DE GRUYTER

Pure Appl. Chem. 2014; 86(10): 1483–1538

IUPAC Technical Report

John A. Robinson* and Kerstin Moehle

Structural aspects of molecular recognition in the immune system. Part II: Pattern recognition receptors

Abstract: The vertebrate immune system uses pattern recognition receptors (PRRs) to detect a large variety of molecular signatures (pathogen-associated molecular patterns, PAMPs) from a broad range of different invading pathogens. The PAMPs range in size from relatively small molecules, to others of intermediate size such as bacterial lipopolysaccharide, lipopeptides, and oligosaccharides, to macromolecules such as viral DNA, RNA, and pathogen-derived proteins such as flagellin. Underlying this functional diversity of PRRs is a surprisingly small number of structurally distinct protein folds that include leucine-rich repeats in Toll-like receptors (TLRs) and NOD-like receptors (NLRs), the DExH box helicase domain in RIG-like receptors (RLRs), and C-type lectin domains (CTLDs) in the C-type lectins. Following PAMP recognition by the PRRs, downstream signaling pathways activate the innate immune system to respond to invading pathogenic organisms. The resulting stimulatory response is also vital for a balanced adaptive immune response to the pathogen, mediated by circulating antibodies and/or cytotoxic T cells. However, an aberrant stimulation of the innate immune system can also lead to excessive inflammatory and toxic stress responses. Exciting opportunities are now arising for the design of small synthetic molecules that bind to PRRs and influence downstream signaling pathways. Such molecules can be useful tools to modulate immune responses, for example, as adjuvants to stimulate adaptive immune responses to a vaccine, or as therapeutic agents to dampen aberrant immune responses, such as inflammation. The design of agonists or antagonists of PRRs can now benefit from a surge in knowledge of the 3D structures of PRRs, many in complexes with their natural ligands. This review article describes recent progress in structural studies of PRRs (TLRs, NLRs, CTLs, and RLRs), which is required for an understanding of how they specifically recognize structurally diverse "foreign" PAMPs amongst a background of other "self" molecules, sometimes closely related in structure, that are present in the human body.

Keywords: immune system; IUPAC Chemistry and Human Health Division; NOD-like receptors (NLRs); pathogen-associated molecular patterns (PAMPs); pattern recognition receptors (PRRs); RIG-like receptors (RLRs); Toll-like receptors (TLRs).

*Corresponding author: John A. Robinson, Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, e-mail: john.robinson@chem.uzh.ch
Kerstin Moehle: Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

CONTENTS

1 Introduction to pattern recognition receptors

The immune system has evolved to protect organisms as diverse as plants, flies, and animals against infection by invading microorganisms. A critical first step in achieving this protection is to quickly recognize when infectious agents are present; the immune system must distinguish what is foreign (or "non-self") from what is normally present in the body (i.e., "self"). It is the main function of the innate immune system to provide this rapid surveillance, using germ-line encoded proteins called pattern recognition receptors (PRRs) [1–4]. PRRs function by recognizing "danger signals" and then initiating an appropriate innate immune response.

Danger signals can be derived from both foreign pathogens as well as endogenous sources, such as lysed cells. The PRRs comprise four main families of proteins, some membrane-bound and others soluble cytoplasmic proteins that recognize molecules (or molecular patterns) associated with microbes – so-called pathogen-associated molecular patterns (PAMPs). The activation of PRRs by binding to PAMPs initiates intra-cellular signaling cascades and altered patterns of gene expression that lead to a rapid first line defense against the invading microorganism, as well as promoting activation and maturation (more slowly – over days to weeks) of the adaptive immune system. The adaptive immune response is mediated mainly by B and T cells. Signals from the innate immune system drive the selective expansion and activation of B- and T-cell populations with specificity for the infectious agent, optimized through cycles of somatic mutation and selection (Fig. 1) [5]. The main effector mechanisms of adaptive immunity include the production of antibodies by B cells (to act as blocking antibodies, or as opsonins for complement- and phagocyte-mediated killing), and the killing of infected host cells by cytotoxic T cells. In addition, activation of adaptive immunity results in the production of memory B and T cells, which can provide long-term specific protection (e.g., over years to a whole lifetime) against subsequent infections by a pathogen bearing the same antigens [6, 7].

Adherence of microbes to the surface of cells of the innate immune system, mediated through binding of PAMPs to cell surface PRRs, actively promotes ingestion of the microbe and entrapment and digestion within phagosomes. On the other hand, soluble cytoplasmic PRRs act as sensors for intracellular pathogens. Through PAMP–PRR interactions, innate immune cells are activated and respond by producing inflammatory cytokines and chemokines that trigger responses from other cell types. Furthermore, activation of the innate immune system leads to deployment of a range of killing factors to attack invading pathogens. These include bactericidal enzymes, toxic chemicals such as reactive oxygen and nitrogen species, and antimicrobial peptides (e.g., defensins). The detection of certain PAMPs may also activate the complement system, a cascade of triggered enzyme reactions that can lead to tagging (or opsonization) of the microbe surface with the complement protein C3b, which marks it out for destruction by other components of the immune system. Finally, PRR–PAMP interactions stimulate cells of the adaptive immune system, including antigen-presenting cells

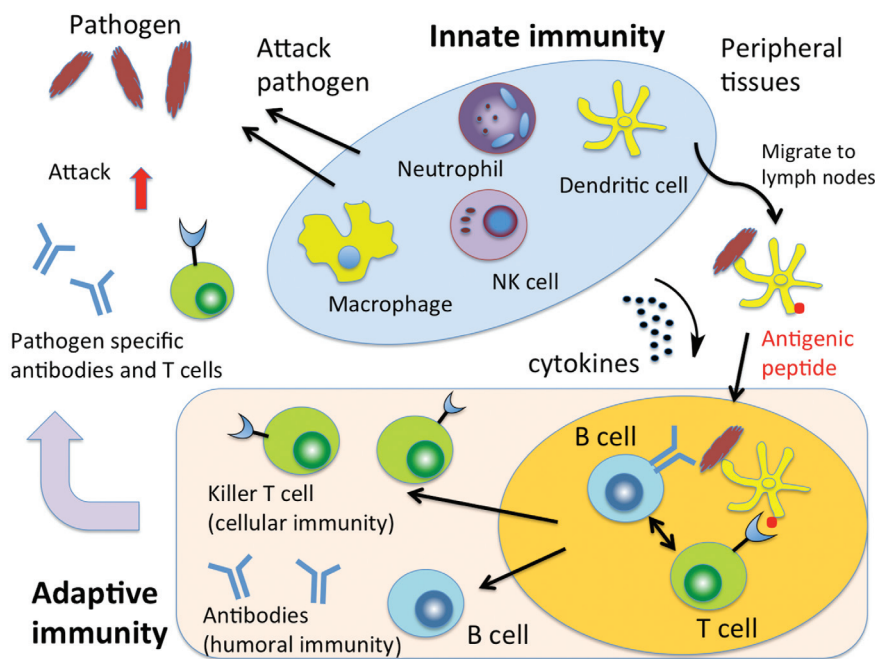**Fig. 1** The innate immune system operates largely in peripheral tissues, where it recognizes foreign pathogens. Upon binding of PAMPs to PRRs, the innate immune system is activated and provides a first line of defence against the invader. Subsequently, the adaptive immune system is activated, and with help from the innate immune system, pathogen-specific antibodies and killer T cells are produced (adapted from [11]).

(APCs), such as dendritic cells (DCs) and B and T cells. DCs reside in a quiescent state in peripheral tissues, where they continuously sample their environment by phagocytosis and pinocytosis. DCs are equipped with a suite of PRRs for binding to PAMPs, thereby facilitating rapid up-take of pathogens by phagocytosis. Inside APCs, foreign proteins are digested into peptide fragments, which include some that act as T-cell epitopes. A dramatic increase in the expression of surface MHC-I and -II is elicited by PAMP–PRR binding. This leads to enhanced surface presentation of T-cell peptide epitopes in complexes with the membrane-bound MHC molecules, ready for recognition by T-cell receptors.

Immature DCs are relatively non-motile, but after activation they become highly motile and migrate through the lymphatic system, carrying pathogen-derived antigens to secondary lymphoid tissues, such as the lymph nodes. Circulating naïve B and T lymphocytes also halt in the lymph nodes, where they scan APCs and intact antigens draining through in interstitial fluids from peripheral tissues. Intact antigens can be recognized by B-cell receptors (BCRs) on the surface of B cells. Viruses and bacteria typically display a repetitive and closely spaced array of epitopes across their surface. Such an arrangement of epitopes is specifically recognized as foreign by B cells [8]. This occurs through cross-linking of multiple BCRs by multivalent binding to surface arrays of antigens, which generates a powerful activation signal, and initiates the process of B-cell activation and maturation [9]. The activated B and T cells then interact within the lymph nodes in a process that leads to the production of antibody-secreting B cells and cytotoxic T cells optimized to recognize and remove foreign pathogens. The activated B and T cells also harbor PRRs, and are further stimulated by repeated interaction with PAMPs during the course of the adaptive immune response [10]. PAMP–PRR interactions thus play a key role in immune responses to invading pathogens, from its beginnings through the later stages of maturation of the adaptive immune response (Fig. 1). For this reason, both natural and synthetic ligands of PRRs are now attracting great interest as immune modulators in the fields of vaccinology and immunotherapy [6, 11–13].

PRRs can be divided into four main families, based upon their structural features as well as the type of PAMP they recognize [14]. These include the Toll-like receptors (TLRs), the C-type lectin receptors (CLRs), NOD-like receptors (NLRs), and RIG-like helicase receptors (RLRs). A review of newly described PRRs appeared recently [15].

1.1 Toll-like receptors (TLRs)

One of the major families of PRRs are the TLRs, so named because of their homology with the *Toll* receptor, first discovered in the 1980s in the fruit fly *Drosophila* [16, 17]. The Toll receptor plays a role in the development of the dorsoventral body axis in *Drosophila*, acting as a cell-surface receptor for the cytokine ligand Spätzle [18]. Later it was discovered that Toll and related receptors also function in the innate immune response to fungal and bacterial infection in the fruit fly, stimulating the release of defensins and related antimicrobial peptides [19]. As genome sequencing advanced, about 10 directly homologous receptors were later discovered in vertebrates, and these became known as the Toll-like receptors (for a recent historical account, see [20]). The ligands for most of these receptors are now known. The ligands display remarkable structural diversity, and include bacterial peptidoglycan (PGN), lipoproteins, lipopolysaccharide (LPS), mycobacterial lipoarabinomannan, yeast zymosan, and bacterial flagellin, as well as pathogen-derived nucleic acids [21].

To date, 10 TLRs have been reported in humans and 13 in mice. The TLRs are type I integral membrane glycoproteins that are located either in the outer cell membrane (TLR1, 2, 4, 5, 6, 10, 11, 12), or in endosomal membranes with their ectodomains facing the internal space of the endosome (TLR3, 7, 8, 9, 13) (Fig. 2). They contain an N-terminal extracellular domain (ECD) of between 550 and 800 residues, a single helical transmembrane domain of ≈20 residues, and a C-terminal intracellular signaling domain of ≈180 residues, called a "Toll IL-1 receptor" (TIR) signaling domain, because it shares homology with the signaling domains of IL-1R family members. The crystal structure of the TIR domain of TLR1/2 was published in 2000 [22]. Some TLRs act as heterodimers, including TLR1/2 and TLR2/6, which respond to different bacterial lipoproteins. All TLRs share a common basic 3D architecture, with multiple leucine-rich repeats (LRRs) arranged in a horseshoe- or crescent-shaped structure that together comprises the N-terminal ectodomain responsible for PAMP-binding. A single
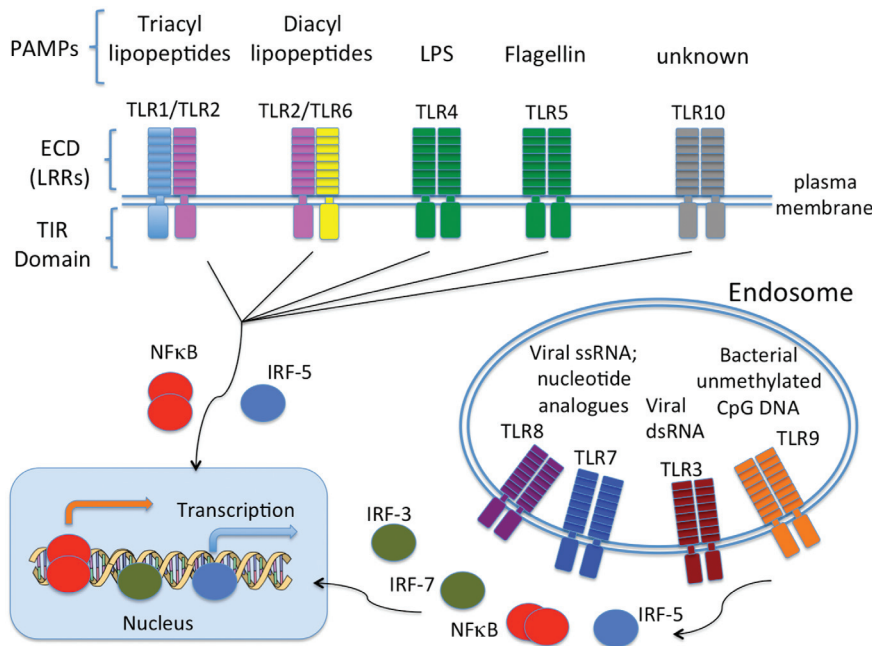
Fig. 2 Ten TLRs have been identified in humans, and are located either on the outer cell membrane or the membrane of endosomes within the cytoplasm of immune cells. When PAMPs bind to the ectodomain of the TLR, signals are transmitted into the cell via intracellular (TIR) domains [22], which bind and activate transcription factors [NFkB and/or members of the interferon-regulated factors (IRFs)]. The transcription factors promote expression of numerous genes that alter the activation state of the cell, as well as for inflammatory cytokines and chemokines that are exported and signal to other cells (see Fig. 1).

transmembrane segment links the ectodomain to the C-terminal domain, which interacts with adapter proteins in the cytoplasm, an event that culminates in activation of transcription factors, such as NFκB and members of the interferon-regulated factors (IRFs) family, which in turn leads to altered patterns of gene expression [23, 24].

1.2 C-type lectin receptors (CLRs)

The C-type lectin receptor (CLR) superfamily is a large group of proteins that are widely distributed in nature [25, 26]. C-type lectins are Ca^{2+}-dependent carbohydrate-binding proteins that contain one or more C-type lectin (or lectin-like) domains (CTLDs), with related folds dictated by a common sequence motif. When such domains bind to carbohydrates, they are called carbohydrate recognition domains (CRDs). A CRD is present in all Ca^{2+}-dependent lectins, but not in other (Ca^{2+}-independent) types of animal lectins. A recent review grouped the CTLDs into 17 different classes [25]. Of special interest here are the class-II asialoglycoproteins on APCs that function as PRRs in the innate immune system. These type-II membrane receptors contain a short cytoplasmic domain, a single transmembrane segment, and an ectodomain comprising a stalk region involved in oligomerization, which is linked to one or more CRDs. Professional APCs, such as macrophages, DCs, and B cells, are equipped with receptors that collectively allow them to recognize, capture, and internalize foreign antigens. Besides the B-cell receptor found only on B cells, other APCs use germ-line encoded membrane proteins for detecting foreign antigens that include Fc receptors, complement receptors, the TLRs, and the CLRs. TLRs are not able to promote internalization of antigens, whereas this is an important function of CLRs. Representatives of the class-II CLRs are ICAM-3, DC-SIGN, MGL, LANGERIN, MINCLE, DECTIN-2, BDCA-2, DCIR, and DCAR. The class-V CLRs are also known as type-II natural killer (NK) cell receptors, and include DCAL-1, DCAL-2, and dectin-I. Finally, the class-VI CLRs include the mannose receptor (MR) and lymphocyte antigen 75 (DEC-205).

CLRs are important in immune recognition of glycoproteins and cell wall glycolipids derived from viruses and microbes, and mediate pathogen uptake and phagocytosis [27, 28]. However, the biological functions of

CTLs are more complex than just facilitating antigen uptake by DCs. Many of the carbohydrates recognized by CLRs are not unique to pathogens but can also be found on many self-glycoproteins [27]. Hence, self- and non-self-structures may be recognized by CLRs, which is a reflection of the multiple and complex functions these receptors, as a whole, have in the maintenance of immune homeostasis [29–32]. Ligand binding to these receptors can also activate signaling pathways that modulate TLR signaling, activate innate immune responses, and influence adaptive immune responses [30, 31].

The C-type (calcium-dependent) lectin (CTL) family of proteins is found on the surface of cells of the innate immune system, including macrophages and DCs [28, 29, 33]. The membrane-bound CTLs bind oligosaccharides, in particular, containing D-mannose and L-fucose residues, using highly conserved carbohydrate recognition domains (CRDs). Many CTLs on APCs contain only a single CRD, well-known examples being DC-SIGN, DCIR, MGL, and Langerin. Others, such as DEC-205 and the mannose receptor (MR), possess 8–10 CRDs within their extracellular regions [25, 34–36]. According to their carbohydrate specificities, CLRs tend to be classified as either mannose/fucose- or galactose-recognizing lectins. Some CLRs share carbohydrate ligands. For example, the trisaccharide Lewis^x can serve as a ligand for DC-SIGN and the MR. However, ligand promiscuity is also seen, for example, with DC-SIGN, which is able to recognize efficiently different Lewis-type epitopes as well as high-mannose glycans. CLRs have now become attractive targets for ligand design, for use in targeting antigens to DCs, and for modulating antigen-specific immune responses [28, 29, 33].

1.3 NOD-like receptors (NLRs)

The nucleotide-binding and oligomerization domain (NOD)-like receptors (NLRs) function as intracellular surveillance sensors to detect microbial products and danger signals inside cells [37–41]. Some NLRs are also involved in inflammatory responses and apoptosis, and thus as a whole the NLRs mediate diverse biological functions in innate immunity [39, 42–46]. NLR family members are found in plants, invertebrates, and mammals, although NLR structure and function has been most thoroughly studied in mice and humans [47]. The NLRs are subdivided into **five subfamilies**, into which the 22 human NLRs are grouped as follows; **NLRA** (CIITA), **NLRB** (NAIP), **NLRC** (NOD1, NOD2, NLRC3, NLRC4, NLRC5, NLRP and NRLX), **NLRP** (NLRP1-14), and **NLRX** (NLRX1) (Fig. 3) [48]. Some NLRs activate a proinflammatory response (such as NOD1 and NOD2), others form a large multiprotein complex called an inflammasome [47, 49] resulting in cleavage of procaspase-1 and formation of activated caspase I (NLRP1, 3, 6, 7, and 12, and NLRC4), others provide immunoregulatory functions (NLRC3 and 5, CIITA, and NLRP10), while others have roles in development (NLRP2, 5, 7, and 14) [48, 50].

The NLRs are soluble cytoplasmic proteins, important for sensing infectious agents that have gained access to the interior of the cell [48, 50–52]. There are 22 NLRs in humans and 34 in mice [53], which exhibit characteristic multidomain structures, typically consisting of a central nucleotide binding domain (NBD) that contains a
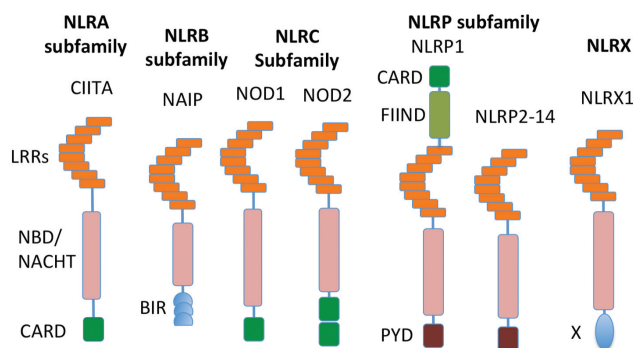

Fig. 3 Domain architecture of representative human NLRs [47, 52]. Abbreviations: LRRs, leucine-rich repeats for ligand sensing; NACHT, a nucleotide binding and oligomerization domain; PYD, pyrin effector domain; BIR, baculovirus inhibitor repeat effector domain; CARD, caspase activation and recruitment effector domain; NBD, nucleotide binding domain; FIIND, function to find domain.

NACHT and other subdomains, an N-terminal effector binding domain most commonly a CARD (caspase activation and recruitment domain), a BIR (baculovirus inhibitor repeat), or a pyrin (PYD) domain, and a C-terminal putative ligand-binding domain comprising multiple leucine-rich repeats (LRRs) (Fig. 3) [39, 40, 48]. The effector domains bind downstream signaling molecules, ultimately leading to activation of protein kinases, transcription factors, proteases, and other components of host defense and inflammatory responses [38, 50].

Some of the PAMPs detected by NLRs have been identified. For example, NOD1 and NOD2 sense fragments of PGN [37]. Macrophages contain intracellular hydrolases that digest bacterial PGN. NOD1 detects a minimal PGN fragment containing γ-D-glutamyl-*meso*-diaminopimelic acid that is found almost exclusively in Gram-negative bacterial PGN [54]. On the other hand, NOD2 detects muramyl-dipeptide (MurNAc-L-Ala-D-isoGln), which is a common motif in PGN in almost all bacteria [55, 56]. NLRP3 and NAIP respond to a remarkably diverse set of PAMPs [39], whereas NLRC4 senses bacterial flagellin [57, 58], and NLRX1 binds to RNA [59].

1.4 RIG-like helicase receptors (RLRs)

How does a cell distinguish viral nucleic acid (RNA) in the cytoplasm from its own? An important clue came by recognizing that viral RNA is structurally different in ways that mark it out as foreign to the cell. Thus, dsRNA is virtually absent in mammalian cells, whereas some viruses generate dsRNA as a byproduct of viral RNA replication, containing transiently an uncapped 5′-triphosphate group. Cellular primary transcripts also contain 5′-triphosphate, however, many self-RNAs undergo processing steps that remove or mask the 5-triphosphate moiety. For example, mRNAs are capped by methylation at the 5′-terminus, tRNAs undergo cleavage of 5′-PPP groups before they reach the cytoplasm, and ribosomal RNAs form complexes with ribosomal proteins. The retinoic acid-inducible gene-I (RIG-I)-like receptors (RLRs) are constitutively expressed in the cytoplasm of most cells, including DCs and macrophages as well as non-immune cells, where they act as cytosolic sensors of viral RNA [60, 61].

RLRs provide a first line of defense against infection by RNA viruses, including influenza, rift valley, measles, vesicular stomatitis, and hepatitis C viruses [61]. There are three members of the RLR family: RIG-I (also known as DDX58), "melanoma differentiation associated factor gene 5" (MDA5), and "laboratory of genetics and physiology 2" (LGP2). Recognition of the 5′-PPP terminus unique to viral RNA, either in a dsRNA or a hairpin-looped structure, a so-called pan-handle-like RNA, is an important function of RIG-I [62, 63]. The prototypical ligand of RIG-I is a short RNA with a blunt-ended base-paired, and uncapped 5′-triphosphate terminus. Until very recently, much less was known about the nature of RNAs recognized by MDA5. MDA5 generally responds to long dsRNA molecules. However, RIG-I and MDA5 may also be activated by self-RNAs that are cleaved by RNaseL. The function of the third RLR, called LGP2, has been less well studied, but seems to participate in RIG-I and MDA5-dependent antiviral responses [60, 61].

RIG-1 and MDA-5 possess a conserved central helicase-like domain, linked to two N-terminal caspase activation and recruitment domains (CARDs), and a Zn^{2+}-containing C-terminal domain (CTD) (Fig. 4). LGP2
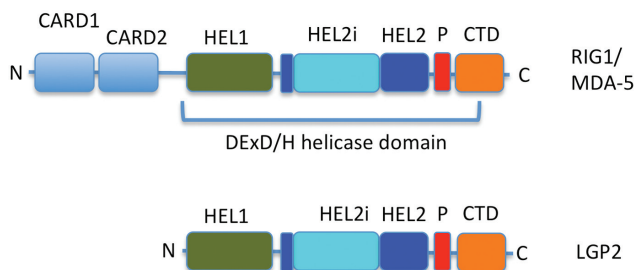

Fig. 4 Domain structures of RLRs. RIG-1 and MDA-5 have the same domain architecture. Abbreviations: CARD, caspase activation and recruitment effector domain; CTD, C-terminal regulatory domain; HEL2i, helicase insertion domain; HEL1/2 helicase 1/ helicase 2; P, pincer domain.

has similar domain architecture, although it contains no CARD domains. The ligands that function as RLR activators are typically virus-specific RNA structures, including 5'-triphosphate linked to single-stranded or blunt-ended, double-stranded RNA, as well as poly I:C (a synthetic analog of viral dsRNA). Upon binding to such PAMPs, RIG-1 initiates a signaling cascade that induces innate immune defenses and inflammatory cytokines to establish an antiviral state [64]. However, knowledge of the complex regulatory mechanisms that control signaling through these PRRs is incomplete. Aberrant RIG-1 signalling can lead to apoptosis, altered cell differentiation, inflammation, autoimmune disease, and cancer. Apart from RLRs, several other members of the helicase superfamily have been proposed to participate in intracellular sensing of pathogen-derived nucleic acids, including DDX1/3/21/36, DHX9/36, and STING [60].

2 Structures of toll-like receptors

2.1 Overview

Several reviews on TLR structural biology have been published recently [16, 65, 66]. TLR ECDs are constructed of tandem copies of a structural motif known as the leucine-rich repeat (LRR). LRRs are found in several thousand proteins involved in a wide variety of physiological functions in all kingdoms of life, including immune responses, signal transduction, cell-cycle regulation and enzyme regulation (for reviews, see [67–69]). These proteins typically contain 2–40 tandem LRR modules, which form curved horseshoe- or solenoid-like structures that seem to be well suited for protein–protein and protein–ligand interactions. The 10 human TLRs contain 19–25 LRR modules. Each LRR typically contains a ≈20–30 residue consensus sequence (Fig. 5a), which includes the characteristic motif LxxLxLxxNxL (x = any amino acid), although other hydrophobic residues can substitute for Leu, and other amino acids capable of donating hydrogen bonds (T, S, C) can substitute the Asn residues [67, 70]. Each LRR module adopts a folded structure comprising one β-strand that includes the entire consensus motif, followed by a bridging loop that often contains helical and/or turn-like elements that fold back over the β-strand, thereby burying the hydrophobic L
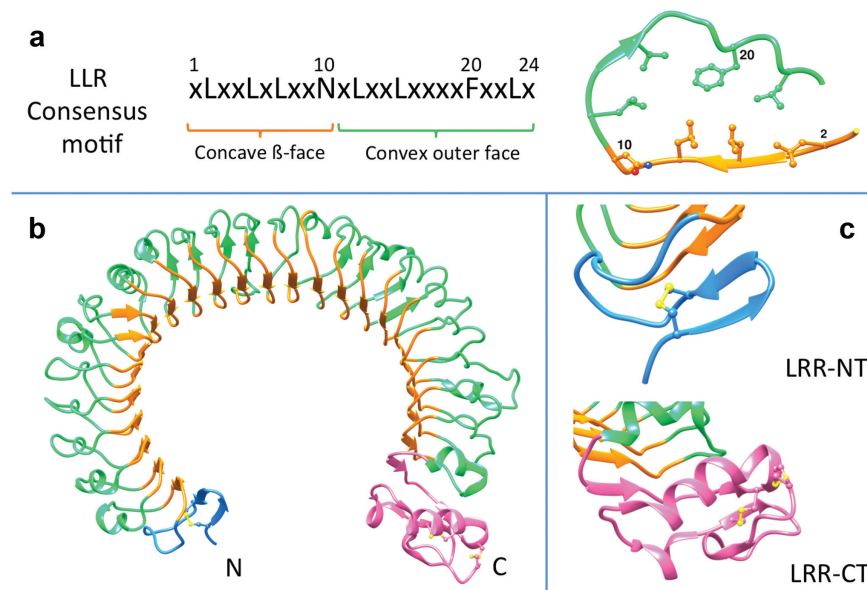

Fig. 5 (a) Characteristic LRR motif, and structure of one LRR module. Orange, β-strand; green, bridging loop. (b) The typical horseshoe, or solenoid-like structure of a TLR ectodomain. (c) The disulfide-rich capping modules at the N- (LRR-NT) and C-terminus (LRR-CT).

residues (Fig. 5a). The bridging loop is linked to the β-strand of the next LRR, which is aligned so that the β-strands form a regular hydrogen-bonded parallel β-sheet. This arrangement can now be repeated by addition of more LRRs, which generate a horseshoe- or solenoid-like tertiary structure, as shown in Fig. 5b. The resulting curved parallel β-sheet forms the inner concave face of the horseshoe, whereas the residues in the adjacent bridging (often helical) loops form the outer convex surface of the horseshoe. The ECDs of most TLRs are glycoproteins, in which surface-exposed Asn residues carry N-linked glycosylation. The side-by-side nestling together of LRRs in these proteins ensures that the hydrophobic residues are buried internally. The situation is different for the N- and C-terminal LRR modules, where one side of the hydrophobic core will not be covered by another LRR module. Instead, capping modules are added to the termini (LRR-NT and LRR-CT) (Fig. 5c). The capping modules do not show the same conserved sequence and structural motifs. They typically comprise smaller disulfide-bonded folds that pack across the hydrophobic core and leave polar surface residues exposed to the solvent.

The LRR horseshoe geometries have been characterized by three parameters – the radius of the horseshoe arc ("radii" R), the mean rotation angle about the central axis relating one β-strand to the next ("twist"), and the tilt angle of the parallel β-strand direction per turn ("tilt") [68]. The observed values of R, for example, range from 1.5 to 4.7 nm, so the arc of LRR family members can possess quite different degrees of curvature. LRR family proteins have been subdivided into seven subfamilies based on sequence and structural comparisons. They are "RI-like", "CC", "bacterial", "SDS-22-like", "plant-specific", "typical", and "TpLRR" [69, 71]. The "typical" subfamily includes all the TLRs. However, phylogenetic analyses suggested that two distinct subclasses of "typical" TLRs exist, the single- and three-domain TLR subclasses, which diverged early in evolution [66, 72].

The ECDs of TLR1, 2, 4, and 6 belong to the three-domain subclass [66]. These TLRs have two structural transitions in the concave β-sheet face of the ECD, thereby creating N-terminal, central, and C-terminal subdomains (Fig. 6). The central domains have radii, twist, and tilt angles that deviate markedly from the standard values of the "typical" single-domain subfamily members. This appears to be caused by several unusual features, including LRRs in the central domain that lack the characteristic Asn residue within the consensus motif. This Asn side chain usually interacts in a ladder of hydrogen bonds with surrounding backbone carbonyl oxygens, thereby stabilizing the horseshoe structure. In addition, the LRR modules of the central domains vary in size from 20 to 33 residues, whereas in the N- and C-terminal domains, the LRR modules are mostly 24 residues. The LRR modules in the central region have bulkier α-helical segments on the exposed convex face, which take up more space, leading in turn to greater curvature and a smaller arc radius in the central domain. These structural features are also important for the biological function of these TLRs because the central domains play an important role in binding ligands or co-receptors. Although no crystal structure is available, TLR10 also appears to contain broken Asn ladders and greater variation in module length in the central part of the ECD, and so is also likely to belong to the three-domain subfamily.

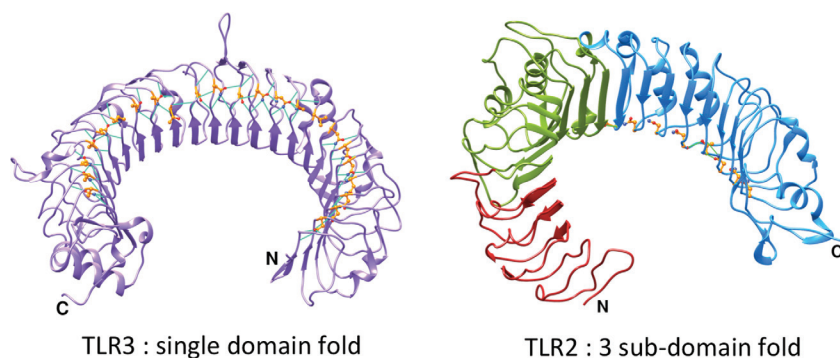TLR3 : single domain fold   TLR2 : 3 sub-domain fold

Fig. 6 Single-domain (left) and three-domain (red, green, and blue, right) TLR subclasses. The characteristic Asn residues (show in orange/red/blue ball-and-stick) in each repeat motif, and the ladder of hydrogen-bonds (in light blue) they form are shown.

