Published in final edited form as:

J Mol Biol. 2008 October 10; 382(3): 779–789. doi:10.1016/j.jmb.2008.07.054.

Construction of a large phage–displayed human antibody domain library with a scaffold based on a newly identified highly soluble, stable heavy chain variable domain

Weizao Chen¹, Zhongyu Zhu^{1,2}, Yang Feng¹, Xiaodong Xiao¹, and Dimitre S. Dimitrov^{1,*}

¹ Protein Interactions Group, CCRNP, CCR, NCI-Frederick, NIH, Frederick, MD 21702, USA

² BRP, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA

Abstract

Currently, almost all FDA approved therapeutic antibodies and the vast majority of those in clinical trials are full-size antibodies mostly in IgG1 format of about 150 kDa size. A fundamental problem for such large molecules is their poor penetration into tissues (e.g. solid tumors) and poor or absent binding to regions on the surface of some molecules (e.g. on the HIV envelope glycoprotein (Env)) which are accessible by molecules of smaller size. We have identified a phage-displayed heavy chain only antibody by panning of a large (size $\sim 1.5 \times 10^{10}$) human naive Fab library against an Env, and found that the heavy chain variable domain (VH) of this antibody, designated as m0, was independently folded, stable, highly soluble, monomeric, and expressed at high levels in bacteria. M0 was used as a scaffold to construct a large (size $\sim 2.5 \times 10^{10}$) highly-diversified phage-displayed human VH library by grafting naturally occurring CDR2s and CDR3s of heavy chains from five human antibody Fab libraries, and randomly mutating four putative solvent-accessible residues in CDR1 to A, D, S or Y. The sequence diversity of all CDRs was determined from 143 randomly selected clones. Most of these VHs were with different CDR2 origins (6 of 7 groups of VH germlines) or CDR3 lengths (ranging from 7 to 24 residues) and could be purified directly from the soluble fraction of the *E. coli* periplasm. The quality of the library was also validated by successful selection of high-affinity VHs against viral and cancer-related antigens; all selected VHs were monomeric, easily expressed and purified with high solubility and yield. This library could be a valuable source of antibodies targeting size-restricted epitopes and antigens in obstructed locations where efficient penetration could be critical for successful treatment.

Keywords

antibody library; phage display; human; VH domain; framework scaffold

Introduction

Monoclonal antibodies (mAbs) with high affinity and specificity are now well established therapeutics and invaluable tools for biological research. The vast majority of these antibodies are full-size typically in an IgG1 format. Antibody fragments which are significantly smaller than full-size antibodies (~ 150 kDa), e.g. Fabs (~ 60 kDa) or single chain Fv fragments (scFvs) (20–30 kDa), have been widely used especially as imaging reagents and candidate therapeutics typically conjugated with toxins or other agents. These antibody fragments can be selected

*Corresponding author. Mailing address: CCRNP, CCR, NCI-Frederick, NIH, Bldg. 469, Rm. 150B, Frederick, MD 21702-1201 Phone: (301) 846-1352 Fax: (301) 846-5598 E-mail: dimitrov@ncicrf.gov

from highly diverse libraries and readily produced in bacterial or yeast cell culture, resulting in improved yields, better quality product and lower costs for production. Moreover, smaller fragments of antibodies are of great interest and advantageous for pharmaceutical applications, for example, cancer targeting and imaging where small antigen binding molecules are needed to penetrate into large solid tumors.

In the late 1980s, the smallest known antigen-binding fragment, which consisted of only the heavy chain variable region (VH) of an antibody, was first isolated when a murine VH repertoire was screened for binding to lysozyme.¹ It has been demonstrated that the variable domains of antibody light chains (VLs) alone can also retain significant binding ability in the absence of heavy chains.² These fragments with size ranging from 11 kDa to 15 kDa were called “domain antibodies” or “dAbs”. The absence of VL or VH domain means that the paratope is concentrated over a smaller area so that the dAbs provide the capability of interacting with novel epitopes that are inaccessible to conventional VH-VL pairs and penetrating into solid tumors even better than Fab and scFv. Before dAbs can be suited for such applications, several issues need to be addressed, including low stability, low or absent solubility, and tendency to aggregate primarily due to the hydrophobic area exposed in the absence of VL or VH. Since it has been known that a unique kind of antibodies is naturally formed only by heavy chains in camels, dromedaries and llamas, dAbs can be also produced directly from these species or camelized for improved solubility.³ However, use of mAbs derived from nonhuman species such as mouse or rabbit may result in immune responses to the foreign immunoglobulin epitopes in humans that could limit the long-term use of these reagents.

Highly diverse antibody libraries have become important sources for selection of antibodies with high affinity and novel properties. Combinatorial strategies provide efficient ways of creating antibody libraries containing a large number of individual clones. These strategies include the reassembly of naturally occurring genes encoding the heavy and light chains from either immune or nonimmune B-cell sources⁴ or introduction of synthetic diversity to either the framework regions (FRs) or the complementarity-determining regions (CDRs) of the variable domains of antibodies.⁵

Here, we describe the identification of a human heavy chain only antibody and its use as a scaffold for construction of a phage-displayed VH library as well as an approach to introduce genetic diversity in this library, where natural human CDR2, CDR3, and synthetic CDR1 repertoires are combined into a single human VH framework scaffold. The usefulness of the library has been demonstrated by the successful selection of high affinity binders to viral and cancer-related antigens.

