

A leucine-rich repeat assembly approach for homology modeling of the human TLR5-10 and mouse TLR11-13 ectodomains

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

Received: 12 December 2009 / Accepted: 24 February 2010 / Published online: 30 March 2010
© Springer-Verlag 2010

Abstract So far, 13 groups of mammalian Toll-like receptors (TLRs) have been identified. Most TLRs have been shown to recognize pathogen-associated molecular patterns from a wide range of invading agents and initiate both innate and adaptive immune responses. The TLR ectodomains are composed of varying numbers and types of leucine-rich repeats (LRRs). As the crystal structures are currently missing for most TLR ligand-binding ectodomains, homology modeling enables first predictions of their three-dimensional structures on the basis of the determined crystal structures of TLR ectodomains. However, the quality of the predicted models that are generated from full-length

templates can be limited due to low sequence identity between the target and templates. To obtain better templates for modeling, we have developed an LRR template assembly approach. Individual LRR templates that are locally optimal for the target sequence are assembled into multiple templates. This method was validated through the comparison of a predicted model with the crystal structure of mouse TLR3. With this method, we also constructed ectodomain models of human TLR5, TLR6, TLR7, TLR8, TLR9, and TLR10 and mouse TLR11, TLR12, and TLR13 that can be used as first passes for a computational simulation of ligand docking or to design mutation experiments. This template assembly approach can be extended to other repetitive proteins.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0697-5) contains supplementary material, which is available to authorized users.

T. Wei · J. Gong (✉) · F. Jamitzky · W. M. Heckl · R. W. Stark
Center for Nanoscience,
Ludwig-Maximilians-Universität München,
80799 Munich, Germany
e-mail: gongj@informatik.uni-muenchen.de

T. Wei · J. Gong · S. C. Rössle · R. W. Stark
Department of Earth and Environmental Sciences,
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 Toll-like receptor · Leucine-rich repeats · Homology modeling · Template assembly · TollML · LRRML

Introduction

Cells of the innate immune system, such as macrophages and dendritic cells, express a limited number of germline-encoded pattern-recognition receptors (PRR) that specifically recognize pathogen-associated molecular patterns (PAMPs) within microbes. These molecular patterns are unique to these microbes and are absent in the host [1]. Toll-like receptors (TLRs) are currently the best-characterized members of the PRRs [2]. The progress of genome sequencing projects has led so far to the identification of 13 groups of TLRs in mammalian genomes, ten in humans and 13 in mice [3], and more than 20 in non-mammalian genomes [4]. All TLRs have a common domain organization, with an extra-cellular ectodomain, a helical transmembrane domain, and

an intracellular Toll/IL-1 receptor homology (TIR) domain [5]. The extracellular domain (ectodomain) is responsible for the recognition of common structural patterns in various microbial molecules. For example, lipoproteins or lipopeptides are recognized by TLR2 complexed with TLR1 or TLR6, viral double-stranded RNAs by TLR3, lipopolysaccharides by TLR4, bacterial flagellins by TLR5, single-stranded RNAs by TLR7 or TLR8, and microbial CpG DNAs by TLR9 [6, 7]. The TIR domains of TLRs are associated with the intracellular signaling cascade leading to the nuclear translocation of the transcription factor NF- κ B [8].

A TLR ectodomain contains 19–27 consecutive leucine-rich repeat (LRR) motifs sandwiched between two terminal LRR modules (LRRNT and LRRCT) [4]. LRRs exist in more than 6000 proteins, and more than 100 crystal structures of these proteins have been deposited in the Protein Data Bank (PDB) [9, 10]. In every case, the protein adopts an arc or horseshoe shape. An individual LRR motif is defined as an array of 20–30 amino acids that is rich in the hydrophobic amino acid leucine. All LRR sequences can be divided into a conserved segment and a variable segment. The conserved segments, LxxLxLxxNxL, generate the concave surface of the LRR arc or horseshoe by forming parallel β -strands, while the variable parts form its convex surface, consisting of helices or loops. The terminal LRRNT and LRRCT modules stabilize the protein structure by shielding its hydrophobic core from exposure to solvent.

To date, only the crystal structures of the ectodomains of human TLR1–4 and mouse TLR2–4 have been determined [11–15]. High-throughput genome sequencing projects, however, have led to the identification of more than 2000 TLR sequences. Thus, the structures of most TLRs are still unknown because structure determination by X-ray diffraction or nuclear magnetic resonance spectroscopy experiments remains time-consuming. Protein structure prediction methods are powerful tools for bridging the gap between sequence determination and structure determination.

