Published in final edited form as:

Nat Struct Mol Biol. 2009 July ; 16(7): 725–730. doi:10.1038/nsmb.1619.

Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen

C Alejandro Velikovsky[1,4], **Lu Deng**[1,4], **Satoshi Tasumi**[2], **Lakshminarayan M Iyer**[3], **Melissa C Kerzic**[1], **L Aravind**[3], **Zeev Pancer**[2], and **Roy A Mariuzza**[1]

[1]Center for Advanced Research in Biotechnology, WM Keck Laboratory for Structural Biology, University of Maryland Biotechnology Institute, Rockville, Maryland 20850, USA

[2]Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, Maryland 21202 USA

[3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894 USA

Abstract

Variable lymphocyte receptors (VLRs) are leucine-rich repeat (LRR) proteins that mediate adaptive immunity in jawless vertebrates. VLRs are fundamentally different from the antibodies of jawed vertebrates, which consist of immunoglobulin (Ig) domains. We determined the structure of an anti-hen egg white lysozyme (HEL) VLRB, isolated by yeast display, bound to HEL. The VLR, whose affinity resembles that of IgM antibodies, uses nearly all its concave surface to bind the protein, in addition to a loop that penetrates into the enzyme active site. The VLR-HEL structure, combined with sequence analysis, revealed an almost perfect match between ligand-contacting positions and positions with highest sequence diversity. Thus, we have defined the generalized antigen-binding site of VLRs. We further demonstrated that VLRs can be affinity-matured to affinities as high as those of IgG antibodies, making VLRs potential alternatives to antibodies for biotechnology applications.

INTRODUCTION

Adaptive immunity in jawless vertebrates (cyclostomes) is mediated by antigen receptors that are fundamentally different from those of jawed vertebrates. Whereas antibodies are composed of immunoglobulin (Ig) domains, the variable lymphocyte receptors (VLRs) of jawless fish (lamprey and hagfish) consist of leucine-rich repeat (LRR) modules1-4. The LRR motif is also found in many innate immune receptors, including Toll-like receptors (TLRs) and plant disease resistance proteins, which underscores its extraordinary competence for microbial recognition5. Evolutionarily, VLRs are the oldest adaptive immune receptors1,2,5. Indeed, VLRs are the only natural antigen receptors to utilize a non-

Correspondence should be addressed to Z.P. (pancer@comb.umbi.umd.edu) or R.A.M. (mariuzza@carb.nist.gov)..

[4]These authors contributed equally to this work.

Ig scaffold, making them potential alternatives to antibodies in applications such as biosensors, bioimaging, and biopurification6-8.

Like antibodies, VLRs of lamprey and hagfish, of which there are two types (A and B), are expressed on the surface of lymphocytes or as secreted proteins. Also like antibodies, VLRs are generated by DNA recombination1,2,9. However, whereas antibodies are assembled from V, D and J gene segments, VLRs are assembled from multiple LRR-encoding cassettes, selected from several hundred that flank each germline VLR gene2,10. This process can generate a vast repertoire of over 1014 unique receptors, which is sufficiently diverse to recognize most, if not all, pathogens2,10,11.

Mature VLRs contain an N-terminal LRR capping module (LRRNT), a variable number of LRR modules, and a C-terminal LRR capping module (LRRCT)1,2,9. The LRR modules are subdivided into the first LRR (LRR1), up to nine variable LRRs (LRRVs), a variable end LRR (LRRVe), and a truncated LRR designated the connecting peptide (CP). A threonine/proline-rich stalk connects the VLR to a glycosyl phosphatidylinositol membrane anchorage motif. Secreted VLRs are assembled into disulfide-linked multimers12.

Although much is known about how I g-based antibodies recognize diverse antigens13, the structural features that endow VLRs with specificity and affinity are far less well understood. Recently, structures of three hagfish VLRs in unbound form were reported3, as well as that of a lamprey VLR (RBC36) bound to H-antigen trisaccharide4. This complex showed that the oligosaccharide is lodged in a cleft between the concave surface of the VLR and a variable insert in LRRCT. To investigate how VLRs recognize protein antigens, we isolated a VLR (VLRB.2D) specific for hen egg white lysozyme (HEL) and determined its structure in free and HEL-bound forms. Additionally, we determined the structure of an affinity-matured variant of VLRB.2D bound to HEL. These structures enabled us to delineate the full size of the antigen-binding site of VLRs, identify the principal ligand-contacting positions, and correlate sequence variability with the recognition of diverse antigens by these LRR-based adaptive immune receptors.

RESULTS

Isolation and characterization of monoclonal anti-HEL VLRBs

We chose HEL as a model for studying antigen recognition by VLRs because of its relatively large size and the availability of multiple structures of Ig-based antibodies in complex with HEL13-16. To isolate monoclonal VLRs specific for HEL, we used yeast surface display17. We generated a VLRB library from lymphocyte cDNA of immunized larvae and displayed it on the yeast surface by fusion to flocculation protein Flo1p (details to be published elsewhere). We sorted the library by flow cytometry and isolated several HEL-binding clones, including VLRB.2D. As measured by isothermal titration calorimetry (ITC), soluble monomeric VLRB.2D bound to HEL with an equilibrium dissociation constant (K_D) of 4.3×10^{-7} M (Supplementary Fig. 1a,b). This relatively low affinity resembles that of IgM antibodies which, like VLRBs, exist as multimers in plasma.

To determine whether the VLR scaffold is capable of higher affinities, VLRB.2D was subjected to *in vitro* random mutagenesis and the resulting library was sorted with HEL to isolate variants with increased reactivity. One of these (VLRB.2DMut13) bound HEL with $K_D = 3.4 \times 10^{-8}$ M (Supplementary Fig. 1c,d), a 13-fold improvement over wild-type. Therefore, no inherent structural features of VLRs preclude affinities in the nanomolar range, comparable to those of IgG antibodies.

