

Supporting Information (SI) Appendix

Structure of the Toll-Spätzle complex, a molecular hub in *Drosophila* development and innate immunity

Christoph Parthier, Marco Stelter, Christian Ursel, Uwe Fandrich, Hauke Lilie, Constanze Breithaupt and Milton T. Stubbs

Institut für Biochemie und Biotechnologie and Mitteldeutsches Zentrum für Struktur und Dynamik von Proteinen, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany

Table of Contents	page
Figure S1. Structure based sequence alignment of the LRR repeats of <i>DmToll</i>	2
Figure S2. Glycosylation sites of <i>DmToll</i>	3
Table S1. Toll (potential) glycosylation sites	4
Figure S3. Multiple sequence alignment of Toll1 orthologs	5-8
Figure S4. Multiple sequence alignment of Spätzle 1 orthologs	9
Figure S5. Crystal packing reveals a 1:1 Toll-Spätzle complex	10
Figure S6. Purification of matured hcSpätzle	11
Table S2. Binding parameters for the interaction of Toll ECD with Spätzle C106 and hcSpätzle monitored by isothermal calorimetry (ITC)	12
Figure S7. Sedimentation velocity analyses of Toll ECD and its complex with C106	13
Figure S8. Conformational differences in the Spätzle protomers	14
Figure S9. Mapping of <i>toll</i> gain of function mutations (7, 8)	15
Figure S10. Comparison of the Toll-Spätzle complex to other LRR-containing receptors with protein ligands	16
Figure S11. Models of Toll dimerization based on the cryoEM reconstruction of Gangloff <i>et al.</i> (2008) (13)	17
Figure S12. Models for the initiation of Toll signaling by Spätzle	18
SI Methods	19
• Expression, complex formation and crystallization	
• Data collection, phase determination, model building and refinement	
• Analytical ultracentrifugation	
• Preparation of Spätzle* and hcSpätzle	
• Isothermal calorimetry (ITC)	
SI Results: Detailed description of the Toll ECD	23
• LRRNT1	
• N-terminal LRR-domain	
• LRRCT1	
• Membrane-proximal LRR domain	
Table S3. MAD datasets of Sm ³⁺ -soaked Toll:Spätzle complex	26
Table S4. datasets of native complexes and refinement	27
SI References	28

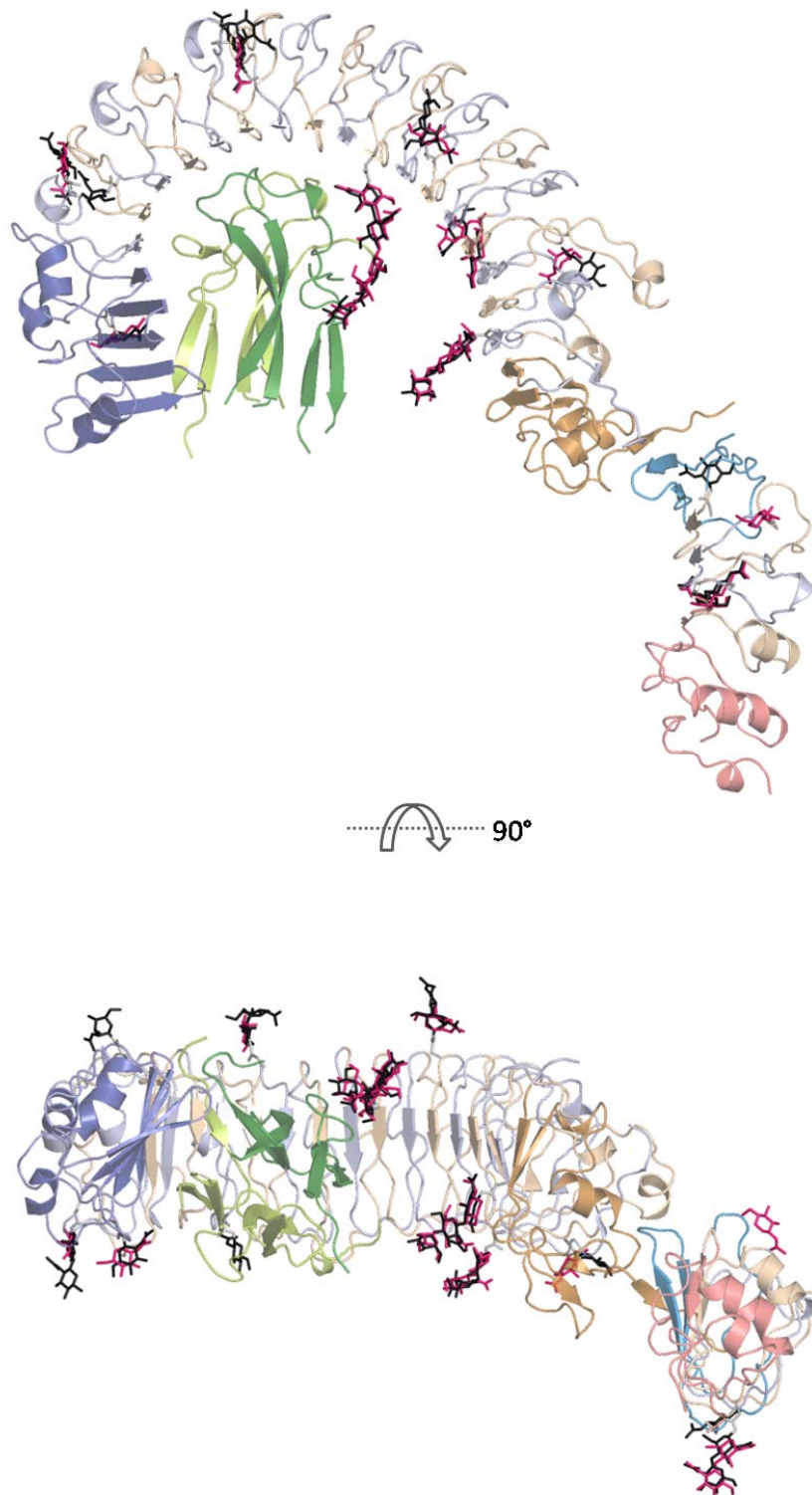
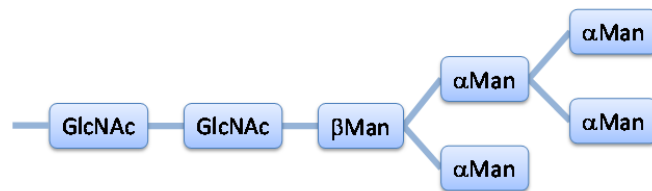


Figure S2. Glycosylation sites of *DmToll*. Overall structure of the (partially) deglycosylated complex of Toll and Spätzle showing the sites of glycosylation (pink sticks, see also **Table S1**) observed in the electron density, superimposed with the glycan residues from the non-deglycosylated complex (black sticks). Orientation and colors as in **Figure 1**.

Table S1: Toll (potential) glycosylation sites



residue	position in LRR	Endo Hf deglycosylated / fully glycosylated
80	ascending	GlcNAc
140	ascending	GlcNAc-GlcNAc
175	descending	GlcNAc
235	descending	GlcNAc
270	descending	GlcNAc-GlcNAc
275	concave	none
346	concave	GlcNAc-GlcNAc-Man ₃
391	descending	GlcNAc-GlcNAc
482	ascending	GlcNAc ₂
508	ascending	GlcNAc
528	concave	GlcNAc ₂ -Man
654	ascending	GlcNAc
677	ascending	none visible
703	ascending	GlcNAc-GlcNAc
715	descending	GlcNAc
730	ascending	none visible
738	descending	none visible

Figure S3. Multiple sequence alignment of Toll1 orthologs. Sequences were retrieved from the NCBI data base using BLASTP (2), accepting sequences from organisms in which orthologs of both Toll and Spätzle were present. These were aligned using CLUSTAL-W (3) followed by manual adjustments based on the structure using Jalview (1), which was also used for visualization and conservation scoring. Residues are colored from red (fully conserved) via blue to green (least conserved). Below the alignment, glycosylation sites and residues of *DmToll* involved in Spätzle binding are marked. In addition, the secondary structural elements and the positions of LRR repeats and capping structures are shown for *DmToll*. Accession numbers are as follows: NP_524518.1 (*Drosophila melanogaster*), XP_001981637.1 (*Drosophila erecta*), XP_002098877.1 (*Drosophila yakuba*), XP_001357966.1 (*Drosophila pseudoobscura pseudoobscura*), XP_004524168.1 (*Ceratitis capitata*), XP_001658507.1 (*Aedes aegypti*), XP_309197.1 (*Anopheles gambiae* str. PEST), XP_396158.1 (*Apis mellifera*), XP_002431445.1 (*Pediculus humanus corporis*), EFX86424.1 (*Daphnia pulex*), XP_002399580.1 (*Ixodes scapularis*), XP_004925675.1 (*Bombyx mori*).

