JMB

The Influence of the Buried Glutamine or Glutamate Residue in Position 6 on the Structure of Immunoglobulin Variable Domains

Annemarie Honegger* and Andreas Plückthun

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

Immunoglobulin V_H domain frameworks can be grouped into four distinct types, depending on the main-chain conformation of framework 1. Based on the analysis of over 200 X-ray structures representing more than 100 non-redundant V_H domain sequences, we have come to the conclusion that the marked structural variability of the V_H framework 1 region is caused by three residues: the buried side-chain of H6, which can be either a glutamate or a glutamine residue, the residue in position H7, which may be proline only if H6 is glutamine, and by H9 (H10 according to a new consensus nomenclature), which has to be either glycine or proline if H6 is a glutamate residue. In natural antibodies, these three residues are encoded in combinations that are compatible with each other and with the rest of the structure and therefore will yield functional molecules. However, the degenerate primer mixtures commonly used for PCR cloning of antibody fragments can and frequently do introduce out-of-context mutations to combinations that can lead to severe reduction of stability, production yield and antigen affinity.

© 2001 Academic Press

Keywords: immunoglobulin variable domains; protein engineering; structure; framework 1; conformation

*Corresponding author

Introduction

Antibody engineering combines the capability of the immune system to produce highly specific binding proteins to almost any antigen with the ease and efficiency of expressing recombinant proteins in *Escherichia coli* and other hosts. With the widespread use of this technology,[2–4] it has become important to predict the structure of immunoglobulin variable domains reliably. For many applications, the stability and production yields of scFv or Fab fragments derived from natural antibody sequences has been found to be insufficient, and various complex strategies have been designed to improve their properties to meet the

specifications (reviewed by Wörn & Plückthun[5]). Both rational engineering[6–8] and evolution schemes[9–11] have been used successfully to improve individual scFvs. For a more global solution of the problem, the design of framework structures and randomization strategies for synthetic combinatorial antibody libraries[12] requires a profound understanding of the structural properties of the scFv fragment to ensure that as little as possible of the encoded diversity is lost on unstable molecules while the maximal antigen-binding diversity is maintained.

Much of the work done so far has been concerned with predicting the conformation of the hypervariable loops, since these interact directly with the antigen and thus have the strongest effect on antigen-binding affinity and specificity.[13–18] Indeed, the technique of "loop grafting",[19] of transferring the loops of an antibody with desired antigen-binding characteristics onto the framework of variable domains with improved stability, better folding yield and reduced immunogenicity rests on the assumption that differences in framework sequence do not affect the binding characteristics of an antibody significantly. This technique was refined, since adjacent residues, termed "vernier"
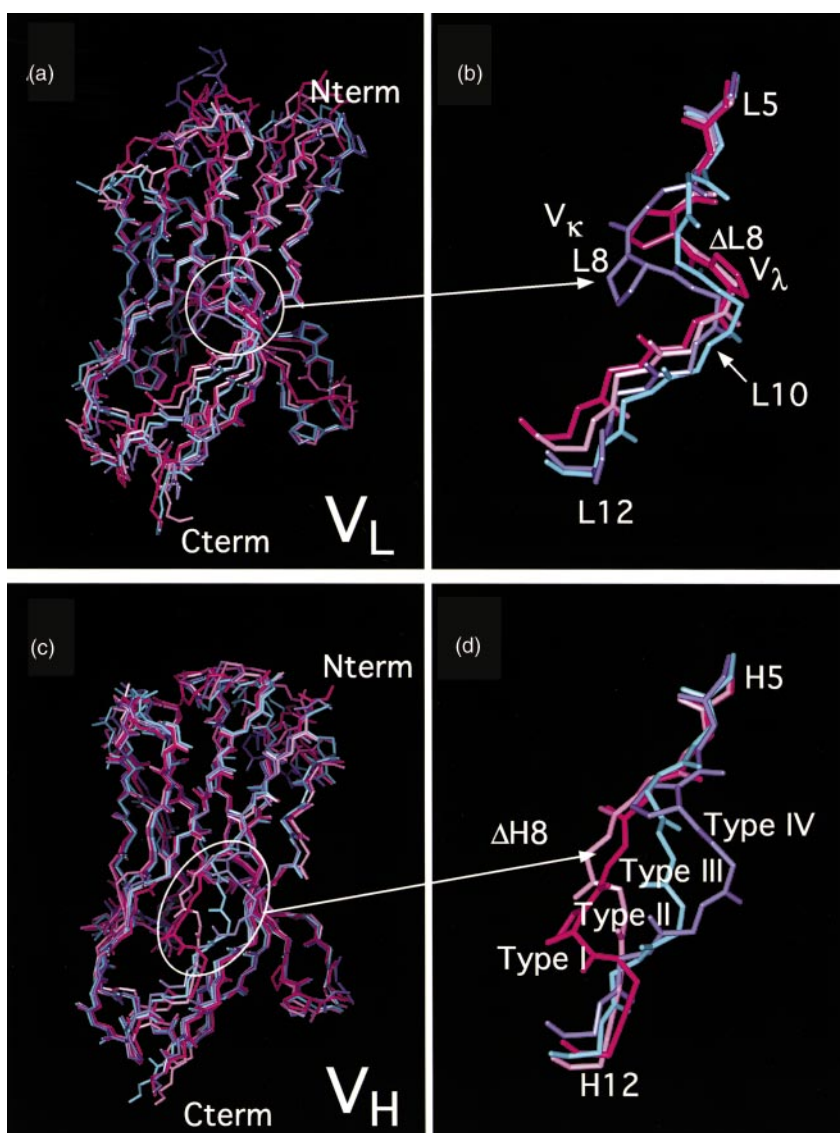
Figure 1. (a) and (b) framework 1 conformation of V_L-domains: magenta, V_λ (L7,L9 Pro) (human Fab B7-15A2,[29] PDB entry 1AQK, 1.84 Å resolution); pink, V_λ (L7,L9 Pro) (murine scFv SE155-4,[49] PDB entry 1MFA, 1.70 Å resolution); cyan, V_κ (L8 Pro) (murine Fab 184.1,[50] PDB entry 1OSP, 1.95 Å resolution; blue, V_κ (L8 Pro) (PDB entry 1FLR,[51] 1.85 Å resolution). (c) and (d) Framework 1 conformation of V_H-domains: magenta, type I (murine Fv fragment of antibody D1.3, C. Marks *et al.*, unpublished results, PDB entry 1A7N, 2.0 Å resolution); pink, type II (human Fab B7-15A2,[29] PDB entry 1AQK, 1.84 Å resolution); cyan, type III (murine Fab 2E8,[30] PDB entry 12E8, 1.9 Å resolution); blue, type IV (first FAB fragment (chain B) in a murine idiotype-anti-idiotype complex,[31] PDB entry 1CIC, 2.5 Å resolution). The Figure was generated with InsightII (MSI/Biosym, San Diego, USA).

residues, were found to affect CDR conformations and to fine-tune antigen recognition.[20] Chothia & Lesk[13] segregated CDR conformations according to "canonical" residues, some of which are located within the CDRs themselves, others in the framework regions. While the CDR grafting[19] approach does work frequently enough to make it a standard technique in antibody engineering, there have been enough failures (often unpublished results, others recognizable by elaborate multi-step strategies used to reach the final goal) to warrant a closer look at how antibody variable domain framework sequences differ from each other, and how these

differences affect structure, stability, folding yield and antigen-binding characteristics.

Within the antibody V_L domains, the structural difference between the lambda and kappa framework is widely appreciated, as the one amino acid residue insertion in the FR 1 of kappa domains relative to the lambda sequences is immediately apparent from a sequence alignment.[21] V_κ domains can be further subdivided according to the presence or absence of a *cis*-proline residue in position L8, while in V_λ the presence or absence of *trans*-proline residues in position L7 and L9 (L8 according to Kabat)† does not significantly affect the main-chain conformation (Figure 1(a) and (b)). V_H domains all have an FR 1 segment of V_λ-like length, but four clearly distinct FR 1 conformations can be observed (Figure 1(c) and (d)). However, only a few authors have appreciated the structural variability in the V_H FR 1 segment.[22–24]

In our study, we introduce a fourth structural subtype, found only amongst murine sequences,

---

† In the AHo numbering scheme,[1] the gap introduced to accommodate the length difference of the FR 1 sequence between V_κ and V_λ is placed in position L8 (V_λ) and not in L10, as in the Kabat scheme. A gap in H8 (V_H) allows use of the same numbering in V_H domains. We use this numbering scheme throughout the paper.
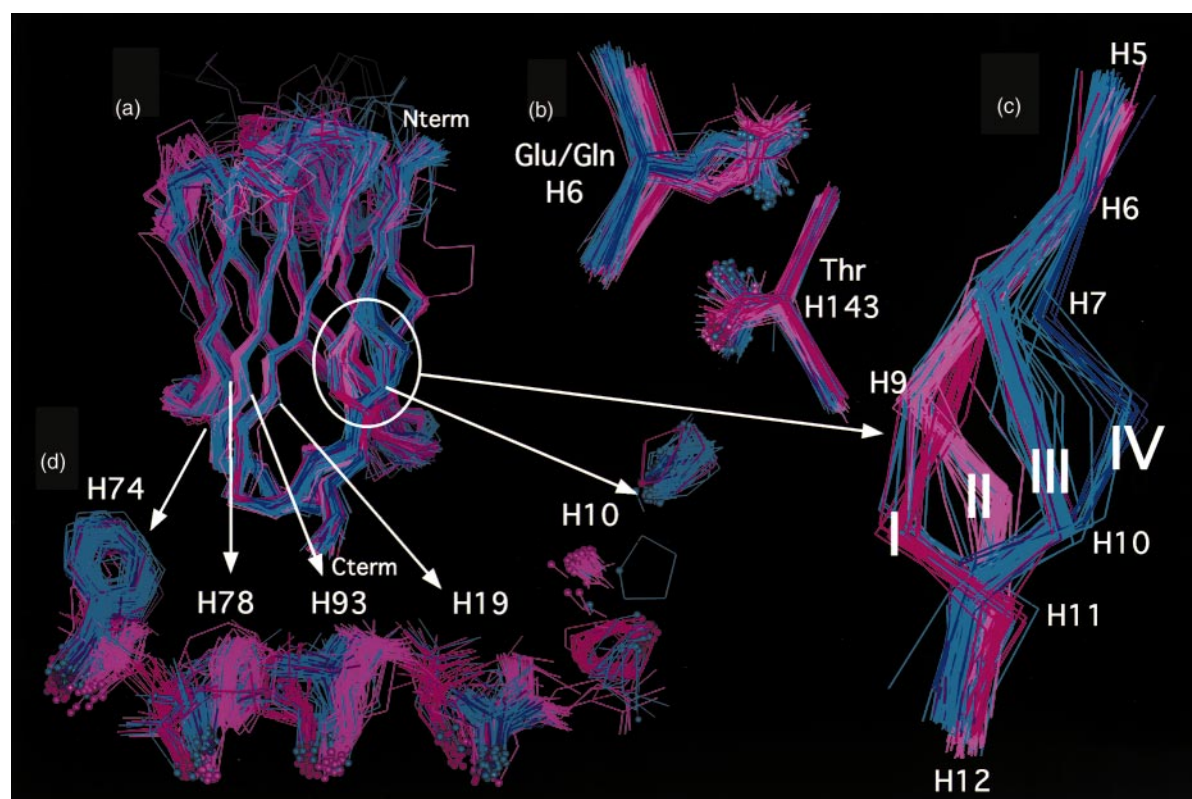
Figure 2. Correlation between predicted subtype and observed structure: 205 V_H domain structures were extracted from antibody Fab and Fv X-ray structures taken from the PDB. They represent 118 non-redundant structures (at least five amino acid residues sequence difference between any pair of sequences). The domain structures were aligned by least-squares superposition of the structurally least variable C^α positions in the V_H domain (H3-H6, H20-H24, H41-H47, H51-H57, H78-H82, H89-H93, H102-H108, H138-H144 in the AHo unified residue numbering scheme,[1] corresponding to residues H3-H6, H19-H23, H35a-H40, H44-H50, H67-H71, H78-H82, H88-H94 and H102-H108 according to the Kabat numbering scheme). The molecules were color-coded according to the structural subtype predicted from the identities of the amino acids in positions H6, H7 and H10 (H9): magenta (type I), H6 E, H7 P, H10 P; pink (type II), H6 E, H7 P, H10 G; cyan (type III), H6 Q, H7 P; blue (type IV), H6 Q, H7 P. The Figure was generated with InsightII (MSI/Biosym, San Diego, USA). (a) C^α trace of the entire V_H domain. (b) Glu/Gln H6 and Thr H143 side-chain conformation. (c) Framework 1 main-chain conformation. (d) Correlated structure and sequence changes across the V_H domain, first described by Saul & Poljak[23] and Saul,[24] who noticed the correlation between framework 1 structure and the types of amino acid found in position H10 (Kabat H9): Gly (as in type II structures) correlated with germline clan V_H3 and Ala/Ser (as in type III and type IV structures), correlated with germline classes V_H1. This in turn correlates with the nature of residue H78 (H67), and with the side-chain orientation of the intervening residues H19 (H18) and H93 (H82). Type I structures, represented in human germline class V_H2 and V_H4, form yet a third class, while type III and type IV structures behave the same with respect to this correlation.

and attempt to establish a causal relationship between the amino acids found in individual positions, particularly positions H6, H7 and H10† (H9) and the conformation assumed by the FR 1 segment of the V_H domain. This causal relationship and the interdependence between the structural requirements of the residues in these three positions offers an explanation for a problem reported in several recent publications. Mutations in the N-terminal part of FR 1 of the V_H domain introduced by PCR cloning using degenerate primers can lead

to severe impairment of production yield, stability and antigen-binding.[25–28] We now provide a structural rationale for the residue combinations allowed in these positions.

Results and Discussion

Comparison of V_H domain structures in the Protein Data Bank

A total of 205 V_H domain structures was extracted from antibody Fab and Fv X-ray structures taken from the PDB‡ and used for statistical analysis (Figures 2 and 3). These domains represent 118 non-redundant structures (at least five amino acid sequence difference between any pair of sequences). The domain structures were aligned

---

† For residue numbering, the unified numbering scheme introduced in the accompanying paper[1] was used. In addition, the traditional Kabat numbering is indicated in parentheses.

‡ http://www.rcsb.org/pdb/

Table 1. Allowed combinations of residues and observed subtype-correlated changes of core residues

	Allowed combinations of residues				Subtype-correlated changes of core residues			
Type	H6	H7	H9	H10	H19	H93	H78	H74
I	Glu	Not Pro	Gly	Pro	Leu	Leu/Met	Ala/Val/Ile/Leu	Leu
II	Glu	Not Pro	Gly	Gly	Leu	Met	Phe	Val
III	Gln	Not Pro	Gly	Any	Leu/Val	Leu	Ala/Val	Phe
IV	Gln	Pro	Gly	Any	Val	Leu	Ala	Phe

by least-squares superposition of the structurally least variable C^α-positions in the V_H domain (H3-H6, H20-H24, H41-H47, H51-H57, H78-H82, H89-H93, H102-H108, H138-H144 in the AHo unified residue numbering scheme[1], corresponding to residues H3-H6, H19-H23, H35a-H40, H44-H50, H67-H71, H78-H82, H88-H94 and H102-H108 according to Kabat). As the inclusion of residue H35a (or 35 or 34, depending on the length of CDR H1) in the set of structurally least variable C^α positions shows, an unambiguous description of this set using Kabat nomenclature becomes rather complicated. The average rms deviation from the average C^α positions for these residues was 0.40 Å. As Figure 2(a) and (c) show, the V_H domain frameworks can be classified according to the main-chain conformation of the framework 1 region into four different subtypes, termed type I to type IV. Typical, well-resolved representatives of the four types of V_H framework structures (shown in Figure 1(c) and (d)) are the V_H domains of the murine Fv fragment of antibody D1.3 (C. Marks *et al.*, unpublished results, PDB entry 1A7N, 2.0 Å resolution) representing the type I structure, the V_H domain the human Fab B7-15A2[29] (PDB entry 1AQK, 1.84 Å resolution) representing type II, the V_H domain of the Fab fragment of the murine antibody 2E8[30] (PDB entry 12E8, 1.9 Å resolution), representing type III and the V_H domain of the first Fab fragment (chain B) in a murine idiotype - anti-idiotype complex[31] (PDB entry 1CIC, 2.5 Å resolution), representing the type IV structure.

Correlation between sequence and structure

Of the 205 V_H structures, 20% correspond to type I, 30% to type II, 41% to type III and 9% to type IV. Type I and type II structures predominantly (87%) have a glutamate residue in position H6, while type III and type IV structures (84%) have a glutamine residue. In most of the structures in which the identity of residue H6 does not match the amino acid expected from the FR 1 main-chain conformation, there is also a mismatch between the V_H sequence family and the identity of the amino acid residue in position H6. While this could be interpreted as an indication that the main-chain

conformation is determined by the overall sequence context rather than by the local sequence, most of these structures were determined using Fab fragments derived from proteolytically digesting the parental monoclonal antibodies, while the sequences were determined at the cDNA level. Therefore, PCR-induced sequence changes would not be reflected at the protein level, but affect only the interpretation of the electron density, and thus for example placing a glutamate rather than a glutamine residue at the critical position. Neither from the shape of the electron density of the side-chain nor from the determination of the protein molecular weight by mass spectrometry would such a substitution be detectable. Results obtained in our lab suggest that deliberately induced mutations of H6, H7 and H10 will indeed change the framework 1 conformation in the manner expected from our analysis, while changes to a sequence pattern different from those observed in nature, e.g. the combination of glutamate at H6 with proline at H7 leads to a significant destabilization.[32]

Type I and II structures segregate with the presence of a proline (type I) or a glycine (type II) residue in position H10 (H9 according to Kabat), while type III and type IV structures are distinguished by the presence (type IV) or absence (type III) of a proline residue in position H7. Pro H7 occurs only in murine and not in human sequences (Figure 2). Proline is never seen in combination with Glu H6, but only in combination with Gln H6. As Figure 3 shows, there is a good correlation between the structural subtype predicted according to the rules summarized in Table 1 from the amino acids found in positions H6, H7 and H10 and the observed structure.

The structural subtypes correlate to different germline families: type I is predominantly represented by human germline families IGHV2 and IGHV4, type II by IGHV3, type III, IGHV1, IGHV5, IGHV6 and IGHV7 (IMGT† and VBase‡). The sequence constellation leading to the type IV structure is not found amongst human germline sequences, but only amongst murine sequences of the germline family V_H1, which contains both type III and type IV sequences (ABG§). In mice, type I comprises representatives of germline families V_H2, V_H3, V_H8 and V_H12, type II, families V_H4, V_H5, V_H6 and V_H7, type III, families V_H1, V_H2, V_H9 and V_H14 and IV, representatives of family V_H1. Although the different germline families
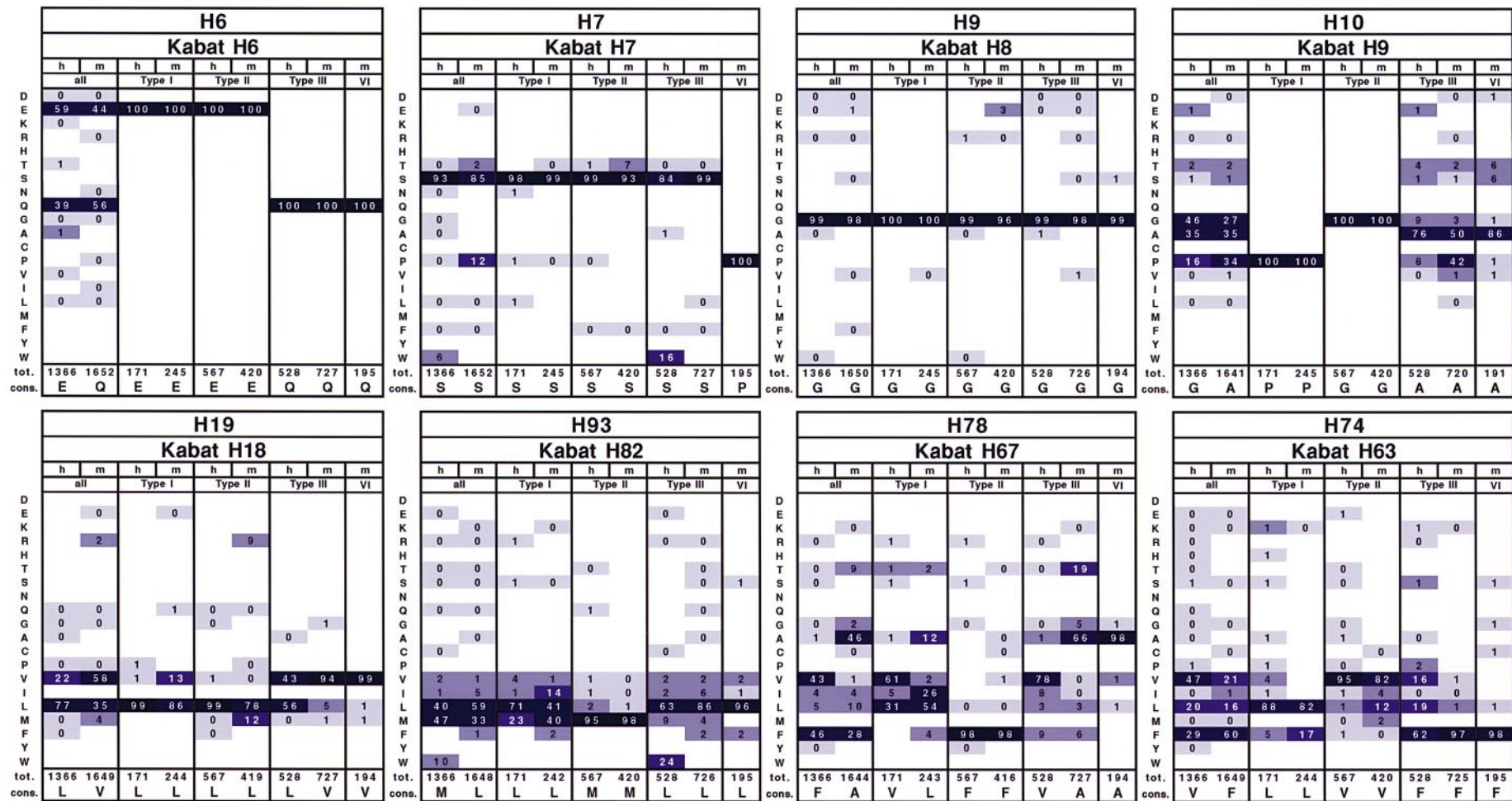
Figure 3. Statistical analysis of the sequence correlations shown in Figure 2. Reasonably complete (>90%) murine (m) and human (h) antibody sequences from the Kabat database were aligned and sorted according to their predicted structural subtype to test for the validity of the correlation with the amino acid sequence in key positions. Fields containing values of exactly zero were colored white and the number was supressed if the amino acid was not observed in any sequence. Those containing higher values were color-coded from light blue (values 0%-1%) to very dark blue (90%-100%) and the amino acid residues sorted according to their properties (charged, uncharged hydrophilic, aliphatic, aromatic side-chains) to facilitate the perception of pattern of position-dependent residue properties and sequence conservation for each position in the domain. *tot.*, Total number of sequences in this category; *cons.*, consensus amino acid for this category.

show correlated sequence changes throughout the domain core, most of the differences are unlikely to cause the striking conformational difference in framework 1. Residues in direct contact with those involved in the altered conformation are either invariant (Gly H9 (H8), Cys H23 (H22), Cys H106 (H92), Trp H43 (H36), Tyr H104 (H90), Gly H140 (H104), Gly H142 (H106) and Thr H143 (H107)) or show no significant correlation with the different structural pattern (Leu, Ile or Val H21 (H20)). Indeed, the conformational difference is most likely caused directly by the local sequence pattern (residue H6-H10), although remote interactions, such as those detailed by Saul and Poljak[23] (Figure 2(d)), mediated by the core packing, may contribute to the observed detrimental effects of away-from-consensus point mutations in this region.

Core hydrogen bonding pattern

Type I and II V_H FR 1 conformations correlate with the presence of glutamate in position H6, type III and IV with glutamine. The different hydrogen bonding requirement of the two different amino acids impose different constraints on the side-chain and main-chain conformation, since the residue H6 side-chain is fully buried in the core of the domain, forming hydrogen bonds to main-chain atoms on the opposite side of the domain. Indeed, the hydrogen bonding requirement of residue H6 may be responsible for breaking the main chain/main chain hydrogen bonding pattern of the "inner" β-sheet (the β sheet that participates in the dimer interface between V_L and V_H), resulting in the framework 4 β-bulge, which is characteristic for all immunoglobulin variable domains (Figures 4 and 5).
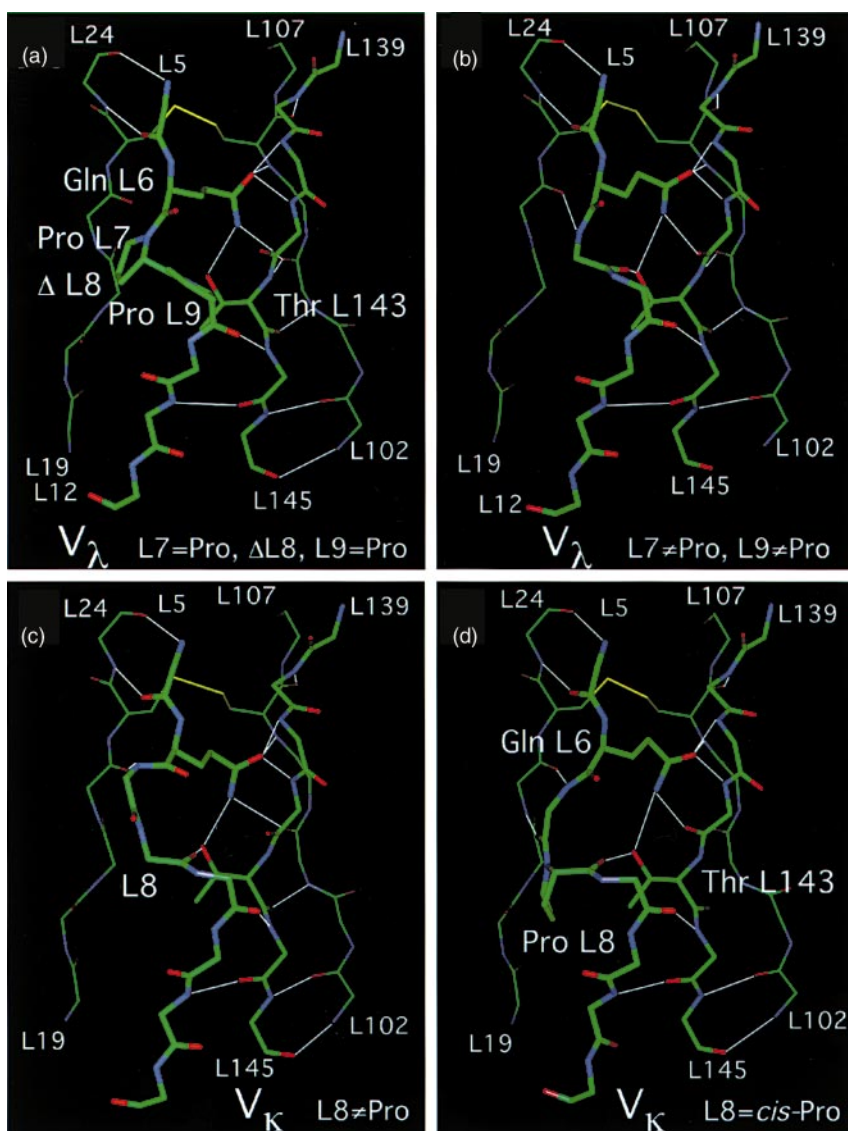

Figure 4. L6 side-chain hydrogen bonding pattern across the V_L domain core. Note that all V_L domains carry a glutamine residue in position L6. The Figure was generated with InsightII (MSI/Biosym, San Diego, USA). (a) V_λ chain with two adjacent proline residues in L7 and L9 (human Fab B7-15A2,[29] PDB entry 1AQK, 1.84 Å resolution). (b) V_λ chain without the two adjacent proline residues L7 and L9 (murine scFv SE155-4,[49] PDB entry 1MFA, 1.70 Å resolution). (c) V_κ chain without a proline residue in L8 (murine Fab 184.1,[50] PDB entry 1OSP, 1.95 Å resolution. (d) V_κ chain with a *cis*-proline residue in position L8 (PDB entry 1FLR,[51] 1.85 Å resolution).
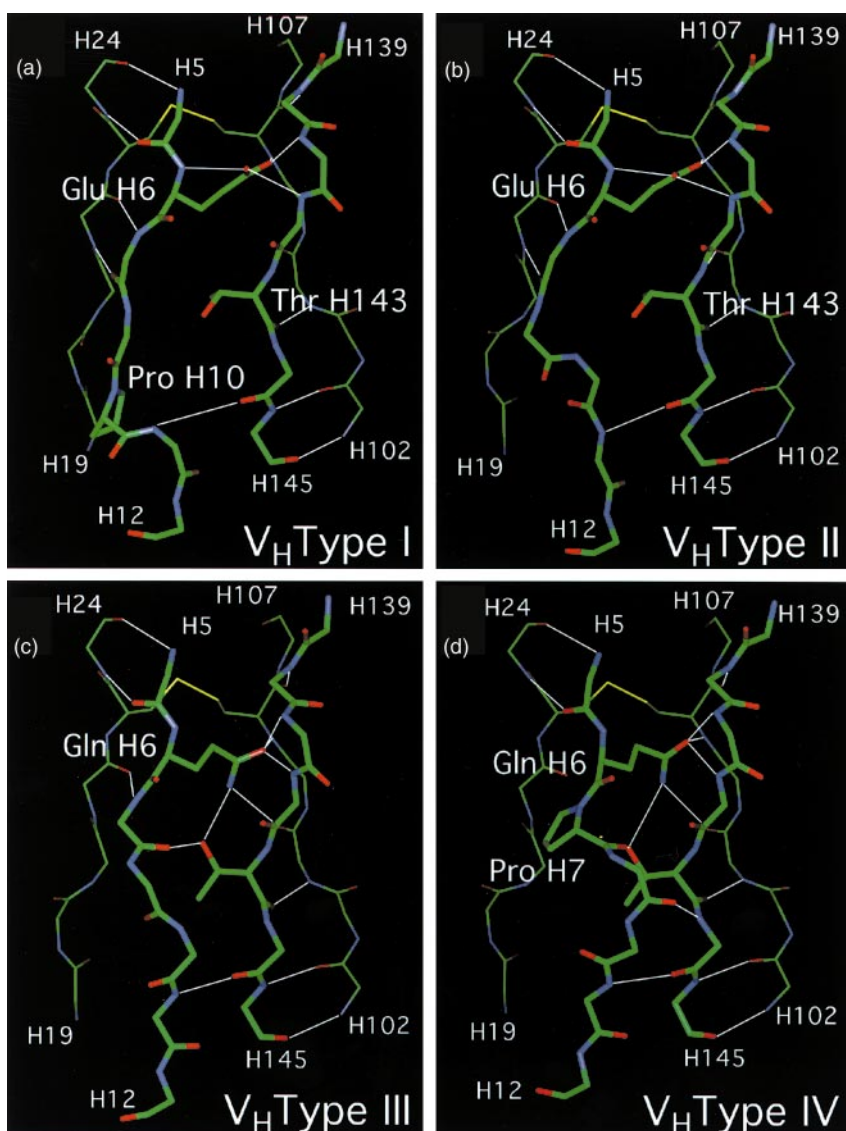
Figure 5. H6 side-chain hydrogen bonding pattern across the V_H domain core. Note that V_H domains carry either (a) and (b) a glutamate or (c) and (d) a glutamine residue in position H6. The Figure was generated with InsightII (MSI/Biosym, San Diego, USA). (a) Type I (murine Fv fragment of antibody D1.3, C. Marks *et al.*, unpublished results, PDB entry 1A7N, 2.0 Å resolution). (b) Type II (human Fab B7-15A2,[29] PDB entry 1AQK, 1.84 Å resolution). (c) Type III (murine FAB 2E8,[30] PDB entry 12E8, 1.9 Å resolution). (d) Type IV (first FabB fragment (chain B) in a murine idiotype-anti-idiotype complex,[31] PDB entry 1CIC, 2.5 Å resolution).

In the case of a glutamate residue in position H6 (Figure 5(a) and (b)), one of the side-chain carboxylate oxygen atoms bridges the two β-sheets by accepting hydrogen bonds from its own main-chain nitrogen atoms (H6) and from the main-chain nitrogen atom of glycine H142 (H106). The other carboxylate oxygen atom accepts hydrogen bonds from the main-chain nitrogen atoms of cysteine H106 (H92) and of residue H141 (H105), although the main-chain nitrogen atom of glycine H142 (H106) would also be within reach. This hydrogen bonding pattern is possible only because residues glycine H140 (H104) and glycine H142 (H106) form the β-bulge in the FR 4 strand of the variable domain structure. Therefore, the main-chain NH groups of residues H106 (H92), H141 (H105) and H142 (H106) are not involved in the main chain/main chain hydrogen bonding of the inner β-sheet and are thus available for interactions with the side-chain of H6. This hydrogen bonding is consistent with the presence of the deprotonated

form of glutamate, as the glutamate side-chain acts only as hydrogen bond acceptor, not as donor.

The side-chain oxygen atom of glutamine H6 (Figure 5(c) and (d)) is similarly within hydrogen bonding range of main-chain NH groups H106 (H92), H141 (H105) and H142 (H106), although, on average, it is somewhat closer to the H142 (H106) NH than the side-chain oxygen atom of glutamate. The glutamine side-chain NH_2 cannot bridge the two sheets by accepting two hydrogen bonds in the way the glutamate oxygen atom does. Instead, it swings around into a different conformation to donate a hydrogen bond to the main-chain C=O of residue H104 (H90). In addition, it donates a second hydrogen bond to the side-chain oxygen atom of threonine H143 (H107), which in turn donates a hydrogen bond to the main-chain C=O of residue H9 (the eighth residue from the N terminus, since in our numbering scheme, H8 is the gap introduced to ensure numbering consistency with other variable domains). If H6 is a glutamate

residue, the Thr H143 (H107) side-chain is rotated in such a way that the side-chain oxygen atom can reach the solvent. In approximately 40 % of the glutamine-containing V_H structures in the PDB, the glutamine side-chain amide group is rotated by 180°, swapping the nitrogen and oxygen atoms. However, this does not make sense, since in this position neither of the two atoms could satisfy its hydrogen bonding potential, and their usual hydrogen bonding partners similarly remain unsatisfied. Since the two atom types cannot be distinguished by their electron density, the correct orientation cannot be determined by X-ray crystallography, but can only be inferred from their hydrogen bonding potential. V_L domains all contain glutamine in position L6, as does the corresponding position in T-cell receptor V_α and V_β domains, and these glutamine side-chains form the same hydrogen bonding pattern as glutamine H6 in the V_H domains (Figure 4(a)(d)).

Effects of the neighboring residues

While the identity of residue H6 is the prime determinant of framework 1 structure, the exact conformation is modified by the structural requirement of the adjacent residues, resulting in four dis-
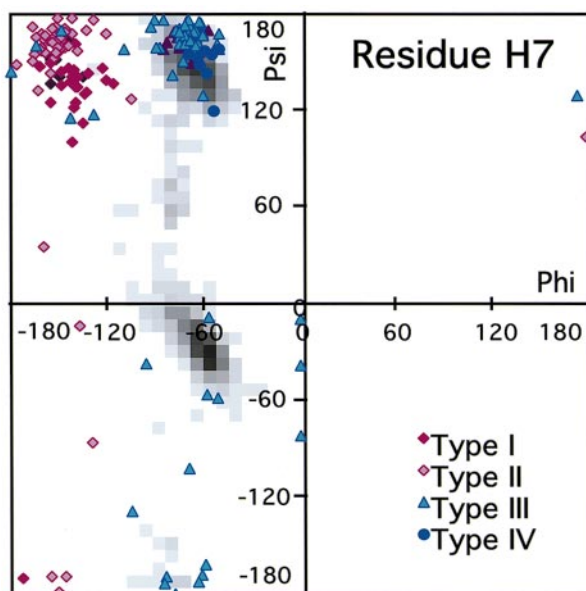

Figure 6. Ramachandran plot of residue H7 main-chain torsion angles. Lists of the main-chain torsion angles of the aligned V_H structures shown in Figure 2 were exported from the program InsightII (Biosym/MSI, San Diego). The phi and psi angles from the different structures were sorted according to the aligned sequence positions using an EXCEL Visual Basic macro. For each sequence position, a Ramachandran plot was drawn in which the points corresponding to the homologous position in the different V_H structures were color-coded according to the predicted subtype for that structure (magenta, type I; pink, type II; cyan, type III, and blue, type IV).

tinct conformations. The main-chain torsion angles assumed by residue H7 clearly differ depending on whether H6 is a glutamate or a glutamine residue (Figure 6). With a glutamine residue in H6, the H7 torsion angles lie within the optimal range for a proline residue, independent of whether this residue is indeed proline (type IV structure) or some other amino acid, predominantly Ala (type III structure). Glutamate in H6 is not compatible with proline in H7, and indeed such a combination leads to a significant destabilization.[32]

The rules determining the FR 1 conformation are summarized in Table 1. If the amino acid residue in H6 is glutamate and H10 (H9) is a proline residue, a type I framework structure is observed. Saul and Poljak[23] described H10 (H9) Pro as the determining feature of this conformation, but Pro H10 in combination with Gln H6 results in a type III or IV structure, as does the combination of Gln H6 with Ala or Ser H10 (H9). The combination of H6 glutamate with Gly H10 (H9) results in a type II framework structure. Any other type of residue in H10 (H9) seems to have a destabilizing effect in combination with H6 Glu. Generally, Glu H6 is associated with positive phi angles in positions H9 (H8) and/or H11 (H10), which show a strong preference for Gly residues. A glutamine residue in H6 leads to a type III structure, unless H7 is a proline. In this case, a type IV structure is observed, which is almost perfectly superimposable upon the structure of V_λ chains characterized by L6 Gln, L7 Pro, L8 gap, L9 Pro (Figures 1(a) and (b), and 4). However, in V_λ, the type IV conformation persists even if L7 and L9 are not proline residues (Figure 1(b)), although with a decreased stability, at least for a lambda-type framework 1 kink in the context of a V_κ domain.[33] In V_H (H6 Gln), we see the type III conformation in the absence of Pro H7, probably due to a conserved glycine residue in position H9, which allows for greater conformational freedom than the proline or serine residue observed in the corresponding position of V_λ.

Relative abundance of the different subtypes in germline and rearranged sequences

Although the relative abundance of H6 Glu and H6 Gln is quite different in the human and murine germline sequences, this difference is much less pronounced in rearranged sequences (Figure 7). Glu H6 (type I and type II framework structures) is observed frequently amongst human germline sequences (66 % of all sequences), but it is rather rare amongst murine germline sequences (21 %). Amongst rearranged sequences (Kabat database, 1366 human and 1652 murine sequences), the type I sequence pattern is underrepresented in human and overrepresented in murine sequences with respect to the germline frequency (human, 20 % of germlines, 13 % of rearranged sequences; mouse, 10 % of germlines, 16 % of rearranged sequences), type II is overrepresented in murine antibodies (human, 46 % of germlines, 38 % of rearranged
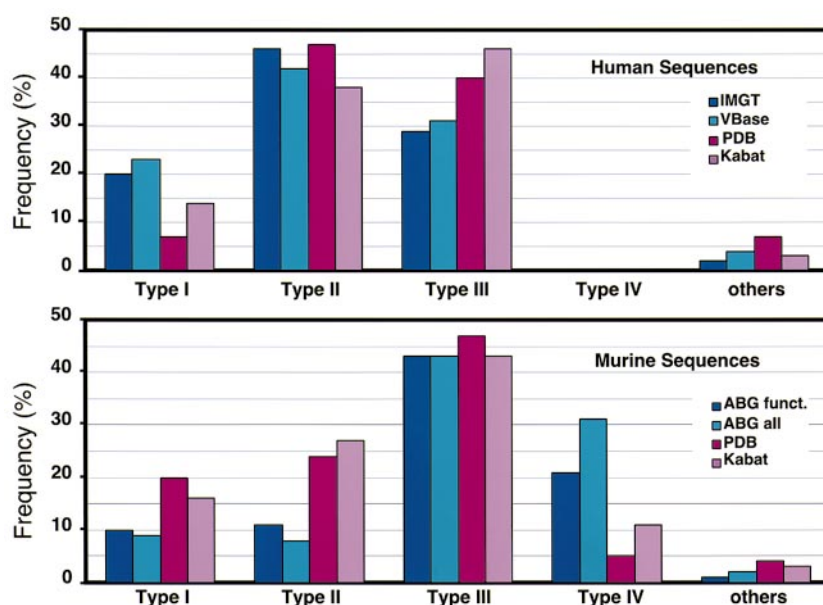
Figure 7. Representation of structural subtypes amongst human and murine germline (IMGT/VBASE/ABG) and rearranged (Kabat, PDB) V_H sequences (see Materials and Methods). Human V_H germline sequences taken from IMGT and from VBASE, murine germline sequences taken from ABG and human and murine rearranged V_H sequences taken from the Kabat database[21,48] were sorted according to their predicted subtype. The relative frequencies of the different subtypes are plotted as a histogram.

sequences; mouse, 11 % of germlines, 27 % of rearranged sequences), type III overrepresented in human rearranged sequences (human, 29 % of germline sequences, 46 % of rearranged sequences; mouse, 43 % of germline sequences, 43 % of rearranged sequences), and type IV, not observed amongst human sequences, is strongly underrepresented amongst murine rearranged sequences (21 % of germline sequences, but only 11 % of rearranged sequences). Furthermore, while for types I to III, the representation amongst functional and total (functional and pseudogenes) is very similar, type IV comprises a disproportionally large fraction of non-functional genes (Figure 7(b)). As a result, the distribution of the subtypes amongst mature antibodies (type I, 14 % in man/16 % in mice; type II, 38 %/27 %; type III, 46 %/43 %; and type IV, 0 %/11 %) is much more similar than could be expected from the distribution of gene sequences. It is not clear whether this effect is just a coincidence, or whether it results from some functional specialization of the different framework classes, which need to be present in a balanced mixture to ensure the proper function of the humoral immune system.

Potential effects on dimer interface orientation and on antigen-binding

The transformation required for a least-squares superposition of the structurally invariant C^α positions of the V_H domain onto the corresponding positions of the V_L domain should represent an objective description of the relative position of the two domains in an Fv fragment. Based on this measure, there exists a significant correlation between domain orientation and V_H subtype. However, more detailed analysis showed that the differences in relative domain orientation between

the different V_H subtypes were not so much due to differences in the relative orientation of the V_L and V_H β-sheets forming the dimer interface, but rather a result of the different relative positions of the inner and outer β-sheet of the V_H domain. This shift of the outer relative to the inner β-sheet of the V_H domain was observed by Saul & Poljak.[23] Jung *et al.*[32] showed that mutating H6, H7 and H10 was not sufficient to produce this effect, indicating that this subtype-dependent shift of the outer relative to the inner β-sheet was due to subtype-correlated sequence differences in the domain core. If the relative orientation of only the inner β-sheets of V_L and V_H was considered, there was little correlation between interface orientation and V_H subtype detectable. In addition, natural type I and type II V_H sequences usually contain longer CDR2 loops than type III and type IV sequences, making it impossible to decide, based on the comparison of natural antibody structures, whether any residual effect of subtype on domain orientation was due to the differences in the FR 1 region or to differences in the CDR 2 region. In our experience, changing the FR 1 subtype has little direct effect on antigen affinity, as long as the resulting scFv is still folding properly.

Experimental consequences of framework 1 variations

Usually, antibody variable domains are cloned from hybridomas or from natural antibody libraries, using commercial or custom-made mixtures of PCR primers designed to cover most of the naturally observed sequence diversity with as few mismatches as possible.[34–39] Different clones derived from the same amplification reaction, even starting from cDNA derived from one particular hybridoma clone, may show considerable sequence

variability in the first few residues at the N terminus of either domain, since different molecules in the degenerate primer set may hybridize productively. The C-terminal primers are usually less problematic, since they are commonly directed to sequences within the constant domains of the antibody or within the very conserved framework 4 region and therefore cannot affect the variable domain sequences. While in many cases the resulting scFv fragments are fully functional,[40] in other cases these PCR-induced sequence changes resulted in impaired antigen-binding, poor production yields and decreased thermodynamic stability.[27]

Particularly the residue in position H6, a buried glutamine or glutamate residue, has been found repeatedly to have a strong effect on antibody functionality. The PCR-induced substitution of Gln H6 by Glu decreased the production yield of the scFv derived from the murine hybridoma OKT3 30-fold, while antigen-binding was not affected.[26] The same PCR-induced substitution also affected the folding efficiency of several anti-hCG (human chorionic gonadotropin) scFvs and Fab fragments belonging to the V_H2A subgroup,[27] impairing both periplasmic production of the scFv in *E. coli* and refolding from inclusion bodies. In this case, antigen-binding was also abolished. The problem could be relieved partially by changing some other FR 1 residues to the V_H1A consensus sequence, which usually contains Glu in position H6. The additional mutations (Lys H14 (H13) to Ser, Lys H20 (H19) to Thr and Met or Ile H21 (H20) to Leu) restored some antigen-binding capability. On the other hand, de Haard *et al.*[27] reported that "analysis of the structures did not reveal any notable differences between related antibodies with either Gln or Glu at position H6", which clearly conflicts with our observations. Langedijk *et al.*[28] noted the effect of the Gln H6 to Glu substitution relative to the putative germline sequence on the stability of the scFv, and went back to determine the protein sequence of the original monoclonal antibody to verify that the substitution was indeed induced in the cloning step and was not the result of a somatic mutation introduced and selected during B-cell maturation.

Since most of the examples described in the literature of such a substitution resulting in impaired functionality have been sequences in which a germline-encoded glutamine residue was replaced by a glutamate residue, it has been suggested that putting a negatively charged glutamate residue into the core of the V_H domain would always have a strongly destabilizing effect, and that this might be the reason for the effect the substitution has on production levels, stability and functionality of these antibodies.[26] However, in our

† http://www.mrc-cpe.cam.ac.uk/imt-doc/
‡ http://imgt.cnusc.fr:8104/
§ http://www.ibt.unam.mx/vir/V_mice.html

hands, some of the best antibody scFv frameworks in terms of production yield (4D5[41]) and stability (disulfide-restored ABPC48 and derivatives[42–44] are sequences that contain a germline-encoded Glu in position H6. The unusual unpaired V_H domains (V_{HH}) of some camelid antibodies,[45,46] which have been proposed to have an exceptionally high thermodynamic stability,[47] all carry a glutamate and not a glutamine residue in this position, contradicting the assumption that a glutamate residue in position H6 is generally unfavorable for an antibody. The stability of the human V_H3 consensus construct encoded in the synthetic HuCAL V_H3 framework,[12] also containing Glu H6, in combination with a suitable CDR H3 sequence, compares favorably with even the best of the camelid V_{HH} domains (S. Ewert *et al.*, unpublished results). These "good" scFv fragments all have a type II FR 1 structure.

Conclusions

As far as can be seen from an analysis of existing variable domain structures, the structural variability observed in the framework 1 region of the immunoglobulin domains is a direct consequence of the hydrogen bonding requirements of the buried hydrophilic residue in position 6. Depending on whether this residue is glutamine or glutamate, the main chain between residues H7 and H11 (H10) is twisted out of the optimal torsion angle range needed to continue the antiparallel β-sheet. The main-chain conformation in positions H7 to H11 (H10) is modified by the presence of glycine and proline residues in these positions. Different hydrogen bonding to Glu/Gln H6 also orients the mostly buried side-chain of Thr H143 (H107), which either forms a hydrogen bond to the main-chain carbonyl group of H7 or to the solvent. Together, these interactions cause the different framework 1 conformations. Out-of-context mutations will lead either to an incompatible combination of amino acids in these positions (e.g. Glu H6-Pro H7 or Glu H6-Ala H10 (H9)) or to steric problems, as an altered FR 1 clashes with the clan-specific core packing, explaining the destabilizing effect of such mutations. Gln H6 is relatively permissive concerning the identity of the amino acids in positions H7 and H10 (H9), while Glu has more stringent requirements: H7 cannot be a proline residue, and H10 (H9) requires a positive phi torsion angle compatible only with a glycine residue in this position. Jung *et al.*[32] have experimentally verified the causal relationship that we postulated on the basis of the observed correlations.

Materials and Methods

Databases

The databases VBASE† (Ian Tomlinson) and IMGT‡ (Marie-Paule Lefranc) for human germline sequences, ABG§ (Juan Carlos Almagro) and IMGT‡ (Marie-Paule

Lefranc) for murine germline sequences and the Kabat database[21,48]† for rearranged sequences were downloaded from the appropriate internet sites, and the V_H and V_L sequences were extracted. Since the Kabat database contains a large number of incomplete sequences and sequences with a large fraction of undetermined residues, these were eliminated before any statistical analysis was performed. All sequences were converted to a GCG (Wisconsin Package V.9.0, Genetics Computer Group, Madison, Wisc., USA) compatible format and aligned and sorted according to their degree of relatedness using the PILEUP program (GCG). The aligned sequences were imported into EXCEL (Microsoft) and the sequence alignment corrected either manually or with the help of Visual Basic macros to conform to the sequence alignment derived from the structural alignment of the immunoglobulin variable domains (see below). Antibody X-ray structures and the corresponding sequences were downloaded from the Protein Data Bank (PDB, at the Research Collaboratory for Structural Bioinformatics (RCSB)).

Renumbering the PDB files to achieve a consistent residue numbering scheme

To facilitate the automation of the structural analysis, the immunoglobulin coordinate files were dissected into individual domains (V_L, C_L, V_H, C_H, antigen) and the residue numbers changed to match the numbering scheme described in the accompanying paper.[1] This renumbering was achieved with a series of EXCEL (Microsoft) Visual Basic macros which, from a series of PDB files, extract the sequences into a sequence alignment. This alignment can be gapped either manually or using alignment programs such as the PILEUP-module of the GCG package. The numbering scheme entered in the header row of the sequence alignment is then applied to all the PDB files represented in the sequence alignment. The PDB files containