITCQMNKT
TRANS   IIGVSVLSVLVVSVVAVLV
CYTOP   YKFYFHLMLLAGCIKYGRGENIYDAFVIYSSQDEDWVR
        NELVKNLEEGVPPFQLCLHYRDFIPGVAIAANIIHEGF
        HKSRKVIVVVSQHFIQSRWCIFEYEIAQTWQFLSSRAG
        IIFIVLQKVEKTLLRQQVELYRLLSRNTYLEWEDSVLG
        RHIFWRRLRKALLDGKSWNPEGTVGTGCNWQEATSI
```

dTLR4

```
        MPLL
        LxxLxLxxNxL
LRR1    NLSLDLSLNPL YFIQPGSFKEIK
LRR2    LHKLTLRSNFN STDVMKTFIQS
LRR3    LAGLKINQLVL GEFKNERKLESFDNSLLEG
LRR4    LCNLTIEKFRI AYFDSFSKDTTNLFNQLVN
LRR5    ISAISLAHLYL DTPKYLPKNLR
LRR6    WQRLEIVNCNL EQFPAWELDS
LRR7    LKEFVLTSNKG MNTFADMKMES
LRR8    LEFLDLSRNRL SFKTCCSHSDFGTTR
LRR9    LKHLDLSFNEI ITMSSNFLGLEQ
LRR10   LEYLDLQHSSL KQASDFSVFLSLRN
LRR11   LRYLDISYTRT EVAFQGIFDGLVS
LRR12   LEVLKMADNSF PDNSLPNIYKGLTN
LRR13   LTILDLSRCHL ERVSQESFVSLPK
LRR14   LQVINMSHNSL LSLDTLAYEPLLS
LRR15   LQILDCSFNRI VAFKEQGQQHFPSN
LRR16   LVSLNLTRNSF ACDCEHQSFLQWVK
        DHRQLLVKVEQMVCAKPLDMKDMPLLSFRNATLSEEAR
        LSISVSVFTVLHGFSGSSFSRYKFYFHLMLLAWLAKGI
        TEGKVPMMHFVIYSSSQDEDWVRNELVKNLEEGVPPFQL
        CLHYRDFIPGVAIAANIIQEGFYKSRKVIVVVSQHFIQ
        SRWCIFEYEIAQTWQFLSSRAGIIFIVLQKVEKSLLRQ
        QVELYRLLSRNTYLEWEDSVLGRHIFWRRLRKALLDGK
        PWSPEGTEDAEKS
```

(4) The *TLR5* family

hTLR5

```
SIGNAL  MGDHLDLLLGVVLMAGPVFG
LRRNT   IPSCSFDGRIAFYRFCNLTQVPQVLNT
        LxxLxLxxNxL
LRR1    TERLLLSFNYI RTVTASSFPFLEQ
LRR2    LQLLELGSQYT PLTIDKEAFRNLPN
LRR3    LRILDLGSSKI YFLHPDAFQGLFH
LRR4    LFELRLYFCGL SDAVLKDGYFRNLKA
LRR5    LTRLDLSKNQI RSLYLHPSFGKLNS
LRR6    LKSIDFSSNQI FLVCEHELEPLQGKT
LRR7    LSFFSLAANSL YSRVSVDWGKCMNPFRNMV
LRR8    LEIVDVSGNGW TVDITGNFSNAISKSQA
LRR9    AFSLILAHHIM GAGFGFHNIKDPDQNTFAGLARSS
LRR10   VRHLDLSHGFV FSLNSRVFETLKD
LRR11   LKVLNLAYNKI NKIADEAFYGLDN
LRR12   LQVLNLSYNLL GELCSSNFYGLPK
LRR13   VAYIDLQKNHI AIIQDQTFKFLEK
LRR14   LQTLDLRDNAL TTIHFIPS
LRR15   IPDIFLSGNKL VTLPKINLT
LRR16   ANLIHLSENRL ENLDILYFLLRVPH
LRR17   LQILILNQNRF SSCSGDQTPSENPS
LRR18   LEQLFLGENML QLAWETELCWDVFEGLSH
LRR19   LQVLYLNHNYL NSLPPGVFSHLTA
LRR20   LRGLSLNSNRL TVLSHNDLPAN
LRR21   LEILDISRNQL LAPNPDVFVS
LRR22   LSVLDITHNKF ICECELSTFINWL
LRRCT   NHTNVTIAGPPADIYCVYPDSFSGVSLFSLSTEGCDEEE
        VLKSLK
TRNS    FSLFIVCTVTLTLFLMTIL
CYTOP   TVTKFRGFCFICYKTAQRLVFKDHPQGTEPDMYKYDAYL
        CFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVP
        GENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYA
        QGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQ
        YLRWPEDLQDVGWFLHKLSQQILK]KEKEKKKDNNIPLQ
        TVATIS
```

(5) The *TLR11* family

mTLR11

```
SIGNAL  MPRMERHQFCSVLLILILLTLVSLTLTGWA
LRRNT   WTIPDCIIADSLLFPNLSYYIPFCTSAPGLHLLASCSNV
        KNLNQTLKRVPR
        LxxLxLxxNxL
LRR1    NTEVLCLQGMV PTLPAKAFIRFHSLQ
LRR2    LLRLQLRTTSV TSRTFQGLDQ
LRR3    LQYLFFDHHAP CCLSLFLSPNCFESLRS
LRR4    LSSLSFQGYCL TYSQSIYLPTS
LRR5    LRHLTLRNSCL TKFQDLQRL
LRR6    FPDLLLSTSST PNIKPGAPF
LRR7    LETLDLSYNLQ LKQAGVRDLYGLT
LRR8    LHSLILDGTPL KALDLTDSGLLH
LRR9    LHFLSLVGTGI EKVPASLTGYSE
LRR10   LRALDLGKNQI QNILENGEIPGYKA
LRR11   LEFLSLHDNHL QTLPTRFLHTLPQ
LRR12   LQKLNLSMNKL GPILELPEGLFSTN
LRR13   LKVLDLSYNQL CDVPHGALSLLSQ
LRR14   LQELWLSGNNI SSLSNESLQGLRQ
LRR15   LRTLDLSWNQI KVLKPGWLSHLPA
```


Figure 7
Sequence alignment of LRR domains within the six families of TLRs. This panel continued in Figure 6 shows jfTLR3 (from LRR25 to CYTOP in the TLR3 family, hTLR4 and dTLR4 in the TLR4 family, hTLR5in the TLR5 family, and mtLR11 (from SIGNAL to LRR15) in the TLR11 family.

```
LRR16   LTTLNLLGTYL EYILGIQLQGPKM
LRR17   LRHLQLGSYPI LDIYPPWPPT
LRR18   LLSLEIQAESC IQFMIHSGQPFLF
LRR19   LENLTLETSIL LLKPDNITIHFPS
LRR20   LRRLTLRGYSF IFSTSQLQRFFPQQLPL
LRR21   LEHFFIWCENS YAVDLYLFGMPRLR
LRR22   VLELGYLNFFY ESSTMKLEMLLKEVPQ
LRR23   LQVLALSHLNL RNLSVSSFKSLQD
LRR24   LKLLLFNSERA LEMNSNLQEFIPQM
LRR25   PQYVYFSDVTF TCQCEASWLESWAT
LRRCT   RAPNTFVYGLEKSICIANASDYSKTLLFSFLATNCPHGT
        EFWGFLTS
TRANS   FILLLLLIILPLIS
CYTOP   CPKWSWLHHLWTLFHTCWWKLCGHRLRGQFNYDVFISYC
        EEDQAWVLEELVPVLEKAPPEGEGLRLCLPARDFGIGND
        RMESMIASMGKSRATLCVLTGQALASPWCNLELRLATYH
        LVARPGTTHLLLLFLEPLDRQRLHSYHRLSRWLQKEDYF
        DLSQGKVEWNSFCEQLKRRLSKAGQERD
```

mTLR12

```
SIGNAL  MGRYWLLPGLLLSLPLVTGWS
LRRNT   TSNCLVTEGSRLPLVSRYFTFCRHSKLSFLAACLSVSNL
        TQTLEVVPRT
        LxxLxLxxNxL
LRR1    VEGLCLGGTVS TLLPDAFSAFPG
LRR2    LKVLALSLHLT QLLPGALRGLGQ
LRR3    LQSLSFFDSPL RRSLFLPPDAFSDLIS
LRR4    LQRLHISGPCL DKKAGIRLPPG
LRR5    LQWLGVTLSCI QDVGELAGMFPDLVQGSSSRVSWT
LRR6    LQKLDLSSNWK LKMASPGSLQGLQ
LRR7    VEILDLTRTPL DAVWLKGLG
LRR8    LQKLDVLYAQT ATAELAAEAVAHFE
LRR9    LQGLIVKESKI GSISQEALASCHS
LRR10   LKTLGLSSTGL TKLPPGFLTAMPR
LRR11   LQRLELSGNQL QSAVLCMNETGDVSG
LRR12   LTTLDLSGNRL RILPPAAFSCLPH
LRR13   LRELLLRYNQL LSLEGYLFQELQQ
LRR14   LETLKLDGNPL LHLGKNWLAALPA
LRR15   LTTLSLLDTQI RMSPEPGFWGAKN
LRR16   LHTLSLKLPAL PAPAVLFLPMYLT
LRR17   SLELHIASGTT EHWTLSPAIFPS
LRR18   LETLTISGGGL KLKLGSQNASGVFPA
LRR19   LQKLSLLKNSL DAFCSQGTSNLFLWQLPK
LRR20   LQSLRVWGAGN SSRPCLITGLPS
LRR21   LRELKLASLQS ITQPRSVQLEELVGDLPQ
LRR22   LQALVLSSTGL KSLSAAAFQRLHS
LRR23   LQVLVLEYEKD LMLQDSLREYSPQM
LRR24   PHYIYILESNL ACHCANAWMEPWVK
LRRCT   RSTKTYIYIRDNRLCPGQDRLSARGSLPSFLWDHCPQTLELK
TRANS   LFLASSALVFMLIALPLL
CYTOP   QEARNSWIPYLQALFRVWLQGLRGKGDKGKRFLFDVFV
        SHCRQDQGWVIEELLPALEGFLPAGLGLRLCLPERDFE
        PGKDVVDNVVDSMLSSRTTLCVLSGQALCNPRCRLELR
        LATSLLLAAPSPPVLLLVFLEPISRHQLPGYHRLARLL
        RRGDYCLWPEEEERKSGFWTWLRSRLG
```

mTLR13