Results

Identification of a heavy chain only antibody and its VH domain, m0

We have recently constructed a human naive Fab library (1.5×10^{10} members) from peripheral blood B cells of 22 healthy donors, spleens of three healthy donors, and lymph nodes of 34 healthy donors (Chen, Zhu and Dimitrov, unpublished). While panning this library against a recombinant soluble HIV-1 envelope glycoprotein (Env) gp140, we identified a phage-displayed binder, which contains a TGA stop codon at the very beginning of the light chain. The stop codon resulted from a reading frame shift due to a nucleotide deletion but the phage-displayed antibody was still isolated and functional as a heavy chain only (Chen and Dimitrov unpublished). We cloned the VH domain of this antibody and found that this completely natural VH domain, designated as m0, was independently folded, stable, highly soluble, monomeric, and expressed at high levels in bacteria (data not shown). The m0 sequence differs from the

closest human germline (VH3-23) sequence by mutations in all FR regions as well as in the CDRs (Fig. 1).

Design and construction of the library

To achieve high diversity and decrease possible immunogenicity and toxicity by using naturally occurring human antibody sequences we grafted CDR2s and CDR3s from five other human Fab libraries (Fig. 1). However, because CDR1s are relatively more conserved than CDR2s and CDR3s we randomly mutated four putative solvent-accessible residues (#27, 29, 31 and 32) (IMGT numbering system) in the m0 CDR1 to A, D, S, or Y (Fig. 1). To make the library more suited for selection of antibodies against a wide range of antigens including infectious and cancer-related antigens, two naïve Fab libraries from cord blood (Chen and Dimitrov, unpublished) and two naïve Fab libraries from adults (Chen, Zhu and Dimitrov, unpublished, and ⁶) as well as an immune Fab library from HIV-1 patients⁷ were used as sources for the grafted CDR2s and CDR3s. In order to graft efficiently highly diverse human VH gene segments, we designed a new set of primers (Table 1) for amplification of CDR2 and CDR3 repertoires based on human germline VH sequences (see, e.g., http://imgt.cines.fr/textes/IMGTrepertoire/Proteins/alleles/human/HuA1_list.html). These primers target highly conserved (within each group of germline sequences) regions in the FRs (indicated by arrows in Fig. 1). In combination with each other (Table 2) they allow efficient amplification of human VH gene segments.

The VH library was constructed in three steps. In the first step, eight and three PCRs were performed for amplification of CDR2 and CDR3 gene segments from each source library, respectively (Fig. 2A), and the products from each source library were pooled. The m0 CDR1 was amplified using a degenerate primer which overlaps with the whole CDR1. The FR3 segment was obtained from m0 for assembly of the VHs. In the second step, overlapping PCRs were performed to join CDR1s to CDR2s and FR3 to CDR3s, respectively (Fig. 2B). In the third step, the VHs were assembled from the products of the second step by overlapping PCR (Fig. 2C). The products were cloned into phagemid pComb3X, and a library of about 2.5×10^{10} members was obtained by performing 100 electroporations as described in Materials and Methods.

Sequence diversity of the library

To assess the library diversity, we analyzed 143 sequences of randomly selected clones, and evaluated the frequency of A, D, S, and Y mutations in the CDR1s, the somatic mutational diversification (number of mutations compared to the corresponding germline) of the CDR2s, the distribution of the CDR2 sequences corresponding to the seven germline VH groups, and the length distribution of the CDR3s. We did not find completely identical VH sequences. Identical CDR1 were found due to the relatively limited number (256) of theoretically possible random CDR1 sequences. Four groups (11, 9, 6 and 4 members, respectively) of sequences each contained identical CDR2s from VH group 1, 4 and 5, respectively. We did not find sequences with identical CDR3s. The frequency of mutations to a specific residue (A, D, S or Y) in CDR1 was not significantly dependent on the mutation position but for all positions the frequencies of D and especially Y were higher than those of A and S although not more than about 2-fold (Fig. 3A). We found CDR2 sequences from all VH groups except 6; about half of all sequences belonged to group 4 (Fig. 3B). About 40% of the CDR2s were identical to germline sequences while 5% contained more than four mutations (Fig. 3C). The CDR3 length varied from 5 to 24 amino acid residues; most of the CDR3 were between 12 and 18 residues long (Fig. 3D). Our domain library contains relatively limited number of CDR3s with length shorter than 8 or longer than 19 residues. There was no preferential grafting of CDR2s from certain groups dependent on the CDR3 length (Fig. 4). These data, although derived from

relatively small number of sequences (143) indicate that the library appears to be highly diversified.

VH expression and solubility

We evaluated the expression and solubility of VHs by measuring the yield of soluble proteins purified from the soluble fraction of the *E. coli* periplasm for three panels of clones. Firstly, we measured the expression of two clones from each group as determined by their CDR2s, one with relatively short CDR3 and the other one – with long CDR3. As shown on the SDS-PAGE 11 of 12 clones could be expressed in the form of soluble proteins and the estimated yield ranged from 0.5 to 24 mg l⁻¹ (Fig. 5A). There was no significant correlation between the yield and CDR2 origin or CDR3 length. Secondly, 12 clones with VH4-derived CDR2s and CDR3 length ranging from 7 to 24 residues were measured for soluble protein expression. Most (10 of 12) of the clones gave yield varying from 0.5 to 15 mg l⁻¹; there was no obvious correlation between the yield and the CDR3 length (Fig. 5B). Moreover, we evaluated the solubility of VHs selected after three and four rounds of biopanning of the library with viral antigens. Almost all clones gave high-level yield ranging from 5.0 to 30 mg l⁻¹. They have highly diversified CDRs and varying CDR3 length (Table 3 and data not shown). These results indicate that the scaffold used for the library construction can support CDR2s derived from a variety of human VH germlines and CDR3s with wide distribution of length.

