



**HAL**  
open science

## **TollML: a database of toll-like receptor structural motifs**

Jing Gong, Tiandi Wei, Ning Zhang, Ferdinand Jamitzky, Wolfgang M. Heckl,  
Shaila C. Rössle, Robert W. Stark

### ► **To cite this version:**

Jing Gong, Tiandi Wei, Ning Zhang, Ferdinand Jamitzky, Wolfgang M. Heckl, et al.. TollML: a database of toll-like receptor structural motifs. *Journal of Molecular Modeling*, 2010, 16 (7), pp.1283-1289. 10.1007/s00894-009-0640-9 . hal-00557308

**HAL Id: hal-00557308**

**<https://hal.science/hal-00557308>**

Submitted on 19 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TollML: a database of toll-like receptor structural motifs

Jing Gong · Tiandi Wei · Ning Zhang ·  
Ferdinand Jamitzky · Wolfgang M. Heckl ·  
Shaila C. Rössle · Robert W. Stark

Received: 18 October 2009 / Accepted: 19 November 2009 / Published online: 19 January 2010  
© Springer-Verlag 2010

**Abstract** Toll-like receptors (TLRs) play a key role in the innate immune system. TLRs recognize pathogen-associated molecular patterns and initiate an intracellular kinase cascade to induce an immediate defensive response. During recent years TLRs have become the focus of tremendous research interest. A central repository for the growing amount of relevant TLR sequence information has

been created. Nevertheless, structural motifs of most sequenced TLR proteins, such as leucine-rich repeats (LRRs), are poorly annotated in the established databases. A database that organizes the structural motifs of TLRs could be useful for developing pattern recognition programs, structural modeling and understanding functional mechanisms of TLRs. We describe TollML, a database that integrates all of the TLR sequencing data from the NCBI protein database. Entries were first divided into TLR families (TLR1-23) and then semi-automatically subdivided into three levels of structural motif categories: (1) signal peptide (SP), ectodomain (ECD), transmembrane domain (TD) and Toll/IL-1 receptor (TIR) domain of each TLR; (2) LRRs of each ECD; (3) highly conserved segment (HCS), variable segment (VS) and insertions of each LRR. These categories can be searched quickly using an easy-to-use web interface and dynamically displayed by graphics. Additionally, all entries have hyperlinks to various sources including NCBI, Swiss-Prot, PDB, LRRML and PubMed in order to provide broad external information for users. The TollML database is available at <http://tollml.lrz.de>.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-009-0640-9) contains supplementary material, which is available to authorized users.

J. Gong · T. Wei · F. Jamitzky · W. M. Heckl · R. W. Stark  
Center for Nanoscience,  
Ludwig-Maximilians-Universität München,  
80799 Munich, Germany

J. Gong · T. Wei (✉) · S. C. Rössle · R. W. Stark  
Department of Earth and Environmental Sciences,  
Ludwig-Maximilians-Universität München,  
Theresienstr. 41,  
80333 Munich, Germany  
e-mail: [tiandi@informatik.uni-muenchen.de](mailto:tiandi@informatik.uni-muenchen.de)

N. Zhang  
Department of Informatics,  
Ludwig-Maximilians-Universität München,  
80333 Munich, Germany

F. Jamitzky  
Leibniz Supercomputing Centre,  
85748 Garching, Germany

W. M. Heckl  
Deutsches Museum,  
80538 Munich, Germany

W. M. Heckl  
TUM School of Education, Technische Universität München,  
80799 Munich, Germany

**Keywords** TollML · Toll-like receptor ·  
Leucine-rich repeats · XML database · Homology modeling

## Introduction

Since the *Drosophila Toll* gene was discovered in the mid-1980s [1], genome projects have led to the identification of 13 receptors in mammalian and more than 20 receptors in non-mammalian genomes that are homologs of *Drosophila Toll*. These receptors have been termed collectively Toll-like receptors (TLRs). TLRs play a key role in innate immunity. They recognize invading microbial pathogens

