Published in final edited form as:
J Immunol. 2011 June 1; 186(11): 6357–6367. doi:10.4049/jimmunol.1100116.

Analysis and Modeling of the Variable Region of Camelid Single Domain Antibodies

Aroop Sircar[1,§], **Kayode A. Sanni**[2], **Jiye Shi**[3], and **Jeffrey J. Gray**[1,†]

[1]Department of Chemical & Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, Maryland 21218

[2]Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250

[3]Research and Development, UCB Inc., 755 Jefferson Road, Rochester, New York 14623

Abstract

Camelids have a special type of antibodies, known as heavy chain antibodies (HCAbs), that are devoid of classical antibody light chains. Relative to classical antibodies, camelid HCAbs (cAbs) have comparable immunogenicity, antigen recognition diversity and binding affinities, higher stability and solubility, and better manufacturability, making them promising candidates for alternate therapeutic scaffolds. Rational engineering of cAbs to improve therapeutic function requires knowledge of the differences of sequence and structural features between cAbs and classical antibodies. Here, amino acid sequences of 27 cAb variable regions (V_HH) were aligned with the respective regions of 54 classical antibodies to detect amino acid differences, enabling automatic identification of cAb V_HH complementarity determining regions (CDRs). CDR analysis revealed that the H1 often (and sometimes the H2) adopts diverse conformations not classifiable by established canonical rules. Also, while the cAb H3 is much longer than classical H3 loops, it often contains common structural motifs and sometimes a disulfide bond to the H1. Leveraging these observations, we created a Monte Carlo based cAb V_HH structural modeling tool, where the CDR H1 and H2 loops exhibited a median root-mean-square-deviation (rmsd) to native of 3.1 and 1.5 Å respectively. The protocol generated 8-12, 14-16 and 16-24 residue H3 loops with a median rmsd to native of 5.7, 4.5 and 6.8 Å respectively. The large deviation of the predicted loops underscores the challenge in modeling such long loops. cAb V_HH homology models can provide structural insights into interaction mechanisms to enable development of novel antibodies for therapeutic and biotechnological use.

Introduction

The *Camelidae* family (camels: one-humped *Camelus dromedaries* and two-humped *Camelus bactrianus*; llamas: *Lama glama, Lama guanicoe, Lama vicugna*; alpaca: *Vicugna pacos*), of suborder *Tylopoda*, of order *Artiodactyla* have a special type of antibody in addition to classical antibodies in their serum (1, 2). These antibodies, called heavy chain antibodies (HCAbs), are unique in their absence of the entire light chain and the first heavy chain constant region (C_H1). Antibodies similar to camelid heavy-chain only antibodies (cAbs) have also been found in wobbegong, nurse sharks and spotted ratfish (3). The actual binding region of the cAbs is the N-terminal variable domain of the antibody, referred to as cAb V_HH (commercially known as a Nanobody) (4). Based on the success of classical

[†]Corresponding author. jgray@jhu.edu; Telephone: (410) 516-5313; Fax: (410) 516-5510.
[§]Current Address: EMD Serono Research Institute Inc., 45A Middlesex Turnpike, Billerica, Massachusetts 01821

therapeutic antibodies in arthritis, oncology, inflammatory and immune disorder treatments (5), one biopharmaceutical company (Ablynx) has developed candidate cAb V_HH domains against more than 150 disease targets, and some like anti-thrombotic cAb V_HH have entered phase II clinical trials (6).

In contrast to classical antibodies, cAbs have been found to be stable and active at high temperatures of 90°C and in high concentration of denaturants (7). Furthermore the absence of the light chain reduces the combinatorial complexity associated with random V_L-V_H recombination requiring smaller phage display libraries (8); combination of smaller libraries and good expression levels in bacteria and yeast systems result in increased yield (7). Impressively, the absence of the light chain and the associated amino acid substitutions do not limit the diversity of the epitopes which can be targeted by cAbs in panning experiments, probably because of the larger structural repertoire of the cAb V_HH CDR H1 and H3 loops (7). Additionally, the cAbs (especially in dromedaries) have longer CDR H1 and H3 loops compared to the respective classical CDRs (2), increasing the paratope size. The longer CDRs bind epitopes which are more concave than those of classical antibodies, and they can also inhibit enzymes by entering clefts in catalytic sites (2), Moreover, cAbs have exhibited binding affinities similar to classical antibodies with reported affinities as low as 100 pM, near the best observed by a classical antibody (9).

The unique properties of the cAbs can be attributed to changes in amino acid compositions at key positions (1, 7, 9-12). Most of these mutations change hydrophobic residues to polar residues and occur at V_H positions that would have interacted with either the V_L or C_H1 domains had they been present in a classical-antibody-like orientation (1, 10). cAb V_HH x-ray crystal structures show the usual immunoglobulin fold, typically most similar to the human variable heavy chain (V_H) of family III (13). However, considerable differences have been observed in the CDRs, and some long CDR H3s bend and make contacts with the framework region of the cAb V_HH which, in a classical antibody, would have been in contact with V_L (12). CDRs play a central role in antibody-antigen recognition, thus cAb V_HH structures with the biologically relevant conformations of the unique CDR loops are required to understand cAb V_HH-antigen interactions.

Unfortunately experimental structure determination using x-ray crystallography or nuclear magnetic resonance is laborious, time consuming and expensive, resulting in a gap between the number of available protein sequences and structures. Furthermore, of approximately 65,000 protein structures present in the Protein Data Bank (PDB) (14), there are only around 1100 antibody structures of which around 50 are cAb V_HH structures. The paucity in cAb V_HH structures combined with the reliance on homology modeling for computational design of humanized antibodies for production of at least eleven marketed classical antibodies (5), including Herceptin (trastuzumab or humanized anti-HER2), Zenapax (daclizumab or humanized anti-Tac) and Avastin (bevacizumab or humanized anti-VEGF), highlights the need for a high-resolution cAb V_HH homology modeling tool.

We previously created RosettaAntibody (15), a homology modeling tool for classical antibody variable regions (F_V). RosettaAntibody assembles the sequence-match-based templates for the heavy and light chain framework and the canonical CDRs L1, L2, L3, H1 and H2 templates followed by *ab inito* modeling of the CDR H3 loop and subsequent optimization of the V_L-V_H relative orientation and all CDR loop conformations. While the median global root-mean-square-deviation (rmsd) for short CDR-H3 loops (< 10 residues) was less than 2.0 Å, the prediction for longer CDR H3 loops was worse with median global rmsds up to 6.0 Å. Some of the best loop modeling protocols like Protein Local Optimization Program (PLOP) (16) and inverse kinematic loop building (17) generate 8-13 residue loops of sub-Ångstrom to 3 Å accuracy. However, these algorithms are computationally expensive and limited to short loops in a native environment. Given the poor performance in modeling long loops and the non-native environment of the CDR H3 loop in a homology model, building cAb V_HH CDR H3 loops, which average length of 16 residues (2) (human and murine CDR H3s average 14 and 12 respectively (2)), is expected to be quite challenging. Another F_V homology modeling program, the Prediction of Immunoglobulin Structure (PIGS) server (18), efficiently grafts a CDR H3 structure with the highest sequence homology, but cannot predict novel conformations. Despite uncertainties in CDR H3 predictions, we have demonstrated that flexible backbone docking strategies like EnsembleDock (19) and SnugDock (20) can sometimes compensate for errors in RosettaAntibody homology models by optimizing the paratope for successful prediction of high-resolution antibody-antigen interaction complexes.

In this paper, we analyze the sequences and structures of known cAb V_HH domains and develop a RosettaAntibody-based cAb V_HH homology modeling tool. We test whether the canonical numbering schemes can be applied to cAb V_HH domains, and where they fail, we seek to identify new distinguishing markers. Similarly, we test the classification of canonical loop conformations and seek to update the repertoire appropriately. For H3 loops, we seek conserved structural features that can alleviate the challenge of the long lengths. Finally we test a combined homology modeling procedure and comment on the usefulness of the models.

Materials and Methods

Test Set

All cAb V_HH structures were downloaded from the PDB (14) as of November 10, 2009. The 46 downloaded structures were filtered for redundancies (three or fewer point mutants). For cases where both the bound and unbound forms of the cAb V_HH were present, only the unbound was retained, resulting in 27 unique cAb V_HH structures (Supplementary Figure 1). Shark, camelized human and humanized camelid HCAbs were not included in the test set. For each downloaded structure, the cAb V_HH domain was manually identified and extracted for subsequent analysis. The test set contains 17 camel and 10 llama cAbs, with CDR H3 loop lengths ranging from 8-24 residues (Table 1). For comparison with V_H, we extracted the heavy chain from the 54 antibodies in the RosettaAntibody test set (15).

Sequence and Structure Analysis

The amino acid sequences in FASTA (21) format were derived from the PDB files of the cAb V_HH and V_H domains and aligned using the MUSCLE (22) (v3.7) multiple sequence alignment feature in the SeaView (23) (v4.2) graphical multiple alignment tool. The alignment was manually edited to ensure that reported conserved antibody residues were aligned correctly, especially for the regions immediately preceding and following the CDR loops. Sequence features were identified by visual inspection of the multiple alignment (Supplementary Fig. 1). The cAb V_HH structures were visualized using PyMOL (24) to identify the structural features. PyMOL was also used to compute the C_α-C_α distances of the cystine residues in CDR H1 and H3 that form a disulfide bond. For all other distance and dihedral angle measurements, the Rosetta biomolecular modeling software was used.