VH folding and oligomerization

Human antibodies binding Staphylococcal protein A (SPA) in the Fab regions are encoded by gene segments belonging to the VH3 group, thus SPA binding has been used as a marker for proper folding of human VH3.⁸ To estimate the folding of the recombinant VHs with VH3 scaffold we randomly picked 95 clones after four rounds of panning against SPA and obtained 79 complete sequences. All have CDR2s derived from VH3 (data not shown). Interestingly, the occurrence of residue D in the CDR1 positions #27, 29 and 32, and S in position 31 was significantly increased (on average about 2-fold) but the Y in all mutated CDR1 positions decreased more than 2-fold compared to their frequency before selection (Fig. 6). Thus it appears that D and S in those positions could stabilize domain antibodies in the library. CDR3s of these clones were significantly diverse with length from 7 to 17 residues (data not shown).

Antibodies in the format of scFv have the tendency to form dimers and higher order multimers in a clone-dependent and relatively unpredictable way. The oligomerization of five VHs randomly selected was measured by size exclusion chromatography and they were all monomeric (data not shown). There is evidence that VHs are prone to aggregate upon concentration or prolonged storage at 4°C.^{9,10} The five VHs tested for oligomerization were concentrated to 10 mg/ml. After being stored at 4°C for more than 8 weeks no precipitation was observed with these 5 protein solutions (data not shown). These data suggest that dAbs selected from this library have such desirable properties for biotechnological applications as high level of expression, solubility, and resistance to aggregation in solution.

Selection of antigen-specific domain antibodies

To evaluate the performance of the library we panned it against the vaccinia protein B5R conjugated to magnetic beads as described in the Materials and Methods. Enrichment was achieved (data not shown) and six positive clones were identified on monoclonal phage-based enzyme-linked immunosorbent assay (mpELISA). They were highly mosaic, especially in CDR1s and CDR3s (Table 3). The CDR2s were derived from two VH groups – three and four. Two clones were expressed and purified, and their binding and specificity were confirmed by ELISA against B5R and several unrelated antigens. One of the clones, designated m301, exhibited high subnanomolar affinity of binding with EC₅₀ of 0.013 µg/ml. Interestingly this highest affinity binder also has the longest CDR3 (17 residues) (Table 3). High-affinity binders

against several other viral and cancer-related antigens were also selected and are being characterized (Chen and Dimitrov, unpublished).

Discussion

Twenty one mAbs are currently approved by the USA FDA for clinical use; almost all of them (except ReoPro which is Fab) are full-size antibodies mostly in IgG1 format of about 150 kDa size. A fundamental problem for such large molecules is their poor penetration into tissues (e.g. solid tumors) and poor or absent binding to regions on the surface of some molecules (e.g. on the HIV envelope glycoprotein) which are accessible by molecules of smaller size. Therefore, a large amount of work especially during the last decade has been aimed at developing novel scaffolds of much smaller size and higher stability.¹¹ Antibody-derived scaffolds, specifically those derived from antibody domains are potentially useful candidate therapeutics. Their small size leads to relatively good penetration into tissues and the ability to bind into cavities or active sites of protein targets which may not be accessible to full size antibodies. This could be particularly important for the development of therapeutics against rapidly mutating viruses, e.g. HIV. Because these viruses have evolved in humans to escape naturally occurring antibodies of large size, some of their surface regions which are critical for the viral life cycle may be vulnerable for targeting by molecules of smaller size including dAbs. In addition, dAbs are typically monomeric, of high solubility and do not significantly aggregate or can be engineered to reduce aggregation. Their half-life in the circulation can be relatively easily adjusted from minutes or hours to weeks. In contrast to conventional antibodies, dAbs are well expressed in bacteria, yeast, and mammalian cell systems. Finally, the small size of dAbs allows for higher molar quantities per gram of product, which should provide a significant increase in potency per dose and reduction in overall manufacturing cost (<http://www.domantis.com>).

In this article we describe the identification of a VH based scaffold which is stable and highly soluble. It was used for construction of a large-size (20 billion clones) dAb phage-display library by grafting CDR3s and CDR2s from five of our Fab libraries and randomly mutagenizing CDR1. Panning of this library against a viral antigen (B5R) resulted in the selection of a high affinity (subnanomolar) binder with EC₅₀ of 0.013 µg/ml. High-affinity binders were also selected against other viral and cancer-related antigens (Chen and Dimitrov, unpublished). These results suggest that the library could be a useful addition to the existing libraries based on antibody domains including those from humans² and camelids.³

Camels produce functional antibodies devoid of light chains, VHHs. Their stabilization compared to human VHs is achieved by replacing aliphatic residues in the former light chain interface with hydrophobic residues, packing against the FR and stabilizing the global VHH fold by long CDR3s. It has been demonstrated that the variable domains of human VH3 are significantly more soluble and stable than those from any other human VH and VL groups in the absence of light chains.² Jirholt et al. used a camelized human VH3-23 germline (DP47) as a master framework for construction of a dAb library of 9×10^6 members but no further characterization was described in terms of solubility and functionality of dAbs from the library.¹² We have recently constructed a large non-immune human Fab library ($\sim 1.5 \times 10^{10}$ members) from the lymph nodes, spleen and peripheral blood lymphocytes of 59 donors. One of these Fabs had a stop codon in the light chain but was still selected from the library by panning against an HIV Env and was functional as a heavy chain alone. We cloned the VH domain of this antibody and found that it exhibits high level of expression and high solubility. This completely natural VH domain, belonging to the VH3 family too, was used as a framework to construct a large human VH domain library ($\sim 2.5 \times 10^{10}$ members) by grafting in vivo-formed CDR2s and CDR3s from our other human antibody libraries and mutating four residues in CDR1. The library exhibited a high degree of variability and most of the tested VHs were

highly soluble. The quality of the library was also validated by selection of VHs against several antigens. In contrast to the previously described library¹² there was no need to camelize the framework which was part of a naturally occurring antibody in humans.

