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

Norio Matsushima*1, Takanori Tanaka2, Purevjav Enkhbayar3, Tomoko Mikami1,4, Masae Taga1,4, Keiko Yamada1 and Yoshio Kuroki5

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

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 LxxLxLxxNxxLxxLxxxF/LxxLxx ("T") and sometimes short motifs including LxxLxLxxNxxLxxLPxx(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.

Published: 21 May 2007
Received: 24 October 2006
Accepted: 21 May 2007

This article is available from: http://www.biomedcentral.com/1471-2164/8/124
© 2007 Matsushima et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
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 cytoplasmic 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 twelve residue stretch, LxxLxLxxCxxL, in which “L” is Leu, Ile, Val, or Phe, “N” is Asn, Thr, Ser, or Cys, and “C” is Cys, Ser or Asn [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 LRR known structures reveal 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.

---

**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.
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(\text{L/I/M})xL, LxxLx-Lxx(R/K/E)xL, LxxLxLxx(N/S/T/A)xL, and LxxLxLxxxxxx in which residues in boldface are irregular. Also there are rare examples, xVxxLxxNxL and PsxxLxxNxL in follicle-stimulating hormone receptor (FSHr), and LxxLxGxxS/Pxl in Inl-C and DLEC1 (Table 1). Furthermore, an irregular LRR with (L/x)xx(L/A)xCxx(L/R)xLxxVPxxLxx, 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, Cynomolagus 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 LxxLxLxx(S/T/N)xL and LxxLxLxx(Q/D)xL or LxxLxLxxLxx, respectively. The VS part is "typical". Both canonical and irregular LRRs are also seen in LRR9 and LRR10. Furthermore, the HCS part of LRR4 shows LxxLxLxxNxL 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-strand 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 LEEEIDASDL 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 prediction 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 Step 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, PTL[N/F/V/L]TL(N/Q)H[1/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, C[GK][AV]YLDHNSE, 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, LxxLxLxxNxLxxLxxxFxxLxx, occurs most frequently followed by shorter motifs including LxxLxLxxXP(x)L.Pxx ("bacterial"; "S") with 19–21 residues. Moreover, all of the C-terminal LRR consists of LxxLxLxxNP(F/L)xCxCxxxx(F/L)xLxxxx. 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-
| Protein | Nb | Position | Amino acid sequence | Irregular LRRs | PDB |
|---------|----|----------|---------------------|----------------|-----|
| TLR3    | 25 | LRR1     | HEVADCSHLKL         | TQVPDDLPTN     | 1ZIW|
|         |    | LRR9     | LFGFLFNNWQL         | GPLSTEKLCELENTS | /2A0Z|
|         |    | LRR13    | VRISILNLKRDFFSFWLKC |                 |     |
|         |    | LRR15    | LKYLSTNNSFT         | SLRTLNETFVSLAHSP |     |
