Determinants of the assembly and function of antibody variable domains

V28

Eva Maria Herold[1,2*], Christine John[1*], Stephan Kremser[3], Jonathan Eras[4], Alexander Bepperling[1,5], Sabrina Deubler[1], Martin Zacharias[3] and Johannes Buchner[1,6]

[1] Center for Integrated Protein Science Munich (CIPSM) at the Department Chemie, Technische Universität München, 85747 Garching, Germany

[2] present address: Pharmaceutical Biotechnology, University of Applied Sciences Biberach, 88400 Biberach, Germany

[3] Center for Integrated Protein Science Munich (CIPSM) at the Physics Department, Technische Universität München, 85747 Garching, Germany

[4] present address: ETH Zürich, Otto-Stern-Weg 5, 8093 Zuerich Switzerland

[5] present address: Hexal AG, 82041 Oberhaching, Germany

[*]these authors contributed equally

[6]address correspondence to: johannes.buchner@tum.de

Tel. +49 89 289 13341

Fax +49 89 289 13345

Abstract

Antibodies are modular proteins. The Fv fragment is the most important module for antigen binding and consists of the two variable domains, V_L and V_H. They exhibit the conserved antibody fold beta sheet structure and comprise highly variable loops (complementarity determining regions, CDRs) for antigen binding. Despite the importance of their association to form the antigen binding site, little is known about the contributions of the framework residues and CDRs on organizing this functional unit. To address this question, we exchanged conserved interface residues as well as CDR loops and tested the effects on structure, stability, association and antigen binding.

Our results show that the affinity between the two MAK33 variable domains studied is not very high, with a K_D of 0.2 µM. For the V_L domain, the exchange of conserved residues had only slight effects on structure, stability and functionality, except for a conserved proline residue which was crucial for both V_H-V_L association and antigen binding. For the V_H domain however, all the conserved residues analyzed exhibited a strong influence on the properties of this domain. The most crucial V_H residue for V_H-V_L association and antigen binding was a conserved leucine. The CDRs modulate the domain framework to a significant extent as shown by swap experiments. This includes effects on antigen binding and domain association, which interestingly, do not always correlate. Thus, the framework region is not necessarily the determining factor for V_H-V_L association, the CDRs also contribute. Vice versa, antigen binding is also influenced by framework mutations. Taken together, for the architecture and function of the variable domains, both, the nature of the interface and the interplay with the CDR loops are of crucial importance.

Introduction

In the humoral immune response, antigen recognition is mediated by immunoglobulins, specifically by the N-terminal variable domains of the light chain (V_L) and of the heavy chain (V_H) which associate noncovalently to form the so-called Fv-fragment. Three hyper-variable regions (complementarity determining regions or CDRs) in V_L and V_H comprise the residues interacting with antigens. They account for approximately 25% of the variable domains [1]. CDR-H3 (i.e., the third CDR of V_H) is the most diverse of these six regions concerning length and amino acid sequence [2]. Apart from the CDRs, both variable domains exhibit a conserved β-barrel framework stabilized by an internal disulfide bridge [1-4]. They are composed of two ß-sheets, one with four strands (A,B,D,E) and one with six strands (A',G,F,C,C',C''), with the strands GFC'C involved in the V_H-V_L interface. The interface ß-sheets are additionally twisted, leading to a three-layer packing. The residues from the edge strands are the central part of the interface [1]. Several studies on the V_H-V_L packing geometry showed that residues within the framework as well as interface contributing residues of the CDRs can influence the interface [1, 3, 5-7]. 75 % of the interface residues are constituted by framework ß-sheets and 25 % by the hypervariable loops (inter-strand links between GF, BC and C'C'', respectively) [1, 3] especially the so-called proximate zone situated at the base of the antigen binding site and comprising residues that do not actively participate in the interface. [1, 3, 5-7]. As the association of V_L and V_H is crucial for antigen binding [3], understanding the underlying principles is of great importance.

For the V_H domain which is composed of about 125 residues and the V_L domain with about 110 residues, Chothia and co-workers suggested that the interface residues at positions 98, 44 and 36 in V_L and 103, 47, 45 and 37 in V_H according to Kabat numbering [8] are conserved [1, 5]. Wang and co-workers [9] aimed at identifying amino acid networks important for V_H and V_L function by covariation analysis. This multiple sequence alignment approach investigates covariations between residues at all possible positions. This allows to reveal conserved amino acids by the correlation of the presence of one particular amino acid with the presence of a second one at a particular sequence position. In their study they included more than 2000 V-class sequences of human, mouse, cow, camel, llama, macaque and chicken with a bias of human sequences (574 out of 2432). Generally, the majority of the most strongly

conserved amino acids identified in this study were positioned at the V_H-V_L interface [9]. For V_L, amino acids Y36, Q37, P44, A43, L46 and F98 were found to be highly conserved, with all residues except Q37 directly in contact with V_H. On the V_H side, amino acids V37, R38, G44, L45, E46, W47 and W103 could be identified in the V_H-V_L interface with all residues except E46 and R38 in direct contact with V_L. V_H W47 seems to be the central node based on the number and strength of its covariations with other interface residues, the same holds true for Y36 and P44 for the V_L domain. Further computational analyses revealed two main modes of interaction for V_H and V_L which is either characterized by a proline or a medium/large hydrophobic residue at position 44 in V_L [7]. Concerning V_H, W47 seems to be essential as it was previously mutated for improved solubility and stability but none of the mutations (W47L, W47R [10]) were favorable. ~~To produce a stable monomeric V_H, three highly conserved hydrophobic interface residues in V_H [44, 45, 47], were exchanged by hydrophilic residues as commonly found in camelid V_HHs [11-14].~~ Interface residues can also affect antigen binding [3, 7, 15-17] by influencing the positioning of hypervariable loops. Additional experimental studies addressed the influence of the exchange of particular conserved amino acids on the association of V_H and V_L via the stabilities of covalently linked scFv and Fab fragments [18-22]. While these studies lay the groundwork, we are still far from a detailed and comprehensive understanding of the organization of the Fv interface. In this context, it is important to determine the affinity of the association of V_L and V_H directly. Strikingly, for the other domain interactions in IgG the K_D values differ by several orders of magnitude. For the C_H3 dimer a $K_D < 10^{-10}$ M was determined using SEC [23, 24] and the K_D for the interaction between C_H1 and C_L was 6.2 µM [25]. This analysis is largely lacking for the V_L/V_H interaction.

Here, we chose to use the V_L and V_H domains of the murine monoclonal antibody MAK33 (κ/IgG1 subclass) as a well-studied model system [16, 18, 30, 32] to analyze the contribution of different factors on V_L and V_H structure and function. Their sequences contain all the conserved residues identified by Wang and coworkers except for an alanine at position 43 in V_L, which is exchanged to serine in MAK33. Since the relative importance of the conserved residues for structure, stability, association and antigen binding is not clear, we mutated every conserved residue against alanine and analyzed the properties of the variants in a comprehensive manner. Additionally, we generated two V_L double point mutations (Y36A/P44A,

Y36A/S43A) to investigate the potential synergistic nature of the mutation. Importantly, we focused in the analysis on the isolated variable domains and the direct influence of point mutations on their interaction, and not on Fab or scFv fragments as in previous studies [16, 20, 22] to draw conclusions concerning their stabilities and antigen binding properties. Furthermore, we performed CDR exchange experiments to address the contribution of these structural elements on domain architecture.

Our results on the effects of mutations on domain structure, stability, association and antigen binding together with CDR exchange experiments reveal complex relationships between structural and functional properties within the V_L and V_H domains.

Results

The association of V_H and V_L is particularly sensitive to mutations in V_H

To determine the influence of specific residues on the association of V_L and V_H, amino acids were selected for mutation which had been predicted to be important [9]. Based on the results of Wang and coworkers we created six single (Y36A, Q37A, S43A, P44A, L45A, and F98A) and two double point mutants for V_L (Y36AP44A and Y36AS43A), as well as seven single point mutations for V_H (V37A, R38A, R44A, L45A, E46A, W47A, and W103A). As shown in figure 1A, these residues lie in or near the interaction interfaces of V_L and V_H. For R38A and E46A in V_H, it turned out that the variants were unstable and aggregation-prone. Therefore, they were not considered further.

In previous studies, the association of the V_L and V_H domains was analyzed using scFvs or Fab fragments [16, 20, 22]. This is a relatively indirect approach, as the domains were either artificially linked (scFvs) or two additional covalently linked domains were present (C_L and C_H1 in the Fab). Here, we used assays that report directly on the formation of the V_L and V_H heterodimer. With sedimentation equilibrium analytical ultracentrifugation (AUC) experiments, a K_D of 0.2 ± 0.05 µM was determined for the wild type (wt) V_L and V_H domains (table 1). Hence, the variable domains interact more strongly than C_H1 and C_L [25], but have a lower affinity compared to the C_H3 homodimer [23, 24].

For all V_H mutants studied, a decrease in the binding affinity was observed. According to the K_D values determined by AUC, W47A and R44A have a 6 - 8 fold higher K_D compared to V_H wt. V37A, L45A and W103A showed a marked decrease in affinity with a K_D that was two orders of magnitude higher and for L45A the interaction with V_L was hardly detectable, resulting in a K_D of ~ 300 µM (table 1). Surprisingly, the mutation of residues in V_L had only minor effects on the affinity for V_H. Only V_L P44A (figure 1; highlighted in red) with a K_D of 15.2 µM strongly decreased the interaction between V_L and V_H. (table 1, figure 1 B; orange). Concerning the two double mutations (Y36A/P44A, Y36A/S43A) which were generated to investigate whether the effects observed for single mutants are additive, unexpected results were obtained. Y36A/S43A, which is the combination of two point mutations with no change in K_D, leads to the expected wt like K_D. But the point

mutation P44A with the highest K_D for the association with V_H (15.2 µM), exhibited improved binding when combined with Y36A, suggesting compensatory effects (table 1).

In conclusion, with the exception of P44, replacement of each of the conserved residues of V_L does not affect dimerization with the V_H domain. In contrast, most of the conserved V_H residues showed a clear influence on association.

The role of conserved residues for variable domain structure and stability

Besides influencing the interaction between the variable domains, the conserved residues might also play a role in the structures of the individual domains. To address this, we recorded CD spectra. The characteristic shape of the V_L FUV-CD spectrum [26] was observed for all the alanine exchange mutants tested (figure 2 A), with only small deviations in the amplitude observed for some variants (e.g. F98A, P44A or Y36A). Thus, the β-sheet structure of the V_L domain seems to tolerate single point mutations of conserved residues. NUV-CD spectra confirm this conclusion (figure 2 C). In the case of the V_H domain, point mutants exhibited more pronounced effects on the FUV-CD spectra (figure 2 A and B) including shifts in the minimum (e.g. W47A). The same holds true for the NUV-CD spectra (figure 2 C and D), suggesting major changes in the tertiary structure. This supports the notion that the structure of V_H is less tolerant than V_L against the exchange of conserved residues.

To determine whether the stability of the respective domain was affected by the mutation of conserved residues, denaturant-induced (GdmCl) unfolding transitions were measured (figure 3, table 1). All transitions were fitted to a two-state model. The V_L domain showed a midpoint for GdmCl-induced unfolding of 1.17 ± 0.68 M GdmCl. Strikingly, V_L S43A was slightly more stable than the wt against GdmCl-induced unfolding (figure 3 A and C, table 1). However, the variants V_L Y36A and V_L L46A exhibited decreased $D_{1/2}$ values, compared to the wt. The cooperativities of these unfolding transitions are in a comparable range for all V_L variants, from 15.8 ± 8.6 to 33.7 ± 18.6 kJ $mol^{-1}M^{-1}$ (figure 3, table 1). For the least stable V_L point mutation Y36A, two double mutations (Y36A/P44A, Y36A/S43A) were generated to investigate the potential synergistic nature of the observed effects. Indeed when incorporating the most stable mutation, S43A, the stability of Y36A/S43A increased slightly

compared to Y36A. For Y36A/P44A, where P44A alone has wt-like stability, almost no difference in stability was visible compared to Y36A (figure 3, table 1).

For V_H wt, we determined a midpoint for GdmCl-induced unfolding of 0.21± 0.2 M GdmCl and thus this domain is much less stable than V_L (figure 3 B and D, table 1). The analysis of the V_H mutants showed that the stability of this domain is very sensitive to manipulations of conserved residues. It is conspicuous that all mutants already start to unfold in the presence of low GdmCl-concentrations (figure 3 B). The point mutations V_H W47A and V_H V37A showed the largest decrease in stability (table 1), while V_H L45A, and V_H W103A were slightly less stable than the wt and V_H G44A was as stable as the wt. The cooperativity of unfolding for different V_H mutations was subject to variation, from 15.3 ± 10.5 to 75.2 ± 19.4 kJ $mol^{-1}M^{-1}$ suggesting that structural changes occurred [27].