The size and sequence diversity are thought to be keys to high-quality libraries. A finding in line with theoretical considerations is that the affinity of antibodies selected is positively correlated to the size of the library, with K_d s ranging from 10^{6-7} for smaller libraries to 10^9 M^{-1} for larger ones, and antibodies with affinity comparable to those obtained from immune libraries can be selected from large naïve libraries. Aiming at generating a highly diverse library, we designed a new set of primers to allow amplification of as many VH genes as possible. The complexity of the library was determined by analyzing the sequences of 143 randomly selected clones (Figs. 3 and 4). They have CDR2s derived from 6 (VH1-5 and 7) of 7 human VH groups (Fig. 3B). Since VH6 is the smallest group of germlines, we assume that CDR2s with VH6 origin could also be found if more clones are sequenced. The length, sequence, and pairing of CDR3s with different germline-derived CDR2s varied (Fig. 3D and 4). These data suggest that this library should have high sequence diversity. In libraries previously constructed, primers were designed for amplification of naturally occurring CDRs but it appears from the resulting antibodies that the grafted CDRs belong to certain (DP47 and DPL3) frameworks only.^{4,12} The primers we designed may cover CDRs derived from all groups of human VH germlines as described above. Thus, they may facilitate more rapid and efficient construction of large human VH domain libraries.

While the combinatorial strategy described here provides a powerful tool for creating tremendous diversity, the VH3 framework scaffold used in the library construction plays a challenging role because it is expected to support the CDRs from non-VH3 germlines and maintain a stable tertiary fold. Structural incompatibility between these foreign CDRs and the fixed framework could potentially prevent the formation of stable and soluble VHs. SPA binding is considered a marker for proper folding of human VH3.⁸ To see whether the recombinant VHs with a fixed VH3 scaffold can retain the activity, we cycled the library through four rounds of selection against SPA. All 79 clones randomly picked from the fourth round of selection have CDR2s derived from VH3 suggesting that VH3-derived CDR2s and their flanking FR regions are essential for their property of SPA binding, in agreement with previous studies.⁸ Interestingly, positions #27 and 29 in CDR1s were significantly biased toward residue D in those clones after SPA selection while residue Y dominated in these positions of VHs with VH3-derived CDR2s before selection (Fig. 6). There was also a slight increase in the usage of D in position #32 but not in position #31. Given that D is a relatively small and polar residue, and Y is an aromatic residue, we propose that the polarity of the residues in these positions significantly impacts the folding of VH3 with respect to SPA binding activity. It has recently been demonstrated that this property can be used to monitor the structural stability and soluble expression of VH3.¹³ Thus, the fact that VHs with non-VH3 CDR2s lack this property further indicates that these VHs may not be soluble and stable. To address this issue, we measured the yield of soluble VHs with non-VH3 CDR2s and CDR3s of 7 to 24 residues long (Fig. 5). The results show that most VHs have favorable yield of soluble proteins and the yield is not significantly related to CDR2 origin and CDR3 length. It is noteworthy that in addition to CDRs, part of FRs was also grafted because the primers were designed based on the FR regions (Fig. 1). Although the FR regions are conserved within each group, they differ among all groups (http://imgt.cines.fr/textes/IMGTrepertoire/Proteins/alleles/human/HuAl_list.html). Thus our data suggest that the VH3 framework scaffold described here not only is capable of presenting diverse CDRs but can also support diverse FRs from non-VH3 groups.

A central question in the evaluation of the library is whether high-affinity functional VHs could be directly selected against a panel of antigens. Our library design was based on the principle

that a specific single framework had the potential to include CDRs derived from other groups of germline genes and the small antigen-binding surface was capable of presenting structural diversity enough to form paratopes for a wide range of antigens. Biopanning of the library showed that VHs with estimated affinity in the nanomolar range could be easily obtained. These binders are significantly diversified in their CDRs (Table 3) suggesting that a number of different antibody specificities could be generated. Thus this library could be useful for selection of high-affinity binders with potential applications for development of therapeutics.

Materials and Methods

Amplification of CDR repertoires and FR fragments

Primers used for PCR amplification of gene fragments are described in Table 1. For amplification of CDR1 repertoire and FR3 fragment, m0 (Fig. 1) was used as the template. The degenerate primer H1R enables randomization of four residues in CDR1 to A, D, S, and Y. Four non-immune human antibody phage display libraries we have recently constructed and one immune library constructed from HIV patients have been used as templates for amplification of CDR2 and CDR3 repertoires (Fig 2A). These are: (a) a naïve human Fab library (5×10^9 members) constructed from peripheral blood B cells of 10 healthy donors;⁶ (b) a naïve human Fab library (1.5×10^{10} members) constructed from peripheral blood B cells of 22 healthy donors, spleens of 3 healthy donors, and lymph nodes of 34 healthy donors (Chen and Dimitrov, unpublished); (c) two naïve human Fab libraries (6×10^8 and 7.2×10^8 members, respectively) constructed from cord blood of 2 healthy babies, respectively (Chen and Dimitrov, unpublished); and (d) an immune human Fab library constructed from bone marrow obtained from 3 long-term nonprogressors whose sera exhibited the broadest and most potent HIV-1 neutralization among 37 HIV-infected individuals.⁷ To maintain maximal diversity, 8 and 3 PCR reactions were carried out for each library template separately to obtain CDR2 and CDR3 repertoires, respectively, using different primer combinations (Table 2). PCR was performed in a volume of 50 μ l using High Fidelity PCR Master (Roche, Indianapolis, IN), 500 pM of each primer and 0.5 μ g of templates (VH scaffold m0 or library plasmid DNA) for 30 cycles (45 sec at 94 °C; 45 sec at 55 °C; and 1 min at 72 °C). Products for each primer combination from one source library were pooled, purified from 2% agarose gel using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA), and quantified by reading the optical density (O.D.) at 260 nm (1 O.D. unit = 50 μ g/ml). Finally all products from the five source libraries were pooled at a molarity ratio calculated by counting the number of donors for construction of these libraries, i.e. 10:59:1:1:3.