|         |    | LRR18    | IFEIYLSYNYK         | LQLTRNSFALYPS  |     |
|         |    | LRR19    | LQRLMLRVRVAL        | KNVDSSPSFPQLRN |     |
| CD14    | 11 | LRR1     | AADVLYGGRG          | SLEYLLKRVTDEADGQFTDIKLS | IWWL|
|         |    | LRR2     | LKRLTVRAAR           | PSRLGALRLVGLSG |     |
|         |    | LRR3     | LQETLENV           | TGTAPPPLLEATGPD |     |
|         |    | LRR4     | LNILNRNVS          | ATRAWALAEQQWKLPG |     |
|         |    | LRR5     | LKVLSAQ           | LNFSEQVRVFPA  |     |
|         |    | LRR7     | LQVLARNAGM         | ETPSGVCALSAAAVQ |     |
| Slit    | 6  | LRR1     | GTTVCDCTGRLGL      | KEIPDIPLH      | IWWA|
| Lingo-1 | 14 | LRR8     | LIVLRLRH            | NAIRDSFKRLYR   | 2ID5|
|         |    | LRR9     | LKVLEISHWPY         | LDTMTFNCPNLGYLN |     |
| Decorin | 12 | LRR1     | LRVVQCSDDLGL        | EKVPKDFLPPD    | 1XCD|
| Biglycan| 12 | LRR1     | LRVVQCSDDLGL        | KAVPKESPD      | 2FT3|
| FSHr    | 10 | LRR6     | LQKVLLDDQDN         | INIHTIERNSFVGLSF | 1XWD|
|         |    | LRR7     | SVILWNLKNGI         | QEIHNCAFPNTQ   |     |
|         |    | LRR9     | PVILDISIRTRI        | HSLPSQGLEN    |     |
| Inl-A   | 16 | LRR1     | VTTLDQARLG           | KSIDGLEYLN    | 1O6V|
| Inl-C   | 7  | LRR1     | VMFNIDNNSN         | QSLAMQGFDTN   | 1XEU|
|         |    | LRR7     | VNWIDLTGQKC         | VNEPKVYQPEL   |     |
| U2snRNP   | 5 | LRR1     | DRERLRGYKY          | PVIRNLGATLDPQ | 1A9N|
| RabGGTα | 5  | LRR1     | VRVLHAKDL           | TVLCHLEQLL    | 1DCE|
| DLC1    | 6  | LRR1     | KVELHGMPPI          | EKMDATLSTLKA  | 1D9S|
| TAP     | 4  | LRR1     | QQLALDKGLRS         | DPDLVQAQIDVVLNRSCMAATLRL2 | 1F8T|
|         |    |          |                     | EEINIPE        |     |
| PGIP    | 10 | LRR1     | VNNLDSLGLN          | PKPYIPSSLANLPLYL | 1OGQ|
| YoPM    | 16 | LRR1     | AHEELNNGL          | SSLPELPFH      | 1G9U|
|         |    | LRR16    | VEDLRMNSEV         | DIFYFAHETTCHEDDDVE |     |
| hRI     | 16 | LRR1     | LEQLVLYYWW         | SEEMEDRLQALEKDKP | 1Z7X|
| pRI     | 17 | LRR1     | WNILDHCEQ           | SDARWTELPQQ   | 2BNH|
| RanGAP  | 10 | LRR2     | LEIAEFSDLFT         | GRVKDEIPALLLQALCCP | 1YRG|
| cTmod   | 5  | LRR1     | LEEVNLNNIMN         | IPVPTKCAEALKTNTY | 1I00|
|         |    | LRR2     | VKKFSVGTRS         | NDPVAFALAEMLKYNNT |     |
|         |    | LRR4     | LIEIRIDNQSQ         | PLGINYMEIANMLKNTT |     |
|         |    | LRR5     | LLKFGYHTQQ         | PRLASNAMNNNDLVRK |     |
| ceTmod  | 5  | LRR1     | LKEVNIENMKR         | VSKERKSLIAAACSNSKH | 1PGV|
|         |    | LRR4     | IVFKNADNQRO         | SVLGQVQEMDMMMMAENES |     |
|         |    | LRR5     | LLVRGSFAASM         | EARHRVSALNERYRVL |     |
| Skp2    | 11 | LRR1     | WQTLDDLGKKNL       | HPDYTVRLLSOG  | 1FQV|
|         |    | LRR4     | LQNLWLFRLR         | SDPVINTLAKSN   | 2ASS|
|         |    | LRR7     | ITQLNLSGYRK         | NNLQKSDLSTLVRCCPN |     |
|         |    | LRR10    | KTLQKVFGVLP         | DGTLQKLLKEA   |     |

