

Comprehensive Analysis of the Factors Contributing to the Stability and Solubility of Autonomous Human V_H Domains*

Received for publication, October 15, 2007, and in revised form, November 8, 2007 Published, JBC Papers in Press, November 28, 2007, DOI 10.1074/jbc.M708536200

Pierre A. Barthelemy[‡], Helga Raab[§], Brent A. Appleton[‡], Christopher J. Bond[‡], Ping Wu[‡], Christian Wiesmann[‡], and Sachdev S. Sidhu^{‡1}

From the Departments of [‡]Protein Engineering and [§]Protein Chemistry, Genentech, Incorporated, South San Francisco, California 94080

We report a comprehensive analysis of sequence features that allow for the production of autonomous human heavy chain variable (V_H) domains that are stable and soluble in the absence of a light chain partner. Using combinatorial phage-displayed libraries and conventional biophysical methods, we analyzed the entire former light chain interface and the third complementarity determining region (CDR3). Unlike the monomeric variable domains of camelid heavy chain antibodies (V_{H^H} domains), in which autonomous behavior depends on interactions between the hydrophobic former light chain interface and CDR3, we find that the stability of many *in vitro* evolved V_H domains is essentially independent of the CDR3 sequence and instead derives from mutations that increase the hydrophilicity of the former light chain interface by replacing exposed hydrophobic residues with structurally compatible hydrophilic substitutions. The engineered domains can be expressed recombinantly at high yield, are predominantly monomeric at high concentrations, unfold reversibly, and are even more thermostable than typical camelid V_{H^H} domains. Many of the stabilizing mutations are rare in natural V_H and V_{H^H} domains and thus could not be predicted by studying natural sequences and structures. The results demonstrate that autonomous V_H domains with structural properties beyond the scope of natural frameworks can be derived by using non-natural mutations, which differ from those found in camelid V_{H^H} domains. These findings should enable the development of libraries of synthetic V_H domains with CDR3 diversities unconstrained by structural demands.

It was long believed that all antibodies are composed of heavy and light chains and consequently that the simplest antigen-binding unit is the variable fragment (Fv),² a heterodimer of heavy and light chain variable domains (V_H and V_L , respectively) (1). This supposition has held true for most species, but members of the Camelidae family have proven to be notable

exceptions (2, 3). Camelids produce not only conventional antibodies but also “heavy chain antibodies” that are devoid of light chains (4). Consequently, the antigen-binding site of a heavy chain antibody is contained within a monomeric variable domain (V_{H^H} , for heavy chain variable domain of a heavy chain antibody), which has evolved to be autonomously stable in the absence of a light chain partner.

The simplicity of the V_{H^H} domain architecture has attracted interest from protein engineers concerned with developing novel antibodies for biotechnological and therapeutic applications. V_{H^H} domains are well suited for use as modular building blocks that can be linked together to assemble multivalent or bispecific antibodies (5). The third complementarity determining loop (CDR3) of camelid V_{H^H} domains is often unusually long, and it has been suggested that these protruding loops may be especially effective for targeting active site clefts (6) and cryptic viral epitopes (7). Furthermore, because of their small size, V_{H^H} domains are rapidly cleared *in vivo* and exhibit enhanced tumor penetration, and these properties may be useful for the delivery of toxins for therapy or radioisotopes for imaging (8–11). Also, unlike most antibody fragments, thermodenaturation of most V_{H^H} domains is a fully reversible process, making them well suited for applications where transient heating may occur (12). Thus, considerable effort has been expended on elucidating the structural basis for the autonomous stability and solubility of V_{H^H} domains, with the ultimate goal of using these insights to engineer autonomous versions of conventional V_H domains.

For a conventional V_H domain, the major consequence of an absent light chain is the exposure of hydrophobic side chains that are normally buried within the interface. In addition, thermodynamic stability is usually compromised because the stability of most V_H domains is augmented by contacts at the light chain interface (13). The elucidation of a number of camelid V_{H^H} domain structures has provided insights into how these challenges to structural integrity are accommodated (6, 14–19). There are a number of hallmark sequence changes in V_{H^H} domains relative to conventional V_H domains, which function to reduce the hydrophobicity of the former light chain interface. The most common and important changes, dubbed the “ V_{H^H} tetrad,” occur at four positions in the second framework region (2, 3, 20, 21). In most cases, Glu-44 and Arg-45 are found in place of the less hydrophilic Gly-44 and Leu-45 of V_H domains, and hydrophilicity at position 47 is also increased by substitution of Trp with smaller residues. The effect of the fourth change is less obvious from the primary sequence alone,

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The atomic coordinates and structure factors (code 3B9V) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence should be addressed. E-mail: sidhu@gene.com.

² The abbreviations used are: Fv, variable fragment; CDR, complementarity determining region; Fab, antigen binding fragment; V_H , heavy chain variable domain; V_{H^H} , heavy chain variable domain of a heavy chain antibody; V_L , light chain variable domain; WT, wild type; PDB, Protein Data Bank.

Autonomous Human V_H Domains

as it involves replacement of Val-37 by a more hydrophobic Phe (or sometimes Tyr). However, structural analysis reveals that the Phe/Tyr residue at position 37 of V_H H domains nucleates a small hydrophobic core, which also typically involves Tyr-91, Trp-103, the aliphatic portion of Arg-45, and hydrophobic residues in the CDR3 loop. The overall effect of these changes is to increase the hydrophilicity of the former light chain interface, either by direct substitution with more hydrophilic side chains or by sequestration of hydrophobic side chains from solvent by interactions between CDR3 and framework residues (6, 14–16, 19).

Insights gained from the analysis of camelid V_H H domains have been applied to efforts aimed at developing autonomous versions of human V_H domains. Even prior to the elucidation of V_H H domain three-dimensional structures, attempts were made to produce soluble V_H domains by a “camelization” process, which involves transferring elements of the V_H H tetrad to a human framework (22). It was found that the incorporation of three hydrophilic substitutions (G44E/L45R/W47G) improved solubility, but protein yields and thermodynamic stability were severely compromised. Substitution by Ile rather than Gly at position 47 improved yields and stability but reduced solubility somewhat. Nevertheless, the second variant was soluble enough to enable elucidation of the solution structure by NMR, albeit only in the presence of detergent (23, 24). Frameworks of this type were used to construct phage-displayed libraries of random CDR3 sequences that yielded soluble V_H domains with novel antigen specificities (25). The structure of the camelized V_H domain revealed deformations in the framework β -sheet, which are likely responsible for reduced stability and protein yields (23). Subsequent attempts at camelization have also been only partially successful because it has been found that the substitutions of the V_H H tetrad do not completely alleviate aggregation (26–28). Thus, it was concluded that, although the camelization strategy yielded moderately stable autonomous V_H domains, additional factors must be responsible for the stability and solubility of camelid V_H H domains (23).

In retrospect, it appears that at least one of these additional factors is the requirement for a CDR3 loop that can supply stabilizing hydrophobic interactions with the framework. The importance of this additional requirement has been underscored by a study that showed that random CDR3 loops selected for stabilization of a llama V_H H domain contained hydrophobic residues at key positions, and structural analysis showed that a selected CDR3 loop engaged in hydrophobic packing interactions that were similar to those of natural camelid CDR3 loops (29). Taken together, these results demonstrate a structural role for CDR3 in the stability and solubility of V_H H domains, which places additional constraints on the CDR3 sequence. In practical terms, these constraints must be taken into account in the design of camelization strategies, and this limits the diversity of CDR3 sequences that are compatible with the camelid V_H H framework.

Despite the importance of the V_H H tetrad for stabilization of camelid V_H H domains, several studies have discovered autonomous V_H domains, from camelids and other species, which do not contain these hallmark sequences. Indeed, even prior to the discovery of camelid heavy chain antibodies, it was found that

certain mouse V_H domains can function autonomously (30). More recently, an unusual mouse V_H domain containing a hydrophilic substitution at the former light chain interface (G44K) has been shown to exist as a stable monomer in solution (31). Surprisingly, a phage-displayed library derived from llama yielded seemingly conventional V_H domains, which lacked elements of the V_H H tetrad but nonetheless could be purified as stable and highly soluble monomers (32). Furthermore, phage-displayed selections have shown that human V_H domain libraries contain a substantial number of autonomous V_H domains with framework sequences typical of conventional antibodies (33, 34). Taken together, these studies suggest that most vertebrate species likely contain some V_H domains that can function autonomously. Although the high sequence diversity among these unusual subpopulations has made it difficult to discern general rules for autonomous V_H domain stabilization (32–34), two structural studies have provided clues to the mechanisms whereby autonomous V_H domains lacking the V_H H tetrad are stabilized and solubilized (35, 36).

The solution structure of a llama V_H domain (BrucD4-4) is similar to those of V_H H domains, as the CDR3 loop folds over Val-37 and thus reduces the hydrophobicity of the former light chain interface (36). However, this interface may be transient because of the conformational flexibility of CDR3, and overall, the hydrophobicity of the BrucD4-4 surface lies between that of conventional V_H domains and autonomous V_H H domains. The crystal structure of an anti-lysozyme human V_H domain (HEL4) reveals that, although the domain is monomeric in solution, the protein crystallizes as a dimer (35). HEL4 was derived from a germ line framework by mutations restricted to the three CDR loops, and the structure revealed that a key change in CDR1 is partially responsible for the favorable solution properties of HEL4. It appears that the substitution of His-35 by Gly produces a cavity that serves to bury the side chain of Trp-47, which is flipped relative to its orientation in structures of V_H/V_L pairs, and this reorganization decreases the hydrophobicity of the former light chain interface. However, because the former light chain interface and CDR3 are involved in crystal packing interactions, it is unclear from the structure whether Trp-47 assumes the same conformation in solution and also whether or not the CDR3 loop plays a stabilizing role similar to that observed in V_H H domains.

All of the studies reported to date have relied on insights gained from natural antibodies to elucidate features contributing to the autonomous nature of camelid V_H H domains and some V_H domains. Camelid V_H H domains have been studied (4, 6, 14–19), features of V_H H domains have been engineered into V_H frameworks (22–25, 37), or subpopulations of autonomous V_H domains have been isolated from libraries of natural frameworks (32–36). These studies have revealed that, in general, the stability and solubility of these domains are dependent on CDR3, and this places significant constraints on the types of CDR3 loops that are compatible with autonomous structural integrity. This limitation poses considerable problems for the engineering of autonomous V_H domains with novel functions, as CDR3 is the major contributor to most antigen-binding sites, and it has been well established that the success of *in vitro* antibody libraries depends on the use of highly diverse CDR3

loops (38–42). Thus, for the purposes of antibody engineering, it would be most advantageous to use V_H domain scaffolds that exhibit the autonomous behavior of camelid V_{HH} domains, but they do so without reliance on the CDR3 loop.

Here we have gone beyond the limits of natural antibody diversity to explore the issue of autonomous V_H domains from a strictly biophysical perspective. Using combinatorial selections followed up by conventional biophysical methods, we scanned the entire former light chain interface and CDR3 to assess the compatibility of different framework and CDR3 combinations with autonomous stability. With this comprehensive approach, we confirm that the evolution of autonomous V_H domains from V_H/V_L pairs depends on mutations that are well tolerated structurally and increase the hydrophilicity of the former light chain interface. However, we find that many sequences that improve the stability and solubility of autonomous V_H domains are rare in natural V_{HH} and V_H domains. Furthermore, we isolated numerous autonomous V_H domains whose structural stability does not depend on the CDR3 loop. Thus, our findings demonstrate that autonomous V_H domains with structural properties beyond the scope of natural frameworks can be derived by the incorporation of substitutions that are rare in nature but, nonetheless, are structurally compatible with the V_H domain fold. These results open the door to the development of libraries of synthetic V_H domains with CDR3 diversities unconstrained by structural demands.

