Supplementary Information

To accompany the manuscript "A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in *Anopheles gambiae*" by Richard H. G. Baxter, Stefanie Steinert, Yogarany Chelliah, Gloria Volohonsky, Elena A. Levashina, & Johann Deisenhofer.

Molecular weight determination of LRIM1/APL1C complex

LRIM1 and APL1C proteins were purified as described above and analyzed by size-exclusion chromatography in 0.2 M NaCl, 20 mM Na-Hepes pH 7.5, using a Superdex200 16/60 column (GE Healthcare) calibrated with a five-protein gel filtration standard (BioRad). LRIM1/APL1C- Δ 26-130 was exchanged into PBS and analyzed by DLS at 5 mg/ml, and by sedimentation velocity (SV) at 0.2–1.0 mg/ml. For the latter, the samples (~390 µl) were placed in dualsectored charcoal-filled Epon centerpieces sandwiched between sapphire windows. The samples were centrifuged at 20 °C, 45,000 rpm (An50 Ti) in a Beckman Optima XL-I analytical ultracentrifuge; the interference optical system was used to acquire the data. Data were analyzed using SEDFIT (1) using the following parameters: partial specific volume $\overline{v} = 0.73$, buffer density $\rho = 1.00523$, buffer viscosity $\eta = 0.01019$, which were estimated using SEDNTERP (2). A c(s) distribution calculated from the SV data demonstrated four peaks: a major one, comprising the LRIM1/APL1C- Δ 26-130 heterodimer, and three minor peaks, which were probably contaminants. These four sedimenting entities were treated as discrete species in a subsequent analysis, allowing for the calculation of the s-value and a 68% confidence interval for the heterodimer: $s = 5.10\pm0.04$ S. Combined with the diffusion constant $D = 3.63\pm0.03 \times 10^{-7}$ cm²/s determined by DLS, the molecular mass was calculated using the Svedberg equation $M = sRT/D(1-\bar{v}\rho) = 131\pm 1$ kDa, where R is the gas constant, and T is the temperature in Kelvin.

Monosaccharide composition analysis by HPAEC

LRIM1/APL1- Δ 26-130 (~375 µg) was hydrolyzed with 2 N trifluoroacetic acid at 100 °C for 4 h. The hydrolysate was lyophilized, redissolved in H₂O, sonicated for 7 min in ice and transferred to an injection vial. Four concentration of standard mix (0.5, 1.0, 2.0, and 4.0 nmol/injection) were prepared to establish a calibration curve. The number of moles of each sugar in the sample was quantified by linear interpolation from the calibration curve. Sugars were analyzed by HPAEC using a Dionex DX500 system equipped with a GP40 gradient pump, an ED40 electrochemical detector, and a Thermo-Separation AS3500 autosampler containing a stainless steel needle. Individual sugars were separated by a Dionex CarboPac PA20 (3 x 150 mm) analytical column with an amino trap. The gradient programs used eluents A, degassed nanopure water; B, 200 mM NaOH; and C, 100 mM NaOH. All methods were based on protocols previously described(3). Analysis of the monosaccharide composition of LRIM1/APL1C- Δ 26-130 is consistent with the empirical formula Fuc₂Nag₁₄Man₁₇ for seven *N*linked glycosylation sites with two sites α 1,3-fucosylated and an average of 2.4 mannose residues per site, with a total mass of 6.5 kDa.

Crystallization of LRIM1 and APL1 LRR domains

The LRIM1-LRR domain that was crystallized was generated by expression of a construct consisting of LRIM1 residues 1–332 with a C-terminal 6xHis tag. The APL1C-LRR domain that was crystallized was generated by limited tryptic digest of purified APL1C. Purified LRIM1-LRR was crystallized at 5 mg/ml in 15–20 %(w/v) PEG 8000, 0.1 M Na-Hepes pH 7.5, 0.2 M calcium acetate, 0–10% glycerol, 21 °C. APL1C-LRR crystallized under similar conditions. Crystals were flash-frozen in liquid nitrogen prior to data collection in a cryobuffer containing

15% glycerol. LRIM1-LRR x-ray diffraction data was collected at beamline 19-ID (SBC-CAT) at the Advanced Photon Source, Argonne National Laboratory. Selenomethionine protein suitable for phasing was obtained by introduction of four additional methionine residues (E99M, Y148M, K216M, L310M) by site-directed mutagenesis. APL1C-LRR-ray diffraction data was collected at beamline ID-14-4 at the European Synchrotron Radiation Facility, Grenoble.