In contrast, TLR3, 5, 7, 8, and 9 belong to the single-domain subfamily. Their LRR modules have conserved Asn residues within the consensus motif and relatively uniform module lengths. Functional and phylogenetic analyses provide support for these correlations, which in the cases of TLR3 and 8 is also supported by structural (crystallographic) information. TLR7, TLR8, and TLR9 are predicted by sequence comparisons to contain "island regions", comprising substantial domains that loop out of the external surface of the horseshoe architecture [73].

Strong evidence exists that active TLRs function as homo- or hetero-dimers, and/or as part of supramolecular complexes with other membrane-associated proteins. For example, TLR2 signals as a heterodimer in a complex with either TLR1 or TLR6. Crystal structures of the ECDs of the TLR1/2 and TLR2/6 heterodimers show how specific lipopeptides are recognized by these receptors (vide supra) [74, 75]. In the case of the TLR2/6 heterodimer, a plasma membrane protein (CD36) of the class B scavenger receptor family, also participates in signaling [76]. Although human TLR3 ECD is found as a monomer in the crystal structure [77, 78], it binds its target double-stranded (ds) RNA as a dimer [79, 80]. Moreover, clustering of multiple TLR3 homodimers may occur at the cell surface through binding to long dsRNAs [79, 81]. FRET studies suggest that TLR9 exists as preformed inactive dimers in the cell membrane that change conformation upon binding to DNA rich in CpG dinucleotides. Ligand binding then results in allosteric changes in the cytoplasmic signaling domains, which lead to receptor activation and signaling [82]. TLR5, the only protein-binding TLR, responds to a monomeric form of bacterial flagellin that, in oligomeric form, constitutes the flagellar tail responsible for swimming locomotion in β- and γ-proteobacteria. A crystal structure of an engineered TLR5-flagellin complex revealed a 2:2 subunit stoichiometry, with two TLR5 ECDs in a tail-to-tail organization [83]. Finally, TLR4 forms a stable 1:1 heterodimer with MD-2, and the resulting TLR4-MD-2 complex is responsible for recognition of lipopolysaccharide (LPS) from Gram-negative bacteria. Cells transfected with TLR4 are not responsive to LPS, but are when complemented with MD-2 [84]. Moreover, another membrane-associated protein, CD14, is required for full ligand recognition and signaling through TLR4. TLR4 may be organized into a supramolecular complex through interactions with MD-2 and CD14 (and perhaps other components) at the cell surface [76, 85, 86].

TLRs interact with an unusually diverse variety of ligands. These ligands range from naturally occurring highly nonpolar lipids, such as LPS and lipopeptides, through to polar macromolecules such as DNA, RNA, and the protein flagellin. It is also important to remember that the same TLRs from different species may exhibit differences in the way they recognize and respond to the same or closely related ligands [87].

2.2 Structures of TLR1/2 and 2/6

Various earlier reports suggested that TLR2 is able to recognize a diverse array of microbial structures, although most are glycolipids (such as lipoteichoic acid (LTA), lipoarabinomannan, and LPS), lipopeptides/lipoproteins, or glycosyl-phosphatidyl-inositol (GPI)-anchored structures [88]. However, some purported TLR2 ligands lack clear structural relationships to lipidic molecules, such as fragments of PGN and the polysaccharide zymosan isolated from yeast. This led to the suggestion that TLR2 is a rather "promiscuous receptor" [89]. This view has now changed, with the realization that some of the potential ligands tested in earlier studies were isolated from biological sources and were likely to contain trace amounts of bacterial lipoproteins/lipopeptides [88]. The extent of the analytical problem can be understood by noting that some PAMPs may activate PRRs at picomolar (pM) concentrations, whereas some cellular studies of TLR2 activation were performed with "purified" ligands in the micromolar range. Ideally, ligands of synthetic origin should be tested in receptor activation assays, although their structural and stereochemical complexity often makes this difficult. Very recently, structural studies have reinforced the view that only lipopeptides/lipoproteins containing Pam_2Cys or Pam_3Cys [or those with related acyl chains (Fig. 7)] represent specific ligands of TLR2, recognized by heterodimers of TLR1/2 or TLR2/6. In contrast, the lipid anchor in LTA, the GPI-anchored lipids in lipoarabinomannan, as well as related glycolipids frequently found in membrane of higher organisms do not activate TLR2 at physiologically relevant concentrations [88].
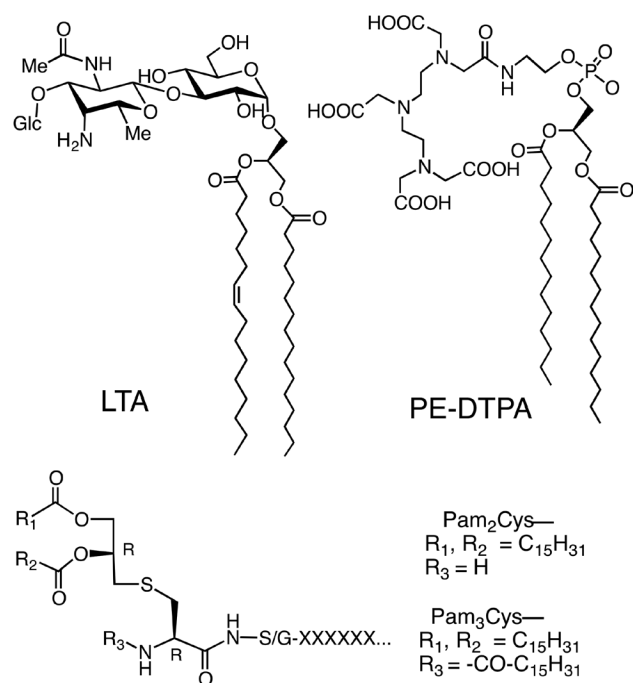
Fig. 7 Structures of lipoproteins Pam₂Cys and Pam₃Cys, and the lipid derivatives PE-DTPA and LTA.

Crystal structures have revealed that the TLR1/2 heterodimer binds to lipoproteins with a triacylated N-terminus (such as Pam₃Cys), while the TLR2/6 heterodimer binds to lipoproteins with a diacylated N-terminus (such as Pam₂Cys). All bacterial lipoproteins contain a glycerol moiety linked via a thioether to the side chain of an N-terminal Cys residue (S-[2,3-bis(acyloxy)-(2R)-propyl]-(R)-cysteinyl), and acylated with two long-chain fatty acids. Lipoproteins anchored in the membrane of Gram-positive bacteria generally have a free N-terminus. In Gram-negative bacteria, however, lipoproteins contain three lipid chains (Fig. 7); a third acyl chain is attached directly to the N-terminus of the peptide/protein. During the biosynthesis of lipoproteins [90]), an inner membrane enzyme (Lgt) transfers the diacylglycerol moiety from membrane phospholipids to the N-terminal cysteine residue of the pre-lipoprotein. The acyl chains are typically palmitoyl groups, although other chain lengths and levels of unsaturation are found in different bacteria [90]. Gram-positive bacteria generally lack the lipoprotein *N*-acyltransferase (Lnt) required to attach the third amide-linked acyl chain to the N-terminus [91]. Mycobacteria are a special case, since they are high-GC Gram-positive bacteria, albeit with a special outer layer of unique lipid components called mycolic acids covalently linked to the PGN. They contain the Lnt enzyme and so are able to produce triacylated lipoproteins carrying mycobacteria-specific fatty acids [92]. Mycoplasma, on the other hand, are low-GC Gram-positive bacteria without cell walls (no PGN or LPS), with a cytoplasmic membrane that is stabilized by the presence of sterols and lipoglycan (long-chain heteropolysaccharides covalently linked to membrane lipids [93, 94]). Lipoproteins in the cell membrane of Mycoplasma are of the Pam₂Cys variety [95, 96].

The residues immediately downstream of the N-terminal Cys in bacterial lipoproteins are variable in sequence, but the first is typically Gly or has a small polar side chain. The following residues appear not to exert a strong influence on signaling by the TLR1/2 heterodimer (vide supra) [97–99]. Structure–activity studies with triacylated lipopeptides have revealed that the immune modulatory activity is strongly dependent on the fatty acid chain length and the correct absolute configuration of the natural 2R-dihydroxypropyl-R-cysteine. Also, replacement of the sulfur by CH₂ reduces biological activity significantly [99].

The Gram-positive bacterial membrane is surrounded by a thick cell wall of PGN, so their lipoproteins are typically embedded under this barrier and are not accessible to cellular receptors. Furthermore, Gram-negative bacteria have an asymmetric outer membrane (OM) with a layer of LPS molecules forming the outer

leaflet and exposed on the outer cell surface; lipoproteins are anchored in the inner leaflet of phospholipids and face the periplasm. As a result, lipoproteins in both types of bacteria are not immediately accessible for binding to PRRs. This accessibility problem is likely solved during the initial stages of the innate immune response, when bacterial cells are engulfed by macrophages and undergo phagocytosis with subsequent degradation of the cell wall. During this process, bacterial lipoproteins are released and can interact with TLRs [90].

2.2.1 The human TLR1/2 heterodimer

Lipopeptide ligand is required to induce formation of heterodimers of TLR1 and TLR2. The full-length receptors on the plasma cell membrane form weakly associated homo- and heterotypic multimers without bound ligand. However, lipopeptide binding induces rearrangement to the more stable heterodimer, which then initiates signaling [100, 101].

To facilitate production and crystallization of human TLR1 and TLR2, a large number of recombinant hybrid fusion proteins were designed and screened to identify constructs suitable for crystallization [74, 102]. This approach led to the identification of two fusion proteins, called hT_1V8 and hT_2V9, in which 105 residues from the C-terminus of human TLR1 and 82 residues from the C-terminus of human TLR2, respectively, were replaced with fragments of hagfish variable lymphocyte receptor (VLR). These fusion proteins formed a stable and crystallizable heterodimeric complex with the lipopeptide Pam_3CysSK_4.

The structure of the TLR1/2-Pam_3CSK_4 complex confirmed the horseshoe-like geometry of each TLR chain [74]. Viewed from the side, the overall shape of the complex resembles the letter "m", with the two C-terminal VLR domains adjacent in the middle, but not in direct contact with each other nor with other TLR residues (Fig. 8). The ECDs of TLR1 and TLR2 contain the distinctive three-subdomain organization (vide infra). The N-terminal domain contains the LRR-NT capping motif and LRRs1-4, each with typical 24-residue LRR repeats. The hydrogen-bonding Asn ladder and a spine formed by consecutive Phe residues are conserved in this region. In contrast, the central and C-terminal subdomains have LRR modules with atypical sequences, and their β-sheet conformations deviate from those seen in standard LRRs. The central subdomain also lacks the Asn ladder and the Phe spline.

The TLR1/2 chains make contact with each other through LRRs in the central domain, near the center of the "m", where the binding site for the lipopeptide is also located. From one orientation, the two ester-linked lipid chains project forwards and interact with a pocket on TLR2, whereas the amide-linked lipid chain projects backwards and inserts into a narrow channel in TLR1 (Fig. 9a and b). The protein–protein interaction between the two TLR chains extends over 850 $Å^2$, with a near circular interface and the bound ligand located near the center.
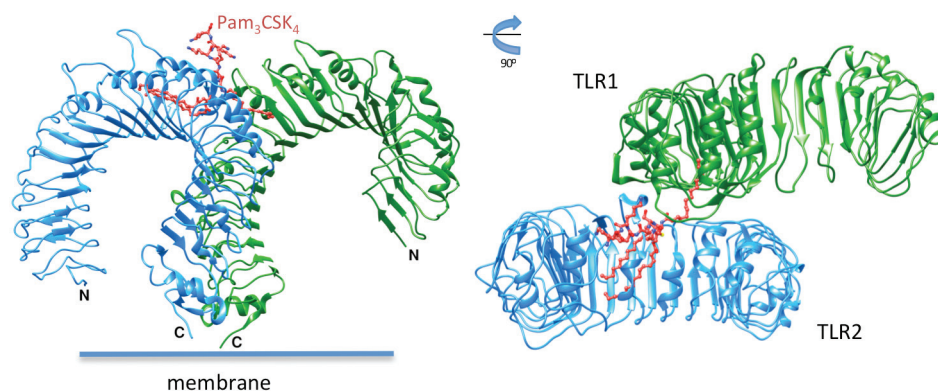

Fig. 8 Structure of the TLR1/2-Pam_3CSK_4 complex; TLR2 (blue), TLR1 (green), lipid (red). View from the side with membrane below (left), and from the top of the complex looking down onto the membrane (not indicated) (right).
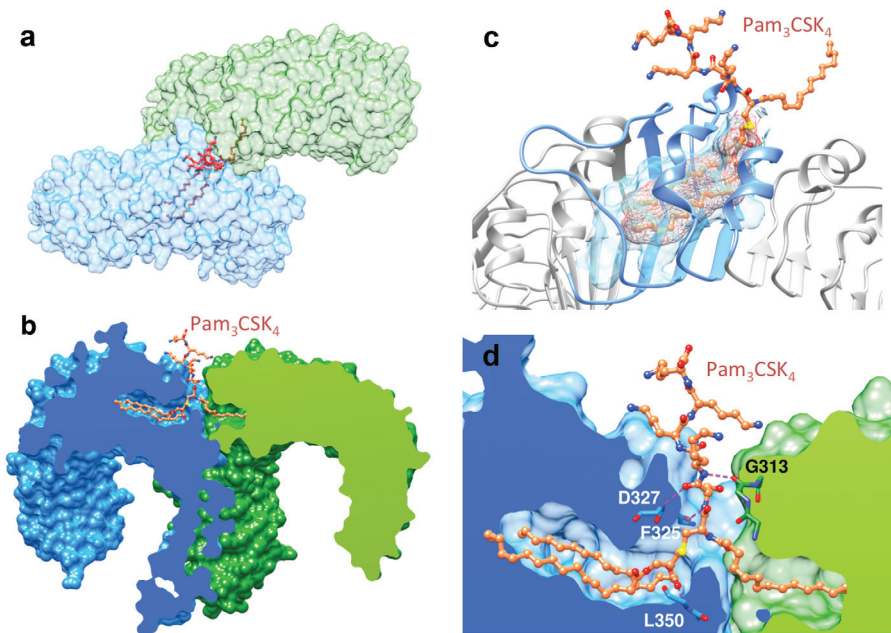
Fig. 9 The TLR1/2_Pam$_3$CSK$_4$ complex. (a) The bound ligand situated near the center of the TLR1/2 protein–protein interface. (b) Slice through the complex. The two ester-linked lipids project into a binding pocket in TLR2 (blue) and the amide-linked lipid into one in TLR1 (green). (c) the lipid-binding cavity (with mesh) and entrance lies near LRR modules 10-12 (light blue). (d) Hydrogen-bonding interactions (pink dotted) with the ligand head-group.

The TLR chains in the heterodimer make contact with each other and the lipopeptide ligand through the sides of each horseshoe. An opening is created in the side of TLR2 by the presence of a large helical loop that projects up and outwards from LRR module 11, thereby allowing the two ester-linked lipid chains to insert into a deeply buried cavity that extends into the region between the β-sheet base and the bridging helical loops of LRR modules 10–12 (Fig. 9c). The two lipid chains occupy over 90 % of the total solvent accessible volume of the pocket. The lack of an Asn ladder and spine of phenylalanine residues within the central LRRs in TLR2 helps to account for the formation of this deep binding pocket. In a similar fashion, an opening between two LRR repeats in the central domain of TLR1 allows the amide-linked lipid chain to insert into a hydrophobic pocket between LRR modules 11 and 12. This pocket is only one-quarter the volume of that in TLR2, but is fully occupied by the amide-linked palmitoyl chain. All three bound lipid chains adopt mostly extended conformations, but gauche backbone conformations also exist at some points along the lipid backbone.

The conserved cysteinyl-glycerol moiety of the ligand exits the TLR1/2 complex through an opening lying between the two TLR chains. A series of hydrogen-bonding interactions between backbone carbonyl oxygen and nitrogen atoms of the TLRs and the ligand, as well as complementary surface contacts, appear to stabilize the complex (Fig. 9d). Small amino acids (Gly/Ser) are preferred in the second position (after Cys) in bacterial lipoproteins. Consistent with this, the second amino acid (Ser) of the synthetic ligand Pam$_3$CSK$_4$ is located in the narrow neck region of the binding pocket, where space for a side chain is limited. The side chains of the four Lys residues make only limited interactions with TLR1 or TLR2, consistent with the low sequence conservation found at these positions in bacterial lipoproteins.

2.2.2 The mouse TLR2 monomer

Crystal structures have also been reported for a mouse TLR2-hagfish VLR fusion protein, in complexes with Pam$_3$CSK$_4$ and Pam$_2$CSK$_4$ [74]. The protein is a monomer in solution and when crystallized with each lipopeptide ligand. The structures of the mouse and human TLR2 ECDs are highly homologous (1.1 Å rmsd over the

backbone atoms), as expected from their high sequence identity (68 %). The two different lipopeptides bind to mouse TLR2 in an almost identical fashion, with the two ester-linked lipid chains inserted into the lipid-binding pocket. The amide-linked peptide chain, however, and the peptide portion of both ligands remain solvent exposed and are not visible in the electron density map. An interesting difference is found in the shape of the lipid-binding pockets of TLR2 between the two species, caused by sequence differences in residues lining the binding pocket. The ligand-binding pocket is slightly shorter in the mouse TLR2 compared to human TLR2. As a result, mTLR2 binds shorter lauroyl chains (C_{12}) more efficiently than hTLR2 [103]. This is reflected in the ability of lauroyl$_3$CSK$_4$ to activate mouse but not human TLR2 [104]. Deletion or more marked truncation of the two ester-bound lipids abolishes TLR2 activation [103].

Additional structures have been reported of mouse TLR2-VLR fusion proteins bound to bacterial lipoteichoic acid from *Streptococcus pneumonia* (pnLTA), and to PE-DTPA, a synthetic derivative of phosphatidylethanolamine linked to a metal-coordinating DTPA group (1,2-dimyristoyl-sn-glycero-3-phospho-ethanolamine-*N*-diethylenetriaminepentaacetic acid) (Fig. 7) [75]. Both pnLTA and PE-DTPA bind to TLR2, but have little or no ability to activate the receptor. LTA typically contains a long chain of ribitol or propanolamine repeating units that are anchored in the cell membrane by linkage to a diacylglycerol head group. pnLTA binds to TLR2, but does not induce heterodimerization and signalling of TLR1/2 or TLR2/6, consistent with the substantially weaker proinflammatory activity of pnLTA compared to bacterial lipoproteins.

In the monomeric TLR2-pnLTA complex, the overall horseshoe-like shape of the ECD and structure of the lipid-binding pocket is unchanged (Fig. 10a and b). However, the LRR10 and LRR11 surface loops and the position of the ligand head group are significantly altered from those in the TLR1/2-Pam$_3$CSK$_4$ and TLR2/6-Pam$_2$CSK$_4$ (discussed below) complexes. This is perhaps not surprising, given that the lipid chains are linked to a peptide chain in Pam$_{2/3}$CSK$_4$ and to a saccharide chain in pnLTA. In the TLR2-pnLTA structure, the position of the sugar head group of LTA is translated by ≈5.2 Å and rotated by ≈110° toward the lateral surface of the horseshoe-like structure (Fig. 10c).

PE-DTPA (Fig. 7) is a synthetic lipid closely related in structure to the abundant phospholipids in cellular membranes. PE-DTPA can bind to TLR2 but cannot induce formation of TLR1/2 or TLR2/6 heterodimers. As in other structures containing lipid-bound TLR2, the acyl chains of PE-DTPA are inserted into the same
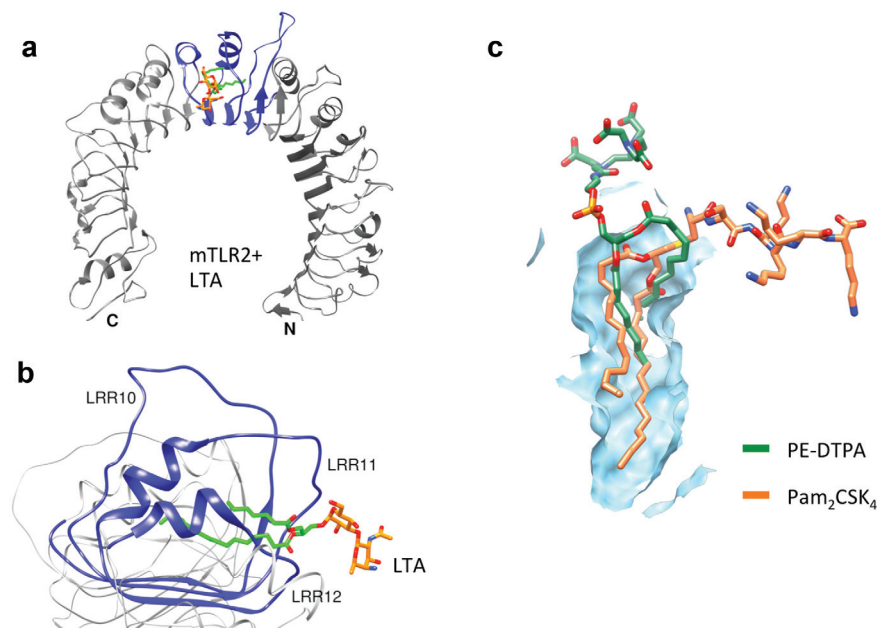

Fig. 10 Structures of the mTLR2-LTA and mTLR2-PE-DTPA complexes. (a) The mTLR2-LTA complex. (b) The LRR10 and 11 loops and position of the LTA ligand head-group are altered from those in the TLR1/2/Pam$_3$CSK$_4$ complex. (c) The head group of bound DTPA (green) resembles that in the complex with LTA rather than that seen in the TLR1/2/Pam$_3$CSK$_4$ complex (orange).

lipid-binding pocket in TLR2. However, the structure of the LRR11 loops and the position of the PE-DTPA head group more closely resemble those seen in the TLR2-pnLTA complex (Fig. 10c). The fact that both pnLTA and PE-DTPA show little or no ability to activate TLR2, and adopt similarly shifted head group binding sites, suggests that the head group structure plays an important role in heterodimerization and hence in TLR2 activation. Bacterial lipopeptides show the strongest TLR2 stimulatory activity [88] most likely because they induce stable heterodimerization of the ECDs of TLR2 with TLR1 or TLR6.

2.2.3 The mouse TLR2/6 heterodimer

To improve protein production and crystallization, C-terminal LRRs in the ECDs of mouse TLR2 and mouse TLR6 were replaced with equivalent regions from the hagfish VLR. Although the engineered TLR2 and TLR6 hybrids each purify as monomeric proteins, they formed a stable heterodimeric complex when incubated with the synthetic lipopeptide Pam_2CSK_4. Binding of the lipopeptide induces the formation of a familiar "m"-shaped heterodimer, with the N termini stretched out to opposite ends and the C-termini converged in the middle (Fig. 11) [75]. Heterodimerization of the ECDs in this way likely enforces juxtaposition and activation of the intracellular TIR domains. Like the TLR1/2 heterodimer, the lipopeptide-binding site in TLR2/6 is located on the sides of each horseshoe at the center of the "m"-shaped complex. As in the TLR1/2-lipopeptide complex, both ester-linked lipid chains are inserted through an opening between LRR11 and LRR12 into a hydrophobic pocket in TLR2 of ≈1200 Å^3 formed by hydrophobic residues from LRR9-12 (Fig. 12a). The ester-bound lipid chains adopt similar structures when bound to TLR2/6 and TLR1/2, but the Me-ends of the chains in the two complexes diverge slightly in different directions at the bottom of the binding pockets, most likely due to sequence differences between mouse and human TLR2s.

Protein–protein interactions between TLR2 and TLR6 are mediated by surface-exposed residues in modules LRR11-14. The LRR11 and LRR12 loops of both TLR2 and TLR6 are located at the center of the interface and contribute key hydrophobic interactions. The surface area of the hydrophobic core of the TLR2/6 interface is increased by ≈80 % compared to that in the TLR1/2 complex (Fig. 11). These hydrophobic contacts are surrounded by a ring of polar hydrogen-bonding and ionic interactions. It seems that the more extensive protein–protein interface in the TLR2/6 heterodimer might help to compensate for the lack of an amide-linked lipid chain, which should help stabilize the TLR1/2-ligand complex.

The structures of the glycerol moiety and the peptide backbones of the ligands in the $hTLR1/2-Pam_3CSK_4$ and $muTLR2/6-Pam_2CSK_4$ complexes can be superimposed almost exactly. However, important structural differences are seen in the positions of surface residues near the LRR11 loop that influence access to the
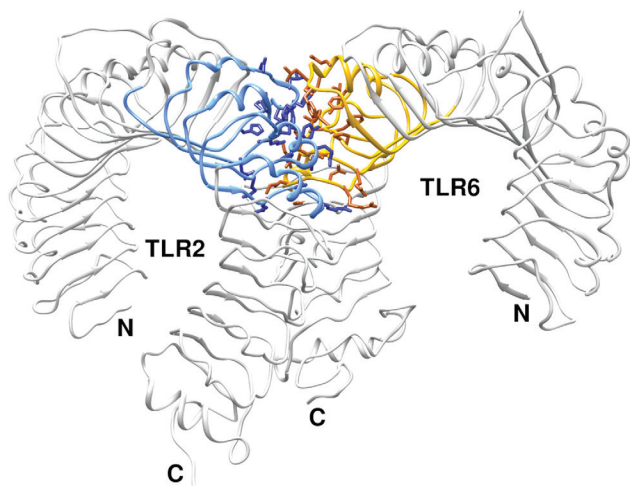

Fig. 11 Structure of the TLR2/6 heterodimer with bound lipopeptide Pam_2CSK_4 (not shown). The hydrophobic core of the protein–protein interface is larger than in the TLR1/2 complex.
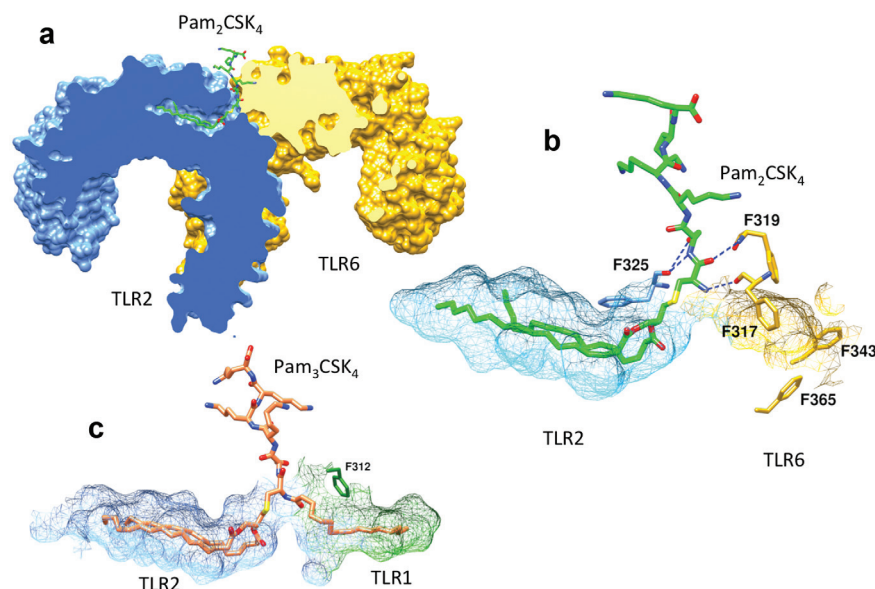
Fig. 12 Structural comparisons of the TLR2/6-Pam$_2$CSK$_4$ and TLR1/2-Pam$_3$CSK$_4$ complexes. (a) In both complexes, the two ester-linked lipid chains in Pam$_2$CSK$_4$ are inserted into a hydrophobic pocket on TLR2. (b) Only a very restricted hydrophobic pocket is present in TLR6. (c) This pocket in TLR1 is larger and contains the amide-linked lipid chain in the ligand.

lipid-binding channels in TLR1 and TLR6, in particular, F317 in TLR6, which corresponds to F312 in TLR1 (Fig. 12b and c). This region on TLR1 forms the opening of the binding site for the amide-linked lipid in Pam$_3$CSK$_4$. In addition, the bulky side chains of F343 and F365 in TLR6, buried in the middle of the pocket, shorten considerably the lipid-binding channel. As a result, the hydrophobic binding pocket in TLR6 is less than half as long as that in TLR1. These changes likely contribute significantly to the markedly reduced affinity of the TLR2/6 heterodimer for triacylated lipoproteins. Thus, typical amide-bound lipid chains are too long to fit into the short TLR6 lipid channel. This conclusion is supported by mutagenesis experiments. For example, an F343M and F365L double mutation of TLR6, which extend the binding pocket, allowed an almost complete response of the TLR2/6 heterodimer to triacylated as well as diacylated lipopeptides [75].

Other differences between the complexes include a repositioning of surface residues in TLR6, which allow a new hydrogen-bonding interaction with the first peptide bond of the ligand, which is not seen in TLR1 (Fig. 12c). On the other hand, many polar and hydrophobic contacts between the polar head group of the ligand and TLR2/6 are conserved, and are likely responsible for the strong stereospecific recognition by both receptor heterodimers of essentially the same head group in both types of lipoprotein/lipopeptide. Using chemically synthesized lipopeptides derived from the lipoprotein MALP2 from *Mycoplasma fermentans*, it was shown that the *R* stereoisomer (S-[2,3-bis(acyloxy)-(2*R*)-propyl-(*R*)-cysteinyl-GNNDESNISFKEK]) exhibits much higher specific activity than the *S*-stereoisomer [105]. The side chains of the first two amino acids (CS) in the ligand also make polar interactions with TLR2/6 (Fig. 12c), and in both cases a small side chain is favored in the second position. The side chains beyond the second residue, however, appear to have higher flexibility in the crystalline state and form only weak interactions with the TLRs. Nevertheless, these side chains are located within ≈0.5 nm of the TLRs and might still be able to interact weakly with the receptors [97, 106].

2.2.4 Applications of TLR1/2 and TLR2/6 agonists

There is now great interest in the discovery of novel small-molecule TLR modulators, because of their potential immune stimulatory activity and value as adjuvants in immunotherapy [107–115]. Several TLR2 ligands

have undergone clinical testing in vaccine formulations [116]. For example, a synthetic lipopeptide related to the naturally occurring MALP2 lipoprotein is currently in clinical trials in combination with the nucleoside analog gemcitabine for the treatment of pancreatic cancer [117, 118]. The lipopeptide should activate DCs to express co-stimulatory molecules and promote type-1 T helper responses. This lipopeptide is diacylated and so interacts with TLR2/6. The structure of the complex is expected to be very similar to that with the synthetic ligand Pam_2CSK_4, since only the peptide side chain after Ser2 is different between the two lipopeptides.

2.2.5 Accessory proteins involved in TLR2 signaling

The mechanism(s) by which TLRs recognize PAMPs on the cell surface appears to involve additional membrane-associated accessory proteins [119]. For example, CD36 was shown by mutagenesis to play an important role in signaling through TLR2 on macrophages by the diacylated lipopeptide MALP2 [120]. CD36 is a member of the scavenger receptor type B family, and has been implicated in the recognition of oxidized LDL particles and the uptake of fatty acids. CD36 also appears to facilitate recognition of diacylated lipoproteins by TLR2/6. CD36 has an established role in the recognition of endogenous ligands and might therefore represent a signaling bridge between endogenous ligands and the PRRs of the innate immune system [121]. No structural data is available so far for CD36, and no direct binding studies with lipopeptides have been reported.