Overview of the VLR-HEL complex

We determined the structure of the VLRB.2D-HEL complex to 2.2 Å resolution. VLRB.2D adopts a horseshoe-shaped solenoid fold characteristic of LRR proteins, including TLRs18 (Fig. 1a,b). The structure comprises an LRRNT, a 24-residue LRR1, two 24-residue LRRVs (LRRV1 and LRRVe), a 13-residue CP, and an LRRCT. The concave surface of VLRB.2D, through which the VLR binds HEL, is composed of six parallel β-strands (two from LRRNT, three from LRRVs, and one from CP) (Fig. 1a).

The VLRB.2D-HEL complex buries a total surface of 1685 $Å^2$, similar to the surface buried in complexes between Ig-based antibodies and protein antigens (1400-2300 $Å^2$)13, including the camel cAb-Lys3 V_HH-HEL and shark PBLA8 IgNAR-HEL complexes (1706 and 1604 $Å^2$, respectively)14,16, in which the camel and shark antibodies, which contain only a single V_H domain, target nearly the same HEL epitope as VLRB.2D (Fig. 1b-d). The VLRB.2D-HEL complex is also reminiscent of the interaction between platelet receptor glycoprotein Ibα (GpIbα), a close structural relative of VLRB.2D (Z-score = 11.9 in a Dali structure homolog search; http://www2.ebi.ac.uk/dali/fssp/), and its protein ligand, von Willebrand factor (VWF)19 (Fig. 2a). However, HEL is shifted towards LRRCT and does not contact LRRNT, whereas VWF engages both N- and C-terminal capping modules of GpIbα. In this respect, the VLRB.2D-HEL complex resembles the VLR RBC36-H-trisaccharide complex4 (Fig. 2b). Of the total buried surface in VLRB.2D-HEL, the LRR (LRR1, LRRV1 and LRRVe), CP and LRRCT modules contribute 27%, 15% and 58%, respectively (Fig. 3a). The interface is characterized by moderate shape complementarity, based on a shape correlation statistic (S_c)20 of 0.67, compared to 0.77 and 0.71 for the cAb-Lys3 V_HH-HEL and PBLA8 IgNAR-HEL complexes14,16, respectively.

The VLR-HEL interface

Nineteen VLRB.2D residues interact with 20 HEL residues (Supplementary Table 1). VLRB.2D binds directly over the catalytic site of HEL, with a protruding loop of LRRCT (residues Pro135-Asp141) penetrating into the carbohydrate-binding cleft (Fig. 1b). Remarkably, this epitope almost totally overlaps those recognized by the anti-HEL antibodies cAb-Lys3 V_HH14 and PBLA8 IgNAR16 (Fig. 3c,d). However, the VLRB.2D epitope is completely distinct from those recognized by mouse V_LV_H antibodies, which bind to flatter surfaces on HEL13,14. Of the 20 HEL residues contacted by VLRB.2D, 17 are contacted by cAb-Lys3 V_HH and 15 by PBLA8 IgNAR. In addition, the camel and shark antibodies, like VLRB.2D, engage HEL through an extended loop that projects into the active site (Fig. 1b-d). The catalytic residue HEL Asp52 makes hydrogen bonds to a loop residue in all three complexes. The interface also includes two clusters of seven salt bridges involving HEL residues at the lips of the catalytic cleft. In one cluster, HEL Arg61 and

Arg73 form four salt bridges to Asp59 and Asp61 of LRRV1; in the other, HEL Arg122 forms three salt bridges to Asp141 and Asp143 of LRRCT (Fig. 4a and Supplementary Table 1).

The LRRCT module of VLRs contains a highly variable insert following its α-helix that is uniquely shared with GpIbα, but is absent from the LRRCTs of other LRR-containing proteins, including TLRs10. In VLRB.2D, a six-residue insert (Val134-Asn139) projects into the active site of HEL (Fig. 4a), where it accounts for 32 of 124 (26%) van der Waals contacts, as well as three of eight hydrogen bonds (Supplementary Table 1). Residues Leu140-Arg146 of LRRCT, which immediately follow the insert, make an additional 16 (13%) contacts and three salt bridges to HEL.

Despite the major contribution of LRRCT to antigen recognition, most (61%) contacts with HEL are mediated by the concave surface of VLRB.2D (Fig. 1a and Supplementary Table 1). This binding mode resembles that between other LRR family members and protein ligands, notably the interaction of TLR4 with MD-218,21. Except for LRRNT, all other LRR modules (LRR1, LRRV1, LRRVe and CP) engage HEL. The antigen-binding concave face of VLRB.2D is composed of three parallel ridges (herein designated R1, R2 and R3) associated with the beginning, middle and end, respectively, of the β-strand of these LRR modules (Fig. 5a). Each β-strand contains a six-residue motif, $X^1L(I)X^3LX^5X^6$, in which leucine or isoleucine at positions 2 and 4 compose the hydrophobic core of the VLR and the remaining four residues form the three ridges on its concave surface that engage the antigen: R1 (position 1), R2 (position 3) and R3 (positions 5 and 6) (Fig. 5a). Significantly, all 11 HEL-contacting positions within these ridges are characterized by high sequence variability (see below).

Conformational changes upon complex formation

To identify conformational changes in VLRB.2D associated with antigen binding, we determined the structure of unbound VLRB.2D to 1.55 Å resolution. Superposition of free VLRB.2D onto VLRB.2D in the VLRB.2D-HEL complex gave an r.m.s. difference of 1.0 Å for all atoms, indicating that the VLR retains nearly the same conformation. However, the LRRCT insert underwent a hinge movement of ~1 Å between Gly136 and Asn139 to accommodate HEL, concomitant with formation of two hydrogen bonds between VLRB.2D Asn139 and HEL Asp52 (Fig. 4b). Additionally, the side chain of Tyr137 was rotated ~30° about the Cα-Cβ axis, optimizing aromatic stacking interactions with HEL Trp62 and Trp63. Although the overall binding mode of VLRB.2D appears to correspond to a “lock-and-key” mechanism, these structural differences may indicate pre-equilibrium conformational diversity, as documented for antibodies22,23. Moreover, we cannot exclude that other VLRs undergo greater conformational adjustments, at least in the LRRCT insert.

