Structure-Based Improvement of the Biophysical Properties of Immunoglobulin V_H Domains with a Generalizable Approach†

Stefan Ewert,‡ Annemarie Honegger, and Andreas Plückthun*

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received July 12, 2002; Revised Manuscript Received October 19, 2002

ABSTRACT: In a systematic study of V gene families carried out with consensus V_H and V_L domains alone and in combinations in the scFv format, we found comparatively low expression yields and lower cooperativity in equilibrium unfolding in antibody fragments containing V_H domains of human germline families 2, 4, and 6. From an analysis of the packing of the hydrophobic core, the completeness of charge clusters, the occurrence of unsatisfied hydrogen bonds, and residues with low β-sheet propensities, positive Φ angles, and exposed hydrophobic side chains, we pinpointed residues potentially responsible for the unsatisfactory properties of these germline-encoded sequences. Several of those are in common between the domains of the even-numbered subgroups, but do not occur in the odd-numbered ones. In this study, we have systematically exchanged those residues alone and in combination in two different scFvs using the V_H6 framework, and we describe their effect on equilibrium stability and folding yield. We improved the stability by 20.9 kJ/mol and the expression yield by a factor of 4 and can now use these data to rationally engineer antibodies derived from this and similar germline families for better biophysical properties. Furthermore, we provide an improved design for libraries exploiting the significant additional diversity provided by these frameworks. Both antibodies studied here completely retain their binding affinity, demonstrating that the CDR conformations were not affected.

Recombinant antibodies are used in an ever increasing number of applications from biological research to therapy. In addition to exhibiting high antigen specificity and affinity, such recombinant antibodies should also be obtainable in high yield, have a low tendency to aggregate, and be stable against high denaturant concentrations, elevated temperatures, and proteases, depending on the requested task. A popular format for many of these applications is the single-chain Fv (scFv)[1] fragment, where the variable domain of the heavy chain (V_H) is connected via a flexible linker to the variable domain of the light chain (V_L) or vice versa (*1*−*3*). This format contains the complete antigen binding site and can be expressed in a wide range of hosts, including bacteria (*4*) and yeast (*5*). While we chose to investigate the effect of mutations with scFvs, as their simple structure makes an untangling of domain interactions much easier, differences in physical properties are also manifest in Fab fragments and whole antibodies, which contain the same domains.

Mutations that are important for the biophysical behavior can influence either the equilibrium thermodynamic stability, the aggregation tendency during folding, or both. Mutations influencing thermodynamic stability can be involved in many different types of interactions, such as packing of the hydrophobic core, secondary structure propensity, charge interactions, hydrogen bonding, desolvation upon folding, and compatibility with the enforced local structure (*6*, *7*). Mutations that influence folding efficiency can also be part of this list, as the stability of intermediates is an important component. Additionally, however, natural proteins use "negative design" (*8*) to avoid aggregation. In its simplest form, this strategy avoids hydrophobic patches on the surface. In the case of antibodies, such hydrophobic patches were found to have almost no effect on the solubility of the native protein, correctly defined as the maximal concentration of the soluble native protein (*9*). The hydrophobic patches can have a very dramatic effect on the folding yield and thus the yield of the functional protein in *Escherichia coli*, which is colloquially but incorrectly often termed "solubility", as the yield describes the overall process of producing soluble protein, but not its solubility.

In the case of scFvs, a further complication for the analysis of stability is introduced by their two-domain nature. The two domains can stabilize each other and unfold either cooperatively or with an equilibrium intermediate, depending on the relative intrinsic stability of the domains and their interface (*10*). However, from these studies of domain interactions and a systematic study of isolated domains and their interactions (*11*), we can now untangle this system. We

† This work was supported by Swiss National Science Foundation Grant 31-65344.01.

* To whom correspondence should be addressed. Telephone: (+41-1) 6355570. Fax: (+41-1) 6355712. E-mail: plueckthun@biocfebs.unizh.ch.

‡ Present address: ESBATech AG, Wagistr. 21, CH-8952 Zürich-Schlieren, Switzerland.

[1] Abbreviations: CDR, complementary determining region; GdnHCl, guanidine hydrochloride; HuCAL, Human Combinatorial Antibody Library; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl β-D-thiogalactopyranoside; scFv, single-chain antibody fragment consisting of the variable domains of the heavy and light chains connected by a peptide linker; V_H, variable domain of the heavy chain of an antibody; V_L, variable domain of the light chain of an antibody.

10.1021/bi026448p CCC: $25.00
Published on Web 01/25/2003

can thus pinpoint the problem spots, and in this study, we provide the evidence that a correction of multiple small defects indeed leads to a marked improvement of phenotypes.

There are three principal mechanisms leading to an increased expression yield of soluble proteins: increasing the total expression level (provided the relative folding yield remains constant), increasing the relative folding yield, or decreasing the level of degradation by proteases. All three mechanisms can be somewhat influenced by extrinsic factors, including the choice of bacterial strain, expression vector, media composition, and expression temperature (summarized in ref *4*) and coexpression of periplasmic chaperones (*12*, *13*). Nevertheless, the major contribution to changes of the expression yield of folded protein is due to changes in the protein sequence itself. In the case of secreted proteins placed in the same vector, the translation initiation region and the beginning of the protein sequence (the signal sequence) are identical among different variants. Therefore, sequence changes are extremely unlikely to influence translation per se. Mutations leading to higher thermodynamic stability often also decrease the extent of protease digestion of the protein, as the *E. coli* proteases usually prefer unfolded protein as a substrate. Nevertheless, mutations removing potential cutting sites for *E. coli* proteases may also prevent degradation. In the case of recombinant antibodies, periplasmic aggregation is probably the major yield-limiting mechanism (*4*).

Mutations can influence the efficiency of folding, independent of influencing the equilibrium thermodynamic stability of the protein. Side reactions of the folding process often lead to aggregated protein, which is enriched in inclusion bodies. The kinetic partitioning into productive folding and aggregation can be influenced by mutations either increasing the thermodynamic stability of intermediates, e.g. removing a solvent-exposed hydrophobic residue, or otherwise making the surface less suitable for aggregate growth [negative design (*8*)]. In addition, the mutations increasing folding efficiency can also indirectly lead to a higher total expression level by preventing the formation of toxic side products, most likely soluble aggregates, which lead to leakiness of the outer membrane and eventually decrease the viability of *E. coli*.

There are different approaches to finding residues that improve the thermodynamic stability and yield of the soluble protein of scFvs [reviewed by Wörn and Plückthun (*7*)]. Previously, most work had concentrated on the optimization of individual antibodies. If the three-dimensional (3D) structure of the antibody to be improved is known, a detailed analysis can identify problematic residues, which can then be exchanged by site-directed mutagenesis (*14*−*16*). A second approach uses random mutagenesis followed by selection with a bias toward the improvement of the desired property (*17*−*19*). The consensus approach as a third approach (*20*) uses the sequence information from antibodies naturally encoded by the immune system. The genes of immunoglobulin variable domains, as is assumed for all gene families, have diverged by multiple gene duplications and mutations. Selected genes are further subjected to an accelerated "local" evolution by somatic mutations that optimize the capacity of the antibody to bind to antigen structures with high affinity, but these mutations are not propagated in the germline. In contrast, mutations acquired during the duplication of the primordial V gene to make the present-day Ig locus are manifest as germline family specific differences.

Destabilizing mutations may be highly probable but are selectively neutral as long as the overall domain stability does not fall below a certain threshold (*20*). Conversely, random mutations resulting in increased thermodynamic stability are highly improbable in the absence of a positive selection. Consequently, the most frequent amino acid at any position in an alignment of homologous immunoglobulin variable domains should be most favorable for the stability of the protein domain. This method was tested on a V_κ domain, and of ten proposed mutations, six increased the stability. Nevertheless, the simplification inherent in this approach is that all frameworks are averaged to a single "ideal" sequence. The different germline genes or frameworks have an important function for antibody diversity. First, framework residues in the outer loop and close to the 2-fold axis can contribute important interactions to protein− and hapten−antigens, respectively. Second, several framework regions can influence the conformation of the CDRs and thereby indirectly modulate antigen binding. Third, different frameworks carry mutually incompatible residues, which cannot simply be exchanged with those of other frameworks. It follows that family specific solutions are needed to create a variety of *different* frameworks with superior properties. In this paper, we provide the basis for this approach.

In this study, we explored a generic approach for improving antibodies for their biophysical properties combining the structure based considerations with our knowledge of the biophysical properties of the germline-encoded V_H, V_κ, and V_λ families (*11*) derived from the Human Combinatorial Antibody Library (HuCAL) (*21*). Since we focus on genes with initially germline-encoded sequences, our approach is not limited to improving individual molecules and thus to removing changes introduced by somatic mutations, but particularly to problematic residues encoded by different germline genes.

