Divergence of the Individual repeats in the leucine-rich repeat domains of human Toll-like receptors explain their diversity and functional adaptations.

Abraham Takkouche, Keita Ichii, Xinru Qui, Lukasz Jaroszewski and Adam Godzik *

Division of Biomedical Sciences, University of California Riverside School of Medicine, Riverside, CA 92521, USA

*Author to whom correspondence should be addressed at adam.godzik@medsch.ucr.edu

Abstract

The receptor domains of Toll-like receptors (TLRs) are characterized by a solenoid-like structure composed of tandem repeats of α/β units known as Leucine Rich Repeats (LRRs). LRR proteins form large paralogous families, with nearly 400 in the human genome alone, all sharing similar semi-regular solenoid-like structures. Despite this structural similarity, they exhibit remarkable diversity in binding specificity. For TLR receptors, this includes a range of pathogen-associated molecular patterns (PAMPs), while other LRR proteins bind an extensive array of ligands, including proteins, DNA, RNA, and small molecules. The LRR domains contain repeats that have similar, yet not identical, 3D structures and patterns of conserved residues. Through in-depth analysis of sequence and structural conservation in individual repeats of human TLRs, we demonstrate that even subtle variations between these repeats alter the local solenoid structure, leading to significant functional changes. Variations in repeat length and defining patterns result in local changes in curvature and the emergence of structural features such as loops, cavities, or specific interaction interfaces. Understanding how divergence in LRR repeats influences their functional roles can provide deeper insights into their binding mechanisms, including interactions with unknown ligands, both in humans and across the diverse phylogenetic spectrum of animals that rely on their TLR repertoires for immune activation.

Introduction

Activation of innate immunity is a critical first step in a defense against invading pathogens. This is facilitated by Pattern Recognition Receptors (PRRs) (1), which recognize Pathogen-Associated Molecular Patterns (PAMPs) and damage-associated molecular patterns (DAMPs) - specific molecular signatures of pathogens and other conserved molecular patterns associated with danger to the cell. Activated receptors initiate a signaling cascade that activates networks of cytokine processing, autophagy, cell death, and phagocytosis (2-5). Named after a Drosophila Toll protein, TLRs consist of three domains: the N-terminal leucine-rich repeat (LRR) domain, a single helix transmembrane domain, and a C-terminal TIR (toll-IL-1 receptor) domain.

The LRR domains of TLRs, composed of multiple repeats with a similar sequence and structure pattern, belong to the class of tandem repeat proteins and specifically to a the α/β solenoids of the RI subgroup, members of class III of tandem repeat proteins (6). LRR domains of TLRs directly or indirectly recognize LPS, flagella and other PAMPs and, through their TIR domains, pass on the signal to the downstream pathways. Toll-like immune receptors are found in all kingdoms of life, but interestingly, they have evolved different strategies for recognizing pathogens. Plants and many invertebrates employ thousands (R-proteins in plants) or hundreds (TLRs in invertebrates) of receptors, most specifically tailored to recognize a single pathogen (7). In contrast, 10-20 TLR receptors in vertebrates can recognize a broad spectrum of pathogens, with many of them being able to bind and be activated by diverse set of both endo- (8) and exo-genous (9) ligands.

Analyses of and comparisons between tandem repeat proteins are difficult because of their periodic structure. With the translational symmetry, both structure and sequence are still similar if the alignment shifts by the multiplicity of the length of the repeat; the periodicity signal competes with the similarity of sequence, so two proteins with similar repeat lengths may appear to be more similar than a close homolog with inserts or deletions in the repeats. This affects phylogeny analysis and leads to many misnamed proteins, where, for instance, apparent orthologs of vertebrate TLRs are identified in invertebrates while more careful phylogenetic analysis shows that they are a part of species-specific expansion (10).