We considered supplementing the amino acid sequences of the test set with cAb sequences for which crystal structures were not available. Although we found an additional 136 llama antibody sequences at http://ncbi.nlm.nih.gov, it was not clear if they were from classical or cAbs. To avoid confusing detection of cAb signatures, only the sequences from PDB structures of cAb V_HH domains were used.

Homology Modeling

The homology modeling protocol follows RosettaAntibody to create models by: 1) identifying homologous framework and loop templates from the RosettaAntibody database by maximum BLAST (25) bit score, 2) grafting CDR templates onto the framework, 3) building the CDR H3 loop and 4) globally refining the paratope. The main differences from the standard RosettaAntibody protocol are highlighted. The cAb $V_H H$ structures were appended to the RosettaAntibody antibody database. Except where noted, we ensured that the homology modeling protocol was blind by removing query crystal structures from the database. Similar to the four-residue C-terminal CDR H3 fragment library used in modeling CDR H3 loops in V_H antibodies (15), we created a six-residue C-terminal CDR H3 fragment library specific to cAbs. The six-residue fragments have been classified as STRETCHED or TWISTED as described in results, and those that could not be classified are referred to as NEUTRAL. The template identification is similar to that in RosettaAntibody. Due to the absence of the light chain, templates for the light chain framework and CDRs are omitted and V_L-V_H assembly is unnecessary.

On successful identification of the respective templates, the CDRs are grafted into the framework as previously described (15). Side chain conformations of the grafted loop and the neighboring residues are optimized by rotamer packing (26). In a few cases, grafting creates a broken loop due to framework deviations (also observed in a few cases in canonical antibodies submitted to the RosettaAntibody server (27)). When grafting breaks loops, the loop is repaired by minimal refinement using a combination of small (28), shear (28) and cyclic co-ordinate descent (CCD) (29) moves with side-chain packing following the high resolution CDR H3 loop refinement in RosettaAntibody (15) without side-chain minimization (30).

CDR H3 Loop Modeling and H1 Refinement—The loop modeling follows that in the RosettaAntibody protocol comprised of: 1) A centroid pseudo-atom side-chain representation (31) low-resolution Monte Carlo stage where diverse loop conformations are sampled by large perturbations by fragment insertion (including the stretched-twisted and twisted fragments), and 2) an all-atom high-resolution Monte-Carlo-plus-minimization stage where all the side-chain conformations are optimized and the loop backbone dihedral angles are perturbed minimally to relieve steric clashes. To model cAb $V_H H$ domains with extremely long CDR H3 loops and the larger diversity of CDR H1 loops, the RosettaAntibody protocol was enhanced as follows.

Two bounded harmonic potential terms are added to the scoring function. The first constraint enforces the disulfide bond if cystines are present in CDR H1 and H3 loops. The second ensures that stretched-twisted structures fold such that the *n*-5 residue of CDR H3 is near to residue 46 in the heavy chain framework. The constraint term is:

$$f(x)=\begin{cases} \left(\frac{x-x_{\min}}{\sigma}\right)^2 & \text{for } x<x_{\min} \\ 0 & \text{for } x_{\min}\le x\le x_{\max} \\ \left(\frac{x-x_{\max}}{\sigma}\right)^2 & \text{for } x_{\max}<x\le x_{\max}+0.5\sigma \\ \frac{1}{\sigma}(x-x_{\max}-0.5\sigma)+0.25 & \text{for } x>x_{\max}+0.5\sigma \end{cases}$$

Where x_{min} and x_{max} are the lower and upper bounds and σ is the half-width of the well. For the disulfide bond formation: x is the distance between the C_α atoms of the cystine residues forming the disulfide bond, x_{min}, x_{max} and σ are 4.0, 6.1 and 0.6 Å respectively (Supplementary Table 1). For the stretched-twist conformation: x is the distance between the

C_α atoms of the *n*-5 CDR H3 residue and the 46th residue, x_{min}, x_{max} and σ are 6.5, 9.1 and 0.7 Å respectively (Table 2 last column *StretchTw Std. Dev.*). In each case x_{min} and x_{max} is the respective minimum and maximum over the range of observed variable and σ is the standard deviation of the variable. The weight of the constraint term is 10 and 100 in the low- and high-resolution stages respectively. The term penalizes deviations from observed structural features during the course of the search, but is not included for final discrimination and ranking of homology models.

Similar to RosettaAntibody, the algorithm initializes a loop by assuming ideal CDR H3 bond lengths and bond angles and stretched torsion angles (φ=-150°, ψ=150°, ω=180°). The six C-terminal residues are given backbone torsion angles from a cAb CDR H3 C-terminus depending on the loop classification (twist, stretched-twist, or neutral). For CDR H3 loops shorter than 16 residues, the loop is built using a three-residue fragment library, and longer loops are built first using a nine-residue fragment library then using a three-residue fragment library. Subsequently, CDR H1 loop is perturbed by $5n_{H1}$ cycles of low-resolution steps, each step comprising of max(5, $n_{H1}/2$) small and shear move perturbations (28) and CCD loop closure (29), where n_{H1} is the length of the H1 loop.

The high-resolution CDR H3 loop refinement is similar to RosettaAntibody's high-resolution stage, and is followed by a similar H3-like refinement of the CDR H1 loop. Finally the backbone dihedral angles of all the CDR loops are simultaneously optimized, using gradient-based minimization in backbone torsion angles and side-chain packing throughout the paratope, but obviously without optimization of the relative orientation of the light and heavy chains. Unlike RosettaAntibody, the high-resolution stage of CDR H3 loop building does not involve minimization of side-chain positions, but side chains are minimized as a final stage before a model is output.

In the standard protocol, 5,000 models are independently built for each target.

Algorithm Availability

The cAb V_HH homology modeling protocol is freely available for academic and non-profit use in version 3.2 of Rosetta (http://www.rosettacommons.org). Command-line syntax used for runs in this manuscript is provided in the Supplementary Information, and cAb CDR loop-recognition scripts are also available at http://antibody.graylab.jhu.edu/antibody/resources.

Results

Before attempting to build homology models, we studied known sequences and structures to identify useful features, particularly in contrast to classical antibodies. We curated a set of 27 non-redundant cAb V_HH structures and compared them to the 54 classical V_H structures used for benchmarking RosettaAntibody. We summarize first observations based on sequence and then those based on structure.

Sequence Analysis of cAb V_HH

The cAb V_HH test set, comprising a modest 27 members, is nevertheless the largest test set used to date for sequence alignment to detect amino acid patterns unique to cAb V_HH domains. These 27 unique cAb V_HH sequences comprise 22, 23 and 23 unique CDR H1, H2 and H3 sequences respectively (Supplementary Figure 1). The low average pair-wise sequence identities for CDRs H1, H2 and H3 in our test set of 37%, 32% and 20%, respectively, suggests that the test set comprises diverse CDR sequences. Additionally the highest pair-wise sequence identity of non-redundant CDR loops for H1, H2 and H3 is 70%, 86% and 42%, respectively.

While many differences have been previously reported in analysis of individual cases (1, 7, 9-12), some new differences arise by visual inspection of the sequence alignment of the cAb V_HH and V_H test sets (Supplementary Figure 1, summarized in Table 3). cAb V_HH positions which differ include residues 23, 49, 84, 87, 105, 108, and 109 (all specific residue numbers in this paper follow the Chothia convention (32)). Their structural locations are shown in Fig. 1 along with a superposition with a classical antibody to show the location with respect to the classical light chain and the classical C_H1 domain (absent in cAbs). Several positions have been reported before, and some new sequence differences are now apparent. At position 29, previously reported to be aspartic acid, glycine, asparagine, or serine (10), we see wider variation and, in a quarter of cAb V_HHs, tyrosine. Position 45 was previously noted to occasionally include cysteine by cDNA analysis (10), but cysteine does not appear at this position in any solved cAb V_HH structure. Position 44 has been previously noted to change from glycine in classical antibodies to glutamic acid or glutamine in cAb V_HH (11, 12); we note that classical antibodies also can have lysine or arginine in position 44, but these bases are never observed in cAb V_HHs.

For several previously unobserved residues, it is possible to speculate about the reasons for the sequential differences. Some seem to be related to stability and solubility of the cAb V_HH domain architecture. Glutamine is found in position 108 of cAb V_HH for all but one case (1OP9 has glutamate), while the corresponding position is occupied predominantly by threonines in V_H. Like the previously reported cAb mutations from hydrophobic to polar residues, the higher polarity of glutamine likely contributes to the enhanced solubility of the cAb V_HH domain (1). Relatedly, classical antibodies frequently have a surface exposed lysine 23 in the first framework region, but the cAb V_HH test set did not have any lysines in this position. The Eris server (33) predicts an average ΔΔG of 1.8 kcal/mol upon mutation of alanine to lysine, suggesting that the mutation away from lysine stabilizes the cAb V_HH relative to a V_H (11). The lysine would be also expected to aid solubility, but apparently it is not necessary in cAbs perhaps due to other solubility-enhancing mutations.

Alanine occupies cAb V_HH position 49 [one llama V_HH (1U0Q) had a glycine] in the middle of an anti-parallel β-sheet, while classical antibodies frequently have glycine there, proximal to the apex of the CDR L3 loop. Depending on whether position 49 is alanine or glycine, the correlated mutation position 69 can be either [IVLT] or [IVLFM], respectively. Thus when position 69 is phenylalanine or methionine (observed only in classicals), position 49 must be glycine to accommodate the larger volume. Another probable pair of correlated mutations is positions 14 and 84 (C_α-C_α distance ~6Å) which exhibit, respectively, predominantly alanine in cAb V_HHs or proline in V_Hs, and proline in cAb V_HHs or serine in V_Hs. An anti-parallel sheet is formed by the β-strands immediately after position 14 and immediately before position 84. The tight turn enabled by proline either at position 14 or 84 (cAb V_HH exceptions: 1U0Q, 1YC7) could be necessary to maintain the required distance between two anti-parallel β-strands for proper hydrogen-bond formation.