*FShr, follic-stimulating hormone receptor; Inl-A, internalin A; Inl-C, internalin C; U2snRNP, 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; Nb: 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

| Familya | Protein | Speciesb | Nc | LRRNTd | LRRCTd | Nb* |
|---------|---------|----------|----|--------|--------|-----|
| TLR1    | TLR1    | h, m, p  | 20(8–9) | No | CxCxs2C3x2C | 1 |
|         |         | t        | 21 | CxCxs3C5x2C | CxCxs5C3x2C | 1 |
| TLR2    | h, m, p, r, d, ra, g, ho, dwb, ha, cm, n | 20(14–18) | Cx5C | CxCxs5C3x2C | 1 |
| TLR2.1  | c       | 20(13)  | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR2.2  | c       | 20(13)  | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR2    | t       | 20      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR2    | jf      | 19      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR2    | z       | 21      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR6    | h, m, r, p, b | 20(13–14) | No | CxCxs5C3x2C | 1 |
| TLR10   | h, p    | 20(12)  | No | CxCxs5C3x2C | 1 |
| TLR14   | t, z    | 21      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR3    | TLR3    | h, m, b, r, bu, rm, z | 25(22–24) | Cx5C | CxCxs5C3x2C | 1 |
|         |         | t        | 25 | Cx5C | CxCxs5C3x2C | 1 |
|         |         | jf      | 27 | Cx5C | CxCxs5C3x2C | 1 |
| TLR    | as      | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         | rt      | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         | go      | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLRII   | rt      | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR4    | TLR4    | h, lg, pc, ob, or | 23(21) | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         |         | m, r, ch, ho, b, ab, n | 23(18–19) | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         |         | ha, ra  | 23 | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR4b   | z       | 23      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR4    | h [QSVZ17] | 16 | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         | d       | 16      | No | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR5    | TLR5    | h, m, r, p, b, jhm | 22(15–16) | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         |         | t       | 22 | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         |         | rt      | 22 | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR5    | t       | 23      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR11   | TLR11   | M       | 25(11) | Cx1CxC1x2C | CxCxs5C3x2C | 1 |
| TLR12   | m       | 24(17)  | Cx1CxC1x2C | CxCxs5C3x2C | 1 |
| TLR13   | m       | 27(21)  | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR21   | t       | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR22   | t       | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR23   | t       | 27      | Cx1CxCx3C | CxCxs5C3x2C | 1 |
| TLR7    | TLR7    | H, m, d  | 27(27–28) | Cx1CxCx3C | CxCxs5C3x2C | 2 |
|         |         | t       | 27 | Cx1CxCx3C | CxCxs5C3x2C | 2 |
| TLR8    | h, m, p  | 27(24–26) | Cx1CxCx3C | CxCxs5C3x2C | 2 |
|         |         | t       | 27 | Cx1CxCx3C | CxCxs5C3x2C | 2 |
| TLR9    | h, m, p, b, d, ca, ho, s, mmn | 27(26) | Cx1CxCx3C | CxCxs5C3x2C | 2 |
|         |         | t, jf, gsb | 27 | Cx1CxCx3C | CxCxs5C3x2C | 2 |
| TLR    | gp      | 28, 26  | ? | CxCxs5C3x2C | 2 |
| Others  | TLRa    | jI      | 21 | Cx1CxCx3C | CxCxs5C3x2C | 1 |
|         | TLRb    | jI      | 21 | Cx1CxCx3C | CxCxs5C3x2C | 2 |
|         | TLR15   | c       | 21 | No | CxCxs3C5x2C | 2 |