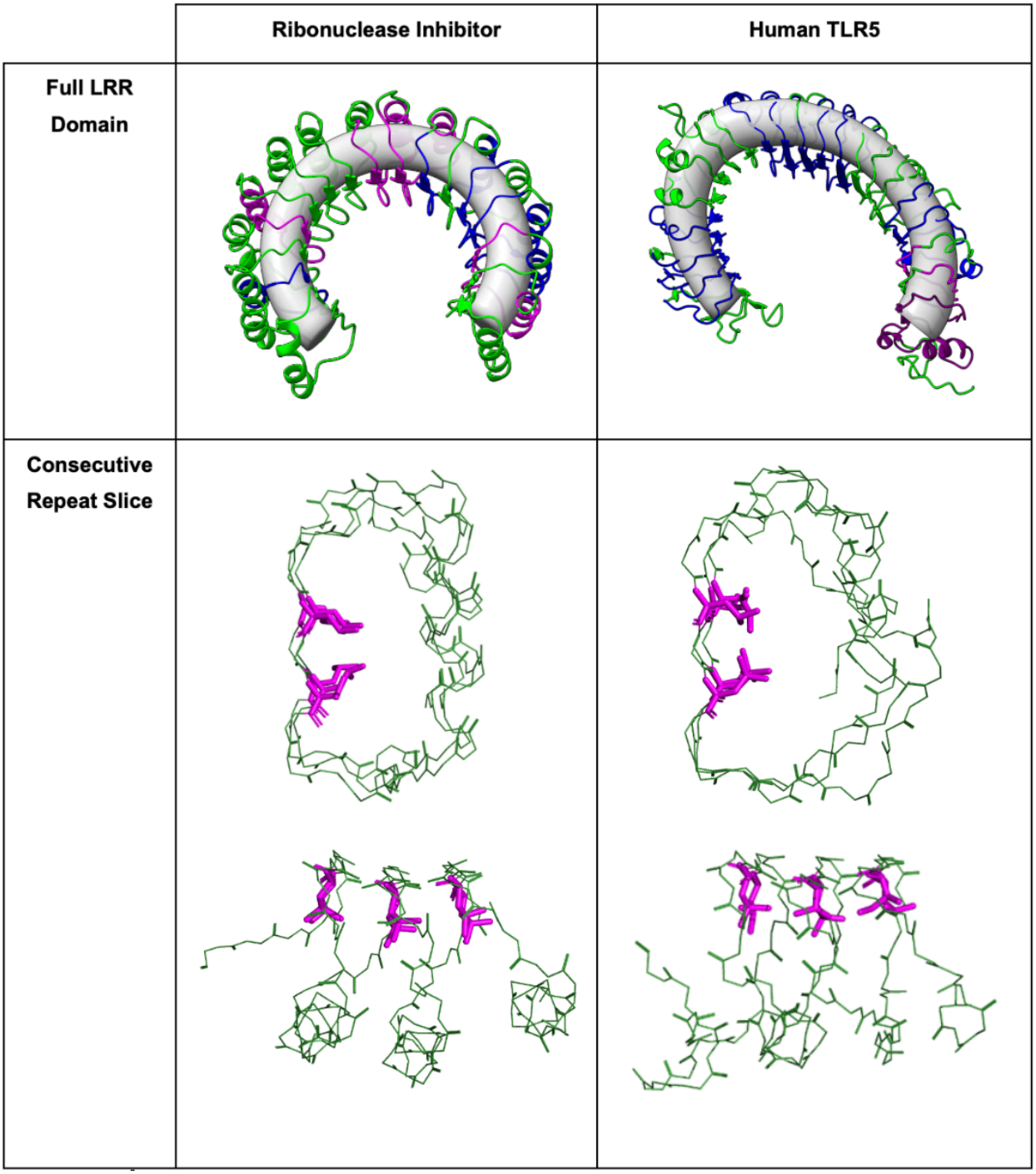

Figure 1: Examples of the regular and irregular LRR domains.

Left Panels: An example of a regular LRR domain, the porcine ribonuclease inhibitor (PDB:2BNH), the first LRR protein with an experimentally solved structure. The LRR domain of the RNI structure is superimposed over a 3D model of a regular torus. The repeats recognized by

an Interpro/Pfam HMM models are highlighted in blue and magenta which are used to characterize different subtypes of LRR repeats – however the classification does not correlate with structural similarity between repeats. 8 repeats out of 17 repeats are recognized by InterPro/Pfam HMMs. The alignment of three consecutive repeats (86-141) with the central LxL pattern are highlighted on the bottom panel.

Right panels: Analogous figures for an example of highly irregular LRR domain of hTLR5 (PDB:3J0A). The LRR domain of hTLR5 LRR is superimposed with an ellipsoid based torus which shows a much greater radius and local curvature divergence than the torus used in RI. The repeats recognized by an Interpro/Pfam HMM models are highlighted in blue and magenta with different colors used to identify different subtypes of LRR repeats – however the classification does not correlate with structural similarity between repeats. Only 7 out of 21 repeats are recognized by InterPro/Pfam HMMs in hTLR5. The repeats identified often overlap in position across LRR subtypes and are inconsistent in length ranging from 19-59 with a mean of 30.5. For reference, manual length analysis showed a range of 20-36 with a mean of 25.9. The alignment of three consecutive repeats (498-560) with the central LxL pattern are highlighted on the bottom panel.

Many algorithms (11), including some developed in our group (12), focus on identifying repeats in tandem repeat proteins. The algorithms based on the analysis of the 3D structure provide the ultimate answer and some of the more recent, AI based algorithms can match them in their ability to identify all the repeats based on the sequence alone (13, 14). In this contribution however we focus on another, rarely discussed feature of tandem repeat proteins – the divergence of the repeats and its consequences. In tandem repeat proteins, after their initial emergence from identical copies of a single repeat, each repeat evolved under individual, function-driven constraints, leading to their sometimes-extreme divergence. This divergence is most visible on the sequence level. As shown in Figure 1, for the hTLR5 InterPro/Pfam HMMs (15) needed 2, out of 16 defined in the database, different HMM models to recognize individual repeats, and even then, more than ½ of the repeats in most LRR proteins are still not recognized. These levels of divergence lead to arguments that in some proteins proper repeats are interrupted by non-homologous sections (16). While this is true for some LRR domains, especially in bacteria and invertebrates, those from the RI class analyzed here have continuous central beta-sheets with conserved hydrogen-bond patterns and recognizable pattern across all repeats, suggesting their evolutionary relationship. The structural diversity of individual repeats is quite subtle and is missed

