Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor

V. A. Streltsov*†‡, J. N. Varghese*, J. A. Carmichael*‡, R. A. Irving*‡, P. J. Hudson*‡, and S. D. Nuttall*†‡

*Division of Health Sciences and Nutrition, Commonwealth Scientific and Industrial Research Organization, and ‡Cooperative Research Centre for Diagnostics, 343 Royal Parade, Parkville, 3052, Australia

Edited by Timothy A. Springer, Harvard Medical School, Boston, MA, and approved July 2, 2004 (received for review May 17, 2004)

The Ig new antigen receptors (IgNARs) are single-domain antibodies found in the serum of sharks. Here, we report 2.2- and 2.8-Å structures of the type 2 IgNAR variable domains 12Y-1 and 12Y-2. Structural features include, first, an Ig superfamily topology transitional between cell adhesion molecules, antibodies, and T cell receptors; and, second, a vestigial complementarity-determining region 2 at the ''bottom'' of the molecule, apparently discontinuous from the antigen-binding paratope and similar to that observed in cell adhesion molecules. Thus, we suggest that IgNARs originated as cell-surface adhesion molecules coopted to the immune repertoire and represent an evolutionary lineage independent of variable heavy chain/variable light chain type antibodies. Additionally, both 12Y-1 and 12Y-2 form unique crystallographic dimers, predominantly mediated by main-chain framework interactions, which represent a possible model for primordial cell-based interactions. Unusually, the 12Y-2 complementarity-determining region 3 also adopts an extended β-hairpin structure, suggesting a distinct selective advantage in accessing cryptic antigenic epitopes.

The rapid diversification (or ''big bang'') of the vertebrate immune system is hypothesized to have occurred >500 million years ago, with the incorporation of a transposon containing a pair of recombinase activating genes into a primitive Ig coding sequence (1, 2). Gene duplication and evolution of the immune effector molecules rapidly followed, along with recruitment of other proteins to maximize antibody diversity, and addition of increasingly sophisticated levels of control and complexity. The resulting immune systems, although varying between classes of animals in organizational strategies (at the genetic level) and gross structure (in the effector organs for generation and maturation of immune cells) all possess hallmarks of true adaptive immunity (2, 3).

The most evolutionary primitive animals to possess this advanced adaptive immune response are the cartilaginous fish (Chondrichthyes: sharks, skates, and rays), which diverged from the bony fish (Osteichthyes) ≈450 million years ago (4). This long evolutionary history is reflected in the diverse array of shark antibodies. These antibodies include the archetypal variable heavy chain/variable light chain (V_H/V_L) antibodies such as IgM monomeric and pentameric forms (most analogous to IgG in higher organisms) and IgW and IgX forms (5). However, recently, a distinctly unconventional antibody isotype was identified in the serum of nurse sharks (*Ginglymostoma cirratum*) and wobbegong sharks (*Orectolobus maculatus*): the Ig new antigen receptors (IgNARs) (6, 7). Current evidence implicates IgNARs as true molecules of the immune armory and as the most probable agent of the shark antigen-driven affinity-maturation antibody response (8–10).

The unconventional nature of IgNARs is apparent in their gross structural organization. First, they are disulfide-bonded homodimers consisting of five constant domains (C_{NAR}) and one variable domain (V_{NAR}) (6). There is no light chain, and the individual variable domains are independent in solution and do not appear to associate across a hydrophobic interface (as seen for conventional V_H/V_L type antibodies) (11). Second, there are three different types of IgNARs characterized by their time of appearance in shark development, and by their disulfide bond pattern (12, 13). The type 1 V_{NAR} topology is characterized by an additional framework disulfide linkage, and (usually) cysteines in the extended complementarity-determining region (CDR)3 loops, which may form intraloop disulfide bonds. Type 1 IgNARs are to date confined to nurse sharks. The type 2 topology is characterized by cysteines in the CDR1 and CDR3 loops in two-thirds of cases, which probably form stabilizing interloop disulfide bonds. Type 3 IgNARs are found predominantly in embryonic sharks, probably as a first line of resistance to pathogens before maturation of the antigen-driven response. Regardless of type, all IgNARs have minimally variable CDR1 and CDR2 loop regions and concentrate diversity in the elongated CDR3s (6, 7, 12). They can vary from 5 to 23 residues in length, although the modal classes are more of the order of 15–17 residues (13). This result is significantly larger than for conventional murine and human antibodies, but approximate to the extended CDR3s loops found in the camelid single-domain V_HH antibodies (14, 15).

The evolutionary origin of IgNARs as single-domain antibodies is open to conjecture (16). There are at least two valid hypotheses (11, 13, 17): first, that V_{NAR}s represent ancestral Igs, akin to or derived from primitive cell-surface molecules coopted to soluble antibodies in serum. In this scenario, IgNARs may represent a unique antibody lineage, or alternatively an extant version of a single-domain antibody before adoption of heterodimeric pairing in the V_H/V_L chain configuration. Alternatively, IgNARs may have reevolved to a single-domain format from a primitive heterodimeric pairing based on antibody (V_H/V_L) or T cell receptor (V_α/V_β) forms. Such reevolution to a single variable domain has been observed for camelid V_HH antibodies, over a far shorter period of evolutionary time (18). To address these questions and to explore the structure/function aspects of these shark antibodies, we have determined three-dimensional structures of two type 2 V_{NAR}s and compared them with a range of immune molecules.