and rapidly initiate intracellular signal transduction pathways to trigger expression of genes, whose products can control innate immune responses [2]. All TLRs have a common domain organization, with an extracellular ectodomain (ECD), a helical transmembrane domain (TD), and an intracellular Toll/IL-1 receptor homology (TIR) domain [3]. The ectodomain is a horseshoe-shaped solenoid structure and is directly involved in the recognition of a variety of pathogens including lipopolysaccharide, lipopeptide, cytosine-phosphate-guanine (CpG) DNA, flagellin, imidazoquinoline and dsRNA [4]. The transmembrane domain determines the subcellular localization of TLRs [5]. The TIR domain is conserved across all TLRs and IL-1 receptors, and is also shared by downstream signaling adaptor molecules. Upon receptor ligation, a TIR signaling complex is formed between the receptor and the adaptor TIR domains [6].

The TLR ectodomain contains varying numbers of leucine-rich repeat (LRR) motifs, which are arrays of 20 to 30 amino acid-long protein sequences that are enriched with the hydrophobic amino acid leucine. All LRR sequences can be divided into a highly conserved segment (HCS) and a variable segment (VS). The HCS consists of an 11 or 12 residue stretch with the consensus sequence LxxLxLxxN(Cx)xL. In this notation, the letter L represents the amino acids leucine, isoleucine, valine or phenylalanine, which form a hydrophobic core, N represents asparagine, threonine, serine or cysteine, and x is any amino acid. The variable segment can vary in both length and consensus sequence. Accordingly, several types of LRRs have been proposed [7, 8]. Of these, typical (T) type (xxLxxxxLxxLxx) and bacterial (S) type (xxLPx(x)LPxx) LRRs have been observed in TLRs [9]. All LRRs in TLRs are capped by N- and C-terminal LRRs that are usually irregular and do not match any type of LRR consensus sequences.

The atomic-detail crystal structure of the human TLR1 and TLR2 TIR domain was published in 2000 and gave the first insight into the molecular basis of TIR signaling [10]. The crystal structures of the ectodomains of human TLR1-4 and mouse TLR2-4 have also been resolved [11–16]. These structures demonstrate how the LRR-based platform is adapted to ligand recognition. Nevertheless, more than 2,000 TLR proteins have been sequenced by high-throughput genome sequencing projects. It is clear that the discrepancy between the rate at which novel protein sequences are discovered and the rate at which detailed structural information on proteins can be obtained from X-ray diffraction or nuclear magnetic resonance spectroscopy will persist for the foreseeable future. Thus, a comparative analysis at the sequence level is a useful approach to identify and characterize structural motifs of TLRs [9, 17] and to gain insight into how receptors and

ligands interact. Due to the variability of LRR motifs in TLRs, however, the indicated repeat number and positions (beginning/end of a repeating unit) for individual TLRs are quite different or missing in established databases. Currently, there is no collection of structural information for features that are contained within LRRs, such as HCS, VS and sequence insertions.

In this paper, we describe a database of TLR structural motifs called TollML. The current release (3.1) includes all known TLR sequences from the NCBI protein database [18]. Structural motifs were identified and annotated by a semi-automatic procedure that included comparison of sequences with the sequences of TLRs that have a known structure, consensus sequence matching, secondary structure prediction and multiple sequence alignments. Three levels of motif elements were generated: (1) signal peptide (SP), ectodomain (ECD), transmembrane domain (TD) and TIR domain of each TLR; (2) LRRs of each ECD; and (3) HCS, VS and insertions of each LRR. Some program application examples are presented in the last section of the paper.

