Review Article

Structural Characterization of Nucleic Acid-sensing Toll-like Receptors

The innate immune system senses pathogen-associated or cell damage-associated structurally conserved molecules through various pattern recognition receptors (PPRs) [1]. Nucleic acids that are released from viruses, bacteria or dead cells during infection or tissue damage are principal ligands to PPRs. PRR involved in nucleic acid recognition can be divided into two groups on the basis of cellular localization: several members of membrane-bound Toll-like receptors (TLRs) and cytoplasmic PPRs such as Nod-like receptors (NLRs), RIG-I like receptors (RLRs), and the DNA recognition receptor cyclic GMP-AMP synthase (cGAS).

TLRs, evolutionarily conserved membrane-spanning receptors homologous to the *Drosophila* Toll protein, are mostly expressed in macrophages and dendritic cells. The recognition of pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs) by TLRs is critical for the activation of the transcription factor NF-κB or IRFs, leading to the production of pro-inflammatory cytokines or type I interferons (Fig. 1). TLRs are glycosylated type I integral membrane receptors with N-terminal extracellular leucine-rich repeats (LRRs), a transmembrane domain, and a C-terminal cytoplasmic domain (Fig. 2(a)). The LRR domain contains binding sites for PAMPs or DAMPs, while the cytoplasmic domain, known as the Toll/IL-1 receptor (TIR) region, activates downstream signaling cascades by interacting with adaptor proteins such as MyD88 and TRIF [2]. Biochemical and structural studies on TLR extracellular domains suggest the following mechanism of TLR activation. Inactive TLRs exist as monomers, and ligand binding induces dimerization of the extracellular TLR domain, producing a typical "m"-shaped structure, where the intracellular C-terminal regions of the two TLR protomers are positioned in close proximity. Subsequent dimerization of the intracellular TIR domain is followed by the recruitment of adaptor proteins that execute signal transduction.

Viral/bacterial nucleic acids are potent stimulators of innate immunity. In humans, nearly half the TLRs recognize nucleic acid ligands. TLR3 responds to double-stranded (ds) RNA, while TLR7 and TLR8, which are closely related, recognize single-stranded (ss)RNA, and TLR9 senses DNA with unmethylated cytosine phosphate-guanosine (CpG) motifs (Fig. 2(b)). In addition, mouse TLR13 detects bacterial 23S ribosomal RNA. In contrast to other TLRs, TLR3 and TLRs 7–9 are expressed on endosomal membranes, limiting the recognition of self-derived ligands that are released from dying cells (Fig. 1). In addition, the members of the TLR7 subfamily comprising TLRs 7, 8, and 9 contain a characteristic long inserted loop region (known as the Z-loop) composed of approximately 30 amino acid residues, and localized in the region between LRR14 and LRR15 (Fig. 2(a,b)). Z-loop processing is required for the activation of these TLRs.

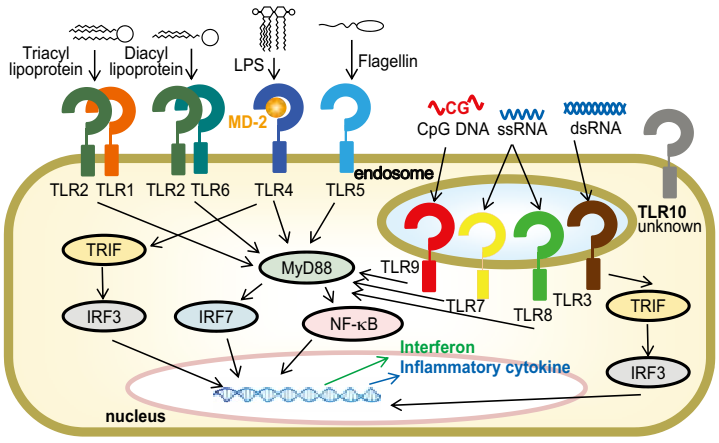

Fig. 1. **Schematic illustration of human TLR signaling**
TLR signaling is initiated by ligand-induced dimerization of receptors, followed by the engagement of TIR-domain-containing adaptor proteins, such as MyD88 and TRIF, which activate downstream signaling cascades. A major consequence of TLR signaling is the induction of proinflammatory cytokines and type I interferons. TLR5 and heterodimers of TLR2 and TLR1 or TLR6 are expressed at the cell surface, whereas TLR3 and TLR7–9 localize to the endosome, where they sense microbial and host-derived nucleic acids. TLR4 localizes to both the plasma membrane and the endosome.