*aThe six families and others. *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; gb, Gilthead sea bream; h, human; ha, hamster; ho, horse; jf, Japanese flounder; jhm, Japanese house mouse; jl, Japanese lamprey; lg, lowland gorilla; m, mouse; mmn, 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. *Nc: the repeat number of LRRs. The number of parenthesis is Nc reported in the Interpro database. *LRRNT, the cysteine clusters on the N-terminal side of LRRs; LRRCT, the cysteine clusters on the C-terminal side of LRRs. *Nc: 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 [Q2PZ94], gTLR2 [AB131733], pTLR2 [Q59H8], hoTLR2 [AA08196], hTLR2 [O606033], cmTLR2 [Q95M53], dTLR2 [Q6891], raTLR2 [AM50059], mTLR2 [Q9QUN7], rTLR2 [Q6GYU2], chTLR2 [Q91F8], chTLR2.1 [Q9DD78], chTLR2.2 [Q9DGB6]. Abbreviations: b, Bovine; n, nilgai; dwb, domestic water buffalo; g, goat; p, pig; ho, horse; h, human; cm, Cynomolgus monkey; d, dog; r, 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/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 known in the recognized NR LRR structures, there are rare examples, xVxxLxLxxNxL, PxxLxLxxNxL and LxxLxGxx/Pxx. 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, LxxLxL/V/M/FxxS/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 vertebrate TLRs [4]. The TLR1 family consists of TLR1, TLR2, TLR6 and TLR10. This family contains 19–21 LRRs and has fewer numbers than 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 RNYYDA/PLS-RNLSS/PGDLKE, that is a 'bacterial' type. Mammalian and Takifugu rubripes TLR2 contains 20 LRRs. Japanese flounder TLR2 lacks one LRR that corresponds to LRR7 in...
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". TLR5 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 belong to the TLR1 family. Two Japanese lamprey TLRs appear to be the TLR8 family. Ten of these 22 LRRs are clearly "typical". LRR15 in TLR5 contains two addi-

The multiple sequence alignment of LRRs within mammalian TLR2 from 14 species. This panel continued from Figure 2 shows the sequences from LRR1 to the C-termini.
Consequently their LRRs are type sixth and the seventh LRRs is longer than the second, the contains also the super-repeat of first domain is homologous to the TLR7 family and thus puffer TLR forms two horseshoe domains of LRRs. The LRRs. All other LRRs are type observed in the first, fourth, and seventh of the 9 tandem LRR, LRR16, is an irregular LRR that is described by (G/

**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, Cx3-14, 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 Cx9 motif in TLR3 and the Cx10 motif of TLR4 form a disulfide bond [52,59], as does the Cx12 motif in GPIbα [41]. The Cx5 motifs presumably form disulfide bonds. The C-terminal clusters, excluding those in three TLRs (Table 2), contain four cysteines consisting of CxCx2-23Cx15-20C. 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 CxCx2-23Cx15-Cx 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 CxCx2-23Cx15-C motif in Nogo receptor [38,39] and the CxCx20Cx19-C motif in Slt1 [49]. The disulfide bond connectivity can be inferred for TLRs. The C-terminal cluster for primate TLR4 (CxCx2-23Cx15-C) is different from that of other mammalian TLR4 (CxCx23Cx15-C). Only in rainbow trout TLR5 and Takifugu rubripes TLR5 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, LxxLxxNxxLxxLxxxFxxLxx and "bacterial"; "S", LRR, LxxLxxNxxLxxPxx(x)LxxPxx. 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 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.
Figure 5

Sequence alignment of LRR domains within the six families of TLRs. (1) hTLR1 [Q15399]; hTLR2 [O60603]; hTLR6 [Q9Y2C9]; hTLR10 [Q9BX35]; tTLR14 [Q5H72N]. (2) hTLR3 [Q15455]; jTLR3 [Q76CT7]. (3) hTLR4 [O00204]; dTLR4 [QBSQH3]. (4) hTLR5 [O60602]. (5) mTLR11 [O6RSP7]; mTLR12 [O6QYN9]; mTLR13 [Q6R5N8]; cTLR21 [NP_001027751]; cTLR22 [Q5H72N]; cTLR23 [AAW70378]. (6) hTLR7 [Q9NYK1]; hTLR8 [Q9NR97]; hTLR9 [Q9NR96]. (7) jTLRα [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., cTLR14 with 871 res., cTLR3 with 904 res., hTLR4 with 961 res., hTLR5 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., cTLR21 with 965 res., cTLR22 with 950 res., cTLR23 with 941 res., jTLRα with 813 res., and cTLR15 with 868 res. Cysteine is highlighted in magenta. Its boldface indicates cysteines in LRRNT or LRRC. Residues of missense mutation are highlighted in blue boldface. SIGNAL, signal peptide sequence; LRNT, the cysteine clusters on the N-terminal side of LRRs; LRRC, 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; j, Japanese flounder. This panel shows hTLR1, hTLR2, jTLRα and TLR6 in the TLR1 family.
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 jTLR3 in the TLR3 family.
**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, hTLR5 in the TLR5 family, and mtLR11 (from SIG-NAL to LRR15) in the TLR11 family.
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 cTLR21 in the TLR11 family.
TLR23