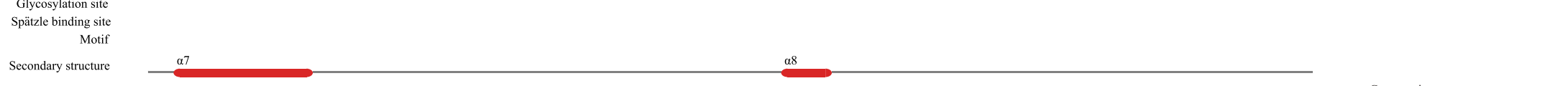
<i>D.melanogaster</i>	558	LNDNPLVCDCTILWFIQLVLRGVHKKPQYSRQFKLRTDRLVCSQPNVLEGTTPVRQIEFPQTLLICPLDFSDDDRERKCPRGCNCHVRTYDKALVINCHSGNLTHTVPR-LPNLH	665
<i>D.erecta</i>	558	LNDNPLVCDCTILWFIQLVLRGVHKTQYADQFKLRTDRLVCTQPNALEGTTPVYQVKFPQTLLICPLDFSDDDRERKCPRGCNCHVRTYDKALVINCHGGNLTHTVPR-LPNLQ	665
<i>D.yakuba</i>	558	LNDNPLVCDCTILWFIQLVLRGVHKKPPYAEQFKLGTDRLVCNQPNALEGTTPVQKVKFPQTLLICPLDFSDDDRDRKCPRGDCCHVRTYDKALVINCHSGNLTHTVPR-LPNLP	665
<i>D.pseudoobscura</i>	560	LNDNPLVCDCTLLWFVQLVRGAHVDPYAKQFKLGTDRLLTCSQPHNLQDLPVRLVFPKELICSIGSAEEPGRQKPCPRGRCWVRTFDKALVIKCHEGNLTKVPE-LPTLY	667
<i>C.capitata</i>	584	LNNNPLHCDCLMLPFLQFISGEFKEKFAANKIELLTAELKCEGPTALQTKPVQELSYMELVCPD-DSSASSERRCPRGCEWVRPHDFMLLVNCSNGNMTRLPA-LPH-E	689
<i>A.aegypti</i>	593	VNDNPLNCCNCVIFSFVQYLLNELDSAVYRRIQFVADDLRCSEPNLEGVHVSQKLTQKDLLCQLD-QPGTEIKHCPANCSCHVRLDRGVIIVNCTRQGLTEIPA-LPQPT	699
<i>A.gambiae</i>	595	LNDNPLHCNCICAYAFAYQYIQNRLATAYDRFELVANELTCSGPEHLEGALIKDVPRTRELLCELD-TPSTAIRHCPAPCRCYIRPEDTGVIVDCSGQALTEVPE-LPRPT	701
<i>A.mellifera</i>	544	IENNPIECDCDIFNFLLYLEGLKLDPNVYKYFHIIMPVCLTQGNPKFKGKEIVKLESKKFICQIS-----NPPCNECTCYSQQSNKEFTVNCSEKNLTSVPR-NIKTL	644
<i>P.corporis</i>	584	LEGNPLKCDQCQIYEMIRYFENKLESEVYVFTLVPGNLTCHSPDVLKGSQVTRVDSSHLMCDLD-----PATYNPCSKCEKLRADWGLIVNCTEKNLTEIPVELPIVK	687
<i>D.pulex</i>	576	LDKNPLVCDCAAYFMAQYIDQS-NPGVRHSWKINPTKLTCEQPASLAGLALSQVNP SQFLCNR-----NEKFWPCEWYVRPVDRTLLFDQCHKNLNEIPTRLPR-H	675
<i>L.scapularis</i>	560	LENNPFDCCNCHIYDFIRYLSDSNQKEIA-LFKN-ASYSKCHDPPSSLSGKSLLTVEFGQLICSIK-----ENCPKGCSCFFRRKDLTTHLDCRGTNLTDLPSTSPS--	657
<i>B.mori</i>	521	VFNNPICVCNMYMVDLLQDVTNRKIM-DARL-----FRCPK-----WQITFCISIRVLT--ISAIILGLGVMYAVITMAYYIYKCRAYETLKWKLSLLSKK	609



<i>D.melanogaster</i>	666	KN---MQLMELHLENNLLRRLPSANTP---GYE---SVTSLHLHLAGNNLT---SIDVDQ-----LPTNLTHLDLISWNHLQMLNATVLGFLNRTMKWRSVKLSGPNPWW	754
<i>D.erecta</i>	666	QN---MQLMELHLEKNTLLRRLPSANTP---GYE---NVTSLYLAGNNLT---NIEVDQ-----LPTNLTHLDVSSNHLQMLNATVLGFLNRTMKWRSVKLSGPNPWW	754
<i>D.yakuba</i>	666	QN---MLMELHLENNLLRRLPSANTP---GYK---SVTSLHLHLAGNNLT---NIDVDQ-----LPTNLNHLDVSWNHLQMLNATVLGFLNRTMKWRSVQLSGPNPWW	754
<i>D.pseudoobscura</i>	668	DN---LHIMELHMDNNTLLGLPAAHSP---GYA---NVTSLHLHLAGNNLT---HIDVDR-----LPTNLKHLDVRRNHLQALNNTVLGFLNRTMPRRSLMLSGNPWI	756
<i>C.capitata</i>	690	PL---LKGIVLFFVENNNLMRPLPGSTS---GYA---DVKQLHAAGNQLR---QIEVNN-----LPVGLKYLDVRHNQLQRLNESVLEYLNSSSSTLEGLLSHNAWA	778
<i>A.aegypti</i>	700	MF---GYNFIELHVEDNNITELPTENLT---GYG---KVAELYARNNAIV---ELLPEN-----LPSSLRILDVTKQKVFVMLNQS VVESLNS SKVLES LHL SGNKWR	789
<i>A.gambiae</i>	702	TF---GYRFIELHLENNISALPDSTTAS---GSSGWA---EVRELLYASNNISIA---ALAADQ-----LPRSLRLLDLSRNRLTTLTDGPLAESLTA STTLT TVRLAHNDWT	795
<i>A.mellifera</i>	645	L---NYKLVIDLTDNKLSEMPSLTEI---GLDNI---QISKLLYSNNDIH---EVSMDG-----LQ-SNIELELHNNISKLEPNVLQFLNRSST---SLTLHGNNPWI	730
<i>P.corporis</i>	688	YT---NHTELFLNGNDLIRTLPNSTFK---GYE---NVTHLYLSNNRIE---NIDTGL-----LSPKIQQLALDHNNTLTKLGDNLIGFLRNSSTEIKY LKLDNPNWI	775
<i>D.pulex</i>	676	ED---YQIQMNLSSNSI---SIGQIHPNSSDCYP---DVTWLDLSHNGMDES---SMSDPQHWAQNLHLRFP-KLNRLDLTHNNFNSIPNGVVDSWNAMHNL-TYNLNGNPWK	774
<i>L.scapularis</i>	658	---NTNILYLQSNSSISLVNLSAP---RWE---NLTEVYLDENLLS---NLDLTT-----MPRRLQILSLTNRLRSLTPQLMGMLSNSSSTLSLSLSGPNPWI	743
<i>B.mori</i>	610	GIRIIEQELIIRYSEKDEEFVMKEILP---TLK---NEKNLKVYTNAIRCPSDI SPDN---LKTCTVKDAKYKTTLLIVFSPNYLTTAYSNVDIKKIHSEMLKAEYTVVYVVD	711



<i>D.melanogaster</i>	755	CDCTAKPLLLFTQD---NFERIIGDRNEMMCV-NAEMP-TRMVE---LSTNDICPAEKGV-----	805
<i>D.erecta</i>	755	CDCTAKPLLLFTQD---NFERIADRNEMMCV-DAEVP-TRMVE---LSTNDICPAERG V-----	805
<i>D.yakuba</i>	755	CCTNAKPLLLFTQD---NFERIIGDRNEMMCV-DAEVP-TRMVE---LSTNDICPAEKGV-----	805
<i>D.pseudoobscura</i>	757	CNCEAKPLLLFTQS---NYERIIGDRGEMLCM-DAEMP-TRMAE---LSTNDICPEEKGV-----	807
<i>C.capitata</i>	779	CDCEAKPLLEFTQNAAILQKLSVRDFNQMHCA TDP TNP-DAAVEWFKDISVADICPAEMGL-----	838
<i>A.aegypti</i>	790	CDCSSAQFMNFVQQ---NHKKIADLQITCETGEEFS-LITVN-----SLCNESVTT-----	837
<i>A.gambiae</i>	796	CQCETQLLTFAHA---NQRRIEDFGRLRCS-DGRPLEQMTLS-----DLCVDRTHMIVLTCV-----	849
<i>A.mellifera</i>	731	CDCDTRDLLSFIQT---KISDIRNSSLITCK-DMNI SMIKMTE-----ADFCPTYTTL-----	779
<i>P.corporis</i>	776	CECSTKSFLELIQL---KFKQVPMLSNVTCFKDNRQLSHLSVS-----DLCPASKTLIA-SCI-----	829
<i>D.pulex</i>	775	CDCTNLALLNFIYG-SWKRLIEDFNQMKCDNGQKISELSVEILCPSVNAAVKYYTIPLPILALLIVCVG-----	841
<i>L.scapularis</i>	744	CD CSTFSFKTWR LGHVYMKDYDPIACGDGVRFNDFHFLYF INEIPDSAYCPVDDSAQRKQLAAVTAICVV-----	814
<i>B.mori</i>	712	IGPENFLYVFLKSRQD TN TMVWVGEANFWLKF LVT LAQYDITYPDL--NKISDKITFAEVEENDKKNCHLCYLGTHRCHSCCYG IILDQSLV	801



Conservation
high █ █ █ █ █ █ █ █ low

Figure S3 (continued)

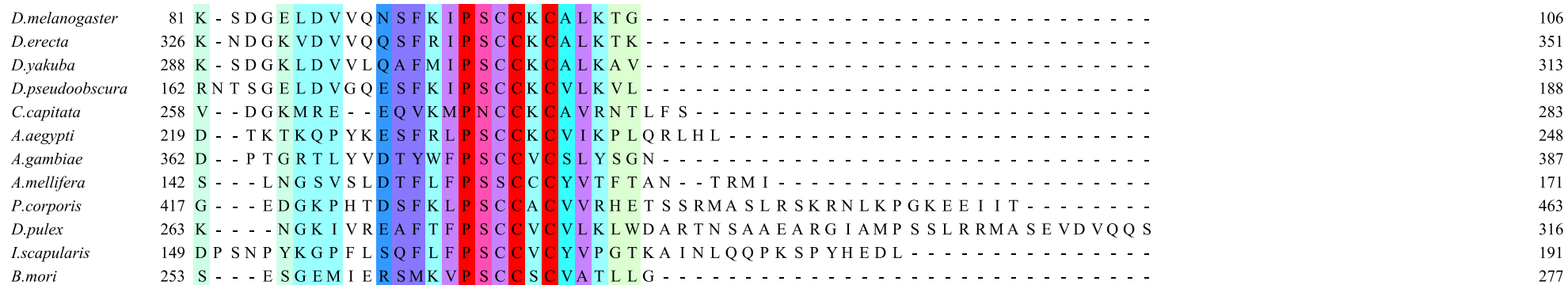
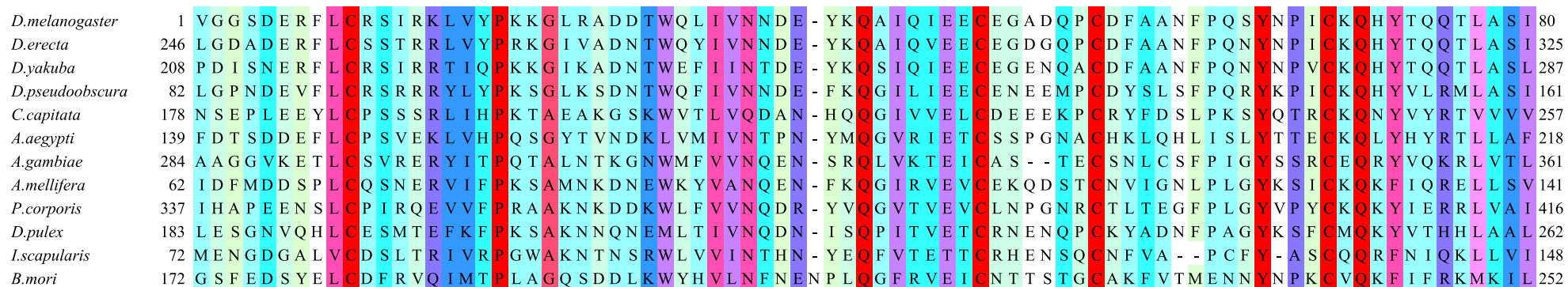