```
LRRNT   MSGLYRILVQLEQSPYVKTVPLNMRRDFFFLVVTWMPKTVK
        MNGSSFVPSLQLLLMLVGFSLPPVAETYGFNKCTQYEFD
        LxxLxLxxNxL
LRR1    IHHVLCIRKKI TNLTEAISDIPRY
LRR2    TTHLNLTHNEI QVLPPWSFTNLSA
LRR3    LVDLRLEWNSI WKIDEGAFRGLEN
LRR4    LTLLNLVENKI QSVNNSFEGLSS
LRR5    LKTLLLSHNQI THIHKDAFTPLIK
LRR6    LKYLSLSRNNI SDFSGILEAVQHLPC
LRR7    LERLDLTNNSI MYLDHSPRSLVS
LRR8    LTHLSFEGNKL RELNFSALSLPN
LRR9    LTNLSASRNGN KVIQNVYLKTLPQ
LRR10   LKSLNLSGTVI KLENLSAKHLQN
LRR11   LRAMDLSNWEL RHGHLDMKTVCHLLGNLPK
LRR12   LETLVFQKNVT NAEGIKQLAKCTR
LRR13   LLFLDLGQNSD LIYLNDSEFNALPS
LRR14   LQKLNLNKCQL SFINNRTWSSLQN
LRR15   LTSLDLSHNKF KSFPDFAFSPLKH
LRR16   LEFLSLSRNPI TELNNLAFSGLFA
LRR17   LKELNLAACWI VTIDRYSFTQFPN
LRR18   LEVLDLGDNNI RTLNHGTFRPLKK
LRR19   LQSLILSHNCL KILEPNSFSGLTN
LRR20   LRSLDLMYNSL SYFHEHLFSGLEK
LRR21   LLILKLGFNKI TYETTRTLQYPPFIKLKS
LRR22   LKQLNLEGQRH GIQVVPSNFFQGLGS
LRR23   LQELLLGKNPS VFLDHHQFDPLIN
LRR24   LTKLDISGTKD GDRSLYLNASLFQNLKR
LRR25   LKILRLENNNL ESLVPDMFSSLQS
LRR26   LQVFSLRFNNL KVINQSHLKNLKS
LRR27   LMFFDVYGNKL QCTCDNLWFKNWSM
LRRCT   NTEEVHIPFLRSYPCQQPGSQSLLIDFDDAMCNFDLGK
TRANS   VYFLCSFSMVLSTMVFSWF
CYTOP   STKMIASLWYGLYICRAWYLTKWHKTEKKFLYDAFVSF
        SATDEAWVYKELVPALEQGSQTTFKLCLHQRDFEPGID
        IFENIQNAINTSRKTLCVVSNHYLHSEWCRLEVQLASM
        KMFYEHKDVIILIFLEEIPNYKLSSYHRLRKLINKQTF
        ITWPDSVHQQPLFWARIRNALGKETVEKENTHLIVVE
```

tTLR21

```
SIGNAL  MWSLSFPLLSVTVLLCATRLVGG
LRRNT   YSFHNCIEFSK
        LxxLxLxxNxL
LRR1    GKTFKCIHRNQ YILGDIIKDLPHS
LRR2    TLDLTIAVNPV SHIPDRSFIHLRN
LRR3    LQQLRLDHNHL GVIDQFAFQGLHQ
LRR4    LKSLNLSFNYI PELSPSVFDGLHN
LRR5    LTFLSLTNNSL KRLPHGIFSHLFN
LRR6    LNTLIIKQNYL TNFSEIAMAVSSLKN
LRR7    LTLLDLCFNRL TSLSHSNVSLPES
LRR8    LNRLYLCRNNL STLGCEPSFLGS
LRR9    IEILDLSYNSE LPTKALEGVNLRR
LRR10   INYLRLRSTKV NIVEFIQNSDIH
LRR11   AGHVDFTSTHL NTEAKLVELCKLLKRKLSR
LRR12   ITKLTLVGNKI ETLTANTLAHCPN
LRR13   ITKTLDLSKVG QKKSDCLQFLKEQRQ
LRR14   ITTFIAEHNHY SSLPTCEDEDPFRQ
LRR15   LEELRYRYNRI LSVNSHAFHHTPN
LRR16   LKTLWLNINTI AFLHQKALSGLRQ
LRR17   LSTLRLDNNLL SDLFADTFEDLFN
LRR18   LNILNLRNNRI SVIFNNTFRNLKN
LRR19   LTTLDLGGNKI THFEPSGLCGLER
LRR20   LSKLYLDGNNL QTIDSSAYHIFQNT
LRR21   LTTLDLRQNMI HFEEDVNFSPFVNLTK
LRR22   LEDLKLDEQKP YGLHILPRTLFRGLYS
LRR23   LRSLYVKNNMI SYLAADVFRDLKH
LRR24   LNFLSLDNCCV GPTHLPAGIFKDLTN
LRR25   LTILTVENMGI QNLSTEVFGNISQ
LRR26   LKKIQLNHNVM QTFPVTVLQSLTK
LRR27   LQYLDIRNVPL SCTCENSLLRNWTV
LRRCT   NNQKVQMIYLYSLPCPHDPKVKFFNFDTSVCNIDLGQY
TRANS   LFFCTAPWIFLFTVWPLLYV
CYTOP   KLYWKIKYSYYVFRSWFSEQWRRLREKEENCKYDAFISY
        NSSDELWVMNELLPNLEGNGSSFKICLHHRDFEPGRYII
        DNIVSAVYSSRKTICVVSRNFLSSEWCSLEIQLASYRLF
        DEHRDVLLLVLLEPISERQLSSYHRMRKVMLKKTYLQWP
        GSECTNPPQAQGLFWSQLRRAIGTTSRIETEEKGTRVAN
        KEDADASDNHV
```


Figure 8
Sequence alignment of LRR domains within the six families of TLRs. This panel continued in Figure 7 shows mtTLR11 (from LRR16 to CYTOP), mTLR12, mTLR13, and tTLR21 in the TLR11 family.

tTLR22

```
SIGNAL  MGSRIKRSTFFFSPAFFLLSFS
LRRNT   QFVIPLKGFALKNACKIS
        LxxLxLxxNxL
LRR1    FNVAICSGNFV KLQDVPQDIPPT
LRR2    VKGFDLSTNKI SRIRTTDFKEFR
LRR3    LLEVLDLKNNI ISQVDEGAFINLRS
LRR4    LKKLNLNRNKL GKLDAGLFDGLQN
LRR5    LTELRVTRNRL KMVEPSALKSLVS
LRR6    LTFLDISNNKL CENIRPLFQLPN
LRR7    LSVLSMAGNNM RFFESGHLTNTSLQ
LRR8    LKSLDLSRNPI TVFRITADILPN
LRR9    LTWLNLRGTFK KRQVTWDVRTFLGN
LRR10   VSLLDISELRL SLRYMTSLLASVNSS
LRR11   LTVLRMDKMSC SLAELINISCSIPT
LRR12   MSALQLQNNKL GSINSSVFHLCTN
LRR13   VKDLDLQRNQI NSTDEGAFRSMKE
LRR14   LKVLTLSDNRL QSVPVATRNLPN
LRR15   LMKLDLSKNKI NALHCDDFANITK
LRR16   LRHLKLNANLI SALPSCVFKEVTK
LRR17   LEVLKLQNNSI SQLNTAFKMYLPN
LRR18   LKQLHLNSNKL VAINHGEFGGLRS
LRR19   LQNLSLHSNQI KKLGMGSFLGLKN
LRR20   LTDIHLQLNMI ETEQIAGGVFNDLIN
LRR21   LRRLDLSNNHI RYYNSRPLRNPPFLHLSL
LRR22   LETLYIPSQRR RGRSQLPSNILKGLSN
LRR23   LLEFTCRNSQL VWLPVDTFSYTPR
LRR24   LQRLDISSNEF QDLSPALFHPIKD
LRR25   VRSLYISRTRL GSLEFLKDARLGK
LRR26   LNFLQSKRNVF SVISEDVLESLPE
LRR27   LVYVDFQGNNF TCDCDNAGFLQWV
LRRCT   NNKTQVFDAYNFECNYPLELKSKKLLDLDTQSCTVNTD
TRANS   FICFISTTCTILLFMAMSF
CYTOP   TYHFLRWQLTYAYYFFLALLADKKRKNQQTPRQYDAFVS
        YNAHDEHWVLRNLLPKLEEEQGWALCLHHRDFEPGKPII
        ENITDAIYGSRKTICVISRRYLESEWCSREIQVASFRLF
        DEQKDVLILIFLEDIPTRQLSPFYRMKKMLKSKTYLSWP
        RAEGHPEVFWEKLRQALLSKDMLDLKPLARGISHKV
```

tTLR23