As there is no correlation between the cooperativity values and the structural data obtained by CD measurements (figure 2) this assumption cannot be confirmed.

In summary, the stability of the V_H domain is more sensitive to mutations than that of the V_L domain. Our analysis identified W47A for V_H and Y36A for V_L as the least stable point mutations. Interestingly, V_L S43A was even more stable than the wt protein. For the double mutants, an additive effect on the stability could be observed.

Antigen binding of V_H and V_L is influenced by conserved interface residues

To test how the mutation of conserved amino acids influences antigen recognition by the Fv fragment, we set up an ELISA for the MAK33 Fv-fragment and human creatine kinase as its antigen. To this end, the respective wt domain was produced with a FLAG-tag at the C-terminus for detection with an anti-FLAG antibody. The tag does not negatively influence stability, folding and the interaction between the variable domains (data not shown). In this ELISA, a concentration-dependent increase in signal is only observed when V_L and V_H are combined (figure 4). As the assay involves several protein interactions, only conclusions on an apparent K_D seem reasonable (figure 4, table 1). For the wt Fv with either V_L or V_H tagged with FLAG similar K_{Dapp}s were determined (data not shown). Antigen binding was found to be abolished for V_H W47A. Apart from that, the V_H mutation L45A led to the strongest reduction in the affinity for the antigen with a 12 fold increased K_D compared to wt, while V37A and W103A exhibited a less pronounced decrease with a 2 fold higher K_D

(table 1). In general, for the V_H point mutations a low antigen binding activity correlates with a high K_D for association of the Fv-fragment (see table 1). However, W47A, showed only a small decrease in the K_D but no binding to the antigen.

For the V_L domain, there is in most cases no apparent correlation between the K_D for association and antigen binding. The V_L mutations L46A, F98A, P44A, as well as the double mutant Y36AP44A have a negative influence on antigen binding (table 1), with a 5 – 12 fold increased K_D. As already observed for V_H/V_L association, P44A exhibited the weakest binding with a 12 fold higher K_D for the antigen. Interestingly, the double mutant Y36A/P44A showed only 50% of the impairment of P44A alone, so there must be a compensating effect of the Y36A mutation. This coincides with the data for the V_H/V_L association (table 1), where P44A exhibited the worst K_D while the double point mutation Y36A/P44A showed a 2 fold higher affinity than P44A alone. Surprisingly, V_L L46A and F98A, which exhibited an affinity for the V_H domain similar to the wt, were defective in antigen binding (table 1). Consequently, the analysis of V_L mutants supports the assumption that the affinity between V_H and V_L is not necessarily correlated with the ability to bind the antigen.

The CDR regions affect domain structure, stability and association

The CDRs of the antibody variable domains are elements of natural variations. How variations in these elements affect their association, structure and stability is therefore of special interest to obtain a comprehensive picture of the factors shaping the Fv-fragment. To address this question, we switched CDRs between MAK33 and unrelated variable domains. We chose human variable domain consensus sequences [28] with CDRs of similar length. For V_L, we selected the 1DH5 domain, and for V_H 1DHU. These human variable domains represent a class of variable domains with a highly stable structure [29].

The K_Ds for the association of grafted variants with wt domains were determined by AUC. For the association of the MAK33 V_H domain containing the 1DHU CDRs (1DHU_MAK V_H) with MAK33 V_L, a K_D of 0.4 µM was obtained. This corresponds to the value determined for the MAK33 wt domains (table 1). Wt 1DHU V_H was an insufficient binding partner for MAK33 V_L - with a K_D of 8.8 µM. Interestingly, the interaction could be improved by grafting the MAK33 CDRs into the 1DHU framework. This chimera (MAK_1DHU V_H) bound to MAK33 V_L with a K_D of 1.0 µM.

These observations lead to the conclusion that while for V_H the framework region is important for the interaction between MAK33 V_L and V_H, the CDRs can exhibit a marked influence. Concerning the V_L grafting constructs, the observations are different: the MAK33 V_L construct containing the 1DH5 CDRs (1DH5_MAK) binds to MAK V_H with a roughly 10 fold lower affinity than wt V_L which is similar to the 1DH5 wt value. In contrast, grafting the MAK33 CDRs onto the V_L-1DH5 framework (MAK_1DH5) gave a wt-like K_D of 0.3 µM. So for the V_L domain, the CDRs and not the framework are the determining factor for the affinity towards V_H wt.

The FUV and NUV spectra of the grafting mutants gave a similar picture. The CDR exchange (1DH5_MAK) showed a FUV-CD spectrum similar to 1DH5 V_L while the spectrum of MAK_1DH5 was different from both wts (figure 5 B). So already on the secondary structure level the CDRs seem to be structurally important. The NUV-CD spectra exhibit a similar pattern (figure 5 D) but here the difference in the number of aromatic amino acids, predominantly concerning the CDRs, could also play a role.

For the V_H domain, both grafting mutants showed similar FUV-CD spectra (figure 5 C). The same was observed for the NUV-CD spectra of the grafting mutants compared to the wt domains; all spectra were similar in shape with variations in amplitude (figure 5 E). As for V_L the observed NUV deviations might also be due to the different numbers of aromatic amino acids. For both domains the CDRs of MAK33 and 1DHU/1DH5 differ by one tryptophan and several tyrosins. Generally for MAK V_H, the CDR exchange does not exhibit the same impact as for the V_L domain.

GdmCl-induced transitions of the different V_L domains showed that, in comparison, the MAK33 V_L domain is least stable against chemical denaturation (figure 6 A). As expected [19], the most stable domain was 1DH5 with a $D_{1/2}$ of 2.39 ± 1.21 M. The grafted mutants showed stabilities in-between MAK33 V_L and 1DH5. Interestingly, the stability of the MAK33 V_L framework was increased when the CDRs were exchanged against the CDRs of 1DH5. 1DH5_MAK V_L was about 20% more stable than MAK33 V_L (figure 6 D). On the V_H side, 1DHU showed the highest stability. The grafting mutants were again in-between (figure 6 B). When the CDRs of 1DHU V_H were transplanted into MAK33, the chemical stability increased, however the transition was less cooperative compared to MAK33 V_H (figure 6 B). Exchanging only CDR H3 in MAK33 V_H led to a slight increase in stability compared to the wt (figure 6 B). The exchange of the CDRs of 1DHU against MAK33 CDRs decreased its stability. In

summary, the analysis of the chemical stabilities allowed the same conclusion for V_L and V_H: the CDRs influence the stability of the variable domains decisively.

When antigen binding of the chimera was analyzed, the exchange of the MAK33 CDRs with that of the human consensus sequences led to a complete abolishment in both cases, as expected. For the grafting of MAK33 CDRs on the human variable domains, an interesting picture emerged. V_L (MAK_1DH5) has a slightly higher K_D of 0.3 µM for creatine kinase whereas there is almost no antigen binding detectable for the V_H grafting domain MAK _1DHU (> 50 µM) (table 1). So for binding of the V_H domain to the antigen not only the CDRs represent a determining factor.

As CDR-H3 (i.e. the third CDR of V_H) is the most flexible of these six regions concerning length and amino acid sequence [2], additionally a mutant carrying a swapped CDR-H3 loop was analyzed. Concerning the V_H/V_L association the CDR-H3 mutation shows a wt-like K_D of 0.4 µM. Interestingly, this construct, MAK CDR-H3 1DHU V_H showed a strongly impaired binding to the antigen with a 10 fold increased K_D (table 1). This demonstrates the importance of CDR-H3 for the antigen binding of MAK33.

Molecular Dynamics Simulations reveal mutation-induced structural alterations

For a subset of V_H and V_L mutations explicit solvent Molecular Dynamics (MD) simulations were performed for the heterodimer and for the individual domains. The set of simulations included point mutations that are associated with a significant reduction in V_H/V_L association (V_L P44A, Y36A/P44A; V_H V37A, L45A and W103A) and, as a control, also substitutions that showed only modest effects on complex affinity (V_L Y36A, S43A; V_H R44A, W47A). Simulations were started from the geometry of the wt structure (pdb-entry:1FH5, see Methods for details). On the time scale of the simulations, none of the V_H/V_L complexes dissociated and the root-mean-square deviation (RMSD) of the complexes did not exceed 0.3 nm from the start structure (figure S1). However, some of the mutations (e.g. V_L P44A, V_H R44A, V_H L45A and V_H W47A, see figure S1) resulted in overall larger final RMSDs compared to the wt indicating mutation-induced structural alterations and increased conformational freedom. Interestingly, simulations of the mutated proteins in the isolated state showed no significant differences in the RMSD (figure S2) except for V_H W103A (see below). The structural distortion of the V_L/V_H complexes due to some

of the mutations is also reflected in overall larger root mean square fluctuations of heavy atoms (RMSF) with respect to the mean structure. For the isolated mutated protein partners no significant difference to the wt was observed (figure S3), again with the exception of W103A.

However, in the complex the mutations with reduced binding affinity (e.g. V_L P44A, V_H R44A,V37A,L45A,W47A) showed increased fluctuations in regions at and near the binding interface but also in loops involved in antigen binding (figure S4, S5). Interestingly, especially for the substitutions that caused the largest drop in affinity between the V_L and the V_H domains an increased solvation at the interface (diffusion of water molecules into the space created by the introduction of a small Ala residue) was observed (illustrated in figure 7, table 2). Especially for V_L P44A, V_H L45A and V_H W47A, the average number of water molecules increased near the mutation site (table 2). The mutation V_H W103A resulted in significant changes of the backbone conformation around the mutation site, specifically the loop formed by residues 93-107 (figure 7), explaining the larger RMSD and RMSF observed for the isolated V_H domain in this case (figure S2 and S3).

Besides of the effect of the mutations on the binding interface, it is interesting to investigate the changes in mobility of the CDR loops involved in antigen binding. We compared the fluctuation pattern observed in the complexes and in the individual (isolated) V_L and V_H partner domains. Even in case of the wt, the RMSF pattern changes significantly in several regions that include regions involved directly in binding the partner domain but also regions involved in antigen binding (figure S3, S4, S5). For example, the antigen binding loop V_H: 93-107 shows large fluctuations in the absence of the V_L binding partner (figure S3) which drop significantly in the complex (figure S4). Hence, complex formation of the V_L and V_H domains appears to lock some of the antigen binding loops into distinct conformations. This effect is qualitatively also observed for several mutants, however, for some loop regions the reduction of CDR loop mobility upon binding is smaller compared to the wt. This is especially seen for the V_H: 93-107 region.

In addition to simulations of point mutations, we also studied a subset of the loop exchange constructs (MAK_1DH5, 1DH5_MAK, MAK_1DHU and 1DHU_MAK). In experiments these variants affected the binding affinity between V_H and V_L domains much less than some of the interface point mutations (see above). During simulations

on the time scale of 100 ns, these variants did not show significant differences of the calculated RMSF compared to the wt (figure S6).

Discussion

The relationship between structure, stability and binding affinity of V_H and V_L is still enigmatic. This is an important aspect for understanding antibody architecture both as the basis of our immune system and also in the context of the engineering of antibodies for therapeutic purposes. In this context, it was found that in mutants an increase in affinity is often accompanied by a decrease in stability and *vice versa* - and these consequences are difficult to predict [30-36]. In our study we analyzed the association of the variable domains for the first time directly. This allowed us to specifically determine the contribution of framework and CDR mutations on the interaction of V_H and V_L. For the wt domains, a K_D of 0.2 µM was determined. Consequently, the two domains interact more efficiently than C_H1 and C_L, but not as efficient as the C_H3 homodimer [23, 25]. This relative weak interaction (in the absence of the C_H1 and C_L domains) necessitates the covalent linkage of Fv fragments via a peptide and thus creating a pseudo-monomeric fusion protein (scFv) which can be used as a therapeutic agent [37, 38]. Simulations of scFvs have shown that the stability of the interface between the two variable domains plays a critical role for the overall stability of an antibody (or fragment) as dissociation preceeds unfolding [39].