Figure S4. Multiple sequence alignment of Spätzle 1 orthologs. Sequence analyses were carried out as described in Figure S3. Color coding based on conservation as described in Figure S3. Residues of DmSpätzle involved in Toll binding and secondary structural elements of DmSpätzle are shown below the alignment. Accession numbers are as follows: NP_733194.1 (*Drosophila melanogaster*), XP_001981600.1 (*Drosophila erecta*), EDW98554.1 (*Drosophila yakuba*), EIM52487.1 (*Drosophila pseudoobscura pseudoobscura*), XP_004523465.1 (*Ceratitis capitata*), EAT48444.1 (*Aedes aegypti*), EAA04507.4 (*Anopheles gambiae* str. PEST), XP_003250675.1 (*Apis mellifera*), EEB19900.1 (*Pediculus humanus corporis*), EFX88879.1 (*Daphnia pulex*), EEC16960.1 (*Ixodes scapularis*), NP_001108066.1 (*Bombyx mori*). Parthier et al. Toll-Spätzle complex crystal structure SI Appendix, page 9

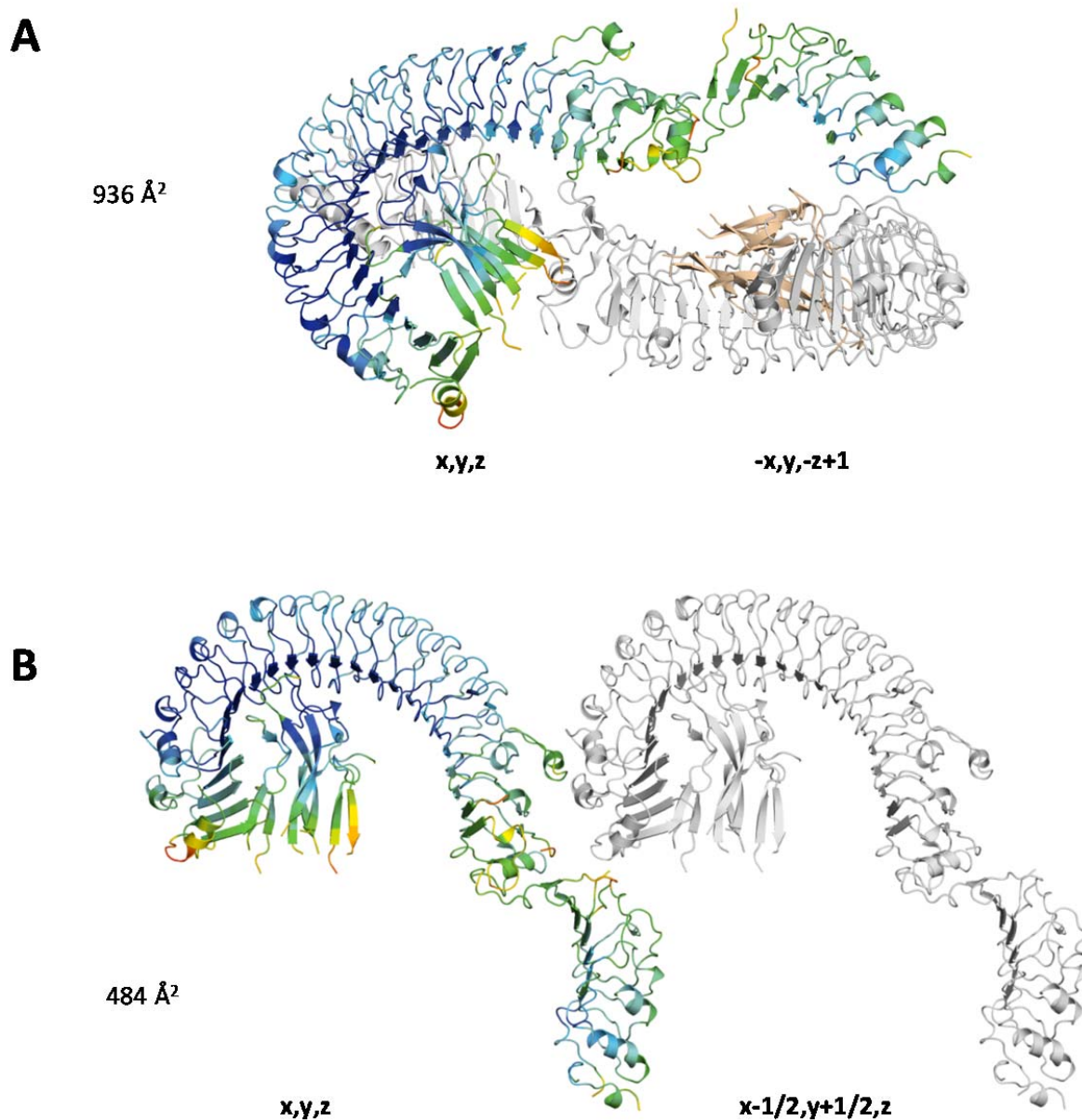


Figure S5. Crystal packing reveals a 1:1 Toll-Spätzle complex. (A and B) Arrangement of symmetry-related molecules in the monoclinic crystals of the Toll Spätzle complex. The symmetry operation noted below the neighboring molecules (colored in grey) is specified in fractional space relative to molecule in the asymmetric unit (x, y, z , colored by B-factor from low/dark blue to high/red). None of the analyzed molecular interfaces deriving from crystal packing (buried surface areas 936 Å² and 484 Å², respectively) are compatible with formation of a putative Toll homodimer.

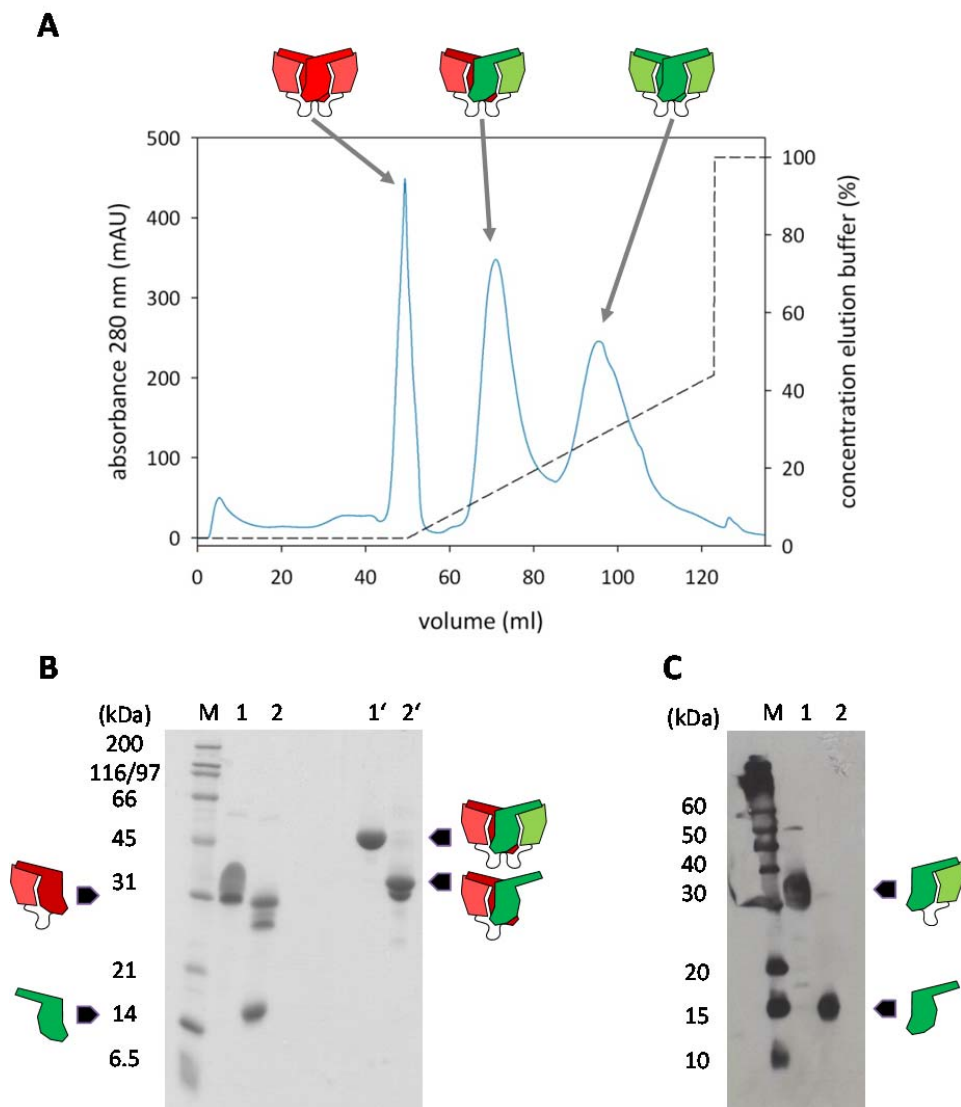


Figure S6. Purification of matured hcSpätzle. (A) Refolding of an equimolar mixture of proSpätzle-His₆ (green pictogram) and ncSpätzle (non-cleavable Spätzle, (4), red pictogram) results in the three possible dimeric species, differing in their His₆-tag content. The homodimeric ncSpätzle₂, heterodimeric hcSpätzle (ncSpätzle-proSpätzle-His₆) and homodimeric wild type Spätzle (proSpätzle-His₆)₂ could be separated using Ni²⁺-affinity chromatography. (B) Maturation with Easter^{Xa} results in the activation cleavage of only one protomer to yield the heterodimeric hcSpätzle. The proform (lane 1 and 1') and the purified, Easter^{Xa}-matured hcSpätzle (lane 2 and 2') are shown under reducing and non-reducing conditions. Reduction of the heterodimeric hcSpätzle leads to separation into one proSpätzle monomer (31 kDa) and one C106 monomer (15kDa). Minor bands at 34 kDa and 29 kDa result from alternative Easter^{Xa} processing as described (4). (C) The corresponding Western blot of hcSpätzle under reducing conditions using an anti-His₆ antibody shows the expected signal for C106, but not for the proSpätzle monomer. (M: molecular weight standard, 1: pro-hcSpätzle, 2: Easter^{Xa}-matured hcSpätzle).

