

## Unsolved Mystery

# A Case Of Convergence: Why Did a Simple Alternative to Canonical Antibodies Arise in Sharks and Camels?

Martin F. Flajnik<sup>1</sup>, Nick Deschacht<sup>2,3</sup>, Serge Muyldermans<sup>2,3\*</sup>

**1** Department of Microbiology and Immunology, University of Maryland at Baltimore, Baltimore, Maryland, United States of America, **2** Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Brussels, Belgium, **3** Department of Molecular and Cellular Interactions, VIB, Brussels, Belgium

## Introduction

The success of our adaptive humoral immune response, mediated by B cells, relies on its capacity to produce multiple antigen-recognition specificities against any possible invader. This antigen-recognition specificity is attributed to antibodies, which display remarkable flexibility and diversity. That diversity is generated by cutting and pasting of immunoglobulin (Ig) gene segments during B-cell development to generate functional variable (V) genes, which are selected from two separate regions of the genome to yield so-called “heavy” (H) and “light” (L) chains (Figure 1; see Section 1 in Text S1). Ig H and L chain V genes have been expanded in various vertebrate taxa to yield multiple families and subfamilies [1]. Given the obvious advantage of the diversity afforded by such multichain antibodies, it was extremely surprising to discover that some species—notably cartilaginous fish (holocephalins, e.g., chimeras, and elasmobranchs, e.g., sharks, skates, and rays) and camelids (e.g., camels and llamas)—developed functional H-chain-only antibodies (HCABs) (Figures 1 and 2) [2,3]. Loss of the combinatorial H-L diversification of heterotetrameric (H<sub>2</sub>L<sub>2</sub>) antibodies should result in a handicapped, less diversified antigen-binding repertoire in HCABs; yet, contrary to expectations, high-affinity, antigen-reactive HCABs can be elicited to all studied antigens [4,5]. Even more peculiar, the HCABs function alongside conventional H<sub>2</sub>L<sub>2</sub> antibodies. What explains this remarkable evolutionary convergence of functional HCABs in sharks and camels? What forces drove such distantly related species to produce antibodies with this simple structure? And, if they are so effective why is their prevalence not more widespread in evolution?

## Occurrence of Functional Heavy Chain-Only Antibodies in Camelids and Shark

The antigen-binding fragment of camelid HCABs consists of a single domain known as a variable domain of the heavy chain of HCABs (VHH) (Figure 1) [2]. It was demonstrated that HCABs from dromedaries (one-humped camels) infected with trypanosomes are capable of associating tightly and specifically with these parasite antigens [2]. Subsequently, HCABs were elicited to many different foreign antigens. HCABs are present in all living species of the mammalian family *Camelidae*, but not in other ungulates (hoofed mammals). Approximately equal levels of HCABs and H<sub>2</sub>L<sub>2</sub> antibodies are present in camel blood [6]. The origin of camelid Ig genes encoding HCABs is unambiguous. Phylogenetic analyses have demonstrated that the HCAB-dedicated  $\gamma$  genes were derived from  $\gamma$  genes coding for conventional antibodies. The genes emerged and diverged ~25 million years ago [6,7], after *Tylopoda* split from other mammals (60–80 million years ago)

and before the camel and llama speciation (~11 million years ago) (Figure 2) [6]. Also, the VHH germline genes descended from the classical variable region of a heavy-chain (VH) genes, and the diversity (D) and JH genes are even shared in the formation of the VH or VHH domains. Thus, the emergence of HCABs was a relatively recent event in these species.

In the nurse shark (*Ginglymostoma cirratum*), a novel secreted antigen receptor was reported and named the new antigen receptor (NAR) [3]. It too is composed of two covalently associated H-chains with no associated L-chains, but having six domains (Figure 1). Since the molecule shares several functional features with Ig isotypes, it was renamed IgNAR [8], and its N-terminal V domain is known as V-NAR. Later studies showed IgNAR to be induced by immunization with protein antigens with similar kinetics as in a typical IgM response [9]. Although the origin of shark IgNAR is poorly understood, IgNAR is found in all elasmobranchs and thus emerged at least 220 million years ago (Figure 2). However, the single-domain V is also found as a T-cell receptor (TCR) in all living cartilaginous fish (so-called NAR-TCR) [10,11], showing that this immune system feature appeared at least 350 million years ago (Figure 2). In addition, V-NAR gene clusters actually are present in the TCR- $\delta$  loci of all cartilaginous fish and thus it remains uncertain whether they originated as a TCR, an Ig, or perhaps a common antigen receptor ancestor [10].