Of special interest is the nature of the domain interface. It has to support the association of the two domains but also allow accommodating different CDRs and their repositioning in the context of antigen binding. Thus, an individual interface residue may be involved in one or more of these processes: 1) formation of the immunoglobulin fold, 2) domain stability, 3) interaction between the variable domains, or 4) antigen binding. Our alanine scanning study of conserved residues allowed us to address each of these possibilities and differentiate between them. Consistent with the results of *in silico* analysis which showed that only very few residues (< 10) are important for adopting an immunoglobulin fold [40], the alanine mutants in the V_L domain had no considerable influence on its secondary or tertiary structure. In contrast, V_H is very sensitive to the exchange of conserved residues in the interface. Two V_H interface residues, E46 and R38, were identified to be essential for the folding of the V_H domain. In the covariation analysis there was a very high ϕ-value for these two residues. In the structure, a salt bridge is formed between them. R38 is buried and E46 does not interact with other V_H interface residues, but might

electrostatically affect V_L binding [9]. Camelids and cartilaginous fish possess naturally occurring heavy chain antibodies lacking the light chain [41-43]. Interestingly a sequence alignment of MAK33 V_H with the variable domain (V_{HH}) of the camelid VHH (PDB entry 2XT1) and the variable domain of “monomeric” shark IgNAR (V_{NAR}) (PDB entry 2I24) shows a match for the residues E46 and R38. Concerning the shark IgNAR these two residues are actually the only ones from the conserved network investigated in this study that can also be identified at corresponding positions of this otherwise highly divergent sequence. For camelid V_{HH}s a mutation of the hydrophobic V_H/V_L interface residues (including the tetrad: V37, G44, L45, W47) in favor of hydrophilic ones was discovered [9, 43]. The increase of hydrophilic residues in the framework also holds true for the V_{NAR} which probably evolved from a cell surface receptor [42, 44].Both antibodies show high biophysical stability and their distinct structural patterns have by now been successfully applied to generate monomeric human V_H domains [13, 45, 46].

Mutations of conserved V_L residues had predominantly less impact than observed for V_H residues. As expected, S43A behaved similar to the wt. This is the only amino acid position of the MAK33 antibody which does not fit to the conserved amino acid network identified by Wang and coworkers [16]. In their covariation analysis an alanine is the conserved amino acid at position 43.

Generally, a change in stability did not necessarily coincide with a change in functionality. For V_L, the proline residue at position 44 showed the most prominent effects with an impaired V_H/V_L association as well as antigen binding. Surprisingly, V_L L46A and F98A exhibited an affinity for the V_H domain similar to the wt but they are clearly less efficient in binding the antigen (table 1). The double point mutations Y36A/P44A and Y36A/S43A indicate that the analyzed mutations do not necessarily act in an additive manner. In terms of stability, the stable mutation S43A is indeed able to improve the low stability of Y36A but, unexpectedly, for the V_H-V_L association, Y36A in combination with the worst binder P44A doubles the affinity for V_H compared to P44A alone. This coincides with the fact that these two residues are supposed to interact [7] and are the most important V_L interface residues in the covariation analysis based on number and strength of covariations with other interface residues [8].

The biophysical properties of isolated V_H domains are in general more affected by mutations compared to V_L [29]. The MAK33 V_H domain shows a low stability with a $D_{1/2}$ value of 0.21 M GdmCl compared to 1.17 M GdmCl for MAK 33 V_L. So already small changes in stability can shift the balance for the isolated V_H domain. For almost all of the V_H point mutations, an increased K_D for association with the V_L domain could be detected, which correlated in most cases with an impaired binding of the antigen. But V_H W47A, for example, showed no antigen binding while the V_H/V_L association was only slightly decreased. In terms of antigen binding, it has to be kept in mind that additionally to the stabilizing effect of the V_L domain, antigen binding itself will stabilize the heterodimer. So there is an additional layer of stabilization. To conclude, for the MAK33 V_H domain, in isolation some of the conserved V_H/V_L interface residues are critical for structure formation and stability. This might cause the impaired association with the V_L domain. In general, association and antigen binding do not necessarily correlate.

It seems that during antibody biogenesis the effect of CDRs on the stability of V_H domains is a decisive, so far underappreciated factor. Especially concerning the observed MAK33 V_H instability the outcome of grafting experiments with stable human consensus sequences was interesting. The grafting constructs revealed that CDRs, in addition to antigen binding, affect variable domain structure strongly. This is especially true for CDR-H3 (table 1). Comparing the MAK33 and 1DHU/1DH5 CDRs, CDR- H3 differed most, in terms of length (14 amino acids for MAK33 versus 11 for 1DHU) as well as charge. According to Morea and coworkers, CDR H3 conformation does not only depend on the environment [2] but both CDRs can additionally be assigned to different conformation types. Since MAK33 possesses a lysine at position 94 and an aspartate at position 101, which can form a salt bridge it is assumed to have a bulged conformation in contrast to 1DHU. As the CDR-H3 also contributes to the interface and interacts with the V_L domain this conformational difference might not only affect antigen binding. In our case, the exchange of the MAK33 CDR-H3 impairs antigen binding (10 fold increased K_D) and slightly increases domain stability. The data for the grafting constructs shows a very different picture for the V_H and V_L domain. For V_L, the CDRs seem to be important for structure and V_H/V_L association but concerning antigen binding also within the 1DH5 framework the affinity is wt like. The V_H domain, though, did not exhibit such a strong CDR

dependence, for V_H/V_L association, only the framework was the determining factor. However, antigen binding was almost not detectable when grafting the CDRs to the 1DHU framework. Since exchanging only CDR-H3 of the V_H MAK33 domain leads to a severe impairment of the antigen binding, this indicates a crucial role of the V_H domain and especially the CDR-H3 in the binding process. An explanation for these observations could be the mentioned differences between MAK33 and 1DHU V_H CDR H3. Interestingly, when grafting either the CDRs or framework from the human consensus sequences on MAK33 V_L or V_H ,always led to an increase in stability. This might be an important aspect for CDR selection and the interplay with domain stability. Previous studies [15, 33] applying CDR graftings for antibody humanization approaches showed the importance and complexity of the influence of specific framework residues in the context of antigen binding and stability improvement. But they focused on stabilization by framework exchange. Here we could show that vice versa, CDRs themselves can be considered as a crucial determinant of stability.

Our MD simulations indicate that the mutations altered conformational fluctuations of the isolated mutated domains which cause structural and mobility changes at the binding interface. The non-optimal packing at the protein-protein interface leads to increased fluctuations at the interface which is also manifested in a reduction of interactions (reduced binding affinity) and also to fluctuations in the antigen binding loop regions which can reduce the binding affinity for antigens. This is in line with previous findings where it was shown that subtle changes in the interface can affect the affinity for antigen [47].

In some cases also increased solvation at the interface was observed. This was especially the case for the mutations with the most impaired V_H/V_L association (V_L P44A, V_H L45A). The presence of water molecules at the interface reduces intermolecular contacts between protein partners by giving the interface an increasing non-specific character. Thus, the experimentally observed changes in binding affinity and stability of the mutations are due to a combination of effects.

A quantitative correlation with the experimentally observed change in antigen binding affinity due to the mutation is, however, not observed. It should be emphasized that this is also not expected since binding to the antigen is affected by each CRD loop differently and increased or decreased loop mobility can in principle cause not only

reduction but possibly also an increase in antigen binding affinity (greater adaptability to the antigen) in an unpredictable way.

Taken together our data indicate that multiple determinants regulate the V_H-V_L association and the affinity for the antigen. The interplay between interface interactions and CDRs turned out to be complex and in consequence V_H/V_L association and antigen binding do not necessarily correlate.

Acknowledgements

This study was supported by grants of the DFG to J.B. and the SFB1035 to M.Z.

Material and Methods

Unless otherwise stated, all experiments were carried out at 25°C. Measurements were performed in 50 mM sodium phosphate buffer at pH 7.5 *Cloning and protein expression:*

V_H and V_L were cloned into pET28 A (Novagen, Darmstadt,Germany) with NcoI and HindIII (NEB, Hitchin, UK) and expressed in E. coli BL21 star (Invitrogen, Carlsbad, USA). The transformed cells were grown in LB medium containing kanamycin at 37°C until an OD_{600} of 0.6-0.8 was reached. The expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). After 12 h, cells were harvested, and preparation of inclusion bodies was carried out as described previously[48]. The purification was performed according to the procedure described for the V_L domain[26] .

Single point mutations were introduced by a quick change PCR approach using the QuikChange® Site-Directed Mutagenesis Kit (Agilent Technologies Inc., Santa Clara, USA) according to the manufacturer's recommendations. Primers were ordered from MWG Operon (Ebersberg, Gemany).

Intact protein was verified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

CD and Fluorescence spectroscopy:

CD measurements were carried out using a Jasco J-720 spectropolarimeter (Jasco, Grossumstadt, Germany) equipped with a Peltier element. Far-UV CD spectra were measured using 10 µM protein in 1.0-mm quartz cuvettes between 260 nm and 198 nm and near-UV CD spectra between 320 nm and 260 nm using 50 µM protein in 5-mm quartz cuvettes. The spectra were accumulated 16 times and buffer corrected.

For denaturant-induced unfolding transitions, structural changes were monitored by fluorescence spectroscopy at 355 nm. Excitation wavelength was 280 nm and slit widths were 1 nm (excitation) and 3 nm (emission) for V_H and 2 nm and 5 nm for V_L, respectively. All measurements were performed with 1 µM protein in a 1-cm quartz cuvette. The samples were incubated overnight at 20°C at the different GdmCl concentrations prior to measurements.

Data evaluation was performed with Origin 8G (OriginLab, Northampton, USA); for GdmCl-transitions a two-state model was applied[49] .

Analytical ultracentrifugation:

Analytical ultracentrifugation was carried out with a ProteomLab XL-I (Beckman, Krefeld, Germany) equipped with absorbance and interference optics. The measurements for sedimentation velocity experiments were performed as described previously [50].

For sedimentation equilibrium experiments samples of V_H or V_L wild type together with the binding partner were prepared in a 1:1 ratio within a concentration range of 0.5 µM - 7.5 µM and incubated on ice during AUC cell preparation. Afterwards samples were loaded into six channel epon centerpieces (12 mm window) and the equilibrium was measured at different velocities (25.000, 30.000, 34.000 or 42.000 rpm). Scans were performed at 280 nm, 250 nm or 230 nm to be in absorption range within 0.1-0.8. 25 replicates were recorded. Data was globally fitted to determine the association constant with a heterogeneous interaction model (1:1 complex, see Equation 1) which is based on the mass action law $C_{AB} = K_a c_A c_B$ with the programme Origin 8.6 (OriginLab, Northampton, MA). Three different velocities and concentrations were included in the fit.

Equation 1:

$$a\,(r) = C_A(r_0)\varepsilon_A d\; exp\left[M_{b,A}\;\frac{\omega^2}{2RT}\;(r^2 - r_0^2)\right] + C_B(r_0)\varepsilon_B d\; exp\left[M_{b,B}\;\frac{\omega^2}{2RT}\;(r^2 - r_0^2)\right] +$$

$$K_a C_A(r_0) C_B(r_0)(\varepsilon_A + \varepsilon_B) d\; exp\left[(M_{b,A} + M_{b,B})\,\frac{\omega^2}{2RT}(r^2 - r_0^2)\right]$$

ELISA:

Binding of the variable domains of MAK33 to the antigen creatine kinase was analyzed by ELISA. Assay components and microwell plates were from Roche (Mannheim, Germany). Samples were prepared in 10 µl volume. Different mutants were tested within a concentration range of 100 nM up to 50 µM against wild type V_L or V_H with a C-terminal Flag-tag for detection. After the addition of 90 µl reaction mix I, the sample was incubated in a streptavidine-coated microwell plate to immobilize human biotinylated creatine kinase. Incubation was performed for 45 minutes with constant agitation at 20°C or 10°C for V_L and V_H mutants, respectively. Afterwards,

the samples were washed with sterile pure water for three times. Then 100 µl/well of reaction mix II were added together with the detection antibody for the Flag-tag coupled to horseradish peroxidase in a 1:15.000 dilution. Afterwards, the samples were washed three times with water again and then 100 µl of reaction mix III was added. The product of the enzymatic reaction was monitored at 405 nm in a GENios plate reader (Tecan, Männedorf, Switzerland) for 0.5 - 3 hours until a plateau was reached

Molecular dynamics simulations:

Start structures of MAK33 V_L/V_H complexes as well as individual V_L and V_H domains were obtained by extracting the corresponding coordinates from the crystal structures 1FH5 [51]. Missing residues were added/corrected using the program PyMol [51] with a final sequence corresponding exactly to the wild type sequence used in the experiments. Model start structures of all mutants were generated based on the wild type structure with residue substitutions generated *in silico*. All Molecular Dynamics (MD) simulations as well as the analysis of root-mean square deviation (RMSD) and fluctuations (RMSF) were performed using the Gromacs4.6 package [52, 53] in combination with the AmberSB99_ILDN force field [54]. Proteins were solvated in dodecahedral boxes including explicit ions (Na^+ and Cl^-) and explicit (TIP3P) water molecules [55]. The simulation systems were first energy-minimized (until the maximum force was smaller than 500 kJ/mol) followed by heating up to 300 K at a constant volume with position restraints on the protein. Subsequently, a pressure equilibration at 1 bar with position restraints on the protein was carried out. All production simulations were performed at a temperature of 300 K and a pressure of 1 bar and extended to 100 ns. Root mean square deviation (RMSD) and root mean square fluctuations (RMSF) with respect to the mean structure were calculated with g_rms and g_rmsf modules of Gromacs. Snapshots were created using VMD [56].

Table and figure legends

Table 1: Characteristics of MAK33 V_H and V_L point mutants and grafting mutants.

K_D values were determined by AUC sedimentation equilibrium experiments. Three concentrations in the range between 0.5 µM and 7.5 µM with V_H and V_L in a 1:1 ratio were measured at three velocities and globally fitted to a heterogeneous interaction

model between two components. All mutants were tested with the wild type V_H or V_L domain. Experiments were performed at 20°C; asterisks indicate that samples were measured at 15°C due to low thermal stability. Errors of the fits for AUC data were in a range between 6% - 50%. Stabilities against the chemical denaturation (GdmCl) of the V_L and V_H mutants. Even though most of the GdmCl-induced unfolding transitions were not reversible, data were evaluated according to a two-state equilibrium unfolding model to derive the midpoint of transitions ($D_{1/2}$), as well as the cooperativity parameter (m), for a qualitative comparison of the data. To determine the functionality of the mutants, an ELISA with isolated variable domains and the antigen creatine kinase was performed. Detection of the V_H and V_L wild type domains was possible via an introduced FLAG-tag at the C-termini. As a reference, the signals of the wild type domains vs. FLAG-tagged domains after 25 minutes of incubation were chosen. Samples were corrected for the signal of the single FLAG-tagged variable domains. Apparent K_D values were obtained by a Boltzmann fit. Experiments were performed at 20°C. NB indicates that no binding could be observed.

Table 2: Average number of water molecules around a point mutation

Water molecules within 7 Å of the C-beta atom of a mutated amino acid were determined during a 100 ns MD simulation. The standard deviation of the number of waters is given in parenthesis.

Figure 1: Conserved residues within the interface between V_L and V_H.

In (A) the positions of conserved amino acids in the Fv fragment of MAK33 are shown. The V_L domain is depicted in light green with the three CDRs highlighted in dark green. V_H is shown in light blue and the CDRs are in dark blue. Conserved residues within the interface are illustrated as spheres and color-coded. On the right, the region within the rectangle is enlarged. For a better orientation CDR-H2 is indicated. The labelled residues were selected for an alanine-exchange mutational analysis. In (B) the top views of the interacting residues of the V_L (left) and V_H (right) domain of MAK33 are shown. Six residues were selected for V_L (Y36, Q37, S43, P44, L46, and F98) and seven for V_H (V37, R38, R44, L45, E46, W47, and W103). Structures are modified from PDB ID 1FH5. In (C), the top-views for V_L and V_H are shown and the residues labelled according to their influence on association. The numbering of the strands (a, b, c, c', c'', d, e, f, g) is also shown. Color-code: orange

= medium influence, red = strong influence; dark green and dark blue = no influence for V_L or V_H mutants, respectively. The grey dotted line indicates CDR-H3 which is not resolved in PDB ID 1FH5.

Figure 2: Secondary and tertiary structure of V_L and V_H alanine exchange mutants.

In (A) and (B), FUV-CD spectra of V_L mutants (A) and V_H mutants (B) are shown. Color code for V_L: Y36A is red, Q37A dark cyan, S43A blue, P44A yellow, L45A purple, F98A black, Y36AS43A orange, Y36AP44A pink and V_L wild type is royal blue. Colour code for V_H: V37A is black, R44A blue, L45A dark cyan, W47A purple, W103A pink and V_H wild type marine blue. In (C) and (D) the NUV-CD spectra of the V_L (C) and V_H (D) point mutants are shown. Color code for (C) and (D) is analogue to (A) and (B), respectively. For the spectra 15 accumulations each were recorded and buffer-corrected (PBS). All measurements were performed at a protein concentration of 20 µM (FUV-CD) and 50 µM (NUV-CD) in 0.5-mm (FUV) or 5-mm (NUV) quartz cuvette at 20°C.

Figure 3: Influence of conserved residues on protein stability.

The stability of the V_H and V_L alanine point mutants towards GdmCl-induced (A and B). (A) V_L mutants: Y36A is red, Q37A dark cyan, S43A blue, P44A yellow, L45A purple, F98A black, Y36AS43A orange, Y36AP44A pink (all closed circles/lines) and the wild type marine blue (open circles/line). (B) V_H mutants: V37A is black, R44A blue, L45A dark cyan, W47A purple, W103A pink (all closed circles/lines) and the wild type royal blue (open circles/broken line).

Figure 4: Influence of conserved residues on antigen binding

To determine the functionality of the mutants an ELISA with Fv fragments was performed. For V_L (left) and V_H (right), titrations of Fvs containing either the wild type, a representative of an inactive, a low active or a mutant with similar activity as the wild type is shown. The absorption at 405 nm at the signal maximum was corrected for the signal of the variable labelled domains alone. Experiments were performed at 20°C.

Figure 5: CDR-grafted mutants.

(A) Schematic representation of CDR-grafted mutants. The framework for V_L grafting mutants (left) was 1DH5, which is the most stable human consensus sequence V_L domain (from HuCAL [28]). The framework for V_H grafting mutants (right) was 1DHU, which is the most stable human consensus sequence for V_H domains. In (B and C), FUV-CD spectra of V_L (left) and V_H (right) CDR-grafted mutants are shown. The color code for V_L is: MAK_1DH5 in red, 1DH5_MAK in blue; wild type 1DH5 is shown in green MAK33 in black and MAK33 CDR3 1DHU in brown. The color code for V_H is: MAK_1DHU in red, 1DHU_MAK in blue and the wild type 1DHU in green and MAK33 in black. In (D and E) NUV-CD spectra of the V_L (left) and V_H (right) point mutants are shown.

Figure 6: The influence of the CDRs on the stability of the variable domains.

To assess the stability of CDR-grafted mutants, GdmCl-induced unfolding experiments were performed. Data for denaturant-induced transitions for V_L (A) und V_H (B) variants are shown. Data were evaluated according to a two-state unfolding model to obtain midpoints and cooperativity parameters of the transitions. Measurements were performed at 20°C at a protein concentration of 1 µM.

Figure 7: Interface hydration near mutation as determined by MD simulations.

The snapshots illustrate the water distribution within 7 Å of the mutation site. Protein chains are shown as cartoon (V_L:green, V_H:blue). Atoms within 7 Å of the mutation site are indicated as van der Waals spheres using atom color code for water molecules (and grey for protein atoms). For comparison the same regions are also shown for the wild type case (left panels). In case of the V_H W103 mutation the average shift in backbone structure is illustrated (green cartoon) and compared to the wild type case (light blue).

Table 1: Characteristics of MAK33 V_H and V_L mutants.

Domain	Mutation	Association K_D [µM] AUC	Stability GdmCl $D_{1/2}$ [M]	Cooperativity [kJ mol^{-1} M^{-1}]	Antigen binding app. K_D [µM]
V_L	wild type	0.2 ± 0.05	**1.17 ± 0.7**	**20.1 ± 8.2**	**0.9 ± 0.04**
	Y36A	0.4 ± 0.1	0.30 ± 0.1	16.1 ± 2.3	0.6 ± 0.04
	Q37A	0.4 ± 0.1	0.94 ± 0.5	16.0 ± 5.2	1.5 ± 0.5
	S43A	0.5 ± 0.1	1.38 ± 1.0	15.8 ± 8.6	1.4 ± 0.06
	P44A	15.2 ± 3.0	0.95 ± 19.0	17.8 ± 25.0	10.5 ± 0.4
	L46A	0.6 ± 0.1	0.71 ± 0.2	21.2 ± 3.5	4.6 ± 1.2
	F98A	0.6 ± 0.05	0.96 ± 0.20	19.0 ± 2.8	8.3 ± 0.5
	Y36A P44A	6.8 ± 1.5	0.52 ± 0.4	33.7 ± 18.6	4.5 ± 0.7
	Y36A S43A	0.4 ± 0.1	0.23 ± 0.2	17.1 ± 6.0	0.7 ± 0.04
	MAK_1DH5	0.3 ± 0.06	1.44 ± 0.66	15.6 ± 4.8	1.2 ± 0.3
	1DH5_MAK	3.5 ± 0.7	1.32 ± 0.21	16.5 ± 1.8	NB
	1DH5_DH5	3.8 ± 0.7	2.39 ± 1.21	17.1 ± 6.1	NB

V_H	wild type	**0.2** ± 0.05	**0.21 ± 0.1**	**68.6 ± 23.3**	**1.4 ± 0.1**
	V37A	33.0* ± 6.6	0.15 ± 0.1	35.3 ± 6.7	3.0 ± 0.4
	R44A	1.8 ± 0.4	0.20 ± 0.1	34.3 ± 3.7	1.5 ± 0.4
	L45A	292.0* ± 50.0	0.17 ± 0.1	21.5 ± 8.5	16.1 ± 0.9
	W47A	1.2* ± 0.2	0.09 ± 0.1	75.2 ±19.4	NB
	W103A	26.4* ± 5.3	0.17 ± 0.1	15.3 ± 10.5	3.7 ± 0.1
	MAK_1DHU	1.0 ± 0.2	0.52 ± 0.6	4.5 ± 0.4	> 50
	1DHU_MAK	0.4 ± 0.1	1.26 ± 0.2	10.7 ± 0.9	NB
	1DHU_1DHU	8.8 ± 1.6	2.48 ± 1.9	6.3 ± 3.6	NB
	MAK_MAK CDR3 1DHU	0.4 ± 0.2	0.27 ± 1.2	7.9 ± 1.4	14.7 ± 0.5

Table 2: Average number of water molecules around a point mutation

Mutation	<water> in mutation	<water> in wild type
V_L P44A	6.8 (2)	4.3 (1)
V_L S43A	14 (4)	15 (3)
V_L Y36A	1.1 (0.3)	0.8 (0.3)
V_H V37A	1.1 (0.4)	0.9 (0.3)
V_H R44A	16 (3)	15 (3)
V_H L45A	9 (3)	5 (2)
V_H W47A	8 (3)	5 (1)
V_HW103	14 (3)	15 (3)