In the case of the V_H domains, we found that the V_H3 germline family specific consensus domain was the most stable V_H domain, followed by the V_H1a, V_H1b, and V_H5 consensus domains with intermediate stabilities and only little or no aggregation-prone behavior. V_H2, V_H4, and V_H6 domains, on the other hand, exhibited low cooperativity during denaturant-induced unfolding, lower yields, and higher tendencies to aggregate. The detailed analysis of hydrophobic core packing and formation of salt bridges revealed that the V_H3 domain had always found the optimal solution while all other V_H domains had some shortcomings explaining the higher thermodynamic stability of V_H3. Furthermore, with the help of a sequence alignment grouped by V_H domains with favorable properties (families 1, 3, and 5) and unfavorable properties (families 2, 4, and 6), residues of the even-numbered V_H domains were identified and structurally analyzed which potentially decrease the folding efficiency.

In this study, we used a structure-based approach exploiting the knowledge of the biophysical properties of the human germline family specific consensus V_H domains (*11*) and, in addition, resorting to tables of published and in-house selection experiments (A. Honegger et al., unpublished results) to improve the V_H6 framework as a model. We chose the V_H6 framework because it shows a somewhat aggrega-

tion-prone behavior and the lowest midpoint of denaturation, compared to the other human V_H domains, indicating that the V_H6 domain is the V_H domain with the lowest thermodynamic stability. These properties were observed with isolated domains as well as in the scFv format with $V_\kappa3$ (*11*). We used two scFvs containing the V_H6 framework which had been selected from the HuCAL (*21*): 2C2, binding the peptide M18 coupled to transferrin, and 6B3, binding myoglobin (see Materials and Methods for details). With site-directed mutagenesis and on the basis of our structural analysis, we introduced six mutations (Q5V, S16G, T58I, V72D, S76G, and S90Y) alone and in several combinations, which were hypothesized to be independently acting and individually exchangeable and were also a feature distinguishing the group of V_H families with favorable properties from the families with less favorable properties. We compared these mutants to the wild-type scFvs for effects on folding yield and, independently, the free energy of unfolding as a measure of the thermodynamic stability and determined the additivity of these mutations.

MATERIALS AND METHODS

Construction of Expression Vectors. scFv 2C2 (A. Hahn et al., Morphosys AG, unpublished results) with the human consensus domains V_H6 and $V_\kappa3$ (CDR-H3, QRGHYGKGYKGFNSGFFDF; and CDR-L3, QYYNIPT) was obtained by panning against the peptide M18 with the sequence CDAFRSEKSRQELNTIASKPPRDHVF coupled to transferrin (Jerini GmbH, Berlin, Germany), while scFv 6B3 (S. Müller et al., Morphosys AG, unpublished results) with V_H6 and $V_\lambda3$ (CDR-H3, SYFISFFSFDY; and CDR-L3, SYDSGFSTV) was obtained by panning against myoglobin from horse skeletal muscle (Sigma). Both scFvs were subcloned via restriction sites *Xba*I and *Eco*RI into the expression plasmid pMX7 (*21*). The different mutations were introduced with the QuikChange site-directed mutagenesis kit from Stratagene according to the manufacturer's instructions. Multiple mutations were constructed by exchanging restriction fragments using unique *Xba*I, *Xho*I, *Bsa*BI, and *Eco*RI sites in the antibody. The final expression cassettes consist of a *phoA* signal sequence, a short FLAG tag (DYKD), and the scFv in the V_H6 domain−$(Gly_4Ser)_4$ linker−V_L domain orientation, followed by long FLAG tag (DYKDDDD) and a hexahistidine tag.

Expression and Purification. Thirty milliliters of dYT medium (containing 30 μg/mL chloramphenicol and 1.0% glucose) was inoculated with a single bacterial colony and shaken overnight at 25 °C. One liter of dYT medium (containing 30 μg/mL chloramphenicol and 50 mM K_2HPO_4) was inoculated with this preculture and incubated at 25 °C (5 L flask with baffles, 105 rpm). Expression was induced at an OD_{550} of 1.0 by addition of IPTG to a final concentration of 0.5 mM. Incubation was continued for 18 h, while the cell density reached an OD_{550} between 8.0 and 11.0. Cells were collected by centrifugation (8000*g* for 10 min at 4 °C), resuspended in 40 mL of 50 mM Tris-HCl (pH 7.5) and 500 mM NaCl, and disrupted by French press lysis. The crude extract was centrifuged (48000*g* for 60 min at 4 °C) and the supernatant passed through a 0.2 μm filter. The proteins were purified using the two-column coupled in-line procedure (*4*). With this strategy, the eluate of an immobilized metal ion affinity chromatography (IMAC) column, which exploits the C-terminal His tag, was directly loaded onto an ion exchange column. Elution from the ion exchange column was achieved with a 0 to 800 mM NaCl gradient. The constructs derived from scFv 2C2 were purified with a HS cation exchange column in 10 mM MES (pH 6.0) and those derived from 6B3 with an HQ anion exchange column in 10 mM Tris-HCl (pH 8.0). Pooled fractions were dialyzed against 50 mM sodium phosphate (pH 7.0) and 100 mM NaCl. Protein concentrations were determined with the OD_{280}. The soluble yield was normalized to a 1 L of bacterial culture with an OD_{550} of 10.

Gel Filtration Chromatography. Samples of purified scFvs were analyzed on a Superdex-75 column equilibrated with 50 mM sodium phosphate (pH 7.0) and 500 mM NaCl on a SMART system (Pharmacia). The samples were injected at a concentration of 5 μM in a volume of 50 μL, and the flow rate was 60 μL/min. Lysozyme (14 kDa), carbonic anhydrase (29 kDa), and bovine serum albumin (66 kDa) were used as molecular mass standards.

Equilibrium Denaturation Experiments. Fluorescence spectra were recorded at 25 °C with a PTI Alpha Scan spectrofluorimeter (Photon Technologies, Inc.). Slit widths of 2 nm were used both for excitation and emission. Protein/GdnHCl mixtures (1.6 mL) containing a final protein concentration of 0.5 μM and denaturant concentrations ranging from 0 to 5 M GdnHCl were prepared from the freshly purified protein and a GdnHCl stock solution [8 M, in 50 mM sodium phosphate (pH 7.0) and 100 mM NaCl]. Each final concentration of GdnHCl was determined by measuring the refractive index. After overnight incubation at 10 °C, the fluorescence emission spectra of the samples were recorded from 320 to 370 nm with an excitation wavelength of 280 nm. With increasing denaturant concentrations, the maxima of the recorded emission spectra shifted from ~340 to 350 nm. The fluorescence emission maximum was determined by fitting the fluorescence emission spectrum to a Gaussian function and was plotted versus the GdnHCl concentration. Protein stabilities were calculated as described previously (*22*, *23*). To compare scFv denaturation curves in one plot, the emission maxima were scaled by setting the highest value to 1 and the lowest to 0 to give normalized emission maxima.

Enzyme-Linked Immunosorbent Assay (ELISA). Myoglobin from horse skeletal muscle (Sigma) and peptide M18 coupled to transferrin (Jerini GmbH) at a concentration of 5 μg/mL in 50 mM sodium phosphate and 100 mM NaCl (pH 7.0) were coated overnight at 4 °C on Maxisorb 96-well plates (Nunc). Plates were blocked in 2.0% sucrose, 0.1% bovine serum albumin (Sigma), and 0.9% NaCl for 2 h at room temperature. After incubation of the samples at concentrations from 2 to 0.125 μM, bound scFvs were detected using an anti-tetra-His antibody (Qiagen) followed by an anti-mouse antibody conjugated with alkaline phosphatase (Sigma).

BIAcore Measurements. BIAcore analysis was performed using a CM5 chip (Amersham Pharmacia) with one lane coated with 2700 resonance units (RU) of myoglobin from horse skeletal muscle (Sigma), one coated with 2500 RU of peptide M18 coupled to transferrin (Jerini GmbH), and one blank lane as a control surface. Each binding−regeneration circle was performed at 25 °C with a constant flow rate of 25 μL/min with different antibody concentrations ranging from 5 to 0.08 μM in 20 mM HEPES (pH 7.0), 150 mM

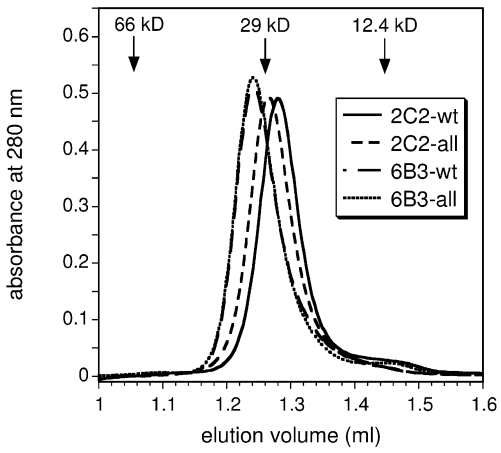

FIGURE 1: Analytical gel filtration of 2C2-wt, 2C2-all, 6B3-wt, and 6B3-all in 50 mM sodium phosphate (pH 7.0) and 500 mM NaCl on a Superdex-75 column at a concentration of 5 μM. 6B3-wt (− − −) and 6B3-all (•••) exhibit similar elution volumes. Arrows indicate elution volumes of molecular mass standards: bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome *c* (12.4 kDa). The mutations carried by 2C2-all and 6B3-all are listed in Table 1 and Figure 4.