Structure of a high-affinity VLR-HEL complex

We determined the structure of the *in vitro* affinity-matured VLRB.2DMut13-HEL complex to 2.4 Å resolution. Superposition of VLRB.2DMut13-HEL onto VLRB.2D-HEL gave an r.m.s. difference of 0.76 Å for all atoms, indicating no substantial structural changes upon affinity maturation. The only noticeable difference involves adjustments in the side chains of

VLR.2D Tyr137 and Asn139, both located in the LRRCT insert (Fig. 4c). Surprisingly, there are no direct contacts between HEL and the two mutated residues, Gly132Trp and Phe142His, neither of which is closer than 6 Å to the antigen. Moreover, the total surface buried in the affinity-matured complex (1574 Å^2) is less than in the wild-type complex (1685 Å^2). However, the higher-affinity interface shows greater shape complementarity: S_c = 0.71 for VLRB.2DMut13-HEL versus 0.67 for VLRB.2D-HEL. This is reminiscent of a study of anti-HEL antibodies at different stages of affinity maturation, where a 35-fold increase in affinity correlated with an increase in S_c from 0.69 to 0.75, with none of the mutations involving HEL-contacting residues24. Hence, similar mechanisms may operate to improve affinities in the VLR and antibody systems. The mutations also alter the electrostatic potential surface of VLRB.2D (Supplementary Fig. 2). Improved shape and electrostatic complementarity, rather than formation of additional hydrogen bonds or van der Waals contacts, probably account for the 13-fold increased affinity of VLRB.2DMut13.

Comparison of the VLR-HEL and VLR-H-trisaccharide complexes

Compared to VLRB.2D, VLR RBC36 contains two additional LRRV modules4. Nevertheless, the VLR RBC36-H-trisaccharide complex buries considerably less surface than the VLRB.2D-HEL complex (1685 Å^2 versus 730 Å^2), owing to the smaller size of the ligand (Figs. 2b and 3a,b). In both complexes, the antigen interacts with LRR1, the LRRVs, CP and LRRCT, but makes no contacts to LRRNT. Whereas VLRB.2D employs all three binding site ridges to engage HEL (Fig. 5a), only R2 and R3 of VLR RBC36 (*i.e.* the middle and end of the β-strand of the LRR modules) are involved in binding the H-trisaccharide antigen. Another important difference between the two complexes is the use of the region between the C-terminal-most strand dyad of LRRCT and the LRRCT insert (residues 140-146 in VLRB.2D), which forms a shallow depression in both VLRs. While this region interacts extensively with antigen in the VLR-HEL complex (Supplementary Table 1), no contacts are seen in the VLR-H-trisaccharide complex (Fig. 2b), suggesting that it might be utilized primarily for large ligands (*e.g.* proteins).

In the case of HEL, the LRRCT insert fits into a complementary pocket in the ligand, as in the GpIbα-VWF complex19 (Fig. 2a). Hence, this mode of interaction with a protein ligand is likely a feature inherited from the common ancestor of the cyclostome VLRs and the pan-vertebrate GpIbα. For the oligosaccharide, the LRRCT insert is again a key feature of the interface, but sandwiches the ligand between itself and the concave surface of the VLR (Fig. 2b). These two distinct interaction modes underscore the versatility of LRRCT inserts in mediating specific contacts with unrelated ligands. In both complexes, a structurally equivalent aromatic residue near the tip of the LRRCT insert (Tyr137 in VLRB.2D; Trp204 in RBC36) is central to the interface. This position may play a key role in antigen recognition across diverse LRRCT inserts.

Antigen binding and sequence variability of VLRs

This comparison of two distinct VLR-ligand complexes suggested several shared features of the antigen-binding site of VLRs. To test their generality, we combined structural information with sequence analysis of a comprehensive database of LRR modules from 588 lamprey VLRB sequences. We prepared multiple alignments of all LRRs corresponding to

the ligand-contacting LRR modules of both structures, and calculated Shannon entropy25,26 for each aligned position as an objective measure of sequence diversity. We then compared the entropy at each position with the contacting residues identified using a sphere of 5 Å around each atom.

In the LRR1s, three of the top four positions in terms of entropy coincided with the ligand-contacting residues (Fig. 5c), while in the LRRVs, LRRVes and the CPs, the top four high-entropy positions also corresponded to the ligand-contacting positions in one or both of the VLR-antigen structures (Fig. 5d-f). Importantly, all four highest-entropy positions in each of these LRR modules (LRR1, LRRV, LRRVe and CP) correspond exactly to residues forming the three parallel ridges on the antigen-binding face of VLRs described above: position 1 in the β-strand segment $X^1L(I)X^3LX^5X^6$ (R1), position 3 (R2), and positions 5/6 (R3) (Fig. 5a). Furthermore, we found that the four highest-entropy positions in LRRCT inserts contacted antigen in the VLRB.2D-HEL structure (Fig. 5g). This nearly perfect correlation between highest-entropy and ligand-contacting positions strongly suggests that these positions likely represent the generalized binding interface utilized by all VLRs. Moreover, the high variability at these positions often results in changes in polarity/hydrophobicity, bulk or charge of the residues in these positions, indicating that they are critical for recognizing antigenic diversity. The residues with the highest average frequency in ligand-contacting positions are Tyr (16.7%), Trp (16.4%) and His (11.2%), which contain aromatic side chains with potential for polar interactions. A similar observation has been made for antibody combining sites13.

We also examined LRRNT, which does not contact antigen in the two complexes. We observed that LRRNT contains four high-entropy positions (Fig. 5b), comparable to the other LRRs, and that the residues at these positions form a continuation of the three ridges involved in ligand contacts in both complexes: Thr12 (R1), Asp14 (R2), and Ser16/Gly17 (R3) in VLRB.2D (Fig. 5a). This implies that LRRNT might be involved in antigen recognition in other VLR-ligand complexes, either through a shift of the antigen away from LRRCT and towards LRRNT or because of increased antigen size.