```
SIGNAL  MLTWSPPVLLCLLLHLRSSIS
LRRNT   FSLKNCTVHAGGASA
        LxxLxLxxNxL
LRR1    EVFVDCASREL LTVPDDVPRD
LRR2    ATTVKLSYNLL TQVKRNDVEHLTK
LRR3    VKFLDLQSNEI AHIDDGSFLHMRS
LRR4    LTKLRLSKNKL SELTAQLFQGLSN
LRR5    LTHLDLSSNII TFIHPSTFKDLPS
LRR6    LQTVVLDANRL KEMADIRPLLILPK
LRR7    LRNLTISSNLF TSFQSKDLLLDQPSS
LRR8    LRVLDVSYSIF EAFSISAATFPH
LRR9    LEMIDLSQSAF TWNITDRTSLQN
LRR10   ITRLFFSHTKV SSRQIQEILQNVAS
LRR11   LKHLRMNYIDE WIREGLLATVCRIPS
LRR12   LRVLELYLNKV PNLSAQVEPCSE
LRR13   LVQLDLSCSDV SEVPPGSFRLMKQ
LRR14   LRLLNLEVNLL TKVPEDIRNISS
LRR15   LQVLYLSDNLI TEVGCEEFSNTSA
LRR16   LVELYLDSNRI TSLQQCSFENLLK
LRR17   LRILDLNNNLL WKIEGVFSRGPAK
LRR18   LQLLDLSRNSV SVYDDGYFQSLGW
LRR19   LTHLDVSSDKV GRVTPGAFVGLHR
LRR20   LKSLHVSIPLD YECDFRGLKQLEN
LRR21   LTISITISDTR SPPKYSQALFHLKS
LRR22   LRSFHVACQGF HYGFPLDVPLSMLSSMTQ
LRR23   LEEFTADNIYL SAPDPETFRSNSR
LRR24   LRSLKISQTDL SDLDPEMFRPIPD
LRR25   LQSLDLSGTQI SSLEFLLQVDFSS
LRR26   LRDLRLCDNDI TSINHTLFQFLPS
LRR27   LTLLDLTNNPL TCDCLNAGFILWV
LRRCT   MDNNQTQVINGHQYSCSFPVAKKGTNLLDFEVQFCWIDVDF
        LC
TRANS   FLSSTCLVVLTLLTSF
CYTOP   TYHFLRWQLLYAFHLFLAFIYDNRNRKQQNPLPYDAFV
        SYNVEDELWVYEEMLPALEDQQGWKLCLHHRDFQPGKP
        IMENITDAIYNSRKTICVISRSYLQSEWCSREIQMASF
        RLFDEQKDVLILLFLEEIPAHQLSPYHRMRKLLKRQTY
        LSWTQAGRHQAGVFWQNVQRALESGDAPHDQVDPLTGPAEP
```

(6) The *TLR7* family

hTLR7

```
SIGNAL  MVFPMWTLKRQILILFNIILISKLLG
LRRNT   ARWFPKTLPCDVTLDVPKN
        LxxLxLxxNxL
LRR1    HVIVDCTDKHL TEIPGGIPTN
LRR2    TTNLTLTINHI PDISPASFHRLDH
LRR3    LVEIDFRCNCV PIPLGSKNNMCIKRLQIKPRSFSGLTY
LRR4    LKSLYLDGNQL LEIPQGLPPS
LRR5    LQLLSLEANNI FSIRKENLTELAN
LRR6    IEILYLGQNCY YRNPCYVSYSIEKDAFLNLTK
LRR7    LKVLSLKDNNV TAVPTVLPST
LRR8    LTELYLYNNMI AKIQEDDFNNLNQ
LRR9    LQILDLSGNCP RCYNAPFPCAPCKNNSPLQIPVNAFDALTE
LRR10   LKVLRLHSNSL QHVPPRWFKNINK
LRR11   LQELDLSQNFL AKEIGDAKFLHFLPS
LRR12   LIQLDLSFNFE LQVYRASMNLSQAFSSLKS
LRR13   LKILRIRGYVF KELKSFNLSPLHNLQN
LRR14   LEVLDLGTNFI KIANLSMFKQFKR
LRR15   LKVIDLSVNKI SPSGDSSEVGFCSNARTSVESYEPQVLEQLH
                    YFRYDKYARSCRFKNKEASFMSVNESCYKY
LRR16   GQTLDLSKNSI FFVKSSDFQHLSF
LRR17   LKCLNLSGNLI SQTLNGSEFQPLAE
LRR18   LRYLDFSNNRL DLLHSTAFEELHK
LRR19   LEVLDISSNSH YFQSEGITHMLNFTKNLKV
LRR20   LQKLMMNDNDI SSSTSRTMESES
LRR21   LRTLEFRGNHL DVLWREGDNRYLQLFKNLLK
LRR22   LEELDISKNSL SFLPSGVFDGMPPN
LRR23   LKNLSLAKNGL KSFSWKKLQCLKN
LRR24   LETLDLSHNQL TTVPERLSNCSRS
LRR25   LKNLILKNNQI RSLTKYFLQDAFQ
LRR26   LRYLDLSSNKI QMIQKTSFPENVLNN
LRR27   LKMLLLHHNRF LCTCDAVWFVWWVNH
LRRCT   TEVTIPYLATDVTCVGPGAHKGQSVISLDLYTCELDLTN
TRANS   LILFSLSISVSLFLMVMM
CYTOP   TASHLYFWDVWYIYHFCKAKIKGYQRLISPDCCYDAFI
        VYDTKDPAVTEWVLAELVAKLEDPREKHFNLCLEERDW
        LPGQPVLENLSQSIQLSKKTVFVMTDKYAKTENFKIAF
        YLSHQRLMDEKVDVIILIFLEKPFQKSKFLQLRKRLCG
        SSVLEWPTNPQAHPYFWQCLKNALATDNHVAYSQVFKETV
```

hTLR8

```
SIGNAL  MENMFLQSSMLTCIFLLISGSCELCA
LRRNT   EENFSRSYPCDEKKQND
        LxxLxLxxNxL
LRR1    SVIAECSNRRL QEVPQTVGKY
LRR2    VTELDLSDNFI THITNESFQGLQN
LRR3    LTKINLNHNPN VQHQNGNPGIQSNGLNITDGAFLNLKN
LRR4    LRELLLEDNQL PQIPSGLPES
LRR5    LTELSLIQNNI YNITKEGISRLIN
LRR6    LKNLYLAWNCY FNKVCEKTNIEDGVFETLTN
LRR7    LELLSLSFNSL SHVPPKLPSS
LRR8    LRKLFLSNTQI KYISEEDFKGLIN
LRR9    LTLLDLSGNCP RCFNAPFPCVPCDGGASINIDRFAFQNLTQ
LRR10   LRYLNLSSTSL RKINAAWFKNMPH
LRR11   LKVLDLEFNYL VGEIASGAFLTMLPR
LRR12   LEILDLSFNYI KGSYPQHINISRNFSKLLS
```


Figure 9
Sequence alignment of LRR domains within the six families of TLRs. This panel continued in Figure 8 shows tTLR22 and tTLR23 in the TLR11 family, and hTLR7 and hTLR8 (from SIGNAL to LRR12) in the TLR7 family.

```
LRR13   LRALHLRGYVF QELREDDFQPLMQLPN
LRR14   LSTINLGINFI KQIDFKLFQNFSN
LRR15   LEIIYLSENRI SPLVKDTRQSYANSSSFQRHIRKRRST
                    DFEFDPHSNFYHFTRPLIKPQCAAY
LRR16   GKALDLSLNSI FFIGPNQFENLPD
LRR17   IACLNLSANSN AQVLSGTEFSAIPH
LRR18   VKYLDLTNNRL DFDNASALTELSD        D543A
LRR19   LEVLDLSYNSH YFRIAGVTHHLEFIQNFTN
LRR20   LKVLNLSHNNI YTLTDKYNLESKS
LRR21   LVELVFSGNRL DILWNDDDNRYISIFKGLKN
LRR22   LTRLDLSLNRL KHIPNEAFLNLPAS
LRR23   LTELHINDNML KFFNWTLLQQFPR
LRR24   LELLDLRGNKL LFLTDSLSDFTSS
LRR25   LRTLLLSHNRI SHLPSGFLSEVSS
LRR26   LKHLDLSSNLL KTINKSALETKTTTK
LRR27   LSMLELHGNPF ECTCDIGDFRRWM
LRRCT   DEHLNVKIPRLVDVICASPGDQRGKSIVSLELTTCVSDVTA
TRANS   VILFFFTFFITTMVMLAALA
CYTOP   HHLFYWDVWFIYNVCLAKVKGYRSLSTSQTFYDAYISYD
        TKDASVTDWVINELRYHLEESRDKNVLLCLEERDWDPGL
        AIIDNLMQSINQSKKTVFVLTKKYAKSWNFKTAFYLALQ
        RLMDENMDVIIFILLEPVLQHSQYLRLRQRICKSSILQW
        PDNPKAEGLFWQTLRNVVLTENDSYNNMYVDSIKQY
```

hTLR9