Homology modeling, also referred to as comparative modeling, is currently the most accurate computational method for protein structure prediction. This approach constructs a three-dimensional model for a target protein sequence from a three-dimensional template structure of a homologous protein. Thus, the quality of the homology model strongly depends on the sequence identity between the target and template. Below 30% identity, serious errors may occur [16]. Due to different repeat numbers and distinct arrangements of LRRs in the TLR ectodomains, a proper full-length template with a sufficiently high sequence identity to the target is often missing. This limitation can be overcome by assembling multiple LRR templates. In this approach, the most similar (on the sequence level) LRR with a known structure is searched for as a local template for each LRR in the target sequence. Such an LRR template may be

derived from TLRs or from other proteins. Thereby, a suitable template may even be found for an insertion-containing irregular LRR. All local template sequences are then combined to generate a multiple sequence alignment for the complete target sequence. Thus, a high-quality model can be created, even if no adequate single template is available. To facilitate the multiple template assembly of LRR proteins, we developed the LRRML database [9], which archives individual LRR structures that were manually identified from all known LRR protein structures. In addition, we have developed TollML [4], a database of sequence motifs of TLRs. In TollML, all known sequences of TLR ectodomains were semi-automatically partitioned into LRR segments and made available for query. For newly sequenced TLRs that are not yet archived in TollML, we have implemented an LRR prediction program named LRRFinder on the TollML webpage. It requires an LRR-containing amino acid sequence as input, and returns the number and positions of the LRRs in the input sequence. LRRFinder recognizes LRR motifs based on a position-specific weight matrix scan, with sensitivity and specificity both higher than 93%. With the help of these two databases, LRR partitions of a TLR ectodomain can be obtained directly, and an optimal structure template for each LRR segment can be quickly found. A schematic flowchart of the modeling procedure is shown in Fig. 1.

In this study, we apply the multiple template assembly approach to TLRs. To demonstrate the potential of the method, we constructed two models of the mouse TLR3 ectodomain as a test case using our LRR template assembly method and a standard profile–profile alignment-aided full-length template recognition method. Both models were then compared with the crystal structure of mouse TLR3. The overall and ligand-binding site conformation of the template assembly-based model is closer to that of the crystal structure than that based on the standard method. We also modeled the human TLR5–10 and mouse TLR11–13 ectodomains, which represent mammalian TLR ectodomains with unknown structures. A comparison of the model for human TLR6 with the very recently reported crystal structure of mouse TLR6 shows a very good structural agreement.

Methods

Template selection and sequence alignment

Amino acid sequences of the mouse TLR3, human TLR5–10, and mouse TLR11–13 ectodomains were extracted from TollML release 3.0 (IDs 627, 531, 571, 992, 575, 1022, 851, 703, 705, and 704). Their LRR partitions were annotated by TollML. For each LRR sequence contained in each target

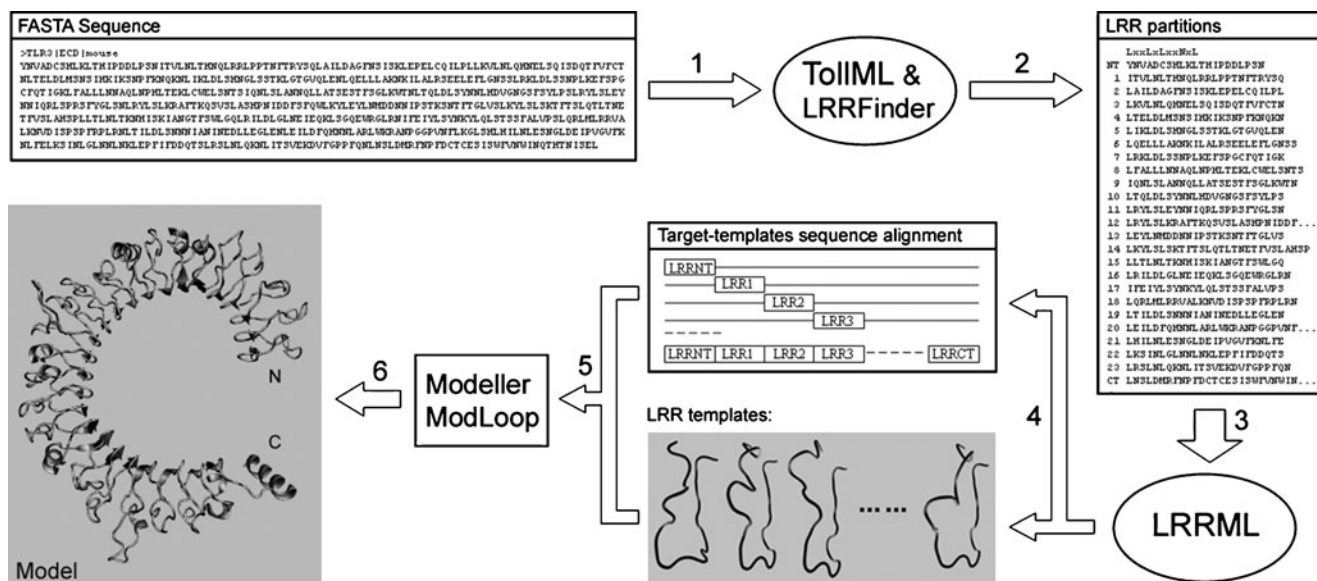


Fig. 1 Flowchart of the LRR template assembly method

TLR, the three-dimensional LRR structure with the highest sequence identity was selected as a template from LRRML through a sequence similarity search. Then, a multiple sequence alignment of a target with all its local LRR templates was generated, with each template comprising one alignment line. For instance, the mouse TLR3 has a total of 25 LRRs and accordingly required 25 templates. The associated multiple sequence alignment then has 26 lines (Fig. 1). Because of the characteristic consensus sequences of LRRs, these alignments were made more accurately manually than automatically. To generate an alternate model with standard methods, the widely acknowledged template recognition program pGenTHREADER [17] was executed to find templates for mouse TLR3. This method calculates sequence profiles from an input sequence and combines profile–profile alignments with secondary structure specific gap-penalties, pair potentials, and solvation potentials using a linear combination. The output is the complete PDB structures that serve as candidate templates ranked by *P* values. Each candidate sequence is aligned with the target sequence.