Materials and Methods

Expression, Purification, and Crystallization of V_{NAR}s 12Y-1 and 12Y-2. Recombinant proteins 12Y-1 and 12Y-2 were expressed into the *Escherichia coli* periplasm as described (19). Protein 12Y-2 (14

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: IgNAR, Ig new antigen receptor; V_{NAR}, variable domain of IgNAR antibody; C_{NAR}, constant domain of IgNAR antibody; V_H/V_L, variable heavy chain/variable light chain; IgSF, Ig superfamily; CAM, cell adhesion molecule; NCAM, neural CAM; CDR, complementarity-determining region; PDB, Protein Data Bank; TCR, T cell receptor.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 1VER and 1VES).

†To whom correspondence may be addressed. E-mail: victor.streltsov@csiro.au (structure) or stewart.nuttall@csiro.au (shark antibody).

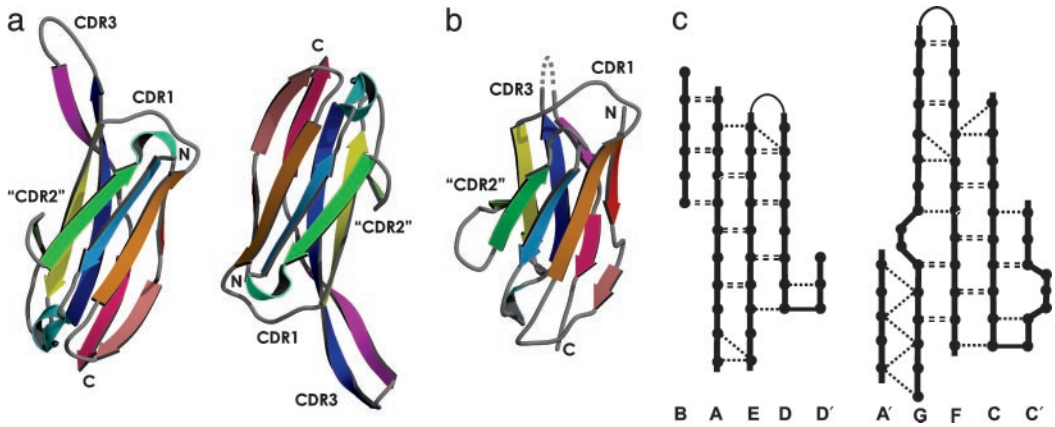

Fig. 1. MOLSCRIPT (42) and RASTER3D (43) diagrams of 12Y-2 with two monomers in the asymmetric unit (*a*) and 12Y-1 monomer without the CDR3 loop (*b*; dashed line). Each chain is shown as a ribbon representation and is colored in rainbow fashion (red for the N terminus to purple for the C terminus). The N and C termini and CDRs are labeled. (*c*) H-bond pattern of the 12Y-2 β-sheets. β-sheets are presented as filled circles and broken lines represent H bonds between main-chain atoms.

mg/ml) was set up in 2-μl hanging drops by using the Hampton Research (Laguna Niguel, CA) sparse matrix crystallization screening kit. Plates were incubated at 25°C. Final crystallization conditions were 0.1 M sodium citrate, pH 4.6/20% vol/vol isopropanol/20% polyethylene glycol 4000. Diffraction quality crystals (space group $I4_122$) were obtained after 48 h. Protein 12Y-1 (6 mg/ml) was set up as 0.2-μl sitting drops by using a Cartesian honey bee robot. Plates were incubated at 25°C. Successful conditions were scaled up to 2-μl hanging drops by using 12Y-1 protein at 13 mg/ml. Final crystallization conditions were 0.1 M 1,3-Bis[tris(hydroxymethyl)methylamino]propane, pH 6.5/45% polypropylene glycol P400. Diffraction quality crystals (space group $I2_12_12_1$) were obtained after 7 days.

Data Collection and Structure Determination. X-ray data from all crystals were measured by using Rigaku RAXIS IV (Rigaku-MSC, Tokyo) and Mar 180 (MAR-Research, Hamburg, Germany) image plate detectors mounted on a Rigaku HR3 HB x-ray generator equipped with monocapillary focusing optics (AXCO, Parkville, Australia). Diffraction data were collected at −160°C (the crystals required no added cryoprotectant) and were processed by using the DENZO/SCALEPACK suite (20). Data statistics are summarized in Table 1, which is published as supporting information on the PNAS web site. Heavy-atom sites were identified and refined with the statistical phasing program SHARP (21), and solvent-flattening DM/SOLOMON procedures were used to resolve the phase ambiguity. The model was manually built by using XTALVIEW (22) with the centroid electron-density map produced by SHARP. The model was then refined against the native 12Y-1 data by using the CCP4 suite (23). During the model refinement, 5% of the data were flagged for crossvalidation to monitor the progress of refinement by using R_{free} statistics (24). The electron density map allowed unambiguous tracing of all residues except the CDR3 (residues 88–98), which was disordered. Water molecules were located automatically with CCP4 ARP (25). After the convergence in standard refinement, a further improvement of >2% in R factors was achieved by refining all protein atoms as one anisotropic domain with the TLS procedure (26) by using CCP4 REFMAC5. The libration tensor showed significant anisotropy. The final R/R_{free} values were 16.6/25.4% for a 6- to 2.82-Å range of refined data, which included 85% of observed data. The low-resolution data were omitted at the final stage of refinement to reduce scattering contribution from the diffuse CDR3. The final 12Y-1 model contains 100 aa (residues 1–87 and 99–111) and 97 water molecules. Of the residues in the 12Y-1 model, 84.5% fall in the most favorable regions of a Ramachandran plot generated by CCP4 PROCHECK with no residues in the generously allowed or disallowed regions.