```
SIGNAL  MGFCRSALHPLSLLVQAIMLAMTLA
LRRNT   LGTLPAFLPCELQP
        LxxLxLxxNxL
LRR1    HGLVNCNWLFL KSVPHFSMAAPRGN
LRR2    VTSLSLSSNRI HHLHDSDFAHLPS
LRR3    LRHLNLKWNCP PVGLSPMHFPCHMTIEPSTFLAVPT
LRR4    LEELNLSYNNI MTVPALPKS
LRR5    LISLSLSHTNI LMLDSASLAGLHA
LRR6    LRFLFMDGNCY YKNPCRQALEVAPGALLGLGN
LRR7    LTHLSLKYNNL TVVPRNLPSS
LRR8    LEYLLLSYNRI VKLAPEDLANLTA
LRR9    LRVLDVGGNCR RCDHAPNPCMECPRHFPQLHPDTFSHLSR
LRR10   LEGLVLKDSSL SWLNASWFRGLGN
LRR11   LRVLDLSENFL YKCITKTKAFQGLTQ
LRR12   LRKLNLSFNYQ KRVSFAHLSLAPSFGSLVA
LRR13   LKELDMHGIFF RSLDETTLRPLARLPM
LRR14   LQTLRLQMNFI NQAQLGIFRAFPG
LRR15   LRYVDLSDNRI SGASELTATMGEADGGEKVWLQPGDL
                    APAPVDTPSSEDFRPNCSTL
LRR16   NFTLDLSRNNL VTVQPEMFAQLSH
LRR17   LQCLRLSHNCI SQAVNGSQFLPLTG       L499P (in mouse)
LRR18   LQVLDLSHNKL DLYHEHSFTELPR
LRR19   LEALDLSYNSQ PFGMQGVGHNFSFVAHLRT
LRR20   LRHLSLAHNNI HSQVSQQLCSTS
LRR21   LRALDFSGNAL GHMWAEGDLYLHFFQGLSG
LRR22   LIWLDLSQNRL HTLLPQTLRNLPKS
LRR23   LQVLRLRDNYL AFFKWWSLHFLPK
LRR24   LEVLDLAGNQL KALTNGSLPAGTR
LRR25   LRRLDVSCNSI SFVAPGFFSKAKE
LRR26   LRELNLSANAL KTVDHSWFGPLASA
LRR27   LQILDVSANPL HCACGAAFMDFLL
LRRCT   EVQAAVPGLPSRVKCGSPGQLQGLSIFAQDLRLCLDEALSWD
TRANS   CFALSLLAVALGLGVPM
CYTOP   HHLCGWDLWYCFHLCLAWLPWRGRQSGRDEDALPYDAFV
        VFDKTQSAVADWVYNELRGQLEECRGRWALRLCLEERDW
        LPGKTLFENLWASVYGSRKTLFVLAHTDRVSGLLRASFL
        LAQQRLLEDRKDVVVLVILSPDGRRSRYVRLRQRLCRQS
        VLLWPHQPSGQRSFWAQLGMALTRDNHHFYNRNFCQGPTAE
```

(7) Others

jlTLRa

```
SIGNAL  MAGWPGMFVAAAVLLCLPHPGSC
LRRNT   VRGEAFFLGRQCSVV
        LxxLxLxxNxL
LRR1    GDVADCSRRGL  SAVPSGLPPS
LRR2    IAQLDLSHNRI  ESLLANDFSDVPL
LRR3    LRVLNLAFNRV  RDIHPGALAHTAL
LRR4    LQHLDLYHNEL  LEIPAEAVGNLRL
LRR5    LQVLNIAMNNY  TSFVLGGAFANLHS
LRR6    LRSLTIGTART  DVLNASDFTALQNVS
LRR7    VTHLNVHTGSPL MKFEPGVFAPFKM
LRR8    LQSFRMNFTVD  DDPVIFSKVLLDLNKTK
LRR9    VSEFQTDRVLL  NPVKNMSIDFFYGLEKCSL
LRR10   LRNLTLVAANF  TDQEITSLLKNVYLSQ
LRR11   ITSVEITNSSY  TDKNVVSFFGLKNVTKLSP
LRR12   LEKVTINQIFH  LNMTYPKFAINFTLFPS
LRR13   FSKLKISHTGM  NKVECFFMKLKF
LRR14   ITWLDFSSNLL  DEEGLWWTNCKYTIILPR
LRR15   TTELYISNNKF  TDLQIISSMVSLMPS
LRR16   ITLLDVGYNYI  TDIDDCSWPPT
LRR17   LETLILRNNDI  SKDSRICTSPQ
LRR18   LKVLDLSYTRM  EAVPYYILDDAKS
LRR19   LRELYLTGNNI  HYIQPEIQSSS
LRR20   LQVLHVDYNTL  GIITKGTFQLLPK
LRR21   IRALKLGNNLF  YCMCDLYWFRQTF
LRRCT   DKSLLVDWPQKYVCSYPENLAEKTLDCFNPSIVSCDKRIA
TRANS   IGLSVVITAVVVALVLGLGYYF
CYTOP   DAVWYIRMGWIWVGAKRRGYKHVTSGEAPPFEYNAFISYS
        HMDSDWVEGTLVPRLENSGSNLKLCVHERDFTPGEWVVDN
        IIRCIEGSSKTLFVLSTNFVKSEWCHYELYFAQHRIMEQH
        QDSLVLVLLESLPKNSLPNKFCRLRRLLNRKTYLEWPAEE
        SKQAIFWASLQAILQTTSSPTNPVT
```

cTLR15