EXPERIMENTAL PROCEDURES

Library Construction and Analysis—For the phage display of VH-4D5, a phagemid (pPAB43778) was constructed from a previously described phagemid (pS1602) (43) by inserting a DNA fragment encoding VH-4D5 in place of the region encoding human growth hormone. Phage-displayed libraries were constructed using methods described previously (29, 44, 45). For library A, NNS degenerate codons were used ($N = A/G/C/T$ and $S = G/C$). For library B, degenerate mutagenic oligonucleotides contained ~70% of the wild-type nucleotide and 10% each of the three other nucleotides at each randomized position. For shotgun alanine scanning, codons were designed as described (46). For quantitative saturation scanning, NNK codons were used ($K = G/T$), and two separate libraries were constructed; one library randomized positions 35, 39, 45, and 50, and the other randomized positions 37, 44, 47, 91, and 105.

Library phage pools were cycled through rounds of binding selection against protein A (Sigma) immobilized on NUNC 96-well Maxisorp immunoplates, as described (29, 44). Bound phage were eluted with 0.1 M HCl, and the eluant was neutralized with 1.0 M Tris base. Eluted phage were amplified in *Escherichia coli* XL1-blue (Stratagene) with the addition of M13-KO7 helper phage (New England Biolabs) and used for further rounds of selection. Following selection, individual clones were subjected to DNA sequence analysis, as described (29).

For shotgun alanine scanning, quantitative saturation scanning, and the analysis of library B, the sequences were analyzed with the program SGCOUNT as described (46–48). SGCOUNT aligned the sequences, tabulated the occurrence of each amino acid at each position, and removed any duplicate clones with identical sequences at all mutated positions. For

alanine scanning, ~100 clones were sequenced for each V_H domain. For the analysis of library B, ~400 clones were sequenced after selection for binding to protein A, and ~200 clones were sequenced from the naive library to determine the nucleotide distribution prior to selection. For saturation scanning, ~200 clones were sequenced from each library, and the data were normalized for codon bias in the NNK degenerate codon (e.g. the NNK codon contains three unique codons for Arg, and thus, the occurrence of Arg was divided by 3).

Protein Purification—To purify protein for biophysical analysis, phage display vectors were modified by the insertion of an amber stop codon between the sequence encoding the V_H domain and the phage coat protein. V_H domains were secreted in *E. coli* BL21 cells (Stratagene) by isopropyl 1-thio- β -D-galactopyranoside induction (0.4 mM) at 30 °C for 3 h. Frozen cell pellets were resuspended in 25 mM Tris, 25 mM NaCl, 5 mM EDTA, pH 7.1, and cells were lysed with an Ultra-Turrax T8 homogenizer (IKA Labortechnik, Staufen, Germany) and a M-110F Microfluidizer® processor (Microfluidics, MA). The lysate was centrifuged, and the supernatant was passed through a 0.2- μ m filter and loaded onto a gravity flow protein A-Sepharose column (GE Healthcare). The column was washed with 10 bed volumes of 10 mM Tris, 1.0 mM EDTA, pH 8.0, and the protein was eluted with 0.1 M glycine, pH 3.0. The eluant was neutralized with 1.0 M Tris, pH 8.0, and protein concentration was determined by the method of Bradford (49).

Analytical Gel Filtration and Light Scattering Analysis—Purified V_H domains were analyzed on a Superdex-75 column equilibrated with phosphate-buffered saline, 500 mM NaCl, pH 7.2, using an Agilent 1100 series high pressure liquid chromatography system (Agilent, Palo Alto, CA) in line with a Wyatt MiniDawn multiangle light scattering detector (Wyatt Technology, Santa Barbara, CA). The domains were injected at a concentration of ~70 μ M (1.0 mg/ml) in a volume of 100 μ l, and the flow rate was 0.5 ml/min. Concentration measurements were made using an on-line Wyatt OPTILAB DSP interferometric refractometer (Wyatt Technology). Astra software (Wyatt Technology) was used for light scattering data acquisition and processing. The light scattering unit and the refractometer were calibrated according to the manufacturer's instructions. A value of 0.185 ml/g was assumed for the dn/dc ratio of the protein. The detector responses were normalized by measuring the signal from monomeric bovine serum albumin (Sigma). The temperatures of the light scattering unit and the refractometer were maintained at 25 or 35 °C, respectively. The column and all external connections were at ambient temperature.

Temperature-induced Denaturation—Thermal denaturation of V_H domain protein was monitored at 207 nm with a J-810 spectrometer (Jasco, Easton, MD), using a 1-cm path length CD cuvette containing protein samples (10 μ M) in phosphate-buffered saline, pH 7.2. The temperature was increased from 25 to 85 °C (or higher, when necessary to achieve complete unfolding) in steps of 2 °C. The reversibility of the temperature-induced denaturation was checked by cooling the sample to 25 °C and repeating the heating program (12). The unfolding curves were assumed to be two-state transitions (50). The fraction folded (α) was calculated using the following equa-

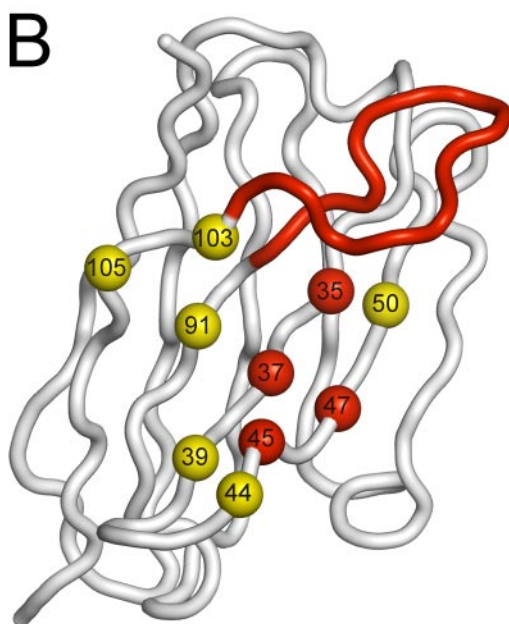
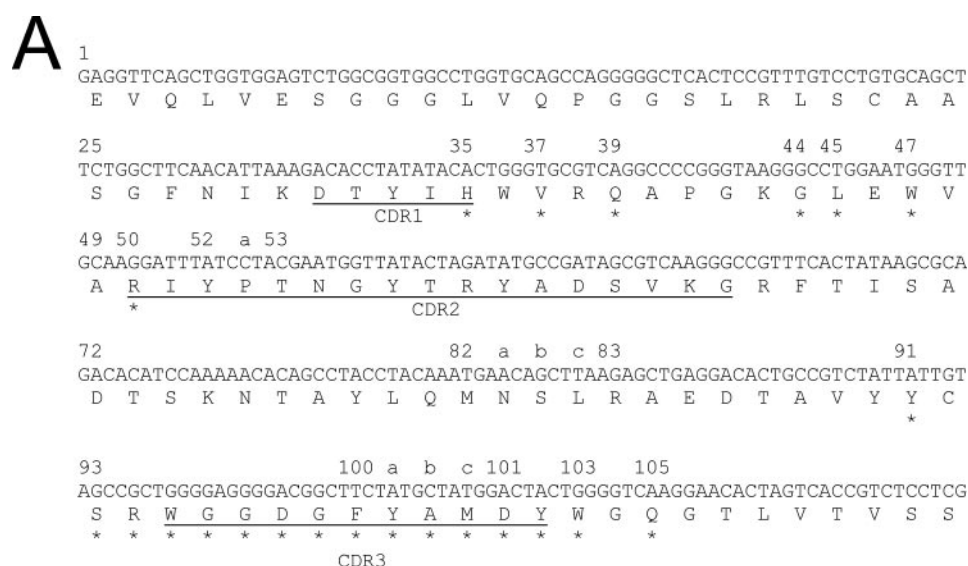


FIGURE 1. Design of V_H domain libraries. *A*, gene encoding for VH-4D5. The DNA sequence of the coding strand is shown along with the protein translation below. The CDRs are *underlined*, and positions that were randomized in the selection experiments are indicated by *asterisks*. CDR definitions and residue numbering are according to the Kabat nomenclature (74). *B*, three-dimensional view of the V_H domain library designs. The VH-4D5 main chain is shown as a tube. CDR3 and the two preceding residues that were randomized in library A are colored *red* (positions 93–102). Other positions chosen for randomization are *numbered* and shown as *red spheres* (randomized in libraries A and B) or *yellow spheres* (randomized in library B only). The structure was rendered with the program PyMOL using coordinates for the crystal structure of Fv-4D5 (PDB code 1FVC).

tion: $\alpha = (\theta_T - \theta_F) / (\theta_U - \theta_F)$, where θ_T is the observed ellipticity at any temperature; θ_F is the ellipticity of the fully folded form of the protein, and θ_U is the ellipticity of the fully unfolded form of the protein (50). The raw ellipticity data of the folded and unfolded states show no temperature dependence outside of the transition zone, and therefore, θ_F and θ_U were considered to be the ellipticity value observed at 25 °C and the lowest ellipticity value observed, respectively. The variation of α was plotted *versus* the temperature to obtain the melting curve for each V_H domain. The melting temperature (T_m) is defined as the temperature at which $\alpha = 0.5$. The fraction of refolded protein recovered following thermal denaturation was estimated as

the α value for the sample cooled down to 25 °C after heating to a temperature that induced complete unfolding.

Crystallization, Structure Determination, and Refinement—For crystallization, VH-B1a was purified by protein A-Sepharose chromatography, as described above, and the eluted protein (10 mg) was loaded on a Superdex™ HiLoad™ 16/60 gel filtration column (Amersham Biosciences Bioscience) with 20 mM Tris, 500 mM NaCl, pH 7.5, as mobile phase at 0.5 ml/min. The protein was concentrated to 10 mg/ml, and crystals were grown at 19 °C in sitting drops using the vapor diffusion method with 2.0- μ l drops containing equal volumes of protein solution and crystallization buffer (1.1 M sodium malonate, pH 7.0, 0.1 M HEPES, pH 7.0, 0.5% (v/v) Jeffamine ED-2001, pH 7.0). Before data collection, crystals were dipped briefly into 2.0 M sodium malonate, pH 7.0, and flash-frozen in liquid nitrogen. Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (Stanford University) and processed using the programs Denzo and Scalepack (51).

The structure was solved by molecular replacement using the program Phaser (52) and the coordinates of the Herceptin Fab (53) (PDB code 1N8Z). The structure was refined using the program REFMAC (54). The model was manually adjusted using the program Coot (55). There was partial density for one or two Jeffamine molecules in the electron density maps, but overall quality was poor in comparison to the protein and was there-

fore not modeled. The coordinates and structure factors for the VH-B1a structure have been deposited in the RCSB Protein Data Bank (PDB code 3B9V).

RESULTS

Strategy for in Vitro Evolution and Analysis of Autonomous V_H Domains—To explore the requirements for an autonomous V_H domain structure, we chose to study the V_H domain of the humanized monoclonal antibody 4D5 (56) (Fig. 1), which binds to the epidermal growth factor receptor family member ErbB2 (57–59) and has been approved for cancer therapy (56, 60). Antibody V_H segments have been classified into three sub-

groups on the basis of sequence homology (61), and the 4D5 V_H domain (VH-4D5) and all camelid V_HH domains belong to subgroup III (20, 21). The 4D5 antibody has been extensively characterized both functionally and structurally, and structures of antibody fragments have been solved both uncomplexed (62) and complexed with antigen (53). Furthermore, the 4D5 antigen-binding fragment (42, 63) (Fab-4D5), single chain variable fragment (62, 64) (scFv-4D5), and V_H domain (65, 66) (VH-4D5) can be produced in *E. coli*, and the Fab and scFv forms have been used as scaffolds for the construction of phage-displayed antibody libraries (42, 64). Importantly, VH-4D5 binds to protein A through interactions that require a native tertiary structure and are mediated entirely by framework regions on the face opposite the light chain interface (66–69), and this property can be used for facile selection and purification of correctly folded V_H domains (29).

Using VH-4D5 as a starting point, we designed experiments to comprehensively explore CDR3 and the framework for features compatible with autonomous V_H domains. We first used phage display to assess the effects of many different combinatorial mutations, as this method allows for the rapid sorting of extremely diverse protein populations. To enrich for stable phage-displayed V_H domains, we took advantage of the interaction with protein A, as it has been shown that binding of a phage-displayed antibody fragment to this ligand is highly correlated with the intrinsic stability of the free protein (29, 47, 70). Phage-displayed libraries were used to not only select autonomous V_H domains from naive libraries but also to rapidly analyze particular variants in detail using quantitative saturation scanning (48) and shotgun alanine-scanning methods (46).