The 12Y-2 structure was determined by molecular replacement using CCP4 MOLREP. The search model was the 12Y-1 structure without CDR3. Two 12Y-2 monomers (A and B) were identified in the asymmetric unit of the $I2_12_12_1$ space group. Iterative model building and refinement with XTALVIEW/REFMAC5 allowed a complete trace of A and B chains, including extended CDR3 loops. The electron density was well defined in the CDR3 (Fig. 6, which is published as supporting information on the PNAS web site). The refinement strategy was as described for 12Y-1. The final refinement included the TLS parameters for each molecule individually and converged to R/R_{free} values of 17.6/24.7% for the 18.12- to 2.18-Å range of data. As for 12Y-1, only the libration tensor was significant, although less anisotropic. The final 12Y-2 model comprises residues 1–113 for each A and B chain and 358 water molecules. In total, 93.4% of residues are in the most favored regions of the Ramachandran plot, with no residues in the generously allowed or disallowed regions. This result indicates that the 12Y-2 model is consistent with a highly refined protein structure.

Results

The V_{NAR} Structures. We previously identified two closely related V_{NAR}s targeting apical membrane antigen 1 of *Plasmodium falciparum* malarial parasites (19). These proteins, designated 12Y-1 and 12Y-2, were isolated from a library containing a broad mixture of type 2 V_{NAR} framework scaffolds derived from the native wobbegong shark repertoire, combined with both naturally occurring and synthetic CDR3 sequences (13). Whereas the 12Y-1 and 12Y-2 CDR3s fit into the synthetic category, their lengths (16 and 18 residues, respectively) and amino acid composition are typical of naturally occurring IgNAR antibodies (Fig. 7*a*, which is published as supporting information on the PNAS web site). Fig. 1 *a* and *b* present the crystal structures of these two proteins. The 12Y-1 asymmetric unit contains one molecule, the 12Y-2 crystal asymmetric unit contains two molecules (chains A and B) differing slightly in their loop structures. The relative disposition of these two 12Y-2 monomers is described by a rotation of 176.2° and screw translation of −1.1 Å. In the following sections, we analyze different structural aspects of these V_{NAR} domains, with a view to determining their evolutionary lineage.

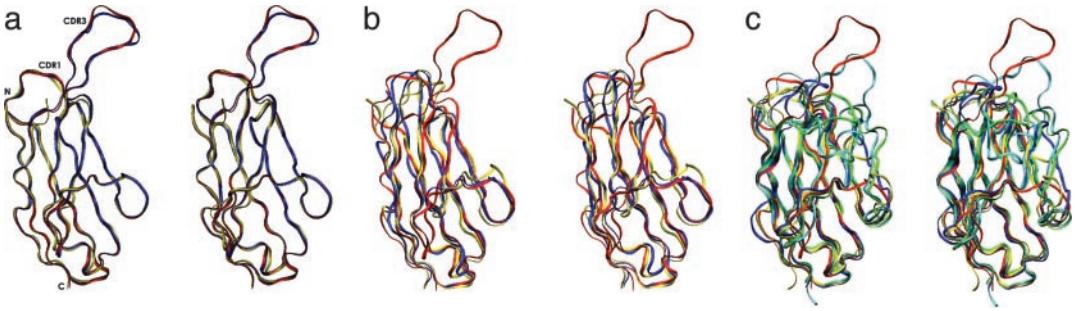

Fig. 2. VMD (44) stereo images of superimposed IgSF domains in ribbon representation. (*a*) The 12Y-2 A chain (red), 12Y-2 B chain (blue), and 12Y-1 (yellow). (*b*) The 12Y-2 A chain (red), telokin (blue) [Protein Data Bank (PDB) ID code 1FHG], and NCAM domain 1 (yellow) (PDB ID code 1QZ1). (*c*) The 12Y-2 A chain (red), human TCR V_α (blue) (PDB ID code 1A07), human V_H (green) and V_L (yellow) (PDB ID code 1IGM), and camel V_HH (cyan) (PDB ID code 1MEL).

V_{NAR}s Are Similar to Cell Adhesion Molecules (CAMs). The folding topologies of both 12Y-1 and 12Y-2 structures show the characteristic Ig superfamily (IgSF) fold, identified by a β-sandwich structure formed by two β-sheets, packed face to face and linked by a disulfide bond between strands B and F (27, 28). IgSF frameworks are further classified into V-, C-, I-, and I2-set categories, based on β-strand number, configuration, and H-bond pattern (27). The 12Y-1 and 12Y-2 frameworks exhibit a folding topology that resembles both the intermediate (I-set) and variable (V-set) folds (Figs. 1*c* and Fig. 7*b*). For example, V-set, I-set, and V_{NAR} proteins all have a typical kink in the first strand (A′) mediated by a conserved cis-proline (Pro-7), which allows the first part of the strand (A) to H bond to one part of the β-sandwich sheet and the second part (A′) to H bond to the extended G strand of the other β sheet (Fig. 1*c* and ref. 29). Like V-set proteins, 12Y-1 and 12Y-2 also have bulges in the C-terminal G strand (conserved Gly-Ala-Gly motif) and in the C′ strand. Most significantly, 12Y-1 and 12Y-2 resemble I-set proteins in their short C′ strand (three H bonds) and a very short C″ strand (labeled as D′ in Fig. 1*c*), which atypically switches from one β-sheet to the other, such that a single H bond links it to the D strand rather than the C′ strand as in V-set domains.