## Construction and content

### Data extraction and pre-processing

Initial TLR sequences were extracted from the NCBI protein database. Two groups of search results were obtained using the search keys *tol\*<sup>\*</sup>* and *tlr\*<sup>\*</sup>*, where the asterisk stands for any suffix, to ensure that all TLRs were included. A manual data pre-processing step was performed before the motif identification of these sequences. We inspected the NCBI annotations of entries one-by-one to exclude TLR related molecules such as adaptors, protein kinases and transcription factors. After we performed these filtering steps, 2,572 TLR entries remained (NCBI release: 1 September 2009). We then categorized the TLRs into different families (TLR1-23) based on their original annotations. In the instances in which entries were not associated with explicit comments, we compared the sequences with well classified TLR sequences using sequence BLAST.

### Motif identification

Three levels of structural motif categories were generated adapting to the structural organization of TLRs. First, each full length TLR sequence was divided into ECD, TD and TIR domain. If a sequence started with a SP, which directs the subcellular transport of a protein, the presence of a SP was also indicated. Second, the ECD of each TLR was partitioned into individual LRRs including

canonical LRRs and N/C-terminal LRRs. Third, each LRR was further divided into a highly conserved segment (HCS) and a variable segment (VS). Insertions within the VS that are longer than three residues were identified and annotated.

The procedure mentioned above was semi-automated. We first sorted the TLR sequences from each family into subgroups, so that an arbitrary pair of sequences from the same subgroup had a local sequence similarity greater than 90%. Then, the three-level structural motifs were identified manually for a selected template sequence from each subgroup. This selected template sequence was a full length sequence that had the most detailed original annotations so that the most accurate motif identifications could be performed. Finally, the other members of the subgroup received their motif partition assignments through multiple alignments with the template. More than 300 subgroups were generated for the 2,572 entries. Thus more than 300 templates were processed manually.

The manual motif identification of a template sequence combined three approaches: consensus sequence matching, secondary structure prediction and reference to original annotations or literatures. The four domains on the first level of the motif categorization (SP, ECD, TD and TIR) have characteristic sequence features, so are usually accurately divided in the original annotation of a selected template. If a template did not have clear annotations, then its sequence was compared to similar sequences with known domains to determine its domain partitions. For the second level of motif categorization, if a selected template was associated with a reliable reference such as a known crystal structure, its LRR partition was then assigned accordingly. Otherwise, we matched the LRR consensus sequence  $LxxLxLxxN(Cx)xL$  to the template sequence amino-acid-by-amino-acid and detected LRR motifs manually. In addition, protein secondary structure predictions (PredictProtein [19], NNPREPDICTION [20], PSIPRED [21] and SSPro [22]) helped to improve the accuracy of LRR detection because all known crystal structures of TLRs show that there is always a short  $\beta$ -strand (3–5 residues) beginning at approximately the third position of an LRR motif [11–15]. After an LRR was identified, consensus sequence matching was used to identify its HCS and VS motifs, as well as any insertions that were longer than three residues for the third level. Simultaneously, each LRR motif was classified into different types (detailed in Database content) according to the VS consensus sequences.

#### Database content

The TollML release 3.1 contains 2,572 TLR entries divided into 23 families (entry distribution shown in Table 1).

Among these, 2,350 of the sequences contain an ectodomain and thus received motif annotations that correspond to the second and third levels of motif categorization. The other entries contain only a TIR domain with or without a transmembrane domain. A total of 46,720 LRR motifs were recognized from the ectodomain containing TLRs. These LRR motifs were classified into five types: typical (T), bacterial (S), N-terminal (NT), C-terminal (CT) and irregular (I). A histogram of LRR length statistics (Fig. 1) shows the characteristic length distribution of each LRR type. The standard length of the T type LRR is 24 amino acids. A large number of T type LRRs have insertions and only some have deletions. These statistics suggest that the evolution of T type LRRs may prefer insertion over deletion. By contrast, the S type LRRs are more highly conserved. Their lengths are concentrated on 20 and 21. N-terminal LRRs vary in length and do not form a peak value in length distribution. Most C-terminal LRRs contain four cysteines that are distantly separated at the sequence level and form disulfide bonds with each other. C-terminal LRRs are generally greater than 35 amino acids in length.