Table S2: Binding parameters for the interaction of Toll ECD with Spätzle C106 and hcSpätzle monitored by isothermal calorimetry (ITC).

Concentrations are for Toll ECD monomer and Spätzle (hetero-)dimer; stoichiometry n refers to the molar ratio of injectant : titrand. Data analysis was performed with the ORIGIN software provided. All measurements were performed at 25°C. Conversion of the commonly used unit calorie (cal) into the SI unit Joule (J) can be done using $1 \text{ J} = 4.182 \text{ cal}$.

Syringe		Cell		n	K_D (nM)	$\Delta_B H$ (kcal·mol ⁻¹)	$T\Delta_B S$ (kcal·mol ⁻¹)
Protein	conc. (μM)	Protein	conc. (μM)				
Toll ECD	131	C106	8.4	0.94 ± 0.003	5.9 ± 1.1	-11.78 ± 0.07	-0.56
C106	68	Toll ECD	9	1.01 ± 0.003	9.2 ± 1.6	-11.92 ± 0.08	-0.96
hcSpätzle	68	Toll ECD	9	1.00 ± 0.003	23.04 ± 2.6	-9.99 ± 0.06	0.43
Toll ECD	151	hcSpätzle	9.9	1.05 ± 0.005	5.85 ± 2.19	-9.02 ± 0.09	2.22

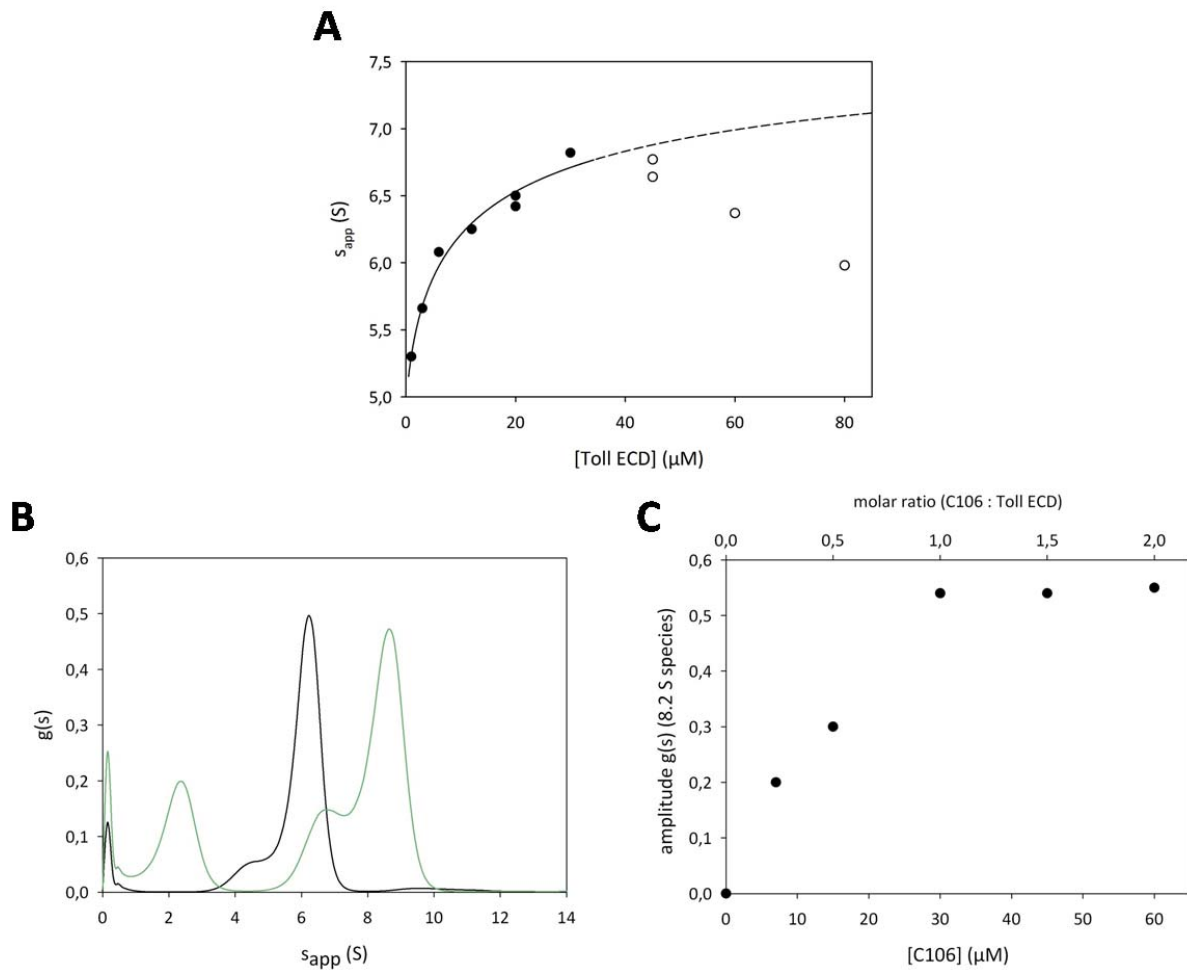


Figure S7. Sedimentation velocity analyses of Toll ECD and its complex with C106. (A) The increase in apparent sedimentation velocity of Toll ECD as a function of protein concentration (1 – 80 μM) is indicative of oligomerization. Data showing ideal behavior (closed circles) were fitted according to a dimer model as described in the materials and methods to give a dissociation constant of $K_D = 6.0 \pm 5.4 \mu\text{M}$ and sedimentation coefficients of $s_M = 4.9 \pm 0.3 \text{ S}$ (monomer) and $s_D = 8.0 \pm 0.6 \text{ S}$ (dimer). At high concentrations (open circles), the data showed significant non-ideality of the solution and were not included in the regression. (B) Sedimentation velocity distributions $g(s)$ for 30 μM free Toll ECD (black) and a sample with 30 μM Toll ECD - 60 μM C106 (green). With an excess of ligand, three species are visible, with $s_{app} = 2.3 \text{ S}$ (free C106), $s_{app} = 8.6 \text{ S}$ (dimeric 2:2 complex) and $s_{app} = 6.9 \text{ S}$ (monomeric 1:1 complex). (C) Titration of 30 μM Toll ECD with increasing concentrations of Spätzle C106 results in an increase of the amplitude $g(s)$ of the 8.2 S species. The inflection point at an equimolar ratio confirms the 2:2 stoichiometry assignment.

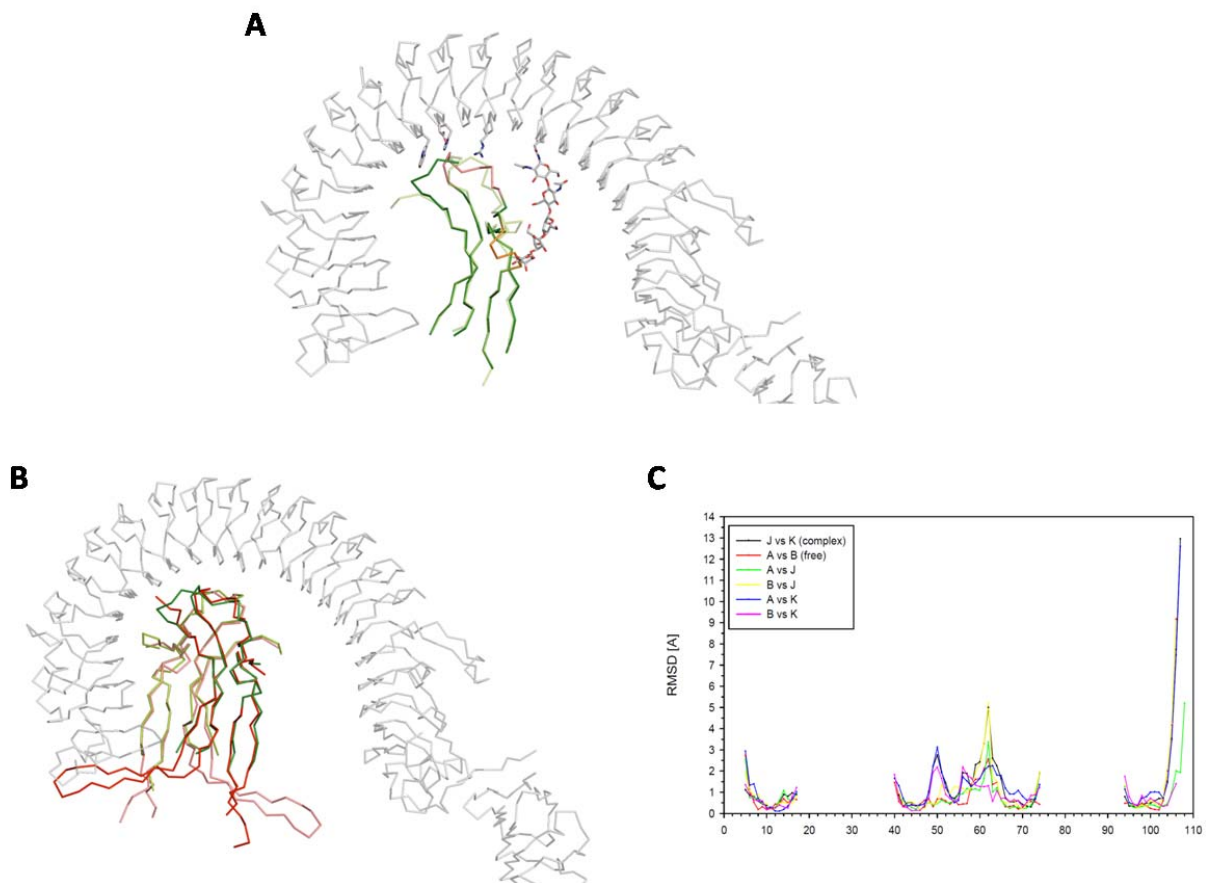


Figure S8. Conformational differences in the Spätzle protomers. (A) Superposition of the leading (chain K, dark green) and trailing (chain J, light green) protomers of Toll-bound Spätzle, revealing structural differences between the two protomers (i) in the region where chain J interacts with Toll Asn³⁴⁶-linked glycan residues Man₃-Man₄ (Spätzle residues orange backbone, sugars as sticks) and (ii) at the base of the Spätzle protomers (shown as pink backbone range), which are involved in the network of interactions to Toll described in **Figure 2**. (B) Comparison of Toll-bound C106 (green) and free Spätzle cystine knot domain (red, PDB code 3E07 (5)). (C) Per-residue rmsd values between C α atom positions of structurally aligned protomer chains (J, K) of bound Spätzle C106 and free Spätzle C106 (chains A, B). Abscissae: Spätzle C106 residue numbers; no rmsd values are given for undefined residues. The structural alignment was carried out using the program LSQKAB from the CCP4 suite (6).

