Published in final edited form as:

Science. 2008 September 26; 321(5897): 1834–1837. doi:10.1126/science.1162484.

Antigen Recognition by Variable Lymphocyte Receptors

Byung Woo Han^{1,2}, Brantley R. Herrin³, Max D. Cooper³, and Ian A. Wilson^{1,2,*}

¹*Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.*

²*The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.*

³*Emory Vaccine Center, and Department of Pathology and Laboratory Medicine, Emory University, 1462 Clifton Road, N.E., Atlanta, GA 30322, USA.*

summary

The Variable Lymphocyte Receptor (VLR) recognizes its antigen using the concave surface of the leucine-rich repeat solenoid structure and a highly variable insert in the LRRCT module.

Variable lymphocyte receptors (VLRs) rather than antibodies play the primary role in recognition of antigens in the adaptive immune system of jawless vertebrates. Combinatorial assembly of leucine-rich repeat (LRR) gene segments achieves the required repertoire for antigen recognition. We have determined a crystal structure for a VLR-antigen complex, the VLR RBC36 in complex with the H antigen trisaccharide from human blood type O erythrocytes at 1.67 Å resolution. RBC36 binds the H-trisaccharide on the concave surface of the LRR modules of the solenoid structure where three key hydrophilic residues, multiple van der Waals interactions, and the highly variable insert of the C-terminal LRR module play the major role in antigen recognition and specificity. The concave surface, which is assembled from the most highly variable regions of the LRRs, along with diversity in the sequence and length of the highly variable insert, can account for the recognition of diverse antigens by VLRs.

In lamprey and hagfish, the only surviving jawless vertebrates, variable lymphocyte receptors (VLRs) play the major role in recognition of foreign antigens (1,2). In contrast to the variable, diverse, and joining gene segments (VDJs) of immunoglobulins in jawed vertebrates, the jawless vertebrates have solved the receptor diversity problem by somatic DNA rearrangement of diverse leucine-rich repeat (LRR) modules into incomplete *vlr* genes. The resulting mature *vlr* genes encode an N-terminal LRR capping region (LRRNT), the first LRR (LRR1), up to seven 24-residue variable LRRs (LRRVs) (3), a terminal or end LRRV (LRRVe), a connecting peptide (CP), a C-terminal LRR capping region (LRRCT), and a threonine/proline-rich stalk region which connects the protein to a GPI anchor and a hydrophobic tail (Fig. 1A) (1,2,4).

From these somatic gene rearrangements, a potential repertoire of about 10^{14} unique VLRs has been estimated (2) that compares favorably with the equivalent diversity attainable through VDJ recombination in antibodies. Different numbers and combinations of LRR modules, coupled with amino acid sequence variation in the LRR segments, thereby contribute to VLR

*To whom correspondence should be addressed. E-mail: wilson@scripps.edu

Supporting Online Material

Materials and Methods

Figs. S1 to S4

Tables S1 to S2

References

diversity. The LRR repeats form a curved solenoid, as in Toll-like receptors (TLRs) (5,6), and its concave surface has been suggested as the antigen-binding site from evolutionary, sequence and mutational analyses (2,7,8). Crystal structures of three unliganded hagfish VLRs with different numbers of LRRV modules have been determined (7), whereas antigen-binding specificity [erythrocyte H-trisaccharide (9) and *Bacillus* collagen-like protein of *anthracis* (BclA) (8)] has been reported only for lamprey VLRs. However, the mode of antigen recognition has not yet been determined in either system nor has it been shown whether complementarity determining region (CDR) equivalents are present in VLRs that would endow them with specificity and affinity for any given antigen, as for antibodies.