Residue 109 in the last (C-terminal) β-sheet of the variable heavy region is occupied by valine in cAb V_HHs, while classical antibodies exhibit mostly valine and some leucines. If leucine occurred at position 109 in a cAb V_HH domain, it would clash with the cAb V_HH-conserved leucine at position 18. In contrast in classical V_H domains, leucines in position 109 are accommodated by a valine at position 18. Thus the 18-109 pair is always leucine-valine in cAb V_HHs while it can be valine-leucine, leucine-valine or valine-valine in classical V_Hs. A few classical V_H domains (1VFA, 2ADG, 2AEP, 2H1P) have an apparent sterically unfavorable leucine-leucine pair. To compensate for the larger volume requirement of the two leucines, the neighboring position 82 for such V_H domains is always methionine. Position 82 can be either isoleucine, leucine or methionine in classical V_Hs, but it is always a methionine in cAb V_HH domains. In summary, the three neighboring residues

18, 82 and 109 are conserved leucine, methionine and valine in cAb V_HH domains, while the classical V_H domains vary. The variation in classical V_H domains is possible because of a slightly larger separation between the strands containing positions 18 and 109 relative to cAb V_HH strands.

The reasons for the following mutations are less apparent. The 87th residue lying in a surface exposed loop is always threonine in cAbs but is equally occupied by both serine and threonine in classical antibodies. Residue 105 in a surface exposed loop C-terminal to the CDR H3 loop (that ends in residue 102) has only glutamines in cAb V_HHs, but frequently exhibits alanine, glutamine and threonine in classicals.

Loop Identification—Standardized numbering systems (32, 34) serve as alignments for clear identification of CDR loop locations. However, the classical numbering systems have not been tested on cAb V_HH sequences, and servers like Abnum (35) that routinely number antibody sequences fail on most cAb V_HH sequences due to differences in the loop stems. Based on visual inspection of the sequences and similarity to the canonical loop definitions, we revised the rules for cAbs. Our new rules are detailed in Table 4.

Structural analysis of CDR loops

Figure 2 compares the structural features of whole cAb V_HH domain and the respective CDRs to their classical V_H counterparts.

CDR H1 and H2—CDR H1 and H2 loop conformations in classical antibodies can typically be identified using the canonical Chothia rules (36) (http://www.bioinf.org.uk/abs/chothia.html), however these rules have not been tested on cAbs. Upon testing these rules for CDR H1, we first notice that some sequences preclude classification because of differences in loop length (Table 4): canonical CDR H1 classes 1, 2, and 3 have 10, 11 and 12 residues respectively, however cAb CDR H1s sometimes contain 7, 8, 9 or 13 residues. Even when loop length matches, the canonical conformations do not cover the span of H1 conformations: Figure 2b shows the 20 of the 27 cAb V_HH CDR H1 loops with length of 10. Most deviate significantly from the two 10-residue canonical structures (class 1, 2FBJ and class 1b, 7FAB/3HFM.

Like the H1 loop, the cAb H2 loop precludes classification based on length: some cAb V_HH H2 loops have 6 or 13 residues, while the classical CDR H2 loop classes 1, 2, 3, 4 comprise 7, 8, 8 and 10 residues respectively. In contrast to the H1 loop, when the H2 loop length does match, the canonical conformations for H2 loops are often useful. Twelve of the nineteen cAbs that have 8-residue CDR H2 loops fit canonical class 2 (represented by 1BBD); four of the five cAbs that have 7-residue CDR H2 loops fit canonical class 1 (1GIG). Figure 2c shows the structural diversity of cAb CDR H2 loops compared to those of classicals.

CDR H3—The CDR H3 is the most variable loop in terms of amino acid composition, length (37), and structure. While CDR H3 loops largely elude structural classification, for classical antibodies Shirai *et al.* detected conserved structural motifs in the C-terminal stem (38). We previously incorporated the rules for CDR H3 modeling in classical antibodies (15). For cAbs, we analyzed the backbone torsion angles for the Shirai motifs. Unfortunately, neither the Shirai kink nor extended conformation is present in cAb H3 loops. We did, however, note the previously observed disulfide bond between CDR loops H1 and H3 (10), and two structural motifs which occur in a subset of H3 loops.

Disulfide Bond between CDRs H1 & H3: Nine antibodies in the test set exhibit a disulfide bond between the CDR H1 and H3 loops (Figure 1, Table 1) (10), and as previously noted, all of these are camel antibodies (7). The mean distance between the C_α atoms of the cystine residues forming the disulfide bond is 5.6 ± 0.6 Å (Supplementary Table 1). The CDR H1 cystine residue involved in the disulfide bond formation is always residue 33, with a single exception (anti-VSG cAbAn33, 1YC7 (39), where it is residue 32). While no such conserved residue was observed for the corresponding cystine residue in the CDR H3 loop, for the seven CDR H3 loops longer than 17 residues, the cystine always occurred in or immediately N-terminal to the middle of the loop.

Stretched-Twist Structural Feature: A structural superposition of the cAb V_HH framework region reveals that the C-terminal region of 17 of the 27 cAb V_HH CDR H3 loops exhibits a conserved structural motif (red segments in Fig. 2a). To quantify the observation, we calculated the φ and ψ backbone dihedral angles of the conserved subset (Table 2). For residues *n*+1 through *n*-4, the backbone φ and ψ angles are very similar for all the members exhibiting the conserved structural motif (CDR H3 residues are numbered 1 to *n*, where H3 residue 1 corresponds to residue 95 and residue *n* to 102). We define the motif by residues with backbone dihedral angles with standard deviation under 30° across the subset. All backbone dihedral angles of these residues meet this criterion except φ of *n*-3 and ψ of *n*-4. The standard deviation of the dihedral angles of residues *n*+1 through *n*-4 of the subset is one-third that of the entire dataset. Since the conserved feature has a sharp twist near the C-terminus and then stretches to reach the heavy framework, we refer to the motif as a *stretched-twist* (Fig. 2a).

Stretched-twisted CDR H3s bend and contact the region of the framework that would have been in contact with the light framework in a classical antibody. The mean (and standard deviation) C_α-C_α distances between the CDR H3 apex residues, *n*-4 and *n*-5, and the nearest cAb V_HH framework residue (residue 46) are 11.3 ± 0.8 Å and 7.9 ± 0.7 Å respectively (black line in Fig. 1; Table 2). The corresponding standard deviations over the entire test set were much higher, 3.5 and 5.5 Å respectively, illustrating that the distances are not conserved over the entire test set. Since the mean (and standard deviations) were obtained from our dataset which includes both antigen-bound and unbound cAbs, the observed C_α-C_α distances do not rely on antigen binding. The similarly conserved C_α-C_α distances in the unbound llama cAb V_HH A52 structure (1I3V) (40) and the corresponding bound structure with dye RR1 (1I3U) (40) further reinforces that the observed C_α-C_α distances are conserved for *stretched-twist*ed CDR H3s and are independent of antigen binding.

Twist Structural Feature: Six additional structures exhibit only the sharp turn of the stretched-twist motif in residues *n*+1, *n*, *n*-1 and *n*-2 (Table 2; green segments in Fig. 2a). All four residues have conserved backbone dihedrals with a standard deviation less than 30° across the set except for the *n*-1 φ angle. The standard deviations are higher in this structural feature than that in the stretched-twist indicating more diversity. We refer to this feature alone as a *twist*. Structures that cannot be classified as either *stretched-twist* or *twist* are referred to as *neutral*.

Sequence-Structure Rules: Upon examining the sequences, structures, and environments of the subsets of H3 loops exhibiting conserved motifs, we found several useful signatures. Examining the loop sequence directly, when *n*-2 is tyrosine, tryptophan or phenylalanine, and *n*-1 is not a tyrosine, a twist is formed. When *n*-2 is not tyrosine, tryptophan or arginine, a twist is observed only when residue *n* is tyrosine. Additionally if the CDR H3 is longer than twelve residues, the CDR H3 adopts a unique stretched-twist conformation and stretches so that the apex of the CDR H3 loop approaches framework residue 46. In the environment, C-terminal residue *n*-2 is near the CDR H3 N-terminal residues 93-94 (Fig.

2d); when these positions have a large aromatic group (tyrosine, tryptophan or phenylalanine), a sharp twist is necessary to prevent steric clashes between the N and C-terminal CDR H3 stems. In 18 cAb V_HH structures (including 5 of the 6 with twists and 13 of the 17 with stretched-twists) at the position where the conserved CDR H3 loop's *n*-5 residue approaches the heavy chain framework, residues 43-44-45-46 form a highly charged and polar patch (lysine-glutamate/aspartate-arginine-glutamate), in contrast to the classical antibodies in which residues 44 and 45 are generally non-polar amino acids (glycine-leucine). In 8 of the 13 cAb V_HHs exhibiting a *stretched-twist* CDR H3 conformation and the polar patch (residues 43-46), a hydrogen bond is formed between arginine 45 and either the *n*-5 or *n*-6 CDR H3 residue. Additionally in 23 of the 27 cAb V_HHs in our dataset, lysine 43 forms a hydrogen bond with glutamine 39, and in 26 cAb V_HH structures glutamate 46 forms a hydrogen bond with arginine 38. The high prevalence of glutamate in the 44th position of cAbs as opposed to glycine in classical antibodies at the same position is due to the region becoming solvent exposed on the loss of the hydrophobic V_L-V_H interface.