## Convergent Evolutionary Steps to Generate Functional Heavy Chain-Only Antibodies

Besides the phylogenetic origins, the most significant question regarding molecular convergence is: “What are the critical events or properties that must be introduced to produce a functional HCAB?” We think that three (and perhaps four) crucial requirements must be fulfilled to obtain homodimeric H-chain

**Citation:** Flajnik MF, Deschacht N, Muyldermans S (2011) A Case Of Convergence: Why Did a Simple Alternative to Canonical Antibodies Arise in Sharks and Camels? *PLoS Biol* 9(8): e1001120. doi:10.1371/journal.pbio.1001120

**Published:** August 2, 2011

**Copyright:** © 2011 Flajnik et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

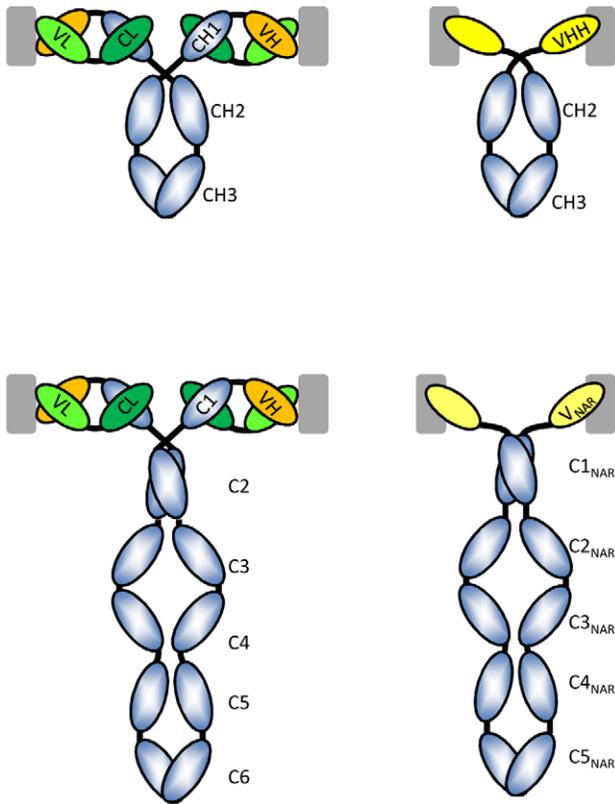
**Funding:** This work was supported by the “Instituut Voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen” to N.D. and the VIB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** C, constant; D, diversity; FR, framework region; H, heavy; HCAB, heavy-chain antibody; Ig, immunoglobulin; J, joining; L, light; NAR, new antigen receptor; TCR, T-cell receptors; VH, variable region of a heavy chain; VL, variable region of a light-chain; VHH, variable region of a heavy chain of a heavy-chain antibody

\* E-mail: svmyulde@vub.ac.be

Unsolved Mysteries discuss a topic of biological importance that is poorly understood and in need of research attention.



**Figure 1. Schematic of classical  $H_2L_2$  and  $H_2$  homodimeric HCABs.** The left column displays the classical antibodies with two identical H-chains and two identical L-chains as they occur in mammals (IgG, top) and cartilaginous fish (IgW, bottom). The L-chain is in green and the antigen-binding site is formed by the paired VH and VL domains. The top right figure is the HCAb as it occurs in sera of camelids, the CH1 domain is missing and there is no L-chain. The antigen-binding site consists of one single domain, known as VHH. The H-chain of the IgW comprises six C domains and a variable domain at the N-terminal end, whereas the IgNAR (bottom right) is a homodimer of a H-chain with five C domains and a V-NAR at its N-terminal end. Note that the equivalent of the first C domain is absent. All of the antibodies are bivalent and the recognition of a possible antigen (gray square) is shown. The VH-VL associated preferentially with flat surfaces on the antigen, whereas the VHH or V-NAR has a preference to interact with cavities on the surface of the antigen.  
doi:10.1371/journal.pbio.1001120.g001

antibodies: (1) the first constant domain (CH1) within the H-chain that normally serves to anchor the L chain should be absent or extensively modified to prevent its retention within the endoplasmic reticulum (see Section 2 in Text S1) [12]; (2) an extensive and diverse primary repertoire of single V-domains must be generated from which to select the most appropriate binders (see Section 3 in Text S1 on how single V domain sequences are diversified) [6], and the selected autonomous V domain should be able to bind antigen with high-affinity-like conventional antibodies; (3) the V domain must be soluble (i.e., nonsticky) in the absence of a variable region of a light-chain (VL) partner (see Section 4 in Text S1) [13]; and (4) although not absolutely required, noncanonical disulfide bonds should be present that stabilize the single V domain in the absence of VL, and to diversify the repertoire of the antigen binding site (see Section 3 in Text S1) [6,14,15].