NaCl, and 0.005% Tween 20 and 2 M NaSCN for regeneration. Determination of the antigen dissociation constant in solution was performed with a competition BIAcore experiment (*24*, *25*) with the same chip, buffer, and regeneration conditions. ScFvs at a constant concentration and variable amounts of antigen were preincubated for at least 1 h at 10 °C and injected in a sample volume of 100 μL. Data were evaluated by using BIAevaluation software (Pharmacia) and SigmaPlot (SPSS Inc.). Slopes of the association phase of linear sensorgrams were plotted against the corresponding total antigen concentrations, and the dissociation constant was calculated as described previously (*26*).

RESULTS

Properties of Wild-Type scFvs

We chose the V_H6 framework as the model system for testing our strategy for improving the biophysical properties by a structure-based design. We used two scFvs selected from the HuCAL as model systems: 2C2, which binds the peptide M18 coupled to transferrin, consisting of V_H6 paired with $V_\kappa3$, and 6B3, which binds myoglobin, consisting of V_H6 paired with $V_\lambda3$. The two antibodies differ in CDR-H3 (see Materials and Methods), but otherwise, the V_H6 sequence is identical. The wild-type (wt) scFvs 2C2 and 6B3 were expressed in the periplasm of *E. coli*. The scFvs were purified from the soluble fraction of the cell extract by immobilized metal affinity chromatography (IMAC), followed by an ion exchange column. The purity of the scFvs was greater than 98%, as determined by SDS−PAGE (data not shown). The yields of soluble protein after purification from a 1 L bacterial culture normalized to an OD_{550} of 10 of 2C2-wt and 6B3-wt were 1.2 ± 0.1 and 0.4 ± 0.1 mg, respectively. Approximately 10 and 25%, respectively, of the total amount of expressed protein was found in insoluble form, as determined by Western blot analysis. The oligomeric state was determined by analytical gel filtration. Both proteins elute with an apparent molecular mass of 29 kDa, indicating that they are monomeric (Figure 1).

The thermodynamic stability of each protein was measured by equilibrium GdnHCl denaturation. Unfolding of the scFvs was monitored by the shift of the fluorescence emission maximum as a function of denaturant concentration. The fluorescence spectra of the native and unfolded scFv are very similar to the ones published previously (*23*). They are characterized by a decrease in fluorescence upon unfolding at 320 nm and an increase at >340 nm. Since the fluorescence intensities of the native and unfolded molecule do not differ greatly, the use of the shift of the emission maximum reflects the molar fractions of the folded and unfolded protein sufficiently well (*27*). Furthermore, the use of the emission maximum leads to much less scattering in the data, since they are robust to concentration errors and any scattering particles. Since we verified that both intensity and emission maximum plots gave essentially the same transitions, we report only the latter.

Figure 2a shows the denaturation curve of 2C2-wt and 6B3-wt. Both curves show only one transition, indicating that V_H and V_L within the scFv denature simultaneously (*10*). Since the fluorescence intensities of the folded and unfolded states are similar, and the maximum changes by only 17 nm, the shift in maximum can be used to determine the population of unfolded molecules (*27*). Under the assumption that the unfolding of the scFvs is a two-state process, the free energy of unfolding (ΔG_{N-U}) can be determined (*22*, *28*). 2C2-wt exhibited a ΔG_{N-U} of 51.3 kJ/mol and 6B3-wt a ΔG_{N-U} of 42.4 kJ/mol with *m* values of 25.2 and 27.4 kJ mol^{-1} M^{-1}, respectively. These *m* values lie in the expected range for proteins of this size, indicating that both scFvs have the cooperativity expected for a two-state process (*30*).

Structural Rationale for the Selection of Mutations

The first set of mutants to improve the properties of scFvs 2C2 and 6B3 containing the human V_H6 framework was chosen from the analysis of the structural model, guided by the sequence alignment of the human consensus V_H domains grouped by V_H domains with favorable biophysical properties (families 1, 3, and 5) and V_H domains with less favorable properties (families 2, 4, and 6) (Figure 3). We focused on residues of the framework and excluded the CDR regions, since we aim to identify generically applicable mutations unlikely to affect antigen binding. The residues that we investigated in 2C2 and 6B3, together with the reasoning behind the specific changes, are as follows.

Q5V. In a selection experiment of the scFv 4D5Flu favoring stability, Val was selected at this position from Val, Gln, Leu, and Glu (*18*). Position 5 is part of the first β-strand, and Val has a higher β-sheet propensity than Gln (*31*). Nevertheless, it was shown previously that mutations of exposed hydrophobic residues have a profound effect on the in vivo folding yield (*9*). Figure 4 shows that Gln in position 5 of the model of a $V_H6-V_\kappa3$ scFv (*21*) [PDB entries 1DHZ (V_H6) and 1DH5 ($V_\kappa3$)] is exposed to solvent, and therefore, the hydrophilic residue Gln or Lys of V_H2, V_H4, and V_H6 might be thought to enhance folding efficiency in contrast to the hydrophobic Val in V_H1a, V_H1b, V_H3, and V_H5. In summary, this mutation increases the β-sheet propensity at the expense of creating an exposed hydrophobic residue.

S16G. V_H2, V_H4, and V_H6 carry a non-glycine residue with a conserved positive ϕ angle at position 16 in the loop of

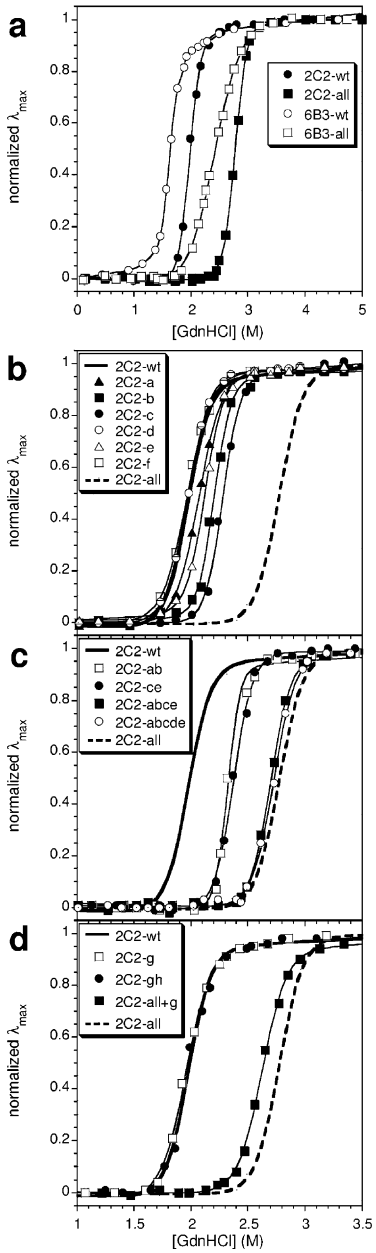

FIGURE 2: Overlay of GdnHCl denaturation curves of (a) 2C2-wt, 2C2-all, 6B3-wt, and 6B3-all, (b) single mutations (a, Q5V; b, S16G; c, T58I; d, V72D; e, S76G; f, S90Y; and all, abcdef), (c) multiple mutations to the consensus of V_H domains with favorable properties, and (d) mutations (g, P10A; and gh, P10A/V74F) to framework 1 subtype III exemplified with the scFv 2C2. In panels b−d, the bold solid line and the bold dotted line represent the fits (*23*) of the experimental data shown in panel a of 2C2-wt and 2C2-all, respectively. All unfolding transitions were measured by following the change in the emission maximum as a function of denaturant concentration at an excitation wavelength of 280 nm.

framework 1 (Figure 4), which probably causes an unfavorable local conformation. Structures that have been determined with a non-Gly residue at position 16 (e.g., PDB entries 1C08, 1DQJ, and 1F58) indeed show that the positive ϕ angle is locally maintained, apparently enforced by the surroundings. In contrast, the odd-numbered V_H domains all have Gly at this position.