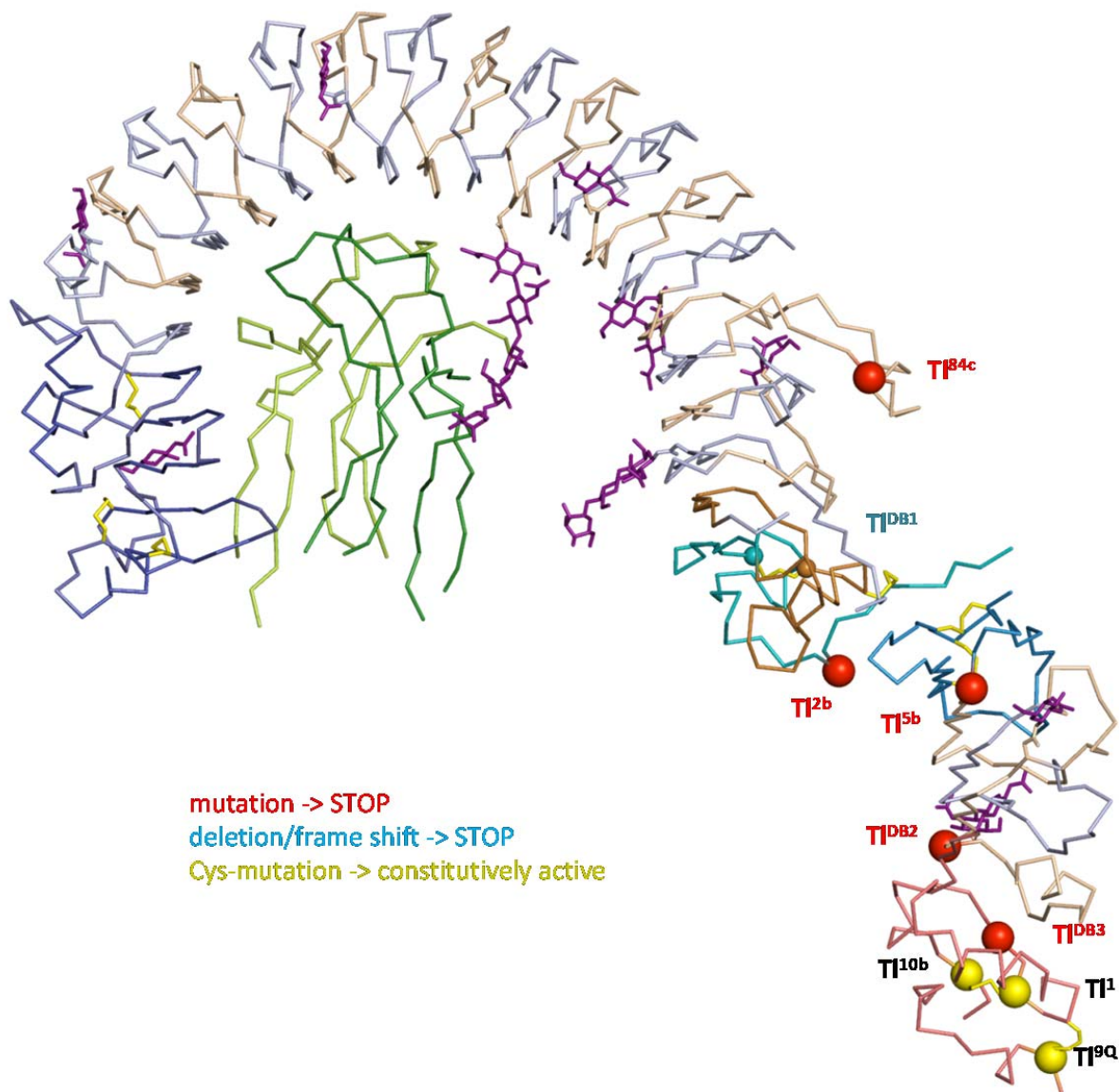


Figure S9. Mapping of *toll* gain of function mutations (7, 8). Coloring of Toll ECD and Spätzle as in **Figure 1**. Red spheres represent positions of premature <STOP> codons, whereas yellow spheres indicate Cys => Tyr mutations. The cyan backbone represents an altered sequence due to a frameshift mutation in *Toll*^{DB1}, whereby the Toll sequence ⁵⁸⁷QFKL RTDRLV⁵⁹⁷... is replaced by the sequence ⁵⁸⁷HRPI GLQPAK⁵⁹⁷CLGG YSSSDRRTD AHL SAGLLR <STOP>, allowing serendipitous disulfide bridge formation to form a stable soluble truncated Toll-mimic protein incorporating only the membrane distal LRR domain. Positions of ⁵⁹⁷C and its disulfide bridge partner ⁵⁶⁵C are indicated with cyan and brown spheres respectively.

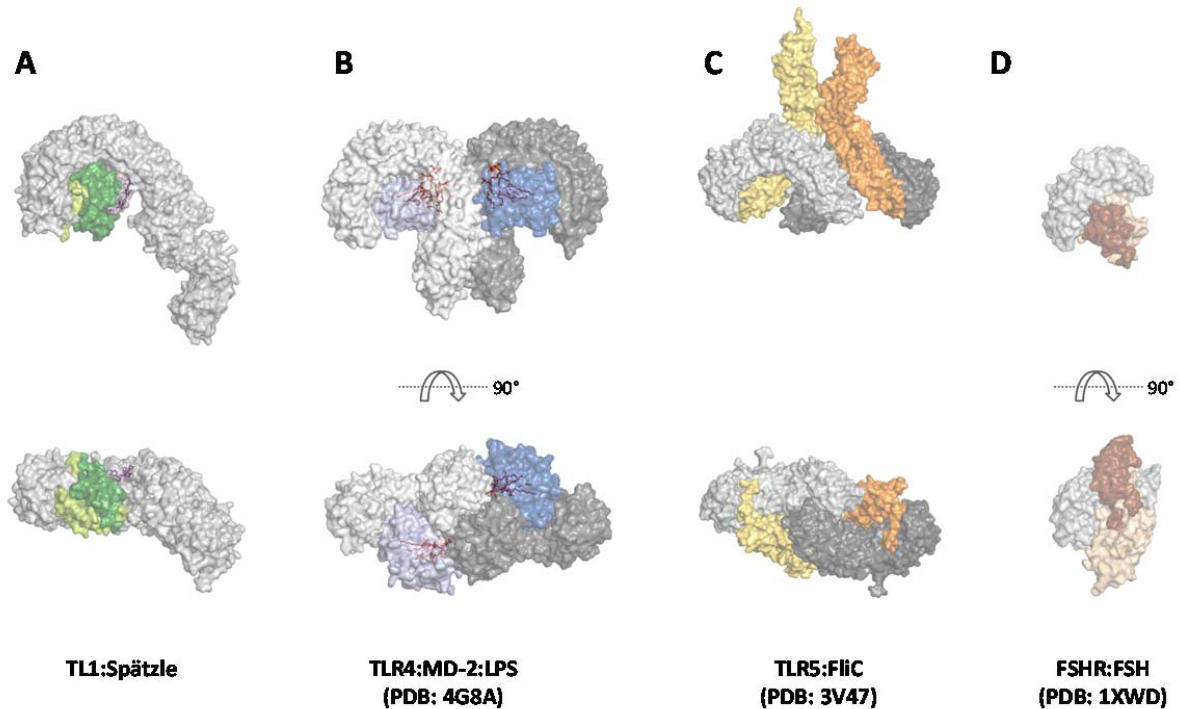


Figure S10. Comparison of the Toll-Spätzle complex to other LRR-containing receptors with protein ligands. Equivalent orientations based on structural alignment of the LRR residues 100 to 200 of each receptor; 90° horizontally rotated view below. **(A)** The Toll-Spätzle complex: the *Dm* Spätzle C106 dimer (comprising two protomers, colored dark green and bright green, respectively) binds to the concave side of monomeric *Dm* Toll-ECD (grey). **(B)** Crystal structure of dimeric human TLR4 (colored in dark grey and white) in complex with its co-receptor human MD-2 (dark and light blue) that recognizes the bacterial lipopolysaccharide (LPS, red sticks) (PDB ID 4G8A (9)). **(C)** Crystal structure of dimeric zebrafish TLR5 (colored in dark grey and white) bound to FliC (colored in orange and yellow), a fragment of of *Salmonella* Flagellin (PDB ID 3V47 (10)). **(D)** Crystal structure of human follicle-stimulating hormone (FSH, colored in brown and beige) in complex with its receptor FSHR (colored in white). Similar in concept to the Toll-Spätzle complex, the binding mode involves a dimeric cystine knot ligand that binds to the concave side of the LRR-containing, monomeric FSHR (PDB ID: 1XWD (11, 12)).