While there are multiple ways to create rules from these observations, we use the following for later structure prediction:

1) A *twist* is formed when

Positions 93 and 94 do not have lysine, glutamine or asparagine.

Additionally,

2) A *stretched-twist* is formed only if in addition to rule (1) the following rules are satisfied:

a. CDR H3 loop length is twelve residues or more

b. *n*-2 position is either tyrosine, tryptophan or phenylalanine

c. *n*-1 position is not histidine

d. *n* position is not glycine

The cAb subsets having *twist* or *stretched-twist* CDR H3 conformations consists of both antigen-bound and unbound structures (Table 1), suggesting that the observed conformational signatures may be preserved upon antigen binding.

Exceptions: The camel antibody cAb-Lys2 (1RJC) (41) against hen egg white lysozyme adopts a slight variation of the stretched-twist structure probably due to the presence of tyrosine at position *n*-3 which forces the C-terminal stem to move away from the respective position of other stretched-twisted structures.

Rule 2c was included to incorporate the camel antibody CAB-CA05 (1F2X) (42). 1F2X is the only antibody in the cAb V_HH test set that, in spite of having a tyrosine at position *n*-2, forms only a twist, but not a stretched-twist. 1F2X is also one of only two antibodies in the cAb V_HH test set that has more than two consecutive aromatic amino acids in the CDR H3 C-terminal positions *n*-2 through *n*+1. By virtue of having a histidine at position *n*-1 (in between two tyrosines), the sterics force the CDR H3 loop to fold atypically towards the framework. The other structure to have more than two consecutive aromatic residues in positions *n*-2 through *n*+1 is the camel antibody CABAMD9 (1KXQ) (43), with all four positions occupied by aromatic amino acids. However, while the three aromatic residues in 1F2X resulted in an atypical bulge, the presence of the fourth consecutive aromatic residue reverses the chain orientation, so 1KXQ adopts an stretched-twist conformation.

Rule 2d was included to incorporate the llama antibody (2BSE) (44), the only antibody in the cAb V_HH test set that has a glycine at position *n*. The flexibility afforded by glycine in the CDR H3 loop's C-terminus enables the loop to extend away from the protein body, contrary to other stretched-twist structures which bend towards framework residue 46. Interestingly, the three other twisted antibodies that satisfy the first clause of Rule (2) and have a tyrosine or tryptophan at *n*-2 position (1F2X, 1G9E, 1HCV) have a glycine in either of N-terminal CDR H3 stem residues 93 and 94. However, in addition to these three antibodies, another antibody in the cAb V_HH test set (1ZVY) has a glycine at position 93 and adopt the stretched-twist conformation.

Possible existence of CDR 4—Because the region between residues 71-78 is close to the other CDRs, it has been suggested that it combines with CDRs H1, H2 and H3 to form a larger paratope (45, 46). Additionally, affinity maturation studies involving mutations in this region affected antigen binding (47, 48) suggesting that that CDR 4 could be recruited in antigen-binding. A structural alignment (Fig. 2e) shows that the loop formed by heavy chain residues 71-78 has a significantly larger structural divergence than in the V_H loops. To determine the importance of CDR 4 in antigen binding, we examined the CDR 4-antigen contacts in the antibody-antigen complexes in the test set. Of the 19 antibody-antigen complexes, only two (1KXQ, 1SJX) show CDR 4 contacts to antigen. Apparently, CDR 4 is capable of but not critical for antigen interactions.

Homology Models

With the new rules and observations about cAb V_HH domains, we created a tailored algorithm to model cAb V_HH domains starting from their sequences. The method, based on RosettaAntibody, uses structural fragments from a database and a Monte Carlo-plus-minimization structure prediction and refinement algorithm, and it incorporates the newly observed structural features either as constraints or by selecting appropriate database fragments in model construction. cAb V_HH coordinates were added to the RosettaAntibody database, and homology models are assembled with templates and loops (when available) from this database. Matching cAb CDR loops are grafted onto a selected framework. For testing each target in a ‘blind’ manner, the native structure and templates with exact sequence match over the entire length of the query sequences (framework, CDRs H1 and H2) were removed from the database. The CDR H3 loop is built using RosettaAntibody's low- and high-resolution Monte Carlo-plus-minimization-based loop building techniques (28). Since the CDR H1 also exhibits larger conformational diversity than that reported in classical antibodies, the CDR H1 loop is also subjected to Monte Carlo perturbation and refinement.

Grafting CDRs—CDRs H1 and H2 are grafted to create cAb V_HH homology models in a manner similar to that used for canonical CDR loops using RosettaAntibody. Due to the non-canonical nature of most cAb V_HH CDR H1 and H2 loops, traditional canonical class-based template selection fails to identify template matches for such cases. Instead, RosettaAntibody selects template loops based on BLAST bit scores, and thus it is not limited by class definitions and identifies templates as long as similar sequences can be found in the database. The median global rmsd for cAb V_HH CDRs H1 and H2 (3.6 and 1.5 Å respectively; Table 1) is higher than that obtained by using RosettaAntibody to graft canonical CDRs into V_H (0.84 and 0.93 Å respectively) (15). The cAb V_HH CDR H1 exhibits larger conformational diversity than classical antibodies, and the available CDR H1 templates deviate from the native structures. To improve the CDR H1 prediction accuracy, we later subject it to explicit perturbations.

While the selection strategy succeeds for most targets, in a few cases the BLAST search for similar loop sequences in the database does not return any matches. Several cases (1YC7, 1ZV5) can be addressed by grafting a length-matched sequence; these often have poor accuracy (median rmsd ~3 Å). The cAb V_HH-R2 anti-RR6 antibody (1QD0) (49) still fails because it is the only antibody in the database to have a 13 residue CDR H1. Similarly, the cAbs with the shortest (6 residue) and the longest (13 residue) CDR H2 loops (camel antibody CAB-CA05, 1F2X (42), and llama antibody A52, 1I3V (40)) are unique and do not have length-matched loop templates. For testing purposes, in these individual cases we grafted the native loop structure. In truly blind tests, de novo methods and grafting methods which do not require length-matching may be useful (50).

Modeling CDR H3—As in classical antibodies, the cAb CDR H3 loop is hypervariable and must be modeled *de novo*. In this section we isolate the loop modeling algorithm performance by building the CDR H3 in the native environment; that is, we start with the crystal structure, remove the CDR H3 residues and rebuild the loop. We illustrate the CDR H3 modeling strategy with a representative cAb V_HH, cabbcII-10 (1ZMY) (51). CabbcII-10 has a 24-residue CDR H3 loop with a disulfide bond between the CDR H1 and H3 loops and exhibits the stretched-twist conformation.

We tested four variants of algorithms for modeling the CDR H3 loop. Initially we used a nine-residue fragment-insertion Monte Carlo strategy and did not use any constraints to model observed structural features. For cabbcII-10, this strategy sampled CDR H3 conformations as low as 4.8 Å global rmsd, with the lowest-energy (LowE) model having a global rmsd of 7.8 Å. Next, to test the usefulness of the observed structural features, we performed the loop building simulations while also incorporating constraints that bias formation of the disulfide bond and the approach of the apex of the stretched-twist loop to the framework (Materials and Methods). The inclusion of constraints improved the CDR H3 global rmsd of the LowE model to 4.6 Å and enabled sampling of conformations as close as 3.6 Å. To further improve the fine sampling, we incorporated additional Monte Carlo steps using three-residue fragment insertions following the nine-residue fragments. The finer sampling produced a more accurate LowE CDR H3 with global loop rmsd of 3.6 Å.

To test whether further improvement is limited by sampling or scoring, we compared the model energies to energies of native structures subjected to the same high-resolution refinement stages. Refined native loops scored better than all model loops (Supplementary Figure 2), implying that better models, if sampled, could be identified by the energy function. Therefore, we tested whether increasing the number of models from 5,000 to 20,000 could produce more accurate models. The increased sampling slightly improved the rmsd of structures sampled (lowest rmsd in the entire set of models improved from 3.0 Å to 2.5 Å), but the difference was not significant enough to improve the LowE performance, which had a median of 4.0 Å. Due to the computational expense of modeling long loops, we limit the number of models to 5,000 loops in the final protocol. The time required to build one model is around 10 minutes on one CPU, which, for one simulation involving 5,000 models, translates to ~35 CPU days or about three wall-clock hours on a 300-CPU cluster.

Results on the full test set are given in Supplementary Table 2 for the four variants of the algorithm, and Table 1 shows the results of CDR H3 modeling using the final protocol. Using the final protocol, the low-energy model loops range from 0.9 to 6.9 Å rmsd, with a median of 3.9 Å. If the best loop in the ten-lowest energy models is allowed, the mean rmsd falls to 3.4 Å. The most-native-like loop sampled during 5000 independent runs ranges from 0.9 to 3.9 Å rmsd, with a median of 2.6 Å. The test set contains 13 loops with CDR H3 loop lengths between 16-24 residues, and incorporating the additional sampling using three-residue fragments improved the median global rmsds for the LowE models from 4.5 to 3.9 Å (Supplementary Table 2). In 23 of 27 cases, refined native structures score better (lower) than the LowE model, suggesting that loop sampling still limits the algorithm's performance. Thus, the final CDR H3 loop building protocol uses the constraints to model the observed structural features, uses fragment-based loop building with nine-residue fragments and three-residue fragments for loops over sixteen residues, and generates 5,000 candidate structures.