Model construction and validation

The initial three-dimensional coordinates of all models were calculated by MODELLER 9v7 [18]. The above-described alignment file and the corresponding template structures of the target model were inputted into the default “model” routine of MODELLER. A given number of three-dimensional models were calculated. The ectodomains of TLRs contain a number of insertion regions. Some of them corresponded to four to 15 amino-acid-long gaps in the alignments because their templates do not contain a

corresponding insertion. During modeling, these gaps produced loop structures in the model, thus deteriorating the model accuracy. ModLoop [19] was used to rebuild the coordinates of these loop regions. Finally, we used the model quality assessment programs ProQ [20] and MetaMQAP [21] to evaluate the output candidate models and select the one with the best scores as the final model. The structure superimpositions and molecular electrostatics involved in the structural analysis were carried out using SuperPose v1.0 [22] and VMD [23], respectively. The docking studies of TLR11 and its ligand profilin were performed with GRAMM-X [24].

Results

LRR templates

The full-length ectodomains of mouse TLR3, human TLR5–10, and mouse TLR11–13 contained 25, 23, 21, 28, 28, 28, 21, 26, 25, and 27 LRRs, respectively. Consequently, a total of 252 individual LRR templates sourced from 41 different PDB structures were selected from LRRML. Their sequence identities with the targets varied from 26.0% to 95.7% (43.8% on average), and similarities ranging from 39.0% to 100% (58.2% on average) were obtained. Remarkably, all cases of relatively low sequence identity (<35%) were caused by highly irregular target LRRs. These highly irregular sequences include LRRNT/CTs, the highly mutated LRR15 of TLR7/8/9, and the insertion-containing LRRs, whose templates do not include a similar insertion. The sources (LRRML IDs) and sequence identities of all LRR templates are listed in Table 1.

Table 1 Sources and sequence identities (%) of the LRR templates

LRR	mTLR3		hTLR5		hTLR6		hTLR10		mTLR12	
	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden
NT	406	47.6	125	34.6	333	72.7	150	40.9	125	28.6
1	65	45.8	284	45.8	334	54.2	334	45.8	90	37.5
2	212	45.8	264	42.3	335	70.8	335	62.5	178	58.3
3	465	41.7	259	54.2	336	81.0	578	42.9	465	40.7
4	151	50.0	464	46.2	337	80.0	337	68.0	578	40.9
5	177	50.0	259	48.0	338	60.9	338	47.8	21	35.5
6	110	46.2	252	34.6	339	36.0	111	33.3	497	56.0
7	275	45.8	494	29.0	340	33.3	340	41.7	202	40.9
8	8	41.4	135	37.0	341	33.3	28	34.6	39	34.6
9	203	42.3	359	26.6	342	48.1	342	44.4	356	45.8
10	484	54.2	357	41.7	343	34.6	343	46.4	179	45.8
11	293	50.0	140	50.0	344	48.3	365	29.0	177	42.3
12	270	33.3	316	50.0	345	45.5	345	40.9	566	54.2
13	357	41.7	151	45.8	346	41.7	346	41.7	66	50.0
14	65	42.9	216	45.0	347	61.5	358	48.0	468	45.8
15	152	50.0	153	38.1	348	40.0	369	41.7	501	44.0
16	259	40.0	253	36.0	349	68.2	349	45.5	581	34.8
17	316	37.5	356	40.0	350	95.7	350	43.5	372	30.8
18	152	38.5	566	41.4	351	52.0	372	45.5	283	37.0
19	239	50.0	72	58.3	128	37.5	513	36.0	396	38.7
20	239	45.8	579	45.5	–	–	–	–	260	37.5
21	92	50.0	484	40.9	–	–	–	–	496	41.4
22	80	41.7	–	–	–	–	–	–	315	45.8
23	628	46.2	–	–	–	–	–	–	110	35.7
CT	575	42.9	149	31.7	149	31.7	514	29.9	149	20.0

LRR	hTLR7		hTLR8		hTLR9		mTLR11		mTLR13	
	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden
NT	257	40.6	250	30.0	125	29.0	125	36.7	125	26.5
1	293	45.8	509	45.8	314	41.7	76	38.5	491	54.2
2	107	36.0	573	48.0	101	41.7	191	37.5	107	50.0
3	106	52.4	581	47.6	219	55.0	288	44.0	254	54.2
4	261	45.8	509	45.8	310	41.7	559	36.4	152	50.0
5	140	44.0	564	44.0	157	41.7	557	40.0	4	53.8
6	583	47.6	106	52.4	106	47.6	238	30.4	419	52.0
7	105	45.8	509	45.8	501	41.7	463	44.0	285	44.0
8	573	45.8	573	50.0	259	37.5	271	45.8	135	41.7
9	79	45.8	494	41.7	357	50.0	157	45.8	488	52.0
10	500	46.2	573	42.3	367	37.0	506	44.0	498	36.7
11	500	52.0	316	54.2	337	40.0	237	45.8	128	33.3
12	488	40.7	261	37.0	274	31.0	509	46.2	288	48.0
13	129	44.0	92	37.5	92	33.3	500	54.2	505	45.8
14	386	37.9	259	44.0	357	34.6	7	50.0	573	54.2
15	36	33.3	300	29.6	139	15.6	111	45.8	260	50.0
16	316	41.7	494	41.7	170	37.5	38	42.3	148	36.0
17	358	46.2	158	40.0	316	40.0	484	43.5	262	45.8