T58I. The residue at position 58, which is the highly conserved Ile, points into the hydrophobic core (Figure 4). Only V_H6 has Thr at this position burying an unsatisfied hydrogen bond donor and acceptor. Therefore, this residue was changed to Ile.

V72D. Solvent-exposed residue 72 (Figure 4) was changed in the antibody McPC603 from Ala to Asp, which increased the ratio of protein found in the soluble periplasmic fraction compared to that in the insoluble periplasmic fraction 20-fold, but did not measurably alter the thermodynamic stability (*15*), indicating that it might have an effect on the folding efficiency. Only the consensus sequence of the most stable V_H family V_H3 has Asp at this position.

S76G. The odd-numbered V_H domains have Gly at position 76 in framework 2 (Figure 4) in contrast to the even-numbered V_H domains, which carry Thr or Ser. In half of the known antibody structures found in the Protein Data Bank, the residue at this position has a positive ϕ angle, indicating that glycine could be a better choice at this position.

S90Y. The semiburied position 90 (Figure 4) of V_H1a, V_H1b, V_H3, and V_H5 is occupied by Tyr, whereas V_H2, V_H4, and V_H6 have Val or Ser. This residue is part of the β-sheet of the immunoglobulin fold and is exchanged with Ser in V_H6, but Tyr has a higher β-sheet propensity than Ser (*31*).

In positions 20 and 88, group specific differences are seen, too (Figure 3). The residues in both positions are solvent-exposed and participate in a β-sheet. At position 20, the odd-numbered V_H domains have basic residues Lys and Arg, while the even-numbered domains have Thr or Ser. At position 88, all domains with favorable properties contain Thr and the domains with unfavorable properties contain Gln. However, as all these residues are hydrophilic and have similar β-sheet propensities, it might be expected that the differences in folding efficiency are small. Therefore, these residues were not exchanged.

Single Mutations

The six mutations (Q5V, S16G, T58I, V72D, S76G, and S90Y) described above were introduced into 2C2-wt and 6B3-wt by site-directed mutagenesis. All scFvs carrying one mutation were expressed and purified in a manner identical to that of wild type scFvs and were monomeric in solution (data not shown). In all single and subsequently constructed multiple mutants, the proportion of soluble to insoluble protein in the periplasm remained constant, even in those cases where the total expression level increased. The biophysical data are summarized in Table 1. To compare the improvements caused by the mutations in 2C2 and 6B3, the expression yield of soluble protein is normalized relative to the yield of the corresponding wild-type scFvs and the free energy of unfolding (ΔG_{N-U}) is given as the difference ($\Delta\Delta G_{N-U}$) from that the corresponding scFv-wt. The denaturant-induced unfolding curves are shown in Figure 2b.

Both single mutations exchanging the non-glycine residues with positive ϕ angles (S16G and S76G) increased the yield of soluble protein by a factor of approximately 2. The thermodynamic stability was also increased in both single mutations with $\Delta\Delta G_{N-U}$ values of 6.2 and 7.3 kJ/mol for 2C2-S16G and 6B3-S16G and $\Delta\Delta G_{N-U}$ values of 3.7 and 3.5 kJ/mol for 2C2-S76G and 6B3-S76G, respectively, compared to wild-type scFvs. The mutation to Gly in a loop region causes greater flexibility, which enables the optimal orientation of the antiparallel β-sheet, stabilizing the whole domain. The higher yield of these mutants is probably due

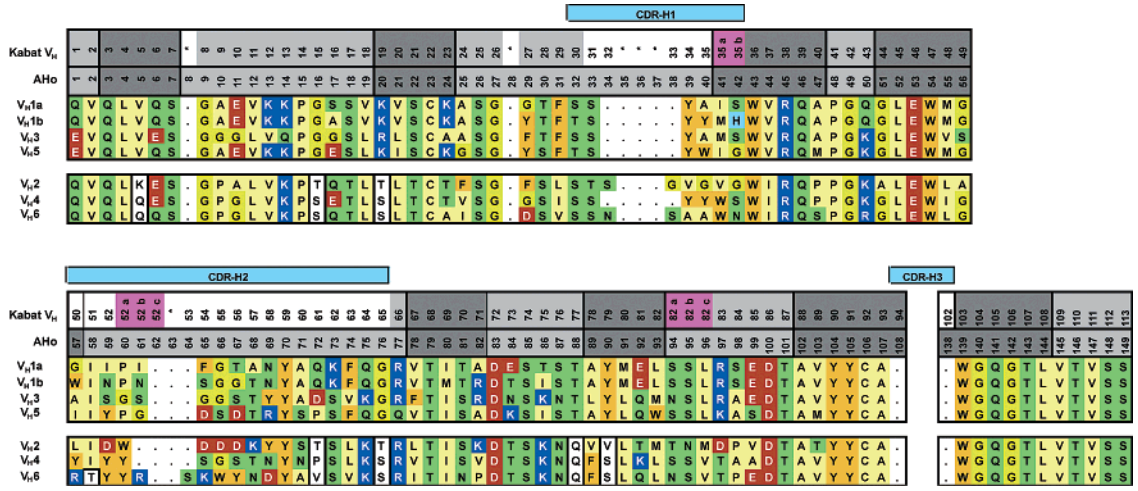

FIGURE 3: Aligned HuCAL V_H sequences. The amino acids are color-coded according to residue type: aromatic residues (Tyr, Phe, and Trp) being orange, hydrophobic residues (Leu, Ile, Val, Met, Cys, Pro, and Ala) yellow, uncharged hydrophilic residues (Ser, Thr, Gln, Asn, and Gly) green, acidic residues (Asp and Glu) red, and basic residues (Arg, Lys, and His) blue. Residues that show correlated sequence differences among the groups of V_H domains with favorable properties (V_H1a, V_H1b, V_H3, and V_H5) and V_H domains with less favorable properties (V_H2, V_H4, and V_H6) are indicated by white boxes. The numbering scheme is according to Kabat et al. (*41*) and Honegger and Plückthun (*33*).

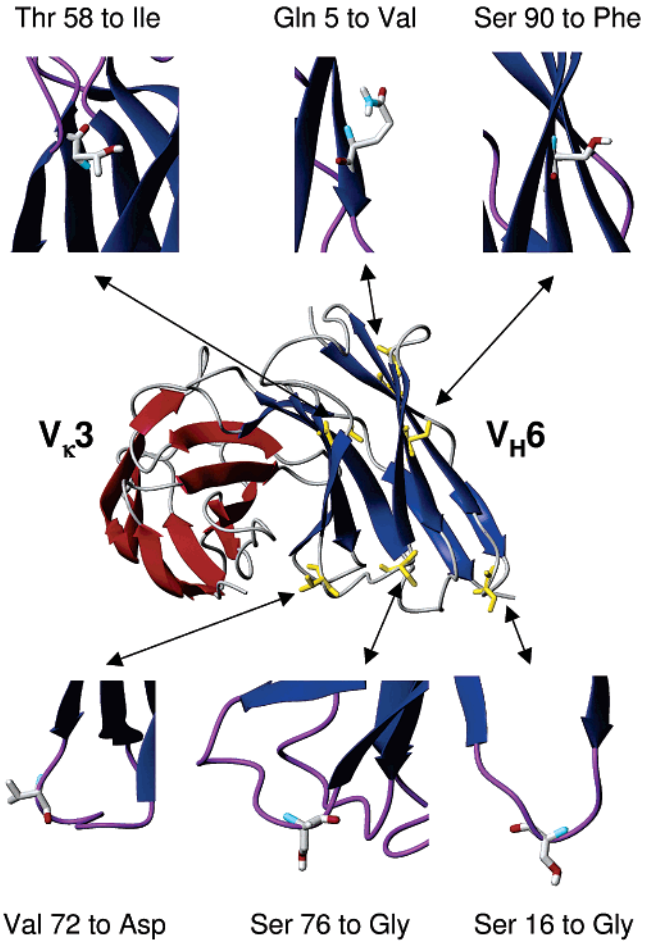

FIGURE 4: Overview of the single mutations to the consensus of those V_H domains with favorable properties. In the middle of the figure, a model scFv consisting of V_H6 (blue ribbon, PDB entry 1DHZ) and the $V_κ$3 domain (red ribbon, PDB entry 1DH5) is shown with the single mutations highlighted in orange. Arrows point to enlargements of the single mutations. All images were generated using the program MOLMOL (*42*). The numbering scheme is according to Honegger and Plückthun (*33*).

to the increased protease resistance and folding efficiency caused by the stabilized folded state of the protein.