Annotations for each TLR entry include:

- (1) Data management information: TollML ID and access/modification date;
- (2) Primary information extracted from the NCBI and related literature: FASTA sequence, biological definition, cell information, glycosylation sites and ligands;
- (3) Protein family classification;
- (4) Database cross links: NCBI, Swiss-Prot [23], PDB [24], LRRML [8] and PubMed [25];
- (5) Three-level motif information.

#### Database comparison

Currently, several protein databases, such as Pfam [26], InterPro [27], SMART [28] and Swiss-Prot, contain information about TLRs. These databases predict the LRR numbers and positions for their TLR entries by various computational methods, thus resulting in a high frequency of false negative predictions. Table 2 illustrates the LRR numbers for human TLR1–10 as reported by these databases. The manual motif identification procedure discussed here provides TollML with the most complete database of LRR motifs. Although Swiss-Prot presents more accurate results than the other three databases that we investigated (Table 2), TollML has four prominent characteristics that distinguish it from Swiss-Prot:

- (1) Comprehensive entry coverage. TollML covers 2,572 TLR sequences from 121 species and all sequences are provided with detailed motif annotations. Swiss-Prot

**Table 1** Entry distribution over Toll-like receptor (TLR) families for mammalian/non-mammalian groups

TLR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	19	20–23	Total
Mammalian	85	146	108	443	132	106	109	101	124	72	10	11	9	0	0	0	0	0	0	1,456
Non-mammalian	356	276	84	77	58	3	102	11	30	0	0	1	4	6	42	2	7	5	52	1,116
Total	441	422	192	520	190	109	211	112	154	72	10	12	13	6	42	2	7	5	52	2,572

covers 636 TLR sequences from 17 species and only 59 sequences have LRR annotations (results obtained on 24 February 2009).

- (2) Structural motifs within an LRR. TollML annotates the HCS, VS and insertion for each LRR. This information is not present in any other published protein databases.
- (3) Uniform LRR definition. The beginning/end positions of LRRs have been defined inconsistently across researchers due to the periodicity of LRR motifs. This variation leads to non-uniform LRR assignments in Swiss-Prot. All LRR motifs in TollML start at the beginning of the HCS and end at the end of the VS, just before the HCS of the next LRR.
- (4) Accessibility of motif sequences. The amino acid sequence of any available motif is directly accessible in TollML, whereas only the full length sequence is directly accessible in Swiss-Prot.