```
SIGNAL  MRILIGSLYFYFISFLFSKVNG
LRRNT   FLTQRTSPVSSFPFYNYSYLNLSSVSQAQAPKT
        LxxLxLxxNxL
LRR1    ARALNFSYNAI EKITKRDFEGFHV
LRR2    LEVLDLSHNHI KDIEPGAFENLLS
LRR3    LVSVDLSFNDK NLLVSGLAPHLKLIPTSGASGPSQIYMYFQ
                    KSAEAALEPSAPAELLPHLEDPPNPGNVNP
                    RFRQRRTEENKTSPPAATLRPDLCGAPI
LRR4    NGLLDLSRTKL SNEELTAKLDADLCQAQLGT
LRR5    VLEFNISHSDL EMDLLSLFILFLPMKD
LRR6    IQSVDASYNRI TINNIDVEAICHFPPSN
LRR7    FSFLNISNNPI NSLETVCLPAS
LRR8    ITVIDLSFTNI STIPANFAKKLSK
LRR9    LERMYVQGNQL IYTVRPENPSATPRPPPGTVQ
LRR10   ISAISLVRNQA GTPIESLPES
LRR11   VKHLKVSNCSI VELPEWFANRMQE
LRR12   LLFLDLSSNRI SMLPDLPIS
LRR13   LQQLDISNSDI KIIPPRFKSLSN
LRR14   VTVFNIQNNKL TEMHPEYFPST
LRR15   LTTCDISKNKL KVLSLTKALEN
LRR16   LESLNVSGNLI TRLEPACQLPS
LRR17   LTNLDSSHNLI SELPDHLGQSLLM
LRR18   LKHFNLSGNKI SFLQRGSLPAS
LRR19   LEELDISDNAI TTIVQDTFGQLTS
LRR20   LSVLTVQGKHF FCNCDLYWFVNIY
LRRCT   IRNPHLQINGKDDLRCSFPPDRRGSLVKSSNLTLLHCSLG
        IQMAITAC
TRANS   MAILVVLVLTGLCW
CYTOP   RFDGLWYVRMGWYWCMAKRRQYKKRPENKPFDAFISYSEH
        DADWTKEHLLKKLETDGFKICYHERDFKPGHPVLGNIFYC
        IENSHKVLFVLPSFVNSCWCQYELYFAEHRVLDENQDSLI
        MVVLEDLPPDSVPQKFSKLRKLLKRKTYLKWSPEEHKQKI
        FWHQLAAVLKTTNEPLVRAENGPNEDVIEME
```


Figure 10
Sequence alignment of LRR domains within the six families of TLRs. This panel continued in Figure 9 shows hTLR8 (from LRR13 to CYTOP) and hTLR9 in the TLR7 family, and jlTLRa and cTLR15.

One or two horseshoe domains of LRRs within TLRs

The *TLR7* family (TLR7, TLR8, TLR9 and green puffer TLR) have 27 LRRs and an additional 58–73 residues at the end of LRR15 (Figures 9, and 10). Such a long region is also observed in chicken TLR15 (Figure 10). Gibbard *et al.*, [61] have considered two horseshoe domains of LRRs for human TLR8. That is, LRR15 has been separated into an LRR motif and 40 residues of undetermined structure. Most of the known LRR structures have a cap, which shields the hydrophobic core at the N- and C-terminii of LRRs. We suggest that these 40 residues function as the cap of the horseshoe structure, an intervening of hydrophobic core of LRR with a specific feature in TLRs. Thus, it can be concluded that the LRRs in vertebrate TLRs form one or two distinct horseshoe structures. Future structure determinations should resolve the question.

The *TLR1* family (TLR1, TLR2, TLR6, and TLR10) and the *TLR4* family share a common feature, the central part of the LRR domain has a more irregular motif compared with those at the N- and C-terminal parts. The LRR structure in the three families of *TLR1*, *TLR4* and *TLR7* might show a structural flexibility at the central part. Alternatively, the central part would play a key role in the function.

The LRR arc of TLRs is flat?

The LRR arc structures can be characterized by three parameters- the inner radius of the arc (R), the mean rotation angle about the central axis relating one β-strand to the next ($\bar{\varphi}$), and the tilt angle of the parallel β-strand direction per turn (θ_t). A 3D circle fitting method to calculate these geometrical parameters has been developed [55]. The TLR3-LRR arc yields R = 26.5–26.6Å, $\bar{\varphi}$ = 10.8–10.9° and θ_t = 24.5–26.7°. The TLR3-LRR belongs to "typical" type. This R value is comparable to 22–36 Å for the LRR arcs in Slit, FSHr, nogo receptor, decorin, and GPIbα with "typical" LRRs [8,55,58]. In contrast, the θ_t value is comparable to only those for Slit (-21°) and FSHr (-40°). Also the θ_t value corresponds to 19–40° for ribonuclease inhibitor and 15° for tropomodulin with "RI-like" LRRs. That is, the TLR3-LRR arc is nearly flat. This indicates that all other TLRs except for the *TLR7* family and TLR15 might adopt flat LRR arc.

Super-motif of LRRs in the TLR7 family

The present analysis reveals that the *TLR7* family consisting of TLR7, TLR8 and TLR9 and green puffer TLR contains the super-motif consisting of ***STT***. Such super motifs have been observed in various LRR proteins [8,11]. One of them is the SLRP family. The SLRP family forms five distinct subfamilies. Class I consists of biglycan, decorin, and

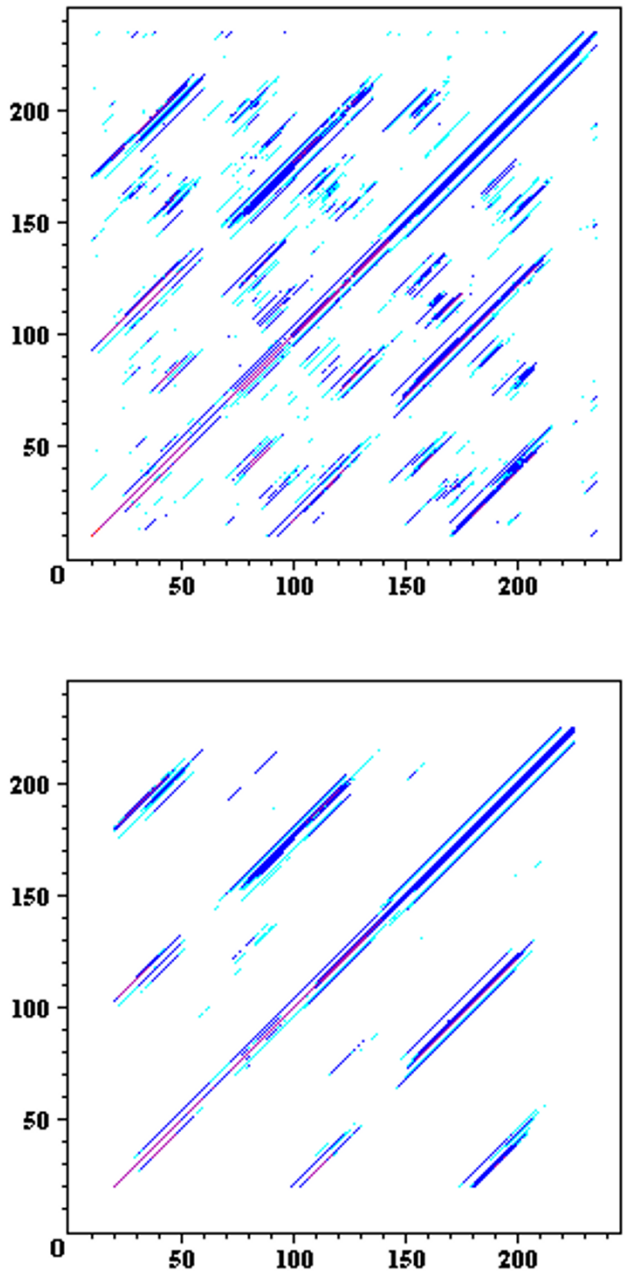

Figure 11
Super-repeat of LRRs in the *TLR7* family of TLR7, TLR8 and TLR9. Forty-two superimposed, cross-dot matrices from human TLR7 [Q9NYK1], mouse TLR7 [P58682], human TLR8 [Q9NR97], mouse TLR8 [P58682], human TLR9 [Q9NR96], mouse TLR9 [Q9EQU3], and green puffer TLR [Q4S0D3] with the widow size of 21 residues and the stringency of 10 (upper) and with the widow size of 41 residues and the stringency of 20 (lower). The summed scores for the 21 ((7 × 6)/2) comparisons are represented by color. The order of higher scores is red > purple > blue > light blue. Residue 46–291, 46–291, 44–288, 44–283, 40–285, 40–285, and 23–268 of human TLR7, mouse TLR7, human TLR8, mouse TLR8, human TLR9, and green puffer TLR, respectively, were used for the cross-dot matrices. The abscissa axis and the ordinate axis are residues number.

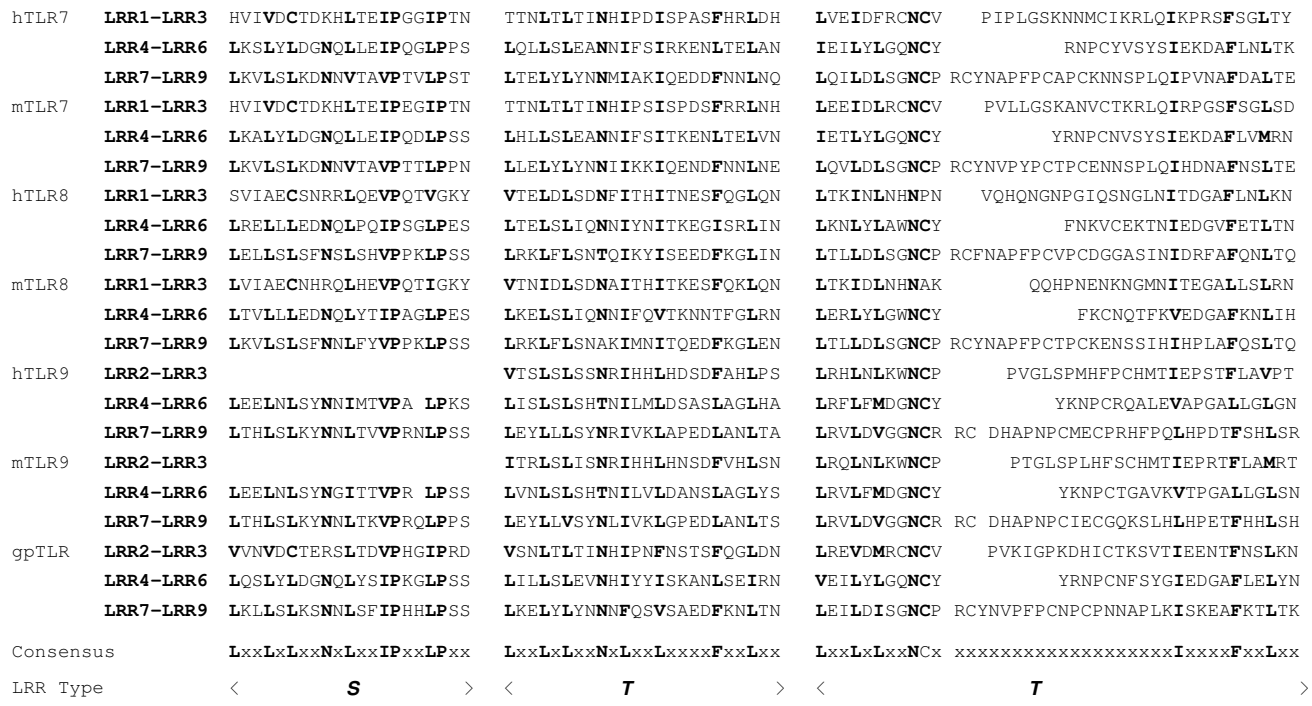

Figure 12
Sequence alignment of super-repeat of LRRs within TLR7, TLR and TLR9 from human and mouse and TLR from green puffer. human TLR7 [Q9NYK1]; mouse TLR7 [P58682]; human TLR8 [Q9NR97]; mouse TLR8 [P58682]; human TLR9 [Q9NR96]; mouse TLR9 [Q9EQU3]; green puffer TLR [Q4S0D3]. Abbreviations: h, human; m, mouse; gp, green puffer.

asporin. Class II has three subclasses: lumican plus fibromodulin (IIA), PRELP plus keratocan (IIB), and osteoadherin (IIC). Class III consists of epiphycan, osteoglycin and opticin. Class IV is more distantly related and consists of chondroadherin and nyctalopin. Class V consists of podocan. Their classes except for class IV contain the super-motif. Super-motifs, ***S*** and ***T***, similar to those in SLRP are also present in asporin-like proteins from human and mouse, mouse fibromodulin-like proteins, biglycan-like proteins from sea lamprey, oligodendrocyte-myelin and glycoprotein (OMGP), the FLRT family from human, mouse and Xenopus, and human ECM2 [8,62]. Furthermore, a preliminary analysis indicates that nephrocan, a novel member of the SLRP family [63], contains an ***STT*** motif. These observations suggest strongly that "bacterial" and "typical" LRRs evolved from a common precursor.

LRR variants in TLRs associated with diseases

A number of amino acid polymorphisms, which occur in LRRs, have been reported in TLRs. Arbour *et al.*, [64] first identified two mutations of human TLR4, D299G and T399I, which were associated with diminished airway responsiveness to inhaled LPS. Since then, these two mutations have been studied for their association with various infectious and inflammatory diseases; results regarding the effects of these mutations have been inconclusive [65-71]. D299G and T399I occur in LRR11 and LRR15, respectively (Figure 7). D299G is near the convex part, while T399I is located on the loop C-terminal to the convex part. Very recently, Ohara *et al.*, [72] reported that one mutation, T135A, was associated with poorly-differentiated gastric adenocarcinomas. T135A in LRR5 occur at position 9 in the HCS part (Figure 7). Such a mutation has been observed in many LRR proteins such as nyctalopin, keratocan, GPIbα, GPIbβ and GPIX, which are associated with human diseases [58]. Position 9 is generally occupied by Asn or Cy and sometimes by Thr or Ser, whose side chains form hydrogen bonds in the loop structure [58]. The T135A mutation may disrupt the hydrogen bond pattern in the loop.

Mouse TLR9 plays a role in defense against systemic mouse cytomegalovirus infection. Mice with the mutation, L499P, are highly susceptible to mouse cytomegalovirus infection and shows low levels of cytokine induction and natural killer activation on viral infection [73]. L499P is located at the short loop that connects the helical structure on the convex part (in LRR17) and the β-strand on concave part (in LRR18) (Figure 10). That is, L499P in