We, therefore, determined the crystal structure of the VLR RBC36 ectodomain (ECD) in complex with the H-trisaccharide derived from the H-antigen of human blood group O erythrocytes at 1.67 Å resolution by molecular replacement using our lamprey VLR2913 crystal structure (unpublished, Protein Data Bank (PDB) ID: 2R9U). Lampreys were previously shown to produce high titer agglutinins against the H-antigens of human O erythrocytes (10,11). Recently, when lampreys were immunized with human blood group O erythrocytes, they elicited VLRs that recognize the dominant H-trisaccharide antigen on CHO cells transfected with 1,2-fucosyltransferase (9). H antigens contain the characteristic disaccharide α -L-Fucp-(1→2)- β -D-Galp-OR, where R is glycoprotein or glycolipid (12). The type II H antigen trisaccharide, α -L-Fucp-(1→2)- β -D-Galp-(1→4)- β -D-GlcNacp-OH, was used as the antigen in the crystal structure with the RBC36-ECD (Fig. 1, 2A-B, and fig. S1) (13).

Lamprey RBC36-ECD (residues 22–238, lacking the N-terminal signal sequence) forms a horseshoe-shaped assembly, that is more abbreviated and crescent-shaped compared to TLRs, and consists of an LRRNT, an 18-residue LRR1, three LRRVs, an LRRVe, a CP, and an LRRCT, that all adopt a right-handed solenoidal structure except for LRRCT. The inner concave surface is formed from eight β -strands (two from LRRNT, five from LRRs, and one from CP), which assemble into a continuous β -sheet. The convex (outer) surface is composed of the more diverse secondary structure elements, including different length loops, one α -helix, and six 3_{10} -helices (Fig. 1). Lamprey RBC36-ECD contains five canonical LRR-signature motifs, $xL^2xxL^5xxL^8xL^{10}xxN^{13}Q^{14}L^{15}xxL^{18}p^{19}xG^{21}V^{22}F^{23}D^{24}$, where L represents obligate hydrophobic residues which, for RBC36-ECD, include leucine (most prevalent), isoleucine, or methionine; and N, Q, P, G, F, and D are conserved asparagine, glutamine, proline, glycine, phenylalanine, and aspartic acid residues respectively (Fig. 2D) (4,14). The side chains of nine conserved residues in each LRR (at relative positions 2, 5, 8, 10, 13, 15, 18, 22, and 23) assemble within the solenoidal structure, and form a tight hydrophobic core which laterally stabilizes the repeating LRR modules (Fig. 2E). In RBC36-ECD, LRRNT comprises residues 22 to 52, in which Thr³², Val³³, and Asp³⁴ initiate an antiparallel β -strand that extends the continuous parallel β -sheet, and LRRNT and LRRCT cover the exposed edges of the hydrophobic core of the solenoidal LRR structure, as observed in other LLR proteins, including TLRs (5,6). In LRRNT, the characteristic four-cysteine motif (Cx_nCxCx_nC) forms two sets of disulfides: Cys²²-Cys²⁸ and Cys²⁶–Cys³⁵, whereas, in LRRCT, a similar motif (Cx_nCx_nCx_nC) gives rise to disulfides Cys¹⁸²-Cys²¹⁷ and Cys¹⁸⁴–Cys²³⁷ (Fig. 1).

After accounting for the protein, extra electron density on the concave surface remained which corresponded to the H-trisaccharide antigen (fig. S2). Specificity between RBC36 and H-trisaccharide is mainly mediated by four hydrogen bonds (Fig. 2A) on the inner concave surface, between Asp¹⁰³ O^{D2} and N-acetylglucosamine O^{AZ}, Asp¹⁵² O^{D1} and galactose O^{4'}, Asp¹⁵² O^{D2} and galactose O^{3'}, and Gln¹⁵³ N^{E2} and fucose O^{AL} (Fig. 2A-B). Asp¹⁰³ is located on LRRV2, and Asp¹⁵² and Gln¹⁵³ on LRRVe.

With the concave surface of RBC36 firmly established as the antigen binding site, we analyzed the variability in amino acids represented on this surface in other VLRs. From BLASTP