Determination of Crystal Structures

LRIM1-LRR data was processed by *HKL3000* (4), 12 sites found by 2-wavelength MAD with *SHELXD* (5). Initial refinement was performed using *REFMAC5* (6) and model building with COOT (7). An initial model comprising the 10 LRR repeats was built in three rounds of manual refinement against the 2.5 Å selenomethionine dataset. Following phase extension to the 2.0 Å native dataset, *ARP/wARP* (8) built the C-terminal region in six rounds with manual editing. Final rounds of refinement including TLS refinement were performed in *PHENIX* (9).

APL1C-LRR and LRIM1/APL1C- Δ 26-130 data was processed with *XDS* (10). Initial attempts to determine the APL1C-LRR structure by molecular replacement (MR) using LRIM1-LRR as search model were unsuccessful, but a solution for the LRIM1/APL1C- Δ 26-130 complex was readily obtained. Further MR trials on the complex with other LRR domains from the PDB yielded a solution for APL1C-LRR with a chimera of two solutions using decorin (1XKU 35–252) and biglycan (2FT3 38–227) as search models. This model provided a solution for the APL1C-LRR dataset which was then built automatically by ARP/wARP in two rounds with manual editing. The final model was returned to the LRIM1/APL1C- Δ 26-130 and the remaining model built automatically by *BUCCANEER* (11) in three rounds with manual editing. Final rounds of refinement including TLS refinement were performed in *PHENIX*.

Crystal structure of LRIM1 and APL1C LRR domains

The structures of LRIM1-LRR and APL1C-LRR were determined to 2.0 Å and 1.85 Å resolution, respectively (Table S2). The final model of LRIM1-LRR contains two complete molecules in the asymmetric unit including the 6xHis tag, with a C_{α} rmsd of 0.8 Å for the LRR domain. The final model of APL1C-LRR contains one molecule in the asymmetric unit comprising residues 140–523, except for the disordered loop 147–152.

LRIM1-LRR (Fig. S3A) commences immediately after the signal peptide cleavage site, prefaced by a short antiparallel β -strand. The LRR domain contains ten repeats (eleven parallel β -strands), with two *N*-linked glycosylation sites: Asn 86 on the concave face of LRR III and Asn 132 on the bottom surface of LRR V. APL1C-LRR (Fig. S3B) begins with an N-terminal capping motif (LRRNT) including a disulfide bond between Cys 145 and Cys 154. The LRR domain contains thirteen repeats (fourteen parallel β -strands) with three of four *N*-linked glycosylation sites resolved: Asn 325 on the concave face of LRR VII, Asn 370 on the bottom face of LRR IX, and Asn 476 at the top of the final parallel β -strand (Asn 187 is resolved in the LRIM1/APL1C complex structure). After the final β -strand, LRIM1 Cys 273 at the base of the convex face commences the LRRCT motif containing two disulfide bonds, one between Cys 273 and Cys 318 and one between Cys 305 and Cys 317 (Fig. 2C). APL1C has a similar LRRCT to LRIM1 with disulfide bonds between Cys 486 and Cys 520 and between Cys 508 and Cys 519, but the length and secondary structure in the interval between the disulfide bonds is different between the two structures.

Each repeat of LRIM1 and APL1C-LRRs V–XIII has a distinct secondary structure on the convex face (APL1C LRRs I–IV are regular repeats). LRIM1 and APL1C have unusually large

radii of curvature compared to other LRR structures, 36 Å and 31 Å, respectively (Table S3). This is caused by a decrease in the length of LRRs in the center of the domain (Table S4); specifically LRIM-LRR IV and APL1C-LRR VII are only 19 residues in length, the shortest repeat in an LRR of known structure. Two *N*-linked glycosylation sites in the LRR domain, LRIM1 Asn 86 and APL1 325, and LRIM1 132 and APL1 370, are shifted by one repeat between LRIM1 and APL1C but conserved in their relative position to this short repeat.