LRR18 occur at position 1 in the HCS part. The side chain of L499 is completely buried in the LRR arc. Such a mutation is also observed in trk-A and nyctalopin, which are associated with human diseases [58]. The mutation of D543A in human TLR8 abolishes the binding of CpG DNA [61]. D543A in LRR19 occur at position 1 in the VS part. Thus, D543A is located at the edge between the convex and the concave parts of the LRR arc. The Cys-to-Ala mutations in the VS part of LRR9 (C257A, C260A, C267A, and C270A) completely abolish signaling by TLR8 [61].

Hidaka *et al*., [74] detected one mutation, F303, in human TLR3 in one of three patients with influenza-associated encephalopathy. This was a loss-of-function mutation. F303S in LRR12 is located at position 4 in the HCS part. The side chain of F303 is completely buried in the LRR arc. Two mutations, H539E and N541A, resulted in the loss of TLR3 activation and ligand binding functions [75]. These two mutations occur in LRR21.

Conclusion

The new method of alignment proposed here rationalizes the difference in the repeat numbers of LRRs and their "phasing" within TLRs in different databases and for various species and isoforms. Moreover, the new method indicates that each of the six TLR families is characterized by their LRR motifs, their repeat numbers, and the motifs of cysteine clusters. The repeat number of LRRs is larger than those previously reported in databases. The central part in the LRR domains within the *TLR1* family and TLR4 has more irregular motifs compared with the N- and C-terminal parts. Moreover, the *TLR7* family contains a region with 58–73 residues in the central part of the LRR domain. The central parts are inferred to play a key role in the structure and/or function of their TLRs. The LRRs in TLRs form one or two horseshoe domains. The LRR arc of TLRs is also predicted to be nearly flat. Furthermore, the LRR supermotif in the TLR7 family suggests strongly that "bacterial" and "typical" LRRs evolved from a common precursor. The present analysis should stimulate and facilitate various experimental studies to understand the molecular mechanism of TLR-ligand interactions.

Methods

Known structures of LRR proteins

The structures of twenty-two different LRR proteins have been determined. They are ribonuclease inhibitor (RI) [2NBH, I1DJ, LA4Y, 1Z7X], GTPase-activating protein (RanGAP) [1YRG, 1K5D, 1K5G], tropomodulin (Tmod) [1IO0, 1PGV], S-phase kinase-associated protein 2 (Skp2) [1FQV], YopM [1G9U], four internalins, Inl-B [1D0B], Inl-H [1H6U], Inl-A [106T, 106V, 106S] and Inl-C [1XEU], spliceosomal U2A' protein [1A9N], mRNA export factor (TAP) [1FT8, 1F01], rab geranylgeranyltransferase α-subunit (RabGGTα) [1DCE, 1LTX], *Chlamydomonas* outer arm dynein light chain 1 (DLC-1) [1DS9], polygalacturonase-inhibiting protein (PGIP) [10GQ], nogo receptor/nogo-66 receptor (NgR) [10ZN, 1P9A], glycoprotein Ibα (GPIbα) [1M0Z, 1GWB, 1QYY, 1M10, 1SQ0, 1P8V, 100K, 1P9A, 1U0N], decorin [1XCD, 1XKU, 1XEC], biglycan [2FT3], Slit [1W8A], CD14 [1WWL], follicle-stimulating hormone receptor (FSHr) [1XWD], TLR3 [1ZIW, 2A0Z], and human lingo-1 [2ID5].

Amino acid sequences

The LRRs alignments within the TLR family were made for TLR1 from four species (human [Q15399, Q5FWG5, Q6FI64, Q32MK3], mouse [Q9EPQ1], pig [Q4LDR7, Q59HI9], Takifugu rubripes [Q5H727]); TLR2 from 17 species (human [O60603], mouse [Q9QUN7, Q8K3D9, Q811T5], pig [Q59HI8, Q5DX20, Q76L24], chicken [Q9DD78 (TLR2.1), Q9DGB6 (TLR2.2)], bovine [Q95LA9], rat [Q6YGU2], dog [Q689D1], rabbit [AAM50059], goat [ABI31733], horse [AAR08196], hamster [Q9R1F8], Cynomolgus monkey [Q95M53], domestic water buffalo [Q2PZH4], Nilgai [Q2V897], Takifugu rubripes [Q5H725], zebrafish [Q6TS42], Japanese flounder [Q76CT8]); TLR3 from 9 species (human [O15455, Q4VAL2, Q504W0], mouse [Q99MB1, Q3TM31, Q499F3], bovine [Q5TJ58, Q5TJ59], rat [Q7TNI8], buffalo [Q1G1A3], Rhesus macaque [Q3BBY1], Takifugu rubripes [Q5H721], zebrafish [Q6IWL5, Q32PW5], Japanese flounder [Q76CT7, Q76CT9]; TLR4 from 17 species (human [O00206, Q5VZI7, Q5VZI8, Q5VZI9], mouse [Q9QUK6, Q5RGT4, Q8K2T5], pig [Q68Y56, Q2TNK4, Q5F4K7, Q401C7], bovine [Q9GL65, Q6WCD5, Q8SQ55], rat [Q9QX05], hamster [Q9WV82], cat [P58727], lowland gorilla [Q8SPE8], horse [Q9MYW3], Pygmy chimpanzee [Q9TTN0], olive baboon [Q9TSP2], orangutan [Q8SPE9], Nilgai [Q2V898], American bison [Q3ZD70], dog [Q8SQH3], rabbit [AAM50060]; zebrafish [Q6NV08, Q6TS41(TLR4b)]; TLR5 from 8 species (human [O60602], pig [Q59HI7], mouse [Q9JLF7], bovine [Q2LDA0], chicken [Q4ZJ82], Japanese house mouse [Q1ZZX0], Takifugu rubripes [Q5H720, Q5H716(TLRS5)], rainbow trout [Q7ZT81]); TLR6 from 5 species (human [Q9Y2C9], mouse [Q9EPW9, Q7TPC5], rat [Q6P690], pig [Q59HI6, Q76L23], bovine [Q704V6, Q706D2]; TLR7 from 4 species (human [Q9NYK1], mouse [P58681, Q548J0], dog [Q2L4T3], Takifugu rubripes [Q5H719]); TLR8 from 4 species ((human [Q9NR97, Q495P4, Q495P6, Q495P7], mouse [P58682], pig [Q865R7], Takifugu rubripes [Q5H718]); TLR9 from 12 species (human [Q9NR96[, mouse [Q9EQU3], pig [Q5I2M3, Q865R8], bovine [Q5I2M5, Q866B2], dog [Q5I2M8], cat [Q5I2M7], Japanese flounder [Q2ABQ3], horse [Q2EEY0], sheep Q5I2M4], Ma's night monkey [Q56R09], Gilthead sea bream [Q3L273, Q3L274], Takifugu rubripes [Q5H717]]; TLR10 from two species (human [Q9BXR5, Q5FWG4, Q32MI7, Q32MI8],

pig [Q4LDR6, Q59HI5]); TLR11 from mouse [Q6R5P0, Q32ME8]; TLR12 from mouse [Q6QNU9]; TLR13 from mouse [Q6R5N8]; TLR14 from Takifugu rubripes [Q5H726] and zebrafish [XP_687315]; TLR15 from chicken [ABB71177], TLR21 from Takifugu rubripes [NP_001027751], TLR22 from Takifugu rubripes [Q5H723], TLR23 from Takifugu rubripes [AAW70378], and TLR from rainbow trout [Q6KCC7, Q4LBC9], Atlantic salmon [Q2A132], goldfish [Q801F9]), Japanese lamprey [Q33E92, Q33E93] and green puffer (Fragment) [Q4S0D3]).

The prediction of secondary structure, signal peptide and membrane-spanning region in protein

The protein secondary structure prediction by SSpro4.0 [13,76,77]and Proteus [12,78] were utilized for the determination and assignments of LRRs within TLRs. Signal peptide prediction was performed by SignalP 3.0 [79,80]. The prediction of membrane-spanning regions in proteins was performed by the TMHMM Program [81,82]. The PFAM program [83] was used to detect LRRs in TLRs.

Multiple sequence alignment, sequence similarity search and dot plot analysis

Multiple sequence alignments and sequence similarity searches were performed at Bioinformatic Center, Insitute for Chemical Research, Kyoto University [84]. Dot-matrix comparisons were performed using the Blosum90 scoring matrix. The program was made in house. Window sizes and stringencies are indicated in figure legends.

Abbreviations

Toll-like receptor: FTLR. Toll IL-receptor: FTIR. LPS: Liopolysacchride. LRR: ECD: Ectodomain. Leucine rich repeat. HCS: Highly conserved segment of LRR. VS: Variable segment of LRR. SLRP: Small leucine-rich repeat proteoglycans: FSHr, Follicle-stimulating hormone receptor.

Authors' contributions

NM (corresponding author) carried out the molecular genetic studies and wrote the manuscript. TT performed the dot plot analysis and contributed to the data analysis. EP performed the geometrical analysis of the known structure of TLR3. TM and MT participated in the sequence alignment. KY and YK conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Robert H. Kretsinger of University of Virginia for his valuable suggestion and comments. This work was supported in part Grant-in-Aid for Scientific Research from the Ministry Education, Science, Sports, and Culture of Japan (16310135) (to N. M.).