Although TLR7 and TLR8 have been considered to primarily recognize ssRNA, they are also activated by small synthetic ligands such as imidazoquinolines and nucleoside analogs, raising a question concerning the molecular basis of the recognition of these structurally and chemically distinct ligands. Furthermore, expression patterns of TLR7 and TLR8 are different, but it remained to be determined whether or not the biochemical features and recognition mechanisms of TLR7 and TLR8 are different. Moreover, the mechanism of the recognition of a pathogenic DNA sequence by TLR9 has not been clarified in detail.

We used brilliant synchrotron X-rays at SPring8 BL41XU and KEK to determine the crystal structures of the LRR domains of TLR7 [3], TLR8 [4,5] and TLR9 [6]. First, the structures of unliganded and small synthetic ligand-bound TLR8 were determined (Fig. 3(a)) [4]. Chemical ligands bind to TLR8 at two equivalent positions on a symmetrical 2:2 complex, triggering the rearrangement of the dimeric configuration. Then, TLR8 C-termini are brought into closer proximity, thus generating a more compact structure than that of the m-shaped dimer. The ligand is positioned at the TLR8 dimer interface between the N-terminal half of one protomer and the C-terminal half of the other protomer, stabilizing the activated form of TLR8. Prior to structural studies of TLRs, it was not known whether the N-terminal fragment in the TLR7 family was required for ligand recognition following Z-loop processing. However, structural observations clearly demonstrated that both N- and C-terminal fragments of TLR8 are necessary for ligand recognition. Three key interactions are important in ligand recognition: (i) stacking interactions between the benzene ring of ligands and Phe405 of TLR8; (ii) hydrogen bonds between the amidine group of ligands and Asp543 of TLR8; and (iii) snug fitting of two substituents of a ligand to the small hydrophobic pocket formed between the two protomers (Fig. 3(b)).

Next, the crystal structure of TLR8 complexed with 20-mer ssRNA was determined [5]. Unexpectedly, the crystal structure revealed that rather than binding directly to the full-length ssRNA, TLR8 binds to RNA degradation products at two distinct sites (Fig. 3(a)). The first site, which accommodates the ssRNA degradation product uridine, is identical to a previously reported binding site for small synthetic ligands [4]. The second binding site located at the interior of the TLR8 ring structure is sandwiched between the concave surface and the Z-loop and holds ssRNA molecules longer than 2 nucleotides. Most recently, the structure of the activated TLR7-guanosine-ssRNA complex has been determined [3] (Fig. 3(a)). Analogous to the TLR8 structure, which recognizes uridine and ssRNA at distinct sites, TLR7 interacts with guanosine and polyU at two different sites. The first site is identical to that of TLR8, and accommodates small agonists; however, it preferentially binds guanosine, while TLR8 accommodates uridine. The second site presents an ssRNA-binding region distinct from that of TLR8 in its spatial position and ligand-recognition mode. As non-terminal uridine was specifically recognized whereas other uridines were loosely recognized, diverse oligonucleotides can be accomodated at the second site as long as they contain uridine. Taken all together, TLR7 is a dual sensor for guanosine and uridine-containing ssRNA, while TLR8 acts as a uridine sensor recognizing diverse nucleotide compositions of ssRNA. These findings indicate that although TLR7 and TLR8 are closely related receptors and demonstrate similar activation mechanisms, they have significant differences in ligand recognition patterns. Biochemical analysis provided further insights into the ligand specificity of TLR7 by showing that not only guanosine but also its modified derivatives (e.g., 7-methylguanosine and 8-hydroxyguanosine) may serve as TLR7 endogenous ligands [7].