DISCUSSION

We used a high-throughput yeast display platform to isolate antigen-specific VLRs from large VLR libraries (10^7-10^8 clones). VLRB.2D bound HEL with relatively low affinity (K_D = 4.3 × 10^{-7} M). This affinity is consistent with the finding that a monomeric VLRB specific for an anthrax glycoprotein also showed weak binding12. By contrast, a multimeric form of this VLR (the natural state of VLRBs in lamprey plasma) bound with high avidity. Hence, VLRBs resemble IgM antibodies, which overcome their weak monomeric affinities for antigen (micromolar K_Ds) by forming high-avidity pentamers. However, our demonstration that VLRB.2D could be affinity-matured to bind HEL with K_D = 3.4 × 10^{-8} M indicates that VLRs can achieve affinities similar to those of IgG antibodies (nanomolar). We anticipate the existence of high-affinity binders in natural VLR repertoires, since lamprey expresses a DNA cytosine-deaminase member of the AID-APOBEC family that could contribute to VLR diversification10.

The past several years have witnessed an intensive search for alternatives to Ig-based antibodies6,7. Various alternative binding scaffolds have been described, including lipocalins, fibronectins, and ankyrin repeats, but none are natural antigen receptors6,7. The fact that VLRs, unlike antibodies, are modular single-chain proteins should facilitate molecular engineering to create novel fusion proteins, or to alter affinity or specificity. Additionally, these extremely stable molecules12 may prove useful as reagents for detecting mammalian antigens that are invisible to antibodies due to self-tolerance.

Crystallographic studies of numerous antigen-antibody complexes have shown that clefts on protein surfaces are generally avoided by V_LV_H antibodies13,14. By contrast, antibodies from camelids and sharks, which contain only a single V_H domain, preferentially target clefts on antigens14-16. The strong proclivity of V_LV_H antibodies for planar epitopes may be explained by their relatively planar binding sites, whereas the convex shape of the binding sites of V_H antibodies, with their protruding CDR3 loops, favors recognition of concave epitopes not accessible to V_LV_H antibodies14.

VLRB.2D recognizes an epitope on HEL that almost completely overlaps those recognized by camelid and shark V_H antibodies. Moreover, the LRRCT insert of VLRB.2D projects directly into the catalytic cleft, in a manner analogous to CDR3 of the V_H antibodies. Nonetheless, several considerations indicate that VLRs are unlikely to be restricted to targeting clefts. First, besides the LRRCT insert, the antigen-binding site of VLRs comprises a large surface composed of up to 12 parallel β-strands. This surface could potentially engage antigens independently of LRRCT, as suggested by other complexes involving LRR proteins (TLR4-MD-2, *Listeria* internalin-E-cadherin)21,27. Second, LRRCT inserts display a broad length distribution, ranging from 0 to 13 residues (Supplementary Fig. 3), such that VLRs with short LRRCT inserts should not contain long protruding loops that could hinder recognition of planar epitopes. Third, even in the case of long LRRCT inserts, these may be oriented to flank the ligand, rather than point directly towards it, as in VLRB.2D-HEL. Indeed, the 14-residue insert in the C-terminal capping module of GpIbα packs against one side of VWF19. Hence, VLRs appear to have evolved to recognize as structurally diverse an array of protein epitopes as V_LV_H antibodies.

Conformational rearrangements in the CDR loops of antibodies (and T cell receptors) have been shown to contribute to antigen recognition13,22,28,29. Because of its structural rigidity, the concave surface of VLRs is unlikely to undergo conformational changes beyond side-chain rearrangements. Larger structural adjustments are likely in the LRRCT inserts, as suggested by the hinge movement in the LRRCT insert of VLRB.2D upon engaging HEL, which may permit multiple binding modes22,23.

The length distribution of the LRRCT insert in different VLRs is a strong discriminant, both in terms of type of VLR (A or B) and the organism from which they are derived. Thus, VLRBs in both lamprey and hagfish show much greater insert length variability than VLRAs (Supplementary Fig. 3), with both cyclostomes displaying fairly large inserts in VLRBs (lamprey median length = 8 residues; hagfish median length = 10 residues). However, the LRRCT inserts of lamprey VLRAs are much longer (median length = 12 residues) than those of hagfish VLRAs (median length = 3 residues). These differences in

insert length distribution suggest possible specialization of the two classes of VLRs in terms of ligand recognition.

While both the VLRB.2D-HEL and VLR RBC36-H-trisaccharide complexes contain lamprey VLRBs, they differ in the length of their LRRCT inserts (6 and 9 residues, respectively). A comparison of these inserts suggests that large inserts are likely to stretch across the concave surface and form a ligand-binding pocket. Thus, VLRs containing larger inserts might specialize in binding smaller ligands, such as sugars and peptides. This pocket would be functionally equivalent to the cleft between the CDR3s of the L and H chains of V_LV_H antibodies, where small ligands bind13.

Because HEL is a large ligand, the VLRB.2D-HEL complex has effectively revealed the full extent of the antigen-binding surface of VLRs. This surface (~850 $Å^2$) is comparable in size to that of anti-protein antibodies and includes a similar number of antigen-contacting residues (~20). Besides LRRCT, the binding site spans all LRR modules of the VLR solenoid, except LRRNT. The antigen-contacting residues are organized into three ridges (R1-3) on the concave face of the VLR, whereby R1, R2 and R3 are formed by residues 1, 3 and 5/6, respectively, of the $X^1L(I)X^3LX^5X^6$ β-strand segment of each LRR module. Whereas sequence variability in antibodies is confined to the CDR loops connecting the β-strands of the V domains, sequence variability in VLRs is concentrated in the β-strands of the LRRs. Additional variability is provided by the LRRCT insert, and by a region between the insert and the C-terminal-most strand dyad of LRRCT. We found a nearly exact match between the most variable positions in LRR1, LRRV, LRRVe and CP (1, 3 and 5/6 of the $X^1L(I)X^3LX^5X^6$ motif) and those contacting HEL. Because this overlap between highest-diversity and ligand-binding positions extends to LRRCT (and probably LRRNT), we conclude that we have pinpointed most, if not all, the key antigen-contacting positions of VLRs, which together constitute the generalized binding site of these ancient immune receptors.