by the protein structure comparisons focusing on global similarity measures such as RMSD or Tm-score (17). Such similarity measures show most LRR proteins to align with each other with RMSDs below 3Å and high Tm-score, indicative of “strong” structural similarity. However, these apparently minor structural differences are critical for the functional diversity of the TLR LRR domains. The focus of this manuscript is the description and the analysis of the differences between repeats in tandem repeat proteins on the sequence, structure level, and function level. In particular, we focus on a correlation between sequence divergence, the local curvature of the concave side of the solenoid, and the presence of functional motifs, such as extra-long loops, binding sites or cavities, sometimes referred to as “embellishments” of the core fold (18). As we will show in this manuscript, mapping and analyzing these differences is critical for understanding the functions of individual tandem repeat proteins and their differences.

Methods

Consensus patterns for different LRR protein subtypes were defined (19), focusing primarily on the namesake conserved leucine pattern, with the central LxL leucines forming a stacked “ladder” stabilizing the central concave beta-sheet. Much less attention was paid to the variations between the repeats within individual proteins. The presence of some other semi-conserved leucines and other hydrophobic residues was noted, but the occasional breaks in the pattern were usually described as random divergence. Here, we argue that changes in this pattern contribute to the nuances of the local structure of the LRR domain, and the repeats breaking the pattern are critical to the specific functions of the individual TLR and other LRR proteins. The consensus sequence pattern for an RI-like class of LRR proteins was defined as LxxLxLxxN/CxL, sometimes including extending it toward the C-terminus by a minor pattern oxxLxxoL (19).

<u>Here, we focus on the following features of individual repeats:</u>

- **Length of a repeat**. To uniquely define repeats’ lengths, we counted the lengths from the first amino acid in the LxL pattern to the equivalent residue in the following repeat, even if in more divergent repeats this residue was replaced by another, typically hydrophobic, residue.
- **Conservation of the consensus pattern**. Number conserved residues out of 5 in the LxxLxLxxN/CxL pattern

- **The local curvature of the inside of the solenoid**. We propose calculating curvature by placing a circle going through Cα atoms of a residues defining a start of three consecutive repeats. We have also tested using Cβ atoms or taking ± 2 repeats or fitting a circle to 5 consecutive repeats, all leading to comparable results (data not shown). A circumradius or its inverse (Menger curvature) was used to define the local curvature (20).
- **Presence of extra-long loops** extending beyond the main solenoid body
- **Presence of cavities** inside the solenoid

The central "LxL" pattern was identified for each structure using a combination of a regex-based pattern search and the solenoid structure repeat searching algorithm ConSole (12) and confirmed through manual analysis with a molecular visualization software, PyMol. The entire sequence of the TLR was parsed into individual repeats based on these positions.

The 3-dimensional coordinates of the structure were obtained from the PDB database (21), or in case of the models, from the AlphaFold database (22) or were calculated locally using a public version of the AlphaFold2 software (23).

To validate our method for measuring local curvature, we examined the structure of ribonuclease inhibitor (RI) (PDB:2BNH) as a baseline, regarded as a classical RN-type LRR in terms of both sequence and structure (24, 25). As expected, the local radius was constant across the entire length of the protein. At the same time, the pattern was conserved in all repeats and their length was equal to an alternating pattern of 30/31.

There are no experimental structures for human TLR6, TLR9 and TLR10. To have a complete set of structures for the analysis, we used AlphaFold2 predicted structures of these proteins. To test the quality of these models to reproduce local structural features we performed a series of comparisons between the local curvature of experimental models versus these AI-predicted structures for the several RI-like classes of LRR domains with experimentally determined structures which showed very little divergence. In a comparison of the experimental and AlphaFold2 models of TLR1 for instance, the two local curvature datasets were found to have no significant difference with a p-value of 0.847.

Results

Describing and measuring LRR segment divergence:

Our analysis focuses on the divergence between individual repeats within a single TLR protein. The first sign of the divergence is the varying length of the individual repeats. The standard deviations of repeat lengths are in the TLRs are between 4 to twenty times higher than in the RI protein, being the highest in the most divergent TLRs (TLR7/8/9) due to their long Z loops. On average, the repeat percent sequence identity is lower within the TLR proteins than the RI; however, the difference is not statistically significant. Excess length leads to the emergence of "embellishments" - additional short loops or hairpins that break outside of the torus outline. Such elements often bind secondary ligands or participate in the dimerization interfaces, as would be discussed in the next section.