Assembly of entire VHs

The primers used in the first round of PCR create identical sequences in the downstream regions of the CDR1s, the upstream regions of CDR3s, and both the downstream and the upstream regions of CDR2s. These identical sequences are homologous to FRs and serve as the overlap for the second- and third-round extensions. In the second-round extension (Fig. 2B), CDR1 fragments containing the whole FR1 and partial FR2 on both sides were joined to the CDR2 fragments containing partial FR2 and partial FR3 on both sides by overlapping PCR performed in a volume of 100 μ l using both templates (in the same molarities) for 7 cycles in the absence of primers and additional 15 cycles in the presence of primers (500 pM of FR1F and 500 pM of H2R1-5 mixture). Under the same condition, FR3 fragments were joined to CDR3 fragments containing partial FR3 and the whole FR4 on both sides by overlapping PCR using primers FR3F and H3R. In the third-round extension (Fig. 2C), the entire VHs were formed by annealing the products in the second-round extension to each other using overlapping PCR with the extension primer HISR and FR1F appended with SfiI restriction sites.

Preparation of the VH library

Gel-purified VH products were digested with SfiI (BioLabs, Ipswich, MA), and cloned into the phagemid vector pComb3X. The SfiI-digested and gel-purified VH fragments and vector pComb3X, 80 µg and 230 µg respectively, were ligated in an 8-ml reaction mixture with 10000 units of T4-DNA ligase (BioLabs, Ipswich, MA) at 16 °C for 72 h. The ligation product was then desalted and concentrated by passing through a 4-ml Amicon Ultra-4 centrifugal filter with a cutoff 3000 MW (Millipore, Billerica, MA) at $4000 \times g$ for 20 minutes at room temperature and washing 3 times with 4 ml of distilled water each. About 100-µl ligation product was recovered from the filter and stored at -20 °C for later use.

For electroporations, 1 L of 2YT medium containing 1% glucose (w/v) was pre-warmed at 37 °C and 100 gene pulser cuvettes with 1 mm gap (Bio-Rad, Hercules, CA) were chilled on ice. At the same time the desalted ligation product and 4 ml of E. coli strain TG1 electroporation-competent cells (Stratagene, La Jolla, CA) were thawed on ice. The TG1 competent cells were divided into 10 pre-chilled 1.5-ml Eppendorf tubes with 400 µl each. Ten µl of ligation product was added to each tube and mixed by pipetting gently. Forty one µl of mixtures were transferred to each cuvette, the cuvette was gently tapped on the bench to make the mixture fill out the bottom. Electroporations were performed at 1.8 kV, 25 µF, and 200 Ω and the cuvettes were flushed immediately with 1 ml and then twice with 1 ml of pre-warmed 2YT medium each and combine the 3 ml in a 2-L flask. After all electroporations were completed, 700 ml of pre-warmed 2YT medium were added to the flask to make a volume of 1 L in total. The cultures were incubated at 37 °C with shaking at 250 rpm for 30 min. Ten µl of the culture were 10-fold serially diluted in 100 µl of 2YT medium, and plated on 2YT agar plates containing 2% glucose (w/v) and 100 µg/ml of ampicillin. The plates were incubated overnight at 37 °C. The total number of transformants was calculated by counting the number of colonies, multiplying by the culture volume, and dividing by the plating volume.

For preparation of the library, 1 ml of 100 mg/ml ampicillin was added to the 1-L culture and the culture was then incubated with shaking for additional 2 h at 37 °C. The cell density was measured by reading the O.D.600 of the culture and the total number of cells was calculated by multiplying the O.D.600 value by 5×10^8 (estimated number of cells in 1 ml culture when O.D.600 reaches 1) and the culture volume (1000 in this case). The culture was infected with 10 M.O.I. of M13KO7 helper phage (BioLabs, Ipswich, MA) and incubated at 37 °C for 30 min, shaking for homogenization every 10 min. The cells were collected by centrifuging at 5000 rpm for 10 min and resuspended in 2-L 2YT medium containing 100 µg/ml of ampicillin and 50 µg/ml of kanamycin. Following incubation at 250 rpm overnight at 30 °C, the culture was centrifuged at 5000 rpm for 15 min at 4 °C. The phagemids were prepared from the bacterial pellet using the Qiagen HiSpeed Plasmid Maxi Kit. For phage precipitation, the supernatant was transferred to 2 clean 2-L flasks, mixed well with ¼ volume of PEG8000 (20%, w/v)/NaCl (2.5 M) solution, and incubated on ice for 3 h. The mixture was centrifuged at $14000 \times g$ for 20 min at 4 °C and the phage pellet was resuspended in 100 ml PBS, pH7.4. The phage suspension was centrifuged at 5000 rpm for 10 min at 4 °C. The supernatant was transferred to a clean 200 ml flask, mixed well with ¼ volume of PEG8000 (20%, w/v)/NaCl (2.5 M) solution, and incubated on ice for 1 h. The mixture was centrifuged at $14000 \times g$ for 20 min at 4 °C and the phage pellet was resuspended in 50 ml PBS, pH7.4. The phage suspension was centrifuged at 5000 rpm for 10 min at 4 °C and the supernatant was transferred to a clean 200 ml flask. The phage concentration was measured by reading O.D.280 ($1 \text{ O.D.280} = 2.33 \times 10^{12}/\text{ml}$). For long-term storage, the phages were mixed with the same volume of autoclaved glycerol, aliquoted to make sure that each contains phage particles at least 100 times of the total number of transformants, and stored at -80 °C.