Figures

Figure 1

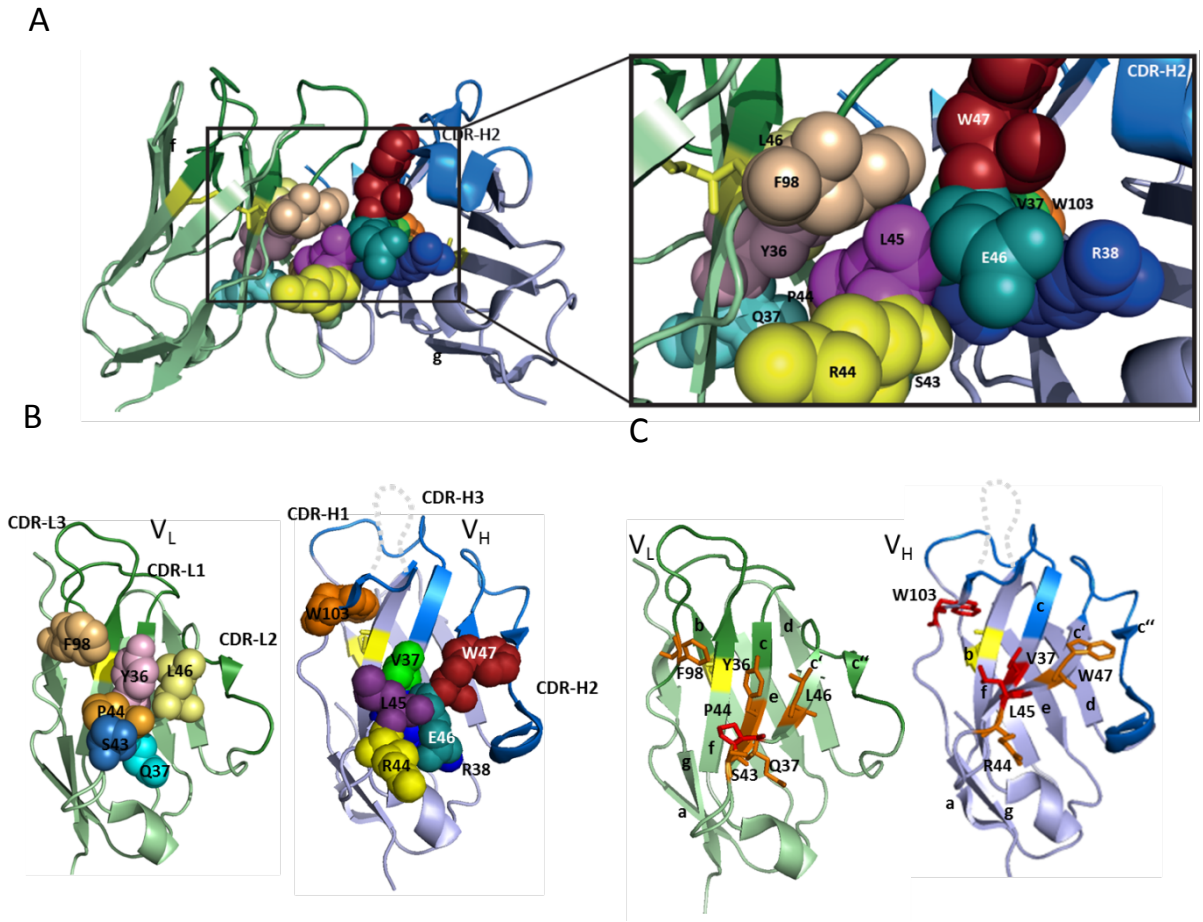


Figure 2

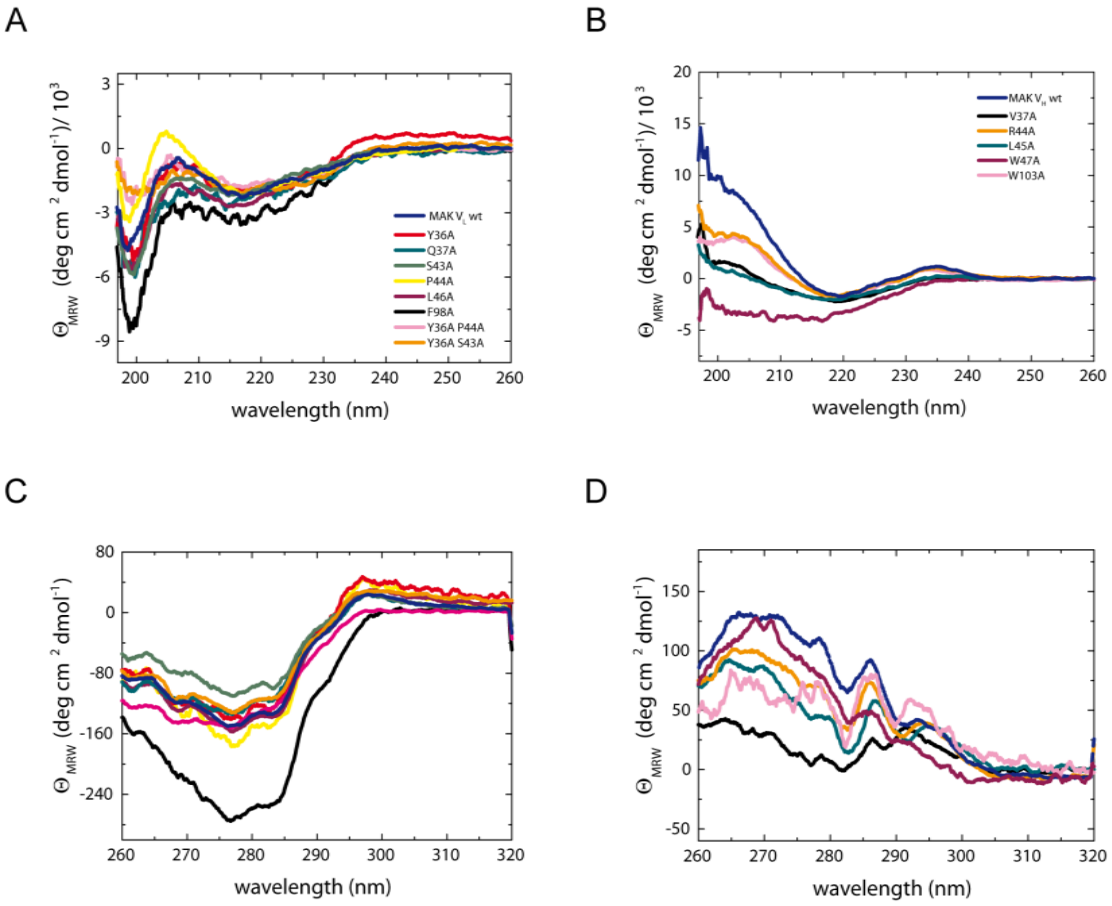
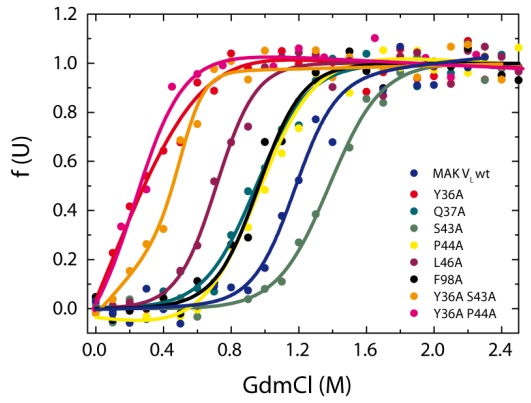

A
B
C
D
wavelength (nm)
MAK V_L wt
Y36A
Q37A
S43A
P44A
L46A
F98A
Y36A P44A
Y36A S43A
MAK V_H wt
V37A
R44A
L45A
W47A
W103A

Figure 3

A


f (U)
GdmCl (M)
MAK V$_L$ wt
Y36A
Q37A
S43A
P44A
L46A
F98A
Y36A S43A
Y36A P44A

B

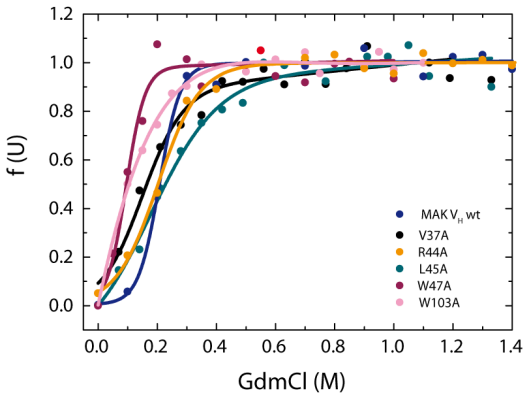
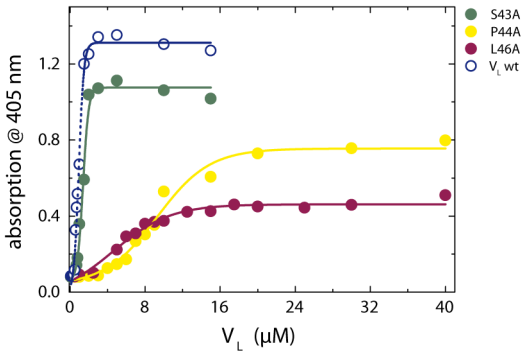

f (U)
GdmCl (M)
MAK V$_H$ wt
V37A
R44A
L45A
W47A
W103A

Figure 4

A


absorption @ 405 nm
0.0
0.4
0.8
1.2
0
8
16
24
32
40
V_L (µM)
S43A
P44A
L46A
V_L wt

B

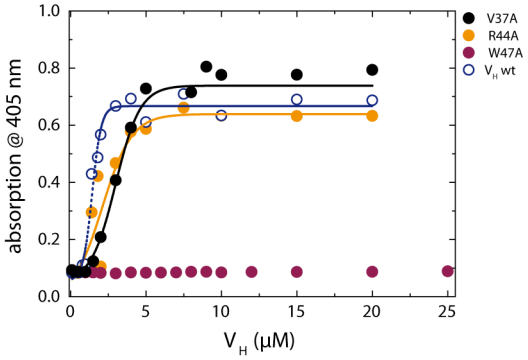
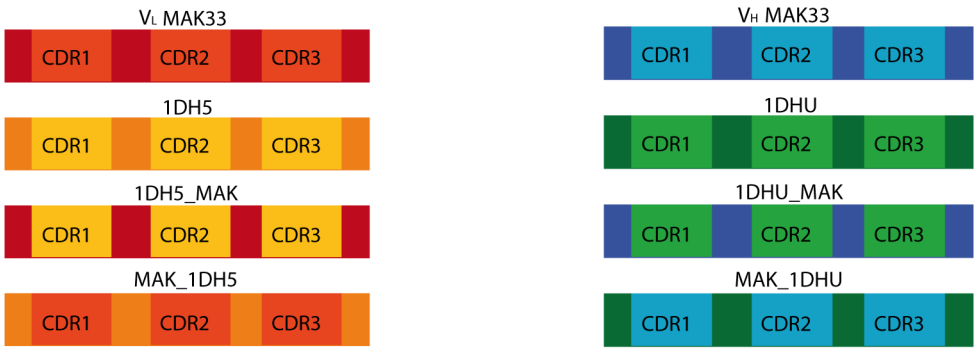
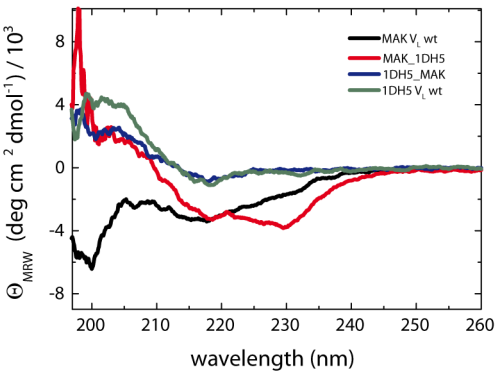

absorption @ 405 nm
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
V_H (µM)
V37A
R44A
W47A
V_H wt

Figure 5

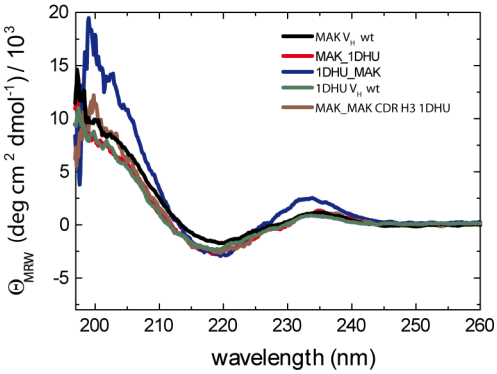
A



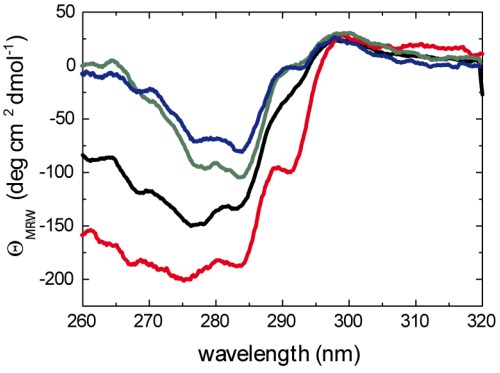
B



C



D



E

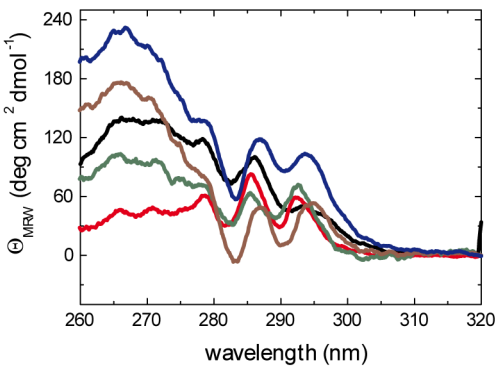
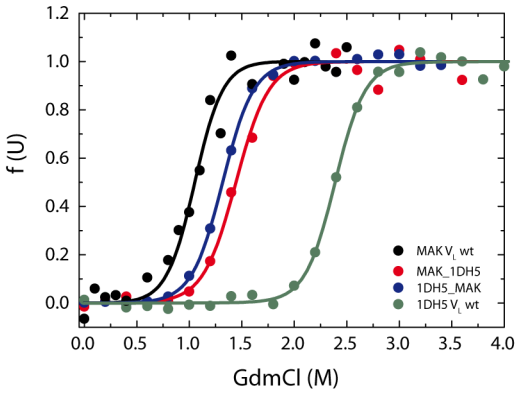


