

RESEARCH ARTICLE

The evolutionary characteristics and structural biology of *Gallus* toll-like receptor 21

Hongping Wu¹  | Hai Wang¹ | Wuqi Jiang^{1,2} | Zhengxing Lian¹

¹Beijing Key Laboratory of Animal Genetic Improvement, China Agricultural University, No.2 Yuanmingyuan West Rd, Haidian, Beijing 100194, China

²College of Animal Science & Veterinary Medicine, Shenyang Agricultural University, No.120 Dongling Rd, Shenhe, Shenyang 110866, China

Correspondence

Zhengxing Lian, Beijing Key Laboratory of Animal Genetic Improvement, China Agricultural University, No.2 Yuanmingyuan West Rd, Haidian, Beijing 100194, China. Email: lianzhx@cau.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 31572389

Abstract

Toll-like receptors (TLRs) are an important part of the innate immune system, acting as a first line of defense against many invading pathogens. The ligand known to bind *Gallus* toll-like receptor 21 (gTLR21) is the unmethylated cytosine phosphate guanine dideoxy nucleotide motif; however, the evolutionary characteristics and structural biology of gTLR21 are poorly elaborated. Our results suggest that gTLR21 is phylogenetically and evolutionarily related to the TLR11 family and is perhaps a close ortholog of the *Mus* TLR13. Structural biology of homology modeling of the gTLR21 ectodomain structure suggests that it has no Z-loop like that seen in *Mus* TLR9. The cytosolic toll-IL-1 receptor region of gTLR21 contains a central 4-stranded parallel β -sheet (β A- β D) surrounded by 5 α -helices (α A- α E) on both sides, a highly conserved structure also seen in other TLRs. Molecular docking analysis reveals that the gTLR21 ectodomain has the potential to distinguish between different ligands. Homodimer analysis results also suggest that Phe842 and Pro844 of the BB loop and Cys876 of the α C helix in gTLR21 are conserved in other cytosolic toll-IL-1 receptor domains of other TLRs and may contribute to the docking of homodimers. Our study on the evolutionary characteristics and structural biology of gTLR21 reveals that the molecule may have a broader role to play in innate immune system; however, further experimental validation is required to confirm our findings.

KEYWORDS

docking, *Gallus* toll-like receptor 21 (gTLR21), homology model, leucine-rich repeat (LRR) motif, toll-IL-1 receptor (TIR) domain

1 | INTRODUCTION

Toll-like receptors (TLRs), which are fundamental sensor molecules in the innate immune system, are membrane-bound receptors located on the cell surface or in endocytic compartments and can recognize a wide range of pathogen-associated molecular patterns (PAMPs).¹ All of TLRs usually contain a series of leucine-rich repeat (LRR) motifs, a

transmembrane region, and a cytosolic Toll-IL-1 receptor (TIR) domain. The LRR motifs in the ectodomains (ECDs) of different TLRs can recognize different PAMPs as "ligands," including molecules such as lipids, lipoproteins, proteins, and nucleic acids derived from a wide range of microbes such as bacteria, viruses, parasites, and fungi.² Meanwhile, the TIR domains of TLRs can also trigger downstream signaling pathways to induce the release and production of inflammatory cytokines, type I interferon, and other immune factors. These reactions not only mediate defensive responses such as inflammation but also activate antigen-specific adaptive immune responses.³

So far, 39 types of TLR family members have been identified in organisms via the LRRfinder database (<http://www.lrrfinder.com/index.php>). A total of 13 TLRs (TLRs 1-13) have been identified in mammals, mainly in mice and humans; however, the expression

Abbreviations: CpG-DNA, cytosine phosphate guanine dideoxy nucleotide motif; ECD, ectodomain; gTLR21, *Gallus* TLR21; IL-1, interleukin-1; IRF7, Interferon regulatory factor 7; LRR, leucine-rich repeat; LRRhs, highly conserved leucine-rich repeat segments; LRRvs, variable leucine-rich repeat segment; MyD88, Myeloid differentiation primary response gene 88; PAMPs, pathogen-associated molecular patterns; ssRNA, single-stranded RNA; TIR, cytosolic Toll-IL-1 receptor region; TLR, Toll-like receptors
Hongping Wu, Hai Wang, and Wuqi Jiang contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors Journal of Molecular Recognition Published by John Wiley & Sons Ltd.

patterns of TLRs 10–13 appear to be species specific.⁴ TLRs 1–9 are conserved in humans and mice, and their immunological functions have received much attention and in-depth study. On the basis of their localization, these TLRs are largely divided into 2 subfamilies—cell-surface TLRs and intracellular TLRs.

Cell-surface TLRs include TLR1, TLR2, TLR4, TLR5, TLR6, and TLR10, which mainly recognize microbial membrane components such as lipids, lipoproteins, and proteins. TLR2 can heterodimerize with TLR1 or TLR6 and recognize lipoproteins and peptidoglycans from Gram-positive bacteria. TLR4 and TLR5 recognize the lipopolysaccharides of Gram-negative bacteria and bacterial flagellin, respectively. TLR10 is a pseudogene in mice, but the human TLR10 can collaborate with TLR2 to recognize ligands from *Listeria* and is involved in sensing influenza A viral infections.⁵

Intracellular TLRs (including TLR3, TLR7, TLR8, TLR9, TLR11, TLR12, and TLR13) are localized in the endosomal compartments and recognize nucleic acids originating from bacteria and viruses, as well as self-nucleic acids in disease conditions such as autoimmunity.¹ TLR3 has been found to recognize viral double-stranded RNA.⁶ Murine TLR7 and human TLR8 predominantly function in detecting GU-rich single-stranded RNA (ssRNA) from viruses.⁷ TLR9 and TLR13 recognize unmethylated cytosine phosphate guanine dideoxy nucleotide (CpG-DNA) motifs and bacterial 23S ribosomal RNA, respectively.^{8,9} Interestingly, TLR11 responds to flagellin, much like TLR5.¹⁰ In a recent study, TLR12 was found to be highly similar to TLR11 and functions in recognizing profilin from *Toxoplasma gondii*.¹¹

Several types of TLRs have been identified in other vertebrates, including fish and birds. Our current knowledge of *Gallus* TLRs (gTLRs) has been greatly advanced by the assembly of the genome sequence of the chicken (*Gallus gallus*). Previous studies have demonstrated the presence of 10 TLRs in *Gallus*. The gTLRs 3, 4, 5, and 7 are close orthologs of the corresponding TLRs found in other vertebrates and have similar immunological functions.⁴ In *Gallus*, the mammalian TLRs 1, 6, and 10 are replaced by TLR1La and TLR1Lb from an evolutionary point of view.¹² The duplicated genes (TLRs 2a and 2b) of *Gallus* are both orthologs of the single TLR2 found in mammals. *Gallus* TLR21 is an ortholog of the TLR21 in fish and amphibians.¹³ It appears that TLR15 is unique to birds and some reptilian species. TLR15 responds to in *Salmonella enterica* infections and is reported to have a unique auto-activation mechanism.¹⁴ The avian TLR21 is an ortholog of the TLR21 proteins from teleost and amphibian species, and clusters into the TLR11 subfamily. Surprisingly, gTLR21 has been shown to localize to the endoplasmic reticulum in cells transfected with the gTLR21 gene and can recognize unmethylated CpG-DNA; therefore, it functions much like TLR9 in mice.^{15–17}

To date, no 3D structure of gTLRs has been obtained by X-ray crystallography with a high level of confidence. Furthermore, no experimental data on the detailed molecular structure of gTLR21 complexes are as yet available. Computational methods may reveal information that is not easy to obtain by experimental means, and therefore, it is necessary for us to use computational methods to facilitate biological research. In this study, we use the computational methods to investigate the genetic evolution of gTLR21 and predict different binding patterns that the gTLR21 protein may have with potential ligands.

The purpose of our study is to improve our understanding of gTLR21 protein biophysics; the techniques we describe here may be used as tools for identifying potential binding locations and catching sight of exploring novel competitive inhibitors, biosensors, network components, and in vaccine development.

2 | MATERIALS AND METHODS

2.1 | Phylogenetic analysis

In this study, sequences of all full-length TLR proteins (not including the protein whose structure is being predicted) derived from known vertebrate species belonging to mammals, reptiles, avians, and teleosts were downloaded from the Uniprot database (<http://www.uniprot.org/>). First, aligned sequences for the TLR proteins were generated with MAFFT (L-INS-i).¹⁸ Following this, a phylogenetic tree was constructed using the neighbor-joining method with the JTT + I + G substitution model developed¹⁹ by Prottest3.4 and bootstrap sampling was performed 1000 times. The display, annotation, and management of phylogenetic trees was performed in iTOLv3.²⁰

2.2 | Analysis of residue conservation and secondary structural elements

The alignment results obtained in the first step were submitted to the ConSurf algorithm for the evaluation of evolutionarily conserved amino acid residue positions.²¹ The conservation scale ranged from 1 to 9 grades representing different degrees of conservation at the residue positions in gTLR21. The secondary structures of TLR proteins are distinguished through LRRfinder.²²

2.3 | Template searching, homology modeling, and interfacing analysis for the LRR region of gTLR21 with differential potential ligands

Because of low sequence identity between the target and template proteins (<40%), we chose to use the best available templates for multiple homology modeling for the LRR region of gTLR21. For this, human TLR3 (PDB ID: 1ziw), monkey TLR7 (PDB ID: 5gmh), human TLR8 (PDB ID: 3wn4), horse TLR9 (PDB ID: 3wpc), and mouse TLR13 (PDB ID: 4z0c) were chosen, as these proteins have structures available with 2.1, 2.2, 1.81, 1.6, and 2.3 Å resolution, respectively. A total of 100 candidate models for the target protein were constructed and optimized by Modeller9.18.²³ To acquire highly reliable model structures, the stereochemical quality and distribution of residual energy for the candidate structures were calculated by SAVES (<http://services.mbi.ucla.edu/SAVES/>)^{24–26} and ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>),²⁷ respectively.

The model of gTLR21 ECD obtained was submitted to the Metapocket2.0 server (<http://projects.biotec.tu-dresden.de/metapocket/index.php>)²⁸ for predicting potential ligand binding sites. CpG-DNA and ssRNA ligands were separated from the proteins with PDB IDs 3wpc and 4z0c, respectively, and saved in the pdb format using Modeller9.18. Ligands and LRR models are submitted to HDock server (<http://hdock.phys.hust.edu.cn/>),²⁹ respectively.

2.4 | Template searching, homology modeling, and homodimer prediction for TIR region of gTLR21

The TIR regions of human TLR6 (PDB ID: 4om7) and TLR10 (PDB ID: 2j67) for which 2.2 Å resolution structures are available were used as templates for modeling the TIR region of gTLR21 in Modeller9.18. The final target model was evaluated as described in the previous section.

The final structure of the gTLR21 TIR domain as a monomer was submitted to the GalaxyGemini server for exploring protein-protein interactions³⁰ (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=GEMINI>). In our study, all representational structures were displayed with PyMOL.³¹

3 | RESULTS AND DISCUSSION

3.1 | gTLR21 belongs to the TLR11 family and is clearly an ortholog to *Mus* TLR13

We used known sequences of full-length TLR proteins from vertebrates to construct phylogenetic relationships based on the neighbor-joining method (Figure 1A). In the phylogenetic tree obtained, all TLR protein sequences were divided into 6 families. Obviously, TLR21 was found to cluster within the TLR11 family and was clearly an ortholog of *Mus* TLR13. This phylogenetic analysis was consistent with the earlier findings.³²

Besides building phylogenetic relationship tree, we also analyze the conserved amino acid residues in the complete TLR21 protein sequences of reptilian, avian, and teleost species (Figure 1B). The gTLR21 owns equally high amino acid identities to Anser TLR21, it is possible that TLR21 among avians are highly orthologous and the avian TLR21s are closely related to reptilian. Furthermore, TLR21 is not found in humans or other mammals. Although TLR21 appears to have been lost in most vertebrates, it is particularly present in a minority of vertebrates like some reptilians, avians, and several fish. These data indicate that the evolution of TLR21 is according to the phylogeny of species and probably subjected to the species-specific constraints.^{13,33}

The TLR11 family includes 2 TLR subfamilies—TLRs 11-13 and TLRs 20-22; these members of the TLR11 family are probably derived from the TLR1 lineage.^{33,34} The results of our phylogenetic relationship analysis indicate that the TLR11 family can be split into 3 subfamilies (including TLRs 11, 13, and 22). Consistent with previous studies, the clade with the TLR subfamilies 4, 11, and 15 subsequently cluster into the TLR1 family, along with the TLR2 subfamilies.^{34,35} The results of the phylogenetic analysis of TLR21 indicate gTLR21 is highly orthologous to other avian TLR21s; TLR21 proteins are not unique to birds, but are also wide spread in reptilian and teleost species.¹³ The analysis of conserved amino acid residues using complete sequences of TLR21 proteins from different species indicates that the sequences share high identity with each other. This result further provides support for the phylogenetic relationships between the avian, reptilian, and teleost TLR21 proteins.

3.2 | Structural analysis of the ECD and intracellular domains of gTLR21

TLR21 proteins are generally 965 to 986 amino acids in length, and each TLR21 protein has a signal peptide, an ECD (which includes an LRRNT, 26 LRR motifs, and an LRRCT sequence), a transmembrane region, and an intracellular TIR domain, as indicated by motif prediction analysis. Our ConSurf results show that different modules in the TLR21 subfamily have different conserved residue positions. Most of the variable residues are located on the signal peptide, whereas most of the conserved amino acid positions are considered as spanning from the LRRNT to LRRCT in the ECD; similar patterns were also found in the TIR domain (Figure 2A). Meanwhile, we also statistics the average evolutionary conservation score about each motif (Figure 2B). It was observed that LRRs 7, 10-13, and 20 have relatively fast evolution rates. On the other hand, LRRs 14-18 have relatively higher conservation scores than other LRR motifs, implying that perhaps they play a role in protein-protein interfacing for dimerization or in binding to other members of the TLR family.

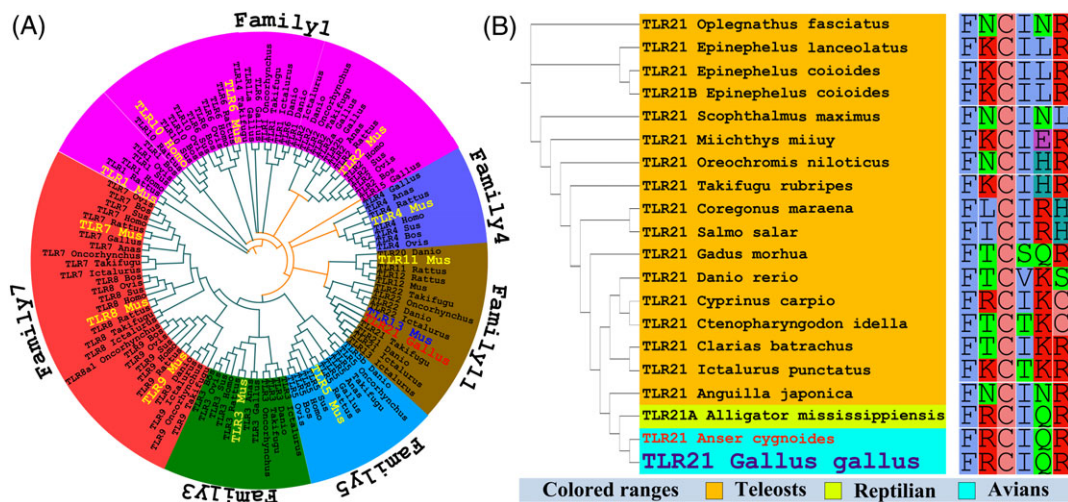


FIGURE 1 Phylogenetic trees of Gallus TLR21 (gTLR21) and other vertebrate TLRs based on the neighbor-joining method. A, This phylogenetic analysis shows that TLRs can be divided into 6 subfamilies in vertebrates and that gTLR21 is assigned to Family11. The each subfamily has own color. B, gTLR21 is highly orthologous to *Anser cygnoides*, and the TLR21 subfamily is conserved in different species

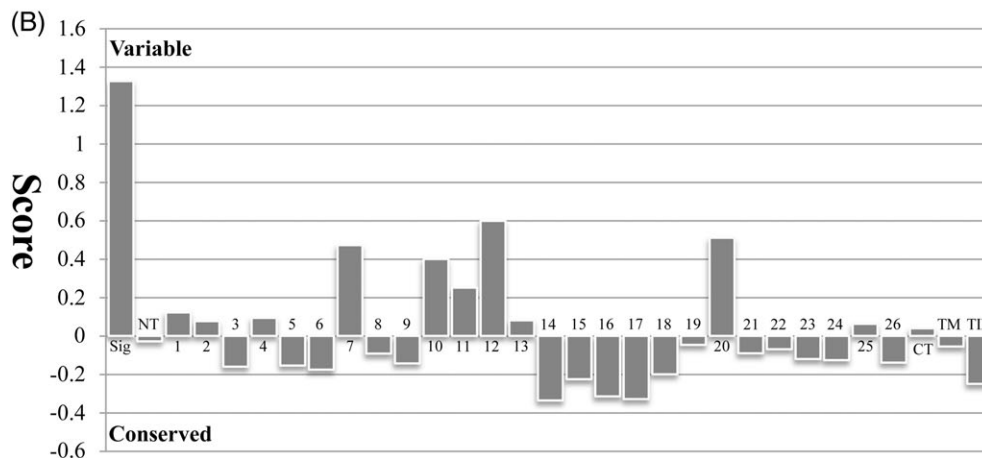
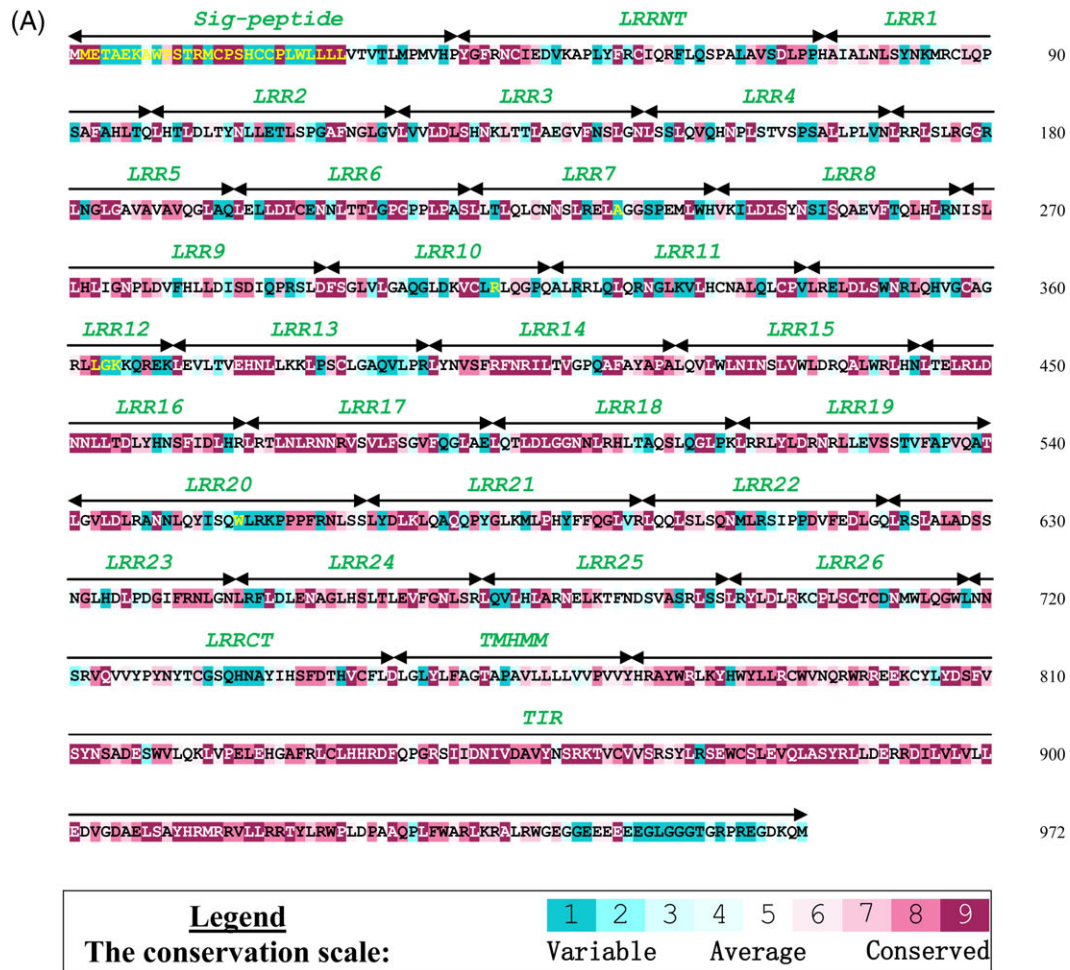


FIGURE 2 Evolutionarily conserved amino acid positions and mean evolutionary conservation rates for each TLR21 motif. A, The conservation scale ranges from 1 to 9; each grade has a different color, and grades 1 and 9 represent the most variable and most conserved residue positions, respectively. The sequence of *Gallus* TLR21 (gTLR21) is used to indicate the 9-color conservation grade system. The signal peptide, each predicted LRR modules of the ectodomain (ECD), transmembrane region (TM), and intracellular domain (TIR) for gTLR21 are labeled. B, The different modules of the gTLR21 protein are represented by their identifying numbers. The lowest score represents the most conserved position in the protein

The structure of gTLR21 ECD was modeled and optimized with Modeller9.18. The scores for the stereochemical quality of the candidate structures calculated by SAVES show that the derived structures of the gTLR21 ECD are reasonable (Tables S1A, S2A, and S3A). Simultaneously, the score of target model is -6.02 in ProSA-web, which indicates that its distribution of residual energies is also

acceptable (Figure S4A). The number of LRR motifs in TLR21 is consistent with those observed in *Mus* TLR13 in the family11 and *Mus* TLR9 in the family7. Each LRR motif comprises approximately 22 to 33 residues, except for the LRRCT and LRRNT sequences in the gTLR21 ECD. The LRR motifs can be classified into 2 types—highly conserved segments (LRRhs) and variable segments (LRRvs) in general. The

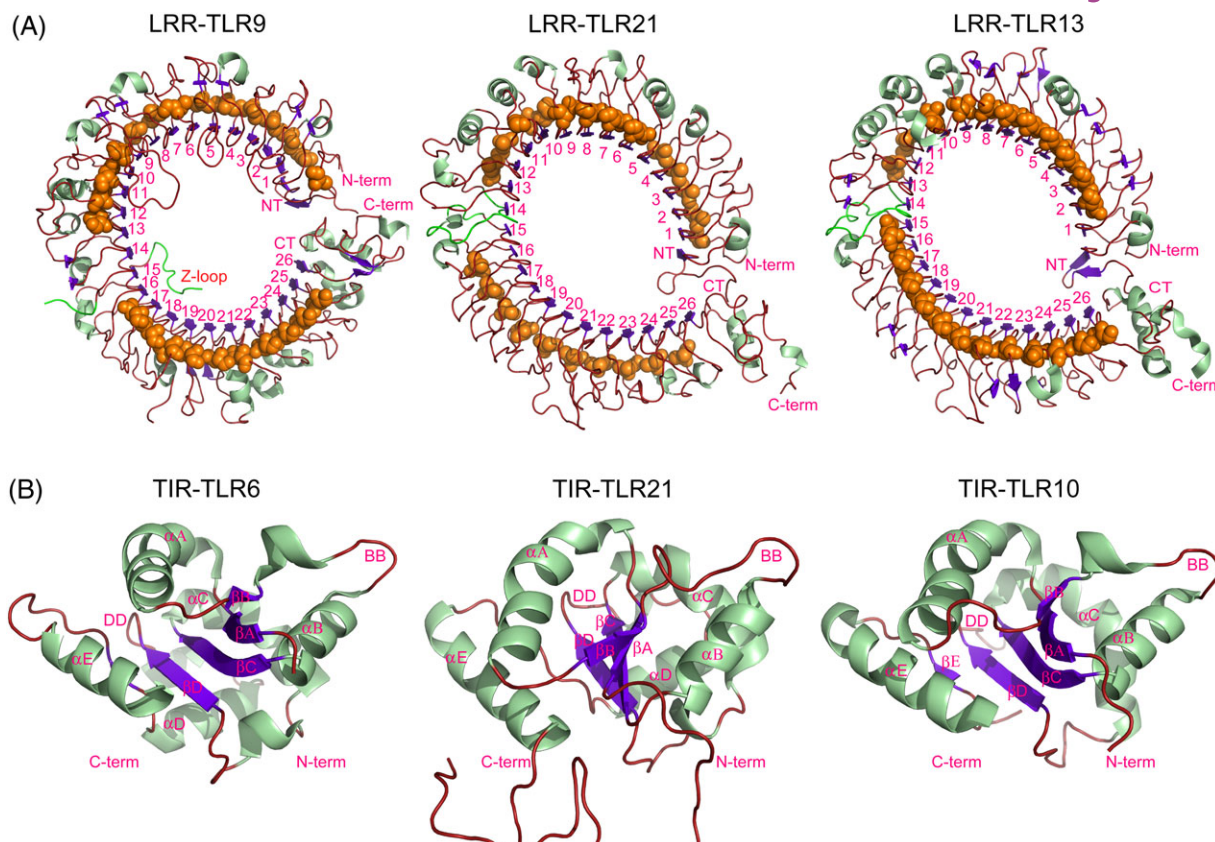


FIGURE 3 Structural comparison of the ectodomains (ECD) and intracellular domains (TIRs) of different toll-like receptors (TLRs). A, No Z-loop was detected between the LRR14 and LRR15 modules in the middle of the gTLR21 ECD; this is markedly different from the crystal structure of the human TLR9 ECD (PDB code: 3wpc), but is similar to the TLR13 ECD (PDB code: 4z0c). B, The overall conformation of the TIR domain of gTLR21 has a typical TIR domain fold like other TLRs (PDB codes: 4om7 and 2j67, respectively)

sequence LxxLxLxxN/C(x)L is an LRRhs-class model sequence, where “L” represents Leu, Ile, Val, or Phe; “N” represents Asp, Thr, Ser, or Cys; and “x” represents any amino acid.³² The LRRhs of gTLR21 are similar to conserved LRR subtypes and can form β -sheets packing the concave surface of the gTLR21 ECD, the remaining “irregular” LRR motifs bear variable similarities to different subtypes and form a convex surface structure (Figure 3A). It is obvious that gTLR21 can form a noncanonical horseshoe-shaped structure where the N- and C-terminal ends of protein extensively interact with each other; this is similar to the “closure” of an oval-shaped structure formed by *Mus* TLR13.⁹ When comparing the LRR modules of gTLR21, *Mus* TLR13, and *Mus* TLR9, an interesting observation was that the 14th LRR motif in gTLR21 lacks the long insertion known as a “Z-loop” (Figure 3A). The Z-loop is necessary for TLR9 dimerization and is involved in recognizing ligands for TLRs 7 and 8.³⁶ These results indicate that gTLR21 may have its own unique patterns for the recognition of ligands.

In this study, we have demonstrated that TLR21 is more likely to form an oval-shaped structure like TLR13, rather than a canonical horseshoe-shaped structure (Figure 3A). We could not detect the presence of a Z-loop between LRRs 14 and 15 through our model-building of the gTLR21 ECD. As previously reported, the Z-loop is considered indispensable for members of the TLR7 subfamily. TLR9 with the uncleaved Z-loop is unable to form a dimer,⁸ and the Z-loop is also required for ssRNA recognition by TLRs 7 and 8.^{7,37}

The structure of the gTLR21 TIR domain was also modeled and optimized with Modeller9.18 in this study. The scores for the stereochemical quality of the candidate structure calculated by SAVES as previously mentioned show that the structure obtained is reasonable (Figures S1B, S2B, and S3B). Simultaneously, the target model has a score of -4.07 in ProSA-web, which indicates that the distribution of residual energies of the model is also acceptable (Figure S4B). Usually, approximately 125 to 200 amino acid residues in the TIR domain are necessary for protein-protein interaction. Conserved residue position statistics indicate that amino acids 778–943 form a highly conserved structure in the cytoplasmic region of gTLR21 (Figure 2A). Our target model suggests that gTLR21 has a typical TIR domain fold with a central 4-stranded parallel β -sheet (β A– β D) surrounded by a total of 5 α -helices (α A– α E) on both sides (Figure 3B). The predicted TIR model has 5 extended loops such as BB and DD loops, which are similar to those in human TLR6. The BB loop is involved in interactions with several TIR domains.^{3,38}

We have identified the overall conformation of the gTLR21 TIR domain and found that it is similar to others in the TLR family. An important part of the structural stability of the TIR domain is likely to be provided by a 4-stranded parallel β -sheet. Recent studies suggest that the DD loop of TLR6 TIR domains and the BB loop of TIR module of TLR10 play critical roles in homotypic TLR interactions during TLR/IL-1R signaling events.^{38,39} This analysis propose that the TIR domain can form various types of dimers.

3.3 | Interaction analysis for potential ligand-ECD complexes formed by the gTLR21 protein

3.3.1 | Potential pockets analysis for LRR domains of gTLR21

The Meta Pocket 2.0 server successfully predicted 5 most likely binding sites in the LRR domains of gTLR21 (Figure 4). The results of the prediction show that binding site A is a large hole in the center of the LRR domains (Figure 4A). A similar positioned pocket has previously been

well studied in many TLRs. In the TLR1/TLR2 complex, this binding pocket of them not only is involved in ligand (lipopeptide) recognition but also is necessary for dimerization of TLRs 1 and 2.⁴⁰ Protein-protein interactions are also observed at a similar site in TLRs 7 and 8, which are required for homo-dimerization of TLRs 7 and 8.⁷ These results reveal that the binding site A may be involved in the biological functions of gTLR21 ECD. Binding site E, which can form a relatively large surface pocket (Figure 4B), is also found in similar regions of the *Mus* TLR9 ECD and may play a role in recognizing the unmethylated CpG-DNA

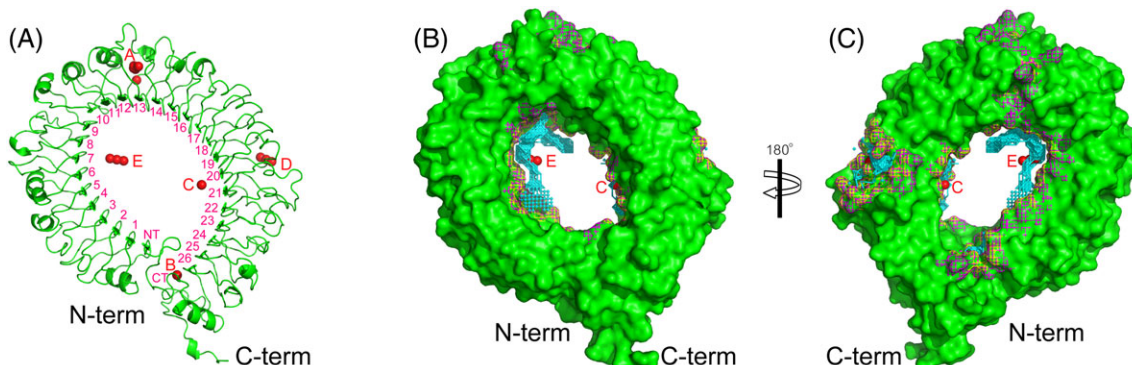


FIGURE 4 Mapping of predicted binding sites on the surface of the Gallus TLR21 (gTLR21) ectodomain (ECD). Five most likely binding sites are highlighted using red dots on the surface of the LRR model of gTLR21 by Meta Pocket2.0. The surfaces of the modeled structure of gTLR21 ECD are colored green, and the potential binding atoms and residues are indicated by yellow and magenta hatches, respectively. The potential binding clusters are indicated by cyan spheres

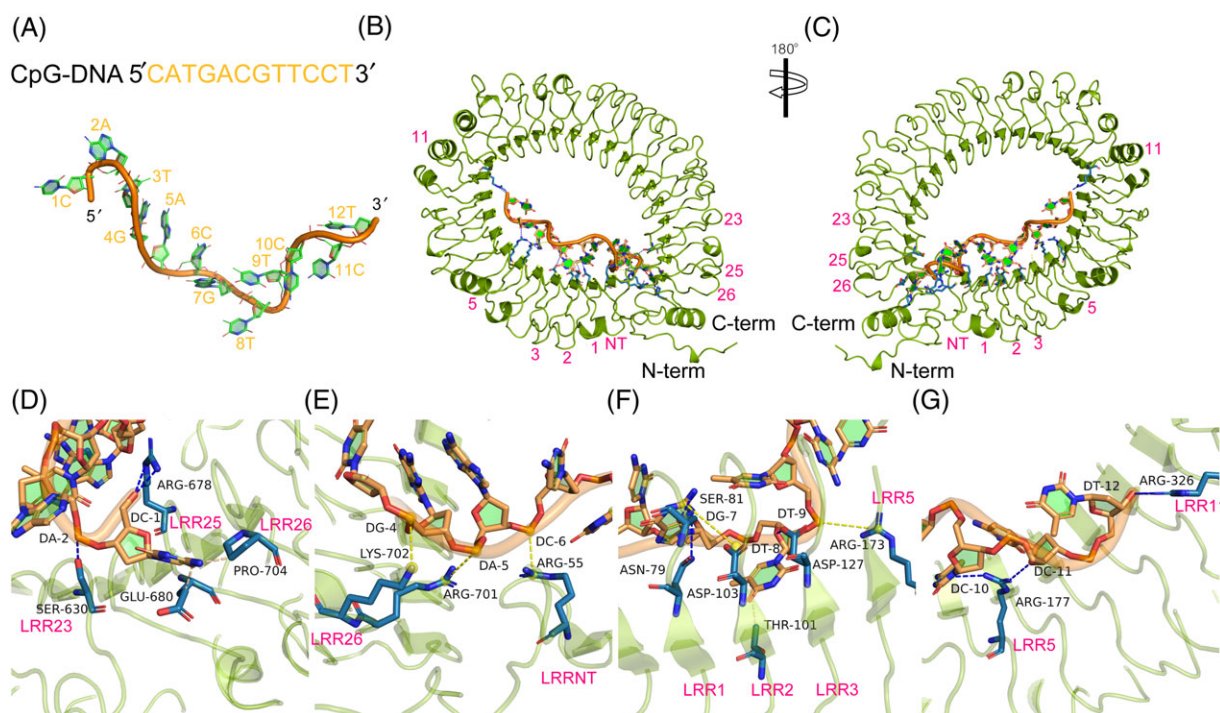


FIGURE 5 The sequence-specific recognition mechanism of unmethylated cytosine phosphate guanine dideoxy nucleotide (CpG-DNA) motif by Gallus TLR21 (gTLR21). A, The sequence and structure of CpG-DNA. B, C, The unmethylated CpG-DNA acts as a “molecular bridge” to penetrate through the concave face formed by the LRRNT domain, LRR1-3, 5, and 11 of gTLR21. D, The base C1 in CpG-DNA motif interfaces with Arg678, Glu680, and Pro704 via hydrogen bonds, and Ser630 in LRR23 forms a direct hydrogen bond with base A2. E, Potential salt bridges may form at the interface of Lys702-G4, Arg701-A5, and Arg55-C6. F, A series of residues located in LRR1, LRR2, and LRR3 may also contribute to the maintenance of the structure of the gTLR21-CpG-DNA complex. G, The side chains of Arg177 of LRR5 and Arg326 of LRR11 are involved in interfacing with the bases C10-T12 of 3' arm of the CpG-DNA. The side chains of gTLR21 are colored blue (in the ball-and-stick model), and the bases of the unmethylated CpG-DNA motif are colored green, respectively. H-bonds, hydrophobic interactions, and salt bridges are indicated by blue, yellow, and wheat dashed lines, respectively

motif.⁸ It is clear that a similar large hole in the inner concave face of gTLR21 ECD also constitutes a binding site C, which was previously identified in *Mus* TLR13 (Figure 4C). It has been proven that the binding site C is used for specific recognition of ssRNA.⁹ The remaining prediction results indicate the presence of small slit structures that may participate in binding other small potential ligands.

3.3.2 | Structural biology of the predicted CpG-DNA interface within the ECD of gTLR21 complex

The gTLR21 protein can recognize unmethylated CpG-DNA as a “danger” ligand to alert the innate and adaptive immune systems; gTLR21 can activate downstream pathways affecting the immune system much like mammalian TLR9.^{15,16} However, the mechanisms by which this specific recognition occurs have not yet been elaborated. Our results suggest that CpG-DNA (Figure 5A) acts as a “molecular bridge” to penetrate through the concave face of gTLR21 LRR domains (Figure 5B). Our results on modeling structural features indicates that the interface comprising the LRRNT and LRRs 1-3, 5, and 11 motifs can enhance the affinity of gTLR21 binding to CpG-DNA (Figure 5C). Recent studies have also shown that the binding region spans from the LRRNT to LRR10 motif of TLR9.⁸

We also tried to identify the amino acid residues potentially involved in the interactions at the binding surface. Results show that the base C1 in CpG-DNA motif forms direct hydrogen bonds with

Arg678, Glu680, and Pro704 (Figure 5D). We have also found that Ser630 in LRR23 interfaces with the base A2 via hydrogen bond (Figure 5D). In the interaction domain, the bases G4, A5, and C6 potentially form salt bridges with Lys702, Arg701, and Arg55 (Figure 5E), respectively. Three amino acid residues are found to the interface with the G7 base via multiple intermolecular forces in which the Asn79 and Ser81 are located in LRR1 and the Asp103 located in LRR2. Thr101 of LRR2 and Asp127 of LRR3 are also devoting to interacting with T8 base. The backbone phosphates of T9 are recognized by Arg173 of LRR5 (Figure 5F). Simultaneously, the side chains of Arg177 and Arg326 conform an interface with the bases C10-T12 of 3' arm of the CpG-DNA (Figure 5G). The structures of the agonistic CpG-DNA bound to gTLR21 described in this study reveal the structural bases of CpG-DNA recognition by gTLR21.

3.3.3 | Structural characteristics of the interaction of the potential ligand, ssRNA, with the LRR modules of gTLR21

We have also investigated the potential binding mechanism of gTLR21 with ssRNA. The ssRNA molecule (Figure 6A) likely fits along the inner concave surface of gTLR21, with its 5' and 3' arms binding to the C- and N-terminal ends of gTLR21 ECD (Figure 6B), respectively. This is consistent with how ssRNA molecules bind TLR13, except for opposite orientation of binding of TLR3–double-stranded RNA complex.⁶ Structural features suggest that ssRNA also forms a stem-loop-like

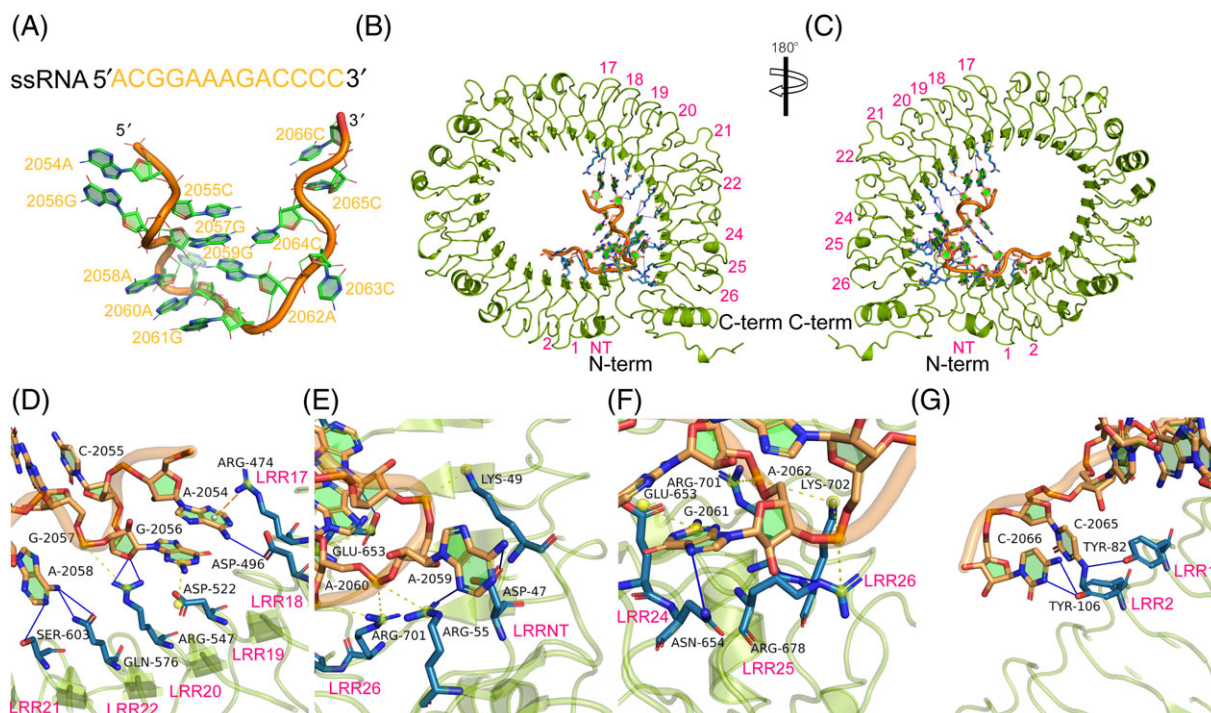


FIGURE 6 The sequence-specific recognition mechanism of single-stranded RNA (ssRNA) motif by Gallus TLR21 (gTLR21). A, The sequence and structure of ssRNA. B, C, The ssRNA could form a stem-loop-like structure so that its 5' and 3' arms mainly fit along the inner concave surface formed by the C- and N-terminal ends of the gTLR21 ECD, respectively. D, The bases A2054 and G2056-2058 of the ssRNA can bind to the interaction surface surrounded by residues located in LRR17-22 via potential π -cation interaction, hydrogen bonds, and salt bridges. E, The backbone phosphates of base A2059 is recognized by Lys49 and form hydrogen bonds with Asp47 and Arg55, whereas the base A2060 interfaces with Arg55, Glu653, and Arg701 via salt bridges. F, The base G2061 in the ssRNA motif interfaces with Glu653-Asn654 of LRR24, whereas Arg678 of LRR25 and Arg701 and Lys702 of LRR26 contribute to a combined interface with the base A2062 via multiple intermolecular forces. G, Tyr82 of LRR1 and Tyr106 of LRR2 are also involved in maintaining the structure of gTLR21-ssRNA complex via hydrogen bonds. The side chains of gTLR21 are colored blue (in the ball-and-stick model), and the bases of ssRNA motif are green. H-bonds, π -cation interactions, and salt bridges are indicated by blue, yellow, and wheat dashed lines, respectively

structure that is highly similar to those observed in TLR13-ssRNA complex (Figure 6C); the stem-loop-like structure is also essential for TLR13 recognition of ssRNA.^{9,41} Our structural biology analysis also strongly indicates that the predicted interaction surface is surrounded by the LRRNT, LRR1-2, LRR17-22, and LRR24-26 motifs of the gTLR21 ECD to fix the ssRNA in position.

The amino acid residues, which could potentially participate in the interactions at the binding surface, are discussed here. Our results indicate that there may be a π -cation interaction between the base A2054 and Arg474 of LRR17. The presence of hydrogen bonds is likely provided by the side chains of Asp496, Arg547, Gln576, and Ser603 in our results. Simultaneously, the bases G2056 and G2057 may potentially form salt bridges with Asp522 and Arg547, respectively, in the combined domain (Figure 6D). With the extension of the ssRNA motif, A2059 is recognized by Asp47, Arg55, and Lys49 via multiple intermolecular interactions. It is

clear that the backbone phosphates of the base A2060 are recognized by Arg55, Glu653, and Arg701 (Figure 6E). The Glu653 to Asn654 of LRR24, Arg678 of LRR25, and Arg701 and Lys702 of LRR26 also contribute to the maintenance of this structure via different intermolecular forces (Figure 6F). The docking results indicate that Tyr82 of LRR1 and Tyr106 of LRR2 also devote to bind to the bases of 3' arms of the ssRNA motif via hydrogen bonds (Figure 6G). Recently, a series of studies has demonstrated that *Mus* TLR13 is a receptor for the vesicular stomatitis virus⁴² and could detect sequence-specific areas of 23S ribosomal RNA from bacteria.⁴¹ The results of our study described here reveal that gTLR21 may have the potential to specifically recognize ssRNA like *Mus* TLR13⁹ and that this could contribute to the development of a strong immune mechanism in *Gallus*.

The TLRs ECD can form a horseshoe-shaped structure with a concave surface that participates in the recognition of various pathogens.

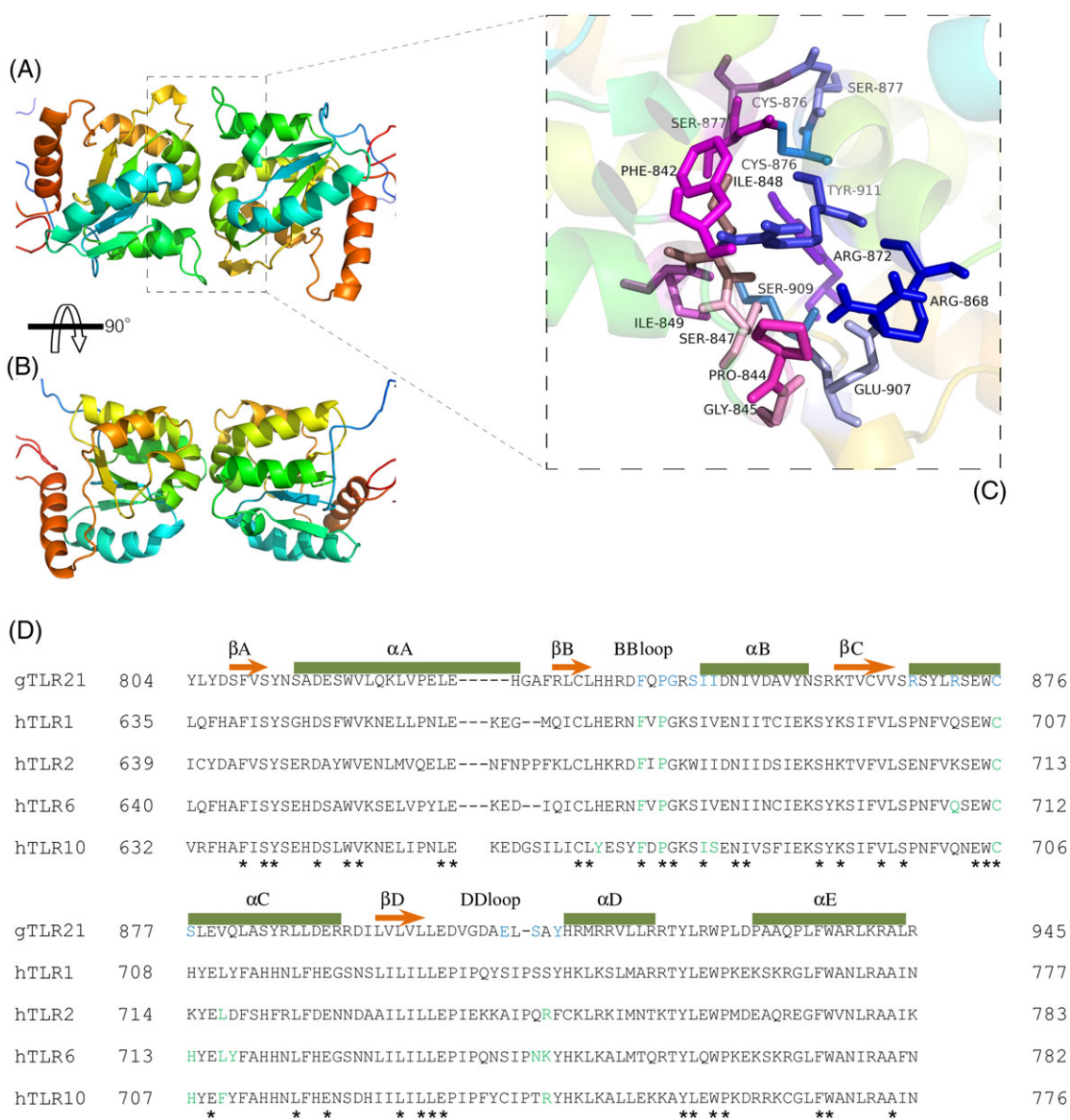


FIGURE 7 Homodimer of the *Gallus* TLR21 (gTLR21) intracellular (TIR) domain. A, The dimeric structure of the gTLR21 TIR domain shows the crystallographic asymmetric unit. B, Potential residues that could interact at the interface of each gTLR21 TIR domain monomer. C, Alignment of representative TIR domain sequences from different TLRs. The elements of secondary structures are indicated above each sequence. Amino acid sequences conserved across TLR families are marked by "*." Previously identified amino acids that are known to be important for TLR function are colored green. The residues predicted to function in the interactions at the interface of gTLR21 are shown in blue

The signaling complex structures of TLRs with their respective ligands reveal diverse mechanisms in the recognition of a wide variety of PAMPs. These varied PAMPs have unique characteristic molecular signatures that are specially detected by the corresponding TLRs that recognize them. The results of LRR-ligand complex prediction analysis indicate that gTLR21 may play an extensive role in mounting immune responses to bacteria and viruses via the recognition of specific nucleic acid ligands.

Because of selection pressures in the vertebrate evolutionary history of vertebrates, different classes of TLRs were generated to recognize the similar ligands and show analogous localization. Although TLR9 is absent in avian, fish genomes have both TLR9 and TLR21. Previous work has demonstrated that TLR9 and TLR21 have different ligand recognition profiles, and cooperatively mediate immune responses to CpG-DNA in zebrafish.¹⁷ Recent studies have also proposed that the TLR21 gene is up-expressed after infection by fish viruses in teleosts.^{43,44} CpG-DNA holds considerable promise as an adjuvant in vaccines target ssRNA viruses such as the Newcastle disease and influenza A virus H5N1.^{45,46} These observations indicate that TLR21 may play an important role in mounting immune responses to the infections of specific pathogens that involve recognition of viral and bacterial RNA and DNA. These data indicate that TLR21 can play multiple roles in the innate immune system. Clearly, a detailed analysis of the structural characteristics of the gTLR21 ECD is needed to discriminate as other TLRs.

3.4 | Homodimer analysis for the TIR domains of gTLR21

The docking results indicate that TIR monomers of gTLR21 are able to form homodimer complexes with one another (Figure 7A). Three-dimensional structures of the gTLR21 TIR domains display a crystallographic asymmetric dimerism. Interaction analyses suggest that 12 amino acid residues (Phe842, Pro844, Gly845, Ser847, Ile848, Ile849, Arg868, Arg872, Cys876, Glu907, Ser909, and Tyr911) can contribute to the formation of the dimeric interface (Figure 7B). The results of the alignment analysis (Figure 7C) show that Phe842, Pro844, and Gly845, which located in the BB loop, are highly conserved in other TLRs and that residues in the DD loop (Glu907, Ser909, and Tyr911) have been reported to play an important role in the dimerization interface.⁴⁰ We also find that Ile848 located in the α B helix and Cys876 located in the α C helix are highly conserved; both residues are likely to participate in forming TIR homodimers similar to those formed by TLR10.³⁹

This study represents an attempt to apply computational methods such as protein-protein docking analysis to explore the interactions of TIR domains to investigate the mechanisms of signaling induced by gTLR21. The homodimer analysis results suggest that the gTLR21 TIR region has the potential to form homodimers much like those formed by TLR6 and TLR10.^{38,39} We have also found 3 potential amino acid residues (Phe842, Pro844, and Cys876) at the homodimer interface that highly conserved in other TLRs; these residues may play an essential role in signal transduction that triggered by TLR and interleukin-1 (IL-1).^{38,39,47} TLR9 was unable to elicit an immune response to unmethylated CpG-DNA in MyD88 knockout mice,⁴⁸ probably because the induction of type I interferons, particularly type I interferon- α , by

TLR9 depends on the MyD88-IRF7 pathway in pDC cells.⁴⁹ TLR13 also appears to induce a MyD88-dependent signaling pathway to trigger the activation of NF- κ B; and TLR13 is also dependent on IRF7 for activating type 1 interferon pathways.⁴² Results of recent studies indicate that gTLR21 ectopic expressed in HEK-293 cells could regulate NF- κ B and furthermore could mediate the expression of cytokines in HD11 cells on both exogenous CpG-DNA stimulation.^{15,16} Taken together, these data indicate that the TIR domain participates in protein-protein interactions with intracellular adaptor proteins for signaling processes.

4 | CONCLUSIONS

By performing a phylogenetic and evolutionary analysis of TLR21 proteins from majority animals, we report that TLR21 is phylogenetically related to TLR11 family and is perhaps a close ortholog of *Mus* TLR13. Our structural biology analysis suggests that there is no Z-loop in the gTLR21 ECD, although the Z-loop is known to play a critical role in activation for the TLR7 family. The molecular structural features of the gTLR21 TIR domain indicate that it contains a central 4-stranded parallel β -sheet surrounded by 5 α -helices on both sides. The TIR domains of gTLR21 are also highly conserved, and similar to other TLRs. Molecular docking analysis has revealed that gTLR21 has the potential to distinguish between different ligands. The results of the homodimer analysis suggest that the TIR domain may be involved in forming gTLR21 homodimers. The Cys876 from the α C helix as well as the Phe842 and Pro844 from the BB loop are identical in the TIR domains of other TLRs and may function as docking sites in TIR domain homodimers.

ACKNOWLEDGEMENTS

We thank Prof Hongbing Han (College of Animal Science and Technology, China Agricultural University, Beijing, China) for generously providing his valuable proposals. We also thank Yao Jiang PhD (College of Animal Science and Technology, China Agricultural University, Beijing, China) for providing necessary software to perform this work.

CONFLICT OF INTEREST

The authors declare to have no conflict of interest.

FUNDING

This work was financially supported by National Natural Science Foundation of China (31572389).

ORCID

Hongping Wu  <http://orcid.org/0000-0003-3827-4595>

REFERENCES

1. Kawasaki T, Kawai T. Toll-like receptor signaling pathways. *Front Immunol*. 2014;5:461
2. Kumar H, Kawai T, Akira S. Toll-like receptors and innate immunity. *Biochem Biophys Res Commun*. 2009;388(4):621-625.
3. Narayanan KB, Park HH. Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. *Apoptosis*. 2015;20(2):196-209.
4. Brownlie R, Allan B. Avian toll-like receptors. *Cell Tissue Res*. 2011;343(1):121-130.

5. Kawai T, Akira S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity*. 2011;34(5):637-650.
6. Liu L, Botos I, Wang Y, et al. Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science*. 2008;320(5874):379-381.
7. Zhang Z, Ohto U, Shibata T, et al. Structural analysis reveals that toll-like receptor 7 is a dual receptor for guanosine and single-stranded RNA. *Immunity*. 2016;45(4):737-748.
8. Ohto U, Shibata T, Tanji H, et al. Structural basis of CpG and inhibitory DNA recognition by toll-like receptor 9. *Nature*. 2015;520(7549):702-705.
9. Song W, Wang J, Han Z, et al. Structural basis for specific recognition of single-stranded RNA by toll-like receptor 13. *Nat Struct Mol Biol*. 2015;22(10):782-787.
10. Mathur R, Oh H, Zhang D, et al. A mouse model of salmonella typhi infection. *Cell*. 2012;151(3):590-602.
11. Koblansky AA, Jankovic D, Oh H, et al. Recognition of profilin by toll-like receptor 12 is critical for host resistance to *Toxoplasma gondii*. *Immunity*. 2013;38(1):119-130.
12. St Paul M, Brisbin JT, Abdul-Careem MF, Sharif S. Immunostimulatory properties of toll-like receptor ligands in chickens. *Vet Immunol Immunopathol*. 2013;152(3-4):191-199.
13. Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW. Evolution of the chicken toll-like receptor gene family: A story of gene gain and gene loss. *BMC Genomics*. 2008;9(1):62
14. Wang J, Zhang Z, Chang F, Yin D. Bioinformatics analysis of the structural and evolutionary characteristics for toll-like receptor 15. *PeerJ*. 2016;4:e2079
15. Brownlie R, Zhu J, Allan B, et al. Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides. *Mol Immunol*. 2009;46(15):3163-3170.
16. Keestra AM, de Zoete MR, Bouwman LI, van Putten JP. Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. *J Immunol*. 2010;185(1):460-467.
17. Yeh DW, Liu YL, Lo YC, et al. Toll-like receptor 9 and 21 have different ligand recognition profiles and cooperatively mediate activity of CpG-oligodeoxynucleotides in zebrafish. *Proc Natl Acad Sci U S A*. 2013;110(51):20711-20716.
18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.
19. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27(8):1164-1165.
20. Letunic I, Bork P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;44(W1):W242-W245.
21. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44(W1):W344-W350.
22. Offord V, Werling D. LRRfinder2.0: A webserver for the prediction of leucine-rich repeats. *Innate Immun*. 2013;19(4):398-402.
23. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. 2016;54: 5 6 1-5 6 37
24. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993;26(2):283-291.
25. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9):1511-1519.
26. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol*. 1997;277:396-404.
27. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35(Web Server issue):W407-W410.
28. Zhang Z, Li Y, Lin B, Schroeder M, Huang B. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*. 2011;27(15):2083-2088.
29. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDock: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res*. 2017;45(W1):W365-W373.
30. Lee H, Park H, Ko J, Seok C. GalaxyGemini: A web server for protein homo-oligomer structure prediction based on similarity. *Bioinformatics*. 2013;29(8):1078-1080.
31. Alexander N, Woetzel N, Meiler J. bcl::Cluster: A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *IEEE Int Conf Comput Adv Bio Med Sci*. 2011;2011:13-18.
32. Matsushima N, Tanaka T, Enkhbayar P, et al. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics*. 2007;8(1):124
33. Oshiumi H, Matsuo A, Matsumoto M, Seya T. Pan-vertebrate toll-like receptors during evolution. *Curr Genomics*. 2008;9(7):488-493.
34. Roach JC, Glusman G, Rowen L, et al. The evolution of vertebrate toll-like receptors. *Proc Natl Acad Sci U S A*. 2005;102(27):9577-9582.
35. Sundaram AY, Consuegra S, Kiron V, Fernandes JM. Positive selection pressure within teleost toll-like receptors tlr21 and tlr22 subfamilies and their response to temperature stress and microbial components in zebrafish. *Mol Biol Rep*. 2012;39(9):8965-8975.
36. Ewald SE, Lee BL, Lau L, et al. The ectodomain of toll-like receptor 9 is cleaved to generate a functional receptor. *Nature*. 2008;456(7222):658-662.
37. Tanji H, Ohto U, Motoi Y, Shibata T, Miyake K, Shimizu T. Autoinhibition and relief mechanism by the proteolytic processing of toll-like receptor 8. *Proc Natl Acad Sci U S A*. 2016;113(11):3012-3017.
38. Jang TH, Park HH. Crystal structure of TIR domain of TLR6 reveals novel dimeric interface of TIR-TIR interaction for toll-like receptor signaling pathway. *J Mol Biol*. 2014;426(19):3305-3313.
39. Nyman T, Stenmark P, Flodin S, Johansson I, Hammarström M, Nordlund P. The crystal structure of the human toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer. *J Biol Chem*. 2008;283(18):11861-11865.
40. Botos I, Segal DM, Davies DR. The structural biology of toll-like receptors. *Structure*. 2011;19(4):447-459.
41. Li XD, Chen ZJ. Sequence specific detection of bacterial 23S ribosomal RNA by TLR13. *Elife*. 2012;1:e00102
42. Shi Z, Cai Z, Sanchez A, et al. A novel toll-like receptor that recognizes vesicular stomatitis virus. *J Biol Chem*. 2011;286(6):4517-4524.
43. Wang W, Shen Y, Pandit NP, Li J. Molecular cloning, characterization and immunological response analysis of toll-like receptor 21 (TLR21) gene in grass carp, *Ctenopharyngodon Idella*. *Dev Comp Immunol*. 2013;40(3-4):227-231.
44. Li S, Wang G, Liu D, Liu Q, Hu G. Cloning and expression analysis of a toll-like receptor 21 (TLR21) gene from turbot, *Scophthalmus maximus*. *Dev Comp Immunol*. 2017;73:163-168.
45. Linghua Z, Xingshan T, Fengzhen Z. Vaccination with Newcastle disease vaccine and CpG oligodeoxynucleotides induces specific immunity and protection against Newcastle disease virus in SPF chicken. *Vet Immunol Immunopathol*. 2007;115(3-4):216-222.
46. Wang Y, Shan C, Ming S, Liu Y, Du Y, Jiang G. Immunoadjuvant effects of bacterial genomic DNA and CpG oligodeoxynucleotides on avian influenza virus subtype H5N1 inactivated oil emulsion vaccine in chicken. *Res Vet Sci*. 2009;86(3):399-405.
47. Poltorak A, He X, Smirnova I, et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: Mutations in Tlr4 gene. *Science*. 1998;282(5396):2085-2088.
48. Häcker H, Vabulas RM, Takeuchi O, Hoshino K, Akira S, Wagner H. Immune cell activation by bacterial CpG-DNA through myeloid differentiation marker 88 and tumor necrosis factor receptor-associated factor (TRAF)6. *J Exp Med*. 2000;192(4):595-600.

49. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. *Nat Immunol*. 2010;11(5): 373-384.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Wu H, Wang H, Jiang W, Lian Z. The evolutionary characteristics and structural biology of *Gallus* toll-like receptor 21. *J Mol Recognit*. 2018;31:e2696. <https://doi.org/10.1002/jmr.2696>