The 12Y-2 chain A and B (rms deviation = 0.53Å for Cα of 113 residues), and the 12Y-1 framework (rms deviation = 0.72Å) are closely related (Fig. 2*a*). Further structural comparison of the 12Y-2 chain A with diverse variable and intermediate set proteins shows significant similarity to I-set molecules such as the neural CAMs (NCAMs) (30–32) and telokin (31, 33). This analysis is heavily biased by the absence of the extended C′ and C″ strands (Fig. 2*b*). In contrast, comparison with conventional T cell receptor (TCR) V_α, and V_H, V_L, and single-domain V_HH antibodies, shows little consistent structural identity, strongly suggesting that these proteins did not evolve from a conventional antibody molecule (Fig. 2*c*).

The V_H/V_L Interface. In heterodimeric immune receptors, i.e., antibodies and TCRs, the paired domains interact across a nonsolvent exposed hydrophobic interface, formed by a conserved patch of residues on the AGFCC′ β-strands, with additional CDR3 interactions. In contrast, many IgSF-based cell surface receptors are single domains in solution, and this face assumes a more charged/polar character. In Fig. 3, we compare this region on the 12Y-2 V_{NAR}, a camel V_HH, a TCR V_α, an antibody V_H, and NCAM and telokin domains. The hydrophobic region of interdomain contact is immediately apparent for the TCR and antibody domains, centered around aromatic residues at the center of the interface (Fig. 3 *c* and *d*). The surface character is altered for V_HH domains, for example, by mutations Leu45Arg and Gly44Glu, to give a more charged character (Fig. 3*b*). However, the relatively short evolutionary time frame since the development of these single-domain antibodies in the Camelidae (34) means that other solutions have also been adopted, for example, the illustrated antibody where part of the CDR3 descends to partly cover the former V_L interface (Fig. 3*b*).

In contrast, for the V_{NAR} domains, this face is dominated by the charged and polar residues Tyr-37, Glu-46, Lys-82, Gln-84, Arg-101, and Lys-104 (Fig. 3*a*). Residues Glu-46, Lys-82, and Lys-104 especially are well conserved, and here, they form a charged pocket with a pattern of H bonds between side chains [i.e., Glu-46(Oε1)–Lys-104(Nζ)] and to adjacent water molecules (i.e., Glu-46(Oε2)–W–Lys-82(Nζ)). The central Tyr-37 is well conserved as an aromatic species across the IgSF, and it and residues Gln-84 and Arg-101 also participate in forming a framework-CDR3 H-bond network [Arg-101(Nη2)–Gln-84(Nζ2]; Tyr-87(Oη)–Arg-101(Nη2). The combined effect is to form a charged pocket ringed by water molecules. A similar situation is observed for NCAM, where this face is dominated by the charged residues Lys-76 and Glu-88 (Fig. 3*e*), and for telokin (Fig. 3*f*), where a charged and polar interface is maintained by a combination of H bonds.

The Hypervariable Loops. In Fig. 4*a* we show the relative positions of the three classically defined antibody CDR or hypervariable loop regions. Sequence alignments show IgNAR antibody variability confined to the CDR1 and CDR3, and this result is confirmed by our structural analysis. CDR1 is the minor loop component, invariant in length and limited in diversity, and is confined to residues 28–33 (12Y-1: ^{N}SYGLES^{C}; 12Y-2: ^{N}SFELKD^{C}), with a topology close to that of canonical structure 2 observed for antibody V_L domains (ref. 35 and Fig. 4*b*). Where a half-cystine is present in the V_{NAR} CDR1, it is exclusively at positions 29 or 32, where the side chains extend outward and upward toward the CDR3 loop, ideally positioned to make contact through an interloop disulfide bridge (Fig. 4*c*). Given the enormous topological latitude inherent in the highly diverse CDR3s, a wide variety of CDR3 conformations can clearly still be adopted, despite the restraints imposed by stabilizing disulfide linkages.

The 12Y-2 CDR3 loop is present in two crystal forms, corresponding to chains A and B, and extends from residues Phe-86 to Glu-103. Unusually, the chain A CDR3 adopts a clear β-hairpin configuration with β-strands from Phe-86–Leu-89, and Leu-98–Glu-103, separated by a flexible loop (Pro-90–Ser-97). For chain B, the β-hairpin extends even further into the CDR3 with residues Phe-86–Asp-93 and Tyr-96–Glu-103 involved in β-strand formation (Figs. 4*d* and 6). Structurally, the β-hairpins are formed by multiple main-chain H bonds, resulting in a CDR3 loop structure that extends $\approx$20 Å outward and upward from the Ig framework, and tipped by the bulky side chains of tyrosine residues (Fig. 4 *a* and *d*). Such extended antigen-binding paratopes have been observed in but a limited number of antibodies, for example, the camel antilysozyme V_HH

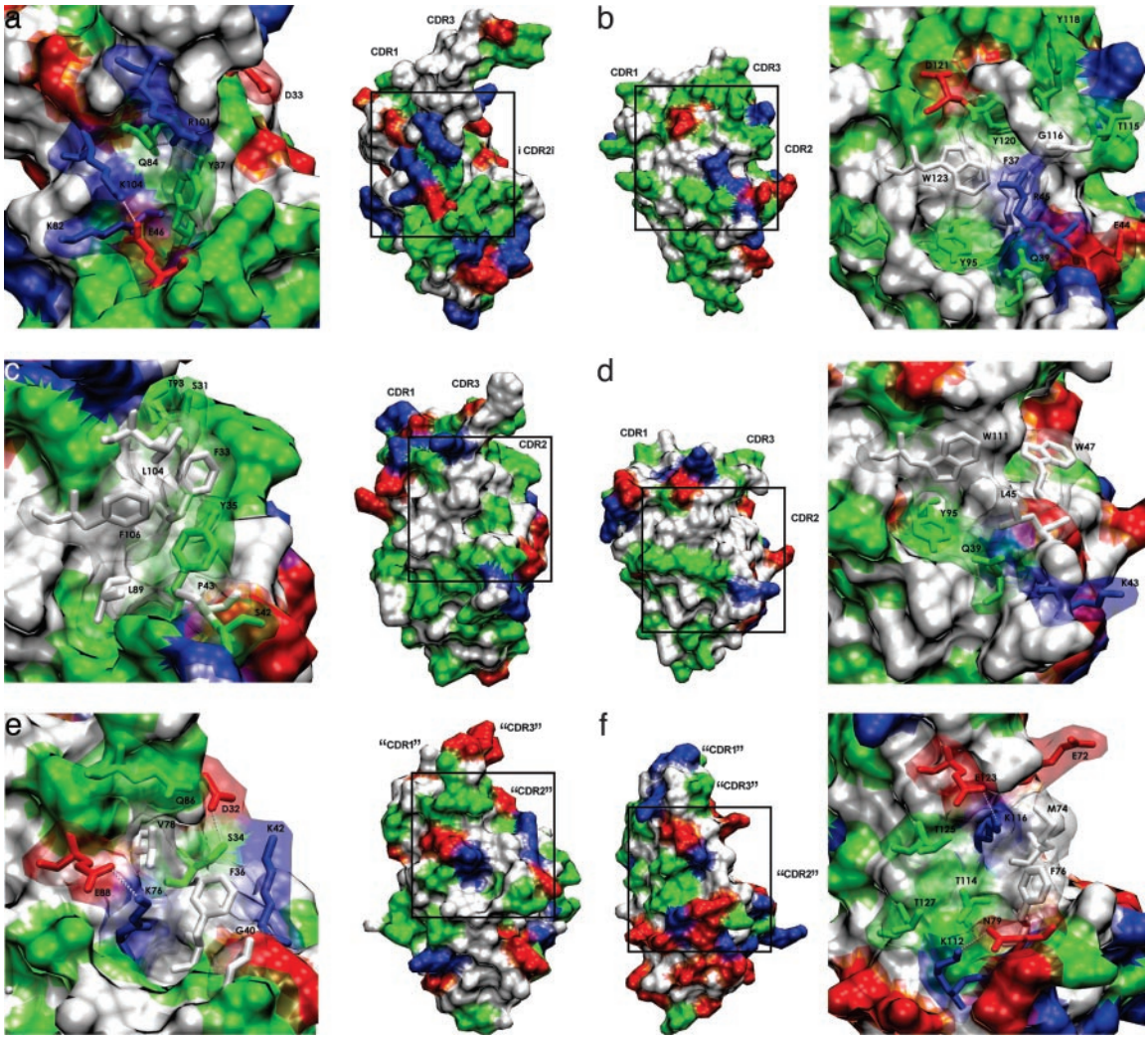

Fig. 3. VMD (44) molecular surfaces for variable domains. (*a*) V_{NAR} 12Y-2 chain A. (*b*) Camel V_HH (PDB ID code 1MEL). (*c*) Human TCR V_α (PDB ID code 1A07). (*d*) Human V_H (PDB ID code 1IGM). (*e*) NCAM domain 1 (PDB ID code1QZ1). (*f*) Telokin (PDB ID code 1FHG). A closeup view of the V_H/V_L type interface is shown for each molecule (boxed area), with some residues visible in licorice-like representation through the transparent surface. Blue, red, and green correspond to basic, acidic, and polar residues, respectively. Dashed lines are H bonds ($\leq$ 3.0 Å).

cAb-Lys-3 (36), and the H3 loop of human antibody b12, which penetrates deeply into the HIV gp120-binding cleft (37). Thus, structures based on the 12Y-2 CDR3 with its extended β-hairpin structure may prove ideal for penetrating buried clefts and cavities, i.e., enzyme active sites, parasite coat proteins, or viral canyons. Additionally, we previously identified two mutations in the 12Y-2 CDR3 that independently enhanced the antigen binding affinity $\approx$10-fold (19). These mutations, Pro90Leu and Phe100Leu, probably acted to increase the flexibility of the β-hairpin (Fig. 4*d*). For example, the three aromatic residues: Phe-29 of CDR1, and Tyr-87 and Phe-100 of CDR3, are involved in stabilizing C-H$\cdots\pi$ interactions. In contrast, these 12Y-2 CDR3 residues are replaced with noninteracting Arg-87 and Pro-98 in 12Y-1, reducing the stability of the CDR3.

Above, we discussed the effect of the unusual V_{NAR} C′C″D strand topology on structural classification, we now turn our attention to its impact on antigen recognition. The V_{NAR}"CDR2" loop is nonexistent, replaced by a short β-turn at the bottom of the molecule. This is graphically illustrated in Fig. 4*e*, where the V_{NAR} CDR2 is aligned with that of a typical human antibody. The "bottom" position of this loop, combined with the low sequence variability, strongly suggests that this region has little impact on the interaction with antigen. However, the loss of the conventional C″ and D strands suggests a possible alternative model for antigen binding, where the extended 12Y-2 CDR3 loop combines with the large concave pocket opened in the absence of the conventional CDR2 (Fig. 4*a*). Additional structural variability is also observed in the 12Y-1 and 12Y-2 structures for the C strand loop, ranging from residues Lys-40 to Glu-46, just before the CDR2 (Fig. 4*e*). Comparison of V_{NAR}s from different shark species shows significant sequence heterogeneity in this region (results not shown), which most likely reflects an area under low selection pressure and susceptible to some degree of structural plasticity.

An Unusual Dimer. Lack of an extended CDR2 loop also has a significant impact on the interaction between isolated V_{NAR} domains. Both 12Y-1 and 12Y-2 form crystallographic twofold symmetry dimers, which form a continuous eight-stranded β-sheet underneath the CDR loops (Fig. 5*a*). Contact areas are highly conserved between the 12Y-1 and 12Y-2 proteins, despite different crystal forms (12Y-1 tetragonal and 12Y-2 orthorhombic). The interdimer relative disposition of monomers can be described as rotation by 6.9° and screw translation by -0.43 Å.

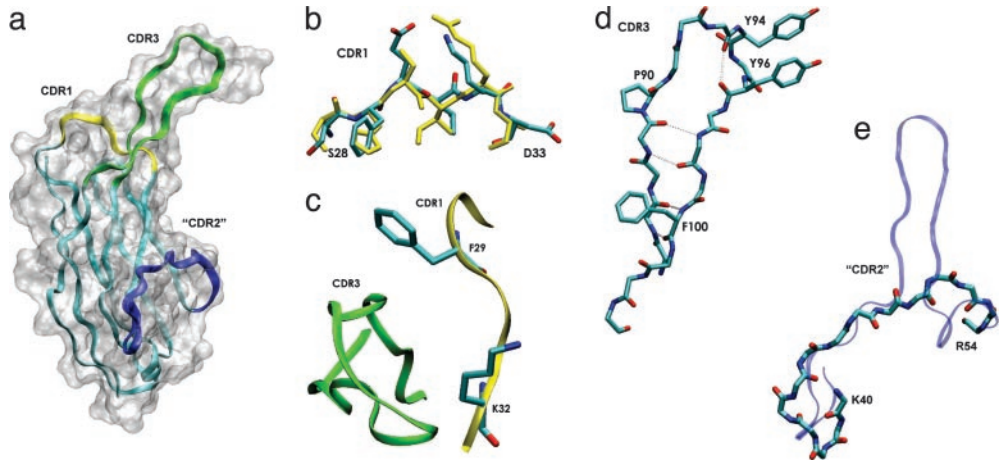

Fig. 4. VMD (44) diagrams for CDR in V_{NAR} 12Y-2. (*a*) The 12Y-2 chain A structure in ribbon representation overlaid with transparent molecular surface. CDR1, CDR2, and CRD3 are in yellow, blue, and green, respectively. (*b*) Overlay of the CDR1 of 12Y-2 chain A (cyan) and human V_L (yellow) (PDB ID code HZH; rms deviation = 1.22 Å^2) in licorice representation. (*c*) Positioning of CDR1 (yellow) and CDR3 (green) in 12Y-2 chain A in ribbon representation. Residues Phe-29 and Lys-32 (possible half-cystine positions) are oriented toward the CDR3, ideally placed to make interloop contacts. (*d*) Backbone of the CDR3 in 12Y-2 chain A (86–103). β-hairpin main-chain H bonds are shown by dashed lines. Residues at the mutation positions, Pro90Leu and Phe100Leu, and residues at the tip of the loop, Tyr-94 and Tyr-96, are shown with side chains. (*e*) Overlay of the backbone CDR2 of 12Y-2 chain A in licorice representation and human V_H (blue) (PDB ID code1HZH) CDR2 in ribbon semitransparent representation.

A comparison of the 12Y-1 and 12Y-2 dimeric forms shows that the interaction surface between the twofold monomers is not continuous and can be subdivided into three areas: (*i*) the main-chain β-sheet interactions between D strands; (*ii*) the interaction between CDR1 loops; and (*iii*) the interactions between CDR3 loops (Fig. 5*b* and Table 2, which is published as supporting information on the PNAS web site). Whereas the contact between CDR3 loops in 12Y-2 is extensive, the dimeric arrangement is preserved in 12Y-1 crystals notwithstanding the more flexible and significantly distorted CDR3, indicating that the conformation of CDR3 is not absolutely required. Thus, the most significant dimer contacts are probably mediated by the CDR1 loops and especially by the D strands, where the main-chain interactions are independent of side-chain variation (Table 2). With a buried surface area of ≈1,760 Å^2, which is ≈27% of the total molecule surface, this appears to be a true protein–protein interaction (38), and the shape correlation statistics S_c (39) of 0.64 is in the range of values for antibody/protein antigen and V_H/V_L chain interfaces. Thus, we suggest that this configuration is a general phenomenon for independent IgNAR variable domains (i.e., not tethered to constant domains), and indeed we have also observed such dimeric species in solution (Fig. 8, which is published as supporting information on the PNAS web site) and in other recombinant V_{NAR}s proteins (ref. 9 and data not shown).