The protein A selection strategy allowed us to rapidly enrich libraries for those members capable of adapting a stable fold in a phage-displayed format, but our ultimate goal was to obtain autonomous V_H domains that would be stable and monomeric at high concentrations. Thus, we next purified selected V_H domains and screened the proteins individually for favorable behavior using biophysical methods that have been applied previously to the study of V_HH and V_H domains (12, 29). These screens assessed whether each V_H domain 1) could be purified in high yield in a correctly folded form, 2) existed as a monomer at high concentrations as evidenced by gel filtration and light scattering analysis, and 3) exhibited reversible folding behavior and high thermostability in temperature-induced denaturation experiments monitored by CD.

First Generation Library for Selection of Autonomous V_H Domains—To select for mutations that stabilize the VH-4D5 domain, we designed a phage-displayed library that targeted for randomization a set of residues that have been implicated in the stabilization of camelid V_HH domains and an autonomous human V_H domain (Fig. 1). Because camelid V_HH domains are stabilized by interactions between CDR3 and the former light chain interface (6, 14–16, 19), we replaced residues 93–102 with random loops of all possible lengths varying from 7 to 17 residues. In addition, we randomized three framework positions that are involved in the V_HH tetrad and are occupied by hydrophobic residues in the VH-4D5 sequence (Val-37, Leu-45, and Trp-47). We did not randomize the fourth position of the V_HH tetrad because VH-4D5 contains a hydrophilic Gly-44

at this position. We also targeted position 35 in CDR1, as a Gly residue at this position has been implicated in the stabilization of the human V_H domain HEL4 (35). We used degenerate codons that encode for all 20 natural amino acids in order to sample all possible sequence combinations and to allow all the variants to compete among each other in a single pool.

The constructed library (library A) contained a total diversity of 2×10^{10} unique members, and the phage pool was cycled through six rounds of selection for binding to protein A. Sequencing of 57 clones revealed 25 unique sequences, which were aligned and inspected for conserved sequence motifs (Fig. 2A). Among the CDR3 loops, there was no obvious consensus in terms of either length or sequence, aside from a modest preference for small residues at positions 93 and 94. There is no bias in favor of hydrophobic sequences, and this contrasts with the results of a previous study in which the CDR3 loop was randomized in the context of a fixed V_HH framework and revealed significant bias for hydrophobic residues that were shown to be involved in stabilizing interactions with the framework (29, 47). At positions 37 and 45, although the sequences were quite variable, the general hydrophobic character of the wild-type residues was conserved. Interestingly, most of the sequences at position 35 contain small residues, which differ markedly from the wild-type His and instead resemble the Gly residue of the HEL4 V_H domain (35). Position 47 is occupied by a hydrophobic residue, a charged residue, or a small residue. The positions involved in the V_HH tetrad are not biased in favor of camelid-like substitutions, and taken together with the lack of a hydrophobic bias in the CDR3 sequences, these results suggest that most of the selected domains are not stabilized by the CDR3/framework interactions typical of V_HH domains. Instead, it appears that increased hydrophilicity may be achieved by interactions between small residues at position 35 and hydrophobic residues at position 47, as observed in the structure of HEL4 (35) or, alternatively, by replacement of the hydrophobic Trp-47 by a charged or small residue.

Biophysical Analysis of Autonomous V_H Domains—We purified 6 of the 25 selected V_H domains from *E. coli* using protein A affinity chromatography and compared the protein yields to that of VH-4D5 (Fig. 2A). The protein yields were all significantly higher than that of VH-4D5, and the proteins were further characterized by gel filtration and temperature-induced denaturation.

Analytical gel filtration was used to profile the quaternary structure of the V_H domains (Fig. 3), and for those which exhibited distinct peaks, the molecular masses were estimated by light scattering (Table 1). VH-4D5 and two of the six V_H domain variants (VH-A2 and VH-A6) eluted in several overlapping peaks and light scattering indicated significant aggregation (Fig. 3A). In contrast, the other four V_H domains (VH-A1, VH-A3, VH-A4, and VH-A5) eluted mainly as single peaks with elution times close to that expected for a monomer (Fig. 3B), and weight average molar masses estimated by light scattering were in good agreement with the formula molecular masses. In addition to the major monomeric peak (peak 1), there were two minor peaks. One of the minor peaks (peak 1a) eluted close to peak 1 and exhibited a weight average molecular mass that also corresponded to that of a monomer, and we speculate that this

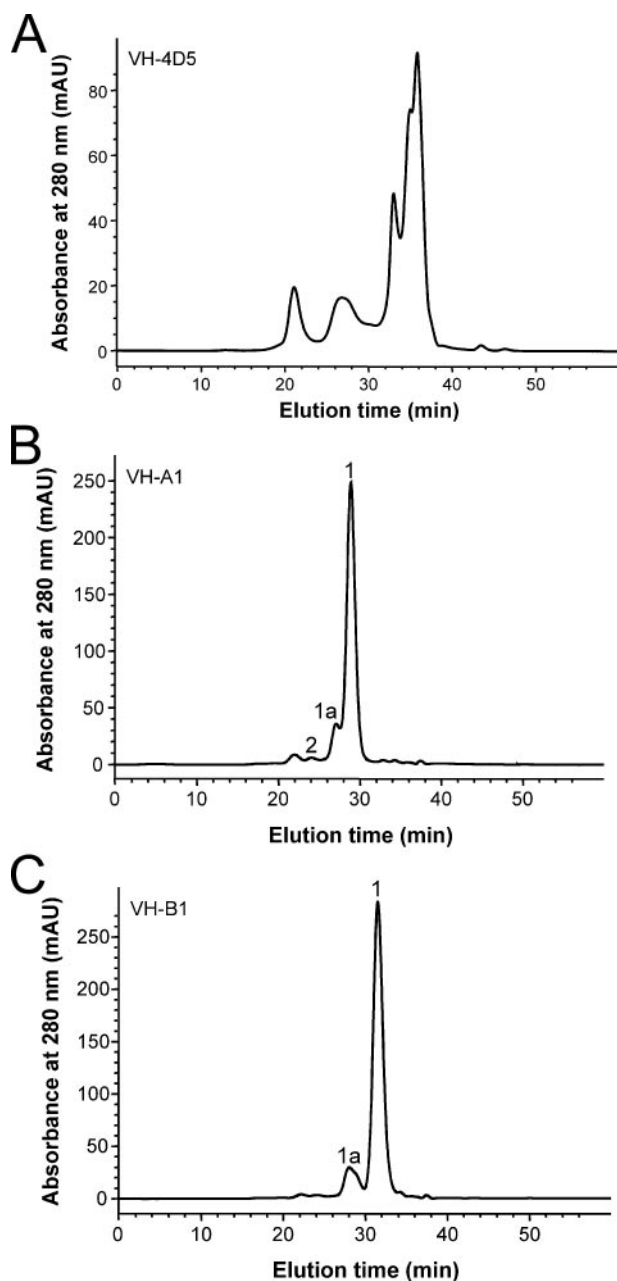


FIGURE 3. **Analytical gel filtration of V_H domains.** Chromatograms are shown for 70 μ M samples of VH-4D5 (A), VH-A1 (B), and VH-B1 domains (C) applied on a Superdex-75 column. Peaks are numbered as described in the text and Table 1. See "Experimental Procedures" for further details.

peak may represent a monomeric form of the protein lacking a disulfide bond. The second minor peak (peak 2) exhibited a weight average molecular mass that is in good agreement with the calculated mass of a dimer. The monomer and dimer peaks were clearly separable, and the two forms did not interconvert when the peaks were purified and reinjected on the gel filtration column (data not shown). Thus, it is unlikely that dimerization is caused by transient interactions involving the former light

chain interface, and instead, we speculate that the dimeric form may result from the swapping of β -strands between two monomers, as has been observed previously for a crystallized V_HH domain (71). By comparing the areas under peaks 1 and 2 to the total peak area, we could estimate the fraction of the total injected protein sample that eluted as a monomer (peak 1) or dimer (peak 2), respectively, and for each of the four nonaggregating domains, the monomeric fraction exceeded 85% (Table 1).

We studied the temperature-induced denaturation properties of the proteins by CD to determine the temperature of unfolding transition (T_m) and to ascertain whether the folding pathways were reversible. In agreement with previous studies (72), Fab-4D5 exhibited high thermostability ($T_m = 83^\circ\text{C}$), but the unfolding was irreversible (Fig. 4A). The unfolding of VH-4D5 was also irreversible, and thermostability was reduced substantially ($T_m = 58^\circ\text{C}$), relative to that of Fab-4D5 (Fig. 4B). Notably, all six V_H domain variants exhibited significant improvements in thermostability, relative to that of VH-4D5, and the melting temperatures of the most stable variants approached that of Fab-4D5 (Table 1). Furthermore, the four variants that behaved as monomers by gel filtration exhibited essentially reversible folding behavior (Fig. 4C).

Taken together, these results show that the phage display selection for binding to protein A selected for domains with improved thermostability, and also enriched for domains with a reduced tendency to aggregate, as evidenced by reversible folding behavior and monomeric gel filtration profiles. Thus, these results validate the use of the protein A selection as a rapid filter that enriches for autonomous V_H domains. By following up with biophysical screening of selected variants, we were able to identify several V_H domains that are highly stable and predominantly monomeric. It appears that monomeric quaternary structure is strongly correlated with reversible folding but less so with thermodynamic stability, as the nonmonomeric variants exhibited high thermostability but only partially reversible folding, whereas the monomeric variants exhibited both high thermostability and reversible folding.

The four monomeric V_H domains contain small residues at position 35, but differ significantly in the nature of the residue at position 47 (Fig. 2A). It is possible that the variants VH-A1 and VH-A3 are stabilized by structural mechanisms similar to those observed for the HEL4 V_H domain (35), as all three proteins contain a small residue at position 35 and a hydrophobic residue at position 47. However, this is clearly not the only structural solution involving these positions, as the other two variants (VH-A4 and VH-A5) contain Ser at position 35 and Ser or Glu at position 47.

Second Generation Library for the Selection of Autonomous V_H Domains—We next investigated whether the stability and quaternary structure of autonomous V_H domains could be improved by targeting additional positions for randomization. For this purpose, we chose VH-A1 as the template, as this

FIGURE 2. **Sequences of autonomous V_H domains.** The V_H domains were isolated from library A (A) or library B (B). A, dashes indicate gaps in the alignment of CDR3, and the numbering is shown for VH-4D5. Asterisks indicate positions that were not randomized. The sequences are colored as follows: red, small residues (G, A, S, and T); black, hydrophobic residues (F, Y, W, L, V, I, M, C, and P); blue, hydrophilic residues (D, E, Q, N, R, K, and H). Gray shading denotes sequences that are identical to VH-4D5 (A) or VH-A1 (B). The protein yields from *E. coli* cultures are shown to the right.

TABLE 1
Biophysical analysis of V_H domains

V_H	Analytical gel filtration							Temperature-induced denaturation	
	Molecular mass ^a		Fraction of total peak area		Elution time (min)		T_m	Fraction refolded ^b (α)	
	Formula	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1			Peak 2
		<i>kDa</i>				<i>min</i>		$^{\circ}\text{C}$	
4D5	14.4	NA ^c	NA	0.52	0.17	NA	NA	58	0.29
A1	14.0	14.6	25.0	0.85	0.02	28.9	24.0	73	0.98
A2	14.0	NA	NA	0.60	0.02	NA	NA	75	0.70
A3	13.7	14.2	29.0	0.87	0.03	29.6	22.0	68	0.89
A4	13.7	14.2	23.4	0.88	0.01	28.3	22.1	65	0.91
A5	14.0	14.5	29.1	0.92	0.01	29.3	22.1	68	0.93
A6	13.7	NA	NA	0.72	0.08	NA	NA	72	0.77
B1	13.8	12.9		0.84		31.5		73	0.94

^a The weight average molar masses of peaks 1 and 2 (Fig. 3) were determined by light scattering.