These requirements underlying the molecular convergence between the HCABs of camelids and the IgNARs of cartilaginous fish have been met, resulting in the appearance of functional

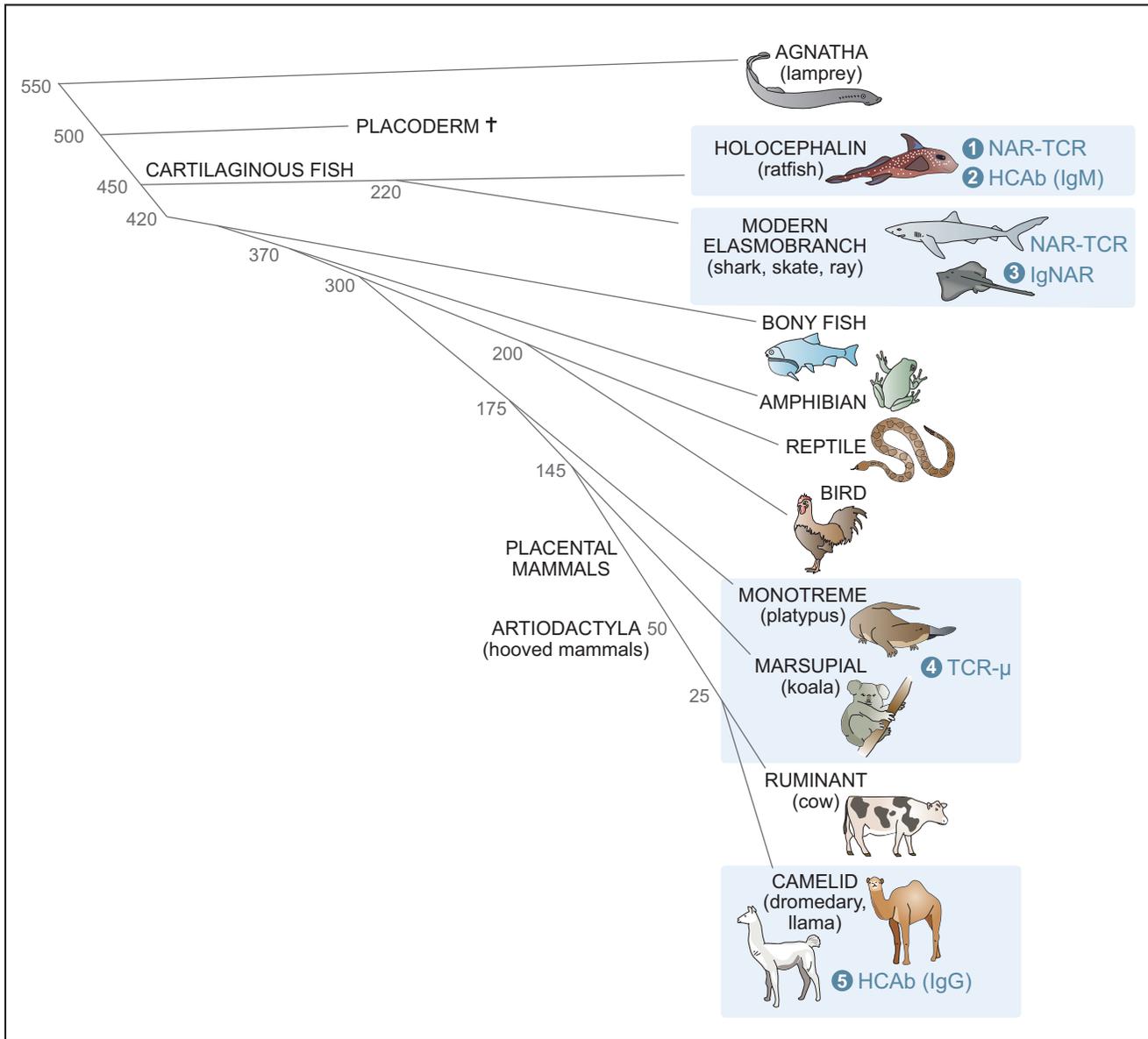
HCABs that have converged at multiple levels: (1) The structural level: Camelid HCABs and Ig-NARs are both homodimeric H-chain antibodies devoid of any L-chain. (2) The sequence level: There is a subtle amino acid sequence convergence of the VHH and V-NAR antigen binding domains (i.e., increased frequency of charged and polar residues at the “former VL interface”). (3) The mechanistic level: The strategy whereby H-chain avoids interaction with L-chain (absence of CH1), and the mechanisms that have been employed to create a highly variable antigen-binding repertoire in absence of the H-L combinatorial diversity. (4) The functional level: The VHH of camelid HCABs and the V-NARs of shark IgNARs target similar epitopes that are distinct from the preferred epitopes of classical  $H_2L_2$  antibodies.