Immunoprecipitation of the TEP1/LRIM1/APL1C ternary complex

Rabbit polyclonal antibodies were raised and affinity purified against purified TEP1r (Genscript). A final stage of purification was performed by passage over secreted *T. ni* proteins conjugated to NHS-Sepharose (GE Healthcare) to remove glycan-specific antibodies. 10 µl of protein mixtures following centrifugation was diluted to 1 ml in IP buffer (50 mM Tris-HCl pH 7.8, 100 mM NaCl, 2 mM EDTA, 0.1 µg/ml BSA, 0.1% Tween-20). IP was performed with either α FLAG-M2 agarose (Sigma) or α TEP1r bound to a mixture of protein A/protein G-sepharose (Sigma). Beads were washed twice each with 50 mM Tris-HCl pH 7.8 ± 0.5 M NaCl and eluted by incubation with 2X Laemli buffer. SDS/PAGE was run with 4-20% minigels, transfer to nitrocellulose and Western blotting with monoclonal α 6xHis/HRP (Clontech), α FLAG-M2/HRP (Sigma) or α TEP1.

Supplementary References

 Schuck P, Perugini MA, Gonzales NR, Howlett GJ, & Schubert D (2002) *Biophys J* 82, 1096-1111.

- Laue TM, Shah BD, Ridgeway RM, & Pelletier SL (1992) in *Analytical* Ultracentrifugation in Biochemistry and Polymer Science, ed. Harding SE, Rowe, A. J., and Horton, J. C (The Royal Society of Chemistry, Cambridge).
- 3. Hardy MR & Townsend RR (1994) *Methods Enzymol* 230, 208-225.
- Otwinowski Z & Minor W (1997) in *Macromolecular Crystallography*, eds. Carter J, Charles W & Sweet RM (Academic Press, New York), pp. 307-326.
- 5. Sheldrick GM (2008) *Acta Crystallogr A* **64**, 112-122.
- 6. Murshudov GM, Vagin AA, & Dodson EJ (1997) Acta Crystallographica D 53, 240-255.
- 7. Emsley P & Cowtan K (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.
- 8. Langer GG, Cohen SX, Lamzin VS, & Perrakis A (2008) *Nat Protoc* **3**, 1171-1179.
- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, & Terwilliger TC (2002) *Acta Crystallogr D Biol Crystallogr* 58, 1948-1954.
- 10. Kabsch W (1988) Journal of Applied Crystallography 21, 916-924.
- 11. Cowtan K (2006) Acta Crystallogr D Biol Crystallogr 62, 1002-1011.

Analyte	nanomoles ¹	Mass, µg	mole %
Fucose	6.68	1.10	6.2
N-acetyl galactosamine	nd	nd	nd
N-acetyl glucosamine	44.84	9.92	41.7
Galactose	nd	nd	nd
Glucose	nd	nd	nd
Mannose	55.99	10.09	52.1

Table S1. HPAEC monosaccharide composition of LR1M1/APL1.

¹Total nanomoles of each residue from the hydrolyzed aliquot (~375 μ g). nd = not detected.