^b The fraction refolded was defined as the fraction folded value (α) for the protein at 25 $^{\circ}\text{C}$, following heating to a temperature that resulted in complete unfolding.

^c NA indicates that the gel filtration data could not be analyzed because distinct peaks could not be discerned.

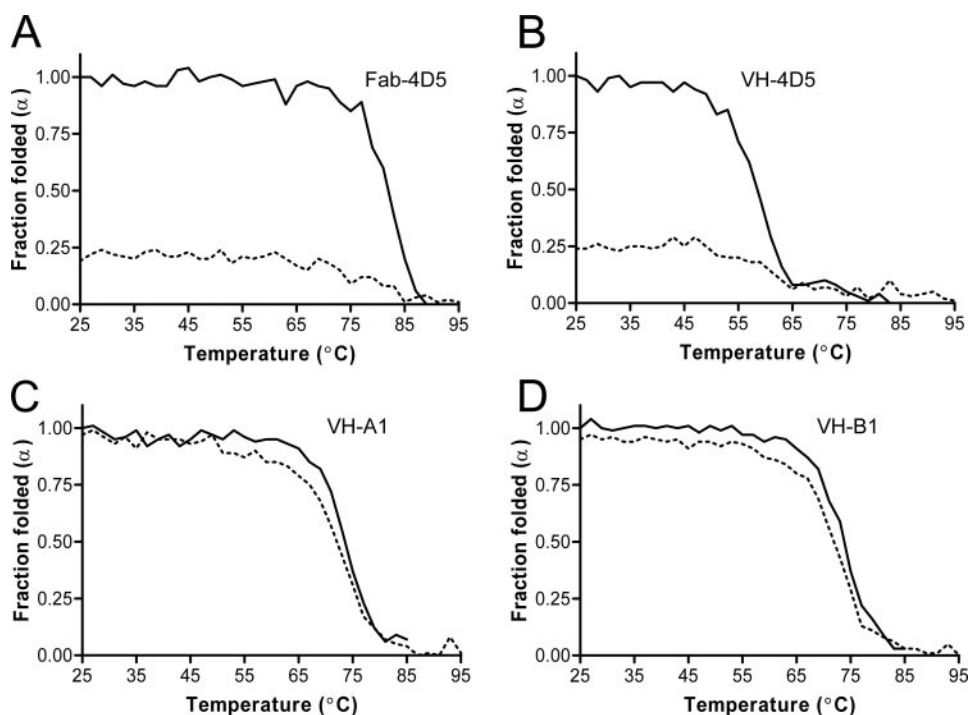


FIGURE 4. Temperature-induced denaturation. Thermal denaturation of 10 μM samples of Fab-4D5 (A), VH-4D5 (B), VH-A1 (C), and VH-B1 (D) was monitored by CD at 207 nm. The sample was heated from 25 $^{\circ}\text{C}$ to $>80^{\circ}\text{C}$ (solid line), cooled to 25 $^{\circ}\text{C}$, and reheated (dashed line). See "Experimental Procedures" for further details.

domain exhibits desirable characteristics in terms of protein yield, quaternary structure, thermostability, and folding behavior. We examined the structure of the Fv-4D5 heterodimer (64) and identified all residues involved in the light chain interface, which we defined as residues that lose greater than 20 \AA^2 of solvent-accessible surface area because of contacts with the light chain. This group consisted of nine positions, including three that were randomized in library A (37, 45, and 47) and six others (39, 44, 50, 91, 103, and 105) (Fig. 1B). Because the residues at these positions are likely to become more solvent-accessible in the absence of the V_L domain, we reasoned that mutations at these sites may improve the stability and/or solubility of autonomous V_H domains. We designed a second generation library (library B) in which these nine positions and position 35 were simultaneously targeted for diversification.

compared these to the observed frequencies among the selected clones (Fig. 5). In this way, we identified statistically significant deviations indicative of positive selective pressure for certain amino acid types at particular positions. Outside of the CDR3 loop, the parental VH-A1 sequence was significantly enriched at most positions. Most notably, Gly was strongly conserved at position 35, suggesting that the mutation H35G contributes to the stabilization of the autonomous V_H domain relative to VH-4D5. However, the parental Trp residue was depleted in favor of Arg at position 103, suggesting that replacement of Trp-103 with Arg may further improve stability.

Biophysical Analysis of a Second Generation V_H Domain—We measured the protein yields for 40 V_H domains selected from the second generation library (Fig. 2B). Yields for most domains were significantly greater than the yield for VH-4D5, but only

CDR3 was also targeted, but the length was fixed as that of the CDR3 of VH-A1.

Because we targeted a large number of positions and were attempting to fine-tune a V_H domain that already exhibited many of the characteristics we desired, we adapted a "soft randomization" approach. At each randomized position, library B was designed to contain roughly 50% wild-type sequence and 50% random sampling of other amino acids. Library B contained a total diversity of 7.5×10^9 unique members and was cycled through six rounds of selection for binding to protein A.

Sequence alignment of 393 clones revealed 384 unique sequences, which were aligned to determine the distribution of amino acids at each mutagenized position. We also calculated the amino acid frequencies at each position within the naive library (prior to selection for binding to protein A), and we compared

	W	F	Y	M	L	I	V	A	G	S	T	R	K	H	N	Q	D	E	P	C
G35	0.0	0.0	0.0	0.0	0.0	0.0	0.3	2.6	93	1.3	0.0	0.0	0.0	0.0	0.0	0.0	2.9	0.0	0.0	0.3
	0.5	1.2	0.8	0.1	1.2	0.8	8.4	6.4	52	5.7	0.7	7.2	0.1	0.6	0.5	0.1	4.7	0.8	0.8	7.2
V37	0.3	2.6	0.3	0.5	2.3	7.8	77	4.7	1.6	0.3	0.8	0.3	0.5	0.0	0.3	0.0	0.0	0.5	0.0	0.3
	0.1	7.5	0.9	0.3	7.9	5.3	53	5.9	6.3	1.5	0.6	0.9	0.1	0.7	0.6	0.1	5.3	0.8	0.7	0.9
Q39	0.0	0.0	0.8	0.3	1.6	0.0	0.0	0.5	0.8	0.8	0.3	11	8.3	6.0	1.8	64	0.0	2.9	1.3	0.0
	0.8	0.3	1.8	0.7	10.7	0.3	0.8	0.5	0.5	1.2	0.7	7.1	4.3	10.5	1.1	41	0.9	3.3	6.4	0.2
G44	0.0	0.3	0.0	0.0	0.0	0.3	1.6	4.9	71	4.7	0.5	10	0.3	0.8	0.0	0.3	2.3	1.6	1.0	0.3
	0.5	1.2	0.8	0.1	1.2	0.8	8.4	6.4	52	5.7	0.7	7.2	0.1	0.6	0.5	0.1	4.7	0.8	0.8	7.2
Y45	0.3	9.6	60	0.3	4.2	0.3	1.6	0.8	0.8	7.8	0.8	0.8	1.0	2.3	1.3	1.3	2.9	1.6	2.1	0.3
	0.4	8.2	45	0.1	2.4	1.0	1.1	0.8	0.8	7.1	0.7	0.8	0.8	5.0	5.1	0.8	5.3	0.8	0.7	5.7
W47	82	1.8	1.3	0.3	7.0	0.0	0.5	0.3	0.8	2.1	0.8	1.6	0.0	0.3	0.0	0.3	0.0	1.0	0.3	0.3
	38	1.8	1.2	0.7	7.9	0.3	1.0	0.8	6.3	7.9	0.8	11	0.5	0.1	0.1	0.5	0.1	0.5	0.7	11
R50	5.2	0.0	0.0	12.8	2.6	3.4	1.3	1.3	5.2	7.6	4.9	47	1.8	0.0	0.3	0.5	0.3	0.0	0.0	6.0
	6.6	0.3	0.2	5.8	2.2	2.3	1.0	0.8	6.4	11	6.2	46	4.2	0.1	1.1	0.5	0.1	0.5	0.8	1.9
Y91	0.3	6.8	69	0.0	0.8	1.0	0.5	1.0	0.0	7.3	0.3	1.0	0.8	7.3	2.3	1.0	0.0	0.3	0.0	0.3
	0.6	7.9	43	0.1	2.7	1.0	1.1	0.8	0.8	7.1	0.7	0.9	1.0	4.8	4.9	1.0	5.1	1.1	0.7	5.4
R93	2.6	0.0	0.3	0.8	0.5	2.9	1.3	2.9	6.3	18	5.7	51	4.7	0.5	0.0	0.8	0.0	0.5	0.0	1.3
	0.8	0.3	0.2	0.7	2.2	7.4	1.0	0.8	6.4	12	6.2	45	4.1	0.1	1.1	0.5	0.1	0.5	0.8	2.0
T94	0.0	0.3	0.3	4.4	1.0	2.3	1.6	8.9	0.5	9.9	65	1.0	1.6	0.0	0.8	0.3	0.0	0.0	1.8	0.0
	0.6	0.3	0.2	6.4	2.4	2.5	1.1	6.5	0.5	10	51	3.8	4.3	0.1	1.1	0.5	0.1	0.6	6.4	0.2
F95	0.3	65	6.8	0.3	13	4.2	3.4	0.0	0.0	5.2	0.0	0.8	0.0	0.3	0.0	0.3	0.0	0.0	0.8	0.0
	0.6	46	5.2	0.6	16	5.8	6.5	0.7	0.8	6.8	0.7	0.9	0.1	0.6	0.6	0.1	0.6	0.1	0.7	5.4
T96	0.0	0.0	0.0	0.3	2.1	3.6	0.8	7.6	3.4	14	48	5.7	4.9	0.3	2.3	1.3	1.0	1.0	4.4	0.0
	0.1	0.4	0.2	0.8	2.4	8.1	1.1	6.5	0.5	10	51	3.8	4.3	0.1	1.2	0.5	0.1	0.5	6.4	0.2
T97	0.5	0.3	0.3	3.6	0.8	0.8	0.0	5.7	1.3	11	45	7.3	6.3	0.0	3.1	1.6	0.0	0.8	12	0.0
	0.6	0.3	0.2	6.4	2.4	2.5	1.1	6.5	0.5	10	51	3.8	4.3	0.1	1.1	0.5	0.1	0.6	6.4	0.2
N98	0.0	0.3	4.2	0.5	0.8	2.6	0.3	0.8	2.1	8.1	5.7	1.0	7.3	6.0	50	1.6	6.3	1.0	1.3	0.0
	0.1	1.4	7.4	0.8	1.4	8.2	1.1	0.8	0.8	6.2	6.0	1.8	8.4	5.0	40	1.0	5.1	1.1	0.8	0.9
S99	0.3	6.8	6.0	0.0	1.8	0.3	0.8	5.5	0.5	59	5.7	0.8	0.3	0.8	1.6	0.0	0.5	0.0	2.9	6.3
	0.3	8.2	5.0	0.1	2.4	1.0	1.1	6.5	0.5	55	6.3	0.6	0.1	0.6	0.6	0.1	0.6	0.1	6.1	3.8
K100	0.3	0.8	1.8	1.6	1.3	2.3	1.0	0.5	0.5	1.3	6.0	11	56	0.3	5.2	5.2	0.3	2.9	1.3	0.0
	0.8	0.3	1.8	6.4	2.4	2.5	1.1	0.8	0.8	2.4	6.0	5.7	38	1.2	10	4.8	1.3	4.9	0.8	0.2
K101	0.0	0.3	2.1	0.0	2.6	15.4	0.8	1.0	0.8	2.1	3.6	7.6	41	1.3	11	5.7	0.8	1.6	1.6	0.3
	0.1	0.4	1.9	0.8	2.4	8.2	1.1	0.8	0.8	2.4	6.0	5.6	38	1.3	10	4.7	1.3	4.8	0.8	0.2
A102	0.0	2.6	1.3	0.0	2.1	1.3	14	55	2.6	7.8	7.3	0.3	0.0	0.5	0.5	0.0	0.5	0.0	2.9	0.5
	0.1	1.2	0.8	0.1	1.4	0.9	9.1	53	4.3	8.9	5.6	0.6	0.1	0.6	0.5	0.1	4.7	1.0	6.6	0.6
W103	43	0.3	2.1	1.8	8.6	0.8	0.5	0.0	2.9	9.1	0.8	24	0.5	0.0	0.0	2.9	0.0	1.0	0.3	2.1
	38	1.8	1.2	0.7	7.9	0.3	1.0	0.8	6.3	7.9	0.8	11	0.5	0.1	0.1	0.5	0.1	0.5	0.7	11
Q105	0.0	0.0	0.3	0.0	6.3	1.0	0.8	0.8	0.0	1.3	2.1	9.1	9.9	12	0.5	46	0.5	1.3	8.6	0.0
	0.1	0.4	1.9	0.1	10.7	0.9	0.8	0.5	0.5	1.3	0.7	7.0	4.3	11	1.2	40	0.9	3.3	6.4	0.2