Figure 6

A


f (U)
GdmCl (M)
MAK V$_L$ wt
MAK_1DH5
1DH5_MAK
1DH5 V$_L$ wt

B

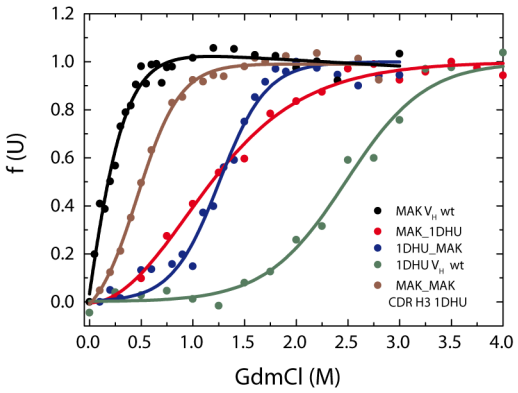
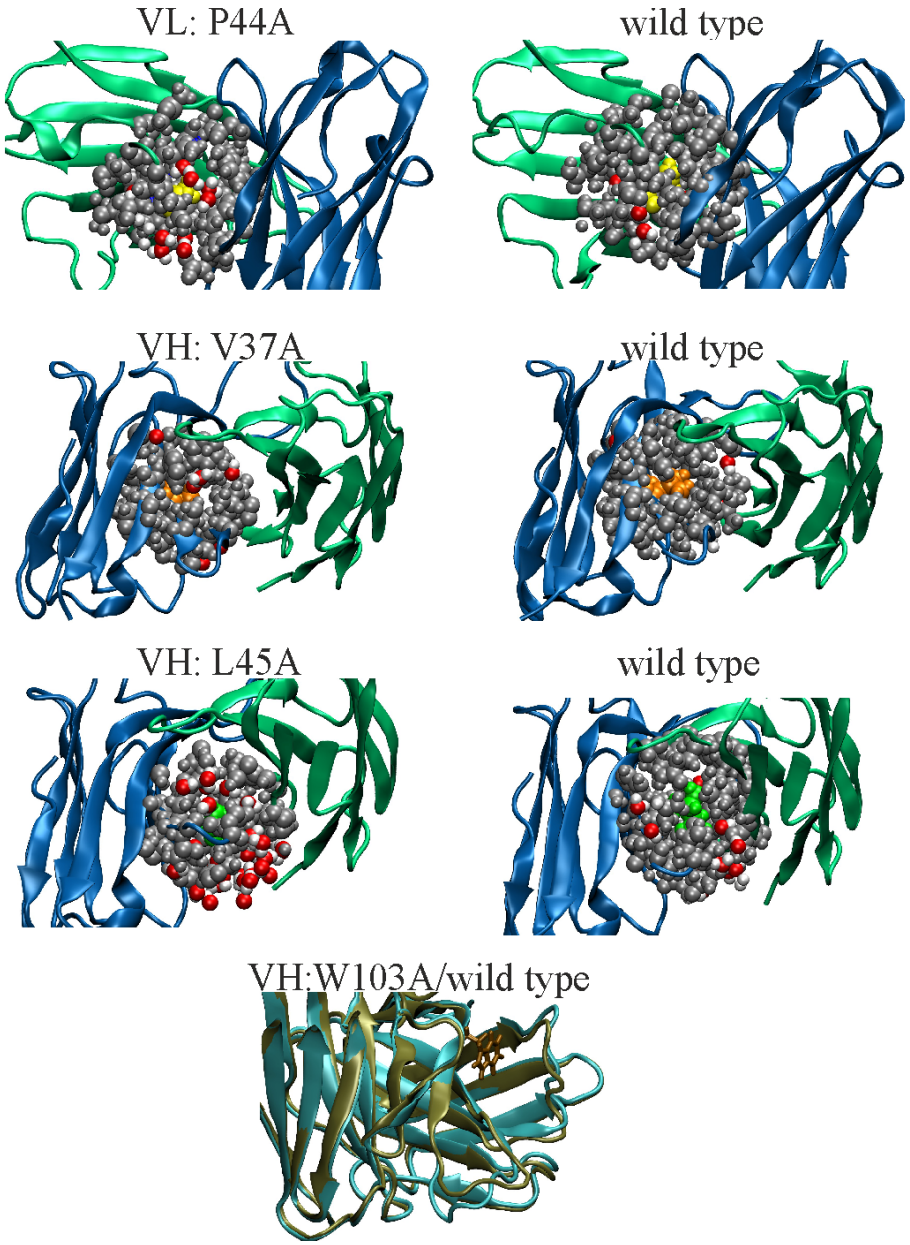

f (U)
GdmCl (M)
MAK V$_H$ wt
MAK_1DHU
1DHU_MAK
1DHU V$_H$ wt
MAK_MAK
CDR H3 1DHU

Figure 7


VL: P44A
wild type
VH: V37A
wild type
VH: L45A
wild type
VH:W103A/wild type

References

1. Chothia, C., et al., *Domain association in immunoglobulin molecules. The packing of variable domains.* J Mol Biol, 1985. **186**(3): p. 651-63.
2. Morea, V., et al., *Conformations of the third hypervariable region in the VH domain of immunoglobulins.* J Mol Biol, 1998. **275**(2): p. 269-94.
3. Vargas-Madrazo, E. and E. Paz-Garcia, *An improved model of association for VH-VL immunoglobulin domains: asymmetries between VH and VL in the packing of some interface residues.* J Mol Recognit, 2003. **16**(3): p. 113-20.
4. Bork, P., L. Holm, and C. Sander, *The immunoglobulin fold. Structural classification, sequence patterns and common core.* J Mol Biol, 1994. **242**(4): p. 309-20.
5. Chothia, C., I. Gelfand, and A. Kister, *Structural determinants in the sequences of immunoglobulin variable domain.* J Mol Biol, 1998. **278**(2): p. 457-79.
6. Abhinandan, K.R. and A.C. Martin, *Analysis and prediction of VH/VL packing in antibodies.* Protein Eng Des Sel, 2010. **23**(9): p. 689-97.
7. Narayanan, A., B.D. Sellers, and M.P. Jacobson, *Energy-based analysis and prediction of the orientation between light- and heavy-chain antibody variable domains.* J Mol Biol, 2009. **388**(5): p. 941-53.
8. Kabat, E.A., et al., *Sequences of Proteins of Immunological Interest.*, in *National Institutes of Health, Bethesda*. 1983.
9. Wang, N., et al., *Conserved amino acid networks involved in antibody variable domain interactions.* Proteins, 2009. **76**(1): p. 99-114.
10. Perchiacca, J.M., M. Bhattacharya, and P.M. Tessier, *Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions.* Proteins, 2011. **79**(9): p. 2637-47.
11. Tanha, J., et al., *Optimal design features of camelized human single-domain antibody libraries.* J Biol Chem, 2001. **276**(27): p. 24774-80.
12. Voordijk, S., et al., *Molecular dynamics simulations highlight mobile regions in proteins: A novel suggestion for converting a murine V(H) domain into a more tractable species.* J Mol Biol, 2000. **300**(4): p. 963-73.
13. Davies, J. and L. Riechmann, *'Camelising' human antibody fragments: NMR studies on VH domains.* FEBS Lett, 1994. **339**(3): p. 285-90.
14. Riechmann, L., *Rearrangement of the former VL interface in the solution structure of a camelised, single antibody VH domain.* J Mol Biol, 1996. **259**(5): p. 957-69.
15. Nakanishi, T., et al., *Critical contribution of VH-VL interaction to reshaping of an antibody: the case of humanization of anti-lysozyme antibody, HyHEL-10.* Protein Sci, 2008. **17**(2): p. 261-70.
16. Chatellier, J., et al., *Functional mapping of conserved residues located at the VL and VH domain interface of a Fab.* J Mol Biol, 1996. **264**(1): p. 1-6.
17. Khalifa, M.B., et al., *Effects on interaction kinetics of mutations at the VH-VL interface of Fabs depend on the structural context.* J Mol Recognit, 2000. **13**(3): p. 127-39.
18. Honegger, A. and A. Pluckthun, *The influence of the buried glutamine or glutamate residue in position 6 on the structure of immunoglobulin variable domains.* J Mol Biol, 2001. **309**(3): p. 687-99.
19. Honegger, A., et al., *The influence of the framework core residues on the biophysical properties of immunoglobulin heavy chain variable domains.* Protein Eng Des Sel, 2009. **22**(3): p. 121-34.
20. Rothlisberger, D., A. Honegger, and A. Pluckthun, *Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability.* J Mol Biol, 2005. **347**(4): p. 773-89.

21. Hugo, N., et al., *Functional aspects of co-variant surface charges in an antibody fragment.* Protein Sci, 2002. **11**(11): p. 2697-705.
22. Tan, P.H., B.M. Sandmaier, and P.S. Stayton, *Contributions of a highly conserved VH/VL hydrogen bonding interaction to scFv folding stability and refolding efficiency.* Biophys J, 1998. **75**(3): p. 1473-82.
23. Isenman, D.E., D. Lancet, and I. Pecht, *Folding pathways of immunoglobulin domains. The folding kinetics of the Cgamma3 domain of human IgG1.* Biochemistry, 1979. **18**(15): p. 3327-36.
24. Bertz, M., J. Buchner, and M. Rief, *Mechanical stability of the antibody domain CH3 homodimer in different oxidation states.* J Am Chem Soc, 2013. **135**(40): p. 15085-91.
25. Feige, M.J., et al., *An unfolded CH1 domain controls the assembly and secretion of IgG antibodies.* Mol Cell, 2009. **34**(5): p. 569-79.
26. Simpson, E.R., E.M. Herold, and J. Buchner, *The folding pathway of the antibody V(L) domain.* J Mol Biol, 2009. **392**(5): p. 1326-38.
27. Englander, S.W., L. Mayne, and J.N. Rumbley, *Submolecular cooperativity produces multi-state protein unfolding and refolding.* Biophys Chem, 2002. **101-102**: p. 57-65.
28. Knappik, A., et al., *Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides.* J Mol Biol, 2000. **296**(1): p. 57-86.
29. Ewert, S., et al., *Biophysical properties of human antibody variable domains.* J Mol Biol, 2003. **325**(3): p. 531-53.
30. Nieba, L., et al., *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, 1997. **10**(4): p. 435-44.
31. Kugler, M., et al., *Stabilization and humanization of a single-chain Fv antibody fragment specific for human lymphocyte antigen CD19 by designed point mutations and CDR-grafting onto a human framework.* Protein Eng Des Sel, 2009. **22**(3): p. 135-47.
32. Honegger, A., *Engineering antibodies for stability and efficient folding.* Handb Exp Pharmacol, 2008(181): p. 47-68.
33. Ewert, S., A. Honegger, and A. Pluckthun, *Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering.* Methods, 2004. **34**(2): p. 184-99.
34. Weidenhaupt, M., et al., *Functional mapping of conserved, surface-exposed charges of antibody variable domains.* J Mol Recognit, 2002. **15**(2): p. 94-103.
35. Worn, A. and A. Pluckthun, *Stability engineering of antibody single-chain Fv fragments.* J Mol Biol, 2001. **305**(5): p. 989-1010.
36. Jung, S., et al., *The importance of framework residues H6, H7 and H10 in antibody heavy chains: experimental evidence for a new structural subclassification of antibody V(H) domains.* J Mol Biol, 2001. **309**(3): p. 701-16.
37. Chen, L., et al., *Anti-hepatoma human single-chain Fv antibody and adriamycin conjugates with potent antitumor activity.* Int Immunopharmacol, 2014. **18**(1): p. 20-6.
38. Nicholson, I.C., et al., *Construction and characterisation of a functional CD19 specific single chain Fv fragment for immunotherapy of B lineage leukaemia and lymphoma.* Mol Immunol, 1997. **34**(16-17): p. 1157-65.
39. Wang, T. and Y. Duan, *Probing the stability-limiting regions of an antibody single-chain variable fragment: a molecular dynamics simulation study.* Protein Eng Des Sel, 2011. **24**(9): p. 649-57.
40. Galitsky, B., *Revealing the set of mutually correlated positions for the protein families of immunoglobulin fold.* In Silico Biol, 2003. **3**(3): p. 241-64.
41. Feige, M.J. and J. Buchner, *Principles and engineering of antibody folding and assembly.* Biochim Biophys Acta, 2014. **1844**(11): p. 2024-2031.