Dete Cellertier (ADC)			
Data Collection (APS)	LRIMI-LRR: native	LRIMI-LKK: SeMet (peak)	LRIMI-LRR: SeMet (infl)
Energy (keV)	12659	12659	12658
Unit cell	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2
a,b(A)	79.265	78.822	78.862
<i>c</i> (A)	238.643	237.012	237.184
Mosaicity (°)	0.23-0.34	0.15-0.57	0.18-0.57
Resolution (A)	50-2.00 (2.03-2.00)	50-2.60 (2.64-2.60)	50-2.60 (2.64-2.60)
Observations	396443	179341	173397
Unique reflections	52399	23950	24003
Redundancy	7.6 (4.4)	7.5 (6.8)	7.2 (5.8)
Completeness (%)	99.5 (95.1)	99.7 (99.4)	99.7 (99.3)
$\langle I \rangle / \langle \sigma \rangle$	37.6 (2.1)	32.4 (4.1)	29.4 (2.5)
$\langle I \rangle / \langle \sigma \rangle > 3 \ (\%)$	79.1 (36.0)	82.5 (48.4)	77.0 (34.2)
$R_{\rm sym}$, (%)	6.7 (65.7)	7.6 (58.2)	8.0 (82.3)
Data Collection (ESRF)		APL1-LRR:native	LRIM1/APL1: native
Energy (keV)		13197	13197
Unit cell $(P4_12_12)$		$P2_{1}2_{1}2_{1}$	$P2_{1}2_{1}2_{1}$
<i>a,b</i> (Å)		38.77, 70.35	109.89, 110.88
c (Å)		161.76	168.32
Mosaicity (°)		0.15-0.57	
Resolution (Å)		50-1.85 (1.90-1.85)	50-2.60 (2.64-2.60)
Observations		273112	233585
Unique reflections		38500	56906
Redundancy		7.1 (6.2)	4.1 (4.2)
Completeness (%)		99.1 (97.2)	99.4 (99.7)
$\langle I \rangle / \langle \sigma \rangle$		22.9 (7.3)	7.4 (2.1)
$\langle I \rangle / \langle \sigma \rangle > 3 (\%)$		88.2 (60.6)	65 9 (39 7)
$\frac{1}{\sqrt{0}}$		55.2(00.0)	14.9(71.5)
M_{sym} , (70)		5.5 (22.7)	14.9 (71.5)
Refinement	LRIM1-LRR	APL1-LRR	LRIM1/APL1
Resolution (Å)	47.7-2.00 (2.03-2.00)	42.8-1.85 (1.90-1.85)	45.6-2.70 (2.74-2.70)
Working set	49666 (2448)	36543 (2349)	53959 (2445)
Test set	2665 (140)	1924 (131)	2880 (115)
$R_{\rm cryst}$ (%)	18.7 (25.9)	16.6 (18.3)	21.4 (26.6)
$R_{\text{frag}}(\%)$	23.3(30.9)	20.6 (26.7)	26.0 (33.1)
Est coord error (Å)	0.009	0.01	017
Est phase error (°)	22.5	20.2	25.3
rmsd from ideal			
Bond lengths (Å)	0.005	0.007	0.012
Bond angles (°)	0.782	0.848	1 159
Wilson B-factor $(Å^2)$	33.8	19.8	51 5
Atoms	0.00	17.0	51.5
non-hydrogen	5575	3501	8861
water	<i>4</i> 91	290	446
Walter Malprobity analysis	791	230	770
All store slash says	5 56	2 22	7.44
All-atom clash score	3.30	3.23	/.44
Bad rotamers (%)	0.5 (3/554)	0.3 (1/354)	1.0 (13/913)
Ramachandran plot			0.1.(1/100.0)
outliers (%)	0.3 (2/626)	0.0 (0/374)	0.1 (1/1006)
tavored (%)	92.5 (579/626)	94.9 (355/374)	92.4 (930/1006)