References

1. Kaisho T, Akira S: **Toll-like receptor function and signaling.** *J Allergy Clin Immunol* 2006, **117(5):**979-987.
2. Galiana-Arnoux D, Imler JL: **Toll-like receptors and innate antiviral immunity.** *Tissue Antigens* 2006, **67(4):**T267-276.
3. Kumar A, Yu FS: **Toll-like receptors and corneal innate immunity.** *Curr Mol Med* 2006, **6(3):**327-337.
4. Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A: **The evolution of vertebrate Toll-like receptors.** *Proc Natl Acad Sci USA* 2005, **102(27):**9577-9582.
5. Kobe B, Deisenhofer J: **The leucine-rich repeat: a versatile binding part.** *Trends Biochem Sci* 1994, **19(10):**415-421.
6. Kobe B, Deisenhofer J: **Proteins with leucine-rich repeats.** *Curr Opin Struct Bio* 1995, **5(3):**409-416.
7. Kobe B, Kajava AV: **The leucine-rich repeat as a protein recognition part.** *Curr Opin Struct Biol* 2001, **11(6):**725-732.
8. Matsushima N, Enkhbayar P, Kamiya M, Osaki M, Kretsinger RH: **Leucine-rich repeats (LRRs): structure, function, evolution and interaction with ligands.** *Drug Design Reviews* 2005, **2(4):**305-322.
9. Kajava AV: **Structural diversity of leucine-rich repeat proteins.** *J Mol Biol* 1998, **277(3):**519-527.
10. Ohyanagi T, Matsushima N: **Classification of tandem leucine-rich repeats within a great variety of proteins.** *FASEB J* 1997, **11:**A949.
11. Matsushima N, Ohyanagi T, Tanaka T, Kretsinger RH: **Super-motifs and evolution of tandem leucine-rich repeats within the small proteoglycans – biglycan, decorin, lumican, fibromodulin, PRELP, keratocan, osteoadherin, epiphycan, and osteoglycin.** *Proteins* 2000, **38(2):**210-225.
12. Kobe B, Deisenhofer J: **A structural basis of the interactions between leucine-rich repeats and protein ligands.** *Nature* 1995, **374(6518):**183-186.
13. Kobe B, Deisenhofer J: **Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A.** *J Mol Biol* 1996, **264(5):**1028-1043.
14. Papageorgiou AC, Shapiro R, Acharya KR: **Molecular recognition of human angiogenin by placental ribonuclease inhibitor-an X-ray crystallographic study at 2.0 A resolution.** *EMBO J* 1997, **16(17):**5162-5177.
15. Hillig RC, Renault L, Vetter IR, Drell T 4th, Wittinghofer A, Becker J: **The crystal structure of rna1p: a new fold for a GTPase-activating protein.** *Mol Cell* 1999, **3(6):**781-791.
16. Seewald MJ, Korner C, Wittinghofer A, Vetter IR: **RanGAP mediates GTP hydrolysis without an arginine finger.** *Nature* 2002, **415(6872):**662-666.
17. Krieger I, Kostyukova A, Yamashita A, Nitanai Y, Maeda Y: **Crystal structure of the C-terminal half of tropomodulin and structural basis of actin filament pointed-end capping.** *Biophys J* 2002, **83(5):**2716-2725.
18. Lu S, Symersky J, Li S, Carson M, Chen L, Meehan E, Luo M: **Structural genomics of Caenorhabditis elegans: crystal structure of the tropomodulin C-terminal domain.** *Proteins* 2004, **56(2):**384-386.
19. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP: **Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex.** *Nature* 2000, **408(6810):**381-386.
20. Hao B, Zheng N, Schulman BA, Wu G, Miller JJ, Pagano M, Pavletich NP: **Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase.** *Mol Cell* 2005, **20(1):**9-19.
21. Evdokimov AG, Anderson DE, Routzahn KM, Waugh DS: **Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit.** *J Mol Biol* 2001, **312(4):**807-821.
22. Marino M, Braun L, Cossart P, Ghosh P: **Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen L. monocytogenes.** *Mol Cell* 1999, **4(6):**1063-1072.
23. Schubert WD, Gobel G, Diepholz M, Darji A, Kloer D, Hain T, Chakraborty T, Wehland J, Domann E, Heinz DW: **Internalins from the human pathogen Listeria monocytogenes combine three distinct folds into a contiguous internalin domain.** *J Mol Biol* 2001, **312(4):**783-794.
24. Marino M, Banerjee M, Jonquieres R, Cossart P, Ghosh P: **GW domains of the Listeria monocytogenes invasion protein InlB are SH3-like and mediate binding to host ligands.** *EMBO J* 2002, **21(21):**5623-5634.
25. Marino M, Banerjee M, Copp J, Dramsi S, Chapman T, van der Geer P, Cossart P, Ghosh P: **Characterization of the calcium-binding**

sites of Listeria monocytogenes InlB. *Biochem Biophys Res Commun* 2004, **316(2):**379-386.
26. Schubert WD, Urbanke C, Ziehm T, Beier V, Machner MP, Domann E, Wehland J, Chakraborty T, Heinz DW: **Structure of internalin, a major invasion protein of Listeria monocytogenes, in complex with its human receptor E-cadherin.** *Cell* 2002, **111(6):**825-836.
27. Price SR, Evans PR, Nagai K: **Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA.** *Nature* 1998, **394(6694):**645-650.
28. Liker E, Fernandez E, Izaurralde E, Conti E: **The structure of the mRNA export factor TAP reveals a cis arrangement of a noncanonical RNP domain and an LRR domain.** *EMBO J* 2000, **19(21):**5587-5598.
29. Ho DN, Coburn GA, Kang Y, Cullen BR, Georgiadis MM: **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* 2002, **99(4):**1888-1893.
30. Zhang H, Seabra MC, Deisenhofer J: **Crystal structure of Rab geranylgeranyltransferase at 2.0 A resolution.** *Structure* 2000, **8(3):**241-251.
31. Pylypenko O, Rak A, Reents R, Niculae A, Sidorovitch V, Cioaca MD, Bessolitsyna E, Thoma NH, Waldmann H, Schlichting I, Goody RS, Alexandrov K: **Structure of Rab escort protein-1 in complex with Rab geranylgeranyltransferase.** *Mol Cell* 2003, **11(2):**483-494.
32. Wu H, Maciejewski MW, Marintchev A, Benashski SE, Mullen GP, King SM: **Solution structure of a dynein motor domain associated light chain.** *Nat Struct Biol* 2000, **7(7):**575-579.
33. Wu H, Blackledge M, Maciejewski MW, Mullen GP, King SM: **Relaxation-based structure refinement and backbone molecular dynamics of the dynein motor domain-associated light chain.** *Biochemistry* 2003, **42(1):**57-71.
34. Di Matteo A, Federici L, Mattei B, Salvi G, Johnson KA, Savino C, De Lorenzo G, Tsernoglou D, Cervone F: **The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense.** *Proc Natl Acad Sci USA* 2003, **100(17):**10124-10128.
35. Barton WA, Liu BP, Tzvetkova D, Jeffrey PD, Fournier AE, Sah D, Cate R, Strittmatter SM, Nikolov DB: **Structure and axon outgrowth inhibitor binding of the Nogo-66 receptor and related proteins.** *EMBO J* 2003, **22(13):**3291-302.
36. He XL, Bazan JF, McDermott G, Park JB, Wang K, Tessier-Lavigne M, He Z, Garcia KC: **Structure of the Nogo receptor ectodomain: a recognition module implicated in myelin inhibition.** *Neuron* 2003, **38(2):**177-185.
37. Huizinga EG, Tsuji S, Romijn RA, Schiphorst ME, de Groot PG, Sixma JJ, Gros P: **Structures of glycoprotein Ibalpha and its complex with von Willebrand factor A1 domain.** *Science* 2002, **297(5584):**1176-1179.
38. Uff S, Clemetson JM, Harrison T, Clemetson KJ, Emsley J: **Crystal structure of the platelet glycoprotein Ib(alpha) N-terminal domain reveals an unmasking mechanism for receptor activation.** *J Biol Chem* 2002, **277(38):**35657-35663.
39. Dumas JJ, Kumar R, Seehra J, Somers WS, Mosyak L: **Crystal structure of the pIbalpha-thrombin complex essential for platelet aggregation.** *Science* 2003, **301(5630):**222-226.
40. Celikel R, McClintock RA, Roberts JR, Mendolicchio GL, Ware J, Varughese KI, Ruggeri ZM: **Modulation of alpha-thrombin function by distinct interactions with platelet glycoprotein Ibalpha.** *Science* 2003, **301(5630):**218-221.
41. Dumas JJ, Kumar R, McDonagh T, Sullivan F, Stahl ML, Somers WS, Mosyak L: **Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Ibalpha complex reveals conformation differences with a complex bearing von Willebrand disease mutations.** *J Biol Chem* 2004, **279(22):**23327-23334.