searches (15) with RBC36, 24 VLR sequences were found with three LRRVs and sequence identity >60 %. The amino acid variation was significantly higher on the concave surface of each LRR module (Fig. 2C, fig. S3, and table S2) and, hence, to some extent, is analogous to the hypervariable regions (CDRs) in antibodies. In each canonical 24-residue LRR module, seven residues, $\text{xxL}^8\text{xL}^{10}\text{xx}$, are located on the concave surface and, of these, only the two obligate hydrophobic residues (L) face inwards to form the hydrophobic core of the solenoidal structure (Fig. 2D-E). Consequently, the other five residues could potentially contribute to antigen recognition, and correspond to the 1st, 2nd, 4th, 6th, and 7th positions of this seven residue segment in each LRR. Asp¹⁰³, Asp¹⁵², and Gln¹⁵³ which contribute significantly to the interaction between RBC36 and H-trisaccharide represent the 4th residue of the LRRV2 and 6th and 7th residues of the LRRV6 concave surfaces. Eight other residues on the concave surface (His⁵⁷, Tyr⁷⁹, Thr¹⁰⁶, Phe¹²⁷, Cys¹²⁹, Ala¹⁵⁰, Tyr¹⁷⁴, and Phe¹⁷⁶) stabilize the H-trisaccharide interaction via 16 van der Waals contacts as calculated with CONTACSYM (16) (Fig. 4B). The carbohydrate antigen buries $\sim 303 \text{ \AA}^2$ on the VLR, whereas the corresponding buried surface on the antigen is $\sim 246 \text{ \AA}^2$ calculated with a $1.4 \text{ \AA}$ probe radius (17,18), which is comparable to buried surfaces of haptens ($\sim 150\text{-}350 \text{ \AA}^2$) with antibodies (19).

Another key interaction with the H-trisaccharide is between Trp²⁰⁴ and the galactose. The Trp²⁰⁴ indole is stacked parallel to the sugar ring, as observed as a common feature in other sugar-protein complexes (Fig. 1 and 2A) (20). Trp²⁰⁴ is located in the middle of LRRCT, where the VLR sequences are extremely diverse and a highly variable insert is often present (21). The variability plot (Fig. 2C) illustrates that highly variable inserts of 2-12 residues occur in LRRCT (fig. S3). In RBC36, a 10-residue insert forms a β -hairpin and Trp²⁰⁴ is located at the end of the first β -strand, prior to the β -hairpin turn (Fig. 1 and 2A). Superposition of the crystal structures of lamprey RBC36 and the three hagfish VLRs reveals not only high sequence variability, but also secondary structure variation in their inserts. The VLRB.59 8-residue insert is a loop, but similar in overall shape to the RBC36 β -hairpin, while the VLRA.29 3-residue insert points towards the horseshoe side rather than its concave surface; VLRB.61 has no insert (Fig. 3A).

The RBC36 insert lies in close proximity to the concave β -sheet, and its overall conformation is remarkably similar to the β -switch in the C-terminal flank region of human Glycoprotein Ib α (GpIb α) that interacts with von Willebrand factor (VWF) A1 domain (22) (Fig. 3B). A search for structural homologs of RBC36, using the DALI server (23), selected GpIb α (PDB ID: 1M10) as one of the top three hits along with hagfish VLRB.59 (PDB ID: 2O6S) and human Slit protein (PDB ID: 2V9T), all of which are LRR-containing proteins. Interestingly, GpIb α and Slit also have crystal structures with their binding partners, in which their mode of interaction is very similar to the RBC36-H-trisaccharide complex where the same 1st, 2nd, 4th, 6th, and 7th residues on the concave surface in each LRR module play the key role in ligand binding, but the insert in the Slit LRRCT does not contact its ligand (Fig. 3B and fig. S4). Considering that the secondary structures of the highly variable inserts of VLRs in the PDB are diverse (Fig. 3A), and the equivalent β -switch of human GpIb α adopts a loop structure when GpIb α is crystallized by itself (22), these structure differences may also play an important role in antigen selection, recognition and affinity. Conformational changes or isomerism in the inserts would also increase possible binding modes in much the same way as induced fit in the CDR loops of antibodies, especially for CDR H3 (24), or equivalent loops ($\alpha 3$, $\beta 3$) in TCRs (25).