Moreover, the research group clarified the crystal structure of three forms of TLR9, namely, unliganded TLR9, TLR9 bound to a CpG-containing DNA, and TLR9 bound to an inhibitory DNA (iDNA) [6]. CpG-containing DNA binds to TLR9 at a stoichiometric ratio of 2:2 in an extended conformation. CpG-

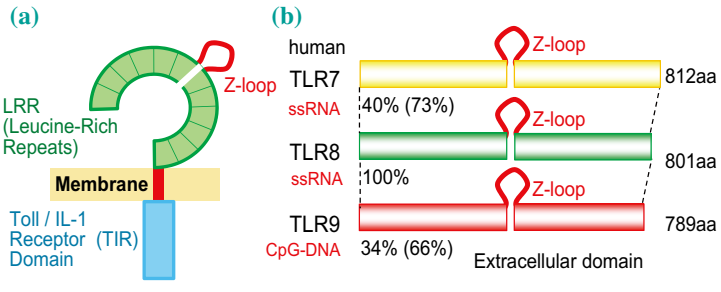

Fig. 2. **Schematic representation of nucleic acid-sensing TLRs** (**a**) Schematic representation of the domain organization in nucleic acid-sensing TLRs. TLRs consist of an extracellular LRR domain, a transmembrane domain, and a cytoplasmic TIR domain. The characteristic Z-loop is shown in red. (**b**) Schematic representation of the extracellular domains of nucleic acid-sensing TLRs. Sequence homology to TLR8 is shown, and numbers in parentheses are the values of sequence identity. The cognate ligands are also shown.

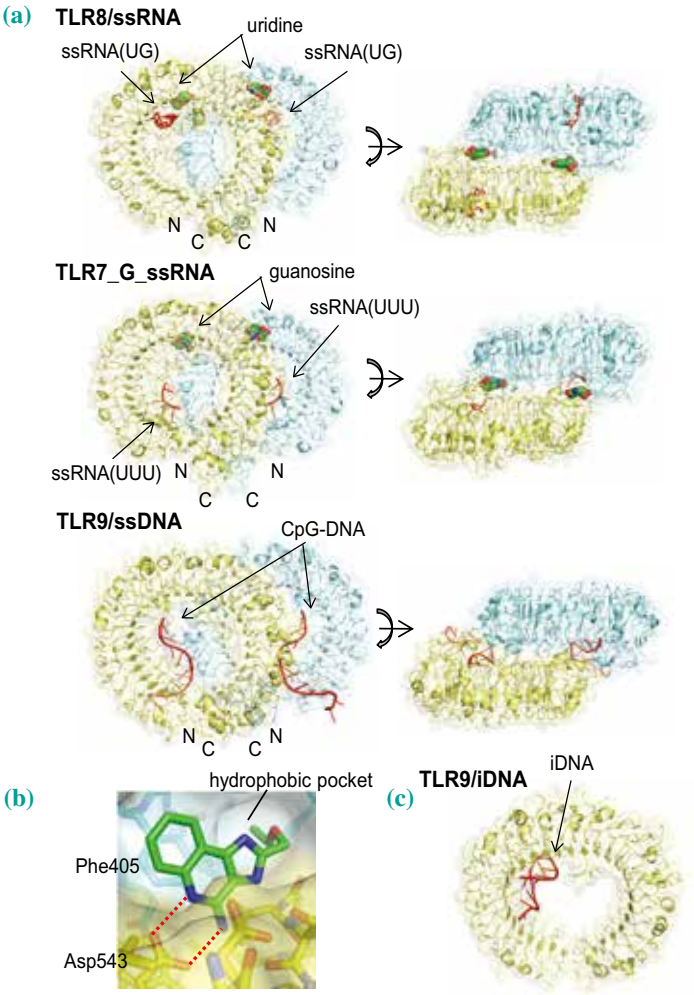

Fig. 3. **Structures of nucleic acid-sensing TLRs**
(**a**) Signaling complexes of nucleic acid-sensing TLRs. Front (left panels) ant top (right panels) views of the signaling complexes: TLR8/ssRNA (PDB: 4R07), TLR7_G_ssRNA (PDB: 5GMF), and TLR9/CpG DNA (PDB: 3WPC). (**b**) Ligand recognition by TLR8. The hydrogen bond is depicted as a red dotted line. (**c**) Structure of TLR9 in complex with iDNA.