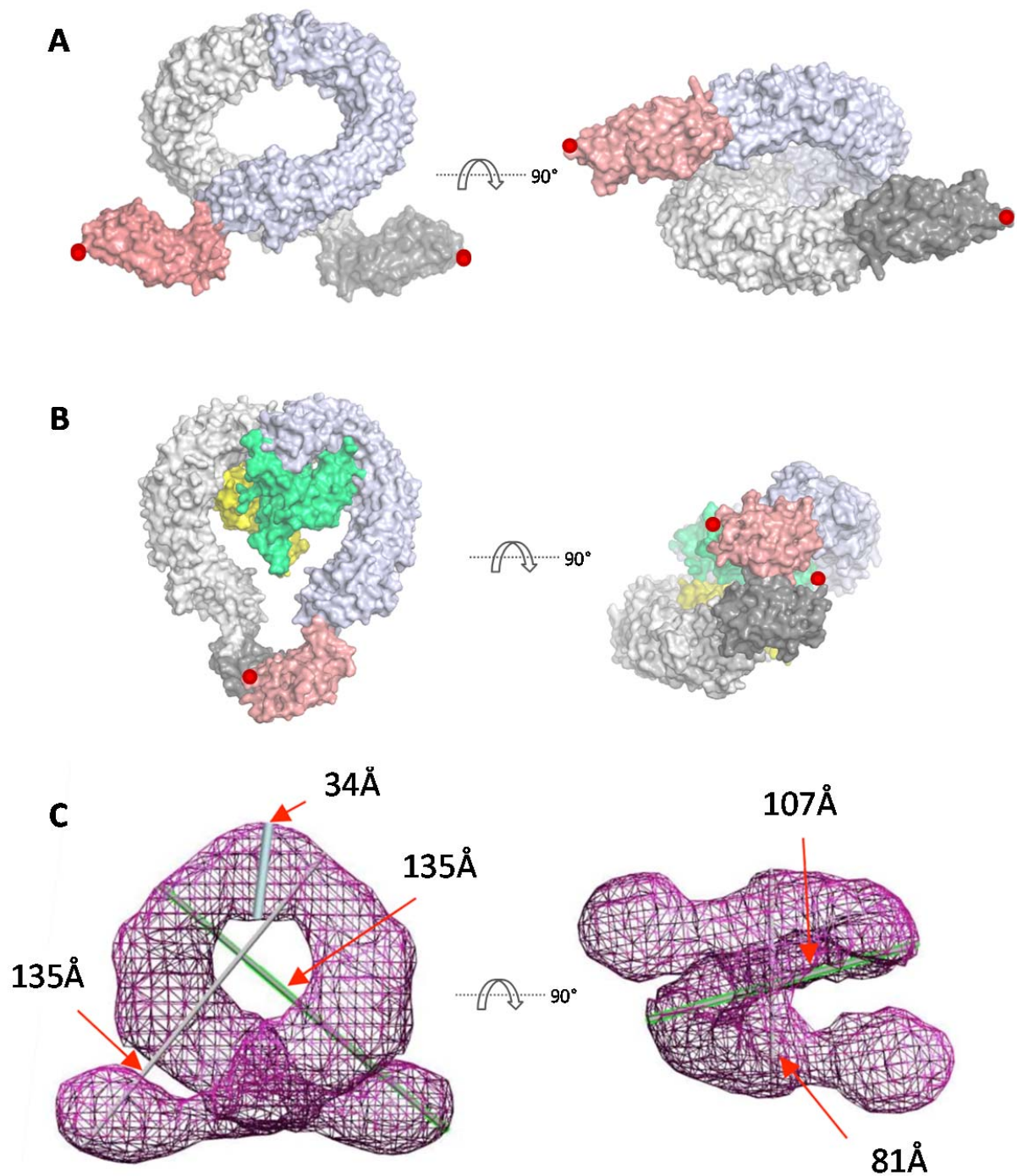
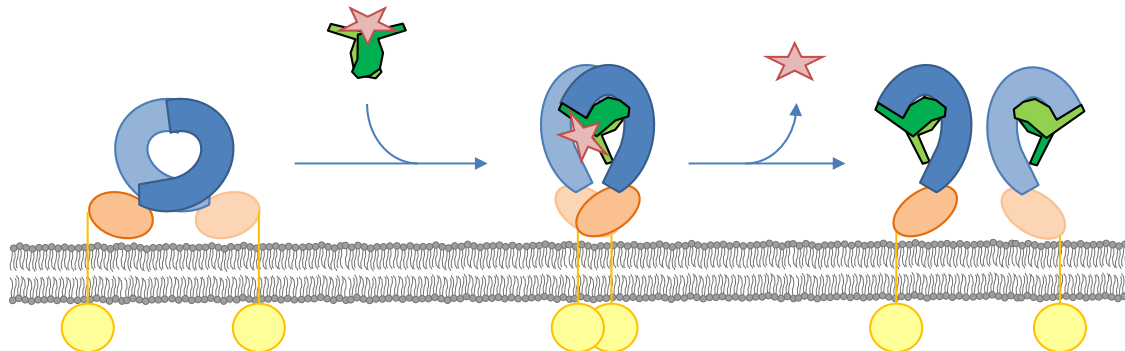


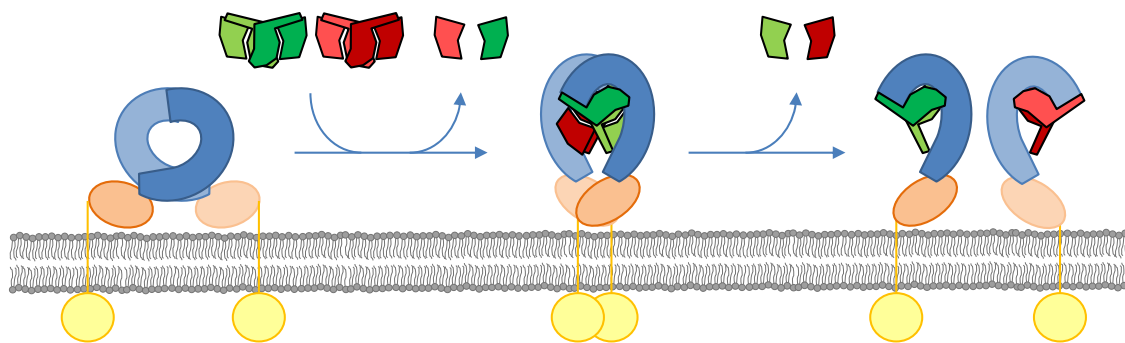
Figure S11. Models of Toll dimerization based on the cryoEM reconstruction of Gangloff *et al.* (2008) (13). (A) The crystal structure of Toll was scaled and oriented to reproduce the features of the “2:2 Toll-Spätzle complex” of Gangloff *et al.* (2008) (13), guided by the curvature of the large N-terminal membrane distal domain of the Toll ECD and the two-fold symmetry. The

“smaller” volume in the 30 Å resolution reconstruction, previously ascribed to Spätzle (which is not possible based on both the present structure and the data of Gangloff *et al.* (2013) (14)), was assigned to the C-terminal membrane proximal LRR domain; red circles mark the locations of the C-termini that link the ECD to the transmembrane sequence. The model suggests steric clashes of the LRRNT1 domains that could be alleviated by a slight “straightening” of the solenoid (also suggested by the reconstruction volume, compare **(C)**) in the absence of Spätzle. Such a head-to-head LRRNT1-LRRNT1 interaction could be disrupted by binding of the Spätzle wings of one C106 dimer to one ECD monomer. **(B)** A hypothetical “active” complex (in this case 2:2) was generated by juxtaposing the ascending flanks of the two Toll receptors in order to bring the C-termini of the membrane-proximal domains in close proximity while avoiding clashes of the two Spätzle dimers (green and yellow). **(C)** Cryo-EM reconstruction of Gangloff *et al.* (2008), modified from Supplementary Figure S1C (13), oriented as in **(A)**. In comparison to **(A)**, the reconstruction suggests additional densities at the concave surface and at the interface between the C-terminal halves of the N-terminal ligand binding domains, which could be accommodated by oligosaccharide chains.

A: co-ligand mediated signaling



B: prodomain mediated signaling



inactive
preformed
dimer

active
signaling
dimer

inactive
dissociated
monomers

Figure S12. Models for the initiation of Toll signaling by Spätzle. In the crystal structure described in this manuscript, only a 1:1 Spätzle dimer : Toll ECD complex is observed, yet Toll signaling is almost certainly to involve Toll receptor dimerization. Starting from disruption of a preformed inactive Toll dimer (**Figure S11**) by Spätzle binding, we suggest two possible routes to signaling and subsequent dissociation: **(A)** an unidentified co-ligand (pink star) could bind to the wings of Spätzle, organizing these to facilitate dimerization (depicted for simplicity as a 1:2 complex). **(B)** Easter maturation of proSpätzle results in the non-covalent complex Spätzle*. The binding mode observed in the crystal can be achieved through displacement of only one prodomain, so that the remaining non-covalently bound prodomain is available to complement receptor dimerization. Removal of the remaining prodomains would result in dimer dissociation.

SI Methods

Expression, complex formation and crystallization. In brief, the stable Schneider 2 cell line containing pMTV5/His-A Toll ECD¹⁻⁸⁰² was cultured in Insect Express Prime medium (PAA, Pasching, Austria). Expression was induced with copper sulfate, the supernatant harvested after 4-6 days and pure receptor (with an N-terminal Ser²⁸ due to cleavage of the signal sequence following secretion) obtained following Ni²⁺-affinity chromatography and size exclusion chromatography. Spätzle C106, produced as inclusion bodies in *E. coli*, refolded and processed with trypsin as described (5, 15), was mixed with purified Toll in a ratio of 3:1. The complex was digested with EndoH_f and applied to a size exclusion chromatography column. The pure deglycosylated complex crystallized in 10% (w/v) PEG 3350, 50 mM HEPES Na-salt, pH 7.5 at 287 K, and the crystals were dehydrated by increasing the PEG concentration of the crystallization buffer to 30%. For heavy metal derivatization, a single crystal of the deglycosylated complex was soaked for one day in 5mM samarium acetate dissolved in the final dehydration buffer at 287 K.

A non-deglycosylated (for simplicity termed "glycosylated") complex (prepared following the same purification and complex formation protocol as described above, but without Endo H_f digestion) crystallized in 100 mM HEPES, 5 % (v/v) propan-2-ol, 10% PEG 4000, pH 7.5 at 287 K within one week. The crystals were dehydrated using the same method as for deglycosylated crystals (16) starting from the crystallization conditions, ending with 30% PEG 4000 (other factors remaining the same).