### What Drove the Evolution and Selection of Homodimeric H-chain Antibodies?

Such a molecular convergent evolution of homodimeric HCABs in species as diverse as camels and sharks suggests that the immune systems of their ancestors faced comparable stress, and in order to be selected, propagated, and fixed in the population these molecules likely provided a strong advantage in protecting the animal. Although multiple hypotheses can be proposed, the driving force remains elusive. Possibilities could be an L-chain variant prone to malfunction (by amyloidosis for example [16]), or an L-chain used as a coreceptor for a virus that was B-cell tropic. In such an event, the humoral immune system would benefit from the development of antibodies without L chains. Once HCABs were designed and widespread in the species, the defective L-chain was no longer expressed; or it was modified leading to alleviation of its malfunctioning; or to prevention of virus recognition so that the viral threat disappeared.

In an alternative and possibly more realistic scenario, HCABs emerged from the need (or advantage) for the adaptive humoral immune system to develop antibodies with a smaller, prolate (rugby ball-like)-shaped, single-domain, antigen-binding V domain that could complement canonical antibodies with their planar or concave binding sites (see Section 5 in Text S1). Such minimal-sized, single-domain antigen-binding fragments have smaller footprints on the antigen and provide access to cryptic epitopes that are inaccessible to VH/VL-based antibodies, or to epitopes with a concave architecture, such as the catalytic clefts of enzymes [17]. Indeed, from several crystal structures of shark V-NAR and camelid VHH bound to antigen it seems clear that the HCABs of camelids and sharks are better equipped to target these epitopes [15,18–20].

While the identification of the driving force for the emergence of homodimeric H-chain antibodies is conjectural, the rationale for the presence of unique mutations in V-NAR and VHH framework regions (FRs) and the presence of disulfide bonds in the hypervariable loops are easier to envision. Amino acids in the second FR (FR2), the region between the first and second hypervariable (i.e., antigen-binding) loops are very conserved and hydrophobic in conventional multichain antibodies and are used in VH/VL pairing (Figure 3) [21,22]. Mutations in these FR2 residues from hydrophobic to charged/polar amino acids are required for the V domain to become independent of its VL partner [4,13]. The generation of non-canonical, intradomain disulfide bonds between the long antigen-binding loops of V-NAR and VHH increases the stability of the domain, which might have been required to cope with the harsh physiological conditions to which these antibodies are exposed [5], i.e., elevated concentrations of urea in shark blood and an arid environment with extreme temperatures for camelids. In addition, the presence of an interloop or a loop-FR disulfide bond probably restricts the conformational flexibility of the antigen-binding loops in



**Figure 2. Simplified vertebrate phylogenetic tree emphasizing those taxa possessing single-domain Igs or TCRs.** Blue boxes display animal groups shown, or predicted, to have single-domain variable regions. There are five known single domains, at least four of which were derived by convergent evolution, holocephalin HCab, elasmobranch IgNAR, monotreme/marsupial Ig $\mu$ , and camelid HCab. The numbers on the left indicate divergence time (millions of years ago) for the various vertebrate taxa. doi:10.1371/journal.pbio.1001120.g002