containing DNA was recognized by its binding to the groove in the N-terminus of TLR9 of one protomer, while the C-terminal domain of the other protomer mainly binds to the DNA backbone. Thus, CpG-containing DNA acts as the molecular glue bridging two TLR9 monomers. In addition, not only the CpG motif but also its flanking regions are recognized by TLR9. It is generally accepted that TLR9 distinguishes pathogenic DNA partially on the basis of methylation status. Our works demonstrated that CpG methylation weakens the affinity of CpG to TLR9 and its ability to cause receptor dimerization, probably because of the disruption of water clusters that mediate interactions between CpG and TLR9. In the TLR9-iDNA complex, iDNA, which has a stem-loop conformation stabilized by intramolecular base-pairing, demonstrates a close fit to the concave surface of TLR9 (Fig. 2(c)). Binding interfaces of CpG-DNA and iDNA are partially overlapped. As iDNA exhibits a stronger binding affinity for TLR9 than do CpG-containing agonists, the overlap would account for the antagonistic effect of iDNA.

Synergistic activation by two ligands, revealed by the structural works in TLR7 and TLR8, is a new concept of the activation mechanism [3,5]. Although TLR7 and TLR8 exhibited the preference for guanosine and uridine, respectively, over other mononucleosides or mononucleotides at the first binding site, the binding affinity to these ligands is still lower than that to synthetic molecules (e.g., TLR8 has K_d values of 55 μM and 0.2 μM for uridine and R848, respectively). The lower affinity for mononucleosides can be attributed to the lack of the alkyl group harbored by the synthetic ligand, protruding into the hydrophobic pocket of the receptor generated by the agonistic form. However, the affinity of TLR7 and TLR8 to mononucleosides could be greatly increased by the binding of ssRNA with a shift in TLR8 K_d from 55 μM to 1 μM. This observation was confirmed in a cell-based immune response–activation assay showing that uridine stimulated substantial NF-κB activity only in the presence of a UG-rich ssRNA. The synergistic activation is also observed for guanosine derivatives and chemical ligands (e.g., loxoribine) in the presence of oligonucleotides in TLR7 [3,7]. Likewise, a synergistic activity between uridine analogs and ssRNA is demonstrated in TLR8 [7]. These findings suggest that an oligonucleotide binding at the second site increases receptor affinity to mononucleosides or chemical ligands at the first site, possibly via allosteric regulation.

Accumulating evidence suggests that proteolytic processing in endolysosomes is required for generating functional, mature TLRs. This requirement as well as localization within endosomal compartments reinforced the prevention of the activation of

TLRs by self-derived nucleic acid. In addition, proteolytic digestion of nucleic acid-sensing TLRs may restrain inappropriate activation in response to host DNA/RNA. The processing at the Z-loop of human TLR8 mediated by furin-like proprotein convertase and cathepsins produces functional TLR8 capable of ligand binding and signaling in endolysosomes. In addition, the cleaved form of TLR8 has been found to be predominant in immune cells. However, after proteolytic cleavage, the extracellular and intracellular domains must remain associated to provide the functional activity of the receptor [8,9]. This notion is supported by structural studies demonstrating that the N- and C-terminal domains of TLR7, TLR8, and TLR9 are connected and involved in ligand binding after Z-loop cleavage [3,4,6]. Biochemical and biophysical studies have revealed that the uncleaved Z-loop prevents the formation of the TLR8 dimer, which is essential for its activation [10]. Crystallographic analysis demonstrated that the uncleaved Z-loop located on the ascending lateral face prevents the approach of the dimerization partner by steric hindrance. Similarly to TLR8, TLR7 and TLR9 also contain the Z-loop, and thus, these proteins might also have this autoinhibition mechanism. This notion is supported by the observation that TLR9 with the uncleaved Z-loop was unable to dimerize irrespective of CpG-DNA presence [6].