Data collection, phase determination, model building and refinement. The crystals, which needed no further cryoprotectant, were kept at 100K. Data sets were collected at the BESSY synchrotron (Helmholtz-Zentrum, Berlin) on beam line BL14.1 (deglycosylated complex, native and samarium derivative) or BL14.2 (glycosylated crystals). Native data sets were collected at a wave length of 0.9184 Å, MAD data sets at 1.8448 Å (peak), 1.8454 Å (inflection), 1.8395 Å (remote) and 0.9184 Å (high energy remote), and processed using the XDS package (17). Heavy atom sites (four) were determined and initial phases obtained using SHELX (18) and SHARP to a maximum resolution of 3.02 Å, followed by density modification using SOLOMON within the autoSHARP pipeline (19). The characteristic horseshoe form of the LRRs was recognizable in the initial density, allowing building of generic leucine rich repeats into the density with COOT (20), and an initial placement of the cystine knot of Spätzle C106 (PDB: 3E07) (5) was achieved

through identification of the disulfide positions. The model was subjected to cycles of refinement using PHENIX.refine (21) of the PHENIX suite (22) and model building using the highest resolution (2.2 Å) native data set. The final refinement cycle included TLS refinement with the automatic TLS group detection of PHENIX.refine. Data and refinement statistics are given in **Tables S3** and **S4** respectively. Structures were validated using Molprobity (23) and figures prepared using PyMOL (Schrödinger, LLC).

Analytical ultracentrifugation. Analytical ultracentrifugation experiments were performed on an Optima XL-A ultracentrifuge (Palo Alto, CA, USA) equipped with an An50Ti rotor and double sector cells. Sedimentation equilibrium measurements of 1 μM Toll ECD in the presence and absence of 1 μM Spätzle C106 or 1 μM hcSpätzle were performed at 6000 rpm, 20°C. Sedimentation velocity measurements (for the self-association of free Toll ECD using monomer concentrations of 1 – 80 μM or for the interaction of 30 μM Toll ECD with increasing concentrations of Spätzle C106 (0 – 60 μM)) were carried out at 40 000 rpm, 20°C. Data were recorded every 15 minutes at 250 nm and 305 nm. Samples were prepared at protein concentrations of 1 - 30 μM (monomer concentration) in 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.005% (w/w) Nonidet P40. Data analysis was performed with the software Sedfit (24). We assumed a monomer-dimer equilibrium of the Toll ECD in which the apparent sedimentation velocity is defined as:

$$s_{app} = \frac{s_M[M] + s_D[D]}{[M] + [D]} \quad (1)$$

where [M] is the concentration of free monomer, [D] of the dimer and s_M and s_D correspond to the sedimentation velocities of the monomeric and dimeric species respectively. The total protein concentration $[M]_T$ is defined as:

$$[M]_T = [M] + 2[D] \quad (2)$$

and the dissociation constant as:

$$K_d = \frac{[M]^2}{[D]} \quad (3)$$

Solving equations (2) and (3) for the free monomer concentration and substituting this term in (1) can be used to fit the concentration dependent sedimentation velocity of the Toll ECD.

Preparation of Spätzle* and hcSpätzle. ProSpätzle isoform Spz11.7 (possessing a C-terminal His₆-tag) was prepared from inclusion bodies and refolded as described (5). In order to generate hcSpätzle, the C-terminal His₆-tag of non-cleavable ncSpätzle (in which the Easter maturation site VSSR-VGG is replaced by VSSA-VGG (4)) was deleted using the primers 5'-G AAG ACT GGG TGA GAG CAC CAC-3' (forward) and 5'-GTG GTG CTC TCA CCC AGT CTT C-3' (reverse). Inclusion bodies were solubilized in 6 M guanidinium hydrochloride (GdmCl), 0.1 M Tris/HCl pH 8.0, 1 mM EDTA, 0.1 M DTT for 24 hours at 4°C, after which the pH was adjusted to 4.0 using concentrated hydrochloric acid and dialyzed four times against 4 M GdmCl, pH 4.0. Denatured proSpätzle-His₆ and ncSpätzle were mixed in a 1:1 ratio and refolded by rapid dilution into chilled, degassed refolding buffer (1 M Tris/HCl pH 8.5, 5 mM EDTA, 5 mM GSH, 2 mM GSSG) to a final concentration of 150 mg/L at 8°C. After incubation for 48 hours, the protein solution was concentrated 10-fold by cross-flow filtration (Vivaflow, Sartorius, Göttingen, Germany) and dialyzed twice against 20 mM Tris/HCl pH 7.4, 20 mM NaCl. The dimeric species was purified via size exclusion chromatography on a Superdex 75 HiLoad 26/60 column (GE Healthcare, Freiburg, Germany) equilibrated with dialysis buffer, resulting in the three possible disulfide-linked dimeric species (proSpätzle-His₆)₂, ncSpätzle-proSpätzle-His₆ and ncSpätzle₂. These were separated on a His-trap HP column (**Figure S7**, GE Healthcare), applying an elution gradient to 20 mM Tris/HCl pH 7.4, 20 mM NaCl, 500 mM imidazole over thirty column volumes. Pooled fractions were dialyzed against 20 mM Tris/HCl pH 7.4, 20 mM NaCl and stored at -20°C.

Spätzle*, the non-covalent complex of prodomain and C106, was obtained by treatment of (proSpätzle-His₆)₂ with active Easter^{Xa} as published (4), and hcSpätzle* obtained by corresponding treatment of the heterodimeric ncSpätzle-proSpätzle-His₆. Briefly, Factor Xa-activated protease Easter^{Xa} was added to the proSpätzle variants (100-fold molar excess) and incubated at 30°C in 20 mM Tris pH 7.4, 20 mM NaCl for 2 hours.

The prodomain of Spätzle* can be reversibly dissociated from the disulfide-linked C106 under mildly denaturing conditions (4). Separation from prodomains was thus effected by adjusting each sample of hcSpätzle* and Spätzle* to 1.5 M GdmCl, 20 mM Tris pH 7.4, 20 mM NaCl, before application to a HisTrap HP column (GE Healthcare) with a linear gradient to 20 mM Tris/HCl pH 7.4, 20 mM NaCl, 1.5 M GdmCl, 300 mM imidazole over 60 column volumes. The resulting hcSpätzle and C106 proteins were subjected to size exclusion chromatography using a

S200 10/300 GL column in 20 mM Tris/HCl pH 7.4, 20 mM NaCl to remove GdmCl and imidazole.

Isothermal calorimetry (ITC). ITC experiments were carried out using a VP-ITC Micro Calorimeter (Microcal LLC, Northampton MA, USA) at 25°C in 20 mM Tris/HCl pH 7.4, 20 mM NaCl (150 mM NaCl) with a stirring speed of 307 rpm and an equilibration time of 300 s between each injection. The samples were dialyzed twice against buffer and all measurements were further corrected for heat of dilution by reference experiments in which the titrant was injected into buffer. Data analysis was performed with the ORIGIN software provided according to the manufacturers' recommendation.

SI Results

Detailed description of the Toll ECD

LRRNT1: An N-terminal amphipathic α -helix covers strand β 1, to which it is disulfide linked. A second disulfide bond connects strand β 1 to the subsequent strand β 2 and a third connects the C-termini of strands β 4 and β 5. The back face of the β -sheet, corresponding to the convex face of the LRR domain, is covered by a 7 amino acid long α -helix that lies nearly perpendicular to the β -sheet and is located in the extended loop between β 4 and β 5 as well as by a two-turn 3_{10} -helix following strand β 5. Whereas the shielded N-terminal β -hairpin and the conserved sequence motif Cx_nCx_mC correspond to a typical LRRNT cap (25), the remainder of the Toll LRRNT1 has not been observed in other LRR proteins.

N-terminal LRR-domain: The consensus sequence $xL_2xxL_5xL_7xxN_{10}xL_{12}xxL_{15}P_{16}xxL_{19}F_{20}xxL_{23}x$ matches that found in eukaryotic TLRs (25, 26). The curved parallel β -sheet of the LRR concave surface consists of 3-5 amino acid β -strands that start at x_4 and is stabilized on the ascending flank by the highly conserved N_{10} which makes regular inter-strand side chain - main chain hydrogen bonds to form the "Asn-ladder" typical of LRR domains. The N-terminal LRRs contribute four 3_{10} -helices to the convex face of the LRR module whereas LRRs 6-13 form a tandem β -turn motif typical for 24-residue LRRs. The tandem of three consecutive β -turns starts with the conserved P_{16} and contains the highly conserved F_{20} , which forms a hydrophobic "spine" to the convex side seen in various LRR proteins (25). The conserved polar residue x_1 (mostly N or Q) is positioned at the very solvent exposed tip of the descending side and includes the glycosylation sites N^{175} , N^{270} and N^{391} .

Only two LRRs exhibit insertions longer than five residues (LRR 14: 35 amino acids, LRR17: 30 amino acids), both of which are distant to the Spätzle binding site. The 14 residue insertion in LRR14 harbors a short hydrophobic α -helix (L⁴⁵⁶ – M⁴⁶⁰) that bulges out of the convex side and covers the following two LRRs. The interface of the α -helix with the tandem β turn of LRR15 is dominated by van der Waals' contacts between helix residues I⁴⁵⁹ and I⁴⁶⁵ and F⁴⁸⁷ of LRR15 as well as between L⁴⁵⁶ and I⁴⁸⁶.

LRRCT1: The first LRR module is terminated by a C-terminal capping motif (amino acids Asn⁵⁵² - Ser⁶²³, LRRCT1) which exhibits the typical disulfide pattern (PxxCx_mCx₂₀C) (conventional: m~20, for Toll m=29 (26)) and topology ($\beta\alpha\beta$) of the C-cap subfamily CF1 (27). The two β -strands add to and extend the parallel LRR β -sheet. The seven-residue α -helix covers the hydrophobic core of the cap, and is followed by an exposed loop that is five to eight residues longer than the respective loops of TLR C-caps (28-30) and expands the descending flank of the LRR domain. This loop, which contains a Toll specific 3_{10} -helix, is situated immediately below a disordered six-residue insertion in LRR17 that protrudes into solvent. The amphipathic cap is completed by a CF1-typical C-terminal loop that is positioned parallel to the α -helix and linked to its N-terminus by two conserved disulfide bonds. The loop (and therewith the N-terminal domain) ends in an extended sequence stretch (amino acids Leu⁶¹⁶-Ser⁶²³) containing the final C-cap Cys⁶¹⁸ that is disulfide linked to Cys⁵⁶⁷. Based on the limited buried surface area, the short antiparallel β -sheet interaction between the final LRRCT1 β -strand (Ile⁶¹⁷-Pro⁶¹⁹) and the N-terminal β -hairpin of LRRNT2 is probably weak and therefore might not be a feature of the Toll ECD in solution.