## Utility

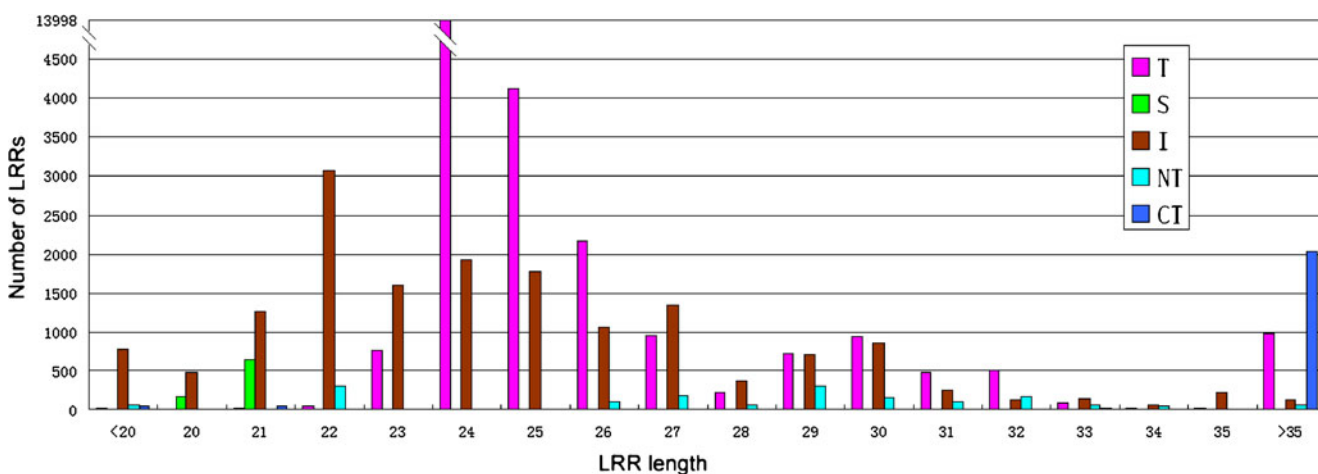
### Web application

The extensible markup language (XML) was standardized in the 1990s and is well established as a format for hierarchical biological data. TollML was designed by using

eXist [29], an XML database management system, and XPath/XQuery [30] for processing queries and web forms. The document type definition (DTD) file of TollML is provided in the electronic supplementary material (ESM; Supplementary file 1).

The entire database is browsable. When browsing, entries appear in a summary table containing ID, definition, family, species and links of motif partitions. Clicking on an entry opens an XML Stylesheet (XSLT) [30] converted HTML web page that describes the entry in detail. The original XML file can also be downloaded. The XSLT file that the program uses is provided in the ESM (Supplementary file 2). Aside from the textual view, the structural motifs of TLRs can be exhibited by three-level dynamic graphics. Figure 2 demonstrates the motif assignment for an example entry (ID: TLR\_561).

On the *advanced search* page of TollML, users can search entries flexibly by inputting keywords, specifying search fields, and defining annotation contents of the output. After selected entries are returned, a *search within result* button allows for further term filtering. The resulting entries, or an arbitrarily selected subset thereof, can be sent to generate multiple sequence alignments supplied by the T-Coffee package [31]. In addition, a Wu-BLAST search tool [32] is available. A query sequence can be BLASTed against the entire database, against a certain TLR family, or against a collection of sequences marked by a user-defined label (available for registered users).

**Fig. 1** Leucine-rich repeat (LRR) length distribution

**Table 2** Comparison of leucine-rich repeat (LRR) numbers of human TLR1–10 in different databases (results obtained on 24 February 2009)

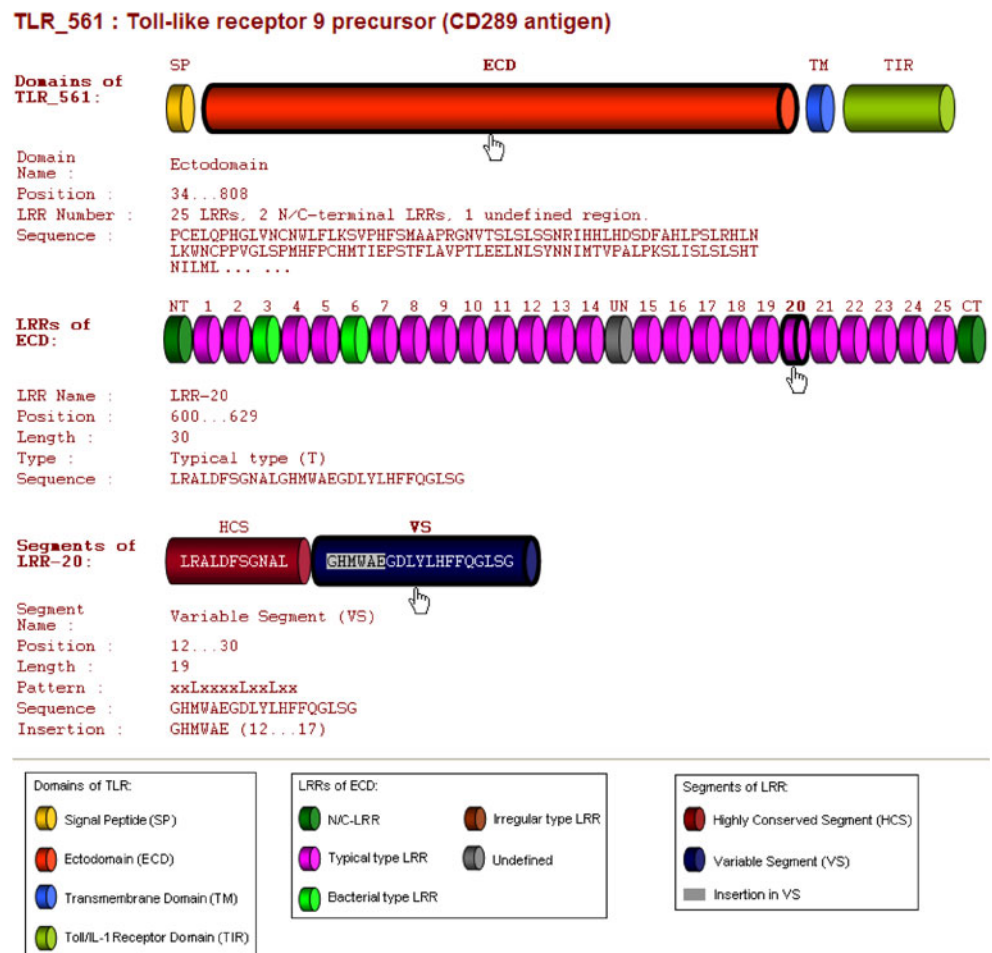
	TLR1	TLR2	TLR3	TLR4	TLR5	TLR6	TLR7	TLR8	TLR9	TLR10
InterPro	4	4	8	7	7	3	9	10	6	4
Pfam	4	4	7	7	8	3	8	9	6	4
SMART	6	9	19	12	10	7	16	17	19	7
Swiss-Prot	8	14	22	21	15	13	27	24	26	15
TollML	21	21	25	23	23	21	28	28	28	21