Complete Homology Models—Complete homology models were generated by building the CDR H3 loop in an environment where the framework and CDRs H1 and H2 have been assembled from sequence-homologous templates. The CDR H1 is also subjected to perturbations, since grafting alone often did not produce near-native structures. Table 1 shows that the median CDR H3 global rmsd for the test set is 5.4 Å for the LowE models. Figure 3a shows the diversity of CDR conformations in the ten lowest-scoring models illustrating that the *ab initio* loop modeling generates a wide variety of CDR H3 conformations while the other CDRs show minimal variation due to the absence of fragment insertions to cause significant backbone conformational change. Although these rmsd values are high, Figure 3b shows that the homology model captures the rough topology of the H3 loop, including the enforced stretched-twisted C-terminal region and the distance between the *n*-5 residue of the CDR H3 loop and residue 46 (8.7 Å, within the observed range of 6.5 to 9.1 Å). The worst-case deviations in CDR H3 loop predictions can be attributed to a poorly modeled environment: in the two structures for which the LowE CDR H3 loop rmsd is more than 10 Å, at least one of the other grafted CDRs in the respective structures deviate by more than 5.0 Å. Additionally the CDR H3 models closest to the native conformations amongst all models built have a median global rmsd of 2.7 Å revealing that the most native-like models still deviate from the native structure. Lower refined-native scores suggest that better sampling is required for more accurate predictions.

The median global rmsds for the LowE models for 8-12, 14-16 and 17-24 residue CDR H3 loops were 5.7, 4.5 and 6.8 Å respectively. The lower rmsd for the 14-16 residue loops compared to the shorter 8-12 residue loops is surprising because the available conformational space for longer loops is much larger than shorter loops, and loop prediction is expected to be more challenging. This anomaly can be explained by the more accurate environment for the 14-16 residue loops as is evident from the median global rmsds of 2.3 and 1.0 Å for the CDRs H1 and H2 respectively, as compared to higher deviations of 3.3 and 1.5 Å for the respective CDRs in 8-12 CDR H3s. Furthermore, seven of the eight CDR loops of length 14-16 residues can be classified as having the stretched-twisted motif and thus the H3 apex constraint can be applied.

To accurately model CDR H3s we not only minimized the conformations of the neighboring CDR H1 and H2 loops, but also explicitly perturbed the CDR H1 loop. The median global CDR H1 rmsd improves from 3.6 to 3.1 Å by incorporating perturbations in addition to grafting.

Discussion

We endeavored to create and test the first homology modeling protocol tailored for cAb V_HH antibodies. The protocol is an extension of RosettaAntibody with several unique features. Explicit perturbations to the CDR H1 loop during CDR H3 modeling enabling sampling of the diverse non-canonical CDR H1 structures found in cAb V_HH. The incorporation of the additional three-residue fragment-based low-resolution loop building following nine-residue fragment-based loop assembly allowed sampling of more native-like low-energy models. Finally, incorporation of constraints to capture the H1-H3 disulfide bond and the approach of the H3 loop apex to the framework improved native-like sampling by preventing predicted loops from adopting conformations with significant deviation from the observed features. The results demonstrate that structural models can be built with gross structural features of the native H3 loop but significant local conformation deviations.

The challenge in modeling very long CDR H3 loops is evident from the high global rmsds, and advancements in computationally efficient loop modeling techniques will be critical to create more accurate cAb V_HH homology models. The diversity in the low-energy loop structures emphasizes the large conformational space available to the CDR H3. The challenge of adequately sampling H3 conformations may be amplified due to the manner in which the H3 and H1 loops compensate for the absence of the light chain and create antigen recognition properties comparable to classical (dimeric) antibodies (7). Forcing the formation of disulfide bonds in applicable cAb V_HH domains reduces the conformational space making more native-like predictions possible and probably also minimizes the entropic penalty of long loops on binding with antigen (12).

Perturbations to the CDR H1 produced more native-like CDR H1 loops as opposed to simple grafting, but differences still remain between the predicted and the native conformation. Building two neighboring loops using *ab initio* methods in a non-native environment is challenging (52, 53), because multiple loop sampling can produce false positive structures and require significantly more computational time. The forcing of the disulfide bond between the CDR H3 and H1 loops in respective CDRs also reduced the conformational space of the CDR H1 loop. As the number of cAbs in the PDB increases, we anticipate better starting conformations for grafting homologous CDR H1 loops.

Antibody surfaces have many charged residues (54) and particularly the cAb V_HH CDR H1 and H3 loops have many charged residues, resulting in dominant inter- and intra-loop electrostatic interactions (55), which might not be captured correctly given the difficulty in modeling such interactions (56). However, the most significant hurdle in higher accuracy loop prediction is inadequate sampling. Other algorithms scale the amount of sampling with the number of residues in the loop, e.g., Protein Local Optimization Program (PLOP) (16) samples 2^n loop conformations up to 10^6, where n is the loop length. The high computational cost of such simulations renders them difficult and more efficient sampling strategies are required.

In contrast to previous studies involving multiple cAb V_HH sequences where it is uncertain whether a sequence is from a V_H or V_HH (10), the test set created here, by virtue of being derived from crystallized cAb V_HH domains, guarantees that only V_HH domains are present. cAb V_HH sequence comparison revealed key residues unique to cAb V_HH regions which probably contribute to their desirable characteristics (7), although additional structures will help confirm or generalize the observations. Resurfacing human antibodies to incorporate the identified key residues may result in more camel-like human cAbs. The observations will for the first time enable automated detection of cAb CDR loops, and subsequently allow Chothia numbering of cAb V_HH antibodies. The ability to number antibody sequences using standardized Kabat (34) or Chothia (32) numbering schemes will ease antibody processing by providing structural insights merely by viewing a properly numbered sequence.

RosettaAntibody approaches can be generalized to model similar molecules of biotechnological importance that can provide alternate scaffolds for therapeutic antibodies like engineered C_H2 domains, nanoantibodies, or MHCs (57). In the future we anticipate incorporating cAb V_HH modeling as a feature in the RosettaAntibody server (27). Although homology models can be used with a flexible backbone antibody-antigen docking protocol like SnugDock (20, 58) to predict the conformation of an antibody-antigen interaction complex, at the moment, loop errors in the cAb predictions likely limit docking accuracy. As modeling abilities improve, predicted structures may provide structural insights that explain consequences of antibody humanization (59), improve antibody specificity by reducing cross-reactivity (60), and improve binding affinity (61-63).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this project was generously provided by the National Institutes of Health (R01-GM078221) and UCB S.A. KS was a Johns Hopkins Institute for NanoBioTechnology summer student funded by the National Science Foundation (EEC-0647755).

References

1. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, Bendahman N, Hamers R. Naturally occurring antibodies devoid of light chains. Nature. 1993; 363:446–448. [PubMed: 8502296]
2. Muyldermans S, Baral TN, Retamozzo VC, De Baetselier P, De Genst E, Kinne J, Leonhardt H, Magez S, Nguyen VK, Revets H, Rothbauer U, Stijlemans B, Tillib S, Wernery U, Wyns L, Hassanzadeh-Ghassabeh G, Saerens D. Camelid immunoglobulins and nanobody technology. Vet Immunol Immunopathol. 2009; 128:178–183. [PubMed: 19026455]
3. Roovers RC, van Dongen GA, van Bergen en Henegouwen PM. Nanobodies in therapeutic applications. Curr Opin Mol Ther. 2007; 9:327–335. [PubMed: 17694445]
4. Wolfson W. Ablynx makes nanobodies from llama bodies. Chem Biol. 2006; 13:1243–1244. [PubMed: 17185218]
5. Schwede T, Sali A, Honig B, Levitt M, Berman HM, Jones D, Brenner SE, Burley SK, Das R, Dokholyan NV, Dunbrack RL Jr, Fidelis K, Fiser A, Godzik A, Huang YJ, Humblet C, Jacobson MP, Joachimiak A, Krystek SR Jr, Kortemme T, Kryshtafovych A, Montelione GT, Moult J, Murray D, Sanchez R, Sosnick TR, Standley DM, Stouch T, Vajda S, Vasquez M, Westbrook JD, Wilson IA. Outcome of a workshop on applications of protein models in biomedical research. Structure. 2009; 17:151–159. [PubMed: 19217386]
6. Ablynx Corporate Overview. http://www.ablynx.com/aboutus/index.htm
7. De Genst E, Saerens D, Muyldermans S, Conrath K. Antibody repertoire development in camelids. Dev Comp Immunol. 2006; 30:187–198. [PubMed: 16051357]
8. Verheesen P, Roussis A, de Haard HJ, Groot AJ, Stam JC, den Dunnen JT, Frants RR, Verkleij AJ, Theo Verrips C, van der Maarel SM. Reliable and controllable antibody fragment selections from Camelid non-immune libraries for target validation. Biochim Biophys Acta. 2006; 1764:1307–1319. [PubMed: 16872921]
9. Harmsen MM, De Haard HJ. Properties, production, and applications of camelid single-domain antibody fragments. Appl Microbiol Biotechnol. 2007; 77:13–22. [PubMed: 17704915]
10. Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng. 1994; 7:1129–1135. [PubMed: 7831284]
11. Davies J, Riechmann L. Single antibody domains as small recognition units: design and in vitro antigen selection of camelized, human VH domains with improved protein stability. Protein Eng. 1996; 9:531–537. [PubMed: 8862554]
12. Muyldermans S, Cambillau C, Wyns L. Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. Trends Biochem Sci. 2001; 26:230–235. [PubMed: 11295555]
13. Muyldermans S, Lauwereys M. Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. J Mol Recognit. 1999; 12:131–140. [PubMed: 10398404]
14. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
15. Sivasubramanian A, Sircar A, Chaudhury S, Gray JJ. Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. Proteins. 2009; 74:497–514. [PubMed: 19062174]

16. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. Proteins. 2004; 55:351–367. [PubMed: 15048827]
17. Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods. 2009; 6:551–552. [PubMed: 19644455]
18. Marcatili P, Rosi A, Tramontano A. PIGS: automatic prediction of antibody structures. Bioinformatics. 2008; 24:1953–1954. [PubMed: 18641403]
19. Chaudhury S, Gray JJ. Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. J Mol Biol. 2008; 381:1068–1087. [PubMed: 18640688]
20. Sircar A, Gray JJ. SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. PLoS Comput Biol. 2010; 6:e1000644. [PubMed: 20098500]
21. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A. 1988; 85:2444–2448. [PubMed: 3162770]
22. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]
23. Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci. 1996; 12:543–548. [PubMed: 9021275]
24. DeLano, WL. The PyMOL Molecular Graphics System. 2002. Available at http://www.pymol.org
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]
26. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A. 2000; 97:10383–10388. [PubMed: 10984534]
27. Sircar A, Kim ET, Gray JJ. RosettaAntibody: antibody variable region homology modeling server. Nucleic Acids Res. 2009; 37:W474–479. [PubMed: 19458157]
28. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. Methods Enzymol. 2004; 383:66–93. [PubMed: 15063647]
29. Canutescu AA, Dunbrack RL Jr. Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein Sci. 2003; 12:963–972. [PubMed: 12717019]
30. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein-protein docking. Protein Sci. 2005; 14:1328–1339. [PubMed: 15802647]
31. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol. 1997; 268:209–225. [PubMed: 9149153]
32. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol. 1997; 273:927–948. [PubMed: 9367782]
33. Yin S, Ding F, Dokholyan NV. Eris: an automated estimator of protein stability. Nat Methods. 2007; 4:466–467. [PubMed: 17538626]
34. Kabat, EA.; Wu, TT.; Bilofsky, H.; Reid-Miller, M.; Perry, H. Sequence of Proteins of Immunological interest. National Institutes of Health; Bethesda: 1983.
35. Abhinandan KR, Martin AC. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. Mol Immunol. 2008; 45:3832–3839. [PubMed: 18614234]
36. Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. Structural repertoire of the human VH segments. J Mol Biol. 1992; 227:799–817. [PubMed: 1404389]
37. Wu TT, Johnson G, Kabat EA. Length distribution of CDRH3 in antibodies. Proteins. 1993; 16:1–7. [PubMed: 8497480]
38. Shirai H, Kidera A, Nakamura H. Structural classification of CDR-H3 in antibodies. FEBS Lett. 1996; 399:1–8. [PubMed: 8980108]
39. Conrath K, Vincke C, Stijlemans B, Schymkowitz J, Decanniere K, Wyns L, Muyldermans S, Loris R. Antigen binding and solubility effects upon the veneering of a camel VHH in framework-2 to mimic a VH. J Mol Biol. 2005; 350:112–125. [PubMed: 15913651]

40. Spinelli S, Tegoni M, Frenken L, van Vliet C, Cambillau C. Lateral recognition of a dye hapten by a llama VHH domain. J Mol Biol. 2001; 311:123–129. [PubMed: 11469862]

41. De Genst E, Silence K, Ghahroudi MA, Decanniere K, Loris R, Kinne J, Wyns L, Muyldermans S. Strong in vivo maturation compensates for structurally restricted H3 loops in antibody repertoires. J Biol Chem. 2005; 280:14114–14121. [PubMed: 15659390]

42. Decanniere K, Muyldermans S, Wyns L. Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes? J Mol Biol. 2000; 300:83–91. [PubMed: 10864500]

43. Desmyter A, Spinelli S, Payan F, Lauwereys M, Wyns L, Muyldermans S, Cambillau C. Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. J Biol Chem. 2002; 277:23645–23650. [PubMed: 11960990]

44. Spinelli S, Desmyter A, Verrips CT, de Haard HJ, Moineau S, Cambillau C. Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol. 2006; 13:85–89. [PubMed: 16327804]

45. Bond CJ, Wiesmann C, Marsters JC Jr, Sidhu SS. A structure-based database of antibody variable domain diversity. J Mol Biol. 2005; 348:699–709. [PubMed: 15826665]

46. Capra JD, Kehoe JM. Variable region sequences of five human immunoglobulin heavy chains of the VH3 subgroup: definitive identification of four heavy chain hypervariable regions. Proc Natl Acad Sci U S A. 1974; 71:845–848. [PubMed: 4522793]

47. Baca M, Presta LG, O'Connor SJ, Wells JA. Antibody humanization using monovalent phage display. J Biol Chem. 1997; 272:10678–10684. [PubMed: 9099717]

48. Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, Rowland AM, Kotts C, Carver ME, Shepard HM. Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc Natl Acad Sci U S A. 1992; 89:4285–4289. [PubMed: 1350088]

49. Spinelli S, Frenken LG, Hermans P, Verrips T, Brown K, Tegoni M, Cambillau C. Camelid heavy-chain variable domains provide efficient combining sites to haptens. Biochemistry. 2000; 39:1217–1222. [PubMed: 10684599]

50. Choi Y, Deane CM. FREAD revisited: Accurate loop structure prediction using a database search algorithm. Proteins. 78:1431–1440. [PubMed: 20034110]

51. Saerens D, Pellis M, Loris R, Pardon E, Dumoulin M, Matagne A, Wyns L, Muyldermans S, Conrath K. Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol. 2005; 352:597–607. [PubMed: 16095608]

52. Sellers BD, Zhu K, Zhao S, Friesner RA, Jacobson MP. Toward better refinement of comparative models: predicting loops in inexact environments. Proteins. 2008; 72:959–971. [PubMed: 18300241]

53. Danielson ML, Lill MA. New computational method for prediction of interacting protein loop regions. Proteins. 78:1748–1759. [PubMed: 20186974]

54. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol. 1999; 285:2177–2198. [PubMed: 9925793]

55. Fenwick MK, Escobedo FA. Hybrid Monte Carlo with multidimensional replica exchanges: conformational equilibria of the hypervariable regions of a llama VHH antibody domain. Biopolymers. 2003; 68:160–177. [PubMed: 12548621]

56. Fitch CA, Whitten ST, Hilser VJ, Garcia-Moreno EB. Molecular mechanisms of pH-driven conformational transitions of proteins: insights from continuum electrostatics calculations of acid unfolding. Proteins. 2006; 63:113–126. [PubMed: 16400648]

57. Dimitrov DS. Engineered CH2 domains (nanoantibodies). MAbs. 2009; 1:26–28. [PubMed: 20046570]

58. Sircar A, Chaudhury S, Kilambi KP, Berrondo M, Gray JJ. A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19. Proteins. 2010; 78:3115–3123. [PubMed: 20535822]

59. Hwang WY, Almagro JC, Buss TN, Tan P, Foote J. Use of human germline genes in a CDR homology-based approach to antibody humanization. Methods. 2005; 36:35–42. [PubMed: 15848073]

60. Humphris EL, Kortemme T. Design of multi-specificity in protein interfaces. PLoS Comput Biol. 2007; 3:e164. [PubMed: 17722975]

61. Karanicolas J, Kuhlman B. Computational design of affinity and specificity at protein-protein interfaces. Curr Opin Struct Biol. 2009; 19:458–463. [PubMed: 19646858]

62. Lippow SM, Wittrup KD, Tidor B. Computational design of antibody-affinity improvement beyond in vivo maturation. Nat Biotechnol. 2007; 25:1171–1176. [PubMed: 17891135]

63. Pantazes RJ, Maranas CD. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. Protein Eng Des Sel. 23:849–858. [PubMed: 20847101]

64. Decanniere K, Transue TR, Desmyter A, Maes D, Muyldermans S, Wyns L. Degenerate interfaces in antigen-antibody complexes. J Mol Biol. 2001; 313:473–478. [PubMed: 11676532]

65. Stanfield R, Cabezas E, Satterthwait A, Stura E, Profy A, Wilson I. Dual conformations for the HIV-1 gp120 V3 loop in complexes with different neutralizing fabs. Structure. 1999; 7:131–142. [PubMed: 10368281]

66. Decanniere K, Desmyter A, Lauwereys M, Ghahroudi MA, Muyldermans S, Wyns L. A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. Structure. 1999; 7:361–370. [PubMed: 10196124]

Abbreviations used in this article

HCAb	heavy chain only antibody
cAb	camelid heavy chain only antibody
CDR	Complementarity Determining Region
V_HH	cAb variable region
rmsd	root mean square deviation
C_H	heavy chain constant domain
V_L	classical antibody light chain variable domain
V_H	classical antibody heavy chain variable domain

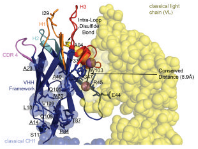

Figure 1.