**TLR2**

**Signal:**

**Sequence:**

**Library:**

**Page 13 of 20**

**Figure 9**

Sequence alignment of LRR domains within the six families of TLRs. This panel continued in Figure 8 shows cTLR22 and cTLR23 in the TLR11 family, and hTLR7 and hTLR8 (from SIGNAL to LRR11) in the TLR7 family.

**hTLR8**

**Signal:**

**Sequence:**

**Library:**

**Page 13 of 20**
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 jTLRa 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 (ϕ), and the tilt angle of the parallel β-strand direction per turn (θt). A 3D circle fitting method to calculate these geometrical parameters has been developed [55]. The TLR3-LRR arc yields R = 26.5–26.6Å, ϕ = 10.8–10.9° and θt = 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 GPIIβα with "typical" LRRs [8,55,58]. In contrast, the θt value is comparable to only those for Slit (-21°) and FSHr (-40°). Also the θt 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 (PS8682), human TLR8 (Q9NR97), mouse TLR8 (PS8682), 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, and green puffer TLR, respectively, were used for the cross-dot matrices. The abscissa axis and the ordinate axis are residues number.
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

**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.

**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 nablacip, keratan, 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.

**asporin.** Class II has three subclasses: lumican plus fibromodulin (IIA), PRELP plus keratan (IIB), and osteoadherin (IIC). Class III consists of epiphysin, 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-
LRR18 occur at position 1 in the HCS part. The side chain of I499 is completely buried in the LRR arc. Such a mutation is also observed in trk-A and nytalolin, 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, 11DJ, LA4Y, 1Z7X], GTPase-activating protein (RanGAP) [1YRG, 1K5D, 1K5G], tropomodulin (Tmod) [1100, 1PGV], S-phase kinase-associated protein 2 (Skp2) [1FQV], YopM [1G9U], four internals, Inl-B [1DB], Inl-H [1H6U], Inl-A [106T, 106V, 106S] and Inl-C [1XEL], spliceosomal U2A' protein [1A9N], mRNA export factor (TAP) [1F87, 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/no-go-66 receptor (NgR) [10ZN, 1PPA], glyco-protein Ibo (GPIbo) [1M0Z, 1GWB, 1QYQ, 1M10, 1SQO, 1PSV, 1Q9A, 1U0N], decorin [1XCD, 1XKL, 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, Q6Fl64, Q32MK3], mouse [Q9EPO1], pig [Q4LD77, Q5H19], Takifugu rubripes [Q5H727]); TLR2 from 17 species (human [Q06063], mouse [Q9QQUN7, Q8K3D9, Q811T5], pig [Q59H18, Q5DX20, Q76L24], chicken [Q9DD78 (TLR2.1), Q9DG86 (TLR2.2)], bovine [Q95L9], rat [Q6GUY2], dog [Q86D91], rabbit [AAM50059], goat [AB131733], horse [AAR08196], hamster [Q9R1F8], Cynomolagus monkey [Q95SM3], domestic water buffal0 [Q2PZI4], Nilgai [Q2V897], Takifugu rubripes [Q5H725], zebrafish [Q6TS42], Japanese flounder [Q76C18]; TLR3 from 9 species (human [O15455, Q4VAL2, Q504W0], mouse [Q99NM1, Q3TM31, Q499F3], bovine [Q5TI58, Q5T159], rat [Q7MN8], buffalo [Q1G1A3], Rhesus macaque [Q3BBY1], Takifugu rubripes [Q5H721], zebrafish [Q6W5L, Q32PW5], Japanese flounder [Q76C17, Q76C19]; TLR4 from 17 species (human [O00206, Q5VZI7, Q5VZI8, Q5VZI9], mouse [Q9QU66, Q5RTG4, Q8K2T5], pig [Q685Y6, Q2TNN4, Q5AFK7, Q401C7], bovine [Q5GL65, Q5WDC5, Q85QQ5], rat [Q5QX05], hamster [Q9WWV2], cat [Q58727], lowland gorilla [Q85PE8], horse [Q9MYW3], Pygmy chimpanzee [Q97NNO], olive baboon [Q9FTSP2], orangutan [Q8SEP9], Nilgai [Q2V898], American bison [Q3ZD70], dog [Q85OH3], rabbit [AAM5060], zebrafish [Q6NV08, Q6TS41(TLR4b)]; TLR5 from 8 species (human [O60602], pig [Q59H17], mouse [Q9L7F7], bovine [Q2LDA0], chicken [Q4Z832], Japanese house mouse [Q1ZZX0], Takifugu rubripes [Q5H720, Q5H716(TLR5)]; TLR6 from 5 species (human [O00206, Q5VZI7, Q5VZI8, Q5VZI9], mouse [Q99NM1, Q3TM31], bovine [Q2LDA0], chicken [Q4Z832], Japanese house mouse [Q1ZZX0], Takifugu rubripes [Q5H720, Q5H716(TLR5)]), rainbow trout [Q7Z181]), TLR7 from 4 species (human [O15455, Q4VAL2, Q504W0], mouse [Q99NM1, Q3TM31, Q499F3], bovine [Q5TI58, Q5T159], rat [Q7MN8], buffalo [Q1G1A3], Rhesus macaque [Q3BBY1], Takifugu rubripes [Q5H721], zebrafish [Q6W5L, Q32PW5], Japanese flounder [Q76C17, Q76C19]; TLR8 from 4 species (human [Q9NR97, Q495P4, Q495P6, Q495P7], mouse [Q58682], pig [Q685R7], Takifugu rubripes [Q5H718]); TLR9 from 12 species (human [O00206, Q5VZI7, Q5VZI8, Q5VZI9], mouse [Q99NM1, Q3TM31, Q499F3], bovine [Q2LDA0], chicken [Q4Z832], Japanese house mouse [Q1ZZX0], Takifugu rubripes [Q5H720, Q5H716(TLR5)]), rainbow trout [Q7Z181]), TLR10 from 17 species (human [O00206, Q5VZI7, Q5VZI8, Q5VZI9], mouse [Q99NM1, Q3TM31, Q499F3], bovine [Q2LDA0], chicken [Q4Z832], Japanese house mouse [Q1ZZX0], Takifugu rubripes [Q5H720, Q5H716(TLR5)]), rainbow trout [Q7Z181]).
pig [Q4LDR6, Q59H15]); 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, Institute 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: Lipopolysaccharide. RRR: 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. Kreisinger 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.).
sites of Listeria monocytogenes InlB. Biochem Biophys Res Commun 2004, 316(2):379-386.

26. Schubert WD, Wurzler H, Clehm T, Beier V, Mascher MP, Domann E, Wehland J, Chakraborty T, Heinz DW: Structure of internalin A, a major invasion protein of Listeria monocytogenes, in complex with its human receptor E-cadherin. Cell 2002, 111(6):925-936.

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. Liu E, Fernández E, Izaurralde E, Conti E: The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. EMBO J 2000, 19(12):3587-3598.