Another mediator of lipoprotein signaling is the GPI-anchored glycoprotein CD14, which is highly expressed on monocytes, macrophages, and neutrophils. CD14 is best known for its role in sensitizing cells to the presence of Gram-negative bacterial LPS, by delivering this molecule to the TRL4 signaling complex (vide infra). Membrane-bound CD14 shuttles LPS to TLR4 complexes, whereas a soluble form of CD14 enables cells that lack endogenous CD14 to respond to LPS [122, 123]. However, CD14 is also capable of binding to a wide variety of natural and synthetic acylated ligands, in addition to LPS. For example, CD14 has been shown to enhance cellular inflammatory responses to a variety of acylated agonists of TLR2, including bacterial lipoproteins. The binding of soluble CD14 to Pam_3CSK_4 has been demonstrated with affinities in the low micromolar range. Moreover, CD14 was shown to enhance signaling by triacylated lipopeptides through the TLR1/2 heterodimer in a whole cell assay [124]. It was suggested that binding of triacylated lipopeptides to CD14 may be the first step in the signaling pathway through TLR1/2. After binding of lipopeptide, the complex with CD14 should interact with TLR1 and TLR2 at the cell surface, with transfer of the lipopeptide to the signaling-competent TRL1/2 heterodimer complex [125]. Recent in vitro studies support roles for CD14 and LPS-binding protein in driving binding of TLR1/2 to lipoproteins [126].

The crystal structures of mouse and human CD14 have been reported, although the latter structure is only to 0.4 nm resolution [127, 128]. Both proteins contain ≈11 tandem LRRs and adopt an overall horseshoe-like shape typical of related LRR-family members. The binding site for LPS on CD14 has been intensively studied by mutagenesis and by epitope mapping of antibodies that block LPS binding [129]. This study highlighted a region near the N-terminus, which contains a large hydrophobic pocket, likely for LPS binding, and possibly also for binding related lipophilic ligands such as triacylated lipopeptides. A co-crystal structure of CD14 bound to one of its lipophilic ligands has not yet been reported, and it is so far unclear structurally how CD14 interacts with TLR1/2.

2.3 Structures of TLR3

Double-stranded RNA (dsRNA) is a molecular pattern associated with viral infection that is specifically recognized by TLR3. dsRNA is produced by most viruses at some point during their replication, although dsRNA of nonviral origin has also been shown to stimulate TLR3 signaling. Other receptors, however, are also capable of responding to viral dsRNA in mice and humans, indicating a redundancy in mechanisms used to recognize this PAMP [130]. Binding of dsRNA to TLR3 leads to stabilization of a homodimeric TLR3-dsRNA complex that signals via TIR domains facing the cytoplasmic side of the endosomal membrane (Fig. 2). TLR3 consists of
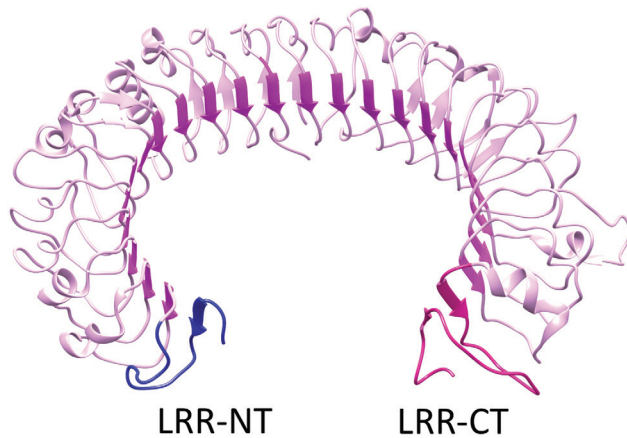
Fig. 13 Structure of TLR3 ectodomain. The LRR-NT and -CT modules, at the N- and C-termini, are highlighted.

an ECD containing multiple LRRs, a single transmembrane domain, and an intracellular TIR domain. TLR3 is distinct from other TLRs in that it does not depend upon MyD88, but rather uses TRIF for signaling [131]. The TRIF-induced intracellular signaling pathway leads to activation of genes for secreted antiviral immunostimulatory cytokines, in particular, the type-I interferons.

The horseshoe-like shape of the TLR3 ECD was first revealed in crystal structures of the monomeric protein [77, 78]. The large curved solenoid has an inner diameter of ≈4.2 nm and an outer diameter of 9 nm (Fig. 13). As with other LRR proteins, the concave inner face consists of a large parallel β-sheet, composed of β-strands from 23 adjacent LRRs. The solenoid has a rather flat surface, devoid of any twist. At its N-terminus, the LRR-NT motif consists of a disulfide cross-linked β-hairpin, whereas the C-terminal capping domain (LRR-CT) forms a compact folded structure stabilized by two disulfide bridges.

dsRNA, but not dsDNA nor single-stranded (ss)-RNA, induces homodimerization of the purified TLR3 ectodomain [79]. The minimum length of RNA needed for binding and homodimerization is ≈45 bp, and the pH optimum for binding is below 6.0, likely reflecting the slightly acidic environment inside endosomes. A crystal structure of the TLR3-dsRNA complex contains the bound RNA in a typical A-DNA-like double helical structure, sandwiched between the two TLR3 ectodomains (Fig. 14a) [80]. The overall structure of the ectodomain in the dimer complex is not changed significantly from that of the monomeric form. The homodimeric complex again has an overall "m" shape, with the N-termini at the outer sides and the C-termini converging at the center of the "m". However, the mechanism of ligand recognition is unique. In the TLR3-dsRNA complex, the straight rod-shaped ligand makes contact with both TLR3 chains at the outer N-termini (LRR-NT and LRR1-3 modules) and at the inner C-termini (LRR19-21 and the LRR-CT modules). The distance between the N-terminal binding sites is ≈14 nm, which corresponds to about 3.5 helical turns of RNA. This explains why the length of the RNA double helix needs to be at least ≈45 bp on each strand. Mostly, the phosphodiester backbone makes critical interactions with both TLR3 chains, whereas the base pairs, and hence the sequence of the RNA, does not (Fig. 14b). This is consistent with the known sequence independence of RNA binding to TLR3. The only place where the two TLR3 monomers make direct contact with each other is through largely polar side chains in the centrally located LRR-CT modules, but the interface surface area is only 574 Å^2.

dsRNA molecules both shorter and longer than ≈45 bp can bind to TLR3, although for shorter dsRNAs the affinity is much reduced [79]. It seems likely that multiple TLR3 dimers can bind to very long dsRNA molecules. However, a different mode of interaction has been proposed to account for binding of short ≈20 nt dsRNA molecules to TLR3, although the models remain to be confirmed by structural analyses [132]. The binding of various synthetic analogues of dsRNA to TLR3 has been studied [133], including so-called polyI:C, which consists of a chain of polyriboinosine (poly I) hybridized with a complementary strand of polyribocytosine (Fig. 15). PolyI:C was initially discovered as a potent inducer of type I interferons (IFNs) [134]. A less-toxic analogue is poly(I:C$_{12}$U), containing a uridine at every 13th monomer position. As an inducer of IFN,
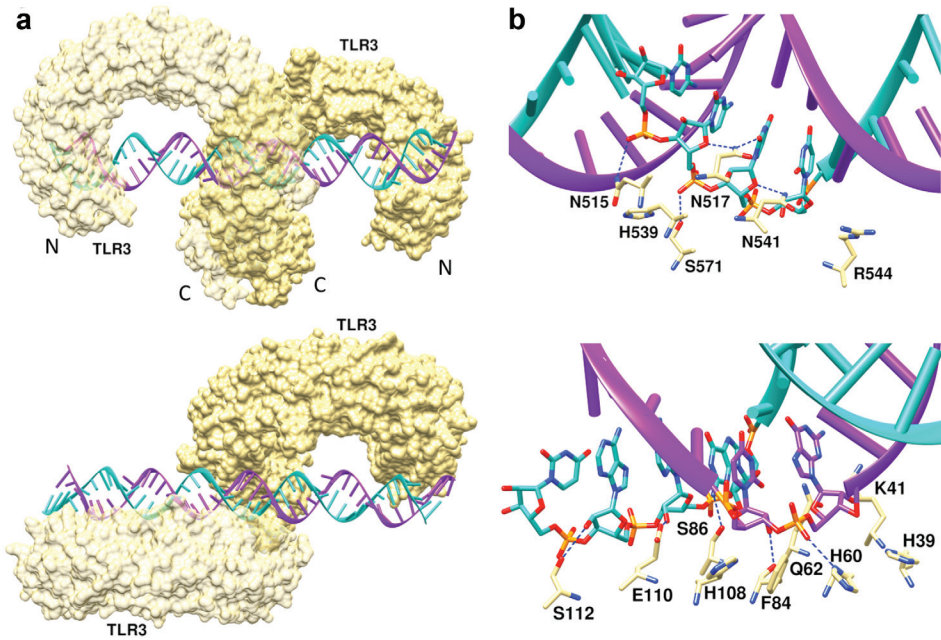
Fig. 14 dsRNA:TLR3 signaling complex. (a) Mouse TLR3 ectodomains form a dimer on the dsRNA. (b) The dsRNA binding sites of TLR3. Residues within contact distance to the dsRNA include, at the C-terminal site, Asn[515], Asn[517], His[539], Asn[541], and Arg[544], which are all well conserved in vertebrates. The N-terminal interaction site is formed by His[39], His[60], Arg[64], Phe[84], Ser[86], His[108], Glu[110].

poly(I:C$_{12}$U) has potent antiviral and immunomodulatory properties, which arise due to its ability to signal through TLR3 [135]. Poly(I:C$_{12}$U), under the names Ampligen and Rintatolimod (Hemispherex Biopharma), has been developed clinically as an experimental immunomodulatory dsRNA drug [118]. A closely related immunostimulatory drug under development is Poly-ICLC, which contains PolyI:C stabilized with poly-lysine (called Hiltonol, Oncovir Inc.) [116].

2.4 Structures of TLR4

TLR4 and the coreceptor MD-2 comprise the physiological receptor for LPS. LPS is a complex glycolipid and the major component of the outer leaflet in the OM of most Gram-negative bacteria (Fig. 16). The structure
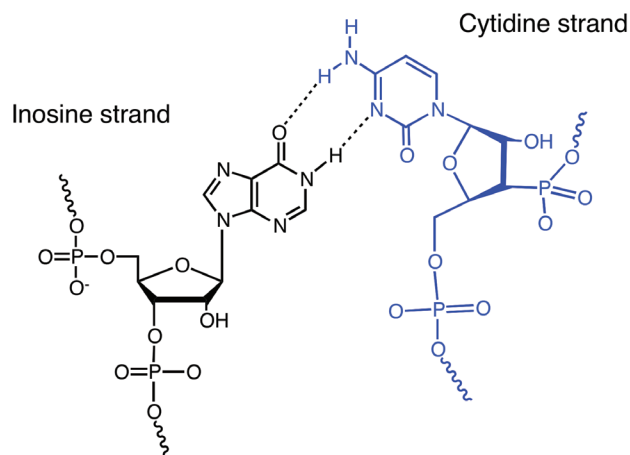

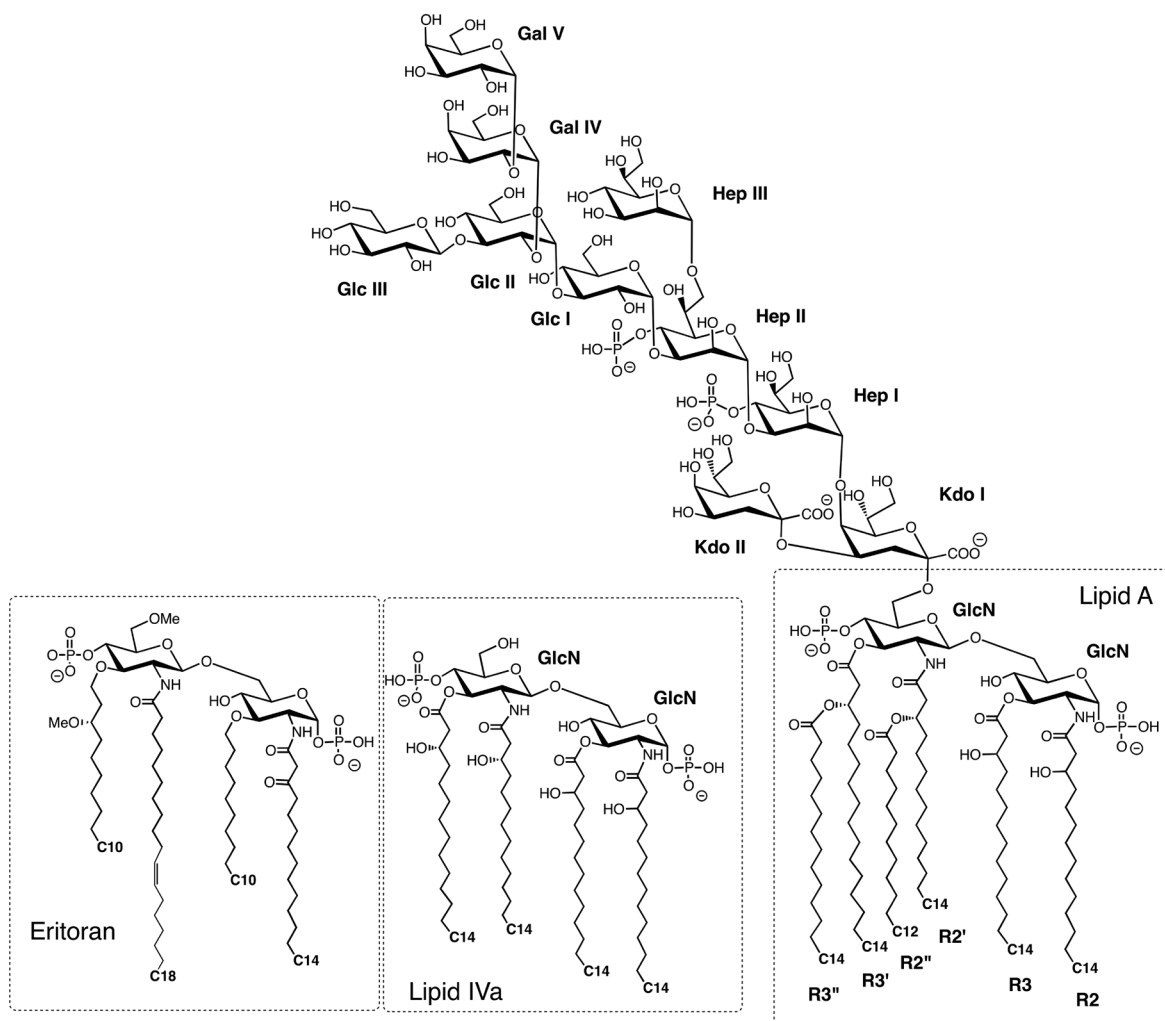**Fig. 15** Structure of a cross-strand I:C base pair.

Fig. 16 Generic structure of LPS and lipid IVa from *E. coli*, and the synthetic antagonist Eritoran. No outer O-antigen polysaccharide is shown. Abbreviations: GlcN, 2-amino-2-deoxy-D-glucose; Kdo, 3-deoxy-D-*manno*-2-octulosonic acid; Hep, L-*glycero*-D-*manno*-heptose; Glc, D-glucose; Gal, D-galactose.

of LPS is highly variable, not only between different Gram-negative bacteria and different strains of the same bacteria, but can also vary within one strain depending upon growth conditions [136]; for example, depleting metal ions in the media, or growth in the presence of cationic host defense peptides (products of the innate immune system) can lead to rapid modification of LPS [137]. However, a common "lipid A" moiety can be recognized in all LPS molecules, which contains a bis-glucosamine disaccharide to which are typically attached 5–6 ester- and amide-linked lipid chains, as well as two phosphate groups, and sometimes other polar groups, such as phosphatidyl-ethanolamine (PE) and 4-amino-4-deoxy-L-arabinose (Ara4N) (Fig. 16). Lipid A is not found naturally, but is released from LPS by mild acid hydrolysis. The lipid A moiety is linked to a carbohydrate core domain, typically comprising ≈6–8 sugar residues, including Kdo (3-deoxy-D-manno-oct-2-ulosonic acid) and Hep (L-glycero-D-mannoheptose), which are not generally found on mammalian cells. The core domain is then linked either to a long oligo-saccharide chain or to a highly immunogenic O-antigen polysaccharide, composed of repeat units that can be highly variable between different serotypes. The O-antigen is the most exposed segment of LPS on the bacterial cell surface. Overall, the LPS layer in the OM constitutes a major permeability barrier, which severely restricts diffusion of substances (e.g., antibiotics) across the OM and into the bacterial periplasm and cytoplasm.

LPS can stimulate a powerful immune response by signaling through TLR4. If the response is overwhelming and uncontrolled, this immune protection mechanism can lead to fatal septic shock. LPS is extracted from the bacterial membrane and transferred to TLR4 by two accessory proteins, the LPS-binding protein (LBP) and CD14 [119]. Membrane-bound CD14 shuttles LPS to TLR4/MD-2 complexes at the cell surface, whereas a soluble form of CD14 enables cells that lack endogenous CD14 to also respond to LPS [122, 123]. Early work showed that the lipid A region is responsible for most of the immunostimulatory activity of LPS. The number of lipid chains has a large influence on inflammatory activity. Lipid A with six lipid chains has optimal inflammatory activity, whereas with five chains the lipid A is ≈100-fold less active. With only four lipid chains, as found in lipids IVa and Eritoran (vide supra), the glycolipid typically has antagonist activity [138]. The mouse TLR4 complex appears to be more promiscuous in recognizing different forms of lipid A than human TLR4, and seems even to interact with the anticancer drug taxol [139]. However, minor changes in LPS structure can have large effects on immunostimulatory activity and even abolish endotoxic effects [140, 141].

2.4.1 The human MD-2/lipid IVa complex

MD-2 is a ≈18 kDa secreted glycoprotein, which alone, as well as in the complex with TLR4, binds LPS with nanomolar affinity [85]. The number and length of the acyl chains have an important influence on signaling through the TLR4/MD-2 complex. For example, *Escherichia coli* lipid A is usually hexa-acylated and acts as a potent activator (agonist) of all mammalian cells. In contrast, tetra-acylated lipid A such as lipid IVa (Fig. 16), the biosynthetic precursor of *E. coli* LPS, acts as an agonist on mouse and as an antagonist (with antiendotoxic) activity on human cells [142]. Lipid IVa does not induce formation of the human TLR4/MD-2 heterodimer complex, as required for signaling activity.

The first structural insight into how lipid A is recognized by MD-2 came from crystal structures of human MD-2 bound to lipid IVa [143]. MD-2 comprises a single immunoglobulin domain stabilized by three disulfide bonds. The protein consists of two β-sheets, one of three antiparallel β-strands, and the other of six antiparallel β-strands (Fig. 17a). Between the two sheets is a large, hydrophobic cavity with a volume of ca. 1710 Å³. The overall structure is rather like a partly open clam, with the two shells represented by the two β-sheets. The four lipid chains of the ligand insert through the opening into the main body of the clam. In a crystal structure of MD-2 without bound lipid IVa, the MD-2 structure is not altered significantly; the lipid A binding pocket is still present, but is occupied by three myristic acid molecules that appear to have been co-purified and co-crystallized with MD-2. The lipid chains at positions 2 and 3 in lipid IVa are mostly fully extended and reach far into the hydrophobic binding pocket of MD-2 (Fig. 17b). The 2′ and 3′ lipid chains have a distinctly curved conformation. The fatty acid chains in the cavity are packed next to each other and interact through van der Waals contacts. The size of the hydrophobic binding cavity is not large enough to accommodate the two additional lipid chains present in lipid A (and LPS). Rather, one of the additional lipid chains is found on the surface of MD-2 where it forms part of the contact interface with TLR4 in the TLR4/MD-2/LPS complex (vide supra).
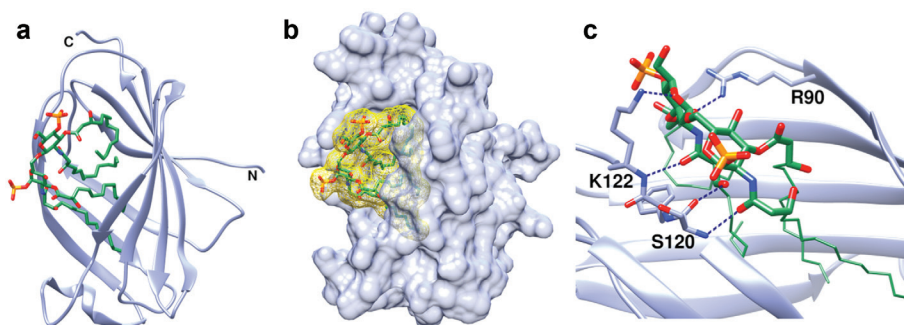

Fig. 17 Structure of MD-2 with lipid IVa. (a) The four lipid chains of lipid IVa are buried between the two β-sheets in MD-2. (b) Surface representation of the complex. (c) Interactions between MD-2 and the head group [sugars (green/red) and phosphates (yellow/red)] of bound lipid IVa.

The 1- and 4′-phosphate groups and the two sugar residues in lipid IVa are at the solvent-exposed surface, in the opening of the clam-like structure. Here, hydrogen-bonding interactions occur between amide and ester carbonyls of the ligand and the edge of the β7 strand in the three-stranded β-sheet (Fig. 17c). Residues from Phe[119]-Gly[123] are important for LPS recognition, and these residues with the exception of Lys[122] are highly conserved in MD-2. Overall, the entrance to the hydrophobic cavity has a positive electrostatic surface, which likely aids binding of the negatively charged ligand. However, neither of the phosphate groups is involved in close electrostatic contacts with atoms of MD-2.

2.4.2 The human TLR4/MD-2 with bound Eritoran

Eritoran (or E5564) is a synthetic molecule related to the lipid A of the nonpathogenic LPS of *Rhodobacter sphaeroides*, with antagonist activity on human TLR4/MD-2. Eritoran contains a bis-glucosamine backbone linked via amide or ether bonds to four lipid chains, one containing a *cis*-double bond, as well as two phosphate groups (Fig. 16). However, neither Eritoran nor lipid IVa induces formation of the TLR4/MD-2 dimeric signaling complex (vide supra). The human MD-2 was crystallized with a hybrid hTLR4-hagfish-VLR fusion protein as a stable 1:1 complex, with Eritoran bound to MD-2 (Fig. 18) [102]. All four lipid chains in Eritoran are buried inside MD-2 and so are not available for interactions with TLR4. The glucosamine backbone in lipid IVa and Eritoran are superimposable when bound to MD-2. The shape of the lipid-binding pocket on MD-2 is essentially unchanged in the complexes with the two antagonists. Eritoran was in clinical development for use as a therapeutic against severe sepsis, however, the compound was recently withdrawn following inconclusive phase III trials [111, 118].

2.4.3 The human TLR4/MD-2/LPS complex

LPS is extracted from the bacterial membrane and transferred to TLR4/MD-2 by LPS-binding protein and CD14. The crystal structure of the TLR4/MD-2/LPS complex has been very valuable in understanding ligand binding specificity and the mechanism of receptor activation. Interestingly, the signaling complex seen in the crystal structure is composed of two copies of TLR4/MD-2/LPS arranged in a symmetrical fashion (Fig. 19) [144].
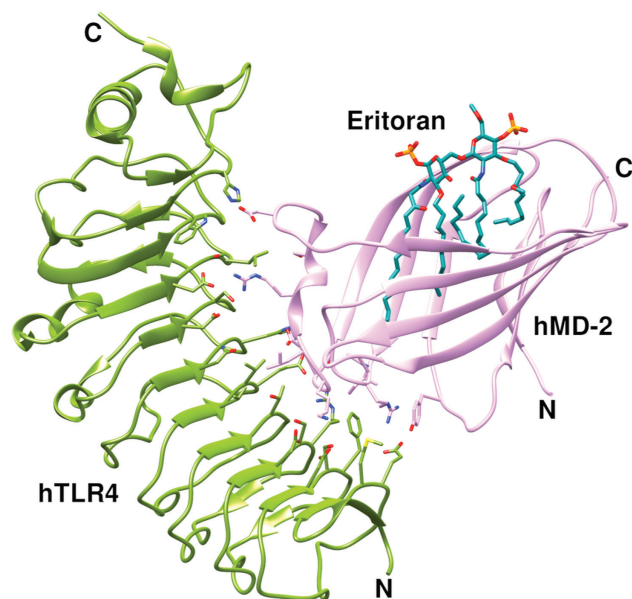

Fig. 18 Structure of the 1:1 TLR4/MD-2 complex with bound Eritoran (see text).
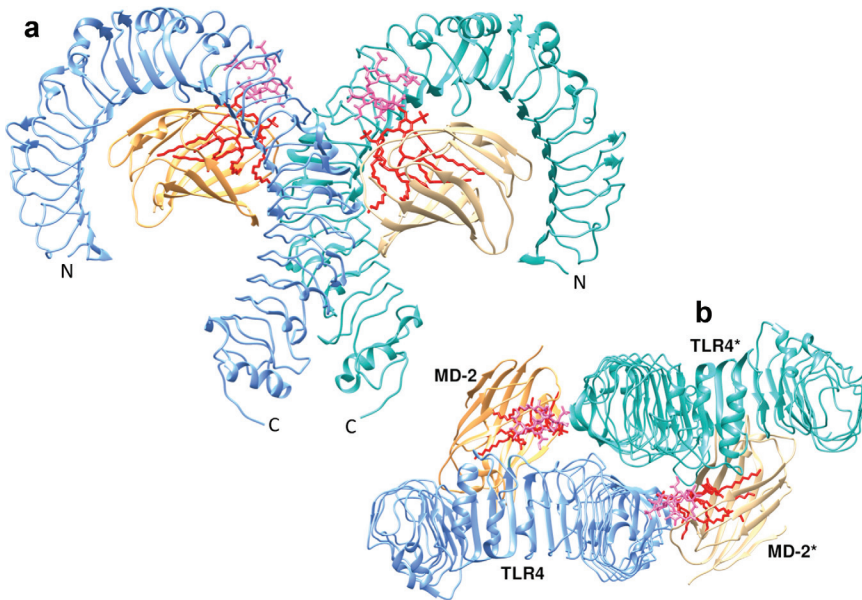
Fig. 19 Structure of the dimeric TLR4/MD-2/LPS signaling complex. TLR4 chains, blue; MD-2 chains, orange; lipid A, red; and outer sugar/phosphates, pink (see Section 2.4). (a) viewed from the side (membrane at the bottom), and (b) viewed from the top.

The two TLR4 chains each adopt a horseshoe-like shape, as seen with other TLRs, resembling in the complex the letter "m", with both C-termini in the center and the N-termini located at each side of the "m" (Fig. 19). The two TLR4 chains contact each other in the center through LRR modules 15–17. MD-2 again has a clam-like shape composed of two antiparallel β-sheets, between which the hydrophobic lipid chains of LPS can bind. The two MD-2 chains are situated close to the center of each arm of the "m". Each MD-2 makes contact with the side of one TLR4 chain and the exposed convex face of the other TLR4 chain. The numbering system used for the lipid chains (R2, R3, R2′, R3′, R2″, and R3″) of the lipid A moiety is shown in Fig. 16.

Five of the six lipid chains of the lipid A are completely buried inside the hydrophobic pocket on MD-2, between the two β-sheets. However, the R2 chain is on the surface of MD-2 (Fig. 20a). This composite
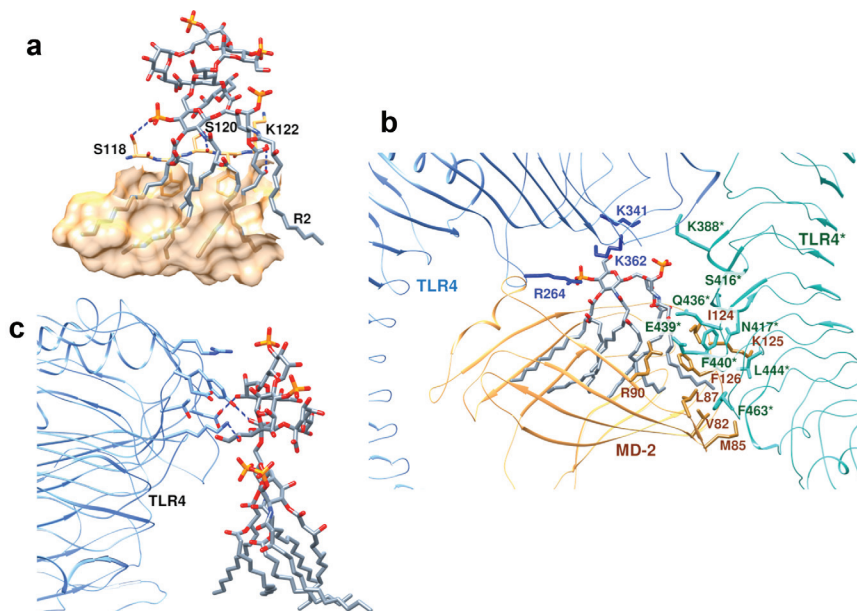

Fig. 20 Structure of the dimeric TLR4/MD-2/LPS signaling complex. (a) Five lipid chains of LPS are buried and one (R2) is on the surface of MD-2. (b) The TLR4/MD-2 interface. (c) LPS sugar–TLR4 interactions.

(protein + lipid) surface forms a largely hydrophobic interface with one TLR4 chain (TLR4[1]). Hence this lipid chain, and its key position, are likely important for the stability of the complex. Some of the ester and amide groups connecting the lipid chains to the sugar residues of lipid A interact at the surface of MD-2 with one exposed β-strand in the 3-stranded β-sheet. The two phosphate groups of lipid A do not appear to occupy specific binding pockets, but sit on the surface, close to a positively charged patch formed by Lys residues on TLR4[2] and MD-2, with one phosphate close to an Arg side chain contributed by the TLR4 chain. An extended loop protruding from LRR module 9 on the concave face of TLR4[2] forms part of an extensive protein–protein interface (PPI) with MD-2 (Fig. 20b). Finally, another extensive PPI is found between the C-terminal regions of both TLR4 chains at the center of the "m" formed by LRR modules 15–17.

Apart from the lipid A moiety, five additional sugar residues are seen in the complex, interacting in particular with a Tyr and several other side chain from TLR4[2] (Fig. 20c). However, the Kdo III sugar of the inner core of LPS and all carbohydrates in the outer core are not clearly visible in the electron density map, presumably due to disorder resulting from a lack of fixed interactions in the crystal lattice.