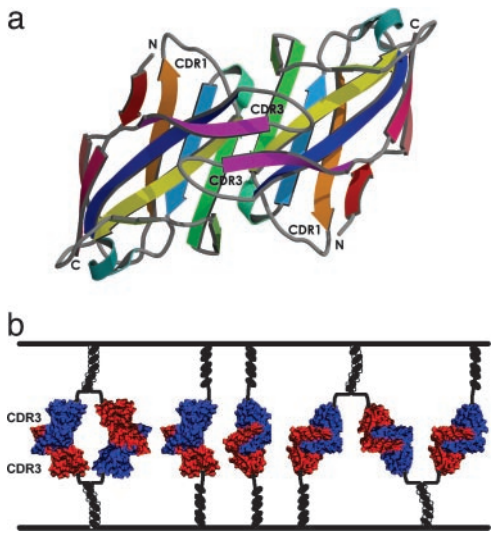

Fig. 5. Structures of the V_{NAR} 12Y-2 twofold symmetry dimer. (*a*) View from the top of the CDR3. Representation is as in Fig. 1. (*b*) A model for cell–cell adhesion mediated by an IgNAR variable domain progenitor. The 12Y-2 dimers are modeled as cell-surface molecules, mediating a cell–cell adhesion event. Individual subunits are represented by red and blue surfaces. Possible contacts include single-variable domains (*Center*), antibody-like dimers (*Left*), or a zipper-like interaction involving multiple molecules (*Right*). CDR3 are indicated, and multiple constant domains connecting the variable domains to the cell surface are represented as black and white ellipsoids.

Discussion

We hypothesize that type 2 IgNAR single-domain antibodies evolved from a primitive cell-surface adhesion molecule, and not from the classical antibody/TCR lineage. Two lines of evidence support this proposal. First, V_{NAR} domains are clearly structurally closer to the I-set of the IgSF, for example, the CAMs, than to the V-set, which contains all known antibodies. The absence of a CDR2 loop is compelling evidence that V_{NAR} domains are a considerable evolutionary distance from other molecules of immune surveillance, particularly because sharks also possess antibody isotypes with intact and functional CDR2 loops. We believe that convergent evolution between IgNAR antibodies and CAMs is a less viable hypothesis, because such convergence would require a conventional antibody to both lose its hitherto functional CDR2, and arrive at a structural loop solution similar to that observed for the CAMs, and not one of the large number of other loop topologies available. In contrast, when such reevolution occurred in the camelids, regions mutated were the highly variable CDR3, and the V_H/V_L interface. The CDR2 was relatively unchanged, both in length and position (15, 18).

Second, the majority of V_{NAR} diversity is concentrated in the extended CDR3 loop, whereas the archetypal V_H/V_L antibodies rely on contributions from up to six CDR structures. With no contribution from a CDR2, we were further surprised by the apparently minor role played by CDR1 in forming the antigen-binding site. This contribution may be little more than to support

the CDR3 loop, which is consistent with our mutational data that it is difficult to affinity-enhance V_{NAR}s by CDR1-directed window mutagenesis (13). Whereas analysis of native IgNARs suggests that affinity maturation processes are biased toward this region (40), we have also found that such changes can destructively impact on protein stability and solubility (13). Certainly, there is not extensive CDR1 variability, and logically, it is simpler to evolve from a monospecific cell-surface molecule to a hypervariable antibody by modification of just one region of the molecule. For V_{NAR} domains this goal has apparently been achieved by acquisition of three germ-line diversity (D) regions to encode the highly diverse CDR3 loop structures (6).

Our hypothesis is supported by the observation that IgNAR variable domains display another type of dimer, which is independent of CDR3 sequence variability. These are not apparent in the quaternary structure of the extant IgNAR antibody, which immune electron microscopy showed as independent variable domains tethered by the interaction between the five constant domains of each antibody chain (11). Rather, we suggest that the dimer form is an evolutionary carryover from a primitive IgNAR, before its subversion to a role as an antibody in the adaptive immune system. In this scenario, the dimer we have identified mediates cell–cell adhesion and/or signaling events, and the precise sequence of the CDR3 loop is immaterial. Fig. 5*b* presents a possible model for such an interaction, which could have occurred either in the context of single variable domains, by formation of homodimeric conformations, or by a continuous zipper-like interaction involving multiple molecules in a surface array. A similar mechanism has been proposed for NCAM, although in that case the dimer does not form across the C and D strands of the Ig framework (32), and for intercellular adhesion molecule 1, which also dimerizes over a hydrophobic interface on the BED β-sheet (41). In our model, the constant domain organization is irrelevant, with interchangeable number and configuration, a plasticity that readily occurs both between leukocyte antigens in an evolutionary sense, and in the most obvious form as the class switch between antibody isotypes.

In summary, we describe the first reported structures, to our knowledge, of IgNAR variable domains, and propose a distinct immune receptor lineage in sharks. Because there has been no identification of IgNARs in the bony fish or in higher vertebrates, this isotype may be limited to the cartilaginous fish, placing the evolutionary split of the order of 450 million years ago. This work also illustrates the potential of V_{NAR} proteins as cleft-binding antibodies, with paratopes formed by extended β-hairpin structures potentially able to penetrate otherwise cryptic antigenic sites.

We thank Dr. Neil McKern, Ms. Pat Pilling, and Mr. Bert van Donkelaar for assistance with protein crystallization and data collection, Ms. Usha Krishnan for recombinant protein production, and Ms. Meghan Hattarki for BIAcore analysis.

1. Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998) *Nature* **394,** 744–751.
2. Litman, G. W., Anderson, M. K. & Rast, J. P. (1999) *Annu. Rev. Immunol.* **17,** 109–147.
3. Flajnik, M. F. (1996) *Vet. Immunol. Immunopathol.* **54,** 145–150.
4. Kumar, S. & Hedges, B. (1998) *Nature* **392,** 917–920.
5. Schluter, S. F., Bernstein, R. M. & Marchalonis, J. J. (1997) *Immunol. Today* **18,** 543–549.
6. Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. & Flajnik, M. F. (1995) *Nature* **374,** 168–173.
7. Nuttall, S. D., Krishnan, U. V., Hattarki, M., De Gori, R., Irving, R. A. & Hudson P. J. (2001) *Mol. Immunol.* **38,** 313–326.
8. Diaz, M., Velez, J., Singh, M., Cerny, J. & Flajnik, M. F. (1999) *Int. Immunol.* **11,** 825–833.
9. Nuttall, S. D., Krishnan, U. V., Doughty, L., Alley, N., Hudson, P. J., Pike, R. N., Kortt, A. A. & Irving, R. A. (2002) *FEBS Lett.* **516,** 80–86.
10. Dooley, H., Flajnik, M. F. & Porter, A. J. (2003) *Mol. Immunol.* **40,** 25–33.
11. Roux, K. H., Greenberg, A. S., Greene, L., Strelets, L., Avila, D., McKinney, E. C. & Flajnik, M. F. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11804–11809.
12. Diaz, M., Stanfield, R. L., Greenberg, A. S. & Flajnik, M. F. (2002) *Immunogenetics* **54,** 501–512.
13. Nuttall, S. D., Krishnan, U. V., Doughty, L., Pearson, K., Ryan, M. T., Hoogenraad, N. J., Hattarki, M., Carmichael, J. A., Irving, R. A. & Hudson, P. J. (2003) *Eur. J. Biochem.* **270,** 3543–3554.
14. Wu, T. T., Johnson, G. & Kabat, E. A. (1993) *Proteins* **16,** 1–7.
15. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A. & Hamers, R. (1994) *Protein Eng.* **7,** 1129–1135.
16. van den Berg, T. K., Yoder, J. A. & Litman, G. W. (2004) *Trends Immunol.* **25,** 11–16.
17. Richards, M. H. & Nelson, J. L. (2000) *Mol. Biol. Evol.* **17,** 146–155.
18. Su, C., Nguyen, V. K. & Nei, M. (2002) *Mol. Biol. Evol.* **19,** 205–215.
19. Nuttall, S. D., Humberstone, K. S., Krishnan, U. V., Carmichael, J. A., Doughty, L., Hattarki, M., Coley, A. M., Casey, J. L., Anders, R. F., Foley, M., *et al.* (2004) *Proteins* **55,** 187–197.
20. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
21. de La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494.
22. McRee, D.E. (1999) *J. Struct. Biol.* **125,** 156–165.
23. Collaborative Computational Project 4 (1994) *Acta Crystallogr. D* **50,** 760–776.
24. Brünger, A. T. (1992) *Nature* **355,** 472–474.
25. Lamzin, V. S. & Wilson, K. S. (1997) *Methods Enzymol.* **277,** 269–305.
26. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001) *Acta Crystallogr. D* **57,** 122–133.
27. Bork, P., Holm, L. & Sander, C. (1994) *J. Mol. Biol.* **242,** 309–320.
28. Chothia, C., Gelfand, I. & Kister, A. (1998) *J. Mol. Biol.* **278,** 457–479.
29. Spada, S., Honegger, A. & Plückthun, A. (1998) *J. Mol. Biol.* **283,** 395–407.
30. Harpaz, Y. & Chothia, C. (1994) *J. Mol. Biol.* **238,** 528–539.
31. Chothia, C. & Jones, E. Y. (1997) *Annu. Rev. Biochem.* **66,** 823–862.
32. Soroka, V., Kolkova, K., Kastrup, J., Diederichs, K., Breed, J., Kiselyov, V., Poulsen, F., Larsen, I. K., Welte, W., Berezin, V., *et al.* (2003) *Structure (London)* **10,** 1291–1301.
33. Holden, H. M., Ito, M., Hartshorne, D. J. & Rayment, I. (1992) *J. Mol. Biol.* **227,** 840–851.
34. Nguyen V. K., Su, C., Muyldermans, S. & van der Loo, W. (2002) *Immunogenetics* **54,** 39–47.
35. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D. R., Tulip, W.R., *et al.* (1989) *Nature* **342,** 877–883.
36. Desmyter, A., Transue, T. R., Ghahroudi, M. A., Dao Thi, M.-H., Poortmans, F., Hamers, R., Muyldermans, S. & Wyns, L. (1996) *Nat. Struct. Biol.* **3,** 803–811.
37. Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R. & Wilson, I. A. (2001) *Science* **293,** 1155–1159.
38. Lo Conte, L., Chothia, C. & Janin, J. (1999) *J. Mol. Biol.* **285,** 2177–2198.
39. Lawrence, M. C. & Colman, P. M. (1993) *J. Mol. Biol.* **234,** 946–950.
40. Diaz, M., Greenberg, A. S. & Flajnik, M. F. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14343–14348.
41. Casasnovas, J. M., Stehle, T., Liu, J-H., Wang, J-H. & Springer, T. A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4134–4139.
42. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
43. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.
44. Humphrey, W., Dalke, A. & Schulten, K. (1996) *J. Mol. Graphics* **14,** 33–38.