Panning of the library

The library of $\sim 10^{13}$ phage particles was blocked in 1 ml PBS containing 2% non-fat dry milk and incubated with 10, 5, 5 and 5 μg of antigen (recombinant B5R obtained from the Biodefense and Emerging Infections Research Resources Repository (www.beiresources.org)) conjugated to magnetic beads (Dynabeads M-270 epoxy, Invitrogen, Carlsbad, CA) for 2 h at room temperature during the first, second, third and fourth rounds of biopanning, respectively. After incubation the beads were washed 5 times for the first round and 15 times for the later rounds with PBS containing 0.05% Tween 20 (PBST) to remove nonspecifically bound phages. Bound phages were rescued by mixing the beads with *E. coli* TG1 cells for 45 min at 37°C and a phage library was prepared for the next round of biopanning. Clones were randomly picked from the infected TG1 cells in the third and fourth round and subjected to mpELISA to identify clones of phage-displayed VHs with binding activity as described.⁶

Expression and purification of VHs

The pComb3X phagemids containing VH genes were prepared and transformed to *E. coli* HB2151 chemical competent cells. Soluble VHs were expressed as described.⁶ The bacterial pellet was collected after centrifugation at $8000 \times g$ for 10 min and resuspended in PBS buffer containing 0.5 mU polymyxin B (Sigma, St. Louis, MO). After 30 min incubation with rotation at 50 rpm at room temperature, it was centrifuged at $25000 \times g$ for 25 min at 4°C, and the supernatant was used for Hexahistidine-tagged VH purification by immobilized metal ion affinity chromatography (IMAC) using Ni-NTA resin (Qiagen, Valencia, CA) according to manufacturer's protocols.

Binding of soluble VHs

ELISA was performed by using Corning high-binding 96-well plates coated with 1 $\mu\text{g}/\text{ml}$ of antigens and blocked with 3% non-fat dry milk in PBS. Microplate wells were then inoculated with 50 μl of serially diluted soluble VHs for 2 h at room temperature. After 4 washes with PBST, FLAG-tagged VHs were detected by adding 50 μl of 1:5000 diluted HRP-conjugated anti-FLAG antibody (Sigma, St. Louis, MO) to each well. Following incubation with the antibody for 1 h at room temperature, the plates were washed 4 times with PBST and the assay was developed at 37°C with ABST substrate (Roche, Indianapolis, IN) and monitored at 405 nm.

Sequence diversity analysis

Clones were randomly selected from the library and sequenced. The amino acid (aa) sequences were deduced from those clones with complete nucleotide sequences. For analysis of CDR1 sequence diversity, the frequency of A, D, S and Y usage in each position mutated was calculated. The origins of CDR2s were determined and the numbers of aa mutations were calculated by comparing to the germlines of human VHs from IMGT database

(http://imgt.cines.fr/textes/IMGTrepertoire/Proteins/protein/human/IGH/IGHV/Hu_IGH_Vallgenes.html). The FR sequences on both sides of CDR2, i.e. residue #53-55 (IMGT numbering) and residue #70-76, are highly diversified among 7 groups of human VH germlines, and therefore, could be used as markers in addition to CDR2 sequences to determine CDR2 origins. The length of CDR3 was calculated one by one. The pairing between CDR2 origin and CDR3 length was also plotted.

Measurement of VH oligomerization

Superdex75 column was calibrated with protein molecular mass standard of 13.7 (ribonuclease A), 25 (chymotrypsin), 44 (ovalbumin), 67 (albumin), 158 (aldolase), 232 (catalase), 440

(ferritin) and 669 (thyroglobulin) kDa. Purified VHs in PBS were loaded onto the column that had been pre-equilibrated with PBS. The proteins were eluted with PBS at 0.5 ml/min.

Measurement of VH folding

The folding of phage-displayed VHs was measured by their binding ability to SPA. The library of $\sim 10^{13}$ phage particles was blocked in 2% non-fat dry milk in PBS for 1 h at room temperature and passed through a chromatography column (Bio-Rad, Hercules, CA) loaded with 300 ml of nProtein A Sepharose 4 Fast Flow (GE Healthcare, Piscataway, NJ). The column was washed 3 times with 10 ml of PBST each. Bound phages were eluted with 1 ml of 100 mM acetic acid (pH 3.0) followed by neutralization with 0.1 ml of 1 M Tris-HCl (pH 9.0). Eluted phages were rescued by infection of *E. coli* TG1 cells and a phage library was prepared for the next round of selection. In the fourth round of selection, TG1 cells were infected with the eluted phages, serially diluted and plated on 2YT agar plates. Clones were randomly selected from these plates and sequenced. The origins of CDR2s were determined as described above.

Acknowledgements

We thank the personnel from the Biodefense and Emerging Infections Research Resources Repository (Beiresources) for kindly providing B5R antigen for library panning and John Owens in our group for technical assistance. This work was supported by the Intramural AIDS Targeted Antiviral Program (IATAP), National Institutes of Health (NIH), by the NIAID (NIH) Intramural Biodefense Program, by the Gates Foundation to DSD, by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, and by the federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400.