 Table S2: X-ray data collection and refinement statistics

PDB	PROTEIN NAME	UNIPROT	SPECIES	R
ID 20147		D05146	T , • .	(A)
20MZ	INTERNALIN A	P25146	Lysteria monocytogenes	21.4
IWWL	CD14_MOUSE	P10810	Mus musculus	22.1
3E6J	VARIABLE LYMPHOCYTE RECEPTOR RBC36		Petromyzon marinus	22.6
1GWB	GPIBA_HUMAN	P07359	Homo sapiens	22.7
2ID5	LINGO-1 ECTODOMAIN (Q96FE5_HUMAN)	Q96FE5	Homo sapiens	23
206R	VARIABLE LYMPHOCYTE RECEPTOR B61	Q4G1L2	Eptatretus burgeri	23.2
2Z80	TOLL-LIKE RECEPTOR 2	O60603	Homo sapiens	23.3
2Z64	TOLL-LIKE RECEPTOR 4	Q9QUK6	Mus musculus	23.3
10GQ	POLYGALACTURONASE INHIBITING PROTEIN	P5882	Phaseolus vulgaris	23.5
2Z7X	TLR1	Q15399	Homo sapiens	23.5
106S	INTERNALIN A	P25146	Lysteria monocytogenes	23.6
2Z81	TOLL-LIKE RECEPTOR 2	QUQ9N7	Mus musculus	24
3FXI	TOLL-LIKE RECEPTOR 4	O00206	Homo sapiens	24.2
206S	VARIABLE LYMPHOCYTE RECEPTOR B59	Q4G1L3	Eptatretus burgeri	24.8
1XEU	INTERNALIN C	Q8Y6A8	Lysteria monocytogenes	25.2
2Z66	TOLL-LIKE RECEPTOR 4	O00206	Homo sapiens	25.3
2FT3	BIGLYCAN (PGS1_BOVIN)	P21809	Bos taurus	25.5
2R9U	VARIABLE LYMPHOCYTE RECEPTOR 2913	Q6E4B4	Petromyzon marinus	26.3
1ZIW	TOLL-LIKE RECEPTOR 3	O15455	Homo sapien	26.6
1G9U	YOPM YERPE	P17778	Yersinia pestis	27.6
3BZ5	INTERNALIN J	Q8Y3L4	L. monocytogenes	27.7
206Q	VARIABLE LYMPHOCYTE RECEPTOR A29	Q32R86	Eptatretus burgeri	28.9
2V70	SLIT HOMOLOG 2 PROTEIN, LRR DOMAIN 3	094813	Homo sapiens	29.2
1XCD	DECORIN (PGS2 BOVIN)	P21793	Bos taurus	29.4
1W8A	SLIT PROTEIN, LRR DOMAIN 3	P24014	Drosophila	29.9
			melanogaster	
	APL1 LRR DOMAIN		Anopheles gambiae	30.9
2HR7	INSULIN RECEPTOR	P06213	Homo sapiens	31.8
	LRIM1 LRR DOMAIN		Anopheles gambiae	36
1XWD	FOLLICLE STIMULATING HORMONE	P23945	Homo sapiens	37.1
	RECEPTOR		1	
1N8Z	ERB2_HUMAN	P04626	Homo sapiens	60.3

Table S3: Radius of curvature for extracellular LRRs of known structure

Table S4: LRIM1-LRR and APL1-LRR sequence¹

LRIM1	Concave surface	Convex surface	Length
LRRNT		AIHEIKQN	
LRR I	³¹ GNR Y K I EK V TD	SS l KQA l AS l RQSAWN	27
LRR II	⁵⁸ VKELDLSGNPL	SQ i saad l ap f tk	24
LRR III	⁸² LELLNLSSNVL	YET L D L ES L ST	22
LRR IV	¹⁰⁴ LRTLDLNNNYV	QE l L V GPS	19
LRR V	¹²³ IETLH A ANNNI	SR V S <mark>C</mark> SRGQGK	22
LRR VI	¹⁴⁵ K-NIYLANNKI	TM l rd l deg <mark>C</mark> rsr	23
LRR VII	¹⁶⁸ VQYLDLKLNEI	DT V N F AE l AASSDT	25
LRR VIII	¹⁹³ lehlnlqynfi	YD V KGQ V V F AK	22
LRR IX	²¹⁵ lktldlssnkl	AF M GPE F QS A AG	23
LRR X	²³⁸ VTWISLRNNKL	VL i eka l r f SQN	23
LRRCT	²⁶¹ lehfdlrgngf	H C GTLRDFFSKNQRVQTVAK	
		QTVKKLTGQNEE <mark>C</mark> TVPTLGH	
		YGAY <mark>CC</mark> EDL	
C-helix		PAPFADRLIAL	