Table 1: Summary of Yield and Stability Measurements

		yield (normalized to wt[a])		stability, $\Delta\Delta G_{N-U}$ (kJ/mol)[b]	
name	abbreviation	2C2	6B3	2C2	6B3
wild type		1	1	0	0
Q5V	a	1.7	2.6	2.4	2.9
S16G	b	1.8	2.3	6.2	7.3
T58I	c	1.0	0.9	7.9	6.8
V72D	d	3.2	1.8	0.1	2.2
S76G	e	2.1	1.5	3.7	3.5
S90Y	f	1.3	1.8	−0.1	1.4
	ab	1.8	3.5	9.8 (8.6)[c]	nd[d]
	ce	1.4	1.4	10.4 (11.6)	nd[d]
	abce	2.3	3.1	18.9 (19.6)	nd[d]
	abcde	3.3	3.7	19.5 (19.7)	nd[d]
all	abcdef	4.3	4.2	20.9 (19.6)	nd[d]
P10A	g	2.9	4.2	0.0	0.0
P10A/V74F	gh	1.9	1.7	0.5	0.4
all+P10A	abcdefg	3.5	2.1	16.8 (19.6)	nd[d]

[a] Yield of soluble protein after IMAC and ion exchange column, normalized to the yield of the respective wild-type scFvs 2C2 and 6B3. Absolute values are as follows: 1.2 ± 0.1 mg/L of bacterial culture of an OD_{550} of 10 for 2C2-wt and 0.4 ± 0.1 mg/L of bacterial culture of an OD_{550} of 10 for 6B3-wt. [b] Absolute values of the free energy of unfolding of wild-type scFvs: ΔG_{N-U} = 51.3 kJ/mol for 2C2-wt and ΔG_{N-U} = 42.4 kJ/mol for 6B3-wt. [c] In parentheses are the sums of the free energy contributions of the individual mutations to equilibrium stability. [d] Not determined because of low cooperativity (see the text for details).

The mutation of the OH-carrying Thr58 to Ile, pointing into the hydrophobic core, did not alter the yield of soluble protein but caused a marked increase in the thermodynamic stability with $\Delta\Delta G_{N-U}$ values of 7.9 and 6.8 kJ/mol for 2C2-T58I and 6B3-T58I, respectively. This remarkable improvement in stability is due to the additional van der Waals interaction of the hydrophobic Ile within the hydrophobic core and to the reduced energetic cost of desolvation. Interestingly, this mutation does not have an effect on the yield of soluble protein, indicating that the folding efficiency is not increased.

Both mutations exchanging a residue in a β-sheet to a residue with higher β-sheet propensity (Q5V and S90Y)

Table 2: Analysis of the Framework 1 Subtype

		subtype-defining residues[a]			subtype-correlated core residues[a]			
name	subtype	H6[b]	H7	H10	H19	H74	H78	H93
	I	Glu	Ser	Pro	Leu	Leu	Ala/Val/Ile/Leu	Leu/Met
	II	Glu	Ser	Gly	Leu	Val	Phe	Met
	III	Gln	Ser	any (Ala)[c]	Leu/Val	Phe	Ala/Val	Leu
wild type	III	Gln (100%)[d]	Ser (84%)[d]	Pro (8%)[d]	Leu (56%)[d]	Ile (1%)[d]	Ile (8%)[d]	Leu (63%)[d]
P10A	III	Gln	Ser	Ala	Leu	Ile	Ile	Leu
P10A/I74F	III	Gln	Ser	Ala	Leu	Phe	Ile	Leu

[a] According to ref *32*. [b] Using the numbering scheme of Honegger and Plückthun (*33*). [c] Ala is used in 76% of subtype III sequences (*32*). [d] Percentage use of the specified amino acid in subtype III sequences, regardless of V_H family (*32*).

resulted in an approximately 1.8-fold increase in the yield of soluble protein. In addition, the thermodynamic stability is slightly increased with the exception of that of 2C2-S90Y, the stability of which shows even a very small decrease in comparison to that of the wild-type scFv. The analysis of these constructs shows that mutations of residues, which participate in a β-sheet, to a residue with higher β-sheet building propensity can increase the yield of soluble protein due to a higher folding efficiency. The thermodynamic stability is also increased probably because of better orientation of the mutated residue, depending on the scFv, facilitating the orientation of stabilizing hydrogen bonds in the β-sheet.

The last single mutation exchanges a solvent-exposed hydrophobic residue with a hydrophilic one (V72D). The yields of soluble protein in 2C2-V72D and 6B3-V72D are increased 3.2- and 1.8-fold, respectively. The thermodynamic stability in 2C2-V72D is not changed, while in 6B3-V72D, it is slightly increased with a $\Delta\Delta G_{N-U}$ of 2.2 kJ/mol.

Multiple Mutations

To determine whether the improvements were additive, we introduced combinations of the single mutations. The scFvs with multiple mutations were expressed and purified as described above and were also monomeric in solution, as demonstrated by analytical gel filtration (2C2- and 6B3-all as examples in Figure 1). The denaturation curves of all multiple mutants of 2C2 that were tested showed one steep, cooperative transition (Figure 2d), indicating that the $V_\kappa 3$ domain is also stabilized with the help of the six mutations in $V_H 6$, probably because the mutated $V_H 6$ domain stabilizes $V_\kappa 3$ through the hydrophobic V_H-V_L interface interactions. In contrast, the transition of the equilibrium unfolding of the double mutants 6B3-Q5V/S16G and 6B3-T58I/S76G revealed a lower cooperativity compared to that of 6B3-wt and gave apparent *m* values of 18.9 and 19.3 kJ mol^{-1} M^{-1}, respectively, indicating that the unfolding is no longer a two-state process. The scFv 6B3 carrying all six mutations derived from the sequence comparison with the group of V_H domains with favorable properties (6B3-all) showed an even lower cooperativity and has an apparent *m* value of 14.3 kJ mol^{-1} M^{-1} (Figure 2a). The $V_\lambda 3$ domain, which has the lowest thermodynamic stability of the isolated V_L domains (*11*), probably starts to unfold first in the scFv 6B3 with multiple mutations, while the mutated, stabilized $V_H 6$ domain is still folded and only unfolds at higher denaturant concentrations. Because of this lack of two-state behavior, the ΔG_{N-U} values could not be calculated for the multiple mutants of 6B3.

The details of the yield of soluble protein and thermodynamic stability determinations are listed in Table 1. In summary, the effect on the stability of the single mutations is almost perfectly additive, while the effect on yield shows qualitative additivity. The scFvs carrying all six mutations, 2C2-all and 6B3-all, show an increase in yield of 4.3- and 4.2-fold, respectively, compared to that of wild-type scFvs. The absolute values for 2C2-all are a yield of 5.1 mg/L, which is 3.9 mg/L more than for 2C2-wt, and a thermodynamic stability of 72.3 kJ/mol. In the case of 6B3-all, a yield of 1.7 mg/L was obtained, which is 1.3 mg/L more than for 6B3-wt.

Analysis of the Framework 1 Subtype

V_H structures can be divided into four distinct framework 1 conformations depending on the type of amino acids at positions 6, 7, and 10 (*32*) [numbering scheme according to Honegger and Plückthun (*33*)]. Residues at positions 19, 74, 78, and 93, which are part of the hydrophobic core of the lower part of the domain and thus influence thermodynamic stability and folding efficiency, are correlated to this structural subtype (*32*). While the V_H domains with the most favorable properties fall into subtype II ($V_H 3$) and subtype III ($V_H 1a$, $V_H 1b$, and $V_H 5$), the V_H domains with less favorable properties ($V_H 2$ and $V_H 4$) fall into subgroup I. $V_H 6$, which we want to improve, can be assigned to subtype III which is defined by Gln at position 6 and the absence of Pro at position 7 (*32*). Analysis of subtype III-defining and -correlated residues of human V_H domains (*32*) shows that the $V_H 6$ fragment carries rarely used residues at positions 10, 74, and 78 (Table 2). Pro in position 10 is used in 8% of the sequences, whereas Ala is used in 76% of the sequences. Pro allows only a more limited number of conformations than Ala. In a mutagenesis experiment (*34*), Pro at position 10 was shown to destabilize a V_H domain in a subtype IV context (only occurring in murine and not in human sequences). Val at position 74 and Ile at position 78 have frequencies of 1 and 8%, respectively, compared to V_H subtype III sequences. Val74 was exchanged in 2C2 and 6B3 to the more frequently found Phe, as the bulky aromatic amino acid probably increases the packing density of the hydrophobic core. Ile78 was not exchanged with the subtype III consensus residues Ala and Val, which are, like Ile, aliphatic residues, as the effect on the packing density would probably be small. In Figure 5a, the framework 1 subtype-determining and -correlated residues are shown in the model of $V_H 6$ (*21*) (PDB entry 1DHZ), and in Figure 5b, the model of the double mutation is shown with P10A (Pro to Ala at position 10) and V74F.