FIGURE 5. Statistical analysis for the detection of sequence biases in stable V_H domains. Each cell shows the frequency of occurrence of each amino acid (columns) at each randomized position (rows) of library B, following selection for binding to protein A (top number) or in the naive library (bottom number). The χ^2 statistic with one degree of freedom was computed to test the statistical significance of differences between the selected and naive frequencies, and shaded squares indicate statistically significant enrichments in the selected population as follows: gray, p value $< 10^{-4}$; black, p value $< 10^{-8}$. The frequencies of the parental sequences are shown in boldface type.

one domain (VH-B1) exhibited a yield greater than that of VH-A1. These results suggest that we have reached the limits for protein yield in our secretion vector system, as extensive screening of the second generation population only resulted in modest improvements relative to the best first generation variant. We speculate that the high thermostability and reversible folding characteristics of the optimized V_H domains may allow for correct folding of essentially all of the secreted polypeptides, and thus, the rate-limiting step for increased yield may now be translation and/or secretion, rather than folding.

We subjected VH-B1 to analytical gel filtration and found the protein to be essentially monomeric (Fig. 3C). We also monitored the thermal denaturation behavior of VH-B1, which exhibited essentially reversible folding behavior and thermostability identical to that of the first generation parent VH-A1 ($T_m = 73$ °C) (Fig. 4D and Table 1).

Shotgun Alanine-scanning Analysis of CDR3 Loops—As the CDR3 loops of camelid V_H domains are involved in stabilization of the protein fold, we investigated whether this is also the case for our *in vitro* evolved V_H domains. To facilitate the rapid analysis of many domains, we used an efficient shotgun alanine-scanning method that was previously validated for the analysis of several camelid V_H domains (29). For each domain, we constructed a library that allowed each CDR3 residue to vary as the wild type or Ala with equal frequency. Following two rounds of selection for binding to protein A, ~100 clones were sequenced, and the WT/Ala ratio was determined for each scanned position. The WT/Ala ratio correlates with the contributions of each wild-type side chain to stability, with WT/Ala ratios greater than or less than 1 indicating side chains that stabilize or destabilize the protein fold, respectively.

Autonomous Human V_H Domains

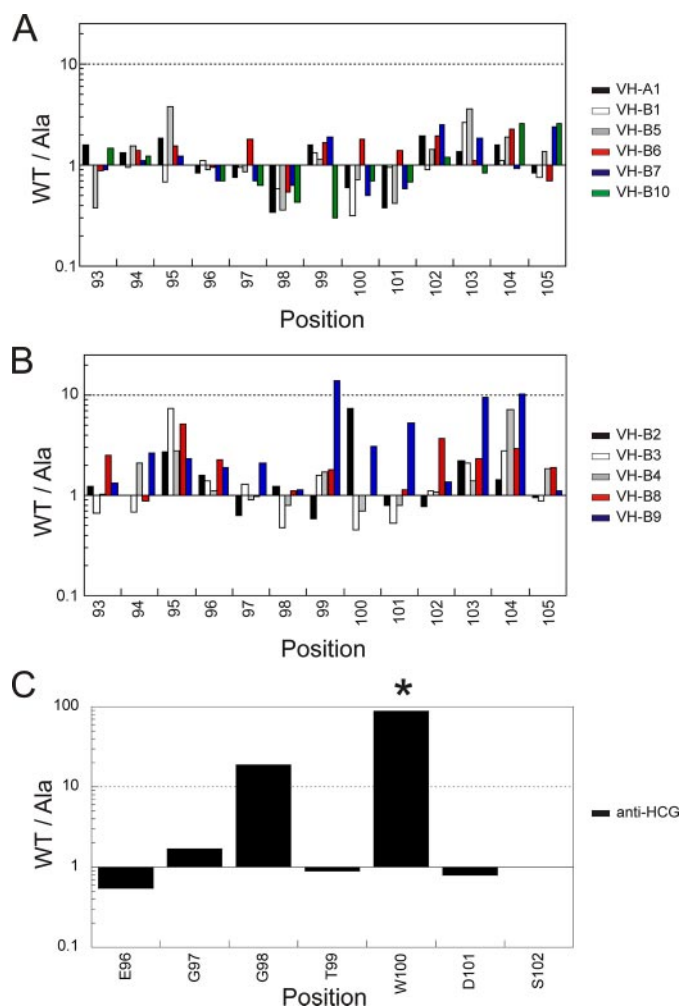


FIGURE 6. Shotgun alanine-scanning analysis of CDR3 loops. The WT/Ala ratios for each residue in CDR3 following selection for binding to protein A are shown for V_H domains whose stability appears to be independent of CDR3 (A), V_H domains whose stability appears to be dependent on CDR3 (B), and the camelid V_H domain α -HCG (C). The asterisk indicates that no Ala substitutions were observed, and thus the WT/Ala ratio is a lower limit.

We scanned 10 variants from the second generation library and also the parental domain from the first generation library (VH-A1). For many of the domains, including VH-A1 and VH-B1, stability is essentially independent of the CDR3 loops, as the WT/Ala ratios were close to 1 at all positions (Fig. 6A). Furthermore, Ala was actually preferred over larger wild-type side chains in the central region of the loops (positions 98–101), suggesting that small Ala side chains at these positions stabilize the fold, perhaps by favoring a turn conformation. However, for several domains, stability does appear to depend significantly upon the CDR3 loop, and in one case (VH-B9), almost half of the residues within CDR3 exhibit WT/Ala ratios significantly greater than 1 (Fig. 6B).

For comparison, we also alanine-scanned the CDR3 loop of an autonomous camelid V_H H domain (anti-HCG), in which a Trp residue within CDR3 (Trp-100) packs against framework residues (Phe-37 and Arg-45) and sequesters the hydrophobic portions of the former light chain interface from the solvent environment (15, 29). Two of the seven scanned residues were intolerant to Ala substitutions (Fig. 6C); Trp-100 was com-

pletely conserved, and Gly-98 was highly conserved. These results confirm the importance of CDR3 for the stability of the camelid V_H H domain and are in good agreement with previous results, which showed that four different CDR3 loops selected for stabilization of the same camelid framework also contained hydrophobic residues that were intolerant to substitution (29). For the anti-HCG V_H H domain, it appears that Trp-100 is required to provide packing interactions with the framework and Gly-98, which is located at the apex of a turn, is required for a conformation that provides the correct orientation of CDR3 relative to the framework. Interestingly, examination of the CDR3 sequence of VH-B9 (Fig. 2B) suggests that the residues that are resistant to Ala substitution may stabilize the domain through a similar mechanism because it is possible that Ile-101 and/or Trp-103 provides hydrophobic packing interactions similar to those of Trp-100 in the camelid V_H H domain, and Gly-99 and Pro-100 may be required for a turn conformation.

Taken together, these results demonstrate that, like camelid V_H H domains, the stabilities of some of our evolved V_H domains are dependent upon particular residues within the CDR3 loops. However, these evolved domains are only moderately dependent on CDR3 because in no case is the wild type more than 10-fold preferred over Ala (Fig. 6B), whereas in contrast, the camelid V_H H domains examined here (Fig. 6C) and elsewhere (29) are highly dependent on hydrophobic CDR3 sequences. Furthermore, the stabilities of most of the selected domains seem to be independent of the CDR3 loop, and this latter class includes VH-A1 and VH-B1, the two domains that exhibit the most desirable characteristics for autonomous V_H domains (Fig. 6A).

Quantitative Saturation Scanning Analysis of Former Light Chain Interfaces—Having ascertained that the stability of the VH-B1 domain is independent of the CDR3 loop sequence, we next investigated whether residues in the former light chain interface contribute to stability. In total, 10 positions outside CDR3 were randomized in the two-stage process that produced VH-B1 from VH-4D5. We conducted a quantitative saturation scan (48) of nine of these positions to assess the tolerance to all possible point mutations (position 103 was not scanned) (Fig. 7). At most scanned positions, modest biases are observed, and the parental sequence is among the most prevalent sequences, but no position is dominated by any single residue type. These results suggest that the stabilization of VH-B1 relative to VH-4D5 was achieved by cumulative effects arising from contributions of multiple mutations. It appears that most of the residues can be substituted by smaller and/or more hydrophilic amino acids, and notably, the small hydrophilic amino acids Ser and Thr are tolerated at all positions. Thus, it may be possible to further improve the solubility and monomeric state of VH-B1 by additional point mutations that improve the hydrophilicity of the former light chain interface without compromising stability (see below). Together with alanine-scanning analysis of CDR3, these results indicate that the stability of VH-B1 does not depend on interactions between the CDR3 loop and residues in the former light chain interface.

For comparison, we also conducted a saturation scan of four positions within the former light chain interface of the anti-HCG V_H H domain (Fig. 7). We randomized Phe-37 and Arg-

WT	W	F	Y	M	L	I	V	A	G	S	T	R	K	H	N	Q	D	E	P	C
G35 (H)	7	4	3	8	2	3	3	10	16	8	5	3	5	2	1	13	4	5	0	0
V37	13	16	17	3	5	3	10	3	0	9	8	2	0	3	3	2	2	2	0	0
R39 (Q)	4	3	3	6	2	4	8	3	3	4	8	8	8	7	4	7	8	11	0	0
G44	4	9	2	5	7	2	6	15	6	6	4	8	4	2	1	7	6	1	5	1
E45 (L)	2	9	7	5	6	6	4	7	5	6	5	5	5	7	6	4	3	5	4	0
W47	6	8	4	10	11	2	9	9	1	4	8	4	1	4	3	3	2	6	5	2
S50 (R)	8	0	1	8	2	3	1	6	18	9	6	15	3	0	2	8	2	9	0	0
Y91	10	13	12	3	6	4	9	4	3	8	4	7	4	4	2	4	3	3	0	0
W103	4	44	11	4	13	4	6	1	1	1	1	4	1	0	0	0	3	0	0	0
Q105	7	9	4	6	7	3	6	4	5	4	7	7	8	5	4	6	0	4	3	0

FIGURE 7. **Quantitative saturation scanning of former light chain interfaces.** The 1st column (WT) shows the sequence of VH-B1 and, in positions that differ, the sequence of VH-4D5 in parentheses. Following selection for binding to protein A, the percent occurrence of each amino acid type at each position was calculated for phage-displayed libraries of VH-B1 (top row) or the α -HCG V_HH domain (bottom row). The wild-type α -HCG sequence is boxed in black, and dashes indicate positions that were not scanned. Point mutations that were shown to be well tolerated in purified V_H protein variants are boxed in red (Table 2). Yellow or blue shading indicates sequences that are abundant (>5%) or rare (<0.5%), respectively, in human/mouse V_H domains (top row) or llama V_HH domains (bottom row) (20).