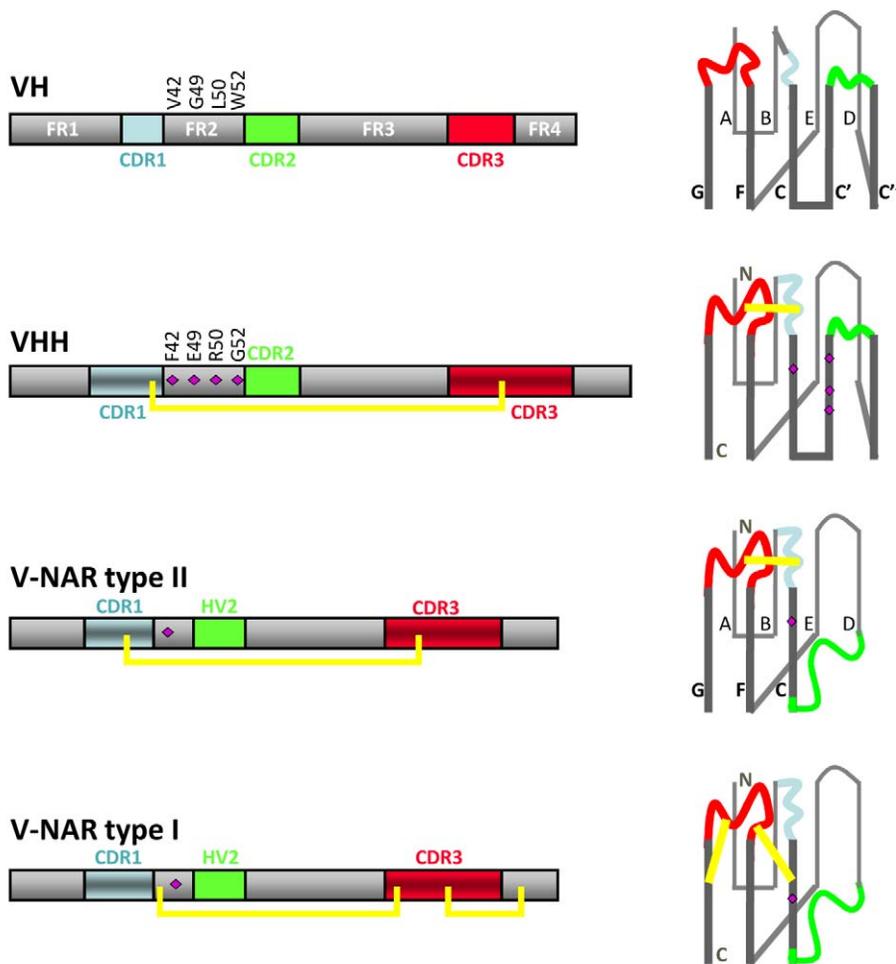
the antigen-free V-domain [19]. This conformational loop constraint in the antigen-free antibody will lead to a lower entropic penalty upon antigen binding, where it is expected that the loops are somehow fixed in a unique position within the antibody-antigen complex.

**Conception of HCAbs in Widely Different Contexts**

Camels and sharks, two evolutionarily remote species (Figure 2), display extreme molecular convergence of their Igs. Both species have a dichotomous adaptive humoral immune system with both conventional and homodimeric antibodies and share the features described above. However, the camelid homodimeric antibodies were derived from conventional genes of the IgH locus by a

relatively recent adaptation [7], while IgNAR of the shark originated from an ancient, undefined evolutionary event. Shark Ig-NARs might display the primordial condition, although it remains enigmatic how the unassociated V regions of a single-chain antibody could evolve into dimeric Vs; perhaps the high salt/urea conditions described above was one driving force in selecting for a dichotomous immune system.

The question must be addressed: if, indeed, the HCAbs offer a complementary binding site to conventional VH/VL antibodies, why haven't they been found in all vertebrates? Their presence in all cartilaginous fish, either as a TCR or an Ig [10], is a testament to the unique organization of their antigen receptor genes. While all other vertebrates employ the so-called translocon-type organization (VnDnJnC), cartilaginous fish Ig genes are encoded in clusters (VDJC)n (Figure 4) [10,11]. Such an organization



**Figure 3. Schematic of the rearranged VH, VHH, and V-NAR type I or type II.** At left are displayed the linear sequence hallmarks such as the complementarity determining region (CDR) within the more conserved framework (FR) residues (gray). The disulfide bonds connecting the CDR3 with the CDR1 or FR in a VHH or V-NAR are shown by the yellow line. The folded structures of the V domains with the nine or seven  $\beta$  strands (named A to G in V-NARs, and A to G with insertion of strands C' and C'' between C and D for VH and VHH) forming two  $\beta$ -sheets are shown on the right. The purple diamonds on the structure denote the VHH hallmark residues in FR2 or the polar charged residues in the V-NARs. The N- and C-terminal ends of the polypeptides are shown in the structure of VHH and V-NAR type I. doi:10.1371/journal.pbio.1001120.g003