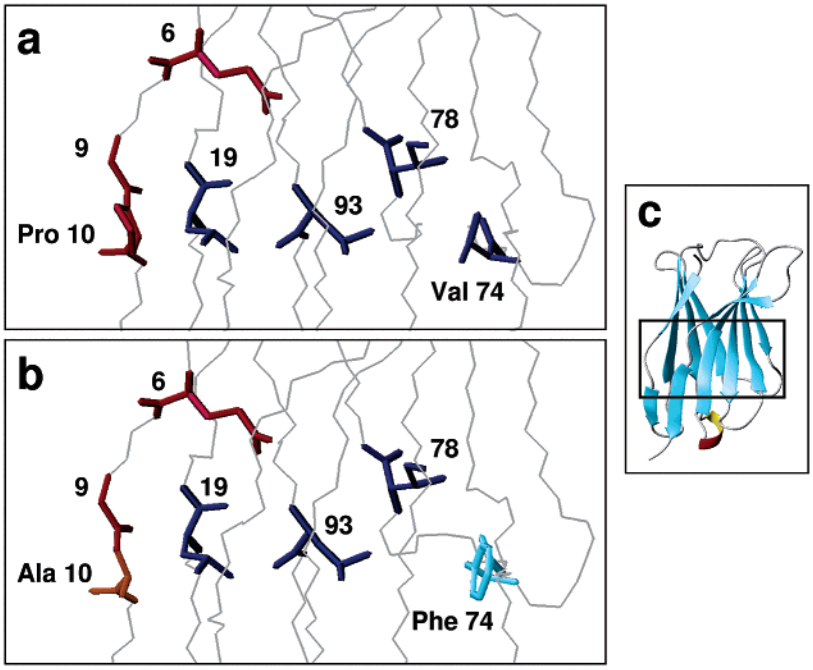

FIGURE 5: Overview of framework 1 subtype III-determining residues (6, 7, and 10) in red and correlated residues (19, 74, 78, and 93) in blue (a) in the wild-type V_H6 domain (PDB entry 1DHZ) and (b) in the model of the doubly mutated form with the P10A (orange) and V74F (cyan) mutations. (c) Ribbon representation of the V_H6 domain with the black frame indicating the enlarged area depicted in panels a and b. All images were generated using the program MOLMOL (*42*). The numbering scheme is according to Honegger and Plückthun (*33*).

The mutations to the framework 1 subtype III consensus P10A alone and in combination with V74F were introduced into the wild-type scFvs by site-directed mutagenesis. 2C2-P10A and 6B3-P10A exhibited 2.9- and 4.2-fold increases in the yield of soluble protein compared to that of the wild-type scFvs, respectively, while the double mutants with P10A and V74F exhibited smaller increases of 1.9- and 1.7-fold, respectively. All biophysical data are summarized in Table 1. The analysis of the soluble and insoluble periplasmic fraction of the single and double mutant showed that both the total expression level and the level of soluble protein increased with the mutations, and thus, the ratio between soluble and insoluble scFv remained constant (data not shown). The thermodynamic stability of the scFvs 2C2 and 6B3 is not increased by the mutation P10A, and is only slightly increased ($\Delta\Delta G_{N-U}$ values of 0.5 and 0.4 kJ/mol, respectively) with the P10A/V74F double mutation (Table 1 and Figure 2d). The biophysical analysis therefore shows that the mutation P10A indeed increases the folding efficiency, as demonstrated by the higher yield of periplasmic protein, but does not change the stability in comparison with that of the wild-type scFvs. In contrast, the mutation V74F may slightly increase the stability because of enhanced stabilizing interactions in the hydrophobic core, probably at the expense of folding efficiency, since the magnitude of the positive effect of P10A on yield is decreased in the double mutant.

Because the yield of the P10A single mutant was higher than that of the P10A/V74F double mutant, which showed only a small increase in thermodynamic stability, we cloned only the mutation P10A into 2C2-all and 6B3-all, resulting in the construct scFv-all+P10A. The yields compared to those of 2C2-all and 6B3-all were decreased 0.8- and 2.1-fold, respectively. In the case of 2C2-all+P10A, the thermodynamic stability with a ΔG_{N-U} of 68.1 kJ/mol was 4.1 kJ/mol lower than the stability of 2C2-all. The midpoint of denaturation, which is a semiquantitative measure for the thermodynamic stability, in 6B3-all+P10A was also at a GdnHCl concentration lower than the midpoint of 6B3-all.

Determination of Binding Activity

The goal of the study was to show that yield and stability of V_H6-containing scFvs can be improved by the structure-based approach, guided by the family specific analysis, while the binding activity is retained. We analyzed the binding activity with two independent methods: ELISA and BIAcore. For the ELISA, we coated the corresponding antigen and applied various concentrations of scFvs. We tested all single mutations, including scFv-P10A, and the multiple mutations scFv-all and scFv-all+P10A. All mutants exhibit similar concentration dependencies, which indicates that they have the same binding affinity (data not shown).

BIAcore experiments were performed with different concentrations of scFvs flowing over an antigen-coated chip. Panels a and b of Figure 6 show an overlay of 2C2-wt and -all and 6B3-wt and -all, respectively, plotted as resonance units (RU) versus time. The association and dissociation curves for binding of scFv-wt and -all to the antigen-coated chip can be superimposed in both cases, indicating that the binding is fully retained. However, the dissociation phase did not reach the background level before injection of scFvs, preventing unambiguous determination of the antigen dissociation constant (K_d). This unspecific binding was observed at different antigen-coating densities (2700 and 370 RU, data not shown). This indicates that this behavior is not due to rebinding on the chip but may be due to a small portion of partially unfolded scFv that sticks nonspecifically to the antigen-coated chip. Therefore, competition BIAcore experiments (*24*, *25*) were performed to determine the K_d in

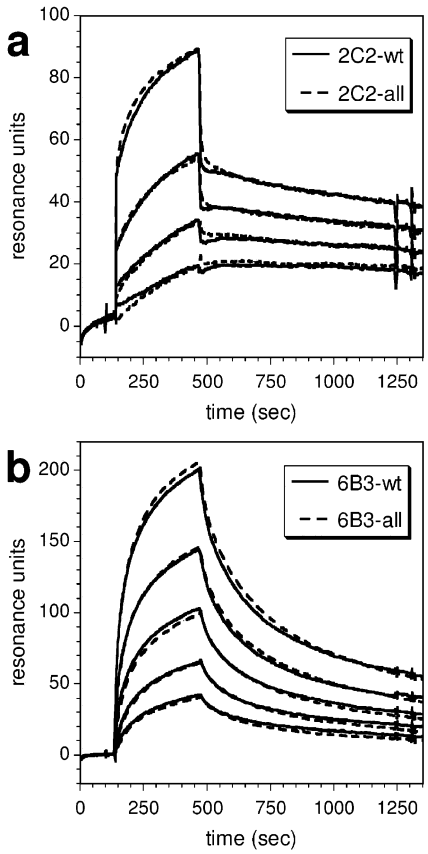

FIGURE 6: Comparison of the binding activities of (a) 2C2-wt and 2C2-all and (b) 6B3-wt and 6B3-all. The results of BIAcore experiments are shown, with resonance units plotted against time after injection of different scFv concentrations over an antigen-coated chip. Solid lines indicate wild-type scFvs, and dotted lines indicate scFvs carrying all six mutations toward the consensus of favorable V_H domains. In panel a 2C2-wt and 2C2-all at concentrations of 1.25, 0.63, 0.31, and 0.16 μM and in panel b 6B3-wt and 6B3-all at concentrations of 1.25, 0.63, 0.31, 0.16, and 0.08 μM are plotted.