Application in LRR prediction

The automatic identification of motifs in a protein sequence is essentially a statistical pattern recognition problem. Therefore, the performance of a prediction method is thus strongly dependent on the quality and scale of the training data set. A recent program, LRRscan [17], demonstrated the feasibility of a statistics-based consensus matching algorithm applied to LRR detection. TollML supplies a large and reliable source to train LRR prediction methods of different algorithms. As an example, we developed a 20×12 position-specific weight matrix of LRR motifs based on

the LRR partitions from TollML (matrix available on the TollML webpage). A matrix element denotes the frequency probability of a certain amino acid occurring at a certain position in the LRR HCS. We have already obtained confident results through this sort of matrix scan with an appropriate cut-off score. A five-fold cross validation against all TollML entries indicates that the sensitivity and specificity are both greater than 93%. This program was implemented on the TollML webpage and has been named LRRfinder. This method can be extended to predict LRR motifs in other LRR containing proteins besides TLRs, since all LRR types possess the same HCS pattern.

**Fig. 2** Online graphic display of three-level structural motifs of a database entry



## Application in homology modeling

Homology modeling is currently the most accurate computational method to predict protein structure. This system constructs a structural model for a target protein sequence from a template structure of a homologous protein. For this method to work, the target and template structures must have a sequence identity over 30%. The known crystal structures of human TLR1–4 and mouse TLR2–4 ectodomains supply valuable templates for the homology modeling of other TLR ectodomains. However, given the variability in repeat numbers and type arrangements of LRRs contained within TLRs, a full length template with a sufficient sequence identity is typically not available. This limitation can be overcome by assembling multiple LRR templates. First, all LRRs in the target sequence are identified. Then, the most similar structure-known LRR based on sequence is found for each LRR in the target sequence as a local template. Finally, all local templates are assembled to generate a multiple sequence alignment for the complete target sequence. TollML enables the direct accessibility of accurate LRR sequence partitions for the first step. The LRRML database [8] can further provide suitable LRR structural templates. This LRR template assembling approach was proven to be both feasible and significant by recent structure modeling research into human TLR7–9 [33].