Membrane-proximal LRR domain: LRRNT2 displays the same capping motif as that found at the N-termini of all known TLR structures (10, 30, 31) as well as other LRR-proteins such as the variable lymphocyte receptors (32, 33) or the Slit2 dimerization domain D4 (34), and is stabilized by an interstrand disulfide bond Cys⁶³⁵ - Cys⁶⁵⁰, a preceding loop structure that covers the hydrophobic core of the second LRR module and a 19 residue loop that connects the second β -strand to the first LRR by adding to the convex surface of the subsequent LRR module. Unlike the TLR structures, the Toll LRRNT2 contains a conserved disulfide bond (Cys⁶³¹ - Cys⁶³⁷) that links the N-terminal shielding loop to the first β -strand. In addition, Toll LRRNT2 shows an unusual six residue insertion (Arg⁶⁴⁰ – Ala⁶⁴⁵) in the β -hairpin loop that approaches LRRCT1 of the N-terminal domain as well as, in contact with the β -hairpin loop, a five residue insertion

(His⁶⁶⁵-Gln⁶⁶⁹) preceding the β -strand of LRR18 that is found in none of the known TLR structures nor in other LRR proteins like VLRB.2D (33), the Nogo Receptor (35), Slit2-D4 (34) or LGR5 (36). Interestingly, the first insertion harbors Arg⁶⁴⁰ and Asp⁶⁴³, two very solvent exposed residues that are highly conserved among Toll receptors. Together with the two Toll specific insertions of LRRCT1 described above, these loops form the surface of the descending side of Toll at the interdomain border. The center of the five stranded β -sheet of the membrane-proximal domain is formed by LRRs 18-20, which adopt typical LRR structures and contribute two 3_{10} helices (LRR18,19) and a seven residue α -helix (LRR20) to the convex surface of the LRR module. The terminal C-flank LRRCT2 corresponds to a typical CF1 cap observed in TLRs (28-30) or VLRLs (33, 37). In summary, the overall structure of the C-terminal domain resembles that of other small LRR domains such as VLRB2.D (33).

Table S3: MAD datasets of Sm³⁺-soaked Toll:Spätzle complex

dataset	high energy remote	peak	inflection	remote
beam line	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
wavelength [Å]	0.9184	1.8448	1.8454	1.8395
space group	C2	C2	C2	C2
cell parameters				
a,b,c [Å]	172.9 78.5 125.8	172.6 78.5 125.7	172.8 78.4 125.8	172.8 78.4 125.8
α,β,γ [°]	90 126.2 90	90 126.2 90	90 126.2 90	90 126.2 90
resolution [Å]	50-3.04 (3.20-3.04)	50-3.5 (3.72-3.50)	50-3.52 (3.73-3.52)	50-3.2 (3.39-3.20)
completeness [%]	97.4 (97.5)	97.6 (95.0)	97.9 (97.6)	97.8 (97.2)
total reflections	99220	63827	63341	83831
unique reflections*	50444 (8110)	32872 (5160)	32572 (5192)	43316 (6974)
R _{merge}	4.7 (40.3)	10.2 (35.1)	5.5 (21.6)	5.1 (28.7)
Wilson B-factor	73.7	61.9	65.1	65.5
I/ σ (I)	13.4 (2.0)	7.3 (2.5)	11.7 (3.6)	12.2 (2.7)
phasing power (anomalous differences)	0.34	0.96	0.90	0.68

*Friedel pairs treated as separate reflections.

Values for highest resolution shell are given in parentheses.

Table S4: datasets of native complexes and refinement

dataset	deglycosylated	glycosylated
beam line	BESSY BL14.1	BESSY BL14.2
wavelength [Å]	0.9184	0.9184
space group	C2	C2
Cell parameter		
a,b,c [Å]	172.0 78.2 124.4	171.3 76.8 123.8
α,β,γ [°]	90 126.3 90	90 126.3 90
resolution	45.75-2.2 (2.32-2.20)	33.9-3.3 (3.4-3.3)
completeness [%]	99.3 (98.8)	93.1 (80.3)
total reflections	207577	38648
unique reflections*	67514 (10786)	18492 (1689)
R _{merge}	5.1 (69.0)	8.9 (49.8)
Wilson B-factor	52.2	64.6
I/ σ (I)	15.02 (1.68)	8.76 (1.8)
refinement		
R values [%]		
<i>work</i>	17.6	20.7
<i>free</i>	21.8	25.8
number of atoms		
protein	7103	6986
carbohydrates	240	310
solvent	504	-
average B factor [Å ²]	44.9	34.9
Rmsd		
<i>bond lengths</i>	0.008	0.005
<i>bond angles</i>	1.14	0.95
Ramachandran [%]		
<i>favored</i>	93.9	86.1
<i>allowed</i>	5.9	12.7
<i>outlier</i>	0.2	1.2
PDB ID	4LXR	4LXS

*Friedel pairs merged.

Values for highest resolution shell are given in parentheses.

SI References

1. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.
2. Altschul, S. F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
3. Larkin, M. A. *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
4. Ursel, C. *et al.* (2013) In vitro maturation of *Drosophila melanogaster* Spatzle protein with refolded Easter reveals a novel cleavage site within the prodomain. *Biol Chem* **394**, 1069-1075.
5. Hoffmann, A. *et al.* (2008) Biophysical characterization of refolded *Drosophila* Spatzle, a cystine knot protein, reveals distinct properties of three isoforms. *J. Biol. Chem.* **283**, 32598-32609.
6. Bailey, S. (1994) The CCP4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* **50**, 760-763.
7. Schneider, D. S., Hudson, K. L., Lin, T. Y., & Anderson, K. V. (1991) Dominant and Recessive Mutations Define Functional Domains of Toll, A Transmembrane Protein Required for Dorsal Ventral Polarity in the *Drosophila* Embryo. *Genes & Development* **5**, 797-807.
8. Hu, X. D., Yagi, Y., Tanji, T., Zhou, S. L., & Ip, Y. T. (2004) Multimerization and interaction of Toll and Spatzle in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9369-9374.
9. Park, B. S. *et al.* (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, 1191-U130.
10. Yoon, S. I. *et al.* (2012) Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859-864.
11. Fan, Q. R. & Hendrickson, W. A. (2005) Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* **433**, 269-277.
12. Jiang, X. *et al.* (2012) Structure of follicle-stimulating hormone in complex with the entire ectodomain of its receptor. *Proc Natl Acad Sci U S A* **109**, 12491-12496.
13. Gangloff, M. *et al.* (2008) Structural insight into the mechanism of activation of the Toll receptor by the dimeric ligand Spatzle. *J. Biol. Chem.* **283**, 14629-14635.
14. Gangloff, M., Arnot, C. J., Lewis, M., & Gay, N. J. (2013) Functional Insights from the Crystal Structure of the N-Terminal Domain of the Prototypical Toll Receptor. *Structure* **21**, 143-153.
15. Hoffmann, A., Neumann, P., Schierhorn, A., & Stubbs, M. T. (2008) Crystallization of Spatzle, a cystine-knot protein involved in embryonic development and innate immunity in *Drosophila melanogaster*. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **64**, 707-710.

16. Stelter, M. *et al.* (2013) High level expression of the *Drosophila* Toll receptor ectodomain and crystallization of its complex with the morphogen Spätzle. *Biol Chem* **394**, 1091-1096.
17. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D-Biological Crystallography* **66**, 125-132.
18. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallographica Section A* **64**, 112-122.
19. Vonrhein, C., Blanc, E., Roversi, P., & Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol* **364**, 215-230.
20. Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.
21. Afonine, P. V. *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367.
22. Adams, P. D. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**, 213-221.
23. Chen, V. B. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.
24. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-1619.
25. Bella, J., Hindle, K. L., McEwan, P. A., & Lovell, S. C. (2008) The leucine-rich repeat structure. *Cell Mol. Life Sci* **65**, 2307-2333.
26. Buchanan, S. G. & Gay, N. J. (1996) Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog. Biophys Mol. Biol* **65**, 1-44.
27. Kobe, B. & Kajava, A. V. (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin. Struct. Biol* **11**, 725-732.
28. Choe, J., Kelker, M. S., & Wilson, I. A. (2005) Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* **309**, 581-585.
29. Kim, H. M. *et al.* (2007) Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* **130**, 906-917.
30. Tanji, H., Ohto, U., Shibata, T., Miyake, K., & Shimizu, T. (2013) Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science* **339**, 1426-1429.
31. Botos, I., Segal, D. M., & Davies, D. R. (2011) The structural biology of Toll-like receptors. *Structure* **19**, 447-459.
32. Mariuzza, R. A., Velikovsky, C. A., Deng, L., Xu, G., & Pancer, Z. (2010) Structural insights into the evolution of the adaptive immune system: the variable lymphocyte receptors of jawless vertebrates. *Biol Chem* **391**, 753-760.
33. Velikovsky, C. A. *et al.* (2009) Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. *Nat Struct. Mol. Biol* **16**, 725-730.
34. Seiradake, E. *et al.* (2009) Structure and functional relevance of the Slit2 homodimerization domain. *EMBO Rep.* **10**, 736-741.

35. Barton, W. A. *et al.* (2003) Structure and axon outgrowth inhibitor binding of the Nogo-66 receptor and related proteins. *EMBO J* **22**, 3291-3302.
36. Peng, W. C. *et al.* (2013) Structure of stem cell growth factor R-spondin 1 in complex with the ectodomain of its receptor LGR5. *Cell Rep.* **3**, 1885-1892.
37. Han, B. W., Herrin, B. R., Cooper, M. D., & Wilson, I. A. (2008) Antigen recognition by variable lymphocyte receptors. *Science* **321**, 1834-1837.