Table 1 (continued)

LRR	hTLR7		hTLR8		hTLR9		mTLR11		mTLR13	
	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden	LRRML ID	Iden
18	316	41.7	314	41.7	135	45.8	360	29.2	293	50.0
19	494	50.0	296	37.5	316	44.0	73	37.5	105	45.8
20	503	41.7	128	45.8	493	40.0	403	28.6	260	48.0
21	92	41.7	77	45.8	252	41.7	102	30.4	466	46.2
22	78	48.0	573	60.0	288	44.0	121	34.5	565	41.7
23	552	45.8	169	36.0	111	37.5	316	50.0	93	50.0
24	210	50.0	283	53.8	289	41.7	35	36.0	79	50.0
25	133	40.7	271	45.8	259	45.8	–	–	573	41.7
26	274	48.3	144	42.3	357	52.0	–	–	–	–
CT	149	36.7	149	33.9	149	28.8	469	26.2	149	36.5

As the modeling of mouse TLR3 was carried out to verify our approach, we assumed that the crystal structure of the mouse TLR3 ectodomain was unknown and excluded the corresponding LRR entries of mouse and human TLR3 ectodomains from LRRML before the template search. The selected individual LRR templates for mouse TLR3 were associated with 18 PDB structures, 14 of which were from non-TLR proteins. The target–template sequence identities ranged from 33.3% to 50.0% (44.1% on average). By contrast, pGenTHREADER provided only complete PDB structures of LRR proteins as candidate templates, with each candidate possessing a pairwise sequence alignment with the target. Because no single template covered the entire sequence of mouse TLR3, we selected the first seven candidates by rank (except mouse and human TLR3) and combined them into a multiple alignment to avoid template gaps. These templates included PDB structures 2Z64, 2Z81, 1O6V, 3FXI, 2Z7X, 1JL5, and 3BZ5, which covered the closest homologs (mouse TLR2/4 and human TLR1/4) to mouse TLR3 among all proteins with known structures. Nevertheless, the sequence identities of the seven templates to mouse TLR3 ranged from 16.1% to 21.2% (18.7% on average), which fell much below the cut-off value of 30% for homology modeling [16].

Structural models

Model of mouse TLR3

Recently, we constructed a model of mouse TLR3 with the LRR template assembly method as a test case for the LRRML database [9]. It revealed a horseshoe-shaped assembly that adopts a regular solenoid structure without disordered regions. The model was superimposed on the crystal structure of the mouse TLR3 ectodomain (PDB code: 3CIG) [14] at both of its ligand-binding regions,

LRRNT–LRR3 and LRR19–LRR21. The backbone root mean square deviations (RMSDs) were 1.96 Å and 1.90 Å, respectively [9]. To verify the improvements in the database and the modeling process, we reconstructed the mouse TLR3 model (Fig. 2b) with up-to-date LRR templates. Compared with the old model, four of the 25 LRRs of TLR3 in the new model were assigned new templates with higher sequence identities (Table 1). Because these four LRRs are not involved in the TLR3 ligand-binding sites, the corresponding RMSD values of the new model were the same as for the previous model. These values indicate that the model predicted with our method closely matches the crystal structure and can be used to predict potential ligand-binding sites [25].

For comparison purposes, the mouse TLR3 ectodomain was also modeled with a standard profile–profile alignment-aided full-length template recognition method. All of the ten output models obtained from MODELLER for the full-length templates-based standard alignment showed a serious structural disorder spanning LRR6–10 (Fig. 2a). The LRR6–10 on the crystal structure form a regular solenoid structure with an α -helix in the variable segment of LRR8 (Fig. 2c). By contrast, the corresponding LRRs in the model completely lost the proper LRR shape and were interwoven with one another. This disorder was caused by mismatches or target gaps in the alignment, where only two to four template LRRs were assigned to five target LRRs (Fig. 2d). The standard alignment could not create a one-to-one correspondence between the target and template LRR units due to the irregularity of the LRRs. ProQ and MetaMQAP were used to evaluate the quality of the different models of mouse TLR3 (Table 2). Both programs make an integrative assessment of the structure quality, considering the geometries, stereochemistries, and energy distributions of the structures. Both of the template assembly-based models received better scores than the model based on the standard method.