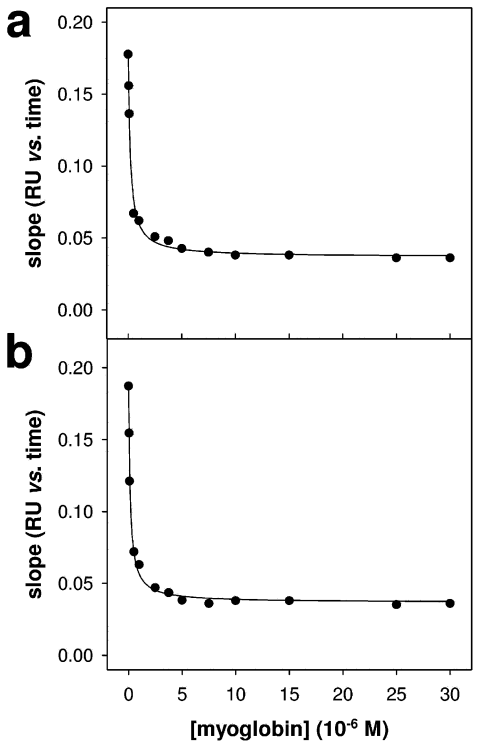

FIGURE 7: Competition BIAcore analysis of 6B3-wt and 6B3-all. (a) 6B3-wt (16 nM) and (b) 6B3-all (10 nM) were incubated with different concentrations of myoglobin for 1 h and injected over a myoglobin-coated sensor chip. From the linear sensograms, the slopes (resonance units vs time in seconds) were plotted against the corresponding total soluble antigen concentration. The slopes correlate to uncomplexed scFv in the injected solutions. The K_d was calculated from a fit according to ref *26*. Each point is the average of three independent measurements.

solution. In this experiment, the scFv protein was incubated with soluble antigen, and the mixture was injected onto a BIAcore chip containing immobilized antigen. Only the free scFv, but not the antigen-bound scFv, could bind to the antigen on the surface. Thereby, the dissociation constant in solution can be determined, independent of any unspecific binding events. From the previous experiments, K_d was estimated to be around 10^{-7} M. Therefore, competition BIAcore experiments were performed with 6B3-wt and 6B3-all at 16 and 10 nM, respectively, in the presence of different concentrations of myoglobin ranging from 50 nM to 30 μM. From a plot of the slope of the association phase versus the corresponding total antigen concentration in solution, the K_d of 6B3-wt was calculated to be $(1.9 \pm 0.5) \times 10^{-7}$ M and that of 6B3-all to be $(1.5 \pm 0.4) \times 10^{-7}$ M as described previously (*26*) (Figure 7). Both K_d values lie in the experimental error range, indicating that the binding is fully retained.

DISCUSSION

The aim of this study was to demonstrate the validity of the structure-based, family consensus-based predictions. We chose scFvs containing the human germline family V_H6 consensus domain as a model system for improving the expression yield of soluble protein and thermodynamic stability. Mutations that potentially improve these biophysical properties were identified from comparison of the residues which define the framework 1 subtype and other residues that interact with the consensus found within the same subtype. The next set of potential mutations was found by an analysis of the structure for potential imperfections, guided by a comparison to the consensus sequences of those V_H domains with known favorable biophysical properties (families 1, 3, and 5). We excluded CDR residues from this analysis. We could pinpoint such residues, as we had previously systematically determined the biophysical properties of consensus sequences of all human variable domain subgroups (*11*). The experiment shows that all seven proposed single mutations fall into three categories. They result in either an increase in the expression yield of soluble protein, in thermodynamic stability, or both. This distinction helps us understand the role of these residues in determining the biophysical properties of this protein. In the case of scFv 2C2 three and in the case of scFv 6B3 even five of these seven mutations result in an improvement in both biophysical properties. These results illustrate that the combination of structure-based analysis, guided by family alignments, is a powerful way of improving the properties of immunoglobulin variable domains. Since our analysis (*11*) covers all human families, we have now a general strategy for this task.

The analysis of different combinations of the single mutations to the consensus of V_H domains with favorable

properties showed that the improvements in free energy were almost perfectly additive, indicating that they act independently. The mutant with the highest yield and thermodynamic stability compared to the wild-type scFvs is indeed the mutant with all six mutations. In the case of scFv 2C2, the properties of the best mutant are comparable to the properties of a model scFv consisting of the most stable V_H domain, V_H3, and the same V_L domain, $V_\kappa3$, with a different CDR-L3, which was part of the systematic biophysical characterization of human variable antibody domains (*11*), indicating that it is indeed possible to turn an antibody with unfavorable properties into one with very favorable properties by changing only a few residues. Most importantly, the CDRs and those framework residues which are important for binding are maintained.

The addition of the mutation P10A to the scFvs carrying six mutations decreases both the expression yield and thermodynamic stability, although in the wild-type scFvs this mutation increased the soluble yield 2.9-fold in the case of 2C2-P10A and 4.2-fold in the case of 6B3-P10A and left the thermodynamic stability unchanged. The mutations Q5V and S16G, which are close to position 10, should still be beneficial to the V_H6 framework as they are independent of the type of amino acid at position 10. The reason for the deleterious effect of this mutation in the context of the improved framework is currently unclear.

The improvements seem to be independent of the V_L domain and of the sequence and length of CDR-H3, as 2C2 with $V_\kappa3$ and 6B3 with $V_\lambda3$ and different CDR-H3 loops gave similar results. There were only two minor exceptions, as the thermodynamic stability of the 6B3 mutants V72D and S90Y is slightly increased, while in 2C2 no stability increase could be observed. It was shown previously that V_λ domains, in contrast to V_κ domains, are able to form very stable V_H-V_L interfaces in the scFv, increasing the stability of the whole scFv even above the intrinsic stabilities of the isolated domains (*11*). The residue at position 72 is not involved in the interface interactions but is in the proximity (Figure 4). It is therefore possible that the mutation V72D may lead to a small change in the orientation of the interface, which has no effect on $V_\kappa3$ domains in 2C2 but a small stabilizing effect through the interface interactions with the $V_\lambda3$ domain of 6B3. The residue in position H90 is on the side of the domain opposite the interface to V_L (Figure 4) and also 29 residues away from CDR-H3, indicating that the slightly increased stability of 6B3 is probably not due to the different V_L domain and CDR H3 sequences compared to 2C2.

Although we did not exchange residues of the CDR with possible direct contact to the antigen, the possibility that changes in the framework might affect the orientation of the CDRs and, thereby, antigen binding could not a priori be excluded. Therefore, we experimentally determined the binding properties. In the case of the mutations that have been examined, antigen binding was fully retained as demonstrated by three independent methods.

In this study, we show that it is possible to rationally transform antibody frameworks with less favorable properties into those with very favorable properties while retaining their binding activity and the binding characteristics of the framework. It could be argued that an easier approach would be to directly use the very stable V_H3 framework with a suitable V_L domain. Nevertheless, framework residues can affect the orientation of CDRs, can be part of the hapten-binding cavity located in the V_H-V_L interface, and build the "outer loop", which was seen in some cases to be involved in antigen binding. These "framework" residues can thereby contribute greatly to affinity and diversity, and it is unlikely that a single framework can provide the ideal solution in all cases. Therefore, we believe that the preferred approach to achieving a structurally diverse library of stable frameworks is to optimize the human consensus antibody frameworks further in the way we presented here, as it would give access to a whole range of stable scaffolds covering all natural families. A part of this analysis has been implemented in HuCAL Gold (Morphosys AG, www.morphosys.com).

In this study, we focused on the improvement of the V_H6 framework. However, because of the sequence similarity, five of the mutations that have been studied (Q5V, S16G, V72D, S76G, and S90Y) should give similar results for V_H domains belonging to family V_H2 and V_H4. While this approach is useful for the design of antibody libraries, in many cases given human antibodies, e.g., from transgenic mice (*35*, *36*), obtained by humanization (*37*) or by phage display from a library of natural sequences (*38*−*40*) may also benefit from improvement. To improve the properties of any given human immunoglobulin heavy chain variable domain while keeping binding activity, we suggest the following procedure. The first task is to compare each residue of the given domain to different subsets of immunoglobulin sequences. As the binding activity should be retained, residues of CDR-H1 (H25−H40), CDR-H2 (H57−H77), CDR-H3 (H109−H137), and the outer loop (H84−H87) are not considered [numbering scheme according to Honegger and Plückthun (*33*)]. After determination of the framework 1 class, the subtype-determining (6, 7, 9, and 10) and subtype-corresponding (19, 74, 78, and 93) residues are compared to the consensus of sequences falling into the same class (*32*). The other residues are then compared to the consensus sequences of the V_H domains with favorable properties (families 1, 3, and 5) (*11*, *21*). In the second task, the differences are analyzed using structure models (PDB entries given above). Mutations, which increase the expression yield of soluble protein and/or thermodynamic stability, as seen in this study, include mutations which replace a non-glycine residue in a loop with a positive ϕ angle to glycine, mutations of residues in a β-strand with low β-sheet propensity to a residue with high β-sheet propensity, and mutations of solvent-exposed hydrophobic residues to hydrophilic ones and replacement of residues with unsatisfied H-bonds.