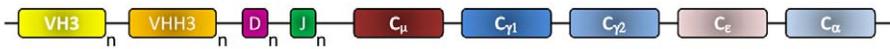
affords the sharks freedom to evolve new Igs “without much evolutionary fuss,” such as germline-joined clusters for generation of binders of conserved antigens, exchange of genes between TCR and Ig loci, and the single-domain V, HCABs described here. Indeed, in the holocephalin *Hydrolagus collicii* (spotted ratfish) there has been CH1 loss in an IgM cluster concomitant with a CH2 exon duplication, which likely resulted in the emergence of another homodimeric IgM (Figure 2) [23]; the protein encoded by this gene has not been examined, but it is likely to function as an L-chainless, H-chain dimer, on the basis of the criteria detailed above. Single-domain Vs have evolved also in monotremes and marsupials via generation of an entirely new locus, which is a chimera between Ig and TCR loci (Figure 2) [24]. By contrast, the mechanism by which both camelid HCABs and conventional antibodies are generated from the same locus during B-cell maturation remains elusive. A novel addition to the classical IgM-IgG class switch must have evolved to allow production of these camelid HCABs [25]. Thus, HCABs have emerged by convergence at least four times in vertebrate evolution, and we believe that they are clearly useful and, anthropomorphically speaking, all species would love to have

them; however, the organization of Ig and TCR loci in all vertebrates besides cartilaginous fish precludes, or makes quite difficult, the capacity to generate them. Consistent with the idea that single-domain Vs are a useful innovation, once they emerged in the ancestral camelid and cartilaginous fish species, HCABs have been perpetuated as a major immune system attribute in all of the descendant species.

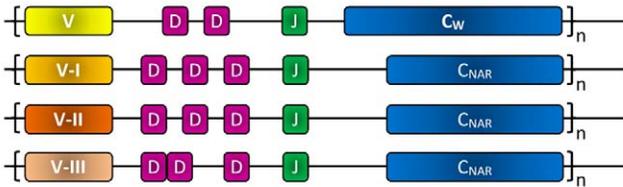
## Conclusions

We have described the convergent features of single-chain antibodies that permitted their emergence and reemergence and the constraints that may have precluded their existence in all vertebrates. On the basis of recent discoveries, it is possible that they will be found in other species as well, either as an Ig or a TCR. Single-domain Vs have a growing repertoire of uses in both clinical and basic research, a testament to their complement of the conventional VH/VL antibodies; basic research on the animals that have naturally evolved such antibodies will provide refined theories of their dedicated functions, and give greater insight to their emergence several times in vertebrate history.

### Translocon organization



### Cluster organization



**Figure 4. Translocon and cluster genome organization of the antibody genes.** The translocon organization (top) of the gene segments as it occurs in camelids to produce the H-chain of classical antibodies and the H-chain of HCABs. For the former antibody type, one member of the different VH families (shown VH4 and VH3) are rearranged with one of the D and one of the J genes to be transcribed with  $C_{\mu}$  and after a class switch with the  $C_{\gamma 1}$ ,  $C_{\epsilon}$ , or  $C_{\alpha}$ . Note that the greater the number of families and the more members per family, the larger the repertoire will be. For the production of HCABs, one of the VHH3 genes (or the VH4) has to be rearranged with one of the D and one of the J genes of the same translocon before transcription occurs with the  $C_{\mu}$ , and after a class switch, with the  $C_{\gamma 2}$  (or  $C_{\gamma 3}$  not shown). To generate the H chain of an  $H_2L_2$  antibody, a VH3 gene (or VH4) has to recombine with one D and one J minigene and should be transcribed with the  $C_{\mu}$ , or after a class switch with any other C isotype except  $C_{\gamma 2}$  or  $C_{\gamma 3}$ . Cartilaginous fish have a cluster organization of their antibody genes whereby each cluster contains a dedicated V element followed by two or three D genes and a J gene. After rearrangement of these minigenes, cotranscription occurs with the W constant gene or with the C-NAR genes to produce classical  $H_2L_2$  antibodies or HCABs, respectively. There are multiple clusters in the genome of shark and the V minigenes of some clusters belong to either type I, type II (both clustered with 3-D segments), or type III (clustered with 2-D segments) minigenes. doi:10.1371/journal.pbio.1001120.g004

