# Structural reconstruction of protein ancestry

Romain Rouet[a,b,1], David B. Langley[a,c,1], Peter Schofield[a,b], Mary Christie[a,b], Brendan Roome[a,b], Benjamin T. Porebski[d], Ashley M. Buckle[d], Ben E. Clifton[e], Colin J. Jackson[e], Daniela Stock[b,c], and Daniel Christ[a,b,2]

[a]Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW 2010, Australia; [b]Faculty of Medicine, St. Vincent's Clinical School, University of New South Wales, Darlinghurst, Sydney, NSW 2010, Australia; [c]Victor Chang Cardiac Research Institute, Darlinghurst, Sydney, NSW 2010, Australia; [d]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia; and [e]Research School of Chemistry, Australian National University, Acton, ACT 2601, Australia

Edited by David Baker, University of Washington, Seattle, WA, and approved January 20, 2017 (received for review August 16, 2016)

**Ancestral protein reconstruction allows the resurrection and characterization of ancient proteins based on computational analyses of sequences of modern-day proteins. Unfortunately, many protein families are highly divergent and not suitable for sequence-based reconstruction approaches. This limitation is exemplified by the antigen receptors of jawed vertebrates (B- and T-cell receptors), heterodimers formed by pairs of Ig domains. These receptors are believed to have evolved from an extinct homodimeric ancestor through a process of gene duplication and diversification; however molecular evidence has so far remained elusive. Here, we use a structural approach and laboratory evolution to reconstruct such molecules and characterize their interaction with antigen. High-resolution crystal structures of reconstructed homodimeric receptors in complex with hen-egg white lysozyme demonstrate how nanomolar affinity binding of asymmetrical antigen is enabled through selective recruitment and structural plasticity within the receptor-binding site. Our results provide structural evidence in support of long-held theories concerning the evolution of antigen receptors, and provide a blueprint for the experimental reconstruction of protein ancestry in the absence of phylogenetic evidence.**

protein evolution | protein structure | directed evolution | antibody | homodimer

The adaptive immune system of jawed vertebrates provides specific responses to pathogens and forms long-lasting immunological memory of past encounters (1). Key components of this system are B and T lymphocytes and their cognate receptors: B-cell receptor (membrane-bound or secreted as antibodies) and T-cell receptor. Both receptors are heterodimeric molecules formed by the association of Ig domain building blocks, which then assemble into higher order complexes. This behavior is exemplified by the canonical Y-shaped IgG antibody molecule, composed of two heavy chains (each containing four Ig domains) and two light chains (each containing two Ig domains) (2). Within the N-terminal Ig domains of the receptors, hypervariable complementarity determining regions (CDRs) mediate contact with antigen (Fig. 1*A*). The diversity and specificity of these receptors is based on the genetic rearrangement of variable (V), diverse (D) and joining (J) gene segments, mediated by recombination-activating gene enzymes, with further diversity introduced through somatic hypermutation (3).

Although the molecular mechanisms and structural features of rearranging Ig antigen receptors are well documented, only limited insights have been gained into their evolutionary origins (4, 5). Reconstructing the evolutionary history of antigen receptors using molecular phylogenetic analysis has proven difficult due to high levels of evolutionary divergence (6–8). Indeed, Ig receptor genes are completely absent from the genomes of invertebrates and jawless vertebrates [from which jawed vertebrates diverged some 500 million years ago and which have developed an alternative means of adaptive immunity based on leucine-rich-repeats (5, 9, 10)]. Limited evidence has been obtained from the sequence analyses of the genomes of slowly evolving jawed vertebrates, such as the elephant shark (11): These analyses suggest a distant genetic link between B- and T-cell receptors, supporting the notion that these molecules diverged from a common primordial Ig-based receptor (12–14).

Although sequence-based phylogenetic reconstruction had failed to revealed detailed insights, more general efforts to understand mechanisms of primordial antigen receptors have been attempted: For instance, it has been suggested that such early receptors were encoded by a basic genetic ensemble consisting of V- and J-gene segments, as observed for modern-day antibody variable light ($V_L$) domains and T-cell receptor variable (Vα) domains (5). Tethered to the surface of primordial lymphocytes, such receptor molecules may have recognized foreign antigens through homodimerization (5). The capacity of the receptors to dimerize would have allowed the formation of a large interaction site capable of forming high-affinity interactions and burying extensive antigen surface. However, in other respects, homodimerization may have limited the utility of primordial antigen receptors: For instance, such an arrangement would have restricted diversity, with the CDRs of both protomers displaying identical composition. Furthermore, it is not directly evident how a symmetrical receptor arrangement would have had the capability of binding the vast number of asymmetrical antigens in nature.

Such limitations do not apply to modern-day heterodimeric Ig receptors, which are believed to have evolved through processes of gene duplication and diversification, leading to distinct heavy and light chains, increased diversity, and a relaxation of

## Significance

**The adaptive immune system forms our primary defense against bacteria and viruses. Key players of this system are antigen receptors, dimeric molecules formed by two different types of immunoglobulin domains. It is generally believed that these receptors evolved from an ancestral dimer formed by only a single type of immunoglobulin. Using laboratory evolution, we have recreated such homodimeric receptors and characterized their interactions by X-ray crystallography. Our findings provide molecular insights and support of long-held theories concerning the evolution of the adaptive immune system. They also provide a blueprint for the experimental reconstruction of ancestral proteins in the large number of cases in which evolution has obscured sequence similarities beyond recognition, and which cannot be analyzed using current sequence-based approaches.**

Author contributions: D.C. designed research; R.R., D.B.L., P.S., M.C., B.E.C., C.J.J., and D.C. performed research; M.C., B.R., B.T.P., A.M.B., C.J.J., D.S., and D.C. analyzed data; R.R., D.B.L., P.S., M.C., C.J.J., D.S., and D.C. wrote the paper; and B.R., B.T.P., and A.M.B. generated illustrations.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4N1E and 4N1C).