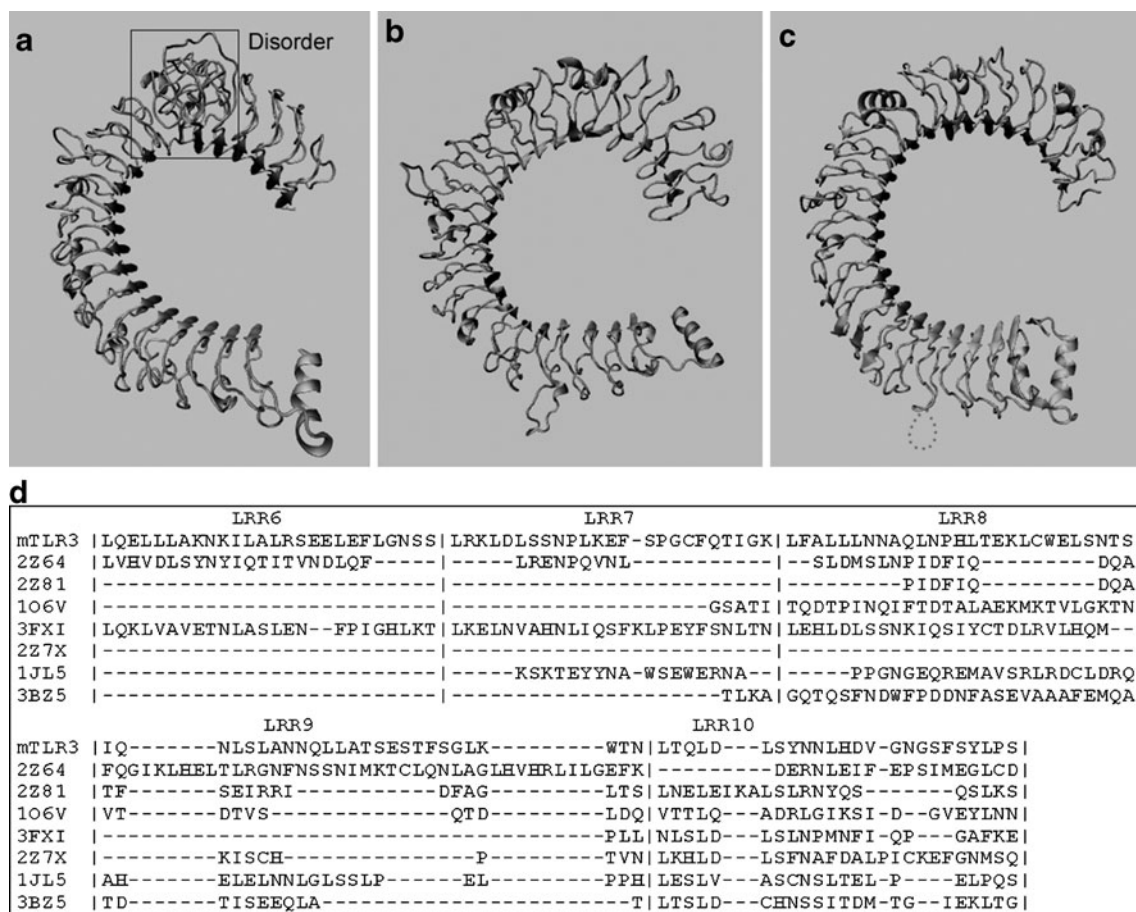


Fig. 2a–d Homology models and crystal structure of the mouse TLR3 ectodomain. **a** The homology model based on the standard method. The framed region exhibits serious disorder. **b** The homology model based on the template assembly method. **c** The crystal structure (PDB code: 3CIG). The dotted region is an insertion on LRR20 that is missing in

the crystal structure. **d** The target-template sequence alignment of the disordered region of the standard method-based model. Mismatches and target gaps resulted in the disorder in **a**, where two to four template LRRs were wrongly assigned to five target LRRs

Models of human TLR5–10 and mouse TLR11–13

Using the LRR template assembly method, we modeled the human TLR5–10 and mouse TLR11–13 ectodomains. All of the resulting models are provided in supplementary file 1. Ramachandran plots of these models were created with PROCHECK [26] and are provided in supplementary file 2. Model evaluation data obtained with ProQ and MetaMQAP are listed in Table 2. The models reveal a horseshoe shape (Fig. 3), where a longer or shorter sequence (more or fewer LRR units) implies a smaller or larger horseshoe opening; e.g., TLR7/8 (smaller opening) and TLR6/10 (larger opening). Their overall structural similarity reflects the phylogenetic relationships among these TLRs. For example, TLR6 is similar to TLR10, while TLR7 is similar to TLR8, consistent with the molecular tree proposed by Roach and co-workers [27]. Mammalian TLRs are distinct from other LRR proteins in that they contain two to seven insertion-containing irregular LRRs, which may be necessary for ligand binding and receptor dimerization. Our

models show that all insertions are located on the face of the horseshoe to which the convex site β -strands point, whereas the other face is completely insertion-free.

To highlight the applicability of these models to analyses of receptor–ligand interaction mechanisms, we performed molecular electrostatics calculations of the mouse TLR11 model with a ligand. TLR11 can recognize the profilin of some apicomplexan protozoan parasites. This protein is involved in parasite motility and invasion [28]. Expression of TLR11 is suppressed in humans [29]. The electrostatic analysis (Fig. 4a) shows that the entire surface of the profilin of *P. falciparum* (PDB code: 2JKF) is predominantly negatively charged, whereas TLR11 exhibits several positively charged patches (Fig. 4b). Protein docking studies using GRAMM-X showed that profilin and the positive patches on TLR11 possess compatible sizes and electrostatic complementarity (supplementary file 3).