This procedure is in principle also applicable to V_L domains. However, the identification of differences is more difficult as V_L domains fall mainly in one major framework 1 class (*32*) and the biophysical properties differ to a smaller extent than between V_H domains (*11*). However, V_κ domains should be compared to the consensus sequence of $V_\kappa3$, as this domain displays the highest thermodynamic stability and expression yield of V_κ domains. Giving advice for the reference domain for V_λ domains is more difficult, as $V_\lambda1$ showed the highest thermodynamic stability but $V_\lambda2$ the highest expression yield of soluble protein. Nevertheless, the physical principles for rational design are the same as with V_H domains described above.

These results also show that some human germline genes do not encode an optimal version of the protein, with regard to its biophysical properties. Since the biophysical properties of natural domains cover a wide range, it cannot be argued that limited stability is a desirable property for the immune system. Rather, the stability of V_H2, V_H4, and V_H6 may simply be good enough to be tolerated by the immune system. For those biomedical or biotechnological applications where it is not good enough, however, we have now provided a pathway for improving these properties in a straightforward way.

REFERENCES

1. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Lee, T., Pope, S. H., Riordan, G. S., and Whitlow, M. (1988) Single-chain antigen-binding proteins, *Science 242*, 423−426.
2. Glockshuber, R., Malia, M., Pfitzinger, I., and Plückthun, A. (1990) A comparison of strategies to stabilize immunoglobulin Fv-fragments, *Biochemistry 29*, 1362−1367.
3. Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R., and Oppermann, H. (1988) Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A. 85*, 5879−5883.
4. Plückthun, A., Krebber, A., Horn, U., Knüpfer, U., Wenderoth, R., Nieba, L., Proba, K., and Riesenberg, D (1996) in *Antibody Engineering, A Practical Approach* (McCafferty, J., Hoogenboom, H. R., and Chiswell, D. J., Eds.) pp 203−252, Oxford University Press, New York.
5. Shusta, E. V., Raines, R. T., Plückthun, A., and Wittrup, K. D. (1998) Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments, *Nat. Biotechnol. 16*, 773−777.
6. Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H., and Pedersen, J. T. (1994) Antibody design: beyond the natural limits, *Trends Biotechnol. 12*, 199−206.
7. Wörn, A., and Plückthun, A. (2001) Stability engineering of antibody single-chain Fv fragments, *J. Mol. Biol. 305*, 989−1010.
8. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature 416*, 507−511.
9. Nieba, L., Honegger, A., Krebber, C., and Plückthun, A. (1997) Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment, *Protein Eng. 10*, 435−444.
10. Wörn, A., and Plückthun, A. (1999) Different equilibrium stability behavior of ScFv fragments: identification, classification, and improvement by protein engineering, *Biochemistry 38*, 8739−8750.
11. Ewert, S., Huber, T., Honegger, A., and Plückthun, A. (2003) Biophysical properties of human variable antibody domains, *J. Mol. Biol. 325*, 531−553.
12. Bothmann, H., and Plückthun, A. (1998) Selection for a periplasmic factor improving phage display and functional periplasmic expression, *Nat. Biotechnol. 16*, 376−380.
13. Bothmann, H., and Plückthun, A. (2000) The periplasmic *Escherichia coli* peptidylprolyl cis,trans-isomerase FkpA. I. Increased functional expression of antibody fragments with and without cis-prolines, *J. Biol. Chem. 275*, 17100−17105.
14. Kipriyanov, S. M., Moldenhauer, G., Martin, A. C., Kupriyanova, O. A., and Little, M. (1997) Two amino acid mutations in an anti-human CD3 single chain Fv antibody fragment that affect the yield on bacterial secretion but not the affinity, *Protein Eng. 10*, 445−453.
15. Knappik, A., and Plückthun, A. (1995) Engineered turns of a recombinant antibody improve its in vivo folding, *Protein Eng. 8*, 81−89.
16. Forsberg, G., Forsgren, M., Jaki, M., Norin, M., Sterky, C., Enhorning, A., Larsson, K., Ericsson, M., and Björk, P. (1997) Identification of framework residues in a secreted recombinant antibody fragment that control production level and localization in *Escherichia coli*, *J. Biol. Chem. 272*, 12430−12436.
17. Sieber, V., Plückthun, A., and Schmid, F. X. (1998) Selecting proteins with improved stability by a phage-based method, *Nat. Biotechnol. 16*, 955−960.
18. Jung, S., Honegger, A., and Plückthun, A. (1999) Selection for improved protein stability by phage display, *J. Mol. Biol. 294*, 163−180.
19. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., and Plückthun, A. (2001) Tailoring in vitro evolution for protein affinity or stability, *Proc. Natl. Acad. Sci. U.S.A. 98*, 75−80.
20. Steipe, B., Schiller, B., Plückthun, A., and Steinbacher, S. (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain, *J. Mol. Biol. 240*, 188−192.
21. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A., and Virnekäs, B. (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides, *J. Mol. Biol. 296*, 57−86.
22. Pace, C. N., and Scholtz, J. M. (1997) in *Protein Structure, A Practical Approach* (Creighton, T. E., Ed.) pp 299−321, Oxford University Press, New York.
23. Jäger, M., Gehrig, P., and Plückthun, A. (2001) The scFv fragment of the antibody hu4D5-8: evidence for early premature domain interaction in refolding, *J. Mol. Biol. 305*, 1111−1129.
24. Karlsson, R. (1994) Real-time competitive kinetic analysis of interactions between low-molecular-weight ligands in solution and surface-immobilized receptors, *Anal. Biochem. 221*, 142−151.
25. Nieba, L., Krebber, A., and Plückthun, A. (1996) Competition BIAcore for measuring true affinities: large differences from values determined from binding kinetics, *Anal. Biochem. 234*, 155−165.
26. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., and Plückthun, A. (1998) Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries, *Proc. Natl. Acad. Sci. U.S.A. 95*, 14130−14135.
27. Eftink, M. R. (1994) The use of fluorescence methods to monitor unfolding transitions in proteins, *Biophys. J. 66*, 482−501.
28. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants, *Biochemistry 27*, 8063−8068.
29. Jäger, M., and Plückthun, A. (1999) Domain interactions in antibody Fv and scFv fragments: effects on unfolding kinetics and equilibria, *FEBS Lett. 462*, 307−312.
30. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
31. Zhu, Z. Y., and Blundell, T. L. (1996) The use of amino acid patterns of classified helices and strands in secondary structure prediction, *J. Mol. Biol. 260*, 261−276.
32. Honegger, A., and Plückthun, A. (2001) The influence of the buried glutamine or glutamate residue in position 6 on the structure of immunoglobulin variable domains, *J. Mol. Biol. 309*, 687−699.
33. Honegger, A., and Plückthun, A. (2001) Yet another numbering scheme for immunoglobulin variable domains: An automatic modeling and analysis tool, *J. Mol. Biol. 309*, 657−670.
34. Jung, S., Spinelli, S., Schimmele, B., Honegger, A., Pugliese, L., Cambillau, C., and Plückthun, A. (2001) The importance of framework residues H6, H7 and H10 in antibody heavy chains: experimental evidence for a new structural subclassification of antibody VH domain, *J. Mol. Biol. 309*, 701−716.
35. Fishwild, D. M., O'Donnell, S. L., Bengoechea, T., Hudson, D. V., Harding, F., Bernhard, S. L., Jones, D., Kay, R. M., Higgins, K. M., Schramm, S. R., and Lonberg, N. (1996) High-avidity human IgG kappa monoclonal antibodies from a novel strain of minilocus transgenic mice, *Nat. Biotechnol. 14*, 845−851.
36. Mendez, M. J., Green, L. L., Corvalan, J. R., Jia, X. C., Maynard-Currie, C. E., Yang, X. D., Gallo, M. L., Louie, D. M., Lee, D. V., Erickson, K. L., Luna, J., Roy, C. M., Abderrahim, H., Kirschenbaum, F., Noguchi, M., Smith, D. H., Fukushima, A., Hales, J. F., Klapholz, S., Finer, M. H., Davis, C. G., Zsebo, K. M., and Jakobovits, A. (1997) Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice, *Nat. Genet. 15*, 146−156.
37. Winter, G., and Harris, W. J. (1993) Humanized antibodies, *Trends Pharmacol. Sci. 14*, 139−143.

38. Hoogenboom, H. R., and Winter, G. (1992) By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro, *J. Mol. Biol. 227*, 381−388.
39. 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., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E., and Winter, G. (1994) Isolation of high affinity human antibodies directly from large synthetic repertoires, *EMBO J. 13*, 3245−3260.
40. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J., and Johnson, K. S. (1996) Human antibodies with subnanomolar affinities isolated from a large non-immunized phage display library, *Nat. Biotechnol. 14*, 309−314.
41. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S., and Foeller, C. (1991) in Sequences of Proteins of Immunological Interest, NIH Publication 91-3242, National Technical Information Service, Springfield, VA.
42. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 29−32, 51−55.

BI026448P