Structural positions of key residues which differ between classical antibodies and cAb V_HHs. The cartoon representation of the camel anti lysozyme-V_HH antibody (1JTO) (64) shows the stick representation of key residues mentioned in Table 3. The newly observed residues are additionally enveloped in transparent structures and indicated by underlined residue labels. The 1JTO heavy chain framework is superimposed on the heavy chain framework of the classical antibody IgG1 58.2 (1F58) (65) which has a 17 residue CDR H3 loop, the same length as in 1JTO. The position of the light chain (yellow transparent spheres) and the C_H1 domain (blue transparent spheres) of the classical antibody 1F58 shows the position of the key residues in the context of a classical antibody. The figure shows the conserved distance (black dashed line) in stretched-twisted loops between the C_α of residue 102 (solid magenta sphere) and the C_α of residue 46 (solid magenta sphere). The yellow spheres representing the V_L region near the conserved distance are lighter for enhanced clarity. CDR H1 (orange), H2 (cyan), H3 (red), 4 (magenta); Cystines forming disulfide bond between CDRs H1 and H3 (yellow sticks).

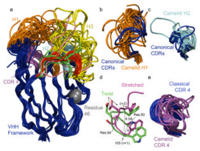

Figure 2.

Structural diversity in cAb V_HH domains. The 27 cAb V_HH structures of the test set were superimposed on their heavy chain framework. a) Structural diversity in cAb V_HH domains with a focus on structural features of CDR H3 C-terminus. The six residues *n*+1, *n*, *n*-1, *n*-2, *n*-3 and *n*-4 form common structural features; stretched-twist (red), twist (green) and neutral (salmon). The apex of the stretched-twist approaches residue 46 (C_α atom: gray spheres). CDR H1 (orange), H2 (cyan), H3 (yellow), 4 (magenta). b) Comparison of structural diversity in cAb V_HH CDR H1 conformations (orange) to representative members of all V_H CDR H1 canonical classes (blue; 2FBD, 7FAB, 3HFM). c) Comparison of structural diversity in cAb V_HH CDR H2 conformations (cyan) to representative members of all V_H CDR H2 canonical classes (blue). Most deviate significantly from the two 10-residue canonical structures (class 1, 2FBJ and class 1b, 7FAB/3HFM). The unique protruding cAb H2 is for the longest, 13-residue, loop in llama V_HH (1I3V) (40). d) High resolution structural details of the turn in both *twist*-ed and *stretched-twist*-ed CDR H3 show that residue *n*+1 is near not only to the sequentially adjacent residue *n*+2, but also to the sequentially distal CDR H3 N-terminal residues 93-94 e) Comparison of structural diversity of the hypothetical CDR 4 region in cAbs (magenta) to the respective region in V_H domains (blue).

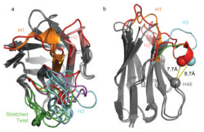

Figure 3.

Representative homology model of cAb V_HH CAB-RN05 (1BZQ) (66). a) Antigen-eye view of the paratope showing the CDR diversity in the ten lowest energy homology models. LowRMS model CDR H3 loop (magenta); crystal structure CDR loops (red). b) Side view of the LowRMS homology model superimposed on the native framework. Residues *n*-5 and 46 C_α atoms are shown in spheres, and the conserved distance between them is indicated by yellow dashed lines. CDR H1 (orange), H2 (salmon), H3 (cyan); Stretched-twist (green); Native crystal CDRs (red); Framework (grey).

Table 1

Global rmsds for CDR loops for homology models. *PDB ID* in regular and *italics* indicates camel and llama antibodies respectively. *Seq. ID* shows the percent identity between the sequences of the query and the template loop used for grafting. A *Seq ID* of 0 means that a sequence-homology based template was not identified and a template CDR of the same length was grafted and 100 means that the native CDR was grafted. *Graft rmsd* measures the loop accuracy after grafting loop coordinates to the homology framework, and *Homology LowE rmsd* measures loop accuracy of the LowE model after CDR H1 refinement. For the CDR H3, measures are given for building loops *In Native Context* and on the *Homolog* framework with predicted CDRs. *LowE* indicates the lowest-energy model. *LowRMS* indicates the model with the lowest global rmsd in the ten lowest-energy models. *LowALL* indicates the lowest global rmsd observed amongst all models. [s]Stretched-twist; [t]Twist; [c]Cystine bond between CDR H1 and H3; [p]Bound to protein antigen; [l]Bound to small molecule; [u]Unbound structure. [1] and [2] CDR H2 canonical class 1 and 2 respectively. [†]The glycine at residue 26 was disordered in the crystal structure of 1RJC, thus the H1 rmsd is computed against the ordered residues of the loop and excluded from the median summary.

PDB ID	CDR Length			CDR H1			CDR H2		CDR H3 rmsd (Å)					
	H1	H2	H3	Seq. ID	Graft rmsd	Homology LowE rmsd	Seq. ID	Graft rmsd	In Native Context			Homolog		
				(%)	(Å)	(Å)	(%)	(Å)	LowE	LowRMS	LowALL	LowE	LowRMS	LowALL
1IEH[u]	10	8	8	89	3.8	3.9	88	4.1	3.1	3.1	1.6	6.0	5.0	3.4
1G9E[tu]	10	8	8	70	4.0	3.3	88	2.1	6.6	1.4	1.3	5.4	4.8	1.8
1HCV[tu]	10	8[2]	8	70	3.1	3.1	88	0.6	1.3	1.2	1.2	5.4	4.5	2.2
3CFI[p]	10	7[1]	9	90	2.4	2.0	64	1.0	5.7	4.0	2.8	7.4	5.7	2.9
1YC7[cu]	8	7	10	0	5.6	5.3	71	3.5	5.5	5.5	3.1	11.5	6.7	4.5
1SJX[tl]	10	7[1]	11	67	1.2	1.2	93	0.7	4.4	2.6	1.8	5.3	4.0	3.0
3EZJ[p]	10	7[1]	12	67	1.2	1.2	93	0.7	1.8	1.8	1.8	3.5	1.4	1.4
1BZQ[sp]	10	8	12	60	3.5	3.4	76	1.5	3.0	2.4	1.3	7.2	4.2	1.8
2P49[sp]	10	8	12	60	3.4	3.1	76	1.7	3.2	2.2	2.0	3.0	2.8	1.8
1ZV5[tcp]	10	8	12	0	5.3	4.9	88	3.2	6.1	3.6	2.5	8.1	5.8	2.1
1KXQ[sp]	7	8[2]	14	0	1.4	1.1	65	0.7	2.1	2.1	2.1	7.4	2.2	2.2
2BSE[tp]	10	8	14	0	5.1	6.0	90	1.2	6.9	3.6	2.8	5.5	4.2	2.7
1OP9[sp]	7	8[2]	15	0	1.3	0.8	76	0.6	3.1	2.9	1.5	3.0	2.1	2.1
1I3V[su]	10	13	15	0	1.6	1.1	100	1.2	4.5	3.6	3.0	4.8	4.5	2.7
1ZVH[sp]	10	8[2]	16	80	3.6	3.3	65	0.8	4.5	3.4	2.3	3.9	3.5	2.7
1U0Q[sl]	10	8[2]	16	71	2.7	2.7	92	0.7	2.8	2.8	2.8	5.0	5.0	1.7
3DWT[su]	11	8	16	67	4.5	3.6	78	1.1	3.3	3.3	2.6	4.1	4.1	2.7

PDB ID	CDR Length			CDR H1			CDR H2		CDR H3 rmsd (Å)					
	H1	H2	H3	Seq. ID	Graft rmsd	Homology LowE rmsd	Seq. ID	Graft rmsd	In Native Context			Homolog		
				(%)	(Å)	(Å)	(%)	(Å)	LowE	LowRMS	LowALL	LowE	LowRMS	LowALL
1QD0^{sl}	13	8	16	100	0.4	1.8	67	1.4	0.9	0.9	0.9	3.3	2.9	2.2
1MVF^{sp}	8	8^2	17	70	3.9	2.9	65	2.9	3.4	3.4	3.4	6.8	5.9	3.1
1RI8^{scp}	10	7^1	17	75	2.7	2.4	73	1.0	5.1	4.0	2.6	2.9	2.9	2.8
1ZVY^{sp}	10	8^2	18	60	2.7	2.3	91	1.6	4.3	3.3	3.0	3.9	3.0	2.8
1F2X^{tcu}	10	6	19	70	4.4	3.4	100	5.5	3.9	3.9	3.3	8.6	7.8	5.4
1RJC^{scp†}	10^d	7	19	60	4.5	3.2	82	4.2	3.6	3.6	2.8	3.2	3.2	2.5
1JTO^{scp}	10	8^2	24	70	4.1	3.3	88	3.0	6.7	4.3	3.4	7.9	6.5	5.0
1MEL^{scp}	10	8^2	24	70	3.9	2.9	88	2.9	4.0	4.0	3.8	7.6	6.5	5.2
1XFP^{scp}	10	8^2	24	70	3.9	5.0	88	1.8	4.2	4.2	3.9	15.7	6.3	4.9
1ZMY^{scp}	10	8^2	24	70	3.9	3.8	88	3.1	3.9	3.9	3.8	5.1	4.9	4.0
Median					3.6	3.1		1.5	3.9	3.4	2.6	5.4	4.5	2.7

Table 2

Structural measurements of cAb CDR H3. Table showing the backbone dihedral angles (φ,ψ) for the CDR H3 C-terminal, and the distance of the CDR H3 loop apex to the closest residue in the heavy chain framework (residue 46). The CDR H3 loop length is denoted by *n* and represents Chothia numbered residue 102. ^{s} and ^{t} indicate structures in the stretched-twist and twist sets, respectively. *Overall*, *StretchTw* and *Twist* mean and standard deviations (Std. Dev.) show the respective values for the entire dataset, the *stretched-twist*ed structures and *twist*ed subsets respectively.