The concave molecular surface area of RBC36 is estimated to be $\sim 1720 \text{ \AA}^2$ calculated with a $1.4 \text{ \AA}$ probe radius (17,18), compared with a buried surface of $\sim 700\text{-}1000 \text{ \AA}^2$ on average for the antigen-binding regions of immunoglobulins for proteins and other large antigens. Considering that the number of LRRVs in VLRs can be as many as seven and the concave

surface area of one LRRV is about $220 \text{ \AA}^2$, the total antigen-binding surface could extend to around $2600 \text{ \AA}^2$ and could potentially accommodate binding sites for multiple antigens. However, this size could well be an overestimate as were the corresponding early predictions for the antigen-binding surfaces of antibodies (26), although the possibility of multiple paratopes in VLRs is certainly an intriguing concept.

With few exceptions, five residues on the concave surface of each LRR module are available for antigen binding (Fig. 2E and 4A). Hence, we mapped these key residues for antigen recognition onto a coordinate system that corresponds to the LRR modules along the x-axis and key residue positions on the concave surface along the y-axis (Fig. 4B). This interaction matrix should be useful when other VLR-antigen complexes are determined.

Antigen recognition by antibodies in the vertebrate immune system is well documented and has revealed how the immunoglobulin fold with its CDR loops can form a high affinity binding site for virtually any antigen it encounters, be it natural or synthetic (24). Similarities and differences can now be assessed for antigen recognition by the immunoglobulin fold and the LRRs of VLRs, as well as TLRs. High sequence variability in the immunoglobulin fold is concentrated in CDRs H1, H2, H3, L1, L2, and L3, whereas that of VLRs is confined to the concave surface of each LRR module (Fig. 2C). In the Ig fold, not only variability in amino acid composition, but also insertions in the CDRs, especially in CDR H3 (27), ensure a wide range of specificities and affinities for the antibody combining sites. However, VLRs contain few insertions on the concave surface of each LRR module, but diversity is attained by variation in the number (up to seven) and amino acid composition of the LRRV modules. The only significant insertion in VLRs is observed in the middle of LRRCT which shows highly variable amino acid composition, length, and secondary structure. To what extent the highly variable insert in LRRCT contributes to the specificity, affinity, and shape of the antigen binding site of VLRs awaits further VLR-antigen complex structures.

Recently, three crystal structures of TLR-ligand complexes, TLR4-MD2-Eritoran (28), TLR1-TLR2-lipopeptide (29), and TLR3-dsRNA (30), have been determined. So far, only the binding mode of the TLR4-MD2 complex is similar to that of antigen recognition by VLRs in that residues on the concave surface of the N-terminal and central domains of TLR4 interact with MD2. However, no interaction is seen between LRRCT of TLR4 and MD2 as observed between the highly variable insert in LRRCT of VLRs and antigens. Since we do not have sufficient VLR and TLR complex structures to yet make statistically significant conclusions, and the number of LRR modules in TLRs is much greater than in VLRs, it may be still too early to infer evolutionary relationships between VLRs and TLRs.

In conclusion, the crystal structure of RBC36-ECD in complex with the H-trisaccharide has provided structural insight into how VLRs recognize their antigens and provides a basis for rational design and modification of other antigen-specific VLRs. Importantly, this VLR-antigen structure sheds light into the adaption and evolution of primordial LRR proteins into their more specialized roles in pathogen recognition (e.g. TLRs) by the mammalian innate immune system.

Acknowledgements

We thank R.L. Stanfield, X. Dai, and X. Zhu for help with data collection and analysis, M.N. Alder for help with VLR cDNA library preparation, J. Paulson and R. McBride for helpful comments and suggestions, and J. Vahnasy, B. Droese, and H.-J. Kim for technical support and advice. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. We gratefully acknowledge support from NIH grants AI42266 (I.A.W.), AI072435 (M.D.C.), the Georgia Research Alliance (M.D.C.), and the Skaggs Institute for Chemical Biology (I.A.W.). This is manuscript no. 19581-MB from the Scripps Research Institute. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code 3E6J.