Structural information on the nucleic acid-sensing TLRs adds to and provides new insights into the areas of the regulation of nucleic acid-sensing TLRs, proteolytic processing of TLR, synergistic activation mechanism through multiple ligand-binding sites, and nucleic acid degradation. In particular, the requirement of DNA and RNA processing for the activation of immune responses is a potentially paradigm-shifting discovery, emphasizing the importance of specific enzymes, such as RNases, DNases, and phosphatases (Fig. 4). Indeed, a recent study shows that DNA digestion by DNase II is required for TLR9 activation [11]. These newly obtained mechanistic insights may be a crucial step towards developing therapeutic drugs (agonists and antagonists) that target TLRs for the modulation of host response to pathogens.

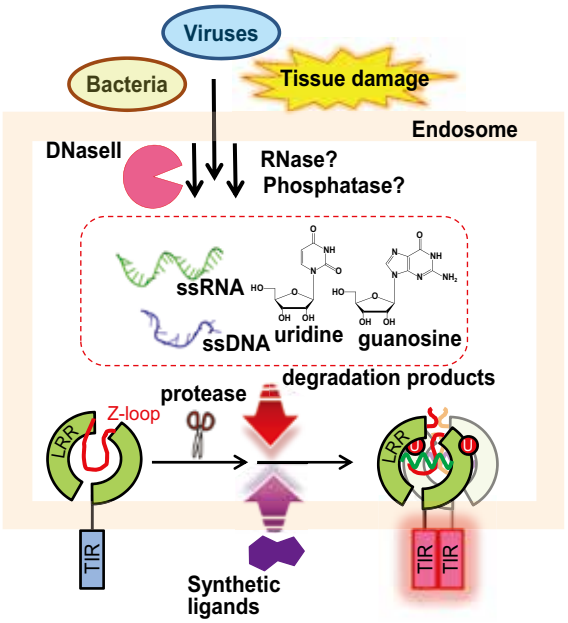

Fig. 4. **Proposed regulation mechanism of nucleic acid-sensing TLRs** Pathogen- and self-derived nucleic acids can be taken up to endolysosomes where they are digested by appropriate enzymes (DNase II, RNase, and/or phosphatase) to generate nucleosides and degradation products that are recognized by TLRs. TLRs 7–9 with the uncleaved Z-loop are monomers. Following Z-loop cleavage, TLRs 7–9 transforms into the activated dimer upon binding to processed nucleic acid agonists. In one example, TLR8 activated by uridine and ssRNA is shown. Synthetic ligands can directly activate TLR7/8.

Toshiyuki Shimizu

Graduate School of Pharmaceutical Sciences, The University of Tokyo

Email: shimizu@mol.f.u-tokyo.ac.jp

References

[1] C.A. Janeway *et al.*: Annu. Rev. Immunol. **20** (2002)197.
[2] T. Kawai *et al.*: Nat. Immunol. **11** (2010) 373.
[3] Z. Zhang *et al.*: Immunity **45** (2016) 737.
[4] H. Tanji *et al.*: Science **339** (2013) 1426.
[5] H. Tanji *et al.*: Nat. Struct. Mol. Biol. **22** (2015) 109.
[6] U. Ohto *et al.*: Nature **520** (2015) 702.
[7] T. Shibata *et al.*: Int. Immunol. **28** (2015) 211.
[8] M.M. Hipp *et al.*: J. Immunol. **194** (2015) 5417.
[9] M. Onji *et al.*: Nat. Commun. **4** (2013) 1949.
[10] H. Tanji *et al.*: Proc. Natl. Acad. Sci USA **113** (2016) 3012.
[11] M.P. Chan *et al.*: Nat. Commun. **6** (2015) 5853.