Very recently, the crystal structure of the TLR6 ectodomain complexed with TLR2 and a Pam2CSK4 ligand was released in PDB (PDB code: 3A79). The TLR6 structure is

Table 2 Evaluation of the crystal structures and models of the TLRs. Higher ProQ_LG/MS and MetaMQAP_GDT values indicate higher model qualities; higher MetaMQAP_RMSD values indicate lower model qualities

Structure/model	ProQ_LG/MS	MetaMQAP_GDT/RMSD
mTLR3 crystal	7.215/0.469	69.840/1.962
mTLR3 pGenTHREADER	4.013/0.357	41.943/4.691
mTLR3 ^a	5.136/0.423	54.068/3.080
mTLR3 new	5.349/0.405	54.542/3.030
hTLR5	4.707/0.358	54.803/3.126
hTLR6	5.807/0.439	73.230/1.907
hTLR7	4.980/0.381	52.816/3.224
hTLR8	5.053/0.408	53.490/3.113
hTLR9	5.025/0.386	53.774/3.181
hTLR10	4.835/0.362	59.918/2.883
mTLR11	4.371/0.351	49.178/3.433
mTLR12	4.546/0.337	46.131/3.493
mTLR13	4.827/0.407	55.709/2.982

^a Model of Wei et al. [9]

a hybrid structure of mouse and inshore hagfish, where 18 mouse LRRs (LRRNT–LRR17) were hybridized with two hagfish LRRs (LRR18 and LRRCT) [30]. This crystal structure served as an additional benchmark for our template assembly approach. The superimposition of the mouse part of the crystal structure with our human TLR6 model yielded a backbone RMSD of 1.94 Å, which indicates that the model is very similar to the crystal structure (Fig. 5). A second model of human TLR6 was generated with pGenTHREADER in a similar procedure to that described for the mouse TLR3 (supplementary file 1). The backbone RMSD between the crystal structure and this model is 1.89 Å. The only full-length template used for this model was the structure of human TLR1 (PDB code: 2Z7X). Because TLR6 and TLR1 possess the same number of LRRs and have a very high sequence identity (63.3%), the structure of TLR1 serves as an excellent full-length template. Under these very good conditions, both the standard and the template assembly approaches provided high-quality models.

Discussion

In template-based protein modeling, the overall sequence identity between the target and template is an important criterion for the selection of suitable templates [31]. For repetitive LRR proteins, however, an appropriate full-length template is often not available, due to different repeat numbers and distinct arrangements. This problem can be solved by combining individual repeating units that are

locally optimal for the target sequence. In the method validation with mouse TLR3, the average target–template sequence identity achieved by our method was 44.8%, which was significantly higher than that (18.7%) achieved by a standard profile–profile alignment-aided template recognition method. However, both the standard and the template assembly methods produced models of human TLR6 that closely matched the crystal structure of mouse TLR6. A comparison between the models obtained with both methods highlights the potential of the template assembly approach. It can produce models of a similar quality to those obtained with the standard profile–profile alignment method. The template assembly method, however, reveals its particular strength in situations where no adequate full-length templates are available. In the case of TLR3, the standard profile–profile alignment method failed to predict a reliable model due to significant gaps in the template. Here, the template assembly method overcomes the difficulties and generates a realistic model.

In a previous work, we constructed models of human TLR7/8/9 ligand-binding domains by combining LRR segments that were extracted from all known crystal structures of TLRs [32]. The average target-sequence similarities for TLR7/8/9 were 47.7%, 47.2%, and 46.8%, respectively. The resulting models supported experimentally determined ligand-binding residues [33, 34] and provided a reliable basis to identify potential ligand-binding residues and potential receptor dimerization mechanisms. Here, we went a step further and extended the scope of the approach by checking LRR segments from all LRR-containing proteins with known structures in the LRRML database, because the same type of LRR can exist in different proteins [9]. Consequently, 33 of the 41 source PDB structures are non-TLR proteins (numbers derived from Table 1). The average target-sequence similarities for TLR7/8/9 increased to 55.9%, 58.2%, and 59.2%.

Another key issue in LRR protein modeling is the sequence-level LRR partition of the target TLR sequence. The indicated number and beginning/end positions of the LRRs in TLRs vary a great deal across different databases or research reports, due to the irregularity and periodicity of the LRRs. TollML reports the most complete and accurate LRR motifs for TLRs as compared with a number of databases [4]. In addition, TollML provides a statistics-based LRR prediction program, LRRFinder, for new TLR entries that are not yet collected in TollML. It can recognize LRRs from an input amino acid sequence with high confidence.

Conclusions

In conclusion, this work depicts an LRR template assembly approach to protein homology modeling. The comparison

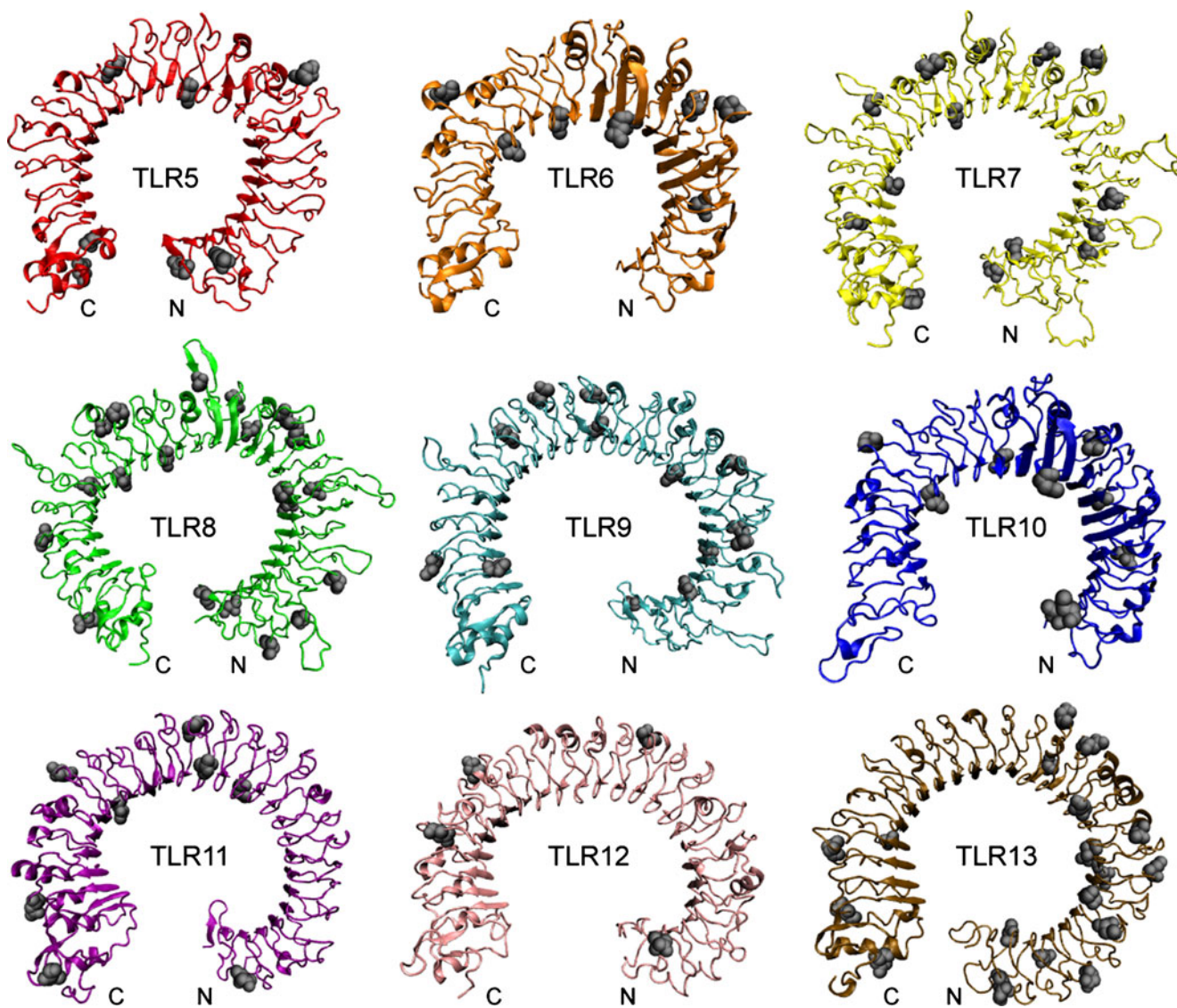


Fig. 3 Models of the human TLR5–10 and mouse TLR11–13 ectodomains. The N-linked glycan sites of these TLRs were obtained from the NCBI protein database and are labeled with *black balls*

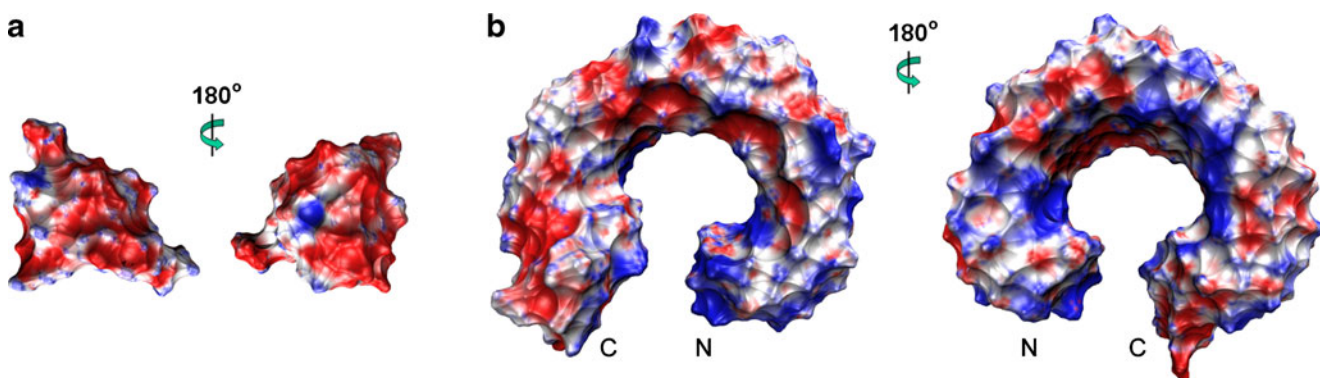


Fig. 4a–b Surface charge analysis (APBS electrostatics) of **a** the crystal structures of profilin (PDB code: 2JKF) and **b** the model of the mouse TLR11 ectodomain. *Blue*, positive charge; *red*, negative charge