Accession codes

Atomic coordinates and structure factors for VLRB.2D, VLRB.2D-HEL and VLRB. 2DMut13-HEL have been deposited in the Protein Data Bank under accession codes 3G39, 3G3A and 3G3B, respectively.

METHODS

Protein preparation

Lamprey larvae were immunized bi-weekly with HEL in Freund's complete adjuvant. Within eight weeks, the lamprey produced a high-titer anti-HEL plasma VLR response. To isolate monoclonal VLRs specific for HEL, a VLRB library (8×10^6 independent clones) was generated from lymphocyte cDNA of immunized larvae and displayed on the surface of yeast by C-terminal fusion to flocculation protein Flo1p. The library was sorted by flow cytometry for binding to fluorescent-labeled HEL and several unique VLRB clones were isolated, including VLRB.2D. The diversity region of VLRB.2D, from LRRNT to LRRCT (residues 1-168), was cloned into the expression vector pT7-7 (Novagen) and expressed as

inclusion bodies in BL21-CodonPlus(DE3)-RIL *Escherichia coli* cells (Stratagene). Bacteria were grown at 37 °C in LB medium to an absorbance of 0.6-0.8 at 600 nm, and induced with 1 mM isopropyl-β-D-thiogalactoside. After incubation for 3 h, the bacteria were harvested by centrifugation and resuspended in 50 mM Tris-HCl (pH 8.0) containing 0.1 M NaCl and 2 mM EDTA; cells were disrupted by sonication. Inclusion bodies were washed with 50 mM Tris-HCl (pH 8.0), 0.1 M NaCl and 0.5% (v/v) Triton X-100, then solubilized in 8 M urea and 100 mM Tris-HCl (pH 8.5). For *in vitro* folding, inclusion bodies were diluted to a final concentration of 10 mg/l into 0.8 M arginine, 100 mM Tris-HCl (pH 8.5), 2 mM EDTA, 3 mM reduced glutathione and 0.3 mM oxidized glutathione. After 3 days at 4 °C, the folding mixture was concentrated, dialyzed against 20 mM Tris-HCl (pH 8.5), and applied to a MonoQ column (GE Healthcare). Further purification was carried out using a Superdex 75 HR column.

For affinity maturation, VLRB.2D was subjected to *in vitro* random mutagenesis. The resulting library (1×10^7 clones) was displayed on surface of yeast by C-terminal fusion to the yeast flocculation protein Flo1p. The library was sorted by flow cytometry using biotinylated HEL to isolate variants with increased reactivity towards HEL (details to be published elsewhere). The affinity-matured VLRB.2DMut13 mutant was produced and purified similarly to the wild-type protein.

Crystallization and data collection

Crystals of unbound VLRB.2D were obtained at room temperature in 70% (w/v) 2-methyl-2,4-pentanediol and 0.1 M Hepes (pH 7.5). The VLRB.2D-HEL complex crystallized in 2.0 M ammonium sulfate and 0.1 M sodium citrate (pH 5.6). Crystals of the VLRB.2DMut13-HEL mutant complex grew in 8% (w/v) polyethylene glycol 8000 and 0.1 M Tris-HCl (pH 8.5). For data collection, VLRB.2D crystals were directly flash cooled in liquid nitrogen. Crystals of VLRB.2D-HEL and VLRB.2DMut13-HEL were cryoprotected in 3.5 M ammonium sulfate and 30% (w/v) glycerol, respectively, prior to flash cooling. X-ray diffraction data were collected at beamline X29 of the Brookhaven National Synchrotron Light Source with an ADSC Quantum-315 CCD detector. All data were indexed, integrated and scaled with the HKL2000 program30. Data collection statistics are presented in Table 1.

Structure determination and refinement

Orientational and positional parameters for unbound VRLB.2D were determined by molecular replacement using Phaser31 and refined using RefMac 5.032. Truncated hagfish VLRB.59 (PDB accession code 2O6S)3 was used as the search model. Rebuilding and modeling were accomplished with XtalView33 using SIGMAA-weighted $2F_o$ - F_c and F_o - F_c maps. The structures of the VLRB.2D-HEL and VLRB.2DMut13-HEL complexes were obtained by molecular replacement using Phaser31. The search models used in the calculations were HEL (PDB code of 2VB1)34 and the refined structure of unbound VLRB.2D. Final refinement statistics are summarized in Table 1. In the VRLB.2DMut13-HEL structure, over half (55%) of one of the four VLR molecules in the asymmetric unit was disordered, possibly due to lack of crystal contacts. Stereochemical parameters of all structures were evaluated with the program PROCHECK35. Contact residues were identified using CONTACT36, and were defined as residues containing an atom 4.0 Å or

less from a residue of the binding partner. Buried surface areas were calculated using SURFACE with a 1.4 Å probe radius. PyMOL (http://www.pymol.org/) was used to prepare figures.

Isothermal titration calorimetry

ITC measurements were carried out using a MicroCal VP-ITC titration microcalorimeter. Purified VLRB.2D, VLRB.2DMut13, and HEL (Roche) were exhaustively dialyzed against 5 mM phosphate (pH 7.2), 136 mM NaCl and 4 mM KCl. In a typical experiment, 3 μl aliquots of 0.638-3.95 mM HEL solution were injected from a 250 μl rotating syringe at 290 rpm into the sample cell containing 1.37 ml of 0.025-0.060 mM VLR solution at 25 °C. For each titration experiment, an identical buffer dilution correction was conducted; these heats of dilution were subtracted from the corresponding binding experiment. Equilibrium dissociation constants (K_Ds) were obtained by non-linear least-squares fit of the ITC data to a single-site binding model. Data acquisition and analysis were performed with the software package ORIGIN.