APL1	Concave surface	Convex surface	Length
LRRNT		QPEYK C IDSNLQYD	
	C V F YD V H I	DMQTQD \mathbf{V} Y \mathbf{F} GFEDITLNN	
LRR I	¹⁸⁰ QKIVTFKNSTM	rk l paa ll ds f rq	24
LRR II	²⁰⁴ VELLNLNDLQI	EE I DTY AF AY A HT	24
LRR III	²²⁸ IQKLY M GFNAI	RY l pph vf QN v pl	24
LRR IV	²⁵² LTVLVLERNDL	SS l prg if hn t pk	24
LRR V	²⁷⁶ LTTLSMSNNNL	ER i edd tf Qa t ts	24
LRR VI	³⁰⁰ LQNLQLSS N RL	TH V D L SL I PS	21
LRR VII	³²¹ LFHANVSYNLL	ST l a ip ia	19
LRR VIII	³⁴⁰ VEELDASHNSI	NV V RGPV N VE	21
LRR IX	³⁶¹ LTILKLQH N NL	TD T A WL LN Y PG	22
LRR X	³⁸³ LVEVDLSYNEL	EK i myhp f vk m qr	24
LRR XI	407 l er l y i sn n rl	VALNLYGQPIPT	23
LRR XII	⁴³⁰ lkvldlshnhl	LH v er n qp qf dr	23
LRR XIII	⁴⁵³ LENLYLDHNSI	VT L KLSTHHT	21
LRRCT	⁴⁷⁴ lknltlshndw	DCNSLRALFRNVARPAVDDA	
		DQH <mark>C</mark> KIDYQLEHGL <mark>CC</mark> KES	

¹Residues in bold face lie upon the inside of the LRR solenoid. N-linked glycosylation sites are colored in green, disulfide-linked cysteines are colored in red.



Figure S1: Analytical ultracentrifugation (AUC) of the LRIM1/APL1C- Δ 26-130 complex. The complex sedimented as a single species with a sedimentation coefficient of *s* = 5.10±0.04 S at 20 °C in PBS. (Inset) Dynamic light scattering (DLS) of LRIM1/APL1C- Δ 26-130 complex, autocorrelation coefficient versus log₁₀(*t*). Fitted curve corresponds to a single species (~5% polydispersity) with diffusion coefficient *D* = 3.6×10⁻⁷ cm²/s.



Figure S2: (A) Structure of LRIM1-LRR and (B) structure of APL1C-LRR. N- and C-termini are labeled; concave β-strands are numbered according to Table S4. Ribbons colored according to secondary structure: α -helices in red, β -strands in green, 3_{10} -helices in yellow, loops in grey. *N*-acetylglucosamines in LRIM1-LRR are shown linked to Asn 86 (LRR III) and Asn 132 (behind LRR V). N-acetylglucosamines in APL1C-LRR are shown linked to Asn 325 (LRR VII), Asn 370 (behind LRR IX) and Asn 476 (LRR XIV).



Figure S3: (A) Precipitate (p) and soluble (s) fractions for the following reactions after 12 h incubation at room temperature: (a) TEP1 + LRIM1/APL1, (b) TEP1(MeNH₂) + LRIM1/APL1, (c) TEP1_{cut} + LRIM1/APL1, (d) TEP1_{cut}(MeNH₂) + LRIM1/APL1, (e) TEP1_{cut}(MeNH₂) + LRIM1-LRR + APL1-LRR, (f) TEP1_{cut}(MeNH₂) + BSA. (B) Elution (d) TEP1_{cut}(MeNH₂) + LRIM1/APL1 on Superose6 10/30 size-exclusion column (GE Healthcare), 0.5 ml fractions. Silver-stained SDS-PAGE gel of selected fractions for TEP1/LRIM1/APL1 ternary complex (peak 3-5) and LRIM1/APL1 heterodimer (peak 14-16). The equivalent staining of bands in the ternary complex suggests a 1:1 ratio of TEP1 to LRIM1/APL1.



Figure S4: Immunoprecipitation (IP) of TEP1/LRIM1/APL1C- Δ 26-130 samples (a-d) (see text). (A) A schematic diagram illustrates the antigens present in solution before IP (bef); α FLAG recognizes and IPs LRIM1 and co-IPs APL1C. α TEP1 recognizes and IPs full-length TEP1 (TEP1-F) and TEP1_{cut} fragments TEP1-N and TEP1-C. α 6xHis recognizes TEP1-F, TEP1-C and APL1C. (B) For α TEP1 IP of (a) TEP1, (b) TEP1(MeNH₂) and (c) TEP1_{cut}, LRIM1 and APL1C remain in the supernatant, but (d) TEP1_{cut}(MeNH₂) co-IPs LRIM1/APL1C. (C) For α FLAG IP of LRIM1/APL1C, (a) TEP1, (b) TEP1(MeNH₂) and (c) TEP1_{cut} remain in the supernatant (su) while (d) TEP1_{cut}(MeNH₂) co-IPs with LRIM1/APL1C.