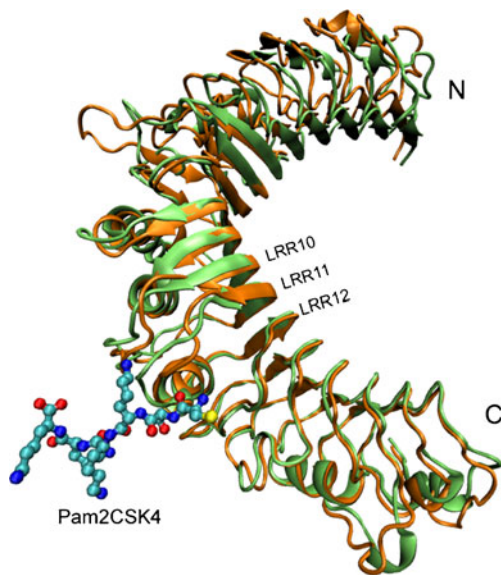


Fig. 5 Superimposition of the homology model and the crystal structure of the TLR6 ectodomain. *Green*, homology model of human TLR6 (LRRNT–LRR17); *orange*, crystal structure of mouse TLR6 (LRRNT–LRR17). The Pam2CSK4 ligand-binding site is located on the variable parts of LRR10–12. The overall backbone RMSD is 1.94 Å and the backbone RMSD of the ligand-binding region is 1.18 Å

of a mouse TLR3 model with its crystal structure underlined the feasibility and reliability of the method. With this method, a series of mammalian TLR ectodomains were modeled. These models can be used to perform ligand docking studies or to design mutagenesis experiments, and hence to investigate TLR ligand-binding mechanisms. Our modeling approach can be extended to other repetitive proteins.

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

References

- Kawai T, Akira S (2008) Toll-like receptor and RIG-I-like receptor signaling. *Ann N Y Acad Sci* 1143:1–20
- Medzhitov R (2007) Recognition of microorganisms and activation of the immune response. *Nature* 449:819–826
- Kumar H, Kawai T, Akira S (2009) Pathogen recognition in the innate immune response. *Biochem J* 420:1–16
- Gong J, Wei T, Zhang N, Jamitzky F, Heckl WM, Rössle SC, Stark RW (2010) TollML: a database of Toll-like receptor structural motifs. *J Mol Model* (in press)
- Brodsky I, Medzhitov R (2007) Two modes of ligand recognition by TLRs. *Cell* 130:979–981
- West AP, Koblansky AA, Ghosh S (2006) Recognition and signaling by toll-like receptors. *Annu Rev Cell Dev Biol* 22:409–437

- Akira S, Hemmi H (2003) Recognition of pathogen-associated molecular patterns by TLR family. *Immunol Lett* 85:85–95
- Leulier F, Lemaitre B (2008) Toll-like receptors-taking an evolutionary approach. *Nat Rev Genet* 9:165–178
- Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC (2008) LRRML: a conformational database and an XML description of leucine-rich repeats (LRRs). *BMC Struct Biol* 8:47
- 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
- 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
- Choe J, Kelker MS, Wilson IA (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* 309:581–585
- 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 U S A* 102:10976–10980
- 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
- Kim HM, Park BS, Kim JI, Kim SE, Lee J, Oh SC, Enkhbayar P, Matushima 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
- Baker D, Sali A (2001) Protein structure prediction and structural genomics. *Science* 294:93–96
- Lobley A, Sadowski MI, Jones DT (2009) PGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* 25:1761–1767
- Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9:1753–1773
- Fiser A, Sali A (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 19:2500–2501
- Wallner B, Elofsson A (2003) Can correct protein models be identified? *Protein Sci* 12:1073–1086
- Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM (2008) MetaMQAP: a meta-server for the quality assessment of protein models. *BMC Bioinformatics* 9:403
- Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* 32:W590–W594
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(33–38):27–38
- Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res* 34:W310–W314
- Kopp J, Bordoli L, Battey JN, Kiefer F, Schwede T (2007) Assessment of CASP7 predictions for template-based modeling targets. *Proteins* 69(Suppl 8):38–56
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291
- Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A (2005) The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* 102:9577–9582
- Yarovinsky F, Zhang D, Andersen JF, Bannenberg GL, Serhan CN, Hayden MS, Hieny S, Sutterwala FS, Flavell RA, Ghosh S, Sher A (2005) TLR11 activation of dendritic cells by a protozoan profilin-like protein. *Science* 308:1626–1629
- Lauw FN, Caffrey DR, Golenbock DT (2005) Of mice and man: TLR11 (finally) finds profilin. *Trends Immunol* 26:509–511
- Kang JY, Nan X, Jin MS, Youn SJ, Ryu YH, Mah S, Han SH, Lee H, Paik SG, Lee JO (2009) Recognition of lipopeptide

- patterns by Toll-like receptor 2–Toll-like receptor 6 heterodimer. *Immunity* 31:873–884
31. Kopp J, Schwede T (2004) Automated protein structure homology modeling: a progress report. *Pharmacogenomics* 5:405–416
 32. Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC (2009) Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. *Protein Sci* 18:1684–1691
 33. Rutz M, Metzger J, Gellert T, Luppa P, Lipford GB, Wagner H, Bauer S (2004) Toll-like receptor 9 binds single-stranded CpG-DNA in a sequence- and pH-dependent manner. *Eur J Immunol* 34:2541–2550
 34. Gibbard RJ, Morley PJ, Gay NJ (2006) Conserved features in the extracellular domain of human toll-like receptor 8 are essential for pH-dependent signaling. *J Biol Chem* 281:27503–27511