Of special interest is a comparison of the TLR4/MD-2/LPS signaling complex with structures of MD-2 bound to the two antagonists, Eritoran and lipid IVa, which reveal similarities as well as some key differences. The structures of MD-2 bound to the two antagonists reveal a large hydrophobic pocket completely filled by the four lipid chains of each antagonist. The size of the pocket is unchanged in the TLR4/MD-2/LPS complex, although the LPS now has six lipid chains. In order to create additional space for lipid binding, the bis-glucosamine backbone is displaced upwards by ≈0.5 nm, which repositions the phosphate groups such that they can interact with positively charged residues of both TLR4 chains (Fig. 21). This appears to be important for dimerization and activation of the complex. However, the bis-glucosamine backbones of the two antagonists are not only translated, but also rotated by ≈180°, thereby interchanging the positions of the two phosphate groups. The two phosphate groups in the lipid A greatly influence the endotoxic activity of LPS [145]. Deletion of either phosphate group reduces endotoxic activity by ≈100 fold. The resulting monophosphoryl lipid A (MPL) is only a weak activator of the innate immune response. An MPL based upon lipid A from *Salmonella minnesota* was shown to selectively activate the less inflammatory TLR4-TRAM-TRIF intracellular signaling pathway rather than the TLR4-Mal-MyD88 pathway [146]. Hence, lipid A modifications can influence innate immune responses by mediating the recruitment of distinct sets of adaptor proteins in the cytoplasm. These results suggest that the MPL ligand has not only a lower affinity, but it also induces a structural rearrangement of the TLR4-MD-2 receptor complex. The effects are likely mediated by the observed interactions of the 1- and 4′-phosphate groups with a cluster of positively charged side chains from TLR4[1], TLR4[2], and MD-2. This likely explains why deletion of one phosphate can have a major impact on the LPS–receptor interaction. It is also interesting to note that many bacteria have evolved enzymes that alter their lipid A structure, not only by modifying the number of acyl chains but also by attaching other polar/charged groups to the phosphates (such as phosphoethanolamine and 4-amino-4-deoxy-L-arabinose). Such modifications have
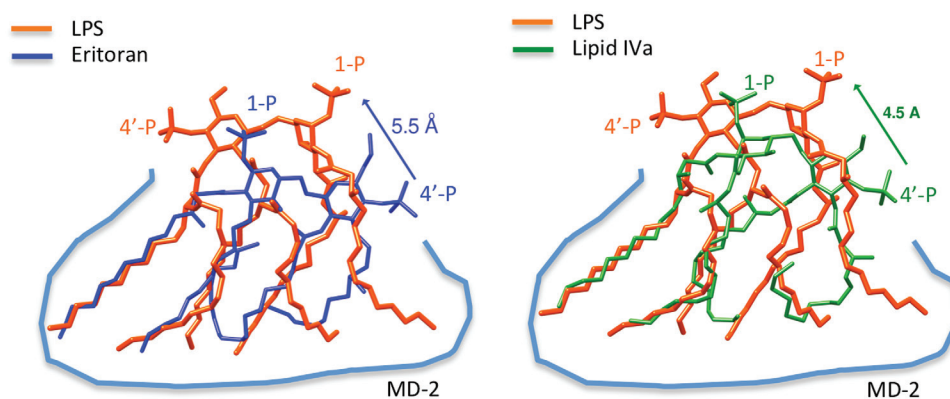

Fig. 21 Structural comparison of LPS and antagonists bound to MD-2. The MD-2 chains are superimposed. The relative positions of LPS and Eritoran (left), and LPS and lipid-IVa (right) are shown.

a major influence on the host response elicited by bacteria and may be an important factor in determining the outcome of infectious disease [139].

Efforts to understand the lipid A–TLR4 interaction have already yielded improved vaccines. Whereas strong induction of the MyD88 pathway is harmful because of high production of proinflammatory cytokines, such as TNF-α and IL-1β, a low level of induction is beneficial for long-lasting immunity elicited by vaccines. One MPL derivative has recently been approved as an adjuvant in vaccines used worldwide [147]. Furthermore, the development of a wide variety of different immunostimulatory lipid A derivatives for vaccine and other therapeutic applications is continuing to attract great interest [107, 110, 114, 118].

2.4.4 The mouse TLR4/MD-2/-LPS and lipid IVa complexes

It was mentioned above that tetra-acylated lipid IVa shows weak agonist activity on mouse but antagonist activity on human cells. As discussed above, the human TLR4/MD-2/lipid IVa complex is a monomeric unit in the complex, whereas the signaling-competent hTLR4/MD-2/LPs complex is dimeric in the crystal. Recently, the crystal structures of mouse TLR4/MD-2 with bound hexaacylated LPS (from *E. coli*) and with lipid IVa were reported [148], which allow interesting comparisons to be made with the human receptor.

The mTLR4/MD-2/LPS complex also appears as a dimeric unit in solution and in the crystalline state, confirming that LPS binding induces dimerization of the TLR4/MD-2 complex for immune signaling (Fig. 22a). In contrast, the mTLR4/MD-2/lipid IVa complex appeared monomeric in solution, but in crystals a dimeric form of the mTLR4/MD-2/lipid IVa complex was observed (Fig. 22b). The appearance of the dimer when the mouse receptor is bound to lipid IVa correlates with the weak agonist activity seen with this ligand in mice. It seems that the dimerization interaction of two 1:1:1 mTLR4/MD-2/lipid IVa complexes is weak in solution, but is favored upon crystallization. Conceivably, dimerization might also be enhanced at the cell surface, compared to free solution, as a result of membrane anchoring, and additional interactions between the transmembrane and/or cytoplasmic domains of TLR4.

Of special interest in the dimeric mTLR4/MD-2/lipid IVa complex is the orientation of the phospho-glucosamine backbone of lipid IVa and the location of the lipid chains (Figs. 22b and c). Lipid IVa has the same
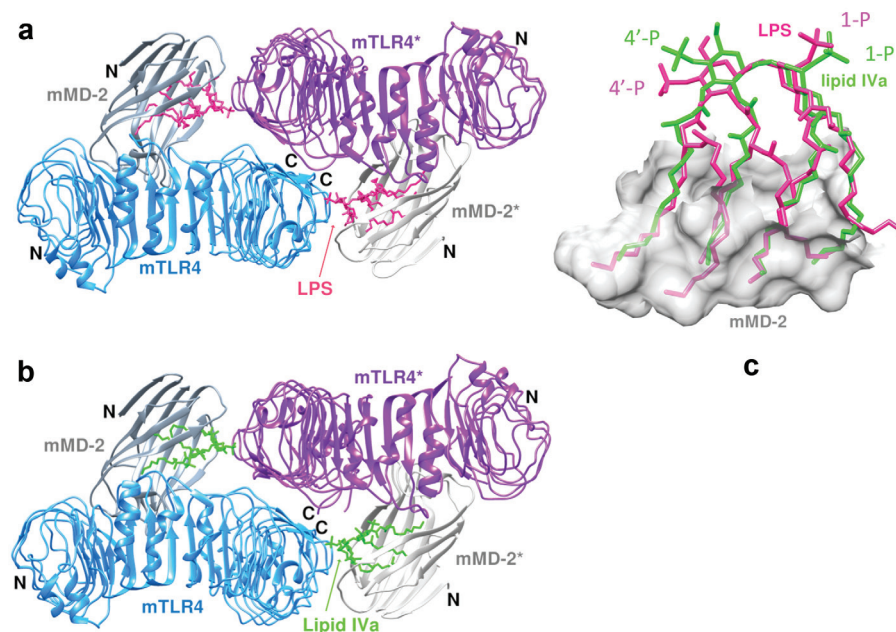

Fig. 22 Structural comparison of mTLR4/MD-2/LPS with mTLR4/MD-2/lipid IVa complexes. (a) Structure of mTLR4/MD-2/LPS. (b) Structure of mTLR4/MD-2/lipid IVa. (c) Orientations of lipid A (in LPS) and lipid IVa bound to mMD-2 in the two complexes.

orientation in the mouse complex as that seen for lipid A in the complex with LPS, and not that seen in the human MD-2/lipid IVa complex (Figs. 21 and 22c). Moreover, in the complex with lipid IVa bound to hMD-2, all four lipid chains are buried in the internal lipid-binding cavity inside MD-2. In the complex with mTLR4/MD-2, however, one of the acyl chains (R2) is displaced towards the surface of MD-2, leaving part of the hydrophobic lipid-binding pocket in MD-2 unoccupied. The R2 chain can then form part of the hydrophobic surface of MD-2 that interacts with a hydrophobic patch on TLR4, like that seen in the complex with LPS (Fig. 22c). It seems that the different hydrophobicities of the dimerization interfaces, and the different charge distributions of residues on mouse and human TLR4/MD-2 have a decisive influence upon the position and orientation of bound lipid IVa, inducing the agonist orientation with mouse TLR4/MD-2 and antagonist orientation with human TLR4/MD-2 [148].

2.5 Structure of TLR5

Gram-positive and -negative bacteria swim by rotating helical flagella filaments up to 15 µm long extending from their external surface. The rotary motor at the base of the filament is embedded in the membrane (or IM) and drives rotation. The filament is mostly a helical assembly of a single protein, flagellin (FliC), which forms a tubular structure comprising 11 strands of protofilaments. The protofilaments contain nearly longitudinal helical arrays of flagellin. Flagellin is, therefore, one of the most abundant proteins on the surface of many bacteria.

The crystal structure of a 41 kDa *Salmonella* flagellin fragment (residues 53–450 of the 494-residue FliC) that does not form filaments, revealed not only the structure of the monomer, but also the likely packing of monomers into pairs of antiparallel straight supercoiled protofilaments in the crystalline state [149]. The FliC structure has an overall boomerang-like shape, which can be divided into three domains (Fig. 23). One wing of the boomerang is the D1 domain, which comprises three α-helices and an extended β-hairpin, where two helices (αND1a, αND1b) and the hairpin are contributed by the N-terminus and the third helix (αCD1) by the extreme C-terminus of the protein. Domain D2 at the center of the boomerang also comprises two segments (Lys177-Gly189 and Ala284-Ala401) and is largely composed of β-structure. A central fragment from Tyr190-Val283 makes up domain D3 and the other wing of the boomerang. Since neither vertebrates nor plants contain flagella, the presence of flagellin serves as a marker of bacterial infection in both classes of organism. TLR5 is the major pattern recognition receptor used to detect flagellin, and is the only protein-binding TLR that is conserved in organisms as diverse as fish, flies, plants, and mammals [150]. TLR5 signals through the adaptor protein MyD88. Activation of the receptor mobilizes the proinflammatory nuclear transcription factor NF-κB in epithelial cells, monocytes, and dendritic cells, and stimulates TNF-α production.

Recently, the crystal structure of an engineered zebrafish TLR5 ectodomain in a complex with a truncated fragment of *Salmonella* FliC was reported [83]. The chimeric receptor contains 14 N-terminal LRR modules
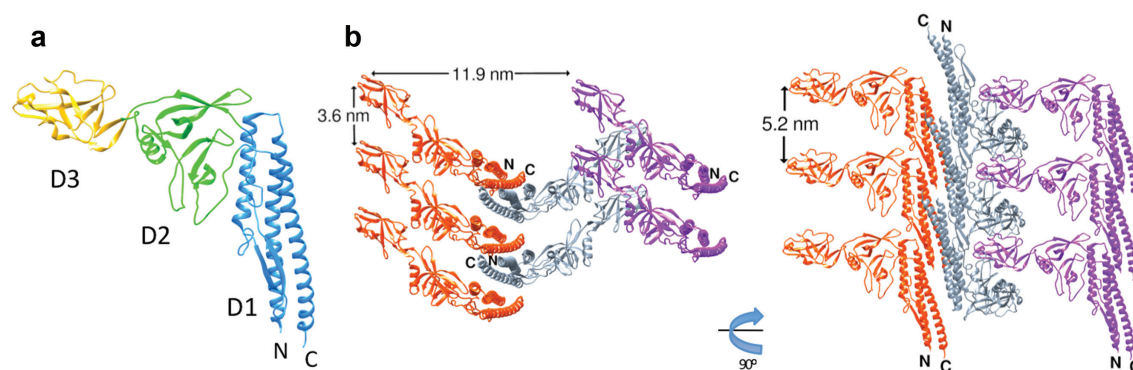

Fig. 23 Structure of flagellin (FliC) fragment (residues 53–450 of the 494 residue FliC). (a) Domain D1 (blue), domain D2 (green), domain D3 (yellow). (b) Crystal packing of the FliC fragment in two orthogonal views. The repeat distances are labeled.

fused to a C-terminal fragment of hagfish VLR. In the TLR5-FliC complex, the FliC D1 domain is located on the lateral side of one TLR5 chain forming an extensive "primary binding interface" that defines a 1:1 heterodimer (Figs. 24a and b). Two such heterodimers further oligomerize to a symmetric 2:2 complex, where TLR5 from the first heterodimer makes additional, less extensive interactions with both FliC′ and TLR5′ from the second heterodimer, so forming a "secondary dimerization interface" (Figs. 24a and b). Thus, flagellin binding results in the assembly of two TLR5 chains into a tail-to-tail organization that juxtaposes the C-terminal regions in the center, as seen in the "m"-like structures formed by other TLR-ligand complexes. However, the ligand-receptor stoichiometry and ligand-binding arrangement observed is unique to the TLR5 signaling complex. The FliC–TLR5 interface also coincides with residues involved in FliC oligomerization in the flagellar filament structure, and is therefore well conserved across bacterial species.

The chimeric TLR5 ectodomain consists of an N-terminal β-hairpin-capping motif (LRR-NT), 13 complete LRR modules, and two residues from LRR14. Only domain D1 of flagellin is involved in direct contacts with TLR5. The primary interface is formed between one side of the three long helices of FliC in domain D1 (αND1a, αND1b, and αCD1) and a lateral surface of TLR5 comprising LRRNT to LRR10, with 1320 Å^2 of buried accessible surface area on each side (Fig. 24d). This interface can be described as two adjacent, spatially separated surfaces, denoted sub-interfaces A and B (and A′ and B′) (Fig. 24b). The interface A/A′ has 530 Å^2 buried surface area provided by LRRNT to LRR6, with mainly hydrophilic interactions, with five H bonds and three salt bridges. The sub-interface B/B′ is provided by LRR7-10 on TLR5 and the upper part of αND1a and αND1b on FliC. This interface is also mainly hydrophilic and is largely located in a protruding loop in LRR9 that undergoes structural rearrangement upon FliC binding. An LRR9-loop deletion mutant of TLR5 is severely impaired in FliC binding [83]. Overall, residues in domain D1 make substantial contributions to
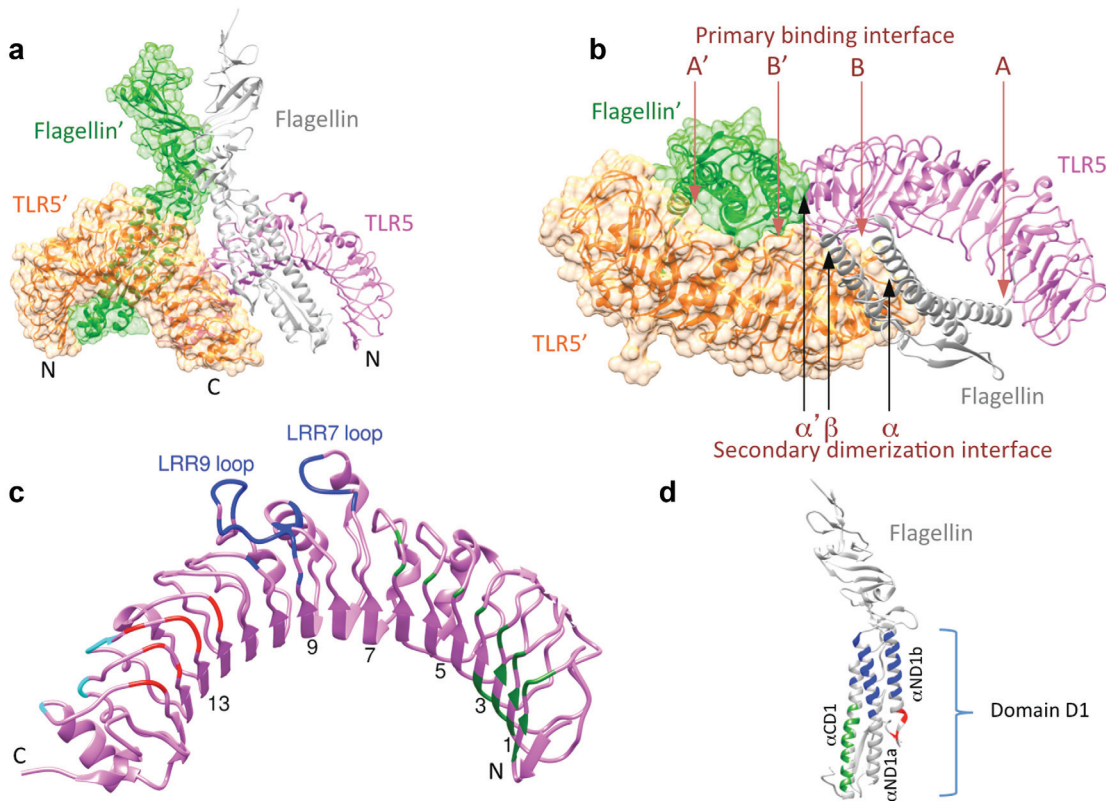
Fig. 24 Structure of the engineered 2:2 TLR5/FliC complex. (a) Two TLR5 chains (pink and orange) oriented tail-to-tail interact with two FliC (green and gray). (b) The 1:1 complex is mediated by the primary binding interface, and its homodimerization to the 2:2 complex by the secondary dimerization interface. (c) The TLR5 structure in the complex with LRR modules numbered. (d) The FliC structure with labeled helices involved in the binding interfaces.

both high affinity binding and TLR5 signaling. The conserved D1 domain is exposed in monomeric flagellin but is buried in the polymerized flagellin fibre, suggesting that TLR5 recognizes flagellin monomers that are released upon depolymerization of flagellin fibres.

A secondary dimerization interface contains interface α (TLR5-FliC), its 2-fold symmetry-related interface α′ (TLR5-FliC′), and interface β (TLR5-TLR5′) (Figs. 24b and c). No direct contacts are seen between the two FliC molecules in the complex. Dimerization interface α is located on the convex side of TLR5′ LRR12/13 and on the FliC C-terminal αND1b region and following loop at the bottom of the D1 domain. Dimerization interface β, with 550 Å^2 buried surface area, is formed mainly by the two TLR5 chains, on the ascending lateral side of LRR12/13.

There has been great interest in exploiting the TLR–flagellin interaction to design new vaccine adjuvants and antagonistic therapeutics for hyperinflammatory diseases. For example, the protein drug CBLB502 is a pharmacologically optimized derivative of *Salmonella* flagellin, designed for use as a radioprotectant in cancer radiotherapy by acting as a TLR5 agonist with reduced immunogenicity and toxicity [151]. This drug contains the complete domains D0 and D1 formed by N- and C-terminal regions of flagellin (1–176 and 402–505), linked by a flexible 16-residue Gly-rich linker. This construct includes the three long helices (αND1a, αND1b, and αCD1) and β-hairpin present in domain D1. The construct can be produced in *E. coli* and retains the NF-κB-inducing activity and stability of flagellin, but is substantially less immunogenic. In another example, flagellin has been fused genetically to various protein antigens to generate chimeric fusion proteins that are significantly more immunogenic than achieved simply by mixing the antigen with flagellin [152–154]. This approach exploits the adjuvant activity expected from a TLR5 agonist, including induction of DC maturation, and up-regulation of cytokines, co-stimulatory signals and antigen-presenting molecules [155].

2.6 TLRs 7, 8, and 9

TLRs 7, 8, and 9 are involved in the recognition of viral and bacterial nucleic acids, and like TLR3 are located in endosomal membranes with the ectodomains facing inside the endosome. This localization is important in directing responses to pathogens while avoiding activation by self-nucleic acids in the cytoplasm. Indeed, recognition of self-nucleic acids by these TLRs can contribute to the pathology of autoimmune diseases, such as rheumatoid arthritis and systemic lupus erythematosus [156]. Perhaps most striking is the complex pathway by which these TLRs are transported from the endoplasmic reticulum to endolysosomal compartments, a pathway that includes an essential processing step by endosomal proteases, involving cleavage of the ectodomain within the so-called Z-loop [157–159].

The amino acid sequences of TLRs 7, 8, and 9 suggest that they form a subfamily, with structural features different from the TLRs discussed above. Each ectodomain contains >800 residues comprising about 25 LRRs, with large insertions in LRRs 2, 5, 8, 11, 18, and 20 after each consensus β-strand. These insertions form loop structures that are predicted to protrude onto the same lateral face of the receptor [70, 160]. In addition, each ectodomain contains an insertion of ≈30 residues between LRR14 and 15, called the Z-loop. The ectodomains of these TLRs are cleaved within the Z-loop by an endosomal protease with the result that no full-length protein is usually detectable in the endosome compartment where ligand recognition occurs [157–159].

Both TLR7 and TLR8 recognize uridine- and guanosine-rich single-stranded RNA (ssRNA) from various viruses, as well as small interfering RNAs [161, 162]. However, TLR7 and 8 also mediate the recognition of self-RNA that is released from dead and dying cells, and so contribute to the pathogenesis of autoimmune diseases. Several small-molecule activators of TLR7 and TLR8 have also been discovered [163, 164]. For example, imidazoquinoline derivatives, such as resiquimod (R848), are recognized by human TLR7 and TLR8 [164], whereas guanosine nucleotide analogs such as loxoribine and other imidazoquinolines, such as imiquimod (R837), activate TLR7 (but not TLR8) (Fig. 25) [163, 165, 166]. These imidazoquinolones are some of the smallest TLR ligands known, and are attractive therapeutic agents for various infectious diseases. By activating
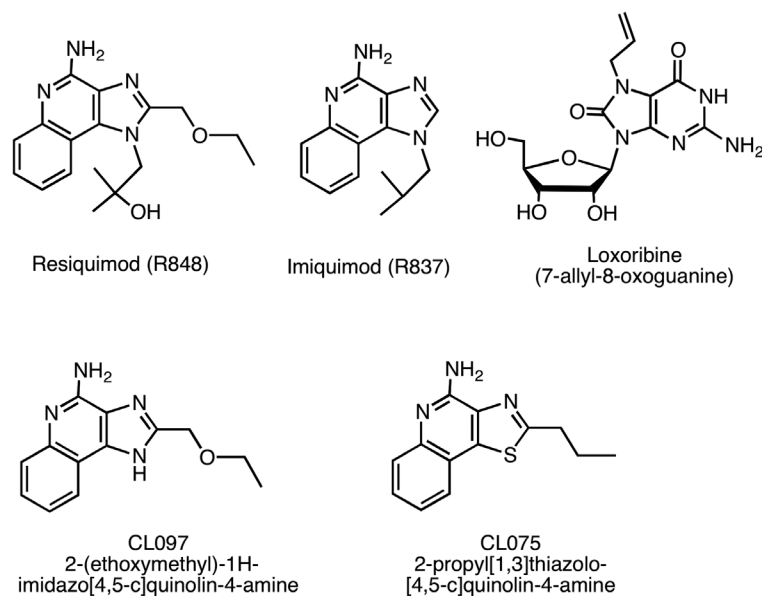
Fig. 25 Structures of synthetic ligands for TLR7 and TLR8.

immune cells they induce the synthesis of IFN-α and other cytokines in a variety of cell types. As TLR activators, they have potent antiviral and antitumor properties [163]. Imiqimod (R837) has been approved for treatment of external genital warts caused by human papillomavirus, actinic keratosis, and basal cell carcinoma, and is also under investigation for treatment of other cancer types. Resiquimod (R848) was under development for the treatment of genital herpes, but these efforts were suspended in late clinical trials due to inadequate efficacy [113].

    TLR9 recognizes single-stranded DNA with unmethylated CpG (the p simply denotes the phosphodiester bond) sequences, which are under-represented and selectively methylated in the mammalian genome and therefore serve to alert DCs to the presence of viral and bacterial DNA [2, 167, 168]. In mice, optimal immune activation requires a CpG motif flanked by two 5′ purines and two 3′ pyrimidines (e.g., GACGTT), whereas in humans the optimal motif is GTCGTT [169, 170]. CpG DNA is known to be an excellent immune adjuvant in various murine disease models and to drive T_H1 immune responses. Synthetic CpG oligodeoxynucleotide (ODN) agonists for TLR9, therefore, have great potential in therapy [170, 171]. Several classes of CpG ODNs are known, varying in nucleotide sequence, single- or double-stranded, varying CpG content and content of phosphorothioate internucleotide linkages. Class B CpG ODNs have attracted most interest clinically. They are single-stranded and fully phosphorothioated in order to confer resistance to nuclease degradation and promote ODN uptake. B-ODNs are now in late-stage clinical trials as vaccine adjuvants to enhance immune responses against infectious diseases and cancer [171]. It is well established that activation by CpG ODNs requires TLR9, but it was unclear how B-ODN administered externally gains access to the receptor ectodomain inside lysosomes. Recent studies suggest that DEC-205 (CD205) is a key cell surface receptor involved in the uptake of B-ODN, thereby facilitating stimulatory activity in DCs and B cells [172].

    Before ligand binding, TLRs 7–9 exist as preformed dimers. Ligand binding then causes conformational changes that result in receptor activation [82]. This preformed dimer model contrasts with the ligand-induced dimerization mechanisms seen with other TLRs [74, 75, 80, 83, 144]. After ligand binding to the ectodomains, downstream signaling is initiated and transmitted to the nucleus, first through recruitment of the cytoplasmic adaptor molecule MyD88. Finally, the transcription factors NF-kB and AP-1 are activated, which leads to up-regulation of proinflammatory cytokine genes, including those for TNFα, IL-6, and IL-12. In addition, signaling can occur via IFN-regulatory factors, which leads to IFN-α gene transcription. However, the TLR types in different cell types can respond to agonists in different ways, resulting in diverse proinflammatory cytokine production profiles.

2.6.1 Structure of TLR8

Recently, the first crystal structure of a member of this TLR subfamily was reported, that for human TLR8, both as unliganded preformed dimer and as a ligand-induced activated dimer [160]. The complete ectodomain of hTLR8 (residues 27–827) used for crystallization appears as two fragments of 60 and 50 kDa by SDS PAGE, corresponding to N-terminal (residues 27–457) and C-terminal (residues 458–827) fragments, due to Z-loop cleavage inside the producing cells (Fig. 26). However, the fragments of each chain remain associated in the native purified protein.

Rather than having the horseshoe-like shape seen for the TLRs described so far, the number of LRRs in each ECD of TLR8 is large enough to generate a complete ring structure with inner and outer diameters of 3.5–4.5 and 7–8 nm, respectively (Fig. 26a). Residues at the N-terminus of TLR8 thus interact directly with others at the C-terminus of the ECD through hydrogen bonding and hydrophobic contacts. The overall shape is rather like a doughnut with a single cut and with the ends remaining in close proximity. The inner concave face comprises a circular β-sheet structure, formed by association of the β-strands in each LRR. This face is decorated with several glycans attached to conserved Asn side chains. Interestingly, the Z loop contains the part of the LRR that would normally exist as a loop/helical segment on the outer convex face of the doughnut. However, the cleavage site is located at the N-terminal side of the Z loop, which allows the main ≈24 residue portion of the loop to tuck back under the inner concave face, rather than being situated on the outer convex face (Fig. 26a). As a result, on the convex face LRR13 interacts directly with LRR15. At the C-terminal end of the Z loop, a single turn of α-helix occurs, which is stabilized by a disulfide bridge between Cys479 and Cys509 (within LRR16). In the unliganded TLR8 dimer, the two doughnut-shaped ECDs bind to each other face-on-face (Fig. 26b), with only a small transposition so that the holes in each doughnut are not exactly lined up. The C-terminal regions of the two TLR8 monomers are separated by ≈5.3 nm, which likely precludes proper association of the cytoplasmic domains; therefore, this dimeric structure should represent the unactivated form of TLR8. The insertions in LRRs 2, 5, 8, 11, 18, and 20 after each consensus β-strand form loop structures that protrude to the same lateral face of the ECD. Upon formation of the receptor dimer, these loops on both ECDs are buried at the dimer interface.

The crystal structures of TLR8 in complex with three different small-molecule agonist ligands [CL097, CL075, and R848 (Fig. 25)] were also determined [160]. Essentially, the same dimeric structure of TLR8 was observed in each of the three complexes. The two TLR8 chains within one dimer are essentially superimposable (RMSD of 0.04 nm). Upon ligand binding, however, the two C-termini in the dimer are brought into closer proximity (≈3 nm) (Fig. 26c), which would likely enable the intracellular TIR domains to interact optimally and initiate downstream signaling. The dimerization interfaces of the unliganded and ligand-induced dimers buries ≈1290 and ≈2150 Å^2 of accessible surface area on each side, with much of the difference accounted for
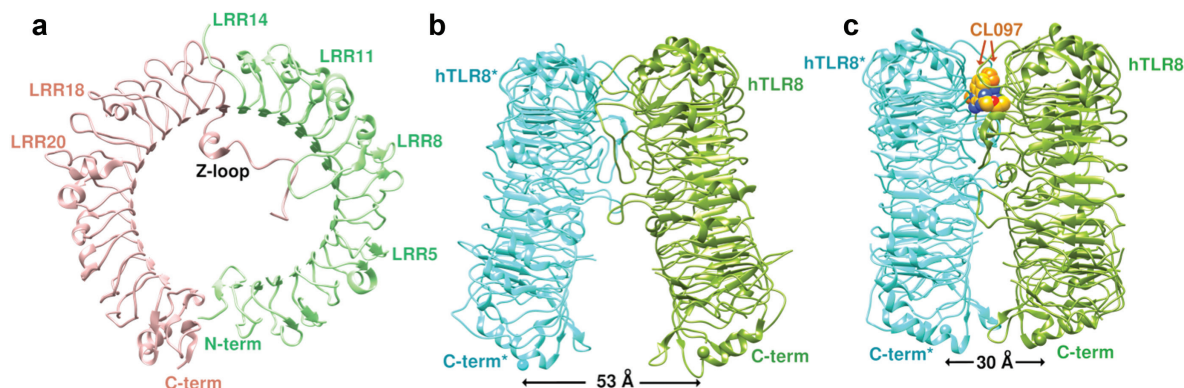

Fig. 26 Structures of human TLR8. (a) Monomer structure of the hTLR8-CL097 complex. (b) Side view of the inactivate TLR8 dimer. (c) Activated state after binding ligand CL097 (see also Fig. 28). The distance between the Cα atoms of the C termini is shown.
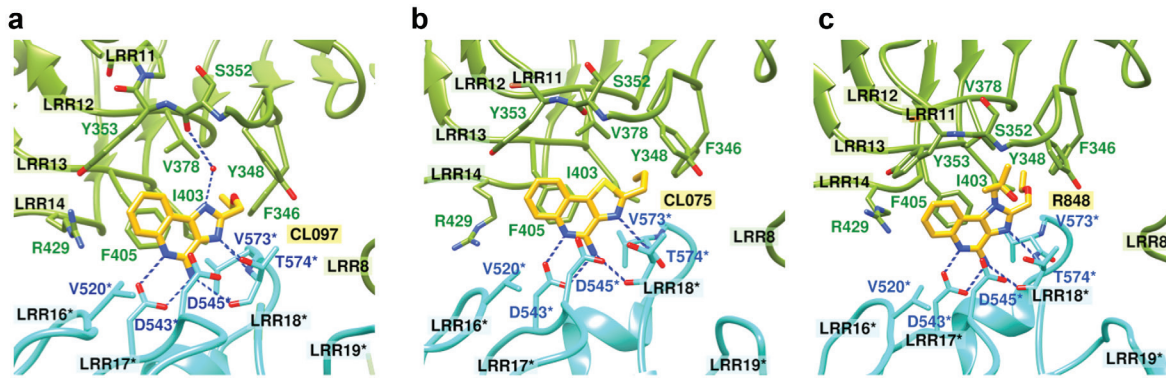
Fig. 27 Ligand recognition by TLR8. Residues involved in the interaction of TLR8 with CL097 (a), CL075 (b), and R848 (c) are shown.

by a ligand-mediated interface (Fig. 27). Ligand-induced local conformational changes occur predominantly in surface loops of LRR5, LRR8, and LRR17–20. The protein–protein interface involves loops on LRR5, LRR8, and LRR14–20, with both van der Waals, hydrophobic, and hydrogen-bonding interactions. The ligand-mediated interface is composed of residues in LRR11–14 and LRR16–18.

Two ligand molecules in each complex interact with the TLR8 dimer at the dimerization interface at positions related by a non-crystallographic 2-fold axis. The first is close to LRR11–14 and LRR16*–18*, and the second is close to LRR11*–14* and LRR16–18. Contact with each ligand is made by multiple residues from TLR8 and TLR8*, including face-to-face pi-stacking between the side chain of Phe405 and a benzene ring in the ligand (R848, CL097, and CL075). Several hydrogen bonds are also detected between heteroatoms in the heterocyclic ligands and residues in TLR8 (Fig. 27). The 2-propyl (CL075) or 2-ethoxymethyl (CL097 and R848) substituents of the ligands protrude into a small hydrophobic pocket on TLR8. These hydrophobic interactions may be important for the agonist activity of ligands targeting TLR8, since R837, an imidazoquinoline similar to R848, specifically activates TLR7 but not TLR8, and lacks the 2-substituent in the imidazole ring.