## Supporting Information

**Text S1** Section 1: The rationale for the  $H_2L_2$  antibody-based adaptive humoral immune system in jawed vertebrates. Section 2: Absence of CH1 domain in H chains for HCABs. Section 3: Generation of a large V domain repertoire in absence of the VH-

VL combinatorial diversification. Section 4: Producing soluble V domains in absence of VL partner. Section 5: Unique antigen recognition by V-NAR and VHH. (RTF)

## References

- Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr (1992) EMBO J 11: 603–609.
- Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, et al. (1993) Naturally occurring antibodies devoid of light chains. Nature 363: 446–448.
- Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, et al. (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374: 168–173.
- Muyldermans S, Cambillau C, Wyns L (2001) Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. Trends Biochem Sci 26: 230–235.
- Dooley H, Flajnik MF (2006) Antibody repertoire development in cartilaginous fish. Dev Comp Immunol 30: 43–56.
- Conrath KE, Wernery U, Muyldermans S, Nguyen VK (2003) Emergence and evolution of functional heavy-chain antibodies in Camelidae. Dev Comp Immunol 27: 87–103.
- Nguyen VK, Su C, Muyldermans S, van der Loo W (2002) Heavy-chain antibodies in Camelidae, a case of evolutionary innovation. Immunogenetics 54: 39–47.
- Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, et al. (1998) Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci U S A 95: 11804–11809.
- Dooley H, Flajnik MF (2005) Shark immunity bites back: affinity maturation and memory response in the nurse shark, *Ginglymostoma cirratum*. Eur J Immunol 35: 936–945.
- Criscitiello MF, Saltis M, Flajnik MF (2006) An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. Proc Natl Acad Sci U S A 103: 5036–5041.
- Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11: 47–59.
- Feige MJ, Groscurth S, Marcinowski M, Shimizu Y, Kessler H, et al. (2009) An unfolded CH1 domain controls the assembly and secretion of IgG antibodies. Mol Cell 34: 569–579.
- Davies J, Riechmann L (1994) ‘Camelising’ human antibody fragments: NMR studies on VH domains. FEBS Lett 339: 285–290.
- Nguyen VK, Hamers R, Wyns L, Muyldermans S (2000) Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. EMBO J 19: 921–930.
- 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: 1770–1773.
- Hurle MR, Helms LR, Li L, Chen W, Wetzel R (1994) A role for destabilizing amino acid replacement in light chain amyloidosis. Proc Natl Acad Sci U S A 91: 5446–5450.
- Lauwereys M, Arbabi Ghahroudi M, Desmyter A, Kinne J, Holzer W, et al. (1998) Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. EMBO J 17: 3512–3520.
- De Genst E, Silence K, Decanniere K, Conrath K, Loris R, et al. (2006) Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc Natl Acad Sci U S A 103: 4586–4591.
- Desmyter A, Transue TR, Ghahroudi MA, Thi MH, Poortmans F, et al. (1996) Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Biol 3: 803–811.
- Stanfield RL, Dooley H, Verdino P, Flajnik MF, Wilson IA (2007) Maturation of shark single-domain (IgNAR) antibodies: evidence for induced-fit binding. J Mol Biol 367: 358–372.
- Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R (1994) Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng 7: 1129–1135.
- Vu KB, Ghahroudi MA, Wyns L, Muyldermans S (1997) Comparison of llama VH sequences from conventional and heavy chain antibodies. Mol Immunol 34: 1121–1131.
- Rast JP, Amemiya CT, Litmans RT, Strong SJ, Litman GW (1998) Distinct patterns of IgH structure and organization in a divergent lineage of chondrichthyan fishes. Immunogenetics 47: 234–245.
- Parra ZE, Baker ML, Schwarz RS, Deakin JE, Lindblad-Toh K, et al. (2007) A unique T cell receptor discovered in marsupials. Proc Natl Acad Sci U S A 104: 9776–9781.
- Muyldermans S, Baral TN, Retamozzo VC, De Baetselier P, De Genst E, et al. (2009) Camelid immunoglobulins and nanobody technology. Vet Immunol Immunopathol 128: 178–183.