[1]R.R. and D.B.L. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: d.christ@garvan.org.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613477114/-/DCSupplemental.

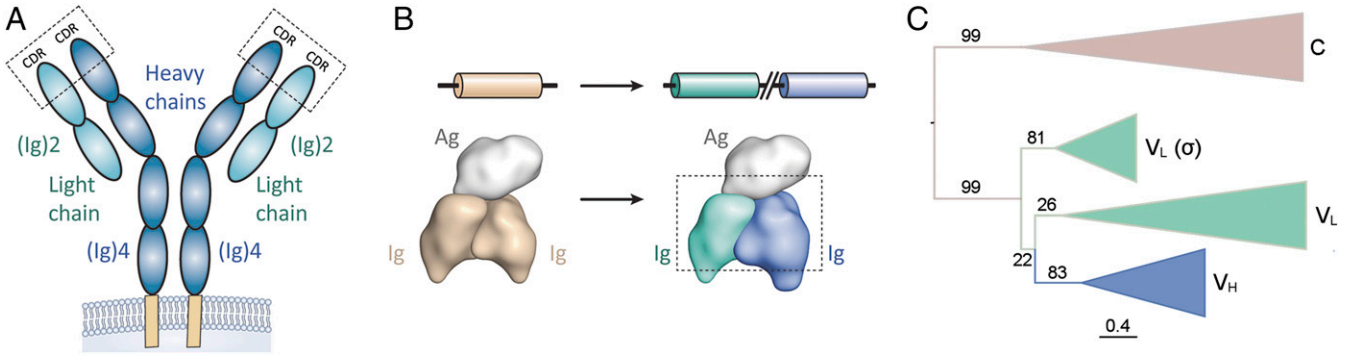


**Fig. 1.** Evolution of rearranging Ig antigen receptors. (*A*) Chain structure of modern-day heterodimeric antigen receptors (IgG antibody isotype shown). The molecule is composed of two heavy chains (each containing four Ig domains, in blue) and two light chains (each containing two Ig domains, in cyan) (2). Hypervariable CDRs, which mediate contact with antigen, are highlighted. (*B*) Gene duplication and diversification events are believed to have converted a homodimeric Ig format (tan) to the heterodimeric format (cyan/blue) observed in the B lymphocyte-based (B-cell receptor, antibodies) and T lymphocyte-based (T-cell receptor) adaptive immune systems of jawed vertebrates. (*C*) Condensed maximum likelihood phylogeny of Ig heavy and light chains of vertebrates. Branches are labeled with bootstrap values from 100 replicates. (Scale bar: mean number of substitutions per site.) C, antibody constant.

symmetry constraints (Fig. 1 *A* and *B*). The diversification of duplicated components is an evolutionary common phenomenon, observed for many other protein families, resulting in the emergence of multimers of paralogous proteins (15, 16). This process is exemplified by structures of the photosystem I core, which is formed by a homodimer in bacteria and by a heterodimer in eukaryotes (17). In the case of rearranging Ig receptors, it has long been proposed that a primordial gene encoding a basic Ig-fold of ∼110 residues gave rise to a two-domain protein structure through an initial duplication event [analogous to the modern-day human antibody light chain (variable light [$V_L$]-constant light [$C_L$])] (18).

A second duplication and diversification event (including the addition of D-gene segments) would then have yielded the four-domain arrangement characteristic of the human antibody heavy chain [variable heavy ($V_H$)-constant heavy ($C_H$)1-$C_H$2-$C_H$3] (18) (Fig. 1*B*). Indeed, it has been suggested that this process may be correlated with two whole-genome duplications, which are believed to have occurred during early vertebrate evolution (9, 19).

We set out to investigate whether a homodimeric Ig receptor could be reconstructed in the laboratory and whether such a molecule would be capable of binding to an asymmetrical protein antigen with high affinity. As a genetic source for the reconstructive process, we used V- and J-segments of the human antibody $V_L$ family to assemble the basic Ig scaffold. This family was chosen because it displays relatively high thermodynamic and colloidal stabilities (20–22), and has long been known for its propensity to dimerize. Indeed, as early as 1848, the English physician Henry Bence-Jones described what were later identified as Ig light chain dimers in the urine of patients with multiple myeloma (23). Structures of such Bence-Jones proteins reveal a face-to-face homodimeric arrangement of the variable domains (24–27), similar to what has been observed for other homomers (15). This symmetrical structural pairing is in excellent agreement with the proposed evolutionary theory, because the homodimeric arrangement of variable domains in Bence-Jones proteins closely resembles the structure of the $V_H$-$V_L$ heterodimer found in modern-day antibodies (5). Moreover, Bence-Jones proteins have long been known to be capable of providing hydrophobic pockets within their CDRs and interface regions, into which haptens (28) and small peptides (29) can be soaked. In contrast, the antibody heavy chain variable domain is less prone to the formation of homodimers, and is encoded by additional D-gene segments, which are believed to be of relatively recent evolutionary origin (30). Taken together, these attributes suggest that light chain more closely resembles the ancestral Ig molecule (18), rendering it a suitable starting point for the reconstruction of a primordial receptor.