References

1. Ward ES, Gussow D, Griffiths AD, Jones PT, Winter G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* 1989;341:544–546. [PubMed: 2677748]
2. Holt LJ, Herring C, Jespers LS, Woolven BP, Tomlinson IM. Domain antibodies: proteins for therapy. *Trends Biotechnol* 2003;21:484–490. [PubMed: 14573361]
3. 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]
4. Soderlind E, Strandberg L, Jirholt P, Kobayashi N, Alexeiva V, Aberg AM, Nilsson A, Jansson B, Ohlin M, Wingren C, Danielsson L, Carlsson R, Borrebaeck CA. Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat. Biotechnol* 2000;18:852–856. [PubMed: 10932154]
5. Lee CV, Liang WC, Dennis MS, Eigenbrot C, Sidhu SS, Fuh G. High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *J. Mol. Biol* 2004;340:1073–1093. [PubMed: 15236968]
6. Zhu Z, Dimitrov AS, Bossart KN, Crameri G, Bishop KA, Choudhry V, Mungall BA, Feng YR, Choudhary A, Zhang MY, Feng Y, Wang LF, Xiao X, Eaton BT, Broder CC, Dimitrov DS. Potent neutralization of Hendra and Nipah viruses by human monoclonal antibodies. *J Virol* 2006;80:891–899. [PubMed: 16378991]
7. Zhang MY, Shu Y, Phogat S, Xiao X, Cham F, Bouma P, Choudhary A, Feng YR, Sanz I, Rybak S, Broder CC, Quinnan GV, Evans T, Dimitrov DS. Broadly cross-reactive HIV neutralizing human monoclonal antibody Fab selected by sequential antigen panning of a phage display library. *J Immunol. Methods* 2003;283:17–25. [PubMed: 14659896]
8. Potter KN, Li Y, Capra JD. Staphylococcal protein A simultaneously interacts with framework region 1, complementarity-determining region 2, and framework region 3 on human VH3-encoded Igs. *J. Immunol* 1996;157:2982–2988. [PubMed: 8816406]
9. Kortt AA, Guthrie RE, Hinds MG, Power BE, Ivancic N, Caldwell JB, Gruen LC, Norton RS, Hudson PJ. Solution properties of *Escherichia coli*-expressed VH domain of anti-neuraminidase antibody NC41. *J. Protein Chem* 1995;14:167–178. [PubMed: 7576085]

10. Ewert S, Cambillau C, Conrath K, Pluckthun A. Biophysical properties of camelid V(HH) domains compared to those of human V(H)3 domains. *Biochemistry* 2002;41:3628–3636. [PubMed: 11888279]
11. Skerra A. Alternative non-antibody scaffolds for molecular recognition. *Curr. Opin. Biotechnol* 2007;18:295–304. [PubMed: 17643280]
12. Jirholt P, Ohlin M, Borrebaeck CA, Soderlind E. Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework. *Gene* 1998;215:471–476. [PubMed: 9714846]
13. Bond CJ, Marsters JC, Sidhu SS. Contributions of CDR3 to V_HH domain stability and the design of monobody scaffolds for naive antibody libraries. *J. Mol. Biol* 2003;332:643–655. [PubMed: 12963373]

Figure 1.

Nucleotide and deduced protein sequences of the human VH scaffold m0. Sequence alignment was performed between m0 and the closest germline VH3-23. Residues were numbered and CDRs were determined according to IMGT numbering system (http://imgt.cines.fr/IMGT_vquest/vquest?livret=0&Option=humanIg). The FR regions that primers targeted for amplification of CDR2s and CDR3s were indicated by arrows. The residues mutated in CDR1 are shown in bold, italic and large-size fonts.

Figure 2.

Library construction procedure. (A) Five human antibody Fab libraries (two naive Fab libraries from healthy donors; two naive Fab libraries from cord blood of two healthy babies, and an immune Fab library from long-term HIV-1 nonprogressors) were used as a source for human CDR2 and CDR3 repertoires. Eight and three PCRs were performed for amplification of CDR2 and CDR3 gene segments from each source library, respectively. Products from different libraries were pooled. CDR1 repertoire was amplified from m0 using a degenerate primer covering the whole CDR1. FR3 segment was also obtained from m0 for assembly of the entire VHs. (B) Overlapping PCRs were performed to join CDR1s to CDR2s and FR3 to CDR3s. (C) The final VHs (domain antibodies) were assembled by overlapping PCR.

Figure 3.

Analysis of the library diversity based on the 143 randomly selected clones. (A) Mutation frequency to A, D, S or Y at the indicated positions in CDR1 (B) Gene usage of CDR2s. (C) The number of aa mutations in CDR2s compared to the closest germlines, respectively. (D) CDR3 length distribution.

Figure 4.

Recombinant pairing between CDR2s originated from different groups of germlines and CDR3s with varying length. The 143 sequences were analyzed for the CDR2 gene usage and length of CDR3s and the pairing between them was plotted for each sequence. The number of individual sequences found for each pairing is shown on top right corner.

Figure 5.

Analysis of soluble VH expression. (A) 2 clones each, one with relatively short CDR3 and the other long CDR3, were selected from those with CDR2s derived from 6 different groups of human VH germlines, respectively and expressed. Soluble VHs were purified as described in Materials and Methods and analyzed on SDS-PAGE. (B) 12 clones with CDR3 length ranging from 7 to 24 residues were picked from those with VH4-derived CDR2s and measured for soluble protein expression. M, protein ladder marker; NA, not available.