29. Ho DN, Coburn GA, Kang Y, Cullen BR, Georgiadis MM: The crystal structure and functional analysis of a novel RNA-binding domain found in the human Tap nuclear mRNA export factor. Proc Natl Acad Sci USA 2002, 99(4):1889-1893.

30. Zhang H, Seabra MC, Deisenhofer J: Solution structure of a dynein motor domain associated light chain. Nat Struct Biol 2000, 7(7):575-579.

31. Wu H, Blackwell M, Majeskie MW, Mullen GP, King SM: Solution structure of a dynein complex involved in plant defense. Proc Natl Acad Sci USA 2003, 100(17):10241-10248.

32. Lu X, Bazan JF, McDermott G, Park JB, Wang K, Tessier-Lavigne M, He G, Garcia KC: Structure of the Nogo receptor ectodomain: a recognition module implicated in myelin inhibition. Neuron 2003, 38(2):177-185.

33. Huzinaga EG, Tsujii S, Romijn BA, Schiporoste MM, 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.

34. Ulf S, Clemston JM, Harrison T, Clemenot KJ, Ensley J: Crystal structure of the platelet glycoprotein Ibalpha N-terminal domain reveals an unmasking mechanism for receptor activation. J Biol Chem 2002, 277(38):35657-35663.

35. Dumas JJ, Kumar R, Seehra J, Somers WS, Mosyak L: Crystal structure of the Ibalpha-thrombin complex essential for platelet aggregation. Science 2003, 301(5630):222-226.

36. Ceccatelli S, Mattie M, Matie M, Alberi M, Tagliafico, R, Capasso A, Bottaro DPM: Structural correlations in the extra-cellular domain of the RhoGEF protein. J Biol Chem 2005, 280(50):48322-48330.

37. Varughese KI, Ruggeri ZM, Ceccatelli S, Bottaro DPM: Structure of internalin A1-glycoprotein Ibalpha complex reveals conformational changes at the receptor-ligand interface. J Biol Chem 2005, 280(19):22257-22264.

38. Vogel SN, Awomoyi AA, Rallabandi 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.

39. Bell JK, Mullen GE, Leifer CA, Mazzoni A, Davies DR, Segal DM: Leucine-rich repeat proteoglycans. J Am Acad Dermatol 2003, 49(6):159-167.

40. Bell JK, Mullen GE, Leifer CA, Mazzoni A, Davies DR, Segal DM: Leucine-rich repeat proteoglycans. J Am Acad Dermatol 2003, 49(6):159-167.

41. Arbour NC, Lorencez S, Schutte BC, Kline JN, Jones M, Frees K, Vangst JL, Schwartz DA, TLR4 mutations are associated with endotoxic hyporesponsiveness in humans. Nat Genet 2000, 25:187-191.

42. Vogel SN, Awomoyi AA, Rallabandi 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.

43. Kumpf O, Hamann L, Schlag P, Schumann RR: Pre- and postoperative cytokine release after in vitro whole blood lipopolysaccharide stimulation in Toll-like receptor 4 polymorphisms. Shock 2006, 25(2):123-128.

44. Coyne JF, Meyer CU, Gemmer C, Russo A, Finn A, Murdoch CS, Sisson W, Maniatis T, Zabel B, Hirt P, Kost H, Kauls LS, Gaspari AA: 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, Means 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 Immunol* 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, Alexopoulos L, Flavell RA, Beutler B: Toll-like receptors 9 and 3 as essential components 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.cs.uiuc.edu/ssp/pro4.html](http://contact.cs.uiuc.edu/ssp/pro4.html).

78. [http://129.128.185.184/proteus/#](http://129.128.185.184/proteus/#).

79. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004, 340(4):783-795.

80. [http://www.cbs.dtu.dk/services/SignalP/](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/](http://www.cbs.dtu.dk/services/TMHMM/).

83. [http://pfam.janelia.org/](http://pfam.janelia.org/).

84. [http://www.genome.ad.jp/](http://www.genome.ad.jp/).