45, which pack with Trp-100 from CDR3 (see above), to form a small hydrophobic core and also two small residues (Ser-47 and Thr-91), which are close to Trp-100 and may accommodate hydrophobic substitutions that could expand upon this core. The results contrast with those for the VH-B1 domain, as there is strong sequence conservation at all four positions, suggesting that the stability of the camelid V_HH domain is dependent upon particular residues at these positions. At position 37, the wild-type Phe was dominant or was replaced by Trp. Similarly, at position 45 the wild-type Arg was dominant or was replaced by hydrophobic residues. Strikingly, the wild-type Thr at position 91 was almost entirely replaced by hydrophobic residues, and at position 47, although the wild-type Ser was quite abundant, the dominant amino acid was Trp and other hydrophobic residues were also prevalent. It is noteworthy that the sequence of the anti-HCG V_HH domain is unusual at positions 47 and 91 because these positions are occupied by hydrophobic residues in most V_HH domain sequences (Fig. 7). Taken together, these results confirm that the packing interactions between Trp-100, Phe-37, and the aliphatic portion of Arg-45 are important for stabilizing the anti-HCG V_HH domain fold and furthermore suggest that the introduction of hydrophobic residues at positions 47 and 91 may lead to further stabilization. Thus, in contrast with VH-B1, the stability of the camelid V_HH domain depends on hydrophobic residues in the former light chain interface and the CDR3 loop.

The Crystal Structure of an Autonomous Version of VH-4D5—Autonomous V_H domains such as VH-A1 and VH-B1 may be ideal scaffolds for the construction of synthetic antibody libraries, as they appear well adapted to support highly diverse CDR3 loops. To further test this supposition, we altered CDR3 and

surrounding residues of VH-B1 (positions 93–103) to resemble the sequence of VH-4D5. The resulting VH domain (VH-B1a) is identical to VH-4D5 aside from four mutations that were accumulated during the selection process (H35G, Q39R, L45E, and R50S) and one mutation preceding CDR3 (S93A). VH-B1a does not bind to Her2 (data not shown), but this result was expected because the 4D5 light chain makes contact with antigen and helps to support the conformation of the heavy chain (53). Also, Arg-50 in VH-4D5 is important for antigen recognition (53, 73), and substitution by Ser in VH-B1a may be detrimental for binding. However, unlike VH-4D5, VH-B1a is well expressed in *E. coli*, exhibits reversible folding kinetics upon thermal denaturation, and is more thermostable ($T_m = 79^\circ\text{C}$) than even VH-B1 (Table 2). Furthermore, the protein elutes predominantly as a monomer in gel filtration experiments, although the

elution time is increased in comparison with that of VH-B1, suggesting some interaction with the column matrix. These results demonstrate that the CDR3 loop does not contribute significantly to the stability of the VH-B1 fold, as the entire loop can be replaced without significantly affecting protein yield, oligomeric state, thermostability, or folding.

To gain insights into the structural basis for the autonomous behavior of VH-B1a, the crystal structure was solved and refined at 1.8 Å resolution with R_{work} and R_{free} of 16.0 and 19.9%, respectively (Table 3). The structure consists of four molecules per asymmetric unit, and unlike the HEL4 structure (35), the crystal packing contacts do not resemble the interactions observed in the interfaces of V_H/V_L pairs. Excluding the CDR3 loop, superimposition of the four independent variable domains and VH-4D5 reveals essentially identical α traces with root mean square deviations of 0.6 or 0.8 Å in comparing VH-B1a monomers to other VH-B1a monomers or VH-4D5, respectively (Fig. 8A). However, there are significant differences in the conformations of the CDR3 loops, which vary even among the VH-B1a structures, indicating that the loop is flexible and may adapt multiple conformations in solution.

Considering the four mutations outside CDR3 that were accumulated during the evolution of VH-B1a from VH-4D5 (Fig. 8B), it appears that the changes at positions 39 and 45 (Q39R/L45E) increase the solubility of VH-B1a because of the increased hydrophilicity of the side chains. In the case of the substitution of Arg by Ser at position 50, it is not clear if and how the change improves solubility because Arg is actually more hydrophilic than Ser. At position 35, the substitution of Gly for His creates a cavity that is occupied by Trp-95 from CDR3 in all four independent VH-B1a structures, despite the

TABLE 2
Biophysical analysis of VH-B1a variants

V _H	Mutations relative to VH-B1a	Analytical gel filtration						Temperature-induced denaturation			
		Molecular mass ^a			Fraction of total peak area		Elution time		T _m	Fraction refolded ^b (α)	
		Formula	Peak 1	Peak 2	Peak 1	Peak 2	Peak 1	Peak 2			
			kDa					min		°C	
B1a ^c		13.0	14.2	25.2	0.86	0.09	46.3	25.2	79	0.93	
B1b	W47A	12.9	12.4	19.4	0.86	0.13	30.1	24.4	76	0.98	
B1c	W47V	12.9	14.1	27.3	0.88	0.10	33.5	23.7	83	0.94	
B1d	W47L	12.9	13.6	32.0	0.91	0.07	31.5	23.7	82	0.95	
B1e	W47E	13.0	13.2	26.3	0.92	0.07	29.8	23.2	75	0.96	
B1f	W47T	12.9	13.4	29.2	0.90	0.08	32.8	23.0	75	0.98	
B1g	W47L/V37S	12.9	12.9	31.0	0.97	0.03	30.9	24.4	72	0.99	
B1h	W47L/V37T	13.0	13.3	29.0	0.89	0.08	31.3	24.3	ND ^d	ND	
B1i	W47L/W103R	12.9	13.2	27.2	0.56	0.21	29.8	23.9	83	1.00	
B1j	W47L/W103S	12.9	12.7	25.3	0.92	0.08	30.5	24.0	ND	ND	
B1k	W47L/W103T	12.9	12.6	26.0	0.88	0.11	30.0	23.7	ND	ND	
B1l	W47L/R39S	12.9	13.1	35.2	0.90	0.08	32.0	24.5	ND	ND	
B1m	W47L/R39T	12.9	13.4	34.1	0.84	0.15	31.6	24.2	ND	ND	
B1n	W47L/R39H	12.9	14.7	9.0	0.85	0.08	33.9	24.1	ND	ND	
B1o	W47L/R39K	12.9	12.6	27.2	0.90	0.08	31.8	24.0	ND	ND	
B1p	W47L/R39D	12.9	12.9	41.9	0.88	0.06	30.1	23.6	ND	ND	
B1q	W47L/R39E	13.0	12.8	34.1	0.88	0.07	30.0	23.4	ND	ND	
B1r	W47L/E45H	13.0	17.7	27.4	0.86	0.06	30.9	24.4	ND	ND	
B1s	W47L/E45S	12.9	15.3	30.1	0.83	0.04	35.9	24.6	ND	ND	
B1t	W47L/E45T	12.9	NA ^e	NA	NA	NA	NA	NA	ND	ND	
B1u	G35H	13.1	15.8	41.4	0.57	0.30	43.4	24.4	80	0.98	
B1v	R39Q	13.0	13.1	26.4	0.79	0.13	46.5	26.0	82	0.95	
B1w	E45L	13.0	13.6	NA	0.80	NA	74.8	NA	83	0.94	
B1x	S50R	13.1	13.4	27.1	0.92	0.07	30.1	24.8	77	0.94	
B1y	W47L/G35H	13.1	15.7	49.8	0.79	0.15	33.5	24.0	77	0.97	
B1z	W47L/R39Q	12.9	13.3	26.8	0.85	0.13	31.4	24.0	77	0.96	
B1aa	W47L/E45L	13.0	17.7	NA	0.87	NA	36.9	NA	84	0.96	
B1ab	W47L/S50R	13.0	13.2	24.9	0.96	0.04	30.0	24.5	78	0.96	
B1ac	W47T/G35H	13.0	14.2	46.8	0.79	NA	31.3	NA	65	0.94	
B1ad	W47T/R39Q	12.9	13.9	26.8	0.88	0.12	32.7	23.9	73	0.96	
B1ae	W47T/E45L	12.9	13.4	26.3	0.79	0.14	32.7	24.4	80	0.96	
B1af	W47T/S50R	13.0	12.8	26.2	0.96	0.10	30.2	24.5	75	0.97	

^a The weight average molar masses of peaks 1 and 2 (Fig. 3) were determined by light scattering.

^b The fraction refolded was defined as the fraction folded value (α) for the protein at 25 °C, following heating to a temperature that resulted in complete unfolding.

^c To obtain estimates of measurement errors, values for VH-B1a were determined by averaging data from two independent experiments, as follows: weight average molar mass of peak 1 (14.2 ± 0.2 kDa), fraction of total peak area for peak 1 (0.86 ± 0.01), elution time for peak 1 (46.3 ± 1.5 min), T_m (79 ± 1 °C), fraction refolded (0.93 ± 0.01).

^d ND indicates that the value was not determined.

^e NA indicates that the gel filtration data could not be analyzed, because distinct peaks could not be discerned.

TABLE 3
Data collection and refinement statistics for the VH-B1a domain

A. Unit cell	
Space group	P1
<i>a</i>	50.9 Å
<i>b</i>	54.1 Å
<i>c</i>	54.2 Å
α	110°
β	95.6°
γ	119°
B. Diffraction data	
Resolution (Å)	50-1.8
No. of reflections	40,233
Redundancy ^a	4.0 (4.0)
Completeness ^a	96.2 (92.6)%
R _{sym} ^{a,b}	0.039 (0.097)
I/σ(I) ^a	43.2 (13.4)
C. Refinement	
R _{work} ^c	0.155
R _{free} ^c	0.200
No. of protein atoms	3,657
No. of waters	405
Root mean square deviation bond length	0.015 Å
Root mean square deviation angles	1.5°

^a Values in parentheses refer to data in the highest resolution shell.

^b R_{sym} = Σ||I - <I>|/ΣI, where <I> is the average intensity of symmetry-related observations of a unique reflection.

^c R_{work} = R_{free} = Σ|F_o - F_c|/ΣF_o, where R_{free} represents 5% of the data selected randomly.

considerable differences in the CDR3 loop conformations. Consequently, compared with VH-4D5, the CDR3 loop of VH-B1a moves closer to the framework, but Trp-47 is exposed to solvent in both structures. These findings contrast with those for the HEL4 V_H domain, where Trp-47 was found in a cavity similar to that occupied by Trp-95 in VH-B1a, and it was concluded that the sequestration of the hydrophobic Trp-47 side chain was in part responsible for increasing solubility (35).

Increasing Hydrophilicity at the Former Light Chain Interface by Rational Design—Using the results of saturation scanning (Fig. 7) and structural analysis (Fig. 8), we analyzed a panel of VH-B1a mutants designed to explore and further optimize autonomous V_H domain behavior (Table 2). Because Trp-47 is exposed to solvent in the crystal structure, we reduced hydrophobicity at this position with substitutions that were abundant in the saturation scanning data set (VH-B1b-f). All five substitutions reduced gel filtration elution times and had only minimal effects on thermostability, which was reduced or increased slightly by hydrophilic (Thr/Glu) or aliphatic (Leu/Val) substitutions, respectively.