## Results

**Sequence-Based Reconstruction.** We first attempted to use ancestral protein reconstruction to gain insights into the evolutionary origins of Ig antigen receptors. Phylogenetic analysis broadly separated constant and variable domains into separate clades; however, the phylogenetic fine structure was not well resolved and was characterized by low bootstrap support (Fig. 1*C* and Fig. S1). This result is consistent with previous studies, which have illustrated the difficulties associated with sequence-based phylogenetic analysis of short and rapidly evolving Ig genes, including poor bootstrap support for major branches (6–8). The incapacity to produce robust phylogenies of antibody genes therefore precluded the direct reconstruction of an ancestral antigen receptor through statistical inference of sequences (31).

**Structural Reconstruction and Selection of Reconstructed Primordial Receptors.** For experimental reconstruction, human V-(O12/O2/DPK9) and J-($J_\kappa$1) gene segments were used, which are among the most common in the human light chain repertoire (32). The segments were recombined, and amino acid diversity was introduced through site-directed mutagenesis of hypervariable CDRs [positions 28, 30–32, 50–51, 53, 91–94, and 96; numbering according to Kabat et al. (33)] to generate a large synthetic Ig repertoire ($5 \times 10^9$ clones) displayed as protein fusions on the tip of filamentous bacteriophage (*SI Text*, Fig. S2, and Table S1). We used a phage display format that presents multiple copies of the receptor on the surface of each phage particle, thereby, in principle, allowing homodimerization (34). This repertoire was panned against the model antigen hen egg-white lysozyme (HEL) immobilized on a solid support; and, after four rounds of selection, binders were identified by antigen ELISA and sequenced (*SI Text*).

**Receptor Characterization.** DNA sequencing revealed 12 unique clones, which clustered into two families. A representative of each family, designated Ig5 and Ig12 (amino acid sequences in Fig. S3*C*) was overexpressed in *Escherichia coli*, and the purified proteins were further investigated by a range of biophysical techniques. Size-exclusion chromatography coupled to multiangle laser light scattering (35) revealed the stoichiometry of both receptor–HEL complexes to be consistent with two Ig domains bound to a single HEL molecule (Fig. S3 *A* and *B*). Moreover, binding studies using biolayer interferometry revealed affinities in the nanomolar range,

with equilibrium-binding constants for HEL binding of 130 nM and 30 nM for Ig5 and Ig12, respectively (Fig. S4).

Crystals of Ig5 and Ig12 in complex with HEL were obtained, and their structures were solved by molecular replacement and refined to resolutions of 2.2 Å and 1.7 Å, respectively (Figs. 2*A* and 3*A*, and Table S2). In both structures, a receptor dimer is bound to a single HEL molecule, although the HEL epitopes targeted are completely different (Figs. 2*A* and 3*A*). Unlike previously reported Ig receptor structures in complex with antigen, in which $V_H$ and $V_L$ paralogs interact in a heterodimeric arrangement (36), both structures reveal a unique homodimeric nature of the receptor components. Strikingly, the CDRs of each receptor protomer, even though identical in amino acid sequence, form discrete and extensive contacts with different parts of the antigen surface (which lacks the twofold rotational symmetry of the receptor). In the case of the $Ig5_2$–HEL structure, the Ig domains individually bury ~400 $Å^2$ and 450 $Å^2$ of the HEL surface (850 $Å^2$ combined). Similar buried surface areas are observed for the $Ig12_2$–HEL structure (~300 $Å^2$, 400 $Å^2$, and 700 $Å^2$ combined). Indeed, the antigen contact surfaces are comparable in size to the antigen contact surfaces of the Ig-Ig homodimer interfaces (at 600 $Å^2$ and 700 $Å^2$ for $Ig5_2$ and $Ig12_2$, respectively). Both receptors form large and well-defined interfaces (Fig. S5), similar in extent to what has been observed for heterodimeric antibody–antigen complexes (560–855 $Å^2$ combined) (37).

## Discussion

**Ancestral Protein Reconstruction in the Absence of Sequence Conservation.** The high levels of sequence diversity within the Ig antigen receptor family had previously precluded detailed analyses of evolutionary origins (as confirmed by further sequence-based ancestral protein reconstruction attempts as outlined above). Here, we have a used different approach based on a combination of in vitro selection and crystallographic structure determination to investigate evolutionary ancestry. More specifically, we have reconstructed antigen receptors that resemble the Ig-fold and homodimeric state of ancestral proteins from which modern-day heterodimeric B- and T-cell receptors are believed to have originated through process of gene duplication and diversification.

**Interaction of the Homodimeric Receptors with Antigen.** How is it possible for the reconstructed symmetrical Ig dimers to bind an asymmetrical target with high specificity and affinity? Closer inspection of the structures reveals two distinct mechanisms. In the $Ig5_2$–HEL structure, which contains four $Ig5_2$–HEL complexes in the asymmetrical unit, superposition of the Ig domains within each complex reveals that the side chains of key residues adopt a range of conformations (Fig. 2*B*). This conformational diversity is particularly pronounced for tyrosine residues at positions 53 and 94, which adopt strikingly different conformations (largely due to rotation about the chi1 dihedral angle, leaving the main chain fold unaffected) (Fig. 2*B*, Table S3, and Movie S1). Hence, in the $Ig5_2$–HEL structure, homodimer symmetry is relaxed through conformational side-chain diversity within CDR loops to complement the asymmetrical antigen.

In contrast, such examples of structural plasticity are not observed in the $Ig12_2$–HEL structure, where superposition of the Ig protomers within the single complex in the asymmetrical unit reveals few detectable differences in side-chain conformations (Fig. S6*A*). Instead, a different mechanism based on the selective recruitment of CDRs is dominant in this second structure. Here, each Ig subunit uses different combinations of contact residues, within otherwise structurally similar CDR loops, to form unique hydrogen-bonding networks with the HEL surfaces (Fig. 3 *C* and *D*). The alternative utilization of otherwise identical CDR surfaces is particularly evident when mapping the contact footprint of the antigen onto the two Ig protomers (Fig. 3*B*). This projection reveals selective use of CDR2, with only one of the Ig12 subunits making extensive contacts through this CDR (detailed in Fig. S6). Such CDR contact selectivity is also observed in the $Ig5_2$–HEL complex (Fig. 2 *C* and *D* and Fig. S6*B*), complementing conformational CDR side-chain plasticity.

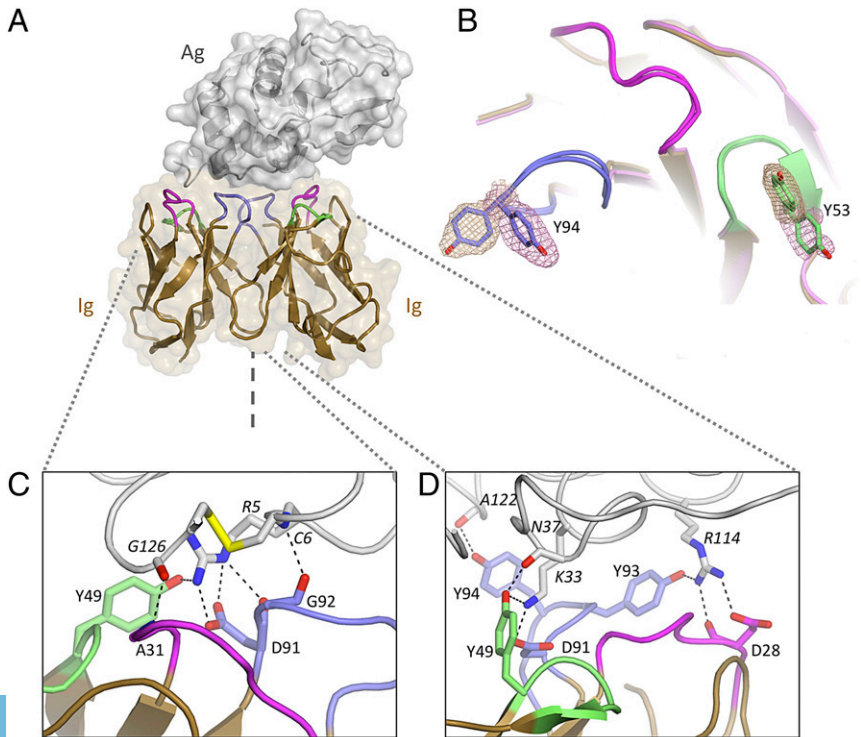


**Fig. 2.** Structure of the Ig5 homodimeric antigen receptor in complex with antigen: Evidence for conformational diversity. (*A*) Homodimer of Ig domains (tan cartoon with CDR1, CDR2, and CDR3 loops colored in purple, green, and blue, respectively) bound to a single HEL molecule (gray cartoon and surface). The symmetry axis is indicated (dashed line). (*B*) Antigen-binding site (superposed Ig protomers, chains C/D): Multiple side-chain conformations are observed at positions Y53 and Y94. Electron density (mesh) is contoured at 1 SD above average (sigma-A weighted 2mFo-DFc map). (*C*) Antigen-binding site 1 (left-hand side). (*D*) Antigen-binding site 2 (right-hand side). Hydrogen bonds are highlighted; distinct contact networks are observed for each Ig subunit.

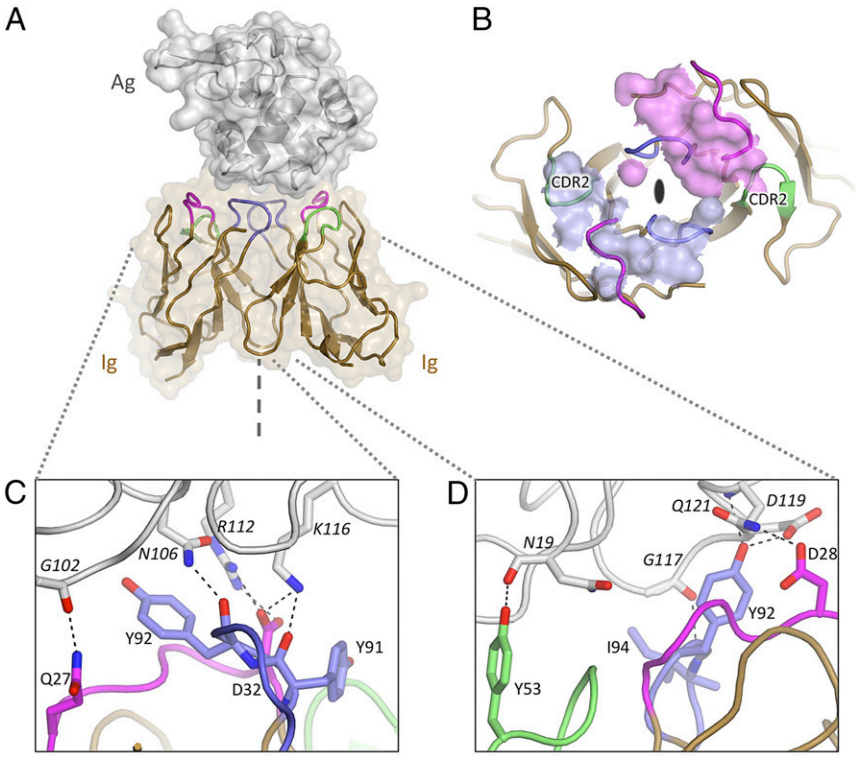


**Fig. 3.** Structure of the Ig12 homodimeric antigen receptor in complex with antigen: Evidence for selective recruitment. (*A*) Homodimer of Ig domains (tan cartoon with CDR1, CDR2, and CDR3 colored in purple, green, and blue, respectively) bound to a single HEL molecule (gray cartoon and surface). The symmetry axis is indicated (dashed line). (*B*) HEL contact footprint mapped onto the surface of the Ig12 dimer (viewed down its twofold axis as indicated). The contact surfaces of the two Ig domains are colored in light blue and purple. The CDR2s are indicated and highlight the selective use of otherwise identical surface loops. (*C*) Antigen-binding site 1 (left-hand side). (*D*) Antigen-binding site 2 (right-hand side). Hydrogen bonds are highlighted; distinct contact networks are observed for each Ig subunit.

**Comparisons with Previously Reported Antigen Receptor Structures.** Further comparison with structures from the Protein Data Bank reveal an intriguing similarity between $Ig12_2$–HEL and a heterodimeric antibody–antigen complex reported in the early 1990s: the 2.4-Å resolution structure of the D11.15 Fv fragment complexed with pheasant egg-white lysozyme (38). Comparison of the two structures (Fig. 4 *A* and *B*) reveals that, despite the absence of CDR sequence similarity, the receptors contact equivalent parts of the lysozyme surface and bury comparable surface areas (Ig12: ~700 $Å^2$ combined; D11.15: ~650 $Å^2$ combined). Similarities are further highlighted by structural superposition of the lysozyme components and examination of CDR3s (Fig. 4*C*), which contribute the majority of binding energy in antibody–antigen interactions (39). Although the CDR3 sequences and molecular details of the interactions are different, strong overall convergence is observed, with one of the Ig protomers resembling $V_H$ and the other resembling $V_L$ (Fig. 4*C*). The observation that Ig12 is capable of interacting with HEL antigen in a manner highly reminiscent of the modern-day D11.15 antibody raises questions relating to which evolutionary advantage duplication and divergence of an ancestral homodimer into a heterodimer may have been conferred. Although the asymmetrical nature of the HEL antigen used here prevents further experimental validation, it is likely that any homodimeric receptor would display an inherent preference for symmetrical antigens. The response would thereby be biased, diminishing neutralization potential. Moreover, pathogens could potentially exploit such epitope preference by expressing symmetrical decoy antigens. This behavior suggests that the need for broad epitope coverage and the avoidance of immune evasion may have shaped the evolution of modern-day heterodimeric antigen receptors.

The epitope contacted by $Ig12_2$ and D11.15 corresponds to a flat part of the lysozyme surface, which is adjacent, but distinctly outside, the active site cleft (Fig. 4 *A* and *B*, the active site is indicated by asterisks); similarly, $Ig5_2$ targets a planar epitope distant from the catalytic site (Fig. 5*A*). Similar preferences for the binding of flat molecular surfaces are observed for other paired Ig receptors, including the model anti-HEL monoclonals HyHEL10 and HyHEL9, both of which bind away from the

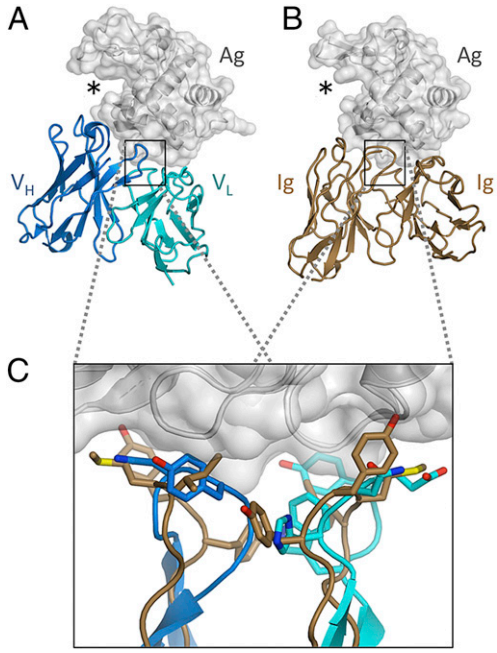


**Fig. 4.** Molecular mimicry at a conserved interface. Comparison of D11.15–pheasant egg-white lysozyme (PEL) and $Ig12_2$–HEL structures. (*A*) D11.15–PEL structure (PDB ID code 1JHL) comprises the classic heterodimer pairing of heavy chain (blue) and light chain (cyan) bound to lysozyme (gray cartoon and surface). The active site cleft is indicated (*). (*B*) $Ig12_2$–HEL structure, viewed from the same perspective as displayed in *A*. (*C*) Superposition of lysozyme molecules reveals the CDR3 loops of the D11.15 and Ig12 components (colored as in *A* and *B*) to be contacting equivalent surface areas of the lysozyme antigens.

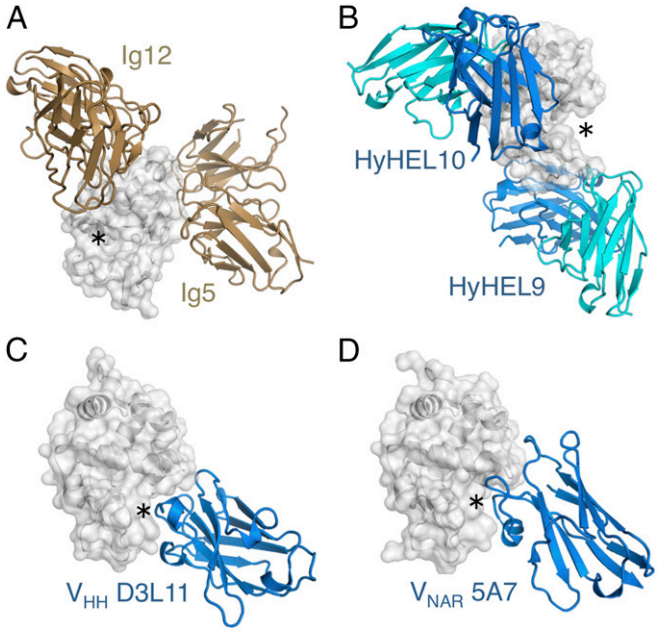


**Fig. 5.** Epitope preferences of Ig receptor formats. (*A*) Ig5 and Ig12 homodimers (tan cartoons) target flat epitopes on the HEL surface (gray cartoon and surface), distant from the HEL active site cleft (indicated by *). (*B*) Flat epitope surfaces apart from the active site are also observed for the model anti-HEL monoclonals HyHEL10 (PDB ID code 1YQV) and HyHEL9 (PDB ID code 3D9A), which are paired receptors formed by $V_H$-$V_L$ domain assembly (blue and cyan cartoons, respectively). Structures of nonpaired Ig antigen receptors of camels (*C*; $V_{HH}$; PDB ID code 1ZVY) and sharks (*D*; shark new antigen receptor ($V_{NAR}$); PDB ID code 1T6V) reveal a 1:1 stoichiometry with antigen, and binding in the HEL active site cleft.

active site cleft (Fig. 5*B*). This behavior is in marked contrast to single-domain Ig formats, such as the heavy-chain-only receptors of camelids, which are naturally devoid of light chain partners (and are believed to be of relatively recent evolutionary origin) (40, 41). Crystal structures of these domains, as well as crystal structures of single domains from sharks [shark new antigen receptors (42)], have been reported in complex with HEL antigen. These previously reported structures reveal monomeric binding, with strict 1:1 stoichiometry and a strong preference for targeting the active site cleft of HEL (43, 44) (Fig. 5 *C* and *D*).

The $Ig5_2$ and $Ig12_2$ complexes reported here clearly outline two distinct strategies by which symmetry mismatches can be tackled: relaxation of symmetry through structural side-chain plasticity and differential recruitment of CDR loops. Such mechanisms are not dissimilar to what has been observed in other cases of symmetrical receptors bound to asymmetrical protein ligands, as exemplified by the structures of the lectin-like homodimeric immunoreceptor NKG2D (natural-killer group 2, member D) bound to three different signaling proteins (45–47). Although these receptors are not Ig-based, they nonetheless structurally and functionally mimic the T-cell receptor Vα-Vβ pairing, recognizing ligands structurally related to MHC counterparts. In these structures, both receptor conformational plasticity and selective recruitment are used, and are supplemented, particularly in the case of one of the ligands, by significant conformational rearrangement of the ligand itself relative to its receptor-free state (ligand hyperplasticity) (45). This observation is in contrast to the two structures reported here, where the structure of the HEL ligand is strongly conserved, with only minor differences observed at surface loop positions. Also, it is also important to note that, unlike antibody–antigen interactions, the NKG2D–ligand complexes are not the products of adaptive immunity, but rather have coevolved over evolutionary time spans.

**Differences to Naturally Occurring Bence-Jones Proteins.** The Ig5 and Ig12 receptors generated here were selected from phage display repertoires based on common human light-chain segments (DPK9 Vκ). Consequently, overall structures of the receptors and interdomain pairing closely resemble the homodimeric arrangement of $V_L$ domains observed in structures of Bence-Jones proteins derived from patients with multiple myeloma (25–27, 48). However, outside the $V_L$ domain–domain interface, the similarities between the in vitro selected receptors and naturally occurring Bence-Jones protein begin to fade. This behavior is exemplified by previously reported structures of the Bence-Jones protein MCG, which has been crystallized with a series of synthetic peptides (Tables S4 and S5). As had been observed for the NKG2D structure, but in contrast to the structures reported here, the capacity of MCG to bind to a range of peptides of increasing length is underpinned by both the plasticity within the antibody-binding site and the conformational malleability of the peptide antigens (29, 49) (Tables S4 and S5). The MCG/peptide interface is small, with buried surface areas ranging from 462–557 $Å^2$ compared with 691–830 $Å^2$ observed for the Ig12 and Ig5 complexes. Moreover, MCG interactions are strongly dominated by hydrophobic contacts, with limited hydrogen bonds and no salt bridges observed (Tables S4 and S5). Several of the hydrophobic interactions in the MCG peptide complex are mediated by antibody framework residues, which do not commonly interact with antigen in antibody–antigen complexes (Fig. S7). Overall, the MCG/peptide structures therefore outline a picture of a hydrophobic antibody pocket into which a wide range of nonpolar ligands can be soaked in a promiscuous fashion. Although details have not been reported, it is likely that affinities of the MCG/peptide are extremely low (micromolar), reflecting the noncognate nature of the ligands and the absence of adaptive selection (29, 49).