	Backbone Dihedral Angles (°)												**Distance to residue 46 (Å)**	
Relative Residue Numbers	***n*+1**		***n***		***n*-1**		***n*-2**		***n*-3**		***n*-4**		***n*-4**	***n*-5**
Measured Variable	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**C_α-C_α**	**C_α-C_α**
PDB ID														
1BZQ^{s}	-123.2	134.8	-117.0	117.0	-105.4	-4.4	-98.8	118.1	-87.1	-4.6	-60.1	-17.8	11.1	7.7
2P49^{s}	-135.0	137.7	-121.2	130.2	-111.4	-9.1	-96.3	132.1	-83.7	-9.9	-62.0	-20.2	10.7	7.3
1KXQ^{s}	-125.2	143.4	-156.0	140.7	-118.7	2.1	-81.3	133.7	86.0	7.1	-106.6	172.1	11.5	8.7
1OP9^{s}	-133.1	135.6	-120.2	134.2	-118.0	-16.8	-81.4	133.3	-87.4	-8.9	-59.9	-21.7	11.8	8.3
1I3V^{s}	-130.1	150.0	-130.3	144.5	-67.8	-48.3	-105.6	136.0	-112.4	48.3	-68.1	-41.2	10.6	7.3
1ZVH^{s}	-133.0	145.9	-131.1	140.5	-106.4	-15.1	-78.9	134.9	-89.7	6.1	-50.1	-47.4	11.1	8.0
1U0Q^{s}	-131.2	146.7	-135.5	149.6	-95.4	-37.1	-88.5	146.1	-84.3	-5.2	-64.2	-21.2	11.3	7.8
3DWT^{s}	-132.7	150.9	-96.6	133.6	-114.4	-32.4	-106.6	154.2	-73.0	-25.1	-51.1	-29.7	11.2	7.8
1QD0^{s}	-128.1	144.1	-119.1	136.4	-83.2	-30.1	-82.2	96.4	81.9	6.5	-89.1	159.0	12.1	8.6
1MVF^{s}	-135.8	142.2	-131.0	136.8	-122.8	0.8	-116.0	149.1	-84.1	-6.7	-59.7	-14.4	11.2	7.7
1RI8^{s}	-125.0	167.6	-129.9	118.5	-110.8	1.6	-120.0	123.3	-94.6	14.8	-68.3	-14.5	10.9	7.5
1ZVY^{s}	-135.4	149.2	-128.9	137.8	-102.0	-16.2	-106.1	138.0	-83.1	-8.8	-66.6	-11.9	10.5	7.0
1RJC^{s}	-169.9	166.2	-122.8	141.3	-95.4	-3.8	-127.8	123.1	-96.0	3.0	-75.4	2.4	9.6	6.5
1JTO^{s}	-116.5	152.2	-123.0	119.8	-96.3	-10.8	-79.1	100.7	98.3	-28.8	-57.1	130.0	12.5	8.9
1MEL^{s}	-115.1	155.1	-103.6	123.1	-84.2	-32.5	-80.8	89.3	98.2	-26.6	-67.8	129.7	12.6	9.0
1XFP^{s}	-136.8	144.2	-116.6	137.2	-90.4	-16.6	-79.4	101.9	98.3	-33.0	-57.1	131.4	11.2	7.7
1ZMY^{s}	-124.1	152.4	-123.8	129.4	-91.3	-11.6	-70.0	98.2	111.2	-42.8	-51.7	129.7	12.7	9.1
1G9E^{t}	-139.2	164.0	-115.0	132.4	-113.6	19.0	-136.7	86.2	-101.8	179.6	105.5	-9.5	16.1	19.0
1HCV^{t}	-135.4	150.8	-130.3	140.5	-127.7	-6.6	-107.0	124.9	-123.2	140.3	79.4	15.3	14.7	18.2
1SJX^{t}	-120.3	144.9	-138.8	139.1	-91.5	-35.1	-71.2	125.0	-129.9	160.9	91.8	-8.3	15.5	18.6

	Backbone Dihedral Angles (°)												**Distance to residue 46 (Å)**	
Relative Residue Numbers	***n*+1**		***n***		***n*-1**		***n*-2**		***n*-3**		***n*-4**		***n*-4**	***n*-5**
Measured Variable	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**φ**	**ψ**	**C_α-C_α**	**C_α-C_α**
PDB ID														
1ZV5^t	-129.3	149.3	-131.9	136.6	-136.6	-12.1	-73.8	139.0	-84.4	100.1	-112.7	115.8	12.0	12.5
2BSE^t	-143.3	116.4	-112.6	141.4	-55.8	-33.2	-117.1	102.9	47.1	63.7	-75.5	149.3	12.2	8.4
1F2X^t	-143.3	137.2	-96.2	139.2	-52.3	-57.1	-77.3	149.8	-97.4	-1.4	-74.8	-8.6	7.9	8.6
1IEH	-126.6	4.4	63.4	112.0	-97.3	-32.0	-97.7	149.5	-142.6	156.9	-160.1	7.6	11.5	14.6
3CFI	-69.8	140.7	50.4	55.8	-164.3	165.9	-129.2	174.2	-109.0	139.7	-126.0	24.7	8.5	12.2
1YC7	-116.7	136.8	-139.8	140.9	-121.1	146.3	-131.1	129.1	-105.3	155.0	-112.5	156.3	24.2	26.4
3EZJ	-143.2	171.0	-126.7	135.8	-100.6	140.4	-147.1	144.4	-125.8	161.2	-90.5	-175.5	22.2	25.5
Overall Mean	-132.2	147.0	-123.1	134.8	-99.6	-17.6	-94.9	123.3	-38.7	23.0	-47.9	37.8	12.5	11.1
Overall Std. Dev.	11.2	11.1	13.4	8.4	21.4	18.0	19.5	20.4	87.8	63.2	57.8	77.8	3.6	5.6
StretchTw Mean	-131.2	148.1	-123.9	133.6	-100.8	-16.5	-94.1	124.0	-23.6	-6.8	-65.6	36.1	11.3	7.9
StretchTw Std. Dev.	11.9	9.2	12.9	9.4	14.8	14.8	17.1	20.0	91.3	21.4	14.2	82.0	0.8	0.7
Twist Mean	-135.1	143.8	-120.8	138.2	-96.2	-20.8	-97.2	121.3	-81.6	107.2	2.3	42.3	13.1	14.2
Twist Std. Dev.	9.0	16.0	15.8	3.3	36.1	26.6	27.1	23.3	65.3	67.8	99.8	71.3	3.0	5.0

Table 3

Observed sequence differences in cAb V_HH domains relative to V_H domains. Underlined residues indicate new observations, and those which occur in most cAb V_HH are summarized in boldface. ^{c}Camel or ^{l}llama residues observed only once. ^{FR}Framework or ^{CDR}Complementarity Determining Region.

Chothia Residue Number	Amino Acids Present (Most/Some/Once)		
	Classical	cAbs	Summary (Reported and Newly Discovered)
11^{FR}	-/LV/-	S/L/-	L→S[10]
14^{FR}	P/-/AL	A/P/-	**P→A**, only two V_HHs have P
18^{FR}	-/LV/-	L/-/-	**[LV]→L**
23^{FR}	K/AT/PQSV	-/AT/E^{l} Q^{c} S^{l}	A, K most common in classicals; K absent in camelids
29^{CDR}	-/FILV/Y	-/ADFGINSVY/-	[FILV]→[DGNS],[10] camelids more varied, 25% Y
37^{FR}	V/I/AF	F/VY/D^{l}	V→F,[11] V→[FY][7]
44^{FR}	G/AKRS/E	E/GQ/A^{c} D^{c}	G→E,[11] G→[EQ],[12] [KR]→other
45^{FR}	-/L/F	R/L/P^{l}	L→R[1]. Occasional C reported in cDNA.[10]
47^{FR}	-/W/LY	G/AFLSW/ Y^{l}	W→G,[11] W→[GSLF][12]
49^{FR}	G/AV/S	A/SV/G^{l}	**G→A**
82^{FR}	-/ILM/-	M/-/-	**[ILM]→M**, I occurs rarely
84^{FR}	S/ANT/-	P/S/G^{l} L^{l} R^{c}	**S→P**, only two V_HHs have S
87^{FR}	-/ST/-	T/-/-	Camelids contain only T
94^{FR}	R/KS/GHPV	A/K/G^{c} I^{c} T^{c} V^{l}	R→non-charged, disrupts salt bridge to res. 101[10]
103^{FR}	-/W/-	-/R,W/-	W→R[9]
105^{FR}	-/AQT/HKR	Q/-/-	**[AQT]→Q**, Camelids contain only Q
108^{FR}	T/L,S/M	Q/-/E^{c}	**[LTS]→Q**
109^{FR}	V/L/-	V/-/-	**L→V**

Table 4

Key residues for CDR detection. Underlined residues denote additional residues that must be incorporated into the classical antibody CDR identification rules (32) in order to detect cAb CDRs. Brackets enclose multiple residues codes that appear at each sequence position. X indicates any of the 20 amino acids.

CDR Loop	Chothia Definition(32)	Starting Position	Preceded By	Length(Residues)		Followed By
				Classical	Camelids	
H1	H26-H35	After first Cys	CXXX	10-12	7-13	W[IVFYAMLND][RKQVNC][QKHELR]
H2	H50-H56	Always 15 residues after H1 end	[WYLAFGS][ILVM][GASV]	7,8,10	6-13	XX[YFLS]XXXX[QKRES]
H3	H95-102	Always 33 residues after H2 end	CXX	3-22	8-24	[WR]GX[GRDS]