Identification	Mean Repeat Length	Median Repeat Length	Repeat Length Standard Deviation	Mean Repeat Percent Identity	Total Number of Repeats	Mean Number of Conserved Sequences
Ribonuclease Inhibitor	29.5	29.5	0.5	31.84 ± 11.85	14	5/5
TLR1	25.82	25	2.81	18.5 ± 5.44	17	3.76/5
TLR2	25.79	25	2.50	20.55 ± 5.21	19	4.47/5
TLR3	23.39	25	2.46	24.15 ± 6.00	23	4.61/5
TLR4	25.52	25	1.99	22.53 ± 6.7	21	4.29/5
TLR5	25.90	25	3.32	21.49 ± 6.23	21	4.33/5
TLR6	25.65	25	2.54	19.20 ± 5.62	17	3.94/5
TLR7	29.32	25	10.09	23.43 ± 5.80	25	4.60/5
TLR8	28.96	25	8.53	24.44 ± 6.86	25	4.56/5
TLR9	29.09	26	8.01	24.56 ± 6.97	21	4.48/5
TLR10	25.42	25	2.39	18.34 ± 5.22	19	4.05/5
CD180	25.48	25	1.43	20.81 ± 5.13	21	4.38/5

Table 1: Overview of repeat divergence in human TLR 1-10 compared against the canonical LRR protein, the ribonuclease inhibitor (PDB:2BNH). The CD180 (PDB:3B2B) protein, a paralog of TLR4 was added for comparison. The start and end of repeats were defined by the first conserved leucine position within the "LxL" consensus sequence as discussed in the Methods section. TLRs colored according to their family.

Local Curvature Analysis

As mentioned in the introduction, the human genome codes for 10 TLR proteins, while other mammals, such as mice, may have up to 13. Figure 2 shows cartoon figures of the structures of the LRR domain of all the 10 human TLRs, including 3 AF2 models, with the local curvature of the internal beta sheet shown and the length of the consecutive repeats in the graphs below each.

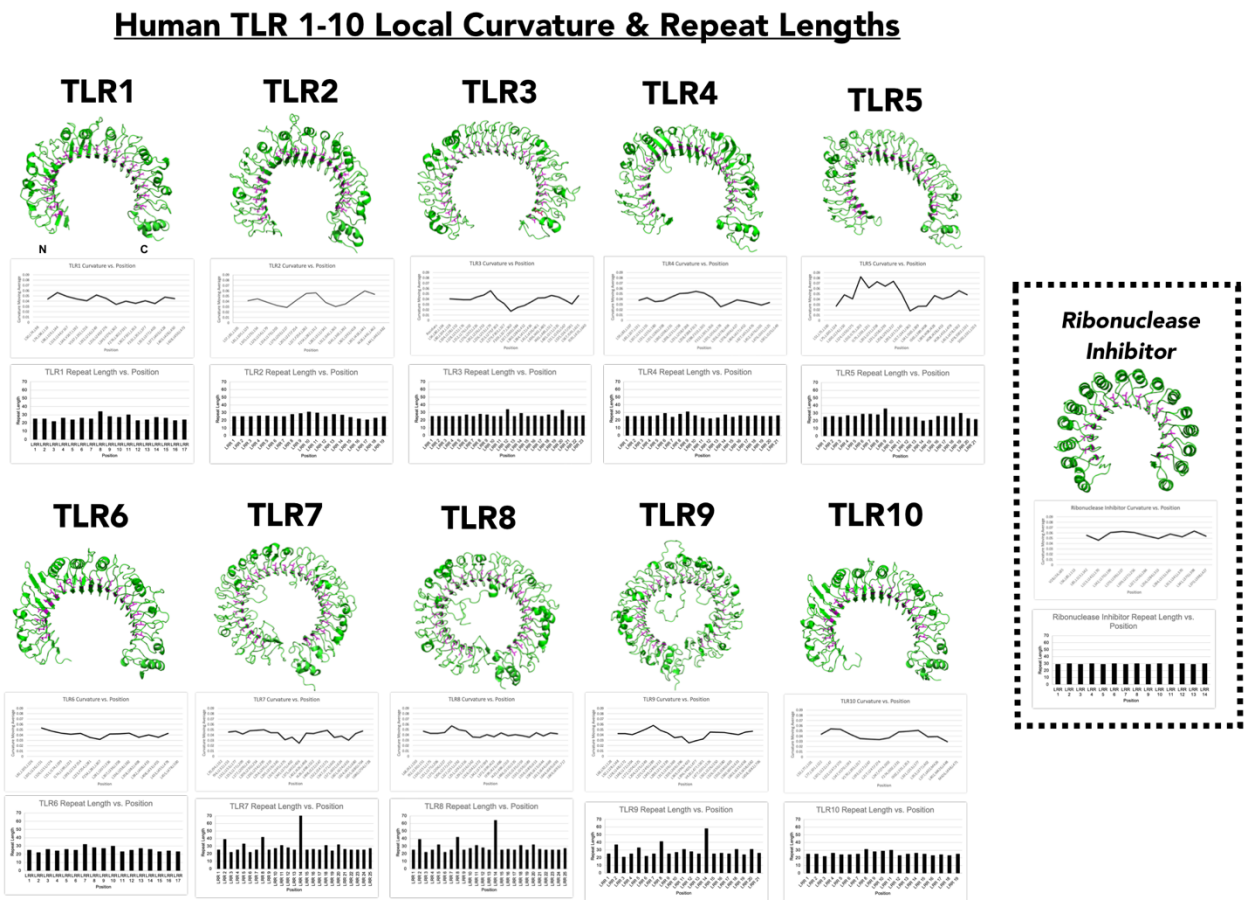

Figure 2: Thumbnail figures of the three-dimensional structures of human TLR 1-10, illustrating the distortions from the ideal solenoid structure (RNI). Local curvature moving averages (interval of three, see the Methods section) and repeat lengths are shown the boxes below. Full data available in the supplemental materials. AlphaFold models of the structures of hTLR5, hTLR6, hTLR9 and hTLR10 are shown, as no experimental structures were available as of this writing.