Ligand binding leads to a reorganization of the dimer structure, and the way in which the doughnuts and individual LRR modules interact (Fig. 28). In the unliganded (inactivated) form, the interactions of LRR8 and LRR18*, as well as LRR11–13 and LRR14*–15*, are the major contributors to the dimerization interface. After ligand binding LRR8 changes its interaction partner to LRR18*–20*, LRR18* to LRR11, and LRR11–13 to LRR17*–18*. Concomitantly, Phe405 in LRR13, which makes contact with Phe494* in LRR15* in the unliganded form, now interacts with the ligand in the liganded form. In addition, after ligand binding new interactions between LRR5 and LRR20* are created. This rearrangement of the complex upon binding of relatively small non-natural ligands is quite remarkable. The aromatic ligands are much smaller than the natural PAMPs the receptor has evolved to detect. Intriguingly, mutations in the receptor that abolish or reduce agonist activity of the small ligands, also abolished or reduced significantly signaling by ssRNA, suggesting that the small ligands and ssRNA share the same binding site on TLR8. However, other TLR8 side chains that surround the small-molecule binding site (Arg429, Tyr353, Val378) are not essential for binding the small molecule ligands, yet their mutation abolishes binding to ssRNA. So far, it remains unclear exactly how ssRNA interacts with the TLR8 dimer. Further structural studies are needed to resolve this problem.

2.7 TLRs 10, 11, 12, and 13

So far, 10 different TLRs have been identified in humans, and 13 in mice. As yet, no ligand has been identified for TLR10 from either species. However, TLR10 has been shown to form homodimers, and to interact with TLRs 1 and 2 [173]. The crystal structure of the intracellular TIR domain of human TLR10 has been reported, in the form of a symmetric dimer most likely corresponding to the signaling competent form [174].
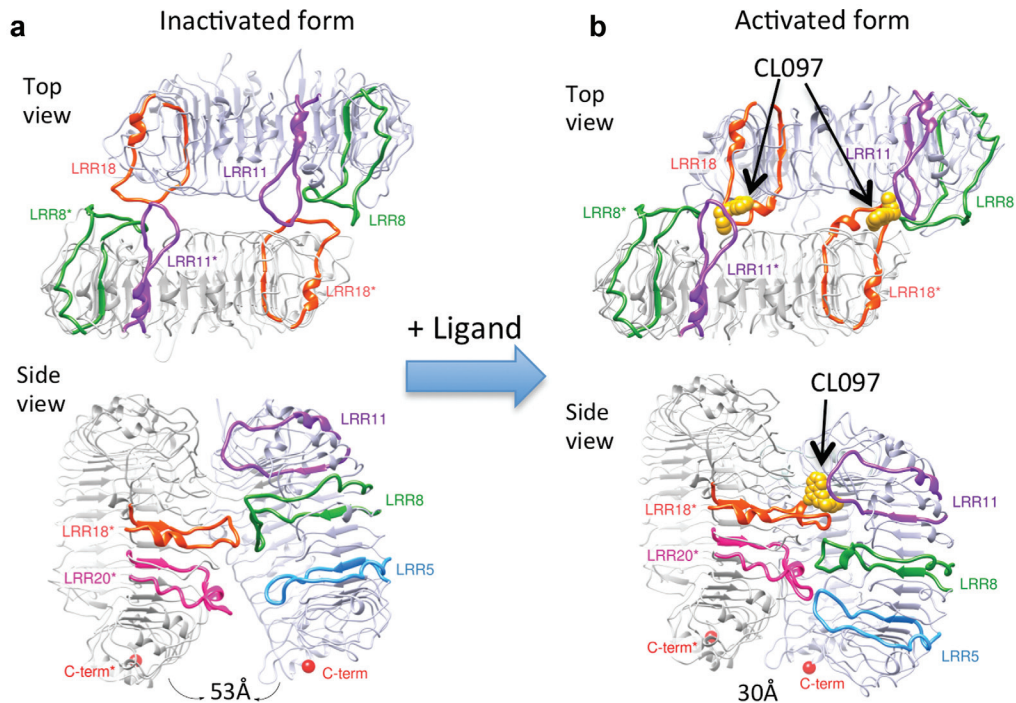
Fig. 28 Conformational changes in TLR8 induced by binding to agonistic ligands. The unliganded inactive form (a) is transformed into the activated form (b) upon ligand binding.

TLR11 and TLR12 are both murine TLRs located in endosomal compartments, which are required to mount an effective protective response against the protozoan parasite *Toxoplasma gondii* and other phylogenetically related microorganisms. These TLRs, working as heterodimers, can recognize profilin-like proteins in these parasites [175, 176]. TLR13 is another endosomal murine TLR, which is involved in recognition of microbial nucleic acids, in particular, a conserved bacterial 23S ribosomal RNA sequence that includes the binding site of macrolide (e.g., erythromycin), lincosamide, and streptogramin group (MLS) antibiotics in Gram-positive bacteria, like *Staphylococcus aureus* [177, 178]. Moreover, modifications to the MLS sequence that give rise to resistance to these antibiotics in bacteria fail to stimulate TLR13, thereby linking mechanisms of antibiotic resistance with defective immune stimulation via TLR13 [178].

3 Structures Of C-type lectin receptors

3.1 Structures of DC-SIGN and DC-SIGNR

DC-SIGN (dendritic cell-specific intercellular adhesion molecule-grabbing nonintegrin) is a prototype type II transmembrane CTL found on macrophages and DCs. There it interacts with endogenous molecules, such as ICAM-2 and ICAM-3, as well as pathogen-associated ligands found on bacteria and viruses. DC-SIGN mediates endocytosis and targeting of foreign antigens to late endosomal/lysosomal compartments for degradation and presentation to T cells. These interactions are mediated through carbohydrates, including N-linked high-mannose structures (typically 5–9 terminal mannose units such as $Man_9GlcNAc_2$), and branched fucosylated oligosaccharides that are common on parasites. Signaling through DC-SIGN influences cytokine production levels in response to binding pathogen-derived carbohydrates [32]. Furthermore, cooperation between innate signaling pathways mediated by different PRRs, such as the CLRs and TLRs, plays an important role in tailoring adaptive immune responses to different pathogens. On the other hand, DC-SIGN mediated internaliza-

tion is exploited by some pathogens to gain entry to target cells. For example, DC-SIGN is exploited by HIV-1, Ebola, and Dengue viruses as well as by *Mycobacterium tuberculosis* to facilitate entry to cells [179–182]. Thus, carbohydrate recognition by DC-SIGN is important for both normal immune responses and opportunistic exploitation of the receptor by pathogens.

DC-SIGNR (or L-SIGN) is a closely related transmembrane receptor found on endothelial cells [183]. DC-SIGN and DC-SIGNR share 77 % sequence identity and are closely similar in overall architecture. However, DC-SIGNR binds only to high-mannose oligosaccharides and does not mediate uptake and degradation of glycoconjugates. Thus, whereas DC-SIGN functions both in adhesion and in endocytosis of pathogens, DC-SIGNR has only the properties of an adhesion receptor and with a more restricted set of ligands. However, DC-SIGNR can also serve as an entry-receptor for many of the same viruses that exploit DC-SIGN.

DC-SIGN and DC-SIGNR consist of an intracellular N-terminal domain for signaling in the cytoplasm and a globular C-terminal CRD that binds carbohydrate ligands. The CRD is separated from the transmembrane region by a neck domain that contains multiple tandem repeats (Fig. 29). The neck domain is required for tetramerization, which is mediated through a coiled coil that forms a parallel 4-helix bundle. This coiled-coil neck region generates stable tetramers and projects the CRDs in a cluster some distance away from the cell surface. However, the CRDs appear not to be held in a fixed position but can reorient to engage with carbohydrate ligands [184]. The coiled coil contains a heptad repeat sequence (denoted *abcdefg*, where *a/d* are mostly hydrophobic residues), which is embedded within multiple tandem repeats of a largely conserved 23-mer motif (Fig. 29) [183]. The side chains of the hydrophobic (*a/d*) residues within each heptad are buried at the center of the 4-helix bundle. Each motif also contains a conserved proline residue that is predicted to interrupt regular helical structure. Crystal structures of the neck region of DC-SIGNR revealed a series of 4-helix bundles, connected by short non-helical linkers containing the conserved prolines (Fig. 30) [185–187].

Crystal structures of the CRDs from DC-SIGN and DC-SIGNR bound to a pentasaccharide (GlcNAc1β-2Manα-[GlcNAc1β-2Manα-6]-3Man) (Fig. 31a) were reported in 2001 [188]. The pentasaccharide contains a central mannose linked through the 3- and 6-positions to the disaccharide GlcNAc1β-2Man1α. Whereas the DC-SIGN-carbohydrate complex contains a pair of CRDs bound to one oligosaccharide, the DC-SIGNR complex reveals a 1:1 CRD–oligosaccharide interaction (Fig. 32). In the DC-SIGN complex, one CRD forms the same contacts to four sugar residues as observed in the DC-SIGNR structure, but the fifth sugar residue (terminal GlcNAc) interacts with the partner CRD in the DC-SIGN complex. The conformation of the pentasaccharide is essentially identical in both structures. The internal α-1,2-linked mannose binds to the Ca^{2+} site found on all

a DC-SIGN/R structure

b DC-SIGN neck region repeats

```
1            IYQNLTQLKAAVGEL
2    SEKSKLQEIYQELTQLKAAVGEL
3    PEKSKLQELYQELTRLKAAVGEL
4    PEKSKLQEIYQELTWLKAAVGEL
5    PEKSKMQEIYQELTRLKAAVGEL
6    PEKSKQQEIYQELTRLKAAVGEL
7    PEKSKQQEIYQELTRLKAAVGEL
8    PEKSKQQEIYQELTQLKAAVERLCHP
```

c CRDs
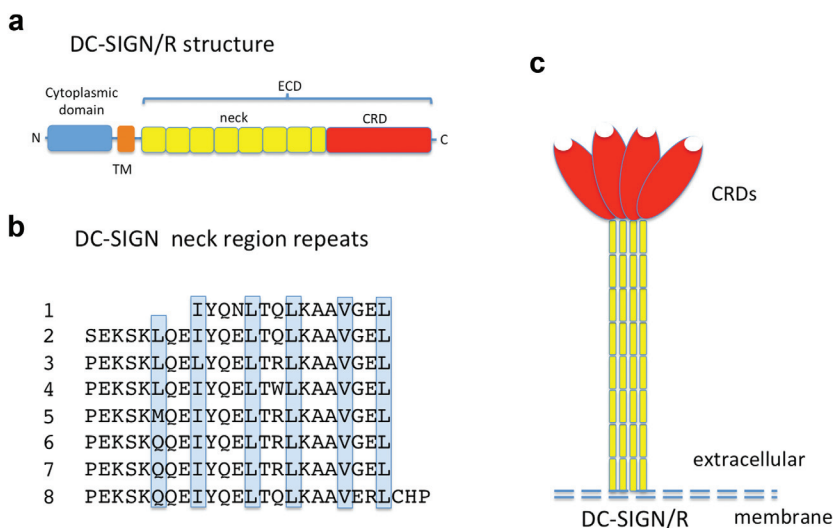
extracellular

DC-SIGN/R   membrane

Fig. 29 (a) Schematic representation of the domain organization of DC-SIGN and DC-SIGNR. (b) Repeat motifs in the neck region. Residues at a/d positions of the heptad repeat are highlighted. (c) The CRDs are projected and held away from the membrane by the neck.
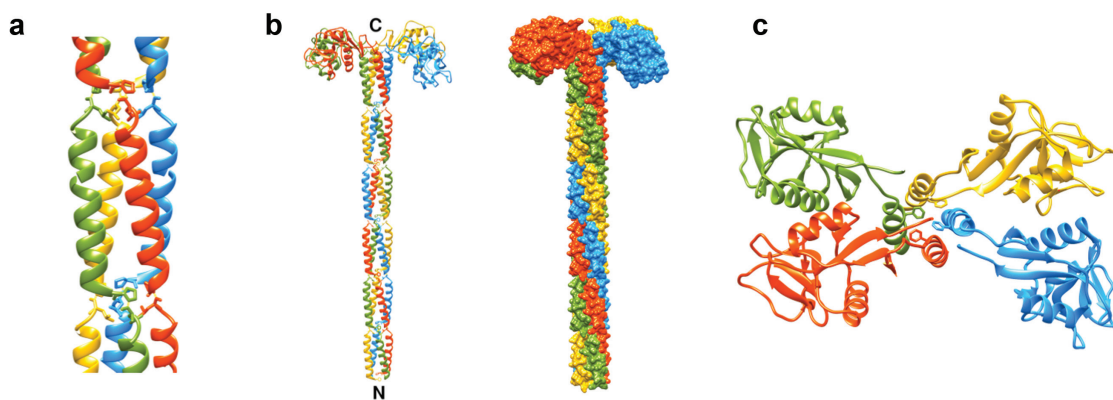
Fig. 30 (a) Four-helix bundle of the 23-amino acid repeat motif of the neck region of DC-SIGNR. The repeat motifs are linked through non-helical segments, including a Pro residue. (b) Model of a neck fragment containing multiple 4-helix bundles, with the C-terminal CRD at the top [187]. (c) Top view of the CRD with the first 4-helix bundle [185].

CTLs, with the equatorial 3- and 4-hydroxyls coordinating with the Ca^{2+} ion (Fig. 32). The oligosaccharide also makes several hydrophobic and hydrogen-bonding interactions with the rather flat ligand-binding site on each receptor protein. The crystal structures reveal the strong preference for binding N-linked glycans where the central mannose is linked in the α-form, a structural motif only found in high-mannose oligosaccharides.

Further insights into the ability of DC-SIGN to bind fucosylated ligands came from a crystal structure of DC-SIGN bound to the pentasaccharide lacto-*N*-fucopentaose-III (LNFP-III), which contains the Lewis[x] trisaccharide (Fig. 33) [189]. The conserved Ca^{2+} at the primary carbohydrate-binding site is coordinated to the 3- and 4-hydroxyls of the α-1,3-linked L-fucose residue in LNFP-III (Fig. 34a). However, since these hydroxyls are equatorial and axial, respectively, rather than both being equatorial as seen in bound mannose (Fig. 32), the fucose sugar is tilted. The compact structure of the Lewis[x] trisaccharide is oriented with the central GlcNAc
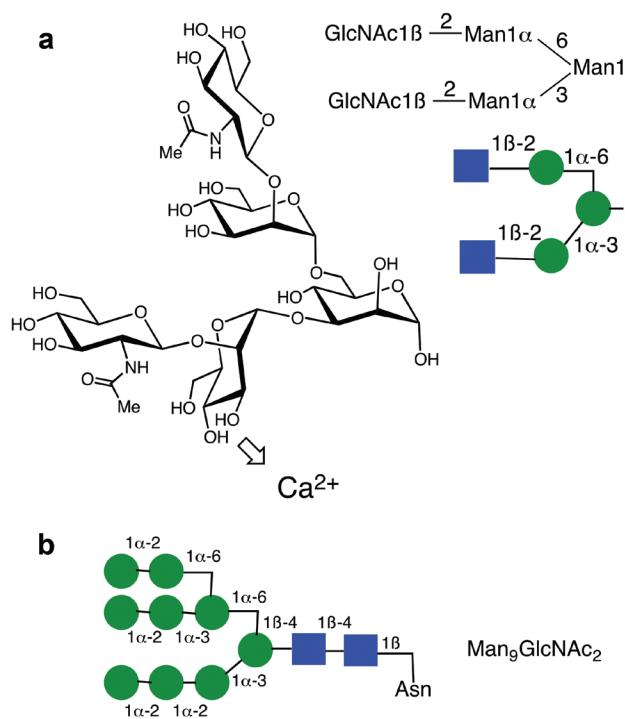
Fig. 31 (a) Structures of oligosaccharides recognized by DC-SIGN and DC-SIGN/R. (b) N-Linked high mannose structure.
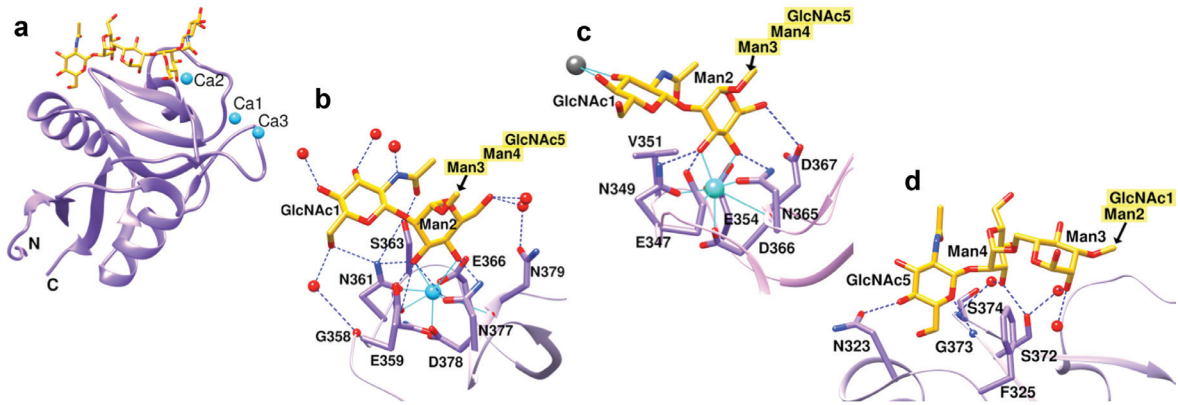
Fig. 32 (a) Structure of the CRD of DC-SIGN bound to GlcNAc$_2$Man$_3$. Cyan spheres are bound Ca^{2+} ions [188]. (b) Interactions of the α-1-3-linked branch with DC-SIGNR. Only key sugars are shown. Ca^{2+} coordination is shown by blue lines, hydrogen bonds as dashed lines. (c) Interactions of α-1-3-linked branch with DC-SIGN. The terminal GlcNAc1 forms a cross-link by forming the typical CTL interaction with the principal Ca^{2+} site of another CRD (only coordination bonds to the Ca^{2+} (gray) are shown). (d) Interaction of the α-1-6-linked branch with DC-SIGNR, including the central (reducing) mannose (Man3).

residue in LNFP-III pointing away from the protein and the terminal galactose residue contacting the protein in a secondary binding site. Various van der Waals as well as hydrogen-bonding interactions are seen between the bound ligand and the receptor. The structure suggests that other fucose-containing ligands may bind with fucose in the primary (principal Ca^{2+}) binding site.
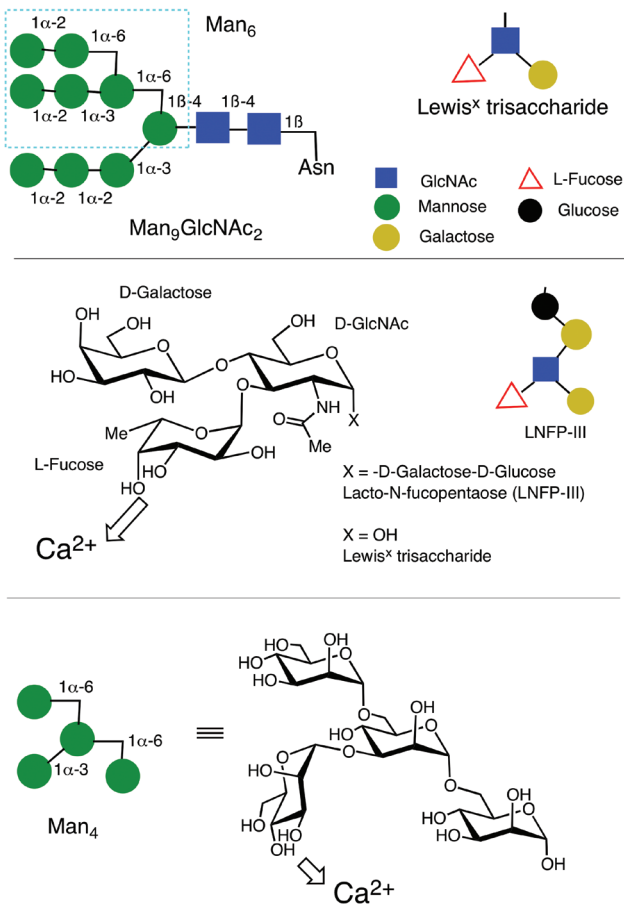
Fig. 33 Structures of ligands used in structural studies of DC-SIGN and DC-SIGNR [189].

Fig. 34 (a) Structure of DC-SIGN bound to LNFP-III (see Fig. 33). (b) Structure of Lewis^x trisaccharide bound to DC-SIGNR. Ca^{2+} coordination is shown by pink lines, hydrogen bonds as dashed lines. Cyan spheres are bound Ca^{2+} ions.

A crystal structure was also reported of a DC-SIGNR fragment bound to the simple Lewis^x trisaccharide [189], although this ligand has only weak affinity for DC-SIGNR ($\approx$mM). The position of the bound ligand is very similar to the position of the corresponding portion 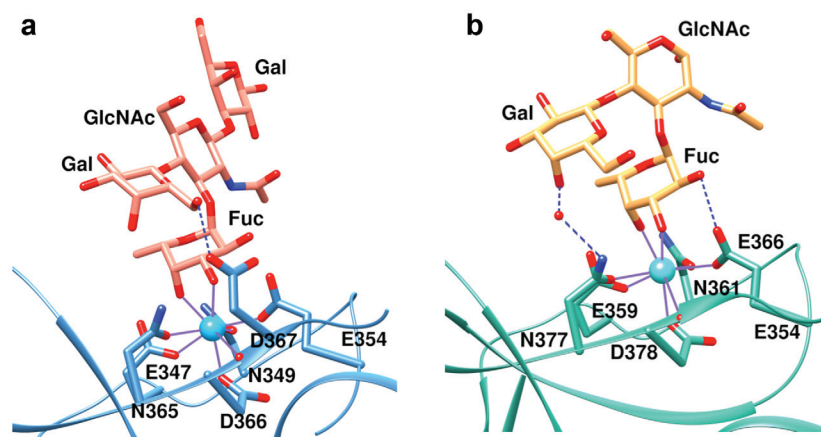of the LNFP-III ligand bound to DC-SIGN. However, several key differences in ligand-protein contacts are also seen between the two complexes (Fig. 34b). A third crystal structure comprises a complex of the CRD of DC-SIGN bound to a branched tetra-mannose (Man$_4$) (Fig. 33) [189]. The α-linked terminal reducing mannose residue is accommodated because it points away from the protein, indicating why binding of branched high-mannose oligosaccharides is permitted by DC-SIGN.

Further insights into binding of branched mannose carbohydrates to DC-SIGN came from crystal structures of its CRD bound to a hexa-mannose fragment (Man$_6$, Fig. 33) of the full N-linked high-mannose oligosaccharide (Man$_9$GlcNAc$_2$), as well as to the disaccharide Manα1-2Man (Table 1) [190]. These structures revealed the carbohydrate ligands occupying two different overlapping orientations in the crystal, in a ratio $\approx$75:25. The major orientation seen with the hexamer corresponds to the arrangement seen in earlier crystal structures [189], but the α-1-6 branch of the carbohydrate is not visible. The penultimate α-1-3-linked mannose that forms one arm of the outer-branched trimannose unit interacts with the primary Ca^{2+} site. In the second (minor) orientation, the same mannose is bound to Ca^{2+}, but its orientation is reversed by a 180° rotation. This rotation exchanges the position of the 3- and 4-OH groups so that they can still coordinate to Ca^{2+}. In this orientation, only two sugars are visible in the complex; the mannose at the Ca^{2+} site and the non-reducing terminal α-1-2-linked mannose. In the complex with the disaccharide Manα1-2Man, binding is only seen at the principal Ca^{2+} site; no other binding site was occupied on the protein. However, the ligand again binds in two orientations, related by a 180° rotation that interchanges the positions of the 3- and 4-OH groups coordinated to Ca^{2+}.

These observations show that high-mannose oligosaccharides can interact with DC-SIGN in multiple orientations, with the primary energetic interaction occurring at the primary Ca^{2+} site, but with multiple secondary interactions occurring through neighboring binding sites. The ability of DC-SIGN (and DC-SIGNR) to bind high-mannose glycans in multiple orientations may facilitate multivalent binding of clusters of CRDs to glycans displayed in various arrangements on the surface of a pathogen. The CRDs of DC-SIGN and DC-SIGNR bind Man$_9$GlcNAc$_2$ oligosaccharide 130- and 17-fold more tightly than mannose (mannose affinity is weak – in the mM range), and the affinity for a glycopeptide bearing two such oligosaccharides is increased a factor of 5- to 25-fold [191]. Moreover, closer spacing of the two oligosaccharides causes a greater increase in affinity to DC-SIGNR than to DC-SIGN, suggesting that the spacing of high-mannose N-linked oligosaccharides on a protein surface influences significantly the selectivity of ligand recognition.

Langerin is another type-II transmembrane PRR found on Langerhans cells. The ECD comprises a neck region that forms trimeric helical bundles, linked to a CRD at the C-terminus. The CRD binds various monosaccharides in a calcium-dependent manner, including mannose, fucose, and *N*-acetylglucosamine, as well

as oligosaccharides like mannan [192]. Binding of the monomeric CRD to monosaccharides is weak (≈mM range). However, oligomerization again facilitates multivalent, higher-affinity binding to oligosaccharides. A role for langerin has been proposed in internalization of both self- and non-self-glycoprotein antigens. Crystal structures show how the CRD from langerin interacts with mannose and the disaccharide maltose (Table 1) [193]. In both structures, an important determinant of monosaccharide recognition is coordination through the pyranose equatorial 3- and 4-OH groups to Ca^{2+} in the primary Ca^{2+} site.

3.2 Mannose-binding lectins

Mannose is found in an extensively branched multivalent format only on the surface of microorganisms. It is not surprising, therefore, that the innate immune system has developed mechanisms for recognizing this form of mannose as a PAMP. However, mannose is also found in many human glycoproteins at non-terminal sites and at a much lower surface density. A key role in innate defense is played by cell surface membrane proteins, such as the mannose receptor, which mediates phagocytosis of microorganisms by macrophages and DCs. In addition, soluble mannose-binding lectins in the circulation are greatly up-regulated during infections, and upon binding to terminal mannose on the bacterial surface, initiate antibody-independent complement activation.

The mannose receptor (MR) belongs to a small subgroup of the C-type lectin superfamily, which also includes DEC-205, Endo180, and PLA_2R [25, 29, 34, 194]. The MR binds carbohydrate moieties on pathogens, such as bacteria, fungi, parasites, and viruses. However, the MR also binds endogenous glycosylated molecules, in particular, pituitary hormones such as luteinizing hormone and thyroid-stimulating hormone, influencing their circulatory half-life. Thus, the MR has multiple roles apart from its functions in the immune system [27, 36, 195]. The MR is preferentially expressed on DCs, tissue macrophages, and specialized endothelial cells, and constitutively recycles between the plasma membrane and the early endosomal compartment.

The MR has an unusual topology, with the N-terminus outside the cell, a single transmembrane segment and a short C-terminal cytoplasmic domain [34]. The ectodomain contains a cysteine-rich domain at the N-terminus, followed by a fibronectin type-II (FNII) domain and then a series of eight C-type lectin domains (CTLDs 1–8) leading down to the cell membrane (Fig. 35). The Cys-rich domain binds to a variety of sulfated carbohydrates, including glycoproteins and hormones terminating in sulfated *N*-acetylgalactosamine, 3-sulfated galactose, and fucose, as well as chondroitin-4-sulfates. The function of the FNII-domain is so far
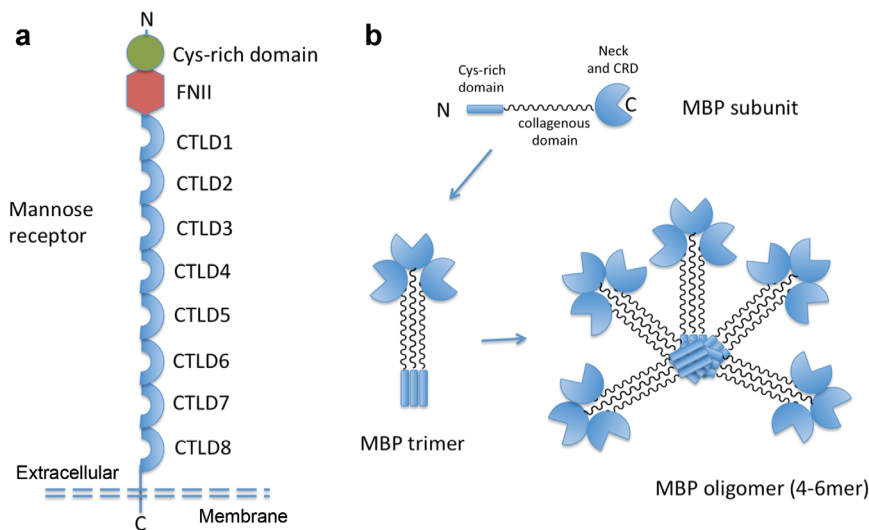

Fig. 35 (a) Domain structure of the mannose receptor. (b) Organization of a typical soluble mannose-binding protein.

unclear, but the domain is predicted to adopt the canonical fold seen in other FNII domains. The MR displays Ca^{2+}-dependent binding to terminal mannose, fucose, and *N*-acetylglucosamine residues. Only CTLD4 shows significant carbohydrate binding as an isolated domain, although CTLD5 retains weak sugar binding activity, which is consistent with the observation that only CTLD4 and CTLD5 contain the conserved residues required for Ca^{2+}-dependent carbohydrate recognition [34].

Soluble mannose-binding proteins (MBPs) recognize a wide variety of branched microbial carbohydrates present on the surface of yeasts, bacteria, mycobacteria, and some viruses such as influenza and HIV-1 [35]. Human MBP is a multimeric protein consisting of a ≈32 kDa subunit, which associates into trimers and then into higher multimers in a "bouquet-like" array (Fig. 35). Each subunit contains an N-terminal cysteine-rich domain, followed by a collagen-repeat region, a neck region comprising a coiled-coil motif, and finally a CRD exposed on the outer surface of the bouquet that requires calcium to bind ligand. The coiled coil mediates formation of a trimeric parallel α-helical bundle, and the resulting trimers then assemble further into higher multimers, likely through association of their collagen-like domains (Fig. 35). The affinities of the isolated CRDs for monosaccharides are relatively low (mM range), however, the apparent affinity of the multimeric MBP for branched multivalent ligands is dramatically higher (nM) [196]. On the other hand, assemblies of lower valency, such as linear carbohydrate chains found on host glycoproteins, interact with MBPs with much lower affinity [197]. Crystal structures proved to be important for understanding how MBPs can distinguish between host and pathogen-derived mannosylated ligands, as described below.

3.2.1 Mannose-binding proteins

The crystal structures of rat and human MBP fragments, comprising the CRD and the coiled-coil neck domain, revealed how MBP trimers form the basic recognition unit for branched oligosaccharides on microorganisms (Fig. 36) [198, 199]. The neck region incorporated into both constructs was 24–30 residues long, which as isolated peptides show no regular secondary structure. In the context of the MBP fragment, this neck region adopts a parallel trimeric helical bundle, with hydrophobic residues at the a/d positions of a heptad repeat. The presence of the CRD domain in each chain stabilizes the trimeric structure through direct contacts with neighboring chains. This packing arrangement gives each CRD a well-defined orientation, and determines the relative spatial organization of the carbohydrate binding sites, which are located about ≈5 nm from each other. This relatively large distance makes it difficult for a single mammalian high-mannose oligosaccharide chain to interact with more than one CRD within the trimer. However, since the carbohydrate binding sites all

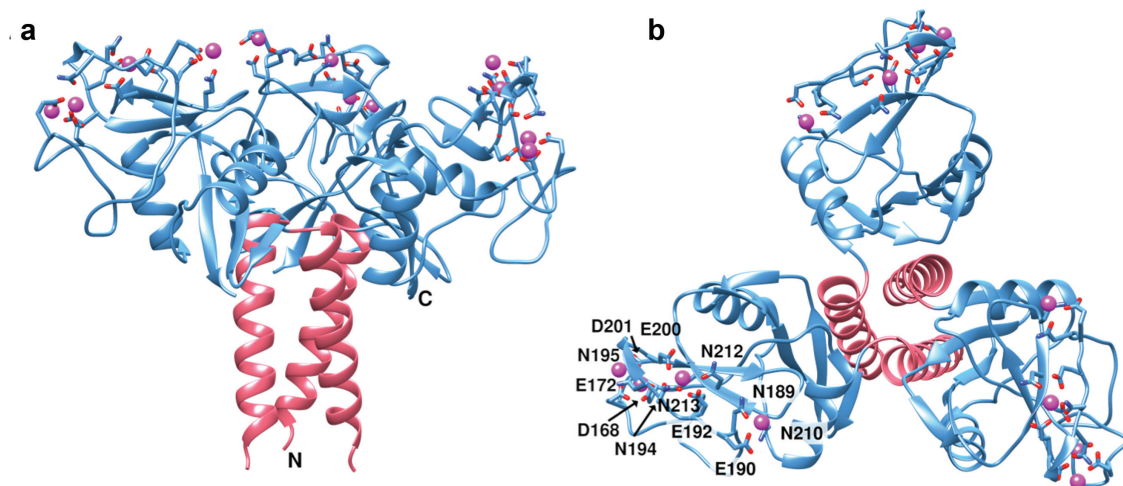**Fig. 36** Trimeric unit of the human mannose-binding protein. Side (a) and top (b) views. Pink spheres are bound Ca^{2+} ions; surrounding residues are highlighted. The coiled-coil region is red. The N- and C-termini of one chain are indicated.

point out from the same side of the molecule, they are ideally positioned to interact with dense sugar arrays on a microbial cell surface.

Within each CRD, bound Ca^{2+} ions are important for the interaction with carbohydrate ligands. The primary Ca^{2+} ion interacts with ligand by coordinating to the 3- and 4-hydroxyls in bound mannose, as revealed in a crystal structure of rat mannose-binding protein A (MBP-A) complexed with a branched oligomannose sugar (Table 1) [200]. Only the terminal mannose of each visible branch of the oligosaccharide interacts with the CRD; the rest of the carbohydrate points away from the CRD and makes limited contacts with the protein. Further crystal structures of MBP-A and MBP-C have been reported with various bound monosaccharides (e.g., mannose, N-acetylglucosamine, and fucose) and disaccharides, as well as the MBP-A trimer cross-linked by a high-mannose oligosaccharide (Table 1) [201, 202]. The structures reveal that the monosaccharides or α-1-6-linked mannose can bind to MBP-A in one orientation, whereas α-1-2- or α-1-3-linked mannose binds in an orientation rotated by 180° around an axis relating the 3- and 4-OH groups. In contrast, a similar set of ligands all bind to MBP-C in a single orientation. Each MBP seems to interact with a particular ligand in preferably one orientation. The protein residues that serve as both Ca^{2+} and sugar ligands are the same in both MBP-A and MBP-C. However, one residue close to the binding site (His 189 in MBP-A) appeared by mutagenesis to have a key influence upon the orientation of bound ligands in MBPs [202–204]. Finally, the crystal structure of MBP-A trimer bound to $Man_6GlcNAc_2Asn$ revealed the two trimers cross-linked by the oligosaccharide (Table 1) [202]. The trimers are arranged in hexagonal packed sheets in the crystal with the sheets cross-lined by the oligosaccharide. It was suggested that this packing arrangement might indicate how a cell surface presenting mannose-terminated sugars (one layer of the crystal) could be recognized by a multimeric form of MBP-A (represented by the opposing layer).

3.2.2 The mannose receptor

The crystal structure of CTLD4 from the MR (Fig. 35) was reported in 2000, although the structure does not contain bound carbohydrate ligand [205]. The overall structure of the CTLD4 domain is similar to that of other C-type CRDs (Fig. 37a). Equivalent residues in this domain can be superimposed on those from MBP-A with an rmsd of 0.15 nm. Unlike the core of the domain, where the similarity is greatest, striking differences are seen in the size and structure of some surface loops. A rather large, extended, and flexible loop that follows the β2 strand is seen in CTLD4 (residues 701–734), which is absent in MBP-A. This region is stabilized in MBP-A by two bound Ca^{2+} ions, including the principal Ca^{2+} site that interacts with bound mannose. In the CTLD4 structure, however, only the primary site located close to the base of the extended loop is occupied with bound Ca^{2+}. Attempts to induce ligand binding by soaking crystals in high concentrations of sugar failed to
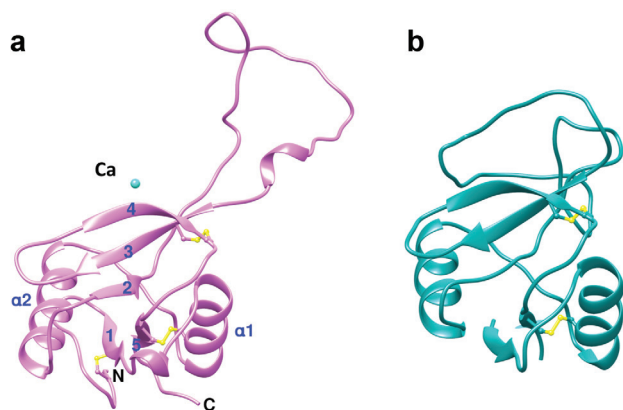

Fig. 37 Structure of the CTLD4 domain of the MR (a) with extended loop, and comparison to rat mannose-binding protein A (b) [205]. Cyan sphere is a bound Ca^{2+} ion. Disulfide bonds are in yellow.

reveal a binding site. It seems, therefore, that the structure present in the crystals is unable to bind the full complement of Ca^{2+} nor mannose ligands. It was speculated that the structure may be related to an intermediate that must occur physiologically in endosomes, where the slightly acidic pH could lead to loss of Ca^{2+} and de-binding of sugar ligands. The ligand and receptor then sort from one another, with the receptor returning to the cell surface for another round of ligand binding. Indeed, isolated MR CTLD4 displays reversible, pH-dependent Ca^{2+} and sugar binding, with a mid-point in the transition around pH 5.0.

Crystal structures have also been reported of the cysteine-rich N-terminal domain of the MR, which plays a critical role in binding sulfated glycoproteins, including pituitary hormones [206, 207]. This globular domain of ≈130 residues adopts a β-trefoil, possessing 12 antiparallel β-strands with six forming an almost β-barrel-like arrangement, with the entire trefoil stabilized by three disulfide bonds (Fig. 38a). The structure of this domain bears no resemblance to the CRDs found in the MR and other CTLDs. The ligand, 4-sulfated N-acetylgalactosamine (GalNAc-4-SO_4) binds with a K_D ≈0.1 mM with its sulfate group inserted into a binding pocket formed on one side by residues 111–116 in the loop between strands 11 and 12, and the other side by the side chains of Asn^{102} and Trp^{117}, and on the bottom by the Asn^{99} side chain (Fig. 38a). The sulfate group makes extensive hydrogen-bonding and van der Waals contacts to the protein. The galactose ring is positioned with the polar A-face exposed to solvent and the nonpolar B-face stacked against the indole of Trp^{117}. This stacking arrangement is also seen in several other protein-galactose complexes [206]. The sulfate group does not interact with positively charged side chains or counter ions. Rather, the interactions with the protein are mediated through multiple hydrogen bonds involving peptide bond NH groups. This domain binds to GalNAc or galactose when it is sulfated at the C3 or C4 positions, and more weakly if the sulfate is at C6. The lack of discrimination between sugars with and without the N-acetyl group is explained by the absence of any direct contact between the NAc group and the protein. The protein recognizes sulfated GalNAc when it is terminal but not internal within an oligosaccharide chain, which again can be nicely explained by the crystal structure. Another sugar residue linked to C3 would prevent access of the sulfate group to the binding pocket. The anomeric center of the bound ligand points away from the protein surface, thus allowing attachment of another sugar at this position without causing steric hindrance.

Crystal structures and affinities of 3-SO_4-Lewis[x], 3-SO_4-Lewis[a], and 6-SO_4-N-acetylglucosamine bound to the Cys-rich domain have also been reported [207]. The conformation of the protein is virtually identical in all crystal structures. Interestingly, a 6-SO_4-Gal binds with K_D ≈ 1 mM, whereas 6-PO_4-Gal showed no detectable binding to the Cys-rich domain, suggesting that the slightly larger phosphate group is not accommodated within the sulfate-binding site.
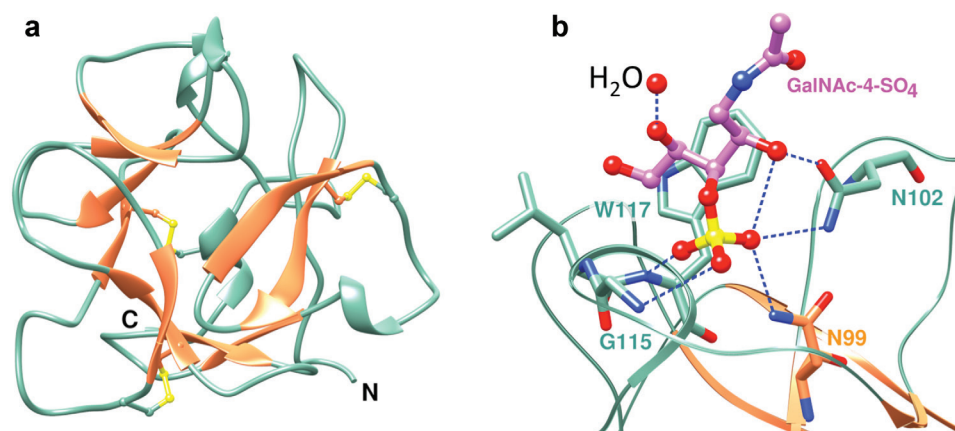

Fig. 38 (a) Cysteine-rich domain from the MR (Cys-MR). Disulfides are yellow ball-and-stick. (b) Cys-MR complexed with 4-SO_4-N-acetylgalactosamine. Sulfur atom in yellow, hydrogen bonds in blue dotted lines.

4 Structures of NOD-like receptors

Relatively few structural studies at atomic resolution have been reported on NLRs to date, nor how they recognize specific PAMPs. The crystal structure of a fragment derived from NLRX1 has been reported, but not bound to its RNA ligand.

4.1 RNA-binding domain of NLRX1

NLRX1 (also known as NOD9) is a proinflammatory activator that localizes to mitochondria, and stimulates the production of reactive oxygen species via TNFα activation [208]. However, this receptor has a more complex immune modulatory function, since it can also down-regulate inflammatory responses as a negative regulator of RIG-1 and TLRs through interaction with mitochondrial signaling adaptors [209, 210]. Recently, a crystal structure was reported of the C-terminal fragment of human NLRX1 (called cNLRX1, residues 629–975), along with evidence that this receptor interacts specifically with an RNA ligand [59].

The cNLRX1 protein contains the entire C-terminal putative ligand-binding domain comprising eight LRRs as well as an N-terminal 2-helix domain (LRR-NT) and a C-terminal antiparallel 3-helix bundle (LRR-CT) (Fig. 39) [59]. The central domain with eight LRR modules folds into a crescent shape with the parallel β-strands oriented on the concave side. The LRR1-LRR7 motifs have the typical LRR fold, except LRR8, which does not contain an α-helix on the convex side, but rather proceeds directly into the LRR-CT domain. The concave face encircles and docks onto one side of the LRR-CT domain. The cNLRX1 protein elutes in two fractions from a gel filtration column, with apparent masses corresponding to monomer (≈39 kDa) and hexamer (≈240 kDa). In the crystal structure, the protein assembles into a hexamer with 2-fold crystallographic symmetry and 3-fold non-crystallographic symmetry. Three cNLRX1 chains associate by intertwining of helical segments of their LRR-NT domains at the center to create the 3-fold axis. Two trimeric subunits then assemble into a hexamer, with the LRR-CT domains buried at the interface (Fig. 39). The LRR-CT domain plays a key role in dimerization and provides structural support for the LRR modules through intra- and inter-subunit interactions.
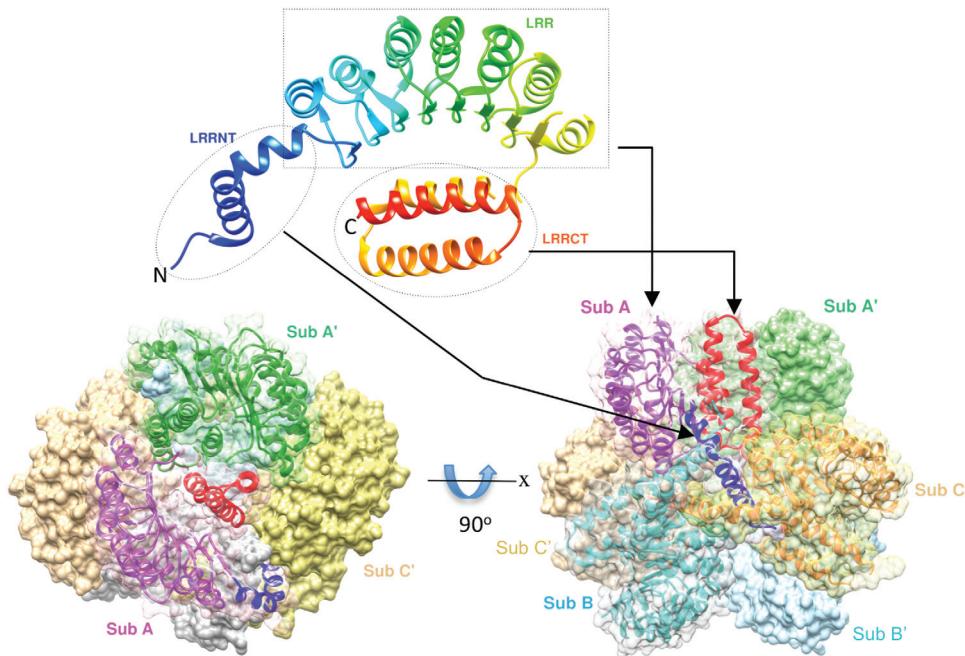

Fig. 39 Structure of cNLRX1. The monomer subunit is shown (top). The hexamer (bottom) is formed by trimerization of dimers (subunits A/A', B/B', and C/C'). The LRR modules in subunit A are labeled (adapted from [59]).

Potential ligands for cNLRX1 have been probed by electrophoretic band-shift assays [59]. In this way, binding to ssRNA and dsRNA was detected with binding affinities around 1 and 0.1 μM, respectively. However, no binding was observed to ssDNA or dsDNA. A possible binding site for the RNA was identified around a large surface-exposed patch of positively charged Arg and Lys residues, which lies on one face of each monomer. Mutation of a key arginine within this region, Arg699, inhibits RNA binding. A model was also proposed for the full-length NLRX1 hexamer, in which NOD-NAD domains extend from the LRR domain and decorate the exterior surface of the hexamer [59]. The NOD domains might further mediate oligomerization into ring-like assemblies, as well as mediating binding to other interacting partners.

4.2 Pyrin domains and CARDs

NLRs containing an N-terminal pyrin domain (PYD) or a caspase activation and recruitment domain (CARD) form the largest group of NLRs (NLRPs), with 14 members in humans (NLRP1-14) [38, 40, 41, 48, 50]. For example, NLRP1 and 3 activate the inflammatory immune response by detecting cytosolic PAMPs via their C-terminal LRR domains. Once activated, NLRP1 and 3 oligomerize via their central NACHT domains. Finally, the N-terminal PYDs recruit the ASC adaptor protein, and then pro-caspase 1 molecules. The resulting multi-component signaling complex, termed the inflammasome [48], results in autoproteolytic activation of caspase 1, which in turn cleaves proinflammatory cytokines such as pro-IL1β and pro-IL18 into their mature forms [47].

Several PYD structures have been determined recently, by crystallography or NMR spectroscopy, including from NLRP3 (also called NALP3) [211], NLRP1 (or NALP1) [212], the ASC adapter protein [213, 214], NLRP12 [215], NLRP7 [216], ASC2 [217], and NLRP4 [218]. These structures reveal a high degree of conservation with a fold comprising a 6-helix bundle. The structures of several CARDs determined by crystallography or NMR spectroscopy have also been reported [214, 219–222]. These domains also adopt a compact 6-helix bundle structure in crystals and in solution.

5 Structures of RIG-like receptors

The RLRs are multidomain proteins (Fig. 4). RIG-I and MDA5 share a similar domain organization, comprising tandem N-terminal CARD domains, followed by a domain that belongs to the DExD/H-box helicase super-family, followed by a small regulatory domain and a C-terminal RNA-binding domain (CTD) [64]. Helicases comprise a large superfamily of enzymes. They are nucleic acid-dependent ATPases that promote unwinding of dsDNA or dsRNA, and are therefore involved in many important cellular processes [223]. Helicase domains convert chemical to mechanical energy by coupling ATP binding and hydrolysis to protein conformational changes. In contrast to bona fide helicases, however, RIG-I translocates on dsRNA and recognizes the 5′-PPP motif, which leads to large conformational changes and RIG-I activation. Recent structural studies have shown that activation of RIG-I occurs through an RNA- and ATP-driven structural switch in the helicase domain, which exposes CARDs that then mediate downstream signaling [224]. This likely occurs through dsRNA binding to the helicase domain, resulting in activation of the ATP-hydrolyzing function and a confor-mational change of the complex that exposes the CARDs. The CARDs then transmit signals downstream by recruitment of other molecules, which leads finally to the expression of antiviral interferon-stimulated genes [225–227]. LPG2 has a similar domain organization, except that it lacks the CARD homology region.

Recent structural studies of RLRs (reviewed recently [228]) have provided important insights into the structural basis for dsRNA and 5′-PPP-RNA sensing by these receptors.

5.1 RNA recognition by RIG-I

A thermodynamic analysis revealed that blunt ended 5′-PPP dsRNA is required for maximal binding affinity to RIG-I [229], and the CTD provides the greatest thermodynamic contribution to binding. The helicase domain,

on the other hand, plays a key role in specific binding to dsRNA, rather than ssRNA. In 2008, crystallographic and NMR studies provided the first atomic models of the CTD domain of RIG-I [230, 231]. These results were followed by crystal structures of the RIG-I CTD bound to the blunt ends of a synthetic 5′-triphosphorylated dsRNA oligonucleotide. Interactions involving the 5′-PPP group were clearly visible in the complex, sequestered in a basic lysine-rich patch on the protein surface (Fig. 40) [232, 233]. The side chain of Phe853 stacks over both bases of the terminal base pair of the dsRNA in the complex, consistent with the requirement for a blunt-end duplex. Later, additional crystal structures were reported of larger fragments of RIG-I, also with bound nucleotide analogues, including ADP or ADP-AlF$_3$ (a mimic of the transition state of ATP hydrolysis) or ADP-BeF$_3$ (likely an ATP structural mimic), and bound 5′-triphosphorylated RNA [234–238]. These results provided key insights into the likely mechanism of RNA- and ATP-dependent receptor activation.

Pyle and co-workers reported crystal structures of human RIG-I in complex with dsRNA (lacking a 5′-PPP) and in a complex with a 5′-triphosphorylated RNA hairpin and ADP (Fig. 41) [237, 238]. This RIG-I protein lacked the N-terminal CARD domain (residues 1–229) and was co-crystallized first with blunt-ended duplex RNA. The RIG-I protein in the crystal completely surrounds the bound RNA and encloses it within an elaborate network of mostly polar interactions (Fig. 41). Numerous contacts are formed between 2′-hydroxyl groups on the RNA and protein amide groups. Helicase domain 1 (HEL1) faces the minor groove of the RNA and interacts with the backbone of both RNA strands. Almost half of the HEL2 domain is disordered in the structure, perhaps due to the absence of the CARD domains and/or ATP. An insertion domain (HEL2i) within HEL2, comprising an α-helical bundle, is a component of the ring that grips the dsRNA by interaction with the minor groove. HEL2i plays an important role in the specific recognition of RNA. Strikingly, HEL2 and the CTD are connected through a V-shaped structure (also called a pincer (P) motif [236]) composed of two long α-helices that make a 65° turn (Fig. 41), which establishes extensive interactions between HEL1, HEL2, and the CTD. Although the pincer is not part of the ATPase site, nor is it a direct RNA-binding site, it likely transmits information essential for RNA sensing between these sites. The CTD interacts with the RNA in a manner similar to that seen in previously reported CTD structures. The Phe853 side chain again stacks over the terminal base pair in the RNA. The dsRNA adopts an A form conformation, with no evidence of destabilization or partial unwinding. The RNA-binding interface involves at least three separate domains. RIG-I undergoes large ATP-driven conformational changes upon binding to dsRNA, with the dsRNA providing a template for RIG-I domain assembly [234].

The crystal structure of RIG-I(Δ-CARD) bound to a designed RNA hairpin containing a 5′-PPP and ADP is similar to the RIG-I:dsRNA structure described above (Fig. 42a) [238]. The RNA contains a 5′-triphosphate and a stem of 8 bp that is linked by a UUCG tetra-loop, an arrangement that mimics the panhandle-like structures formed by the genomes of many negative-stranded RNA viruses. The structure is likely to be biologically

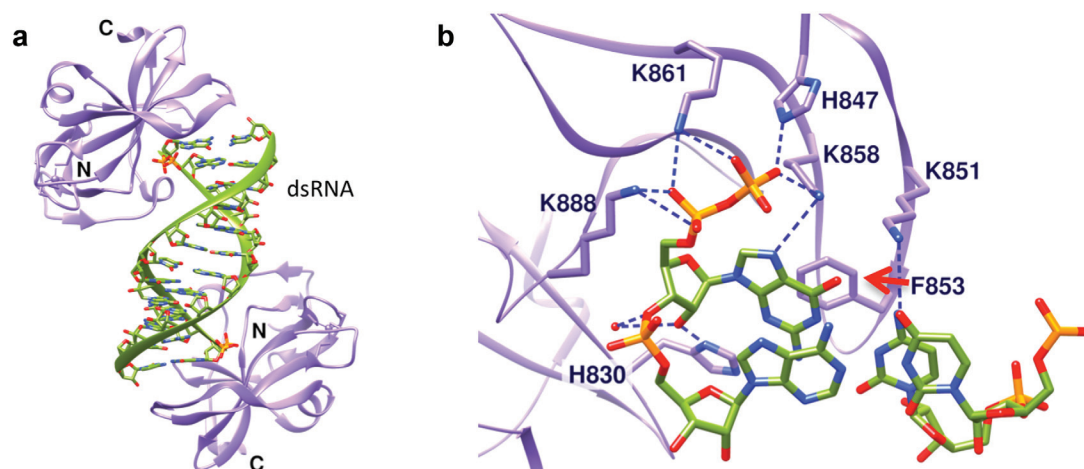**Fig. 40**   (a) Structure of RIG-I CTD bound to blunt-end 5′-PPP 12-mer ds-RNA. (b) Interactions between the 5′-PPP and the CTD. Phosphates are yellow/red. The Phe853 side chain, which stacks with the bases of the terminal base pair, is highlighted.
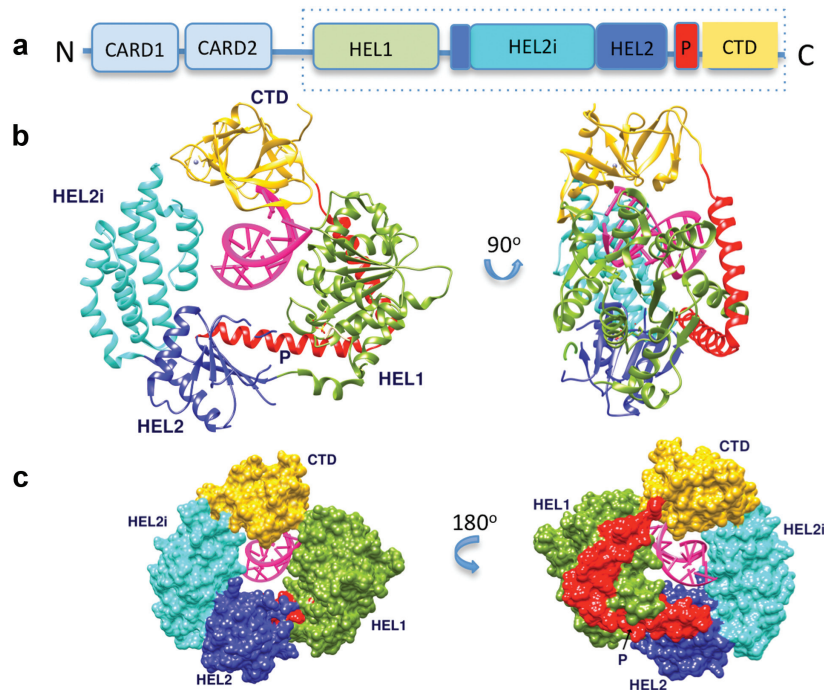
Fig. 41 (a) Domain organization of the RIG-I protein. The boxed region (ΔCARDs) was crystallized. (b) Overall structure of the RIG-I dsRNA complex. The domain coloring is that shown in A. C, Solvent accessible protein surface of the complex.

relevant since the RNA stimulates efficient ATP hydrolysis by RIG-I. The CTD now encapsulates the 5′-PPP at the duplex terminus through a network of electrostatic interactions involving K861, K888, K858, and H847 (Fig. 42b). The α and β phosphates at the 5′ terminus are particularly critical for RIG-I binding, which together with the favorable π–π stacking of F853 side chain onto the terminal base pair, helps RIG-I to select the correct pathogenic RNA from the vast pool of capped cellular RNAs. The γ-phosphate does not form direct contacts with the protein, suggesting that it is not a major recognition determinant. The position of the bound nucleotide (ADP) was identified within conserved ATPase motifs in HEL1. HEL2 is not involved in RNA or ADP binding, however, the HEL2i domain again makes an extensive contact interface with the RNA duplex. The UUCG tetra-loop at the other hairpin terminus is absorbed into an RNA-binding funnel and does not establish any base-specific contacts with RIG-I. The phosphates of ADP are bound to RIG-I through a bridging Mg^{2+}.

Fig. 42 (a) Ternary complex of RIG-I (ΔCARDs), 5′-PPP hairpin-RNA and ADP-Mg^{2+}. (b) Interactions between the 5′-PPP and the CTD. Phosphates are yellow/red. In the hairpin RNA, G1 base pairs with C20.

A comparison of available RIG-I:nucleotide structures indicates that the helicase domain is in an open conformation in the absence of RNA substrate [235–237]. In the presence of RNA and the ATP analogue ADP-AlF$_3$, the helicase domain adopts a closed conformation (Fig. 4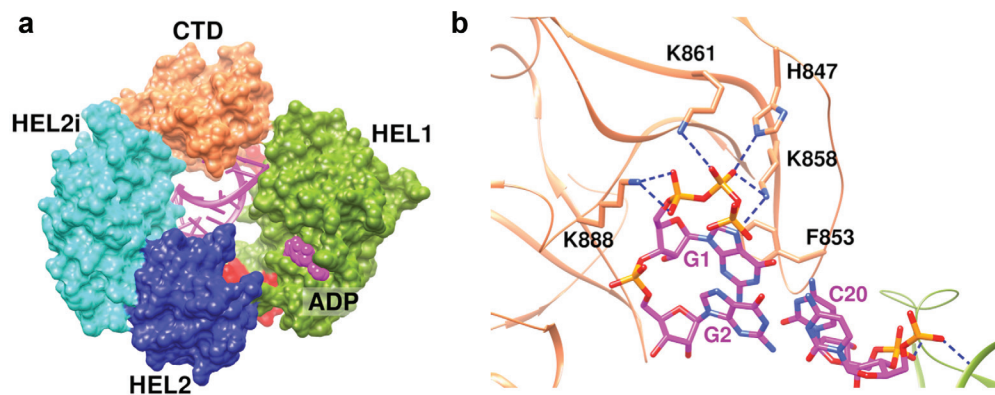3). In the complex with ADP-Mg^{2+}, the product of ATP hydrolysis, RIG-I adopts an intermediate semi-closed state. The RIG-I conformation thus appears to be very sensitive to ATP binding, hydrolysis, and product release. This leads to a model for RIG-I signaling, where after RNA binding, ATP hydrolysis moves the CTD and HEL2i domains in opposite directions, which likely allows the CARDs to be released from HEL2i into a position where they can signal to downstream molecules (Fig. 43) [238].

5.2 RNA recognition by MDA5

MDA5 is a second cytoplasmic viral RNA receptor, which shares high sequence similarity to RIG-I, and interacts with a common signaling adapter MAVS, which activates interferon signaling pathways [239]. However, these sensors recognize largely distinct groups of viruses and viral RNAs. MDA5 detects long-duplex RNAs in the genome of dsRNA viruses or dsRNA replication intermediates of positive strand viruses. Nevertheless, MDA5 and RIG-I have similar domain architectures, with N-terminal CARD domains that interact with MAVS, a central helicase domain for RNA-dependent ATP hydrolysis, and a C-terminal domain (CTD) (Fig. 4). However, the CTD of MDA5 displays affinity for dsRNA that is orders of magnitude lower than that of full-length MDA5, suggesting different mechanisms of dsRNA recognition by RIG-I and MDA5 [240].

MDA5 forms a higher-order structure during RNA recognition and signaling, by cooperatively assembling into a filamentous oligomer composed of a repeating segmental arrangement of MDA5 dimers along the length of dsRNA [240]. Binding of MDA5 to dsRNA stimulates ATP hydrolysis, which stimulates assembly and disassociation of MDA5 from dsRNA in a manner inversely proportional to the filament length. Thus, ATP
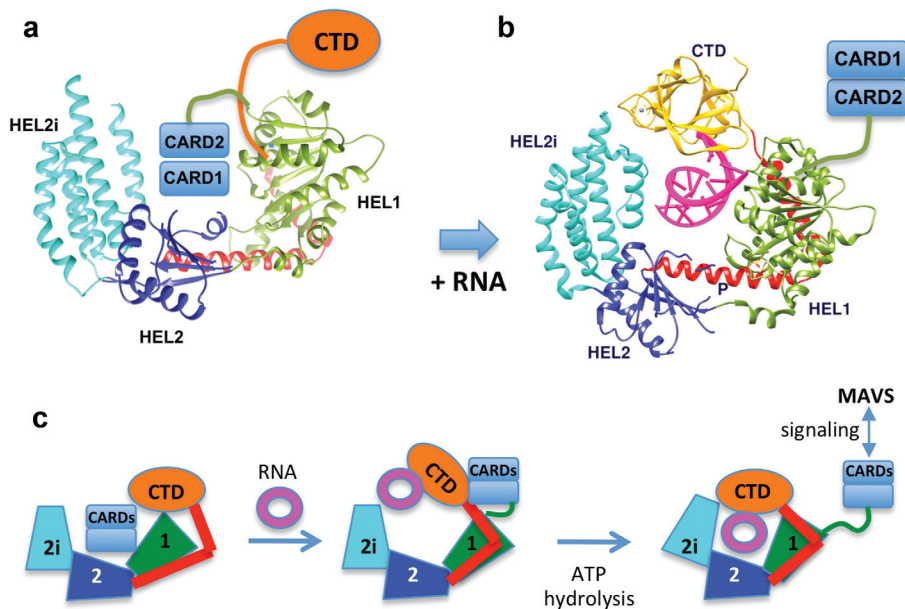
Fig. 43 (a) Structure of duck RIG-I-Hel (lacking CARDs and CTD domains) in an open conformation (PDB 4A2P), in which the HEL1 and HEL2 domains are not in direct contact [236]. In full-length RIG-I, the CARDs would be linked to the N-terminus of HEL1 and the CTD to the C-terminus of HEL2 (see Fig. 41a). The bridging or pincer domain comprising two long α-helices connected by an ≈80° bend is shown in red. (b) Structure of RIG-I lacking only the CARDs with dsRNA (PDB 2YKG) [237]. The RIG-I protein completely surrounds the dsRNA, which is bound in a central cavity. (c) A model for RIG-I activation (adapted from [236] and [237]). In the absence of RNA, RIG-I is in an autoinhibited open state. 5′-PPP-RNA is first captured by the CTD. Upon binding and/or hydrolysis of ATP the activated closed form is generated, competent for signaling through the CARDs.

hydrolysis has a key role in the control of filament assembly and disassembly, thereby autoregulating the interaction of MDA5 with dsRNA, and likely providing a basis for dsRNA length-dependent antiviral signaling.

Recently, a crystal structure was reported of human MDA5 lacking the CARD domains (residues 298–1018), an internal loop (residues 645–662) and the C-terminal tail (residues 1018–1025) [241]. These deletions do not affect RNA-binding affinity or ATP hydrolysis activity of the MDA5, but afforded crystals of the protein (MDA5ΔN′) complexed to a 12 bp dsRNA and an ATP analog [5′-adenyl-β-imidodiphosphate (ADPNP)].

The structure of the ternary complex reveals that MDA5ΔN′ forms a ring-like structure around the 12 bp dsRNA (Fig. 44). MDA5 makes contact with the dsRNA using the HEL1, Hel2i, Hel2, and CTD domains, but not the pincer (P) domain. Most protein residues interact primarily with the RNA phosphate backbone and 2′-hydroxyl groups, consistent with sequence-independent recognition of dsRNA. Mutational studies suggest that all domains make important contributions to high-affinity dsRNA recognition [241]. Similar interactions are found upon RNA binding to RIG-I, with the exception of interactions involving the CTD. The bound ADPNP is located in a cleft between HEL1 and HEL2, in the outer rim of the MDA5 ring structure. A structurally similar binding site for the nucleotide is found in RIG-I.

A major difference between MDA5 and RIG-I is seen in the orientation of the CTD [241]. Whereas in RIG-I the CTD encapsulates the 5′-PPP end of the RNA, in the complex with MDA5 the CTD (residues 900–1017) is oriented differently so that it makes contact to the side of the dsRNA (Fig. 45). This difference in CTD orientation results in an open, C-ring-shaped structure of MDA5 that binds to the stem of dsRNA, as opposed to the O-ring structure formed by RIG-I, which caps the dsRNA end. On the other hand, the conformation of the helicase domain is well preserved between RIG-I and MDA5 (Cα RMSD 1.9 Å). In both complexes with RNA, the CTDs form an extensive interaction with HEL2i, with buried surface areas of ≈510 Å^2. The convex surface of the Zn-binding site in the CTD is complementary to the concave face of HEL2i. Interestingly, a surface loop in HEL2 (residues 758–767) of MDA5 inserts into the major groove of the bound dsRNA stem, whereas in RIG-I the equivalent loop binds to the end of the RNA (Fig. 45). In the MDA5-dsRNA complex, the major groove is widened from 1.2 to 1.8 nm at the site of the HEL2 loop insertion, but the interaction between the inserted loop and dsRNA is limited to the phosphate backbone and does not lead to sequence-dependent base recognition. A ΔHEL2-loop mutant, however, still bound strongly to dsRNA, but lacked the dsRNA-dependent ATP hydrolysis activity. Thus, insertion of the HEL2 loop into the major groove seems to be important for ATP hydrolysis and subsequent signaling. Overall, close interaction between the helicase domain and the CTD provides the molecular basis for the divergent RNA specificity of MDA5 and RIG-I.

EM studies have revealed that MDA5 rings bind and stack along dsRNA to form filamentous oligomers [240, 242, 243]. Structural models of the filaments have been built using both EM and crystallographic results.
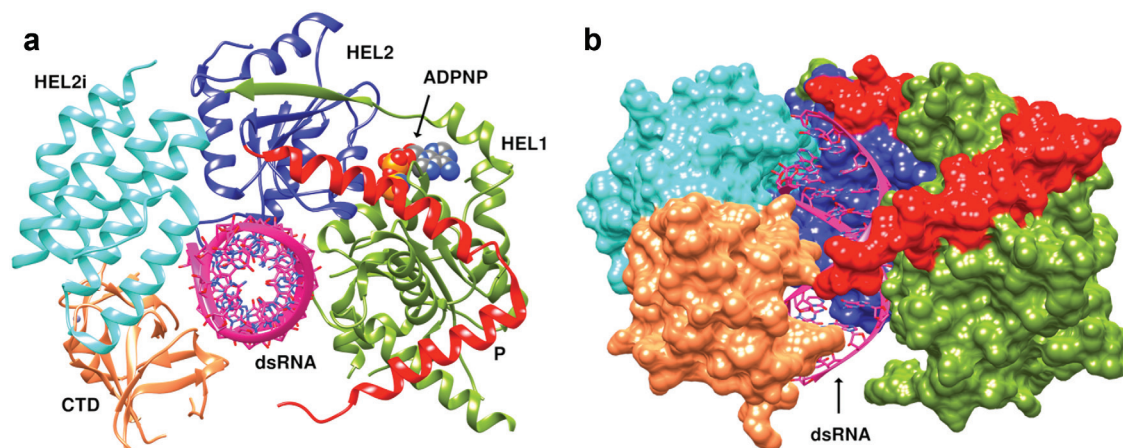

Fig. 44 Structure of the MDA5ΔN′ protein bound to dsRNA (PDB 4GL2) [241]; left, top view; right, side view. The domains are colored as in Fig. 4. The MDA5 forms a ring around dsRNA, bound in the central cavity. The position of bound nucleotide analogue ADPNP is indicated.
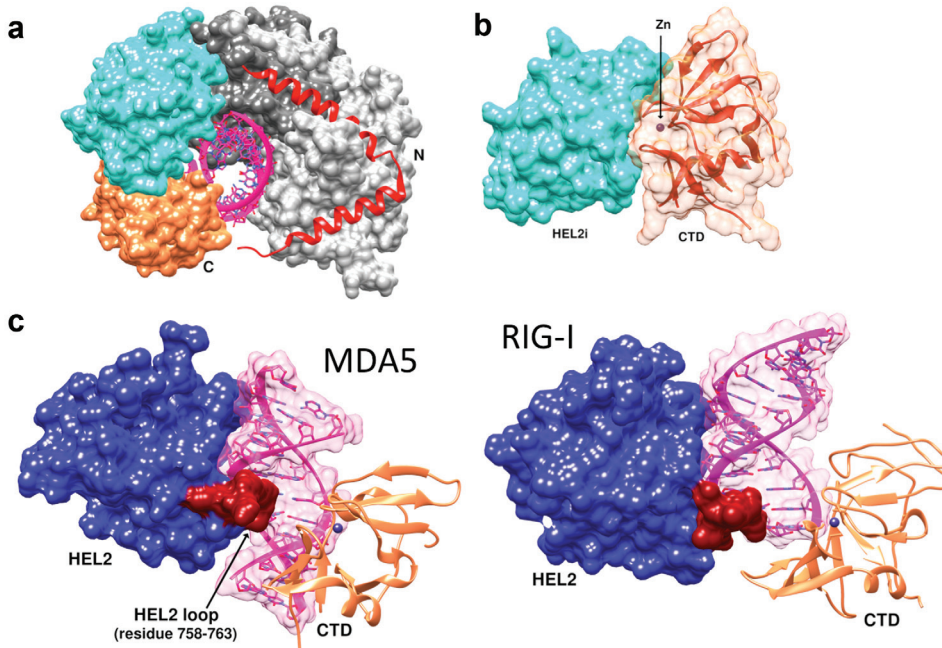
Fig. 45 (a) Top view of the MDA5ΔN′-dsRNA complex showing the extensive interactions between HEL2i and CTD (HEL1 and HEL2 are gray, pincer domain red). (b) The HEL2i-CTD interface with bound Zn²⁺. (c) Left, the HEL2 loop (red) in MDA5ΔN′ is inserted into the major groove of dsRNA, which widens from 1.2 to 1.8 nm. Right, the RIG-I loop binds to the end of the 5′-PPP-RNA.

High-resolution structural models predict a head-to-tail association of MDA5 subunits in the filaments, a repeating unit of ≈14 bp, and a turn of ≈70° between each monomer (Fig. 46) [241]. This arrangement leads to a monomer–monomer interface comprising largely the helicase domain, with a total buried surface area of ≈1500 Å². Defining the complete architecture of full-length MDA5, however, requires knowledge of the arrangement of the two CARDs within the filament. Biochemical studies have shown that isolated 2CARDs can self-oligomerize at high protein concentration. This oligomerization of the CARDS is required to activate the downstream signaling molecule MAVS. Favored models of full-length MDA5 within filaments bring the 2CARD from each monomer together to form discrete filaments or patches of oligomers along the central MAD5 filament (Fig. 46) [241]. The size, frequency, and positioning of individual patches are stochastically determined during MDA5 filament assembly and ATP hydrolysis. Thus, MDA5 binding to dsRNA, filament formation, and additional conformational rearrangements during ATP hydrolysis, appear to be required for efficient oligomerization of the CARDs and interaction with MAVS. ATP hydrolysis does not simply trigger assembly or disassembly of the filament, but is required to allow 2CARD oligomerization and binding to MAVS.

6 Note added in proof

A recent study describes the crystal structure of a fragment derived from mouse NLRC4 [244]. The NLRC4 inflammasome is activated in mice by bacterial flagellin or the components of type 3 secretion system. In another study, the crystal structure was reported of a RIG-I:dsRNA complex. A minimal RNA duplex was identified that binds one RIG-I molecule, stimulates robust ATPase activity, and elicits a RIG-I-mediated interferon (IFN) response in cells [245]. Other studies establish a clear role for the ligand-induced ATPase activity of RIG-I in the stimulation of an IFN response [246]. Finally, new studies suggest that RNA:DNA hybrids arising during viral replication are a novel molecular pattern sensed by TLR9 [253], and that TLR10 in humans is involved in induction of innate immune responses to influenza infection [254].
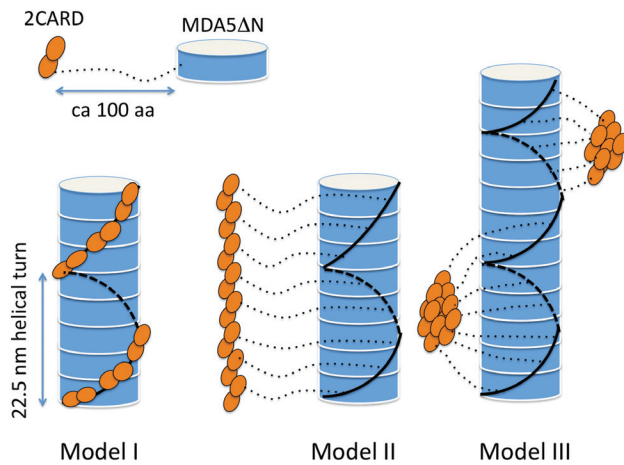
Fig. 46 Oligomerization of full-length MDA5 on long dsRNA. Three models for the oligomers (I, II, and III) that are consistent with experimental data. In the first two models, 2CARDs form a filament along or around the central MAD5 filament. Alternatively, 2CARDs may form discrete patches along the central MAD5 filament. Adapted from [241].

Table 1 Structures of pattern recognition receptors and complexes with PAMPs available in the protein database (PDB).

PRR	Source	Resol. (×10 nm)	PDB code	Reference
TLR1/2	human TLR1-TLR2 + Pam$_3$CSK$_4$ complex	2.10	2Z7X	[74]
	mouse TLR2 dimer with Pam$_3$CSK$_4$ or Pam$_2$CSK$_4$	1.80	2Z81	[74]
		2.60	2Z82	
TLR2	mouse TLR2 with PE-DPTA or pn-LTA	2.4	3A76	[75]
		2.5	3A7B	
TLR2/6	mouse TLR2-TLR6 + Pam$_2$CSK$_4$	2.9	3A79	[75]
CD14	mouse CD14	2.5	1WWL	[128]
	human CD14	4.0	4GLP	[127]
TLR3	human TLR3	2.4	2A0Z	[78]
		2.1	1ZIW	[77]
	mouse TLR3 alone and with dsRNA	2.66	3CIG	[80]
		3.41	3CIY	
MD2	human MD-2 aloneand with lipid IVa	2.00	2E56	[143]
		2.21	2E59	
TLR4/MD-2	mouse TLR4 with MD-2 and bound endotoxin	2.84	2Z64	[102]
	human TLR4 with MD-2 and bound LPS	3.10	3FXI	[144]
	mouse TLR4/MD-2 with lipid-IVa	2.70	3VQ1	[148]
	with LPS	2.48	3VQ2	
FliC	*Salmonella* sp. flagellin	2.00	1IO1	[149]
TLR5	zebra fish TLR5	2.83	3V44	[83]
	TLR5 with flagellin	2.47	3V47	
TLR8	human TLR8, unliganded	2.30	3W3G	[160]
	hTLR8 with CL097	2.00	3W3J	[160]
	hTLR8 with CL075	2.30	3W3K	
	hTLR8 with R848,	2.33	3W3L	
	3 forms	2.70	3W3M	
		2.10	3W3N	
DC-SIGN	human DC-SIGN and DC-SIGNR CRD with bound sugar 5mer	2.50	1K9I	[188]
DC-SIGNR		1.90	1K9J	
	DC-SIGN + Man$_4$	1.55	1SL4	[189]
	DC-SIGN + LNFP-III	1.80	1SL5	
	DC-SIGN + Lewis^x	2.25	1SL6	
	DC-SIGN + Man$_6$	2.40	2IT5	[190]
	DC-SIGN + Man$_2$	1.95	2IT6	
	DC-SIGNR neck region	2.25	1XAR	[185]

(**Table 1** Continued)

PRR	Source	Resol. (×10 nm)	PDB code	Reference
		1.41	1XPH	[186]
	DC-SIGNR neck region	2.20	3JQH	[187]
Langerin	apo-human Langerin	2.50	3P7F	[193]
	Langerin + disaccharide	2.30	3P7H	
	Langerin + mannose	1.50	3P7G	
Mannose-binding	rat MBL CRD + neck	1.80	1RTM	[199]
proteins (lectins)	hMBL CRD + neck	2.50	1HUP	[198]
	rat MBP CRD	2.30	1MSB	[248]
	rat MBP-A + mannose	1.70	2MSB	[200]
	rat MBP-A CRD	1.80	1YTT	[249]
	rat MBP-C + sugars	1.70	1RDO	[201]
		1.80	1RDN	
		1.90	1RDM	
		1.70	1RDL	
		1.80	1RDK	
		1.80	1RDJ	
		1.80	1RDI	
	MBP-A + galactose	2.00	1AFA	[204]
		1.90	1AFB	
		2.00	1AFD	
	MBP-A + Lewis^x oligos	2.10	1KMB	[250]
		2.00	2KMB	
		1.95	3KMB	
		2.00	4KMB	
	rat MBP-C	1.85	1BV4	[251, 252]
		1.90	1BUU	
	rat MBP-C + GalNAc	2.10	1BCJ	[203]
		2.00	1BCH	
	rat MBP-A	1.95	1KWT	[202]
	+αMe-O-Man	1.95	1KWU	
	+αMe-O-GlcNAc	2.00	1KWV	
	+αMe-O-Fuc	1.90	1KWW	
	+βMe-O-Fuc	2.00	1KWX	
	+Manα1-3Man	2.00	1KWY	
	H189V + Manα1-3Man	1.90	1KWZ	
	H189V/I207V + Manα1-3Man	2.00	1KX0	
	+Man$_6$GlcNAc$_2$Asn	2.80	1KX1	
	rat MBP-C + Manα1-3Man	1.74	1KXA	
	+ Manα1-3[Manα1-6]Man	1.80	1KXB	
	+ Manα1-2Manα1-6Man	1.85	1KXC	
	+ GlcNAcβ1-2Manα1-3[GlcNAcβ1-2Manα1-6]-Man	1.90	1KZD	
	+ N-Ac-Tyr-Asp-[Gly-Gly-NH-(CH$_2$)$_6$-O-Man]$_2$	1.80	1KZE	
Mannose receptor	human MR CTLD4	2.30	1EGG	[205]
		2.30	1EGI	
	mouse MR Cys-rich domain	1.70	1DQG	[206]
	+GalNAc-4-SO$_4$	2.20	1DQO	
	+3-SO$_4$-Lewis^a	2.20	1FWV	[207]
	+3-SO$_4$-Lewis^x	1.90	1FWU	
NODs	human NRLX1	2.65	3UN9	[59]
RLRs	human RIG-I CTD	3.00	2QFB	[230]
			2QFD	
		NMR	2RMJ	[231]
	+5′-ppp dsRNA	2.55	3NCU	[232]
		2.60	3LRN	[233]
		2.15	3LRR	
	hRIG-I (ΔCARD) + dsRNA	2.50	2YKG	[237]
	+dsRNA + ADP	2.58	4BPB	
	ATPase domain of RIG-I	2.80	4AY3	[238]

1532 — J.A. Robinson and K. Moehle: Molecular recognition in the immune system. Part II

DE GRUYTER

(**Table 1** Continued)

PRR	Source	Resol. (×10 nm)	PDB code	Reference
	+AMP-PNP	2.14	TBK2	[234]
	duck RIG-I CTD	1.44	4A2V	
	dRIG-I helicase	3.00	4A2P	[236]
	dRIG-I CARDs + helicase	3.40	4A2Q	
	dRIG-I full length	3.70	4A2W	
	dRIG-I CTD + dsRNA	4.00	4A2X	
	dRIG-I + dsRNA + ADP-AlF$_3$	3.70	4A36	
	hRIG-I + dsRNA + BeF$_3$	2.90	3TMI	[235]
MDA5	human MDA5 + dsRNA	3.56	4GL2	[241]

7 Abbreviations

APC	antigen-presenting cell
BCR	B-cell receptor
BIR	baculovirus inhibitor repeat
CARD	caspase activation and recruitment domain
CIITA	class II transactivator
CRD	carbohydrate recognition domain
CTD	C-terminal domain
CTL	C-type lectin
CTLD	C-type lectin (or lectin-like) domain
CTLR	C-type lectin receptors
DC	dendritic cell
DC-SIGN(R)	dendritic cell-specific intercellular adhesion molecule-grabbing nonintegrin (and the R variant)
DTPA	1,2-dimyristoyl-sn-glycero-3-phospho-ethanolamine-N-diethylenetriaminepentaacetic acid
ECD	extracellular domain
FliC	flagellin
FNII	fibronectin type II
GPI	glycosyl-phosphatidyl-inositol
IFN	interferon
IRF	interferon-regulated factor
HEL1	helicase domain 1
LBP	LPS binding protein
LGP2	laboratory of genetics and physiology 2
LNFP-III	lacto-N-fucopentaose-III
LPS	lipopolysaccharide
LRR	leucine-rich repeat
LTA	lipoteichoic acid
MBP	mannose-binding protein
MHC-I and -II	major histocompatibility complex-I and -II
MDA5	melanoma differentiation associated factor gene 5
MPL	monophosphoryl lipid A
MR	mannose receptor
MLS	macrolide-lincosamide-streptogramin group
NACHT	a conserved domain found in NAIP, CIITA, HET-E, and TPI-1
NAIP	NLR family apoptosis inhibitory protein

NK	natural killer
NLR	NOD-like receptor
NOD	nucleotide-binding and oligomerization domain
ODN	oligodeoxynucleotide
OM	outer membrane
PAMP	pathogen-associated molecular pattern
PE	phosphatidyl-ethanolamine
PGN	peptidoglycan
PPI	protein–protein interface
PRR	pattern recognition receptor
PYD	pyrin domain
RIG	retinoic acid-inducible gene-I
RLR	RIG-like receptor
TIR	Toll IL-1 receptor
TLR	Toll-like receptor
VLR	variable lymphocyte receptor

8 Membership of sponsoring body

Membership of the IUPAC Chemistry and Human Health Division Committee for the period 2014–2015 is as follows:

President**:** T. J. Perun (USA); Vice President**:** R. Cornelis (Belgium); Secretary**:** M. Schwenk (Germany); Titular Members**:** E. Differding (Belgium); J. Fischer (Hungary); V. Gubala (Slovakia); P. Illing (UK); L. Johnston (Canada); H. Møller Johannessen (Denmark); W. A. Temple (New Zealand); Associate Members**:** Vincenzo Abbate (UK); M. Kiilunen (Finland); Y. C. Martin (USA); S. Mignani (France); D. Rotella (USA); National Representatives**:** N. Nahar (Bangladesh); M.-X. Wang (China/Beijing); R. Jih-Ru Hwu (China/Taipei); S. Alihodžić (Croatia); A. Rahatgoanker (India); G. Bee Teh (Malaysia); R. Leurs (Netherlands); B. Haug (Norway); S. Bachurin (Russia); P. Ploypradith (Thailand).

References

[1] J. M. Blander, L. E. Sander. *Nat. Rev. Immunol.* **12**, 215 (2012).
[2] H. Kumar, T. Kawai, S. Akira. *Int. Rev. Immunol.* **30**, 16 (2011).
[3] O. Takeuchi, S. Akira. *Cell* **140**, 805 (2010).
[4] H. Kumar, T. Kawai, S. Akira. *Biochem. J.* **420**, 1 (2009).
[5] A. Iwasaki, R. Medzhitov. *Science* **327**, 291 (2010).
[6] R. E. W. Hancock, A. Nijnik, D. J. Philpott. *Nat. Rev. Microbiol.* **10**, 243 (2012).
[7] P. J. Delves, S. J. Martin, D. R. Burton, I. M. Roitt. *Roitt's Essential Immunology*, 12th ed., Wiley-Blackwell, Hoboken (2011).
[8] M. F. Bachmann, U. H. Rohrer, T. M. Kundig, K. Burki, H. Hengartner, R. M. Zinkernagel. *Science* **262**, 1448 (1993).
[9] N. E. Harwood, F. D. Batista. *Annu. Rev. Immunol.* **28**, 185 (2010).
[10] Z. Hua, B. Hou. *Cell. Mol. Immunol.* **10**, 103 (2013).
[11] S. Akira. *Phil. Trans. R. Soc. B* **366**, 2748 (2011).
[12] C. Olive. *Exp. Rev. Vaccines* **11**, 237 (2012).
[13] B. Guy. *Nat. Rev. Microbiol.* **5**, 505 (2007).
[14] E. M. Pålsson-McDermott, L. A. J. O'Neill. *Biochem. Soc. Trans.* **35**, 1437 (2007).
[15] P. Broz, D. M. Monack. *Nat. Rev. Immunol.* **13**, 551 (2013).
[16] N. J. Gay, M. Gangloff. *Annu. Rev. Biochem.* **76**, 141 (2007).
[17] B. Beutler. *Blood* **113**, 1399 (2009).

[18] A. N. R. Weber, S. Tauszig-Delamasure, J. A. Hoffmann, E. Lelievre, H. Gascan, K. P. Ray, M. A. Morse, J.-L. Imler, N. J. Gay. *Nat. Immunol.* **4**, 794 (2003).

[19] B. Lemaitre, E. Nicolas, L. Michaut, J.-M. Reichhart, J. A. Hoffmann. *Cell* **86**, 973 (1996).

[20] L. A. J. O'Neill, D. Golenbock, A. G. Bowie. *Nat. Rev. Immunol.* **13**, 453 (2013).

[21] M. Sasai, M. Yamamoto. *Int. Rev. Immunol.* **32**, 116 (2013).

[22] Y. Xu, X. Tao, B. Shen, T. Horng, R. Medzhitov, J. L. Manley, L. Tong. *Nature* **408**, 111 (2000).

[23] T. Kawai, S. Akira. *Nat. Immunol.* **11**, 373 (2010).

[24] D. H. Song, J.-O. Lee. *Immunol. Rev.* **250**, 216 (2012).

[25] A. N. Zelensky, J. E. Gready. *FEBS J.* **272**, 6179 (2005).

[26] H.-J. Gabius. *Biochim. Biophys. Acta* **1572**, 163 (2002).

[27] J. J. García-Vallejo, Y. Van Kooyk. *Immunol. Revs.* **230**, 22 (2009).

[28] W. W. J. Unger, Y. van Kooyk. *Curr. Opin. Immunol.* **23**, 131 (2011).

[29] T. B. H. Geijtenbeek, S. I. Gringhuis. *Nat. Rev. Immunol.* **9**, 465 (2009).

[30] E. P. McGreal, J. L. Miller, S. Gordon. *Curr. Opin. Immunol.* **17**, 18 (2005).

[31] M. J. Robinson, D. Sancho, E. C. Slack, S. LeibundGut-Landmann, C. Reis e. Sousa. *Nat. Immunol.* **7**, 1258 (2006).

[32] S. I. Gringhuis, J. den Dunnen, M. Litjens, M. van der Vlist, T. B. H. Geijtenbeek. *Nat. Immunol.* **10**, 1081 (2009).

[33] T. B. H. Geijtenbeek, S. J. van Vliet, A. Engering, B. A. 't Hart, Y. van Kooyk. *Annu. Rev. Immunol.* **22**, 33 (2004).

[34] L. East, C. M. Isacke. *Biochim. Biophys. Acta* **1572**, 364 (2002).

[35] W. K. Eddie, K. Ip, R. Takahashi, L. M. Alan Ezekowitz, Stuart. *Immunol. Revs.* **230**, 9 (2009).

[36] U. Gazi, L. Martinez-Pomares. *Immunobiology* **214**, 554 (2009).

[37] D. J. Philpott, M. T. Sorbara, S. J. Robertson, K. Croitoru, S. E. Girardin. *Nat. Rev. Immunol.* **14**, 9 (2014).

[38] J. H. Fritz, R. L. Ferrero, D. J. Philpott, S. E. Girardin. *Nat. Immunol.* **7**, 1250 (2006).

[39] K. Geddes, J. G. Magalhás, S. E. Girardin. *Nat. Rev. Drug Discov.* **8**, 465 (2009).

[40] N. Inohara, M. Chamaillard, C. McDonald, G. Nuñez. *Ann. Rev. Biochem.* **74**, 355 (2005).

[41] N. Inohara, G. Nunez. *Nat. Rev. Immunol.* **3**, 371 (2003).

[42] S. E. Girardin, I. G. Boneca, J. Viala, M. Chamaillard, A. Labigne, G. Thomas, D. J. Philpott, P. J. Sansonetti. *J. Biol. Chem.* **278**, 8869 (2003).

[43] N. Inohara, Y. Ogura, A. Fontalba, O. Gutierrez, F. Pons, J. Crespo, K. Fukase, S. Inamura, S. Kusumoto, M. Hashimoto, S. J. Foster, A. P. Moran, J. L. Fernandez-Luna, G. Nuñez. *J. Biol. Chem.* **278**, 5509 (2003).

[44] S. E. Girardin, I. G. Boneca, L. A. M. Carneiro, A. Antignac, M. Jéhanno, J. Viala, K. Tedin, M.-K. Taha, A. Labigne, U. Zäthringer, A. J. Coyle, P. S. DiStefano, J. Bertin, P. J. Sansonetti, D. J. Philpott. *Science* **300**, 1584 (2003).

[45] M. Chamaillard, M. Hashimoto, Y. Horie, J. Masumoto, S. Qiu, L. Saab, Y. Ogura, A. Kawasaki, K. Fukase, S. Kusumoto, M. A. Valvano, S. J. Foster, T. W. Mak, G. Nunez, N. Inohara. *Nat. Immunol.* **4**, 702 (2003).

[46] Y. Fujimoto, A. R. Pradipta, N. Inohara, K. Fukase. *Nat. Prod. Rep.* **29**, 568 (2012).

[47] F. Di Virgilio. *Pharmacol. Rev.* **65**, 872 (2013).

[48] F. Martinon, A. Mayor, J. Tschopp. *Annu. Rev. Immunol.* **27**, 229 (2009).

[49] K. W. Chen, K. Schroder. *Curr. Opin. Microbiol.* **16**, 311 (2013).

[50] T. P. Monie. *Trends Biochem. Sci.* **38**, 131 (2013).

[51] J. K. Krishnaswamy, T. Chu, S. C. Eisenbarth. *Trends Immunol.* **34**, 224 (2013).

[52] K. Kersse, M. J. M. Bertrand, M. Lamkanfi, P. Vandenabeele. *Cytokine Growth Factor Rev.* **22**, 257 (2011).

[53] J. P. Y. Ting, B. K. Davis. *Annu. Rev. Immunol.* **23**, 387 (2005).

[54] H. Laroui, Y. Yan, Y. Narui, S. A. Ingersoll, S. Ayyadurai, M. A. Charania, F. Zhou, B. Wang, K. Salaita, S. V. Sitaraman, D. Merlin. *J. Biol. Chem.* **286**, 31003 (2011).

[55] C. L. Grimes, L. D. Z. Ariyananda, J. E. Melnyk, E. K. O'Shea. *J. Am. Chem. Soc.* **134**, 13535 (2012).

[56] J. Mo, J. P. Boyle, C. B. Howard, T. P. Monie, B. K. Davis, J. A. Duncan. *J. Biol. Chem.* **287**, 23057 (2012).

[57] E. F. Halff, C. A. Diebolder, M. Versteeg, A. Schouten, T. H. C. Brondijk, E. G. Huizinga. *J. Biol. Chem.* **287**, 38460 (2012).

[58] L. M. Delbridge, M. X. D. O'Riordan. *Curr. Opin. Immunol.* **19**, 10 (2007).

[59] M. Hong, S.-I. Yoon, I. A. Wilson. *Immunity* **36**, 337 (2012).

[60] J. D. Desmet, K. J. Ishil. *Nat. Rev. Immunol.* **12**, 479 (2012).

[61] Y.-M. Loo, M. Gale Jr. *Immunity* **34**, 680 (2011).

[62] M. Schlee, A. Roth, V. Hornung, C. A. Hagmann, V. Wimmenauer, W. Barchet, C. Coch, M. Janke, A. Mihailovic, G. Wardle, S. Juranek, H. Kato, T. Kawai, H. Poeck, K. A. Fitzgerald, O. Takeuchi, S. Akira, T. Tuschl, E. Latz, J. Ludwig, G. Hartmann. *Immunity* **31**, 25 (2009).

[63] M. Schlee, E. Hartmann, C. Coch, V. Wimmenauer, M. Janke, W. Barchet, G. Hartmann. *Immunol. Rev.* **227**, 66 (2009).

[64] E. Dixit, J. C. Kagan. *Adv. Immunol.* **117**, 99 (2013).

[65] M. S. Jin, J.-O. Lee. *Immunity* **29**, 182 (2008).

[66] J. Y. Kang, J.-O. Lee. *Annu. Rev. Biochem.* **80**, 917 (2011).

[67] J. Bella, K. L. Hindle, P. A. McEwan, S. C. Lovell. *Cell. Mol. Life Sci.* **65**, 2307 (2008).

[68] N. Matsushima, N. Tachi, Y. Kuroki, P. Enkhbayar, M. Osaki, M. Kamiya, R. H. Kretsinger. *Cell. Mol. Life Sci.* **62**, 2771 (2005).

[69] B. Kobe, A. V. Kajava. *Curr. Opin. Struct. Biol.* **11**, 725 (2001).

[70] J. K. Bell, G. E. D. Mullen, C. A. Leifer, A. Mazzoni, D. R. Davies, D. M. Segal. *Trends Immunol.* **24**, 528 (2003).

[71] N. Matsushima, T. Tanaka, P. Enkhbayar, T. Mikami, M. Taga, K. Yamada, Y. Kuroki. *BMC Genomics* **8**, (2007).

[72] T. Mikami, H. Miyashita, S. Takatsuka, Y. Kuroki, N. Matsushima. *Gene* **503**, 235 (2012).

[73] N. Matsushima, T. Mikami, T. Tanaka, H. Miyashita, K. Yamada, Y. Kuroki. *Biochim. Biophys. Acta.* **1790**, 1217 (2009).

[74] M. S. Jin, S. E. Kim, J. Y. Heo, M. E. Lee, H. M. Kim, S.-G. Paik, H. Lee, J.-O. Lee. *Cell* **130**, 1071 (2007).

[75] J. Y. Kang, X. Nan, M. S. Jin, S.-J. Youn, Y. H. Ryu, S. Mah, S. H. Han, H. Lee, S.-G. Paik, J.-O. Lee. *Immunity* **31**, 873 (2009).

[76] B. Beutler, Z. Jiang, P. Georgel, K. Crozat, B. Croker, S. Rutschmann, X. Du, K. Hoebe. *Annu. Rev. Immunol.* **24**, 353 (2006).

[77] J. Choe, M. S. Kelker, I. A. Wilson. *Science* **309**, 581 (2005).

[78] J. K. Bell, I. Botos, P. R. Hall, J. Askins, J. Shiloach, D. M. Segal, D. R. Davies. *Proc. Natl. Acad. Sci. USA* **102**, 10976 (2005).

[79] J. N. Leonard, R. Ghirlando, J. Askins, J. K. Bell, D. H. Margulies, D. R. Davies, D. M. Segal. *Proc. Natl. Acad. Sci. USA* **105**, 258 (2008).

[80] L. Liu, I. Botos, Y. Wang, J. N. Leonard, J. Shiloach, D. M. Segal, D. R. Davies. *Science* **320**, 379 (2008).

[81] J. Luo, G. Obmolova, T. J. Malia, S.-J. Wu, K. E. Duffy, J. D. Marion, J. K. Bell, P. Ge, Z. H. Zhou, A. Teplyakov, Y. Zhao, R. J. Lamb, J. L. Jordan, L. R. San Mateo, R. W. Sweet, G. L. Gilliland. *J. Mol. Biol.* **421**, 112 (2012).

[82] E. Latz, A. Verma, A. Visintin, M. Gong, C. M. Sirois, D. C. G. Klein, B. G. Monks, C. J. McKnight, M. S. Lamphier, W. P. Duprex, T. Espevik, D. T. Golenbock. *Nat. Immunol.* **8**, 772 (2007).

[83] S.-I. Yoon, O. Kurnasov, V. Natarajan, M. Hong, A. V. Gudkov, A. L. Osterman, I. A. Wilson. *Science* **335**, 859 (2012).

[84] R. Shimazu, S. Akashi, H. Ogata, Y. Nagai, K. Fukudome, K. Miyake, M. Kimoto. *J. Exp. Med.* **189**, 1777 (1999).

[85] S. Viriyakosol, P. S. Tobias, R. L. Kitchens, T. N. Kirkland. *J. Biol. Chem.* **276**, 38044 (2001).

[86] Y. Nagai, S. Akashi, M. Nagafuku, M. Ogata, Y. Iwakura, S. Akira, T. Kitamura, A. Kosugi, M. Kimoto, K. Miyake. *Nat. Immunol.* **3**, 667 (2002).

[87] D. Werling, O. C. Jann, V. Offord, E. J. Glass, T. J. Coffey. *Trends Immunol.* **30**, 124 (2009).

[88] U. Zähringer, B. Lindner, S. Inamura, H. Heine, C. Alexander. *Immunobiology* **213**, 205 (2008).

[89] S.-Y. Seong, P. Matzinger. *Nat. Rev. Immunol.* **4**, 469 (2004).

[90] H. Nakayama, K. Kurokawa, B. L. Lee. *FEBS J.* **279**, 4247 (2012).

[91] M. I. Hutchings, T. Palmer, D. J. Harrington, I. C. Sutcliffe. *Trends Microbiol.* **17**, 13 (2009).

[92] A. Tschumi, C. Nai, Y. Auchli, P. Hunziker, P. Gehrig, P. Keller, T. Grau, P. Sander. *J. Biol. Chem.* **284**, 27146 (2009).

[93] M. T. Madigan, J. M. Martinko, D. A. Stahl, D. P. Clark. *Brock Biology of Microorganisms*, 13th ed., Addison-Wesley Longman, Amsterdam (2011).

[94] A. Ray, M. Cot, G. Puzo, M. Gilleron, J. Nigou. *Biochimie* **95**, 33 (2013).

[95] P. F. Mühlradt, M. Kieß, H. Meyer, R. Süßmuth, G. Jung. *J. Exp. Med.* **185**, 1951 (1997).

[96] K.-I. Shibata, A. Hasebe, T. Into, M. Yamada, T. Watanabe. *J. Immunol.* **165**, 6538 (2000).

[97] U. Buwitt-Beckmann, H. Heine, K.-H. Wiesmüller, G. Jung, R. Brock, S. Akira, A. J. Ulmer. *Eur. J. Immunol.* **35**, 282 (2005).

[98] U. Buwitt-Beckmann, H. Heine, K.-H. Wiesmüller, G. Jung, R. Brock, S. Akira, A. J. Ulmer. *J. Biol. Chem.* **281**, 9049 (2006).

[99] R. Spohn, U. Buwitt-Beckmann, R. Brock, G. Jung, A. J. Ulmer, K.-H. Wiesmüller. *Vaccine* **22**, 2494 (2004).

[100] S. Akira, K. Takeda. *Nat. Rev. Immunol.* **4**, 499 (2004).

[101] M. Triantafilou, F. G. J. Gamper, R. M. Haston, M. A. Mouratis, S. Morath, T. Hartung, K. Triantafilou. *J. Biol. Chem.* **281**, 31002 (2006).

[102] H. M. Kim, B. S. Park, J.-I. Kim, S. E. Kim, J. Lee, S. C. Oh, P. Enkhbayar, N. Matsushima, H. Lee, O. J. Yoo, J.-O. Lee. *Cell* **130**, 906 (2007).

[103] U. Buwitt-Beckmann, H. Heine, K.-H. Wiesmüller, G. Jung, R. Brock, A. J. Ulmer. *FEBS J.* **272**, 6354 (2005).

[104] A. Grabiec, G. Meng, S. Fichte, W. Bessler, H. Wagner, C. J. Kirschning. *J. Biol. Chem.* **279**, 48004 (2004).

[105] O. Takeuchi, A. Kaufmann, K. Grote, T. Kawai, K. Hoshino, M. Morr, P. F. Mühlradt, S. Akira. *J. Immunol.* **164**, 554 (2000).

[106] T. Okusawa, M. Fujita, J.-I. Nakamura, T. Into, M. Yasuda, A. Yoshimura, Y. Hara, A. Hasebe, D. T. Golenbock, M. Morita, Y. Kuroki, T. Ogawa, K.-I. Shibata. *Infect. Immun.* **72**, 1657 (2004).

[107] S. Basith, B. Manavalan, G. Lee, S. G. Kim, S. Choi. *Exp. Opin. Therapeut. Patents* **21**, 927 (2011).

[108] M. Czarniecki. *J. Med. Chem.* **51**, 6621 (2008).

[109] M. Hedayat, M. G. Netea, N. Rezaei. *Lancet* **11**, 702 (2011).

[110] M. Hedayat, K. Takeda, N. Rezaei. *Med. Res. Rev.* **32**, 294 (2012).

[111] E. J. Hennessy, A. E. Parker, L. A. J. O'Neill. *Nat. Rev. Drug Discov.* **9**, 293 (2010).

[112] H. Kanzler, F. J. Barrat, E. M. Hessel, R. L. Coffman. *Nat. Med.* **13**, 552 (2007).

[113] A. Kuznik, G. Panter, R. Jerala. *Curr. Med. Chem.* **17**, 1899 (2010).

[114] F. Romagne. *Drug Disc. Today* **12**, 80 (2006).

[115] E. Vacchelli, L. Galluzzi, A. Eggermont, W. H. Fridman, J. Galon, C. Sautés-Fridman, E. Tartour, L. Zitvogel, G. Kroemer. *Oncoimmunology* **6**, 894 (2012).

[116] F. Steinhagen, T. Kinjo, C. Bode, D. M. Klinman. *Vaccine* **29**, 3341 (2011).

[117] J. Schmidt, T. Welsch, D. Jager, P. F. Muhlradt, M. W. Buchler, A. Marten. *Br. J. Cancer* **97**, 598 (2007).

[118] D. J. Connolly, L. A. J. O'Neill. *Curr. Opin. Pharmacol.* **12**, 510 (2012).

[119] K. Miyake. *J. Endotoxin Res.* **12**, 195 (2006).

[120] K. Hoebe, P. Georgel, S. Rutschmann, X. Du, S. Mudd, K. Crozat, S. Sovath, L. Shamel, T. Hartung, U. Zahringer, B. Beutler. *Nature* **433**, 523 (2005).

[121] S. Gordon. *Cell* **111**, 927 (2002).

[122] S. D. Wright, R. A. Ramos, P. S. Tobias, R. J. Ulevitch, J. C. Mathison. *Science* **249**, 1431 (1990).

[123] R. S. Jack. *CD14 in the Inflammatory Response*, Vol. 74, Karger, Basel, (1999).

[124] T. Nakata, M. Yasuda, M. Fujita, H. Kataoka, K. Kiura, H. Sano, K. Shibata. *Cell. Microbiol.* **8**, 1899 (2006).

[125] M. Manukyan, K. Triantafilou, M. Triantafilou, A. Mackie, N. Nilsen, T. Espevik, K.-H. Wiesmüller, A. J. Ulmer, H. Heine. *Eur. J. Immunol.* **35**, 911 (2005).

[126] D. R. E. Ranoa, S. L. Kelley, R. I. Tapping. *J. Biol. Chem.* **288**, 9729 (2013).

[127] S. L. Kelley, T. Lukk, S. K. Nair, R. I. Tapping. *J. Immunol.* **190**, 1304 (2013).

[128] J.-I. Kim, C. J. Lee, M. S. Jin, C.-H. Lee, S.-G. Paik, H. Lee, J.-O. Lee. *J. Biol. Chem.* **280**, 11347 (2005).

[129] M. D. Cunningham, R. A. Shapiro, C. Seachord, K. Ratcliffe, L. Cassiano, R. P. Darveau. *J. Immunol.* **164**, 3255 (2000).

[130] S.-Y. Zhang, M. Herman, M. J. Ciancanelli, R. Pérez de Diego, V. Sancho-Shimizu, L. Abel, J.-L. Casanova. *Curr. Opin. Immunol.* **25**, 19 (2013).

[131] J. Hyun, S. Kanagavelu, M. Fukata. *Microb. Infect.* **15**, 1 (2013).

[132] M. E. Kleinman, K. Yamada, A. Takeda, V. Chandrasekaran, M. Nozaki, J. Z. Baffi, R. J. C. Albuquerque, S. Yamasaki, M. Itaya, Y. Pan, B. Appukuttan, D. Gibbs, Z. Yang, K. Kariko, B. K. Ambati, T. A. Wilgus, L. A. DiPietro, E. Sakurai, K. Zhang, J. R. Smith, E. W. Taylor, J. Ambati. *Nature* **452**, 591 (2008).

[133] S. Okahira, F. Nishikawa, S. Nishikawa, T. Akazawa, T. Seya, M. Matsumoto. *DNA Cell Biol.* **24**, 614 (2005).

[134] A. K. Field, A. A. Tytell, G. P. Lampson, M. R. Hilleman. *Proc. Natl. Acad. Sci. USA* **58**, 1004 (1967).

[135] B. B. Gowen, M.-H. Wong, K.-H. Jung, A. B. Sanders, W. M. Mitchell, L. Alexopoulou, R. A. Flavell, R. W. Sidwell. *J. Immunol.* **178**, 5200 (2007).

[136] B. D. Needham, M. S. Trent. *Nat. Rev. Microbiol.* **11**, 467 (2013).

[137] C. R. H. Raetz, C. M. Reynolds, M. S. Trent, R. E. Bishop. *Annu. Rev. Biochem.* **76**, 295 (2007).

[138] A. Teghanemt, D. Zhang, E. N. Levis, J. P. Weiss, T. L. Gioannini. *J. Immunol.* **175**, 4669 (2005).

[139] S. I. Miller, R. K. Ernst, M. W. Bader. *Nat. Rev. Microbiol.* **3**, 36 (2005).

[140] C. R. H. Raetz, C. Whitfield. *Annu. Rev. Biochem.* **71**, 635 (2002).

[141] C. Erridge, E. Bennett-Guerrero, I. R. Poxton. *Microbes Infect.* **4**, 837 (2002).

[142] T. K. Means, D. T. Golenbock, M. J. Fenton. *Cytokine Growth Factor Rev.* **11**, 219 (2000).

[143] U. Ohto, K. Fukase, K. Miyake, Y. Satow. *Science* **316**, 1632 (2007).

[144] B. S. Park, D. H. Song, H. M. Kim, B.-S. Choi, H. Lee, J.-O. Lee. *Nature* **458**, 1191 (2009).

[145] E. T. Rietschel, T. Kirikae, F. U. Schade, U. Mamat, G. Schmidt, H. Loppnow, A. J. Ulmer, U. Zähringer, U. Seydel, F. Di Padova. *FASEB J.* **8**, 217 (1994).

[146] V. Mata-Haro, C. Cekic, M. Martin, P. M. Chilton, C. R. Casella, T. C. Mitchell. *Science* **316**, 1628 (2007).

[147] C. R. Casella, T. C. Mitchell. *Cell. Mol. Life Sci.* **65**, 3231 (2008).

[148] U. Ohto, K. Fukase, K. Miyake, T. Shimizu. *Proc. Natl. Acad. Sci. USA* **109**, 7421 (2012).

[149] F. A. Samatey, K. Imada, S. Nagashima, F. Vonderviszt, T. Kumasaka, M. Yamamoto, K. Namba. *Nature* **410**, 331 (2001).

[150] F. Hayashi, K. D. Smith, A. Ozinsky, T. R. Hawn, E. C. Yi, D. R. Goodlett, J. K. Eng, S. Akira, D. M. Underhill, A. Aderem. *Nature* **410**, 1099 (2001).

[151] L. G. Burdelya, V. I. Krivokrysenko, T. C. Tallant, E. Strom, A. S. Gleiberman, D. Gupta, O. V. Kurnasov, F. L. Fort, A. L. Osterman, J. A. DiDonato, E. Feinstein, A. V. Gudkov. *Science* **320**, 226 (2008).

[152] C. Cuadros, F. J. Lopez-Hernandez, A. L. Dominguez, M. McClelland, J. Lustgarten. *Infect. Immun.* **72**, 2810 (2004).

[153] J. J. Treanor, D. N. Taylor, L. Tussey, C. Hay, C. Nolan, T. Fitzgerald, G. Liu, U. Kavita, L. Song, I. Dark, A. Shaw. *Vaccine* **28**, 8268 (2010).

[154] D. N. Taylor, J. J. Treanor, E. A. Sheldon, C. Johnson, S. Umlauf, L. Song, U. Kavita, G. Liu, L. Tussey, K. Ozer, T. Hofstaetter, A. Shaw. *Vaccine* **30**, 5761 (2012).

[155] T. K. Means, F. Hayashi, K. D. Smith, A. Aderem, A. D. Luster. *J. Immunol.* **170**, 5165 (2003).

[156] A. M. Krieg, J. Vollmer. *Immunol. Rev.* **220**, 251 (2007).

[157] S. E. Ewald, A. Engel, J. Lee, M. Wang, M. Bogyo, G. M. Barton. *J. Exp. Med.* **208**, 643 (2011).

[158] S. E. Ewald, B. L. Lee, L. Lau, K. E. Wickliffe, G.-P. Shi, H. A. Chapman, G. M. Barton. *Nature* **456**, 658 (2008).

[159] F. E. Sepulveda, S. Maschalidi, R. Colisson, L. Heslop, C. Ghirelli, E. Sakka, A.-M. Lennon-Duménil, S. Amigorena, L. Cabanie, B. Manoury. *Immunity* **31**, 737 (2009).

[160] H. Tanji, U. Ohto, T. Shibata, K. Miyake, T. Shimizu. *Science* **339**, 1426 (2013).

[161] S. S. Diebold, T. Kaisho, H. Hemmi, S. Akira, C. Reis e Sousa. *Science* **303**, 1529 (2004).

[162] F. Heil, H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, S. Bauer. *Science* **303**, 1526 (2004).

[163] H. Hemmi, T. Kaisho, O. Takeuchi, S. Sato, H. Sanjo, K. Hoshino, T. Horiuchi, H. Tomizawa, K. Takeda, S. Akira. *Nat. Immunol.* **3**, 196 (2002).

[164] M. Jurk, F. Heil, J. Vollmer, C. Schetter, A. M. Krieg, H. Wagner, G. Lipford, S. Bauer. *Nat. Immunol.* **3**, 499 (2002).

[165] F. Heil, P. Ahmad-Nejad, H. Hemmi, H. Hochrein, F. Ampenberger, T. Gellert, H. Dietrich, G. Lipford, K. Takeda, S. Akira, H. Wagner, S. Bauer. *Eur. J. Immunol.* **33**, 2987 (2003).

[166] J. Lee, T.-H. Chuang, V. Redecke, L. She, P. M. Pitha, D. A. Carson, E. Raz, H. B. Cottam. *Proc. Natl. Acad. Sci. USA* **100**, 6646 (2003).

[167] H. Hemmi, O. Takeuchi, T. Kawai, T. Kaisho, S. Sato, H. Sanjo, M. Matsumoto, K. Hoshino, H. Wagner, K. Takeda, S. Akira. *Nature* **408**, 740 (2000).

[168] O. Joffre, M. A. Nolte, R. Spörri, C. Reis e. Sousa. *Immunol. Rev.* **227**, 234 (2009).

[169] G. Hartmann, A. M. Krieg. *J. Immunol.* **164**, 944 (2000).

[170] A. M. Krieg. *Nat. Rev. Drug Discov.* **5**, 471 (2006).

[171] C. Bode, G. Zhao, F. Steinhagen, T. Kinjo, D. M. Klinman. *Exp. Rev. Vaccines* **10**, 499 (2011).

[172] M. H. Lahoud, F. Ahmet, J. G. Zhang, S. Meuter, A. N. Policheni, S. Kitsoulis, C. N. Lee, M. O'Keeffe, L. C. Sullivan, A. G. Brooks, R. Berry, J. Rossjohn, J. D. Mintern, J. Vega-Ramos, J. A. Villadangos, N. A. Nicola, M. C. Nussenzweig, K. J. Stacey, K. Shortman, W. R. Heath, I. Caminschi. *Proc. Natl. Acad. Sci. USA* **109**, 16270 (2012).

[173] U. Hasan, C. Chaffois, C. Gaillard, V. Saulnier, E. Merck, S. Tancredi, C. Guiet, F. Brière, J. Vlach, S. Lebecque, G. Trinchieri, E. E. M. Bates. *J. Immunol.* **174**, 2942 (2005).

[174] T. Nyman, P. Stenmark, S. Flodin, I. Johansson, M. Hammarström, P. Nordlund. *J. Biol. Chem.* **283**, 11861 (2008).

[175] W. A. Andrade, M. do C. Souza, E. Ramos-Martinez, K. Nagpal, M. S. Dutra, M. B. Melo, D. C. Bartholomeu, S. Ghosh, D. T. Golenbock, R. T. Gazzinelli. *Cell Host Microbe* **13**, 42 (2013).

[176] R. Pifer, A. Benson, C. R. Sturge, F. Yarovinsky. *J. Biol. Chem.* **286**, 3307 (2011).

[177] A. Hidmark, A. von Saint Paul, A. H. Dalpke. *J. Immunol.* **189**, 2717 (2012).

[178] M. Oldenburg, A. Krüger, R. Ferstl, A. Kaufmann, G. Nees, A. Sigmund, B. Bathke, H. Lauterbach, M. Suter, S. Dreher, U. Koedel, S. Akira, T. Kawai, J. Buer, H. Wagner, S. Bauer, H. Hochrein, C. J. Kirschning. *Science* **337**, 1111 (2012).

[179] G. Simmons, J. D. Reeves, C. C. Grogan, L. H. Vandenberghe, F. Baribaud, J. C. Whitbeck, E. Burke, M. J. Buchmeier, E. J. Soilleux, J. L. Riley, R. W. Doms, P. Bates, S. Pöhlmann. *Virology* **305**, 115 (2003).

[180] L. Tailleux, O. Schwartz, J.-L. Herrmann, E. Pivert, M. Jackson, A. Amara, L. Legres, D. Dreher, L. P. Nicod, J. C. Gluckman, P. H. Lagrange, B. Gicquel, O. Neyrolles. *J. Exp. Med.* **197**, 121 (2003).

[181] B. Tassaneetrithep, T. H. Burgess, A. Granelli-Piperno, C. Trumpfheller, J. Finke, W. Sun, M. A. Eller, K. Pattanapanyasat, S. Sarasombath, D. L. Birx, R. M. Steinman, S. Schlesinger, M. A. Marovich. *J. Exp. Med.* **197**, 823 (2003).

[182] Y. van Kooyk, T. B. H. Geijtenbeek. *Nat. Rev. Immunol.* **3**, 697 (2003).

[183] U.-S. Khoo, K. K. Chan, V. F. Chan, C. L. S. Lin. *J. Mol. Med.* **86**, 861 (2008).

[184] S. Menon, K. Rosenberg, S. A. Graham, E. M. Ward, M. E. Taylor, K. Drickamer, D. E. Leckband. *Proc. Natl. Acad. Sci. USA* **106**, 11524 (2009).

[185] H. Feinberg, Y. Guo, D. A. Mitchell, K. Drickamer, W. I. Weis. *J. Biol. Chem.* **280**, 1327 (2005).

[186] G. A. Snyder, M. Colonna, P. D. Sun. *J. Mol. Biol.* **347**, 979 (2005).

[187] H. Feinberg, C. K. W. Tso, M. E. Taylor, K. Drickamer, W. I. Weis. *J. Mol. Biol.* **394**, 613 (2009).

[188] H. Feinberg, D. A. Mitchell, K. Drickamer, W. I. Weis. *Science* **294**, 2163 (2001).

[189] Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis, K. Drickamer. *Nat. Struct. Mol. Biol.* **11**, 591 (2004).

[190] H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger, W. I. Weis. *J. Biol. Chem.* **282**, 4202 (2007).

[191] D. A. Mitchell, A. J. Fadden, K. Drickamer. *J. Biol. Chem.* **276**, 28939 (2001).

[192] N. S. Stambach, M. E. Taylor. *Glycobiology* **13**, 401 (2003).

[193] L. Chatwell, A. Holla, B. B. Kaufer, A. Skerra. *Mol. Immunol.* **45**, 1981 (2008).

[194] J. A. Willment, G. D. Brown. *Trends Microbiol.* **16**, 27 (2007).

[195] P. Allavena, M. Chieppa, P. Monti, L. Piemonti. *Crit. Rev. Immunol.* **24**, 179 (2004).

[196] J. H. Lu, S. Thiel, H. Wiedemann, R. Timpl, K. B. Reid. *J. Immunol.* **144**, 2287 (1990).

[197] R. T. Lee, Y. Ichikawa, T. Kawasaki, K. Drickamer, Y. C. Lee. *Arch. Biochem. Biophys.* **299**, 129 (1992).

[198] S. Sheriff, C. Y. Chang, R. A. B. Ezekowitz. *Nat. Struct. Mol. Biol.* **1**, 789 (1994).

[199] W. I. Weis, K. Drickamer. *Structure* **2**, 1227 (1994).

[200] W. I. Weis, K. Drickamer, W. A. Hendrickson. *Nature* **360**, 127 (1992).

[201] K. K. S. Ng, K. Drickamer, W. I. Weis. *J. Biol. Chem.* **271**, 663 (1996).

[202] K. K. S. Ng, A. R. Kolatkar, S. Park-Snyder, H. Feinberg, D. A. Clark, K. Drickamer, W. I. Weis. *J. Biol. Chem.* **277**, 16088 (2002).

[203] A. R. Kolatkar, A. K. Leung, R. Isecke, R. Brossmer, K. Drickamer, W. I. Weis. *J. Biol. Chem.* **273**, 19502 (1998).

[204] A. R. Kolatkar, W. I. Weis. *J. Biol. Chem.* **271**, 6679 (1996).

[205] H. Feinberg, S. Park-Snyder, A. R. Kolatkar, C. T. Heise, M. E. Taylor, W. I. Weis. *J. Biol. Chem.* **275**, 21539 (2000).

[206] Y. Liu, A. J. Chirino, Z. Misulovin, C. Leteux, T. Feizi, M. C. Nussenzweig, P. J. Bjorkman. *J. Exp. Med.* **191**, 1105 (2000).

[207] Y. Liu, Z. Misulovin, P. J. Bjorkman. *J. Mol. Biol.* **305**, 481 (2001).

[208] C. B. Moore, D. T. Bergstralh, J. A. Duncan, Y. Lei, T. E. Morrison, A. G. Zimmermann, M. A. Accavitti-Loper, V. J. Madden, L. Sun, Z. Ye, J. D. Lich, M. T. Heise, Z. Chen, J. P. Y. Ting. *Nature* **451**, 573 (2008).

[209] I. C. Allen, C. B. Moore, M. Schneider, Y. Lei, B. K. Davis, M. A. Scull, D. Gris, K. E. Roney, A. G. Zimmermann, J. B. Bowzard, P. Ranjan, K. M. Monroe, R. J. Pickles, S. Sambhara, J. P. Y. Ting. *Immunity* **34**, 854 (2011).

[210] X. Xia, J. Cui, H. Y. Wang, L. Zhu, S. Matsueda, Q. Wang, X. Yang, J. Hong, Z. Songyang, Z. J. Chen, R.-F. Wang. *Immunity* **34**, 843 (2011).

[211] J. Y. Bae, H. H. Park. *J. Biol. Chem.* **286**, 39528 (2011).

[212] S. Hiller, A. Kohl, F. Fiorito, T. Herrmann, G. Wider, J. Tschopp, M. G. Grütter, K. Wüthrich. *Structure* **11**, 1199 (2003).

[213] E. Liepinsh, R. Barbals, E. Dahl, A. Sharipo, E. Staub, G. Otting. *J. Mol. Biol.* **332**, 1155 (2003).

[214] E. de Alba. *J. Biol. Chem.* **284**, 32932 (2009).

[215] A. S. Pinheiro, C. Eibl, Z. Ekman-Vural, R. Schwarzenbacher, W. Peti. *J. Mol. Biol.* **413**, 790 (2011).

[216] A. S. Pinheiro, M. Proell, C. Eibl, R. Page, R. Schwarzenbacher, W. Peti. *J. Biol. Chem.* **285**, 27402 (2010).

[217] A. Natarajan, R. Ghose, J. M. Hill. *J. Biol. Chem.* **281**, 31863 (2006).

[218] C. Eibl, S. Grigoriu, M. Hessenberger, J. Wenger, S. Puehringer, A. S. Pinheiro, R. N. Wagner, M. Proell, J. C. Reed, R. Page, K. Diederichs, W. Peti. *Biochemistry* **51**, 7330 (2012).

[219] F. Ferrage, K. Dutta, E. Nistal-Villán, J. R. Patel, M. T. Sánchez-Aparicio, P. De Ioannes, A. Buku, G. G. Aseguinolaza, A. García-Sastre, A. K. Aggarwal. *Structure* **20**, 2048 (2012).

[220] F. Manon, A. Favier, G. Núñez, J.-P. Simorre, S. Cusack. *J. Mol. Biol.* **365**, 160 (2007).

[221] T. Jin, J. Curry, P. Smith, J. Jiang, T. S. Xiao. *Proteins* **81**, 1266 (2013).

[222] N. P. Coussens, J. C. Mowers, C. McDonald, G. Nuñez, S. Ramaswamy. *Biochem. Biophys. Res. Commun.* **353**, 1 (2007).

[223] M. R. Singleton, M. S. Dillingham, D. B. Wigley. *Annu. Rev. Biochem.* **76**, 23 (2007).

[224] M. Yoneyama, M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira, T. Fujita. *Nat. Immunol.* **5**, 730 (2004).

[225] Q.-X. Jiang, Z. J. Chen. *EMBO Rep.* **13**, 7 (2012).

[226] H. J. Ramos, M. Gale Jr. *Curr. Opin. Virol.* **1**, 167 (2011).

[227] M. Yoneyama, T. Fujita. *Immunol. Rev.* **227**, 54 (2009).

[228] D. W. Leung, G. K. Amarasinghe. *Curr. Opin. Struct. Biol.* **22**, 297 (2012).

[229] A. Vela, O. Fedorova, S. C. Ding, A. M. Pyle. *J. Biol. Chem.* **287**, 42564 (2012).

[230] S. Cui, K. Eisenächer, A. Kirchhofer, K. Brzózka, A. Lammens, K. Lammens, T. Fujita, K.-K. Conzelmann, A. Krug, K.-P. Hopfner. *Mol. Cell* **29**, 169 (2008).

[231] K. Takahasi, M. Yoneyama, T. Nishihori, R. Hirai, H. Kumeta, R. Narita, M. Gale Jr., F. Inagaki, T. Fujita. *Mol. Cell* **29**, 428 (2008).

[232] Y. Wang, J. Ludwig, C. Schuberth, M. Goldeck, M. Schlee, H. Li, S. Juranek, G. Sheng, R. Micura, T. Tuschl, G. Hartmann, D. J. Patel. *Nat. Struct. Mol. Biol.* **17**, 781 (2010).

[233] C. Lu, H. Xu, C. T. Ranjith-Kumar, M. T. Brooks, T. Y. Hou, F. Hu, A. B. Herr, R. K. Strong, C. C. Kao, P. Li. *Structure* **18**, 1032 (2010).

[234] F. Civril, M. Bennett, M. Moldt, T. Deimling, G. Witte, S. Schiesser, T. Carell, K.-P. Hopfner. *EMBO Rep.* **12**, 1127 (2011).

[235] F. Jiang, A. Ramanathan, M. T. Miller, G.-Q. Tang, M. Gale, S. S. Patel, J. Marcotrigiano. *Nature* **479**, 423 (2011).

[236] E. Kowalinski, T. Lunardi, A. A. McCarthy, J. Louber, J. Brunel, B. Grigorov, D. Gerlier, S. Cusack. *Cell* **147**, 423 (2011).

[237] D. Luo, S. C. Ding, A. Vela, A. Kohlway, B. D. Lindenbach, A. M. Pyle. *Cell* **147**, 409 (2011).

[238] D. Luo, A. Kohlway, A. Vela, A. M. Pyle. *Structure* **20**, 1983 (2012).

[239] H. Kato, K. Takahasi, T. Fujita. *Immunol. Rev.* **243**, 91 (2011).

[240] A. Peisley, C. Lin, B. Wu, M. Orme-Johnson, M. Liu, T. Walz, S. Hur. *Proc. Natl. Acad. Sci. USA* **108**, 21010 (2011).

[241] B. Wu, A. Peisley, C. Richards, H. Yao, X. Zeng, C. Lin, F. Chu, T. Walz, S. Hur. *Cell* **152**, 276 (2013).

[242] I. C. Berke, Y. Modis. *EMBO J.* **31**, 1714 (2012).

[243] I. C. Berke, X. Yu, Y. Modis, E. H. Egelman. *Proc. Natl. Acad. Sci. USA* **109**, 18437 (2012).

[244] Z. Hu, C. Yan, P. Liu, Z. Huang, R. Ma, C. Zhang, R. Wang, Y. Zhang, F. Martinon, D. Miao, H. Deng, J. Wang, J. Chang, J. Chai. *Science* **341**, 172 (2013).

[245] A. Kohlway, D. H. Luo, D. C. Rawling, S. C. Ding, A. M. Pyle. *EMBO Rep.* **14**, 772 (2013).

[246] J. R. Patel, A. Jain, Y. Y. Chou, A. Baum, T. Ha, A. Garcia-Sastre. *EMBO Rep.* **14**, 780 (2013).

[247] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin. *J. Comp. Chem.* **25**, 1605 (2004).

[248] W. I. Weis, R. Kahn, R. Fourme, K. Drickamer, W. A. Hendrickson. *Science* **254**, 1608 (1991).

[249] F. T. Burling, W. I. Weis, K. M. Flaherty, A. T. Brünger. *Science* **271**, 72 (1996).

[250] K. K. S. Ng, W. I. Weis. *Biochemistry* **36**, 979 (1997).

[251] K. K. S. Ng, S. Park-Snyder, W. I. Weis. *Biochemistry* **37**, 17965 (1998).

[252] K. K. S. Ng, W. I. Weis. *Biochemistry* **37**, 17977 (1998).

[253] R. E. Rigby, L. M. Webb, K. J. Mackenzie, Y. Li, A. Leitch, M. A. M. Reijns, R. J. Lundie, A. Revuelta, D. J. Davidson, S. Diebold, Y. Modis A. S. MacDonad, A. P. Jackson. *EMBO J.* **33**, 542 (2014).

[254] S. M. Y. Lee, K.-H. Kok, M. Jaume, T. K. W. Cheung, T.-F. Yip, J. C. C. Lai, Y. Guan, R. G. Webster, D.-Y. Jin, J. S. Malik Peiris. *Proc. Natl. Acad. Sci. USA* **111**, 3793 (2014).