We also analyzed the effects of hydrophilic substitutions for two other exposed hydrophobic residues by adding mutations in the background of VH-B1d (VH-B1a with a W47L mutation). The tolerance of the V_H domain to these mutations was

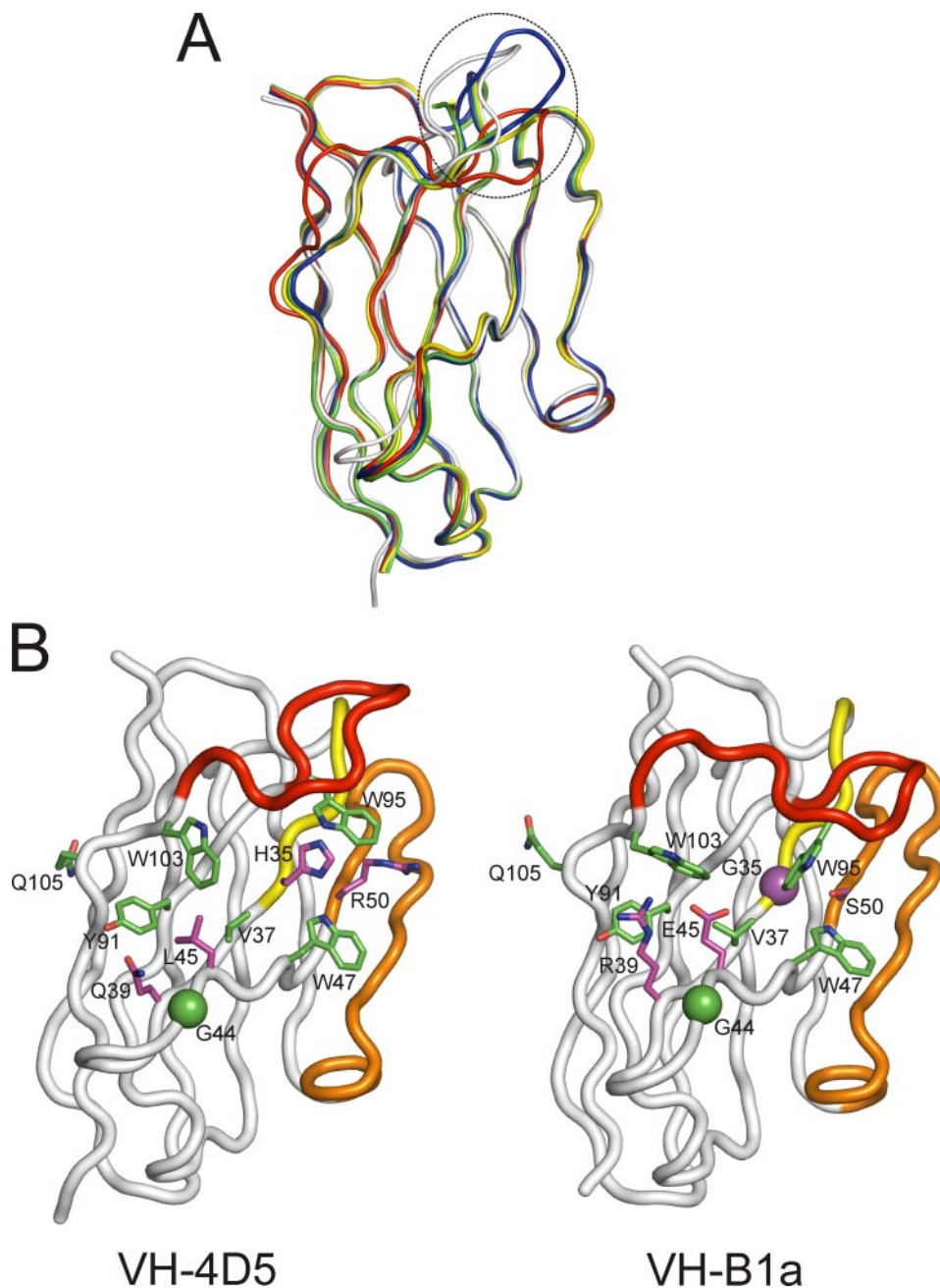


FIGURE 8. The crystal structure of VH-B1a. *A*, superimposition of the main chain of VH-4D5 (gray) and the four independent VH-B1a molecules in the asymmetric unit (colored). The dashed oval demarcates the CDR3 loops. *B*, comparison of the light chain interface of VH-4D5 (left) and the former light chain interface of VH-B1a (right). The model of VH-4D5 was derived from the Fab-4D5 structure coordinates (PDB code 1N8Z). The main chain is shown as a tube and the CDR loops are colored as follows: CDR1 (yellow), CDR2 (orange), and CDR3 (red). Gly residues are shown as spheres and side chains are rendered as sticks. Atoms are colored as follows: nitrogen, blue; oxygen, red; carbon, green or magenta for residues that are the same or different in the two domains, respectively.

assessed by the effects on the elution time and area of the monomer peak observed by gel filtration. Substitution of Val-37 by Ser or Thr was well tolerated, and in particular, the Ser-substituted protein (VH-B1g) was very well behaved in gel filtration, as 97% of the sample eluted as a sharp single peak with an elution time and apparent molecular mass consistent with a monomer. The thermostability of VH-B1g ($T_m = 72^\circ\text{C}$) was reduced relative to that of VH-B1a ($T_m = 79^\circ\text{C}$), but nonetheless, it was still greater than those of typical camelid V_H H

domains ($T_m = 57\text{--}67^\circ\text{C}$) (12, 29). Although position 103 was not subjected to saturation scanning, the analysis of library B suggested that the solvent-exposed Trp-103 may be replaced by Arg (Fig. 5). Incorporation of this substitution produced a domain (VH-B1i) that was very thermostable ($T_m = 83^\circ\text{C}$) and unfolded reversibly, but a significant proportion of the protein eluted from the gel filtration column as a dimer (peak 2). However, Ser or Thr substitutions at position 103 were also well tolerated (VHB1j and VHB1k) and the purified proteins exhibited monomeric behavior by gel filtration.

Finally, we explored the effects of hydrophilic substitutions for two hydrophilic residues (Arg-39 and Glu-45) that were acquired during the evolution of VH-B1 from VH-4D5. At position 39, six different substitutions were well tolerated (VH-B1l-q), indicating that general hydrophilic character at this position is sufficient to promote monomeric behavior of the V_H domain. At position 45, His was well tolerated (VH-B1r), but substitution by Ser increased the elution time significantly (VH-B1s), and substitution by Thr (VH-B1t) resulted in significant aggregation. Thus, it appears that hydrophilic residues at positions 39 and 45 promote solubility of the V_H domain, but position 45 is more discriminating with regard to the chemical nature of the residues.

Taken together, these results validate the utility of the saturation scanning analysis for predicting substitutions that are well tolerated structurally by the V_H domain fold. By judiciously incorporating hydrophilic substitutions that are prevalent in the quantitative scanning

data set (Fig. 7), we were able to increase the hydrophilicity of the former light chain interface without significantly compromising thermostability. It is notable that many of the best substitutions for improving autonomous V_H domain behavior are rare in natural V_H and V_H H domains and thus could not have been identified by analysis of natural sequences and structures.

Back Mutation Analysis of the Former Light Chain Interface—To further characterize how mutations contribute to autonomous V_H domain behavior, we analyzed variants of VH-B1a to

Autonomous Human V_H Domains

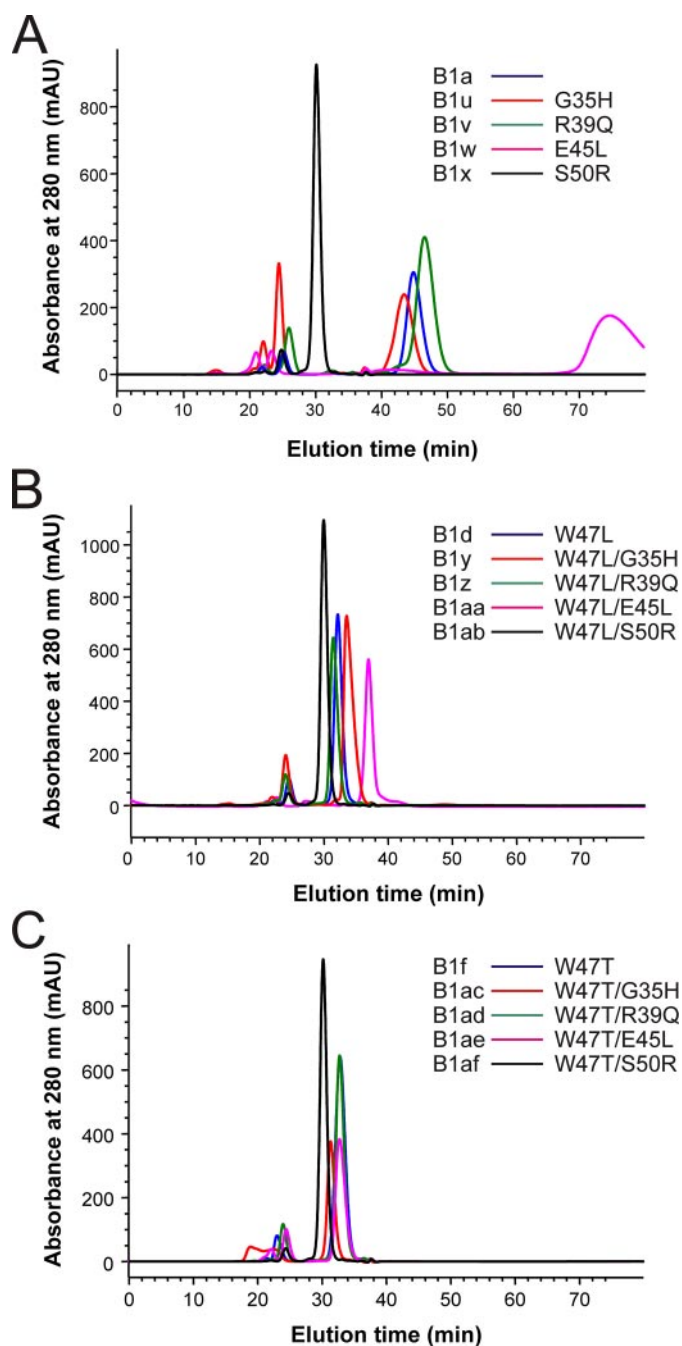


FIGURE 9. Analytical gel filtration of VH-B1a variants. Chromatograms are shown for 70 μM samples of V_H domains applied on a Superdex-75 column. The domains contain back mutations to the VH-4D5 sequence in the background of VH-B1a (A), VH-B1d (B), and VH-B1f domains (C). Chromatograms are colored as indicated in the legend, which shows the name of each domain at the left (Table 2) and the mutations relative to VH-B1a at the right.

assess the effects of back mutations at four positions that changed during the evolution of VH-B1 from VH-4D5 (Fig. 9A and Table 2). The substitution of Gly-35 by His (VH-B1u) did not significantly alter thermostability or gel filtration elution time, but it did greatly reduce the fraction of the protein eluting as a monomer. Similarly, the substitution of Arg-39 by Gln (VH-B1v) reduced the monomeric fraction but only marginally increased the elution time and even improved thermostability slightly ($T_m = 82^\circ\text{C}$). The substitution of Leu for Glu-45 (VH-

B1w) also increased thermostability, but in contrast with the other back mutations, the elution time was increased dramatically, indicating that the greater hydrophobicity of the Leu side chain promotes interactions with the hydrophobic column matrix. Interestingly, the protein containing the fourth back mutation, substitution of Ser-50 by Arg (VH-B1x), exhibited slightly reduced thermostability but eluted as a sharp peak with an elution time and calculated molecular mass consistent with a well behaved monomer.

We also performed the back mutation analysis in the background of VH-B1a variants containing either Leu or Thr in place of Trp position 47 (VH-B1d and -f). Compared with VH-B1a, VH-B1d and VH-B1f exhibit gel filtration profiles more typical of monomeric proteins, and thus, the effects of the back mutations were muted in these backgrounds (Fig. 9 and Table 2, VH-B1y to -af). However, in all cases, the back mutation at position 50 improved monomeric behavior, whereas those at the other three positions had detrimental effects.

Taken together, these results indicate that substitutions that reduce hydrophobicity at positions 39, 45, and 47 act additively to improve the solubility of V_H domains. The presence of a Gly residue at position 35 is beneficial when Trp is present at position 47 but is not necessary when Trp-47 is replaced by less hydrophobic residues. Finally, it appears that the Ser for Arg substitution in VH-B1a compared with VH-4D5 is a spurious mutation that is actually detrimental to autonomous V_H domain behavior.

DISCUSSION

Our analysis of sequence features that promote autonomous V_H domain behavior reveals that, unlike camelid V_HH domains, most solutions do not require interactions between the CDR3 loop and the former light chain interface. Instead, hydrophilicity of the former light chain interface is increased by replacing exposed hydrophobic residues with more hydrophilic residues that are compatible with the structure of the V_H domain fold. Importantly, our quantitative saturation scanning analysis reveals that many different substitutions, both hydrophobic and hydrophilic, are structurally viable across the former light chain interface (Fig. 7). Many of these substitutions are rare in both natural V_H and V_HH domains and thus could only be discovered by the comprehensive and unbiased strategy that we employed. This data set expands the options available for the engineering of autonomous V_H domains beyond the scope of natural sequence diversity.