Individual TLR 1-10 local curvature Analysis

As Figure 2 shows, all the structures of the LRR domains of human TLRs can be described as curved solenoids or tauruses. Their global structure could be described by the radius of the torus and its level of completeness toward a full circle. At the same time, even casual observation reveals significant deviations from the ideal torus shape, with the local curvature of the central beta-sheet different in different sections of the LRR domain. This is captured by the local curvature introduced in the Methods section and shown in the graphs below each thumbnail in Figure 2. Interestingly, regions with locally modified curvatures are not located in the same region in different TLRs. In TLR2,3, and 9, they are located approximately in the middle of the protein, in TLR10 at both ends and in TLR5 toward the N-terminus. TLR1, 6, 7, and 8 contain long regions with no significant curvature deviations.

Individual LRRs exhibit such high divergence on the sequence level, that some authors suggested that these are non-LRR inserts into the LRR domains (16). This is true for some LRR proteins, especially in bacteria or invertebrates. However, in hTLRs, despite local deviations of the curvature on the concave side of the beta-sheets, the presence of non-canonical residues at some of the usually conserved positions, and sometimes extreme variations in length, the overall LRR signature pattern is conserved and the network of hydrogen bonds in the central beta-sheet is continuous. More careful sequence and phylogenetic analysis, especially between more and less divergent paralogs, such as TLR1 and TLR2, or TLR4 and CD180, suggests that the “unusual” repeats arose by divergence rather than non-homologous inserts.

Functional consequences of the repeat divergence.

The divergence of individual repeats is obviously connected to the specific functions of individual TLRs. In a first approximation, we can classify the repeats in each TLR into “structural” repeats with the classical RI class pattern, constant length, and local radius vs. “functional” repeats that show significant departure from the classical pattern and varied length and local curvature. Below, we will discuss the specific structural/functional adaptation on the examples from each TLR subfamily and show how it connects to the LRR repeat divergence. The vertebrate TLRs are typically divided into 6 subfamilies, differing in their functions and structural features (26). Five of these families are represented in the human genome and were the subject of the analysis as described above.

TLR1/2/6/10 family: Special structural feature - open cavities in TLR2 driven by the divergence in the repeats LLR9 to LLR12. Interestingly, no cavities are present in TLR1/10 despite strong sequence similarity to TLR2/6.

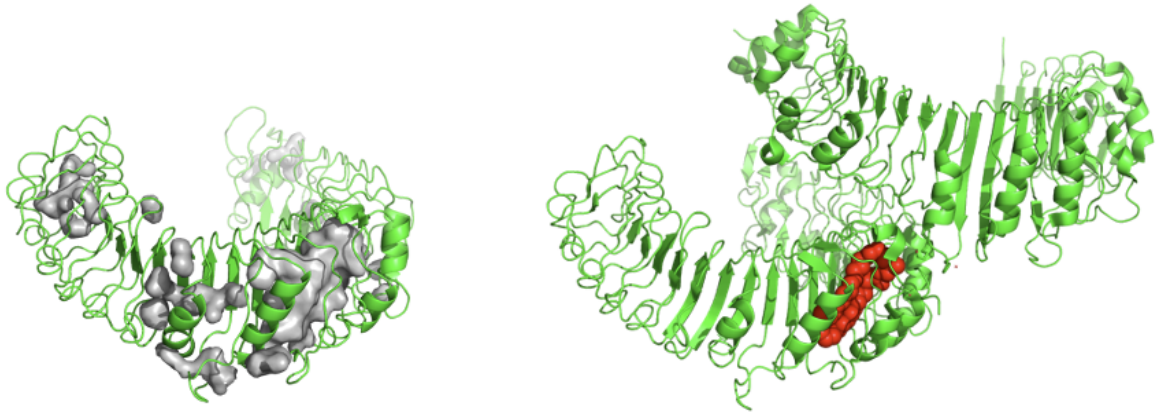

Figure 3: Cavities in hTLR2 (left panel) and their role in lipoproteins binding and creating the dimerization interface in the TLR1/2 heterodimer (PDB ID 2Z7X, chain A (left), full dimer with tri-acylated lipopeptide (right). LRR repeats forming the opening of the cavity (LRR12) and its sides (LRR9-10) are the longest and most divergent of all the repeats present and they form a kink in the internal radius of the solenoid. The conserved positions between LRR9 and LRR12 are almost entirely occupied by amino acids with similar properties to the **L**x**L**xx**N**x**L**xx**L** pattern but not precise matches.

TLR3: Endosomal receptor recognizes viral nucleic acid. Special structural feature - nucleic acid interaction interface formed by loops in repeats.

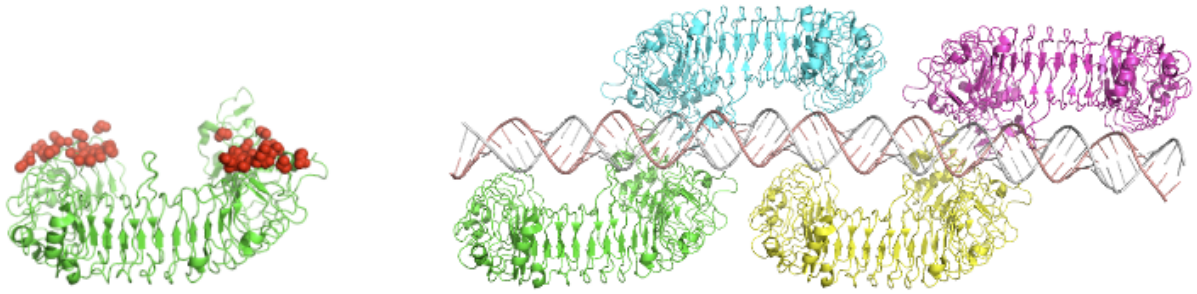

Figure 4: TLR3 double-stranded RNA binding interface (red spheres in the left panel). Most of the interaction driving residues are located on LRR20, the longest and most divergent repeat in hTLR3.

TLR4: Recognizes LPS by binding to the LPS loaded MD2 protein. Special structural feature - conserved MD2 interaction and dimerization interfaces, loop interacting with LPS filled cavity in MD2.

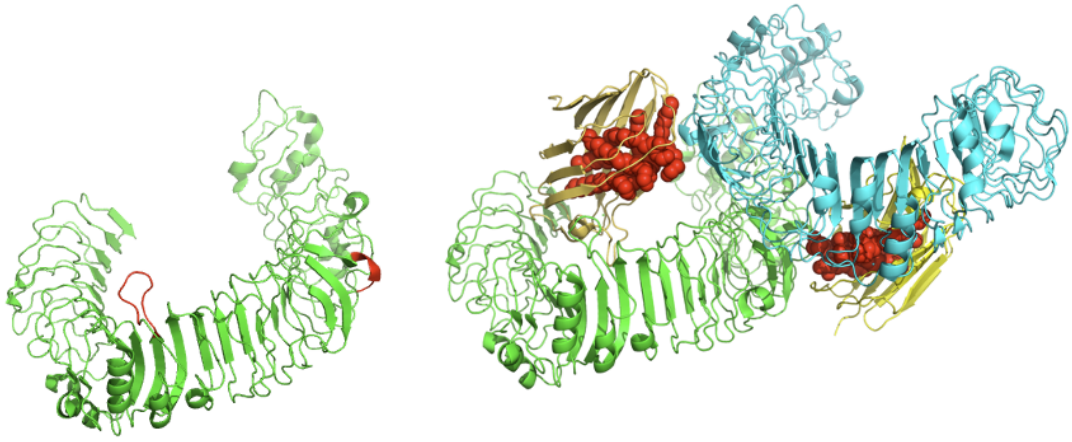

Figure 5: Functional loops in hTLR4 (left panel) and their role in ligand binding and creating the dimerization interface in the TLR4/MD2 dimer of dimers. LRR repeat with the loop stabilizing TLR4/MD2 interface (LRR9) and one with the loop providing the “lid” to the LPS binding cavity in MD2 (LRR17) are the longest and the most divergent of all the repeats, respectively.

TLR5:

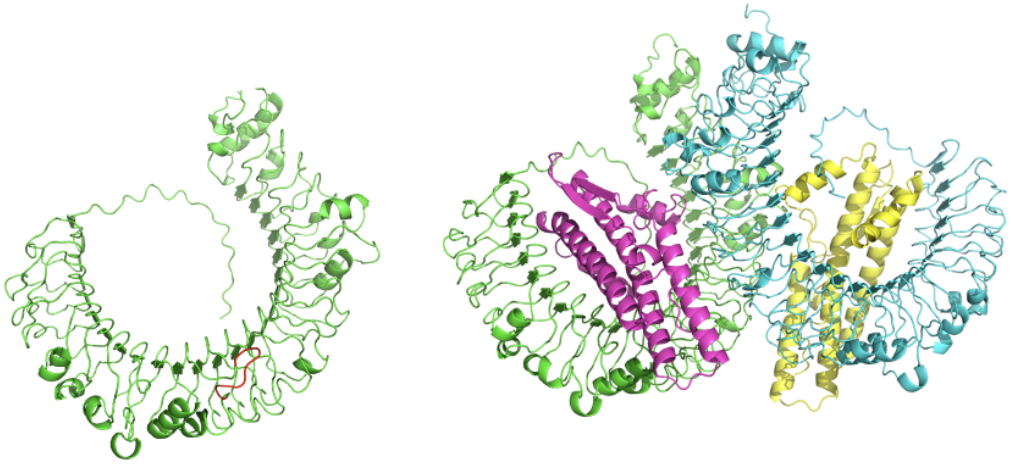

Figure 6: TLR5 binds bacterial flagellin. Conserved interaction interfaces with flagellin and critical loop forming TLR5 homodimer interface upon binding to flagellin.

TLR 7/8/9: Endosomal, recognize viral nucleic acid. Special structural feature - long loop (Z loop) in the LRR14 position in TLRs 7/8/9 contributing to this subfamily's large repeat length standard deviation. xxx

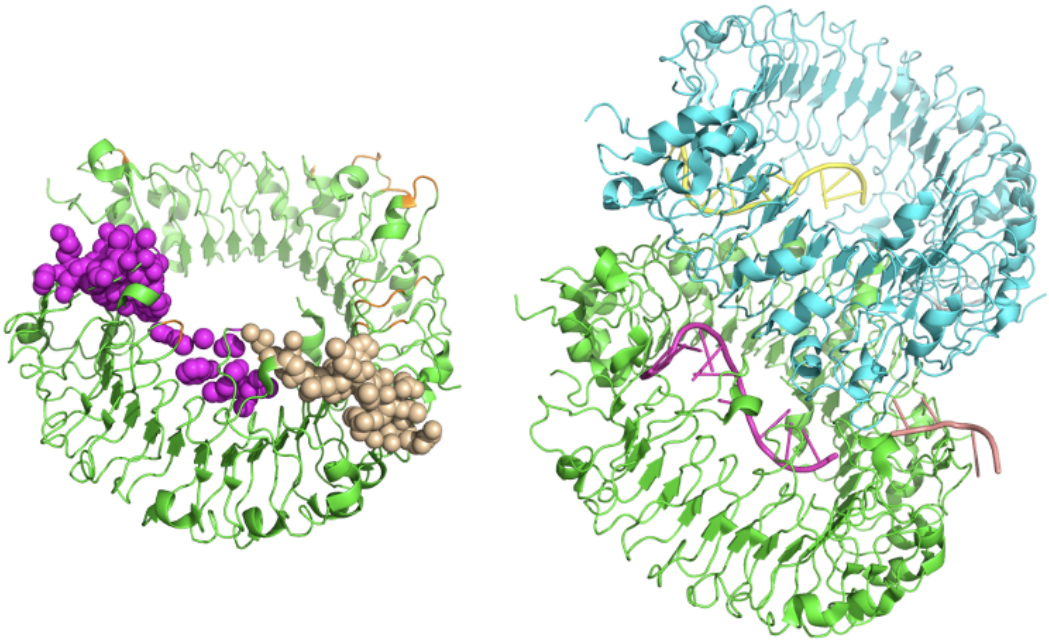

Figure 7: two conserved interaction interfaces in hTLR7, one (magenta), includes the Z-loop binding the single strand RNA and contributing to the homodimer interface. The second interface binding secondary ligands

Discussion

The results of our study highlight the significant structural diversity between the individual repeats in the leucine-rich repeat (LRR) domains of human toll-like receptor (hTLRs) proteins. This diversity between the repeats in individual TLRs is evident on the sequence level in the varying lengths and sequence similarity between individual repeats within each TLR protein. We show that the standard deviation of repeat lengths in TLRs and the level of conservation of the signature sequence pattern is substantially higher than in regular LRR proteins like ribonuclease inhibitor (RNI) or even paralogs in individual TLRs, such as TLR4 paralog, CD180. The high sequence divergence observed in the LRRs of TLRs is driving their functional specialization. At the same time, structural differences between different TLRs not captured well by the global similarity measures such as RMSD or TM-score. For instance, the TM-score between TLR4 and TLR5, which have different binding mechanisms and functions, have an RMSD of 0.5Å demonstrating that global similarity measures do not tell the full story. The functional consequences of variability in repeat length and sequence can only be fully appreciated by analyzing the structural diversity of the repeats. Despite the presence of non-canonical residues and variations in length, the overall LRR signature pattern and the hydrogen bond network in the central beta-sheet remain

conserved. This conservation suggests that the sequence divergence has evolved to enhance functional specificity without compromising the structural integrity of the LRR domain. Variations in the local curvature and structural embellishments such as loops and cavities contribute to what we propose to call function-defining regions in the TLRs, which are involved in ligand binding or dimerization interfaces and are essential for the specialized functions of individual TLRs.

The local curvature of the LRR domains, which we measured by fitting circles to sets of three consecutive residues, provides insight into the structural adaptations of TLRs. Our findings reveal that regions with modified curvature are positioned differently in individual TLRs. For instance, TLR2, TLR3, and TLR9 exhibit locally modified curvature in the middle of the protein, while TLR10 shows modifications at both ends, and TLR5 at the N-terminus. The functional consequences of changes in the canonical pattern on LRRs or effects of inserts and the resulting changes in the local curvature or specific structural features are evident in the adaptations of different TLRs. To their functions. For example, the cavities in TLR2, driven by divergence in repeats LRR9 to LRR12, play a crucial role in ligand binding and dimerization with TLR1/6. Similarly, the nucleic acid interaction interface in TLR3 is formed by loops in the highly divergent repeats. These findings illustrate the importance of understanding the relationship between sequence divergence, local structural changes, and functional specialization of individual TLRs.

Our study provides a comprehensive analysis of the structural and functional diversity within the LRR domains of the human TLRs. The observed variations in repeat length, sequence conservation, and its consequences in local structural features such as curvature and extra loops or cavities highlight the intricate mechanisms by which TLRs have evolved to recognize a wide array of pathogen-associated molecular patterns. This knowledge enhances our understanding of TLR-mediated innate immunity and opens avenues for developing targeted therapeutics that modulate TLR activity.

The LRR domains are not unique to the immune receptors; LRR is a motif found in hundreds of thousands of proteins from all kingdoms of life, typically serving as a ligand binding or protein-protein interaction domain in diverse pathways (19). The human genome codes for about 400 LRR-containing proteins active in different processes, including immune response, apoptosis, autophagy, ubiquitin-related processes, nuclear mRNA transport, and neuronal development (27). Toll-like receptors are also widely distributed in all kingdoms of life. Direct orthologous lines of the human TLRs can be traced to early vertebrates, but large repertoires of TLRs are also present in most invertebrates, where they mostly emerged by lineage specific expansions.

Interestingly, also large families of proteins involved in plant equivalent of innate immunity, the R proteins, have receptor domains similar to the LRR domains of TLRs. As the binding specificities are likely to at least partly overlap between all the groups of TLRs, it is possible that similar function driven solutions have emerged by parallel evolution, which makes analyses like presented here applicable to a much larger sets of proteins.

References

1. N. W. Palm, R. Medzhitov, Pattern recognition receptors and control of adaptive immunity. *Immunol Rev* **227**, 221-233 (2009).
2. T. Duan, Y. Du, C. Xing, H. Y. Wang, R. F. Wang, Toll-Like Receptor Signaling and Its Role in Cell-Mediated Immunity. *Front Immunol* **13**, 812774 (2022).
3. M. A. Delgado, R. A. Elmaoued, A. S. Davis, G. Kyei, V. Deretic, Toll-like receptors control autophagy. *EMBO J* **27**, 1110-1121 (2008).
4. Y. Yang *et al.*, Toll-like receptors: Triggers of regulated cell death and promising targets for cancer therapy. *Immunol Lett* **223**, 1-9 (2020).
5. M. A. Sanjuan *et al.*, Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature* **450**, 1253-1257 (2007).
6. E. I. Deryusheva, A. V. Machulin, O. V. Galzitskaya, Diversity and features of proteins with structural repeats. *Biophysical Reviews* **15**, 1159-1169 (2023).
7. Y. Belkhadir, R. Subramaniam, J. L. Dangl, Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Curr Opin Plant Biol* **7**, 391-399 (2004).
8. L. Yu, L. Wang, S. Chen, Endogenous toll-like receptor ligands and their biological significance. *Journal of cellular and molecular medicine* **14**, 2592-2603 (2010).
9. Z. Chang, Important aspects of Toll-like receptors, ligands and their signaling pathways. *Inflammation research* **59**, 791-808 (2010).
10. M. G. Tassia, N. V. Whelan, K. M. Halanych, Toll-like receptor pathway evolution in deuterostomes. *Proceedings of the National Academy of Sciences* **114**, 7055-7060 (2017).
11. M. Pellegrini, Tandem Repeats in Proteins: Prediction Algorithms and Biological Role. *Front Bioeng Biotechnol* **3**, 143 (2015).
12. T. Hrabe, A. Godzik, ConSole: using modularity of contact maps to locate solenoid domains in protein structures. *BMC Bioinformatics* **15**, 119 (2014).
13. K. Qiu, S. Dunin-Horkawicz, A. Lupas, Exploiting protein language model sequence representations for repeat detection. *bioRxiv* 10.1101/2024.06.07.596093, 2024.2006.2007.596093 (2024).
14. K. Ichii, A. Takkouche, e. al., Protein Repeat Recognition Through Unsupervised Clustering of Protein Large Language Model Embeddings. *Bioinformatics* **in preparation** (2024).

15. M. Blum *et al.*, The InterPro protein families and domains database: 20 years on. *Nucleic acids research* **49**, D344-D354 (2021).

16. N. Matsushima *et al.*, Analyses of non-leucine-rich repeat (non-LRR) regions intervening between LRRs in proteins. *Biochim Biophys Acta* **1790**, 1217-1237 (2009).

17. Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702-710 (2004).

18. S. Das, N. L. Dawson, C. A. Orengo, Diversity in protein domain superfamilies. *Curr Opin Genet Dev* **35**, 40-49 (2015).

19. B. Kobe, A. V. Kajava, The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* **11**, 725-732 (2001).

20. P. Strzelecki, H. Von Der Mosel, Integral Menger curvature for surfaces. *Advances in Mathematics* **226**, 2233-2304 (2011).

21. S. K. Burley *et al.*, RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res* **51**, D488-D508 (2023).

22. M. Varadi *et al.*, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444 (2022).

23. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

24. B. Kobe, J. Deisenhofer, A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**, 183-186 (1995).

25. B. Kobe, J. Deisenhofer, Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. *Journal of molecular biology* **264**, 1028-1043 (1996).

26. J. C. Roach *et al.*, The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences* **102**, 9577-9582 (2005).

27. A. C. Ng *et al.*, Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proceedings of the National Academy of Sciences* **108**, 4631-4638 (2011).