42. Feige, M.J., et al., *The structural analysis of shark IgNAR antibodies reveals evolutionary principles of immunoglobulins.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. 8155-60.
43. Ewert, S., et al., *Biophysical properties of camelid V(HH) domains compared to those of human V(H)3 domains.* Biochemistry, 2002. **41**(11): p. 3628-36.
44. Streltsov, V.A., et al., *Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12444-9.
45. Kovalenko, O.V., et al., *Atypical antigen recognition mode of a shark immunoglobulin new antigen receptor (IgNAR) variable domain characterized by humanization and structural analysis.* J Biol Chem, 2013. **288**(24): p. 17408-19.
46. Muyldermans, S., *Nanobodies: natural single-domain antibodies.* Annu Rev Biochem, 2013. **82**: p. 775-97.
47. Fransson, J., et al., *Human framework adaptation of a mouse anti-human IL-13 antibody.* J Mol Biol, 2010. **398**(2): p. 214-31.
48. Feige, M.J., S. Walter, and J. Buchner, *Folding mechanism of the CH2 antibody domain.* J Mol Biol, 2004. **344**(1): p. 107-18.
49. Bolen, D.W. and M.M. Santoro, *Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of delta G degrees N-U values in a thermodynamic cycle.* Biochemistry, 1988. **27**(21): p. 8069-74.
50. Feige, M.J., et al., *Dissecting the alternatively folded state of the antibody Fab fragment.* J Mol Biol, 2010. **399**(5): p. 719-30.
51. *The PyMOL Molecular Graphics System, Version 1.6. x Schrödinger, LLC*.
52. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of Chemical Theory and Computation, 2008. **4**(3): p. 435-447.
53. Berendsen, H.J.C., D. Vanderspoel, and R. Vandrunen, *Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation.* Computer Physics Communications, 1995. **91**(1-3): p. 43-56.
54. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins-Structure Function and Bioinformatics, 2010. **78**(8): p. 1950-1958.
55. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
56. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

Supplementary figures

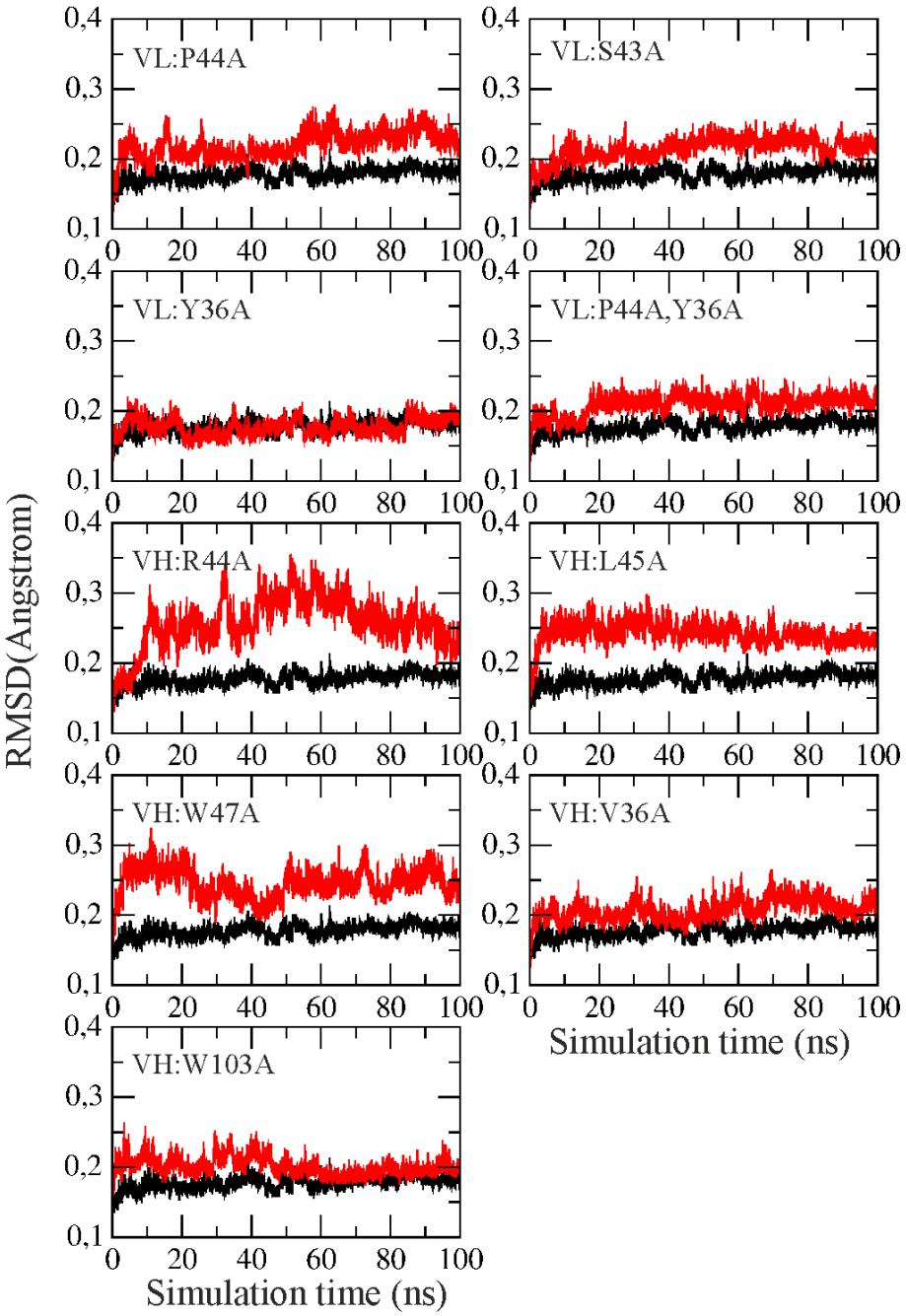


Figure S1. Root-mean-square deviation (RMSD) of the protein backbone (V_L/V_H complex) from the experimental start structure vs. simulation time. The RMSD recorded for the wild type complex is indicated in black (the RMSD of each mutation is shown in red).

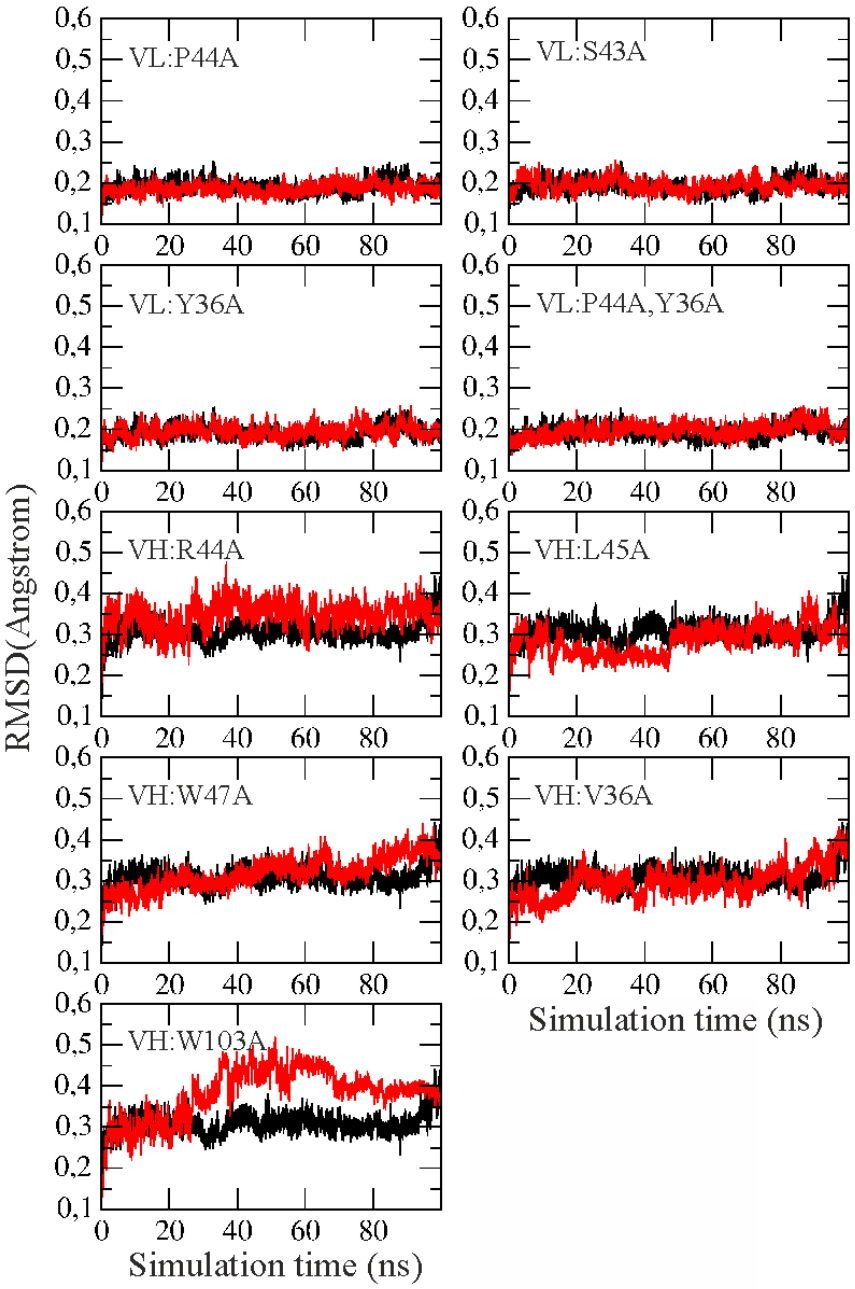


Figure S2. Root-mean-square deviation (RMSD) of the protein backbone of isolated V_L or V_H protein partners from the experimental start structure vs. simulation time. The RMSD recorded for the wild type protein domain is indicated in black (the RMSD of each mutation is shown in red).

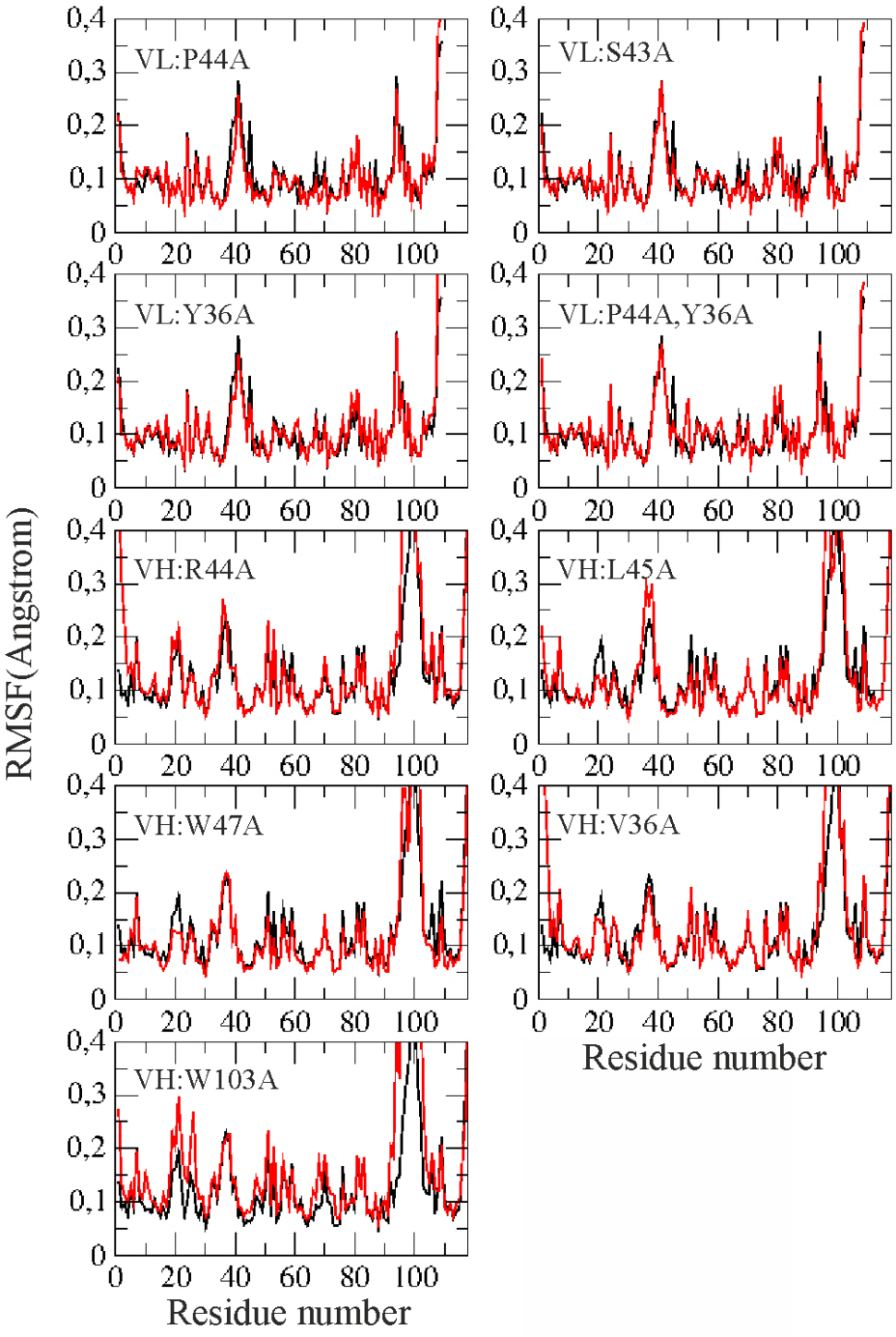


Figure S3. Root-mean-square fluctuation (RMSF) with respect to the mean protein structure during each 100 ns simulation. The RMSF was calculated for all atoms of a residue and plotted vs. residue number. The RMSF obtained for the wild type V_L or

V_H domain is indicated in black (the RMSF of each isolated domain mutation is shown in red).