42. Varughese KI, Ruggeri ZM, Celikel R: **Platinum-induced space-group transformation in crystals of the platelet glycoprotein Ib alpha N-terminal domain.** *Acta Crystallogr D Biol Crystallogr* 2004, **60(3):**405-411.
43. Fukuda K, Doggett T, Laurenzi IJ, Liddington RC, Diacovo TG: **The snake venom protein botrocetin acts as a biological brace to promote dysfunctional platelet aggregation.** *Nat Struct Mol Biol* 2005, **2(2):**152-159.
44. Scott PG, McEwan PA, Dodd CM, Bergmann EM, Bishop PN, Bella J: **Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan.** *Proc Natl Acad Sci USA* 2004, **101(44):**15633-15638.
45. Scott PG, Dodd CM, Bergmann EM, Sheehan JK, Bishop PN: **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* 2006, **281(19):**13324-13332.
46. Howitt JA, Clout NJ, Hohenester E: **Binding site for Robo receptors revealed by dissection of the leucine-rich repeat region of Slit.** *EMBO J* 2004, **23(22):**4406-4412.
47. Fan QR, Hendrickson WA: **Structure of human follicle-stimulating hormone in complex with its receptor.** *Nature* 2005, **433(7023):**269-277.
48. Kim JI, Lee CJ, Jin MS, Lee CH, Paik SG, Lee H, Lee JO: **Crystal structure of CD14 and its implications for lipopolysaccharide signaling.** *J Biol Chem* 2005, **280(12):**11347-11351.
49. Bell JK, Botos I, Hall PR, Askins J, Shiloach J, Segal DM, Davies DR: **The molecular structure of the Toll-like receptor 3 ligand-binding domain.** *Proc Natl Acad Sci USA* 2005, **102(31):**10976-10980.
50. Choe J, Kelker MS, Wilson IA: **Crystal structure of human toll-like receptor 3 (TLR3) ectodomain.** *Science* 2005, **309(5734):**581-585.
51. Mosyak L, Wood A, Dwyer B, Buddha M, Johnson M, Aulabaugh A, Zhong X, Presman E, Benard S, Kelleher K, Wilhelm J, Stahl ML, Kriz R, Gao Y, Cao Z, Ling HP, Pangalos MN, Walsh FS, Somers WS: **The structure of the Lingo-1 ectodomain, a module implicated in CNS repair inhibition.** *J Biol Chem* 2006, **281(47):**36378-36390.
52. Montgomerie S, Sundararaj S, Gallin WJ, Wishart DS: **Improving the accuracy of protein secondary structure prediction using structural alignment.** *BMC Bioinformatics* 2006, **7:**301-313.
53. Pollastri G, McLysaght A: **Porter: a new, accurate server for protein secondary structure prediction.** *Bioinformatics* 2005, **21(8):**1719-1720.
54. Kneller DG, Cohen FE, Langridge R: **Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network.** *J Mol Biol* 1990, **214:**171-182.
55. Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, Matsushima N: **Structural principles of leucine-rich repeat (LRR) proteins.** *Proteins* 2004, **54(3):**394-403.
56. Asahina Y, Yoshioka N, Kano R, Moritomo T, Hasegawa A: **Full-length cDNA cloning of Toll-like receptor 4 in dogs and cats.** *Vet Immunol Immunopathol* 2003, **96(3–4):**159-167.
57. Hirono I, Takami M, Miyata M, Miyazaki T, Han HJ, Takano T, Endo M, Aoki T: **Characterization of gene structure and expression of two toll-like receptors from Japanese flounder, Paralichthys olivaceus.** *Immunogenetics* 2004, **56(1):**38-46.
58. Matsushima N, Tachi N, Kuroki Y, Enkhbayar P, Osaki M, Kamiya M, Kretsinger RH: **Structural analysis of leucine-rich-repeat variants in proteins associated with human diseases.** *Cell Mol Life Sci* 2005, **62(23):**2771-2791.
59. Nishitani C, Mitsuzawa H, Sano H, Shimizu T, Matsushima N, Kuroki Y: **Toll-like receptor 4 region Glu^{24}-Lys^{47} is a site for MD-2 binding: Importance of Cys^{29} and Cys^{40}.** *J Biol Chem* 2006, **281(50):**38322-38329.
60. Bell JK, Mullen GE, Leifer CA, Mazzoni A, Davies DR, Segal DM: **Leucine-rich repeats and pathogen recognition in Toll-like receptors.** *Trends Immunol* 2003, **24(10):**528-533.
61. Gibbard RJ, Morley PJ, Gay NJ: **Conserved features in the extracellular domain of human toll-like receptor 8 are essential for pH dependent signaling.** *J Biol Chem* 2006, **281(37):**27503-27511.
62. McEwan PA, Scott PG, Bishop PN, Bella J: **Structural correlations in the family of small Leucine-rich repeat proteins and proteoglycans.** *J Struct Biol* 2006, **155(2):**294-305.
63. Mochida Y, Parisuthiman D, Kaku M, Hanai JI, Sukhatme VP, Yamauchi M: **Nephrocan, a novel member of the small leucine-rich repeat protein family, is an inhibitor of transforming growth factor-beta signaling.** *J Biol Chem* 2006 in press.
64. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA: **TLR4 mutations are associated with endotoxin hyporesponsiveness in humans.** *Nat Genet* 2000, **25:**187-191.
65. Vogel SN, Awomoyi AA, Rallabhandi P, Medvedev AE: **Mutations in TLR4 signaling that lead to increased susceptibility to infection in humans: an overview.** *J Endotoxin Res* 2005, **11(6):**333-339.
66. Kang SS, Kauls LS, Gaspari AA: **Toll-like receptors: applications to dermatologic disease.** *J Am Acad Dermatol* 2006, **54(6):**951-983.
67. Pandey S, Agrawal DK: **Immunobiology of Toll-like receptors: Emerging trends.** *Immunol Cell Biol* 2006, **84(4):**333-341.
68. Kumpf O, Hamann L, Schlag PM, Schumann RR: **Pre- and postoperative cytokine release after in vitro whole blood lipopolysaccharide stimulation and frequent toll-like receptor 4 polymorphisms.** *Shock* 2006, **25(2):**123-128.
69. Faber J, Meyer CU, Gemmer C, Russo A, Finn A, Murdoch C, Zenz W, Mannhalter C, Zabel BU, Schmitt HJ, Habermehl P, Zepp F, Knuf M: **Human toll-like receptor 4 mutations are associated with susceptibility to invasive meningococcal disease in infancy.** *Pediatr Infect Dis J* 2006, **25(1):**80-81.

70. Mockenhaupt FP, Cramer JP, Hamann L, Stegemann MS, Eckert J, Oh NR, Otchwemah RN, Dietz E, Ehrhardt S, Schroder NW, Bienzle U, Schumann RR: **Toll-like receptor (TLR) polymorphisms in African children: Common TLR-4 variants predispose to severe malaria.** *Proc Natl Acad Sci USA* 2006, **103(1):**177-182.
71. Montes AH, Asensi V, Alvarez V, Valle E, Ocana MG, Meana A, Carton JA, Paz J, Fierer J, Celada A: **The Toll-like receptor 4 (Asp299Gly) polymorphism is a risk factor for Gram-negative and haematogenous osteomyelitis.** *Clin Exp Immuno* 2006, **143(3):**404-413.
72. Ohara T, Morishita T, Suzuki H, Hibi T: **Heterozygous Thr 135 Ala polymorphism at leucine-rich repeat (LRR) in genomic DNA of toll-like receptor 4 in patients with poorly-differentiated gastric adenocarcinomas.** *Int J Mol Med* 2006, **18(1):**59-63.
73. Tabeta K, Georgel P, Janssen E, Du X, Hoebe K, Crozat K, Mudd S, Shamel L, Sovath S, Goode J, Alexopoulou L, Flavell RA, Beutler B: **Toll-like receptors 9 and 3 as essentialcomponents of innate immune defense against mouse cytomegalovirus infection.** *Proc Natl Acad Sci USA* 2004, **101(10):**3516-3521.
74. Hidaka F, Matsuo S, Muta T, Takeshige K, Mizukami T, Nunoi H: **A missense mutation of the Toll-like receptor 3 gene in a patient with influenza-associated encephalopathy.** *Clin Immunol* 2006, **119(2):**188-194.
75. Bell JK, Askins J, Hall PR, Davies DR, Segal DM: **The dsRNA binding site of human Toll-like receptor 3.** *Proc Natl Acad Sci USA* 2006, **103(23):**8792-8797.
76. Cheng J, Randall AZ, Sweredoski MJ, Baldi P: **SCRATCH: a protein structure and structural feature prediction server.** *Nucleic Acids Res* 2005, **33:**W72-76.
77. [http://contact.ics.uci.edu/sspro4.html].
78. [http://129.128.185.184/proteus/#].
79. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 30.** *J Mol Biol* 2004, **340(4):**783-795.
80. [http://www.cbs.dtu.dk/services/SignalP].
81. Moller S, Croning MD, Apweiler R: **Evaluation of methods for the prediction of membrane spanning regions.** *Bioinformatics* 2001, **17(7):**646-653.
82. [http://www.cbs.dtu.dk/services/TMHMM/].
83. [http://pfam.janelia.org/].
84. [http://www.genome.ad.jp/].