Sequence analysis

All VLR sequences used here were derived from previous studies1,2,9,10. Protein sequence searches were performed using the BLASTPGP and PSI-BLAST programs37, with a profile threshold of 0.01. Similarity-based clustering of protein sequences was performed with BLASTCLUST (ftp://ftp.ncbi.nih.gov/blast). Multiple sequence alignments were constructed with KALIGN38. Structural alignments were performed using MUSTANG39 and the structures were rendered with PyMOL (http://www.pymol.org/) and SWISS-PDBviewer (http://spdbv.vital-it.ch/). Ligand-binding residues were determined using a custom script (L. Aravind, unpublished) and different sphere radii, such as 3.0 Å or 5.0 Å. Entropy calculations were performed with a custom script using the alignments generated by KALIGN (L. Aravind, unpublished). The R package was used for preparing plots.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEGMENTS

This study was supported by the National Institutes of Health (AI065610 to R.A.M.) and the National Science Foundation (MCB-0614672 to Z.P.). R.A.M. and Z.P. were also supported by an Intercenter Collaboration Grant from the University of Maryland Biotechnology Institute. We thank H. Robinson (Brookhaven National Synchrotron Light Source) for X-ray data collection. Support for beamline X29 comes from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. L.M.I. and L.A. were supported by the National Library of Medicine of the National Institutes of Health. L.D. is a Cancer Research Institute Postdoctoral Fellow.

REFERENCES

1. Pancer Z, et al. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature. 2004; 430:174–180. [PubMed: 15241406]
2. Alder MN, et al. Diversity and function of adaptive immune receptors in a jawless vertebrate. Science. 2005; 310:1970–1973. [PubMed: 16373579]

3. Kim HM, et al. Structural diversity of the hagfish variable lymphocyte receptors. J. Biol. Chem. 2007; 282:6726–6732. [PubMed: 17192264]
4. Han BW, Herrin BR, Cooper MD, Wilson IA. Antigen recognition by variable lymphocyte receptors. Science. 2008; 321:1834–1837. [PubMed: 18818359]
5. Pancer Z, Cooper MD. The evolution of adaptive immunity. Annu. Rev. Immunol. 2006; 24:497–518. [PubMed: 16551257]
6. Binz HK, Amstutz P, Plückthun A. Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. 2005; 23:1257–1268. [PubMed: 16211069]
7. Skerra A. Alternative non-antibody scaffolds for molecular recognition. Curr. Opin. Biotechnol. 2007; 18:295–304. [PubMed: 17643280]
8. Pancer Z, Mariuzza RA. The oldest antibodies newly discovered. Nat. Biotechnol. 2008; 26:402–403. [PubMed: 18392023]
9. Pancer Z, et al. Variable lymphocyte receptors in hagfish. Proc. Natl. Acad. Sci. USA. 2005; 102:9224–9229. [PubMed: 15964979]
10. Rogozin IB, et al. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat. Immunol. 2007; 8:647–656. [PubMed: 17468760]
11. Alder MN, et al. Antibody responses of variable lymphocyte receptors in the lamprey. Nat. Immunol. 2008; 9:319–327. [PubMed: 18246071]
12. Herrin BR, et al. Structure and specificity of lamprey monoclonal antibodies. Proc. Natl. Acad. Sci. USA. 2008; 105:2040–2045. [PubMed: 18238899]
13. Sundberg EJ, Mariuzza RA. Molecular recognition in antigen-antibody complexes. Adv. Protein Chem. 2003; 61:119–160. [PubMed: 12461823]
14. De Genst E, et al. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc. Natl. Acad. Sci. USA. 2006; 103:4586–4591. [PubMed: 16537393]
15. Stanfield RL, Dooley H, Flajnik MF, Wilson IA. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. Science. 2004; 305:1770–1773. [PubMed: 15319492]
16. Stanfield RL, Dooley H, Verdino P, Flajnik MF, Wilson IA. Maturation of shark single-domain (IgNAR) antibodies: evidence for induced fit binding. J. Mol. Biol. 2007; 367:358–372. [PubMed: 17258766]
17. Gai SA, Wittrup KD. Yeast surface display for protein engineering and characterization. Curr. Opin. Struct. Biol. 2007; 17:467–473. [PubMed: 17870469]
18. Jin MS, Lee J-O. Structures of the Toll-like receptor family and its ligand complexes. Immunity. 2008; 29:182–191. [PubMed: 18701082]
19. Huizinga EG, et al. Structures of glycoprotein Ibα and its complex with von Willebrand factor A1 domain. Science. 2002; 297:1176–1179. [PubMed: 12183630]
20. Lawrence MC, Colman PM. Shape complementarity at protein-protein interfaces. J. Mol. Biol. 1993; 234:946–950. [PubMed: 8263940]
21. Kim HM, et al. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell. 2007; 130:906–917. [PubMed: 17803912]
22. James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. Science. 2003; 299:1362–1367. [PubMed: 12610298]
23. James LC, Tawfik DS. Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. Proc. Natl. Acad. Sci. USA. 2005; 102:12730–12735. [PubMed: 16129832]
24. Li Y, Li H, Yang F, Smith-Gill SJ, Mariuzza RA. X-ray snapshots of the maturation of an antibody response to a protein antigen. Nat. Struct. Biol. 2003; 10:482–488. [PubMed: 12740607]
25. Durbin, R.; Eddy, SR.; Krogh, A.; Mitchison, G. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press; Cambridge: 1988.
26. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics. 2007; 23:1875–1882. [PubMed: 17519246]

27. Schubert W-D, et al. Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. Cell. 2002; 111:825–836. [PubMed: 12526809]
28. Wilson IA, Stanfield RL. Antibody-antigen interactions: new structures and new conformational changes. Curr. Opin. Struct. Biol. 1994; 4:857–867. [PubMed: 7536111]
29. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annu. Rev. Immunol. 2006; 24:419–466. [PubMed: 16551255]
30. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.
31. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. Acta Crystallogr. D. 2004; 60:432–438. [PubMed: 14993666]
32. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. 1997; 53:240–255. [PubMed: 15299926]
33. McRee DE. XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. 1999; 125:156–165. [PubMed: 10222271]
34. Wang J, Dauter M, Alkire R, Joachimiak A, Dauter Z. Triclinic lysozyme at 0.65 Å resolution. Acta Crystallogr. D. 2007; 63:1254–1268. [PubMed: 18084073]
35. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereo chemical quality of protein structures. J. Appl. Crystallogr. 1993; 26:283–291.
36. Collaborative Computational Project, No. 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. 1994; 50:240–255.
37. Altschul SF, et al. Gapped BLAST and PSI-PLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
38. Lassmann T, Sonnhammer ELL. Kalign - an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics. 2005; 6:298. [PubMed: 16343337]
39. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: A multiple structural alignment algorithm. Proteins. 2006; 64:559–574. [PubMed: 16736488]

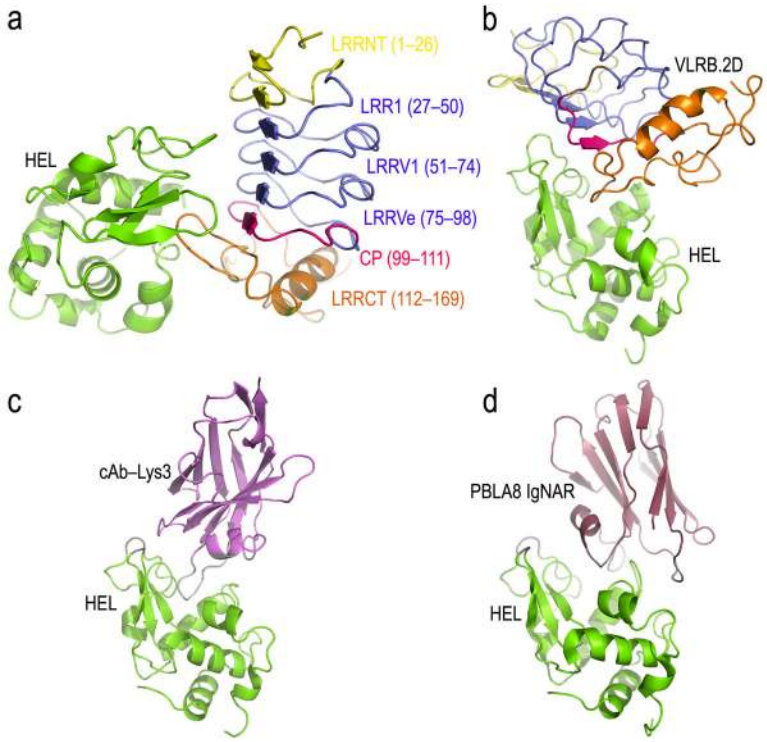

Figure 1.

Structure of the VRLB.2D-HEL complex and comparison with antibody-HEL complexes. (**a**) Ribbon diagram of the VLRB.2D-HEL complex showing the concave antigen-binding surface of the VLR solenoid. LRRNT (yellow); LRR1, LRRV1 and LRRVe (blue); CP (red); LRRCT (orange); HEL (green). In parentheses are the residue numbers corresponding to each module. (**b**) The complex is oriented to highlight the interaction between the LRRCT insert and the active site cleft of HEL. (**c**) Structure of the complex between the camel single-domain V_H antibody cAb-Lys3 (violet) and HEL (green) (Protein Data Bank accession code 1MEL). The orientation of HEL is similar to that in (**b**). (**d**) The complex between the shark single-domain V_H antibody PBLA8 IgNAR (raspberry) and HEL (green) (1T6V).

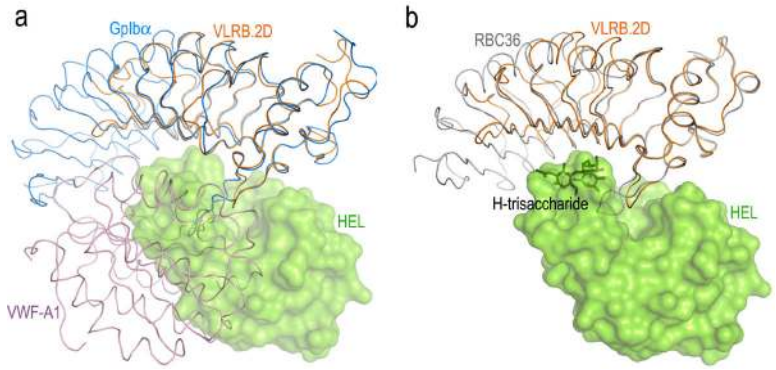

Figure 2.

Comparison of ligand recognition by LRR family proteins. (**a**) Superposition of the VLRB.2D-HEL and GpIbα-VWF (1M10) complexes. VLRB.2D is delineated as an orange α-carbon trace, GpIbα as a blue trace, VWF-A1 domain as a rose trace and HEL as a surface representation in green. (**b**) Superposition of the VLRB.2D-HEL and VLR RBC36-H-trisaccharide (3E6J) complexes. RBC36 is shown as a gray α-carbon trace; H-trisaccharide is drawn in stick presentation.

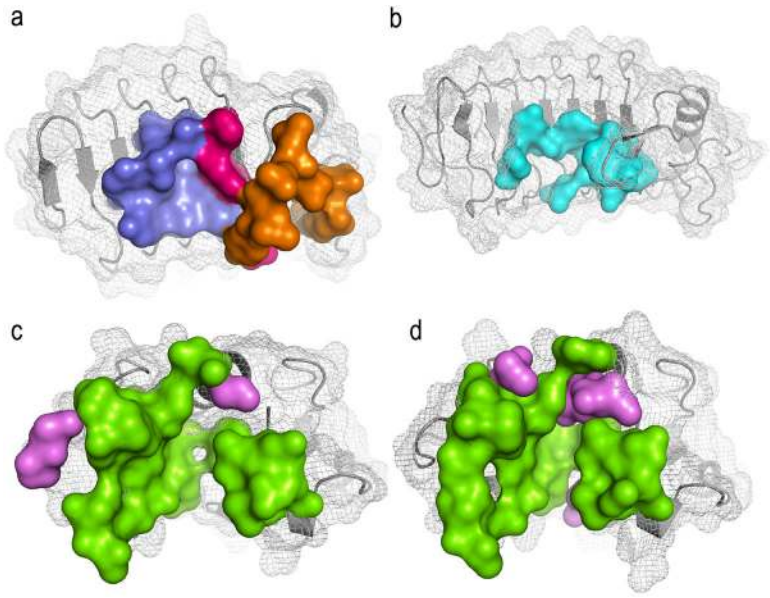

Figure 3.

Surface analysis of VLR-HEL and antibody-HEL binding interfaces. Residues involved in contacts are depicted as surface representation. (**a**) Binding surface of VLRB.2D with contacting residues from LRR1, LRRV and LRRVe in blue, from CP in red, and from LRRCT in orange. (**b**) Binding surface of VLR RBC36 (cyan) in the complex with H-saccharide4. (**c**, **d**) Comparison of VLR and antibody epitopes on HEL. Binding surfaces of HEL in the VLRB.2D-HEL (**c**) and camel cAb-Lys3-HEL (**d**) complexes. Residues forming contacts in both complexes are green; residues that interact exclusively with VLRB.2D or cAb-Ly3 are violet.

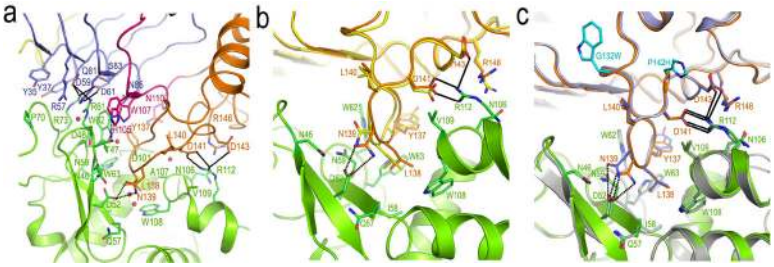

Figure 4.

The VLRB.2D-HEL binding interface. (**a**) Close-up view of contacting residues in the VLRB.2D-HEL interface. Colors (including residue labels) are as in Figure 1. The side chains of interacting residues are drawn in stick representation, with nitrogen atoms in blue and oxygen atoms in red. Salt bridges and hydrogen bonds (only those between side chains are shown) are indicated by solid lines and dashed lines, respectively. Bound waters mediating hydrogen bonds between VLRB.2D and HEL are represented as red spheres. (**b**) Structural rearrangements in VLRB.2D induced by binding to HEL. Bound VLRB.2D is orange; unbound VLRB.2D is yellow. (**c**) Superimposition of wild-type VLRB.2D-HEL and affinity-matured VLRB.2DMut13-HEL complexes. VLRB.2DMut13 is light blue; the mutated side chains, Trp132 and His142, are cyan.

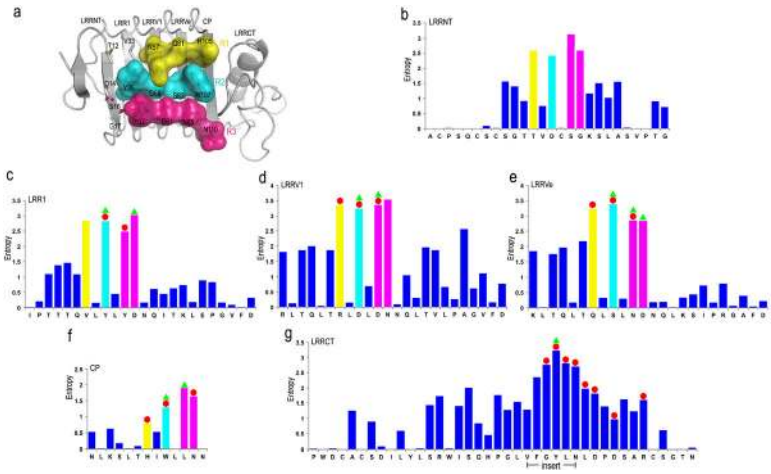

Figure 5.

Molecular architecture of the antigen-binding site of VLRs. (**a**) Surface representation of the three HEL-contacting ridges of VLRB.2D: R1 (yellow), R2 (cyan) and R3 (magenta). Contacting residues are labeled. The side chains of additional residues in R1-R3 that could potentially contact antigen are shown in stick representation. (**b-g**) Antigen-contacting positions and sequence variability. The entropy per aligned position is shown separately for LRRNT, LRR1, LRRVs, LRRVe, CP and LRRCT. Entropy bars corresponding to R1, R2 and R3 are yellow, cyan and magenta, respectively; all other bars are blue. Red circles above the entropy bars mark the HEL-contacting positions of VLRB.2D; green triangles indicate the H-trisaccharide-contacting positions of VLR RBC36. The sequence shown below is that of VLRB.2D. An insert seen in a small subset of LRRNTs, but neither in VLRB.2D nor VLR RBC36, has not been shown and does not contain any position with entropy >2.

Table 1

Data collection and refinement statistics

	VLRB.2D-HEL	VLRB.2D13-HEL	Unbound VLRB.2D
Data collection			
Space group	$P2_1$	$P2_12_12_1$	$P6_5$
Cell dimensions			
a,*b*,*c* Å	82.30, 133.65, 90.44	62.87, 106.24, 218.54	49.54, 49.54, 97.91
α,β,γ(°)	90.00, 114.69, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 120.00
Resolution Å	50.0-2.2 (2.25-2.2)*	30.0-2.4 (2.48-2.4)*	30.0-1.55 (1.61-1.55)*
R_{sym} or R_{merge}	11.7 (25.2)	9.0 (41.8)	5.8 (44.6)
$I/\sigma I$	12.6 (3.2)	29.1 (6.7)	55.8 (6.9)
Completeness (%)	98.4 (96.2)	99.9 (100)	98.4 (96.9)
Redundancy	4.6 (3.7)	14.4 (14.0)	22.2 (20.0)
Refinement			
Resolution Å	50.0-2.20	30.0-2.40	30.0-1.55
No. reflections	84142	55191	18422
R_{work} / R_{free}	22.0/27.1	24.2/28.6	16.3/19.9
No. atoms			
Protein	8959	8291	1295
Water	578	180	120
B-factors			
Protein	25.9	44.8	18.4
Water	28.0	37.7	30.3
R.m.s. deviations			
Bond lengths (Å)	0.019	0.016	0.018
Bond angles (°)	1.755	1.615	1.790

*Values in parentheses are for highest-resolution shell.