By incorporating sequence changes suggested by biophysical analysis, rather than those suggested by the analysis of natural sequences, we evolved frameworks that should be ideal scaffolds for engineering autonomous V_H domains with novel specificities (Tables 1 and 2). These domains can be purified at high yields from *E. coli* (Fig. 2) and, like camelid V_HH domains, remain monomeric at high concentrations (Fig. 3) and fold reversibly (Fig. 4). Moreover, the thermostabilities of many of these domains exceed those of typical V_HH domains (12, 29) and approach those of highly stable Fabs. Furthermore, the autonomous nature of many of our *in vitro* evolved domains stems from only a few framework changes and, unlike camelid domains, is independent of the CDR3 loop (Fig. 6). Thus, we are

optimistic that these novel V_H domain scaffolds will be able to support diverse CDR3 sequences and should enable the development of naive phage-displayed libraries that can be used to derive specific antibodies against diverse antigens.

Acknowledgments—We thank Robert Kelley and Charles Eigenbrot for helpful discussions and Ivan Bosanac for assistance in x-ray data collection. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the United States Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the NIGMS, National Institutes of Health.

REFERENCES

- Padlan, E. A. (1994) *Mol. Immunol.* **31**, 169–217
- Holt, L. J., Herring, C., Jespers, L. S., Woolven, B. P., and Tomlinson, I. M. (2003) *Trends Biotechnol.* **21**, 484–490
- Muyldermans, S., Cambillau, C., and Wyns, L. (2001) *Trends Biochem. Sci.* **26**, 230–235
- Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N., and Hamers, R. (1993) *Nature* **363**, 446–448
- Els Conrath, K., Lauwereys, M., Wyns, L., and Muyldermans, S. (2001) *J. Biol. Chem.* **276**, 7346–7350
- Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R., Muyldermans, S., and Wyns, L. (1996) *Nat. Struct. Biol.* **3**, 803–811
- Stijlemans, B., Conrath, K., Cortez-Retamozo, V., Van Xong, H., Wyns, L., Senter, P., Revets, H., De Baetselier, P., Muyldermans, S., and Magez, S. (2004) *J. Biol. Chem.* **279**, 1256–1261
- Cortez-Retamozo, V., Lauwereys, M., Hassanzadeh Gh, G., Gobert, M., Conrath, K., Muyldermans, S., De Baetselier, P., and Revets, H. (2002) *Int. J. Cancer* **98**, 456–462
- Rosebrough, S. F. (1996) *Q. J. Nucl. Med.* **40**, 234–251
- Sundaresan, G., Yazaki, P. J., Shively, J. E., Finn, R. D., Larson, S. M., Raubitschek, A. A., Williams, L. E., Chatziioannou, A. F., Gambhir, S. S., and Wu, A. M. (2003) *J. Nucl. Med.* **44**, 1962–1969
- Zuckier, L. S., and DeNardo, G. L. (1997) *Semin. Nucl. Med.* **27**, 10–29
- Ewert, S., Cambillau, C., Conrath, K., and Pluckthun, A. (2002) *Biochemistry* **41**, 3628–3636
- Worn, A., and Pluckthun, A. (1999) *Biochemistry* **38**, 8739–8750
- Desmyter, A., Decanniere, K., Muyldermans, S., and Wyns, L. (2001) *J. Biol. Chem.* **276**, 26285–26290
- Spinelli, S., Frenken, L., Bourgeois, D., de Ron, L., Bos, W., Verrips, T., Anguille, C., Cambillau, C., and Tegoni, M. (1996) *Nat. Struct. Biol.* **3**, 752–757
- Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M. A., Muyldermans, S., and Wyns, L. (1999) *Structure (Lond.)* **7**, 361–370
- Spinelli, S., Frenken, L. G., Hermans, P., Verrips, T., Brown, K., Tegoni, M., and Cambillau, C. (2000) *Biochemistry* **39**, 1217–1222
- Spinelli, S., Tegoni, M., Frenken, L., van Vliet, C., and Cambillau, C. (2001) *J. Mol. Biol.* **311**, 123–129
- Desmyter, A., Spinelli, S., Payan, F., Lauwereys, M., Wyns, L., Muyldermans, S., and Cambillau, C. (2002) *J. Biol. Chem.* **277**, 23645–23650
- Harmsen, M. M., Ruuls, R. C., Nijman, I. J., Niewold, T. A., Frenken, L. G. J., and de Geus, B. (2000) *Mol. Immunol.* **37**, 579–590
- Nguyen, V. K., Hamers, R., Wyns, L., and Muyldermans, S. (2000) *EMBO J.* **19**, 921–930
- Davies, J., and Riechmann, L. (1994) *FEBS Lett.* **339**, 285–290
- Riechmann, L. (1996) *J. Mol. Biol.* **259**, 957–969
- Riechmann, L., and Davies, J. (1995) *J. Biomol. NMR* **6**, 141–152
- Davies, J., and Riechmann, L. (1995) *Bio/Technology* **13**, 475–479
- Kortt, A. A., Guthrie, R. E., Hinds, M. G., Power, B. E., Ivancic, N., Caldwell, J. B., Gruen, L. C., Norton, R. S., and Hudson, P. J. (1995) *J. Protein Chem.* **14**, 167–178
- Martin, F., Volpari, C., Steinkuhler, C., Dimasi, N., Brunetti, M., Biasiol, G., Altamura, S., Cortese, R., De Francesco, R., and Sollazzo, M. (1997) *Protein Eng. Eng.* **10**, 607–614
- Voordijk, S., Hansson, T., Hilvert, D., and van Gunsteren, W. F. (2000) *J. Mol. Biol.* **300**, 963–973
- Bond, C. J., Marsters, J. C., and Sidhu, S. S. (2003) *J. Mol. Biol.* **332**, 643–655
- Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T., and Winter, G. (1989) *Nature* **341**, 544–546
- Reiter, Y., Schuck, P., Boyd, L. F., and Plaxin, D. (1999) *J. Mol. Biol.* **290**, 685–698
- Tanha, J., Dubuc, G., Hiram, T., Narang, S. A., and MacKenzie, C. R. (2002) *J. Immunol. Methods* **263**, 97–109
- Jespers, L., Schon, O., Famm, K., and Winter, G. (2004) *Nat. Biotechnol.* **22**, 1161–1165
- To, R., Hiram, T., Arbabi-Ghahroudi, M., MacKenzie, R., Wang, P., Xu, P., Ni, F., and Tanha, J. (2005) *J. Biol. Chem.* **280**, 41395–41403
- Jespers, L., Schon, O., James, L. C., Vepintsev, D., and Winter, G. (2004) *J. Mol. Biol.* **337**, 893–903
- Vranken, W., Tolkachev, D., Xu, P., Tanha, J., Chen, Z., Narang, S., and Ni, F. (2002) *Biochemistry* **41**, 8570–8579
- Riechmann, L., and Muyldermans, S. (1999) *J. Immunol. Methods* **231**, 25–38
- Barbas, C. F., III, Amberg, W., Simoncsits, A., Jones, T. M., and Lerner, R. A. (1993) *Gene (Amst.)* **137**, 57–62
- Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Griffiths, A. D., Williams, S. C., Harley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E., and Winter, G. (1994) *EMBO J.* **13**, 3245–3260
- Hoet, R. M., Cohen, E. H., Kent, R. B., Rookey, K., Schoonbroodt, S., Hogan, S., Rem, L., Frans, N., Daukandt, M., Pieters, H., van Hegelsom, R., Neer, N. C., Natri, H. G., Rondon, I. J., Leeds, J. A., Hufton, S. E., Huang, L., Kashin, I., Devlin, M., Kuang, G., Steukers, M., Viswanathan, M., Nixon, A. E., Sexton, D. J., Hoogenboom, H. R., and Ladner, R. C. (2005) *Nat. Biotechnol.* **23**, 344–348
- Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellhofer, G., Hoess, A., Wolle, J., Pluckthun, A., and Virnekas, B. (2000) *J. Mol. Biol.* **296**, 57–86
- Lee, C. V., Liang, W. C., Dennis, M. S., Eigenbrot, C., Sidhu, S. S., and Fuh, G. (2004) *J. Mol. Biol.* **340**, 1073–1093
- Sidhu, S. S., Weiss, G. A., and Wells, J. A. (2000) *J. Mol. Biol.* **296**, 487–495
- Fellouse, F. A., and Sidhu, S. S. (2007) in *Making and Using Antibodies* (Howard, G. C., and Kaser, M. R., eds) pp. 157–180, CRC Press, Inc., Boca Raton, FL
- Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000) *Methods Enzymol.* **328**, 333–363
- Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A., and Sidhu, S. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8950–8954
- Bond, C. J., Wiesmann, C., Marsters, J. C., Jr., and Sidhu, S. S. (2005) *J. Mol. Biol.* **348**, 699–709
- Pal, G., Kouadio, J.-L. K., Artis, D. R., Kossiakoff, A. A., and Sidhu, S. S. (2006) *J. Biol. Chem.* **281**, 22378–22385
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Greenfield, N. J. (2006) *Nat. Protoc.* **1**, 2527–2535
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 458–464
- Cho, H. S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., Jr., and Leahy, D. J. (2003) *Nature* **421**, 756–760
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255
- Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystal-*

Autonomous Human V_H Domains

- logr.* **60**, 2126–2132
56. Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B., Henner, D., Wong, W. L., Rowland, A. M., Kotts, C., Carver, M. E., and Shepard, H. M. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4285–4289
 57. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987) *Science* **235**, 177–182
 58. Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., and Ullrich, A. (1989) *Science* **244**, 707–712
 59. Yarden, Y., and Sliwkowski, M. X. (2001) *Nat. Rev. Mol. Cell Biol.* **2**, 127–137
 60. Tokunaga, E., Oki, E., Nishida, K., Koga, T., Egashira, A., Morita, M., Kakeji, Y., and Maehara, Y. (2006) *Int. J. Clin. Oncol.* **11**, 199–208
 61. Kirkham, P. M., Mortari, F., Newton, J. A., and Schroeder, J. A., Jr. (1992) *EBMO J.* **11**, 603–609
 62. Eigenbrot, C., Randal, M., Presta, L., Carter, P., and Kossiakoff, A. A. (1993) *J. Mol. Biol.* **229**, 969–995
 63. Carter, P., Kelley, R. F., Rodrigues, M. L., Snedecor, B., Covarrubias, M., Velligan, M. D., Wong, W. L., Rowland, A. M., Kotts, C. E., and Carver, M. E. (1992) *Bio/Technology* **10**, 163–167
 64. Sidhu, S. S., Li, B., Chen, Y., Fellouse, F. A., Eigenbrot, C., and Fuh, G. (2004) *J. Mol. Biol.* **338**, 299–310
 65. Meininger, D. P., Rance, M., Starovasnik, M. A., Fairbrother, W. J., and Skelton, N. J. (2000) *Biochemistry* **39**, 26–36
 66. Starovasnik, M. A., O'Connell, M. P., Fairbrother, W. J., and Kelley, R. F. (1999) *Protein Sci.* **8**, 1423–1431
 67. Graille, M., Stura, E. A., Corper, A. L., Sutton, B. J., Taussig, M. J., Charbonnier, J. B., and Silverman, G. J. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5399–5404
 68. Potter, K. N., Li, Y., and Capra, J. D. (1996) *J. Immunol.* **157**, 2982–2988
 69. Potter, K. N., Li, Y., Pascual, V., and Capra, J. D. (1997) *Int. Rev. Immunol.* **14**, 291–308
 70. de Wildt, R. M., Mundy, C. R., Gorick, B. D., and Tomlinson, I. M. (2000) *Nat. Biotechnol.* **18**, 989–994
 71. Spinelli, S., Desmyter, A., Frenken, L., Verrips, T., Tegoni, M., and Cambillau, C. (2004) *FEBS Lett.* **564**, 35–40
 72. Kelley, R. F., O'Connell, M. P., Carter, P., Presta, L., Eigenbrot, C., Covarrubias, M., Snedecor, B., Bourell, J. H., and Vetterlein, D. (1992) *Biochemistry* **31**, 5434–5441
 73. Kelley, R. F., and O'Connell, M. P. (1993) *Biochemistry* **32**, 6828–6835
 74. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H., and Gottesman, K. (1987) *Sequences of Proteins of Immunological Interest*, 4th Ed., pp. 1–2387, National Institutes of Health, Bethesda