Before selection						
aa	Frequency (%)					
	27	28	29	30	31	32
	-	F	-	F	-	-
A	11.1		22.2		29.6	22.2
D	33.3		18.5		18.5	40.0
S	18.5		11.1		7.4	11.1
Y	37.0		48.1		44.4	25.9

After selection						
aa	Frequency (%)					
	27	28	29	30	31	32
	-	F	-	F	-	-
A	16.5		11.4		31.6	11.4
D	62.0		53.2		20.3	65.8
S	10.1		13.9		31.6	11.4
Y	11.4		21.5		16.5	11.4

Figure 6.

A, D, S, and Y occurrence in CDR1 before and after selection against SPA. Library was cycled through four rounds of selection against SPA. Clones were randomly picked from the fourth round and analyzed for A, D, S and Y occurrence in CDR1, which was compared to that of clones with VH3-derived CDR2s from the original library.

Table 1

Primers used for library construction

Primer description	Name	Sequence	Target
H1 antisense	H1R	GCG GAC CCA GCT CAT TTC ATA <u>AKM AKM</u> GAA	CDR1
H2 sense	H2F1	<u>AKM</u> GAA <u>AKM</u> AGA GGC TGC ACA GGA GAG GAA ATG AGC TGG GTC CGC CAG GCT CCA <u>GGA CAA</u> <u>SGS CTT GAG TGG</u>	VH1-2 ^a , VH1-3, VH1-8, VH1-18, VH1-45, VH1-46, VH1-58, VH1-69, VH1-C, VH6-1, VH7-* ^b VH2-*
	H2F2	GAA ATG AGC TGG GTC CGC CAG GCT CCA <u>GGG AAG</u> <u>GCC CTG GAG TGG</u>	
	H2F3	GAA ATG AGC TGG GTC CGC CAG GCT CCA <u>GGG AAG</u> <u>GGN CTR GAG TGG</u>	VH1-24, VH1-F, VH3-*, VH4-*, VH5-*
H2 antisense	H2R1	ATT GTC TCT GGA GAT GGT <u>GAC CCT KYC CTG RAA</u> <u>CTY</u>	VH1-*
	H2R2	ATT GTC TCT GGA GAT GGT <u>GAA TCG GCC CTT CAC</u> <u>NGA</u>	VH3-*, VH6-1
	H2R3	ATT GTC TCT GGA GAT GGT <u>GAC TMG ACT CTT GAG</u> <u>GGA</u>	VH2-*, VH4-*
	H2R4	ATT GTC TCT GGA GAT GGT <u>GAC STG GCC TTG GAA</u> <u>GGA</u>	VH5-*
	H2R5	ATT GTC TCT GGA GAT GGT <u>AAA CCG TCC TGT GAA</u> <u>GCC</u>	VH7-*
H3 sense	H3F1	ACC CTG AGA GCC GAG GAC <u>ACR GCY TTR TAT TAC</u> <u>TGT</u>	VH3-9, VH3-20, VH3-43
	H3F2	ACC CTG AGA GCC GAG GAC <u>ACA GCC AYR TAT TAC</u> <u>TGT</u>	VH1-45, VH2-*, VH5-*, VH7-81
	H3F3	ACC CTG AGA GCC GAG GAC <u>ACR GCY GTR TAT TAC</u> <u>TGT</u>	Other than above
H3 antisense	H3R	GTG GCC GGC CTG GCC ACT <u>TGA GGA GAC GGT GAC</u> <u>C</u>	FR4
FR3 sense	FR3F	ACC ATC TCC AGA GAC AAT TCC	FR3
FR3 antisense	FR3R	GTC CTC GGC TCT CAG GGT G	FR3
Extension #1	FR1F	TGG TTT CGC TAC CGT GGC CCA GGC GGC CCA GGT GCA GCT GGT G	FR1
Extension #2	HISR	GTC GCC GTG GTG GTG GTG GTG GTG GCC GGC CTG GCC ACT TG	5 end of H3R primer

The sequences underlined enable amplification or randomization of CDR repertoires.

^a Sub-groups of human antibody heavy chain genes.^b All members in the groups.

Table 2

Primer pairings used for amplification of gene segments

Primer pairings	Products	Targets
FR1F- H1R	FR1-CDR1-FR2 ^a	CDR1
H2F1- H2R1	FR2 ^a -CDR2-FR3 ^a	VH1-* ^b (except VH1-24 ^c , VH1-F)
H2R2		VH6-1
H2R5		VH7-*
H2F2- H2R3		VH2-*
H2F3- H2R1		VH1-24, VH1-F
H2R2		VH3-*
H2R3		VH4-*
H2R4		VH5-*
H3F1- H3R	FR3 ^a -CDR3-FR4	VH3-9, VH3-20, VH3-43
H3F2		VH1-45, VH2-*, VH5-*, VH7-81
H3F3		Other than above
FR3F- FR3R	FR3 ^a	FR3

^aThese products partially cover FRs.^bAll members in the group of VH germlines^cSub-group within the group of VH germlines.

Table 3

CDR diversity of unique phage-displayed VHs selected against B5R after the third round of panning

Clone	CDR1	CDR2		CDR3 length (aa)
		Sequence	Origin	
m301	S-Y-DY--	IYHSGST	4-30-2	17
m302	S-D-AD--	INSSSSYI	3-21	10
m303	D-D-YS--	ISGDGGAT	3-23	10
m304	D-S-YD--	IYYSGST	4-39	8
m305	D-S-YY--	IKQDGSVV	3-7	10
m306	Y-D-YA--	ISYDGSNK	3-30, 3-30-3	10