In marked contrast, the Ig5 and Ig12 structures outlined here are characterized by affinities that are orders of magnitude higher (nanomolar) and an extensive interface encompassing numerous hydrogen bond and charge interactions (Tables S3–S5), reflecting selective pressure and their origins from combinatorial immune repertoires.

## Conclusion

For many protein families, evolutionary ancestry can be reconstructed through analyses of modern-day sequences from a range of organisms. However, in other cases, evolutionary traces have been obscured through mutation and divergence, thereby hindering sequence-based reconstruction. Nevertheless, it is evident that, even in such cases, structural features of proteins are often highly conserved throughout evolution, despite the absence of detectable sequence identity (50, 51). Here, we have used this characteristic, by selecting antigen receptors resembling the fold and homodimeric nature of ancestral proteins from a repertoire of structural Ig building blocks, followed by biophysical confirmation and structure determination. Much as experimental archeology has allowed the validation of our perceptions of ancient materials and machinery (52, 53), the approach outlined here provides insights into molecular mechanisms that are believed to have shaped antigen receptor evolution. More specifically, the structures reported here provide definite validation that a symmetrical receptor is capable of interacting with asymmetrical ligands in an adaptive fashion, not only in principle but also in a specific and high-affinity manner typical of bona fide antibody–antigen interactions.

## Methods

Library design and construction, selection of antigen binders, determination of binding stoichiometry, affinity measurements and crystal growth, structure determination and refinement, and phylogenetic analysis are described in detail in *SI Text*.

**ACKNOWLEDGMENTS.** We thank the staff at the Australian Synchrotron MX2 beamline. This work was supported by the Australian National Health and Medical Research Council and the Australian Research Council (Grants DP140103465, DE160100608, APP1050146, and APP1022688).

1. Zinkernagel RM, Hengartner H (2001) Regulation of the immune response by antigen. *Science* 293(5528):251–253.
2. Bork P, Holm L, Sander C (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* 242(4):309–320.
3. Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* 18(1):1–11.
4. Cooper MD, Alder MN (2006) The evolution of adaptive immune systems. *Cell* 124(4): 815–822.
5. Eason DD, et al. (2004) Mechanisms of antigen receptor evolution. *Semin Immunol* 16(4):215–226.
6. Das S, Nikolaidis N, Klein J, Nei M (2008) Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc Natl Acad Sci USA* 105(43):16647–16652.
7. Criscitiello MF, Flajnik MF (2007) Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. *Eur J Immunol* 37(10):2683–2694.
8. Richards MH, Nelson JL (2000) The evolution of vertebrate antigen receptors: A phylogenetic approach. *Mol Biol Evol* 17(1):146–155.
9. Cooper MD, Herrin BR (2010) How did our complex immune system evolve? *Nat Rev Immunol* 10(1):2–3.
10. Herrin BR, Cooper MD (2010) Alternative adaptive immunity in jawless vertebrates. *J Immunol* 185(3):1367–1374.
11. Venkatesh B, et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505(7482):174–179.
12. Hirano M, Das S, Guo P, Cooper MD (2011) The evolution of adaptive immunity in vertebrates. *Adv Immunol* 109:125–157.
13. Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nat Rev Genet* 11(1):47–59.
14. Litman GW, et al. (1993) Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol Biol Evol* 10(1):60–72.
15. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453(7199):1262–1265.
16. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA (2007) Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol* 8(4):R51.
17. Schubert WD, et al. (1998) A common ancestor for oxygenic and anoxygenic photosynthetic systems: A comparison based on the structural model of photosystem I. *J Mol Biol* 280(2):297–314.
18. Hill RL, Delaney R, Fellows RE, Lebovitz HE (1966) The evolutionary origins of the immunoglobulins. *Proc Natl Acad Sci USA* 56(6):1762–1769.
19. Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3(10):e314.
20. Ewert S, Huber T, Honegger A, Plückthun A (2003) Biophysical properties of human antibody variable domains. *J Mol Biol* 325(3):531–553.
21. Dudgeon K, et al. (2012) General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc Natl Acad Sci USA* 109(27):10879–10884.
22. Rouet R, Lowe D, Christ D (2014) Stability engineering of the human antibody repertoire. *FEBS Lett* 588(2):269–277.
23. Bernier GM, Putnam FW (1963) Monomer–dimer forms of Bence Jones proteins. *Nature* 200:223–225.
24. Novotný J, Haber E (1985) Structural invariants of antigen binding: Comparison of immunoglobulin VL-VH and VL-VL domain dimers. *Proc Natl Acad Sci USA* 82(14): 4592–4596.
25. Colman PM, Schramm HJ, Guss JM (1977) Crystal and molecular structure of the dimer of variable domains of the Bence-Jones protein ROY. *J Mol Biol* 116(1):73–79.
26. Fehlhammer H, et al. (1975) The structure determination of the variable portion of the Bence-Jones protein Au. *Biophys Struct Mech* 1(2):139–146.
27. Epp O, et al. (1974) Crystal and molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI. *Eur J Biochem* 45(2):513–524.
28. Edmundson AB, et al. (1974) Binding of 2,4-dinitrophenyl compounds and other small molecules to a crystalline lambda-type Bence-Jones dimer. *Biochemistry* 13(18):3816–3827.
29. Edmundson AB, et al. (1993) Principles and pitfalls in designing site-directed peptide ligands. *Proteins* 16(3):246–267.
30. Litman GW, Anderson MK, Rast JP (1999) Evolution of antigen binding receptors. *Annu Rev Immunol* 17:109–147.
31. Thornton JW (2004) Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat Rev Genet* 5(5):366–375.
32. Tomlinson IM, Cox JP, Gherardi E, Lesk AM, Chothia C (1995) The structural repertoire of the human V kappa domain. *EMBO J* 14(18):4628–4638.
33. Kabat E, Wu TT, Perry HM, Kay S, Gottesman CF (1992) Sequences of Proteins of Immunological Interest (DIANE Publishing, Collingdale, PA), 5th Ed.
34. McCafferty J, Griffiths AD, Winter G, Chiswell DJ (1990) Phage antibodies: Filamentous phage displaying antibody variable domains. *Nature* 348(6301):552–554.
35. Ye H (2006) Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography-multiangle laser light scattering. *Anal Biochem* 356(1):76–85.
36. MacCallum RM, Martin AC, Thornton JM (1996) Antibody-antigen interactions: Contact analysis and binding site topography. *J Mol Biol* 262(5):732–745.
37. Davies DR, Cohen GH (1996) Interactions of protein antigens with antibodies. *Proc Natl Acad Sci USA* 93(1):7–12.
38. Chitarra V, et al. (1993) Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex. *Proc Natl Acad Sci USA* 90(16):7711–7715.
39. Mian IS, Bradwell AR, Olson AJ (1991) Structure, function and properties of antibody binding sites. *J Mol Biol* 217(1):133–151.
40. Hamers-Casterman C, et al. (1993) Naturally occurring antibodies devoid of light chains. *Nature* 363(6428):446–448.
41. Rouet R, Dudgeon K, Christie M, Langley D, Christ D (2015) Fully human VH single domains that rival the stability and cleft recognition of camelid antibodies. *J Biol Chem* 290(19):11905–11917.
42. Greenberg AS, et al. (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* 374(6518):168–173.
43. De Genst E, et al. (2006) Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci USA* 103(12):4586–4591.
44. Stanfield RL, Dooley H, Flajnik MF, Wilson IA (2004) Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* 305(5691):1770–1773.
45. Li P, et al. (2001) Complex structure of the activating immunoreceptor NKG2D and its MHC class I-like ligand MICA. *Nat Immunol* 2(5):443–451.
46. Li P, McDermott G, Strong RK (2002) Crystal structures of RAE-1beta and its complex with the activating immunoreceptor NKG2D. *Immunity* 16(1):77–86.
47. Radaev S, Rostro B, Brooks AG, Colonna M, Sun PD (2001) Conformational plasticity revealed by the cocrystal structure of NKG2D and its class I MHC-like ligand ULBP3. *Immunity* 15(6):1039–1049.
48. Edmundson AB, Ely KR, Abola EE, Schiffer M, Panagiotopoulos N (1975) Rotational allomerism and divergent evolution of domains in immunoglobulin light-chains. *Biochemistry* 14:3953–3961.
49. Edmundson AB, Manion CV (1998) Treatment of osteoarthritis with aspartame. *Clin Pharmacol Ther* 63(5):580–593.
50. Christ D, Winter G (2003) Identification of functional similarities between proteins using directed evolution. *Proc Natl Acad Sci USA* 100(23):13202–13206.
51. Löwe J, Amos LA (1998) Crystal structure of the bacterial cell-division protein FtsZ. *Nature* 391(6663):203–206.
52. Reibold M, et al. (2006) Materials: Carbon nanotubes in an ancient Damascus sabre. *Nature* 444(7117):286.
53. Juleff G (1996) An ancient wind powered iron smelting technology in Sri Lanka. *Nature* 379(6550):60.
54. Hoogenboom HR, et al. (1991) Multi-subunit proteins on the surface of filamentous phage: Methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res* 19(15):4133–4137.
55. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA* 82(2):488–492.
56. Rouet R, Dudgeon K, Christ D (2012) Generation of human single domain antibody repertoires by Kunkel mutagenesis. *Methods Mol Biol* 907:195–209.
57. Lee CM, Iorno N, Sierro F, Christ D (2007) Selection of human antibody fragments by phage display. *Nat Protoc* 2(11):3001–3008.
58. Rouet R, et al. (2012) Expression of high-affinity human antibody fragments in bacteria. *Nat Protoc* 7(2):364–373.
59. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):271–281.
60. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72–82.
61. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.
62. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
63. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255.
64. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
65. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
66. de Castro E, et al. (2006) ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34(Web Server issue):W362–W365.
67. Lefranc MP, et al. (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* 29(3):185–203.
68. Lefranc MP, et al. (2009) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 37(Database issue):D1006–D1012.
69. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
70. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
71. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.
72. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
73. Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59(Pt 7):1131–1137.