## Conclusions

We have developed a specialized database of TLR structural motifs called TollML. It is supported by an XML database management system and can be searched and browsed with an easy-to-use web interface. This interface is suitable for use with most graphical web browsers and has been tested on the Windows, Mac and Linux operating systems. TollML includes all TLR sequences that are published in the NCBI protein database and semi-automatically creates three levels of motif annotations. This database can help to develop motif prediction programs, to model three-dimensional structures of TLRs and to design new mutation experiments to better understand receptor–ligand or receptor–receptor interactions. We plan to update TollML every 2–3 months since the number of sequenced TLR proteins increases constantly.

## Availability and requirements

This database is freely available at <http://tollml.lrz.de>. Any internet user can search and download data from the database, but only registered users can define and save labels for arbitrary entries.

**Acknowledgments** This work was supported by Graduiertenkolleg 1202 of the Deutsche Forschungsgemeinschaft (DFG) and the DFG excellence cluster Nanosystems Initiative Munich (NIM).

## References

- Anderson KV, Bokla L, Nusslein-Volhard C (1985) Establishment of dorsal-ventral polarity in the *Drosophila* embryo: the induction of polarity by the Toll gene product. *Cell* 42:791–798
- Takeda K, Akira S (2005) Toll-like receptors in innate immunity. *Int Immunol* 17:1–14
- Brodsky I, Medzhitov R (2007) Two modes of ligand recognition by TLRs. *Cell* 130:979–981
- Gay NJ, Gangloff M (2007) Structure and function of Toll receptors and their ligands. *Annu Rev Biochem* 76:141–165
- Barton GM, Kagan JC, Medzhitov R (2006) Intracellular localization of Toll-like receptor 9 prevents recognition of self DNA but facilitates access to viral DNA. *Nat Immunol* 7:49–56
- O'Neill LA, Bowie AG (2007) The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* 7:353–364
- Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* 11:725–732
- Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Roessle SC (2008) LRRML: a conformational database and an XML description of leucine-rich repeats (LRRs). *BMC Struct Biol* 8:47
- Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, Kuroki Y (2007) Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* 8:124
- Xu Y, Tao X, Shen B, Horng T, Medzhitov R, Manley JL, Tong L (2000) Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* 408:111–115
- Bell JK, Botos I, Hall PR, Askins J, Shiloach J, Segal DM, Davies DR (2005) The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proc Natl Acad Sci USA* 102:10976–10980
- Choe J, Kelker MS, Wilson IA (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 309:581–585
- Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee H, Lee JO (2007) Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* 130:1071–1082
- Kim HM, Park BS, Kim JI, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, Lee JO (2007) Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* 130:906–917
- Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, Segal DM, Davies DR (2008) Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* 320:379–381
- Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 458:1191–1195
- Dolan J, Walshe K, Alsbury S, Hokamp K, O'Keefe S, Okafuji T, Miller SF, Tear G, Mitchell KJ (2007) The extracellular leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. *BMC Genomics* 8:320
- Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvermin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E (2008) Database resources of the

- National Center for Biotechnology Information. Nucleic Acids Res 36:D13–D21
19. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic Acids Res 32:W321–W326
  20. Kneller DG, Cohen FE, Langridge R (1990) Improvements in protein secondary structure prediction by an enhanced neural network. J Mol Biol 214:171–182
  21. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at University College London. Nucleic Acids Res 33:W36–W38
  22. Pollastri G, Przybylski D, Rost B, Baldi P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins 47:228–235
  23. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187–D191
  24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
  25. PubMed: <http://www.ncbi.nlm.nih.gov/pubmed>
  26. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. Nucleic Acids Res 36: D281–D288
  27. Mulder N, Apweiler R (2007) InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol 396:59–70
  28. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 28:231–234
  29. eXist: <http://exist-db.org>
  30. The World Wide Web Consortium: <http://www.w3.org>
  31. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217
  32. WU-BLAST: <http://blast.wustl.edu>
  33. Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rossle SC (2009) Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. Protein Sci 18:1684–1691