References and Notes

1. Pancer Z, et al. *Nature* 2004;430:174. [PubMed: 15241406]
2. Alder MN, et al. *Science* 2005;310:1970. [PubMed: 16373579]
3. When the distribution of LRRV modules per transcript was analyzed from 517 unique VLR sequences, VLRs have an average of 1.31 LRRV modules between the canonical LRR1 and LRRVe modules (ref. 2): 109 VLRs (0 LRRV), 228 VLRs (1 LRRV), 119 VLRs (2 LRRVs), 45 VLRs (3 LRRVs), 6 VLRs (4 LRRVs), 8 VLRs (5 LRRVs), 1 VLR (6 LRRVs), and 1 VLR (7 LRRVs).
4. Pancer Z, et al. *Proc. Natl. Acad. Sci. USA* 2005;102:9224. [PubMed: 15964979]
5. Choe J, Kelker MS, Wilson IA. *Science* 2005;309:581. [PubMed: 15961631]
6. Bell JK, et al. *Proc. Natl. Acad. Sci. USA* 2005;102:10976. [PubMed: 16043704]
7. Kim HM, et al. *J. Biol. Chem* 2007;282:6726. [PubMed: 17192264]
8. Herrin BR, et al. *Proc. Natl. Acad. Sci. USA* 2008;105:2040. [PubMed: 18238899]
9. Alder MN, et al. *Nat. Immunol* 2008;9:319. [PubMed: 18246071]
10. Boffa GA, Fine JM, Drilhon A, Amouch P. *Nature* 1967;214:700. [PubMed: 4167424]
11. Pollara B, Litman GW, Finstad J, Howell J, Good RA. *J. Immunol* 1970;105:738. [PubMed: 4195499]
12. At least four sub-types of H antigens are known which can be further converted into the A-antigen of human blood group A and B-antigen of human blood group B by glycosyltransferases A and B, respectively (ref. 31, fig S1).
13. Materials and methods are available as supporting material on *Science* Online
14. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
15. Altschul SF, et al. *Nucleic Acids Res* 1997;25:3389. [PubMed: 9254694]
16. Sheriff S, Hendrickson WA, Smith JL. *J. Mol. Biol* 1987;197:273. [PubMed: 3681996]
17. Gelin BR, Karplus M. *Biochemistry* 1979;18:1256. [PubMed: 4271111]
18. Connolly ML. *J. Mol. Graphics* 1993;11:139.
19. Wilson IA, Stanfield RL. *Curr. Opin. Struct. Biol* 1993;3:113.
20. Aromatic residues, such as Trp and Tyr, are frequently observed in the protein sugar-binding sites. For example, in the maltose/malto oligosaccharide-binding protein (ref. 32), aromatic residues play a major role in carbohydrate recognition and binding by providing not only hydrophobic stacking interactions between aromatic residues and the hydrophobic face of the sugar rings, but also polar contacts to the sugar hydroxyl groups.
21. Rogozin IB, et al. *Nat. Immunol* 2007;8:647. [PubMed: 17468760]
22. Huizinga EG, et al. *Science* 2002;297:1176. [PubMed: 12183630]
23. Holm L, Sander C. *J. Mol. Biol* 1993;233:123. [PubMed: 8377180]
24. Wilson IA, Stanfield RL. *Curr. Opin. Struct. Biol* 1994;4:857. [PubMed: 7536111]
25. Garcia KC, Teyton L, Wilson IA. *Annu. Rev. Immunol* 1999;17:369. [PubMed: 10358763]
26. von Gara P. *Z. Immunitäts* 1931;71:1.
27. Wu TT, Johnson G, Kabat EA. *Proteins* 1993;16:1. [PubMed: 8497480]
28. Kim HM, et al. *Cell* 2007;130:906. [PubMed: 17803912]
29. Jin MS, et al. *Cell* 2007;130:1071. [PubMed: 17889651]
30. Liu L, et al. *Science* 2008;320:379. [PubMed: 18420935]
31. Yamamoto F, McNeill PD, Hakomori S. *Biochem. Biophys. Res. Commun* 1992;187:366. [PubMed: 1520322]
32. Duan X, Quiocho FA. *Biochemistry* 2002;41:706. [PubMed: 11790091]
33. Molecular Operating Environment (MOE), version 2000.09. Chemical Computing Group; Quebec, Canada: 2003.

Fig. 1.

Overall architecture of the VLR RBC36-ECD in complex with the H-trisaccharide. **(A)** Schematic diagram of RBC36. Regions from left to right: signal peptide (SP), N-terminal LRR (LRRNT), five LRRs, connecting peptide (CP), C-terminal LRR (LRRCT), threonine/proline-rich stalk region, GPI anchor, and hydrophobic tail. **(B)** Ribbon diagram of RBC36-ECD in complex with H-trisaccharide. LRRNT, LRRs, and LRRCT are colored in blue, green, and red, respectively. Carbons, nitrogens, and oxygens of the H-trisaccharide are colored in yellow, blue, and red, respectively. Disulfide bridges are shown in orange. Green dotted lines represent hydrogen bonds, and black dotted lines indicate hydrophobic effects. **(C)** View rotated 90° from (B) that highlights the continuous β -sheet and the H-trisaccharide binding site on the concave surface.

Fig. 2.

The H-trisaccharide binding site of the RBC36. **(A)** VLR residues involved in recognizing the H-trisaccharide. After refinement, a 2Fo-Fc electron density map was calculated and contoured at 2σ as a blue mesh around the H-trisaccharide. Colors are as in Fig. 1. **(B)** H-trisaccharide interaction with RBC36 including solvent molecules, modified from the ligand interaction calculation by the program MOE 2000.09 (33). O in a circle represents waters. Hydrogen bonds with RBC36 residues and solvent molecules are drawn with green and pale green lines, respectively. Trp²⁰⁴ which is important for stabilizing the galactose via hydrophobic and stacking effects is shown in a green circle beside the galactose sugar ring. **(C)** Sequence variability plot for amino acid residues from LRRNT to LRRCT of known VLRs. Green bars on the bottom represent residues on the concave surface, and the red bar in the LRRCT shows the location of the highly variable insert. **(D)** Sequence alignment of LRR modules of RBC36. The green bar shows the residues on the concave surface. Blue and red star symbols (*) represent residues forming the β -sheet, and side chains that face the concave surface, respectively. Letters on yellow, blue, and red backgrounds show conserved hydrophobic residues, asparagine residues, and residues in the highly variable insert. Key residues on the concave surface (Asp¹⁰², Asp¹⁵², and Gln¹⁵³) for the H-trisaccharide interaction are shown as red letters and Trp²⁰⁴ in the highly variable insert is indicated by a black star symbol on the left. **(E)** The conformation of the LRRVe module highlights the tight packing of the conserved hydrophobic residues with their van der Waals radii outlined in dots. The seven residues that form the concave surface are numbered from N- to C-terminus of the LRR.

Fig. 3. Highly variable inserts of VLRs. **(A)** Crystal structures of lamprey RBC36 and three hagfish VLRs are superposed together. Ribbon diagrams are used for different VLRs; RBC36 in green, VLRA.29 in blue (PDB ID: 2O6Q), VLRB.59 in orange (PDB ID: 2O6S), and VLRB.61 in magenta (PDB ID: 2O6R), respectively. Highly variable inserts are drawn in cartoon presentation. **(B)** Superposition of RBC36-H-trisaccharide complex and GpIbα-VWF A1 domain complex (PDB ID: 1M10). Overall RBC36-ECD structure is rotated 180° vertically from (A) to highlight the comparison of the highly variable insert of RBC36 and the β-switch of GpIbα. RBC36 is depicted as a green ribbon, GpIbα as orange ribbon, VWF A1 domain as

a surface presentation in cyan, and the H-trisaccharide as in Fig. 1. The β -hairpin of the highly variable insert of RBC36 and the β -switch of GpIb α are shown in cartoon presentation.

Fig. 4.

The Interaction matrix of VLRs. **(A)** Seven residues on the concave surface of each LRR modules are shown as main-chain stick model in the same view with Fig. 1C, and the Cα atoms are connected by black lines. Five residues of the seven on the concave face (1, 2, 4, 6, and 7), which are available for the antigen recognition are connected by black lines, and 3rd and 5th residues which face inwards to the hydrophobic core are connected by yellow lines. **(B)** The simplified interaction matrix of (A). Residues involved in hydrogen bonds or in van der Waals contacts are labeled in a circle or in a square, respectively.