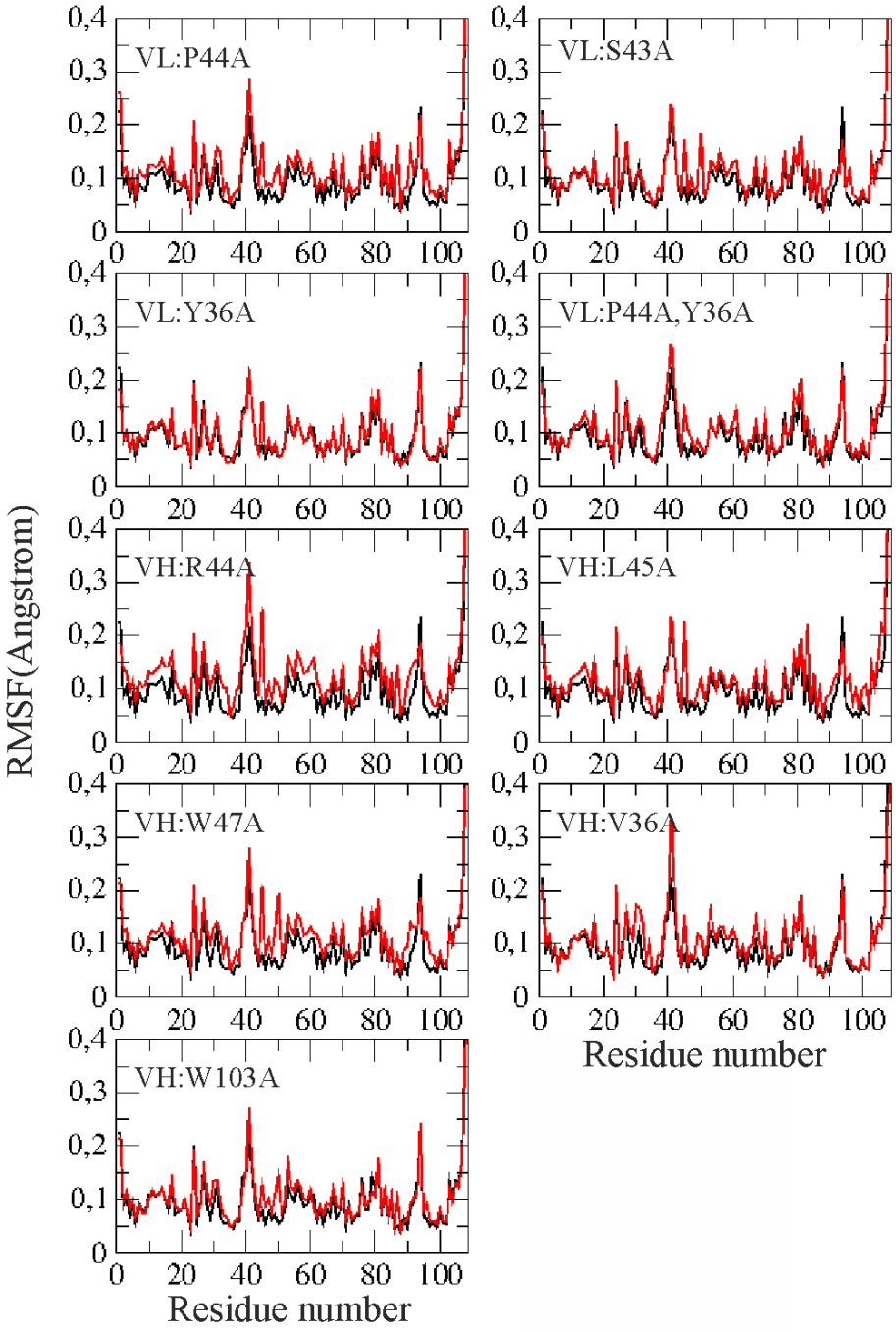


Figure S4. Root-mean-square fluctuation (RMSF) with respect to the mean V_L/V_H complex structure during each 100 ns simulation. The RMSF was calculated for all atoms of a residue and plotted vs. residue number of the V_L domain. The RMSF

obtained for the wild type V_L domain (in complex with the V_H) is indicated in black (the RMSF of each isolated domain mutation is shown in red).

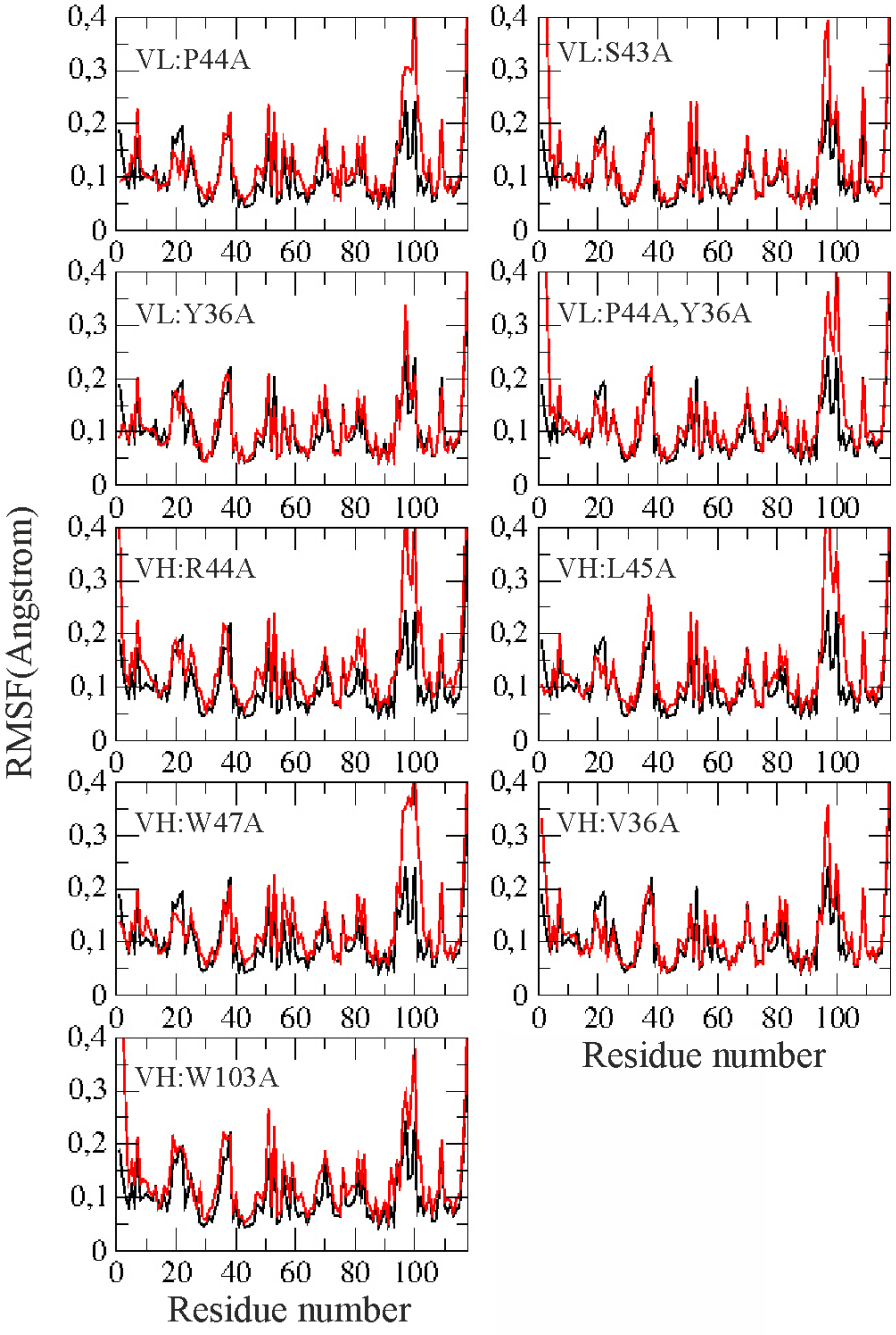


Figure S5. Root-mean-square fluctuation (RMSF) with respect to the mean V_L/V_H complex structure during each 100 ns simulation. The RMSF was calculated for all

atoms of a residue and plotted vs. residue number of the V_H domain. The RMSF obtained for the wild type V_H domain (in complex with the V_H) is indicated in black (the RMSF of each isolated domain mutation is shown in red).

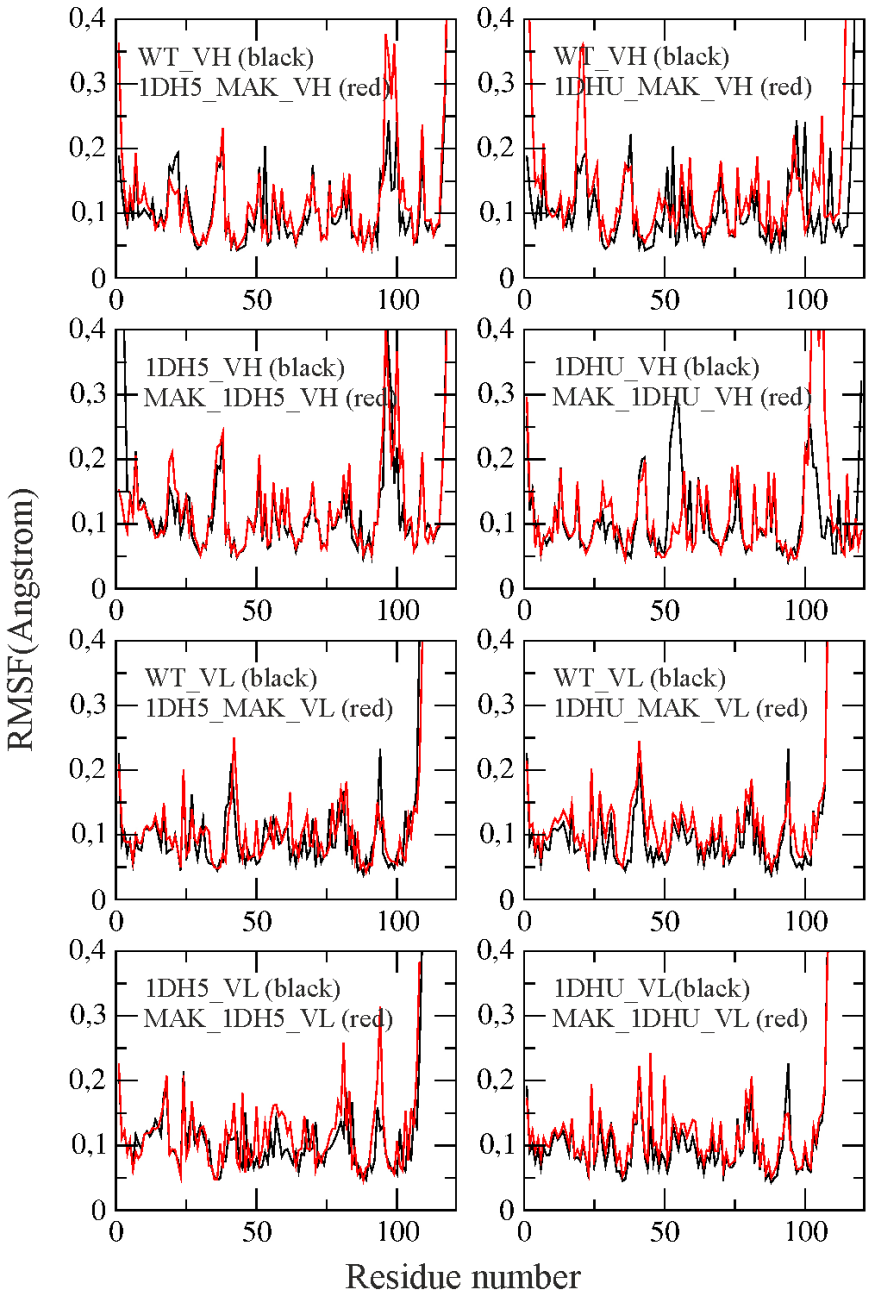


Figure S6. Root-mean-square fluctuation (RMSF) with respect to the mean V_L/V_H complex structure during each 100 ns simulation. The RMSF was calculated for all atoms of a residue and plotted vs. residue number of the V_H domain (upper four

panels) and for the V_L domain (lower four panels). The RMSF obtained for the wild type case is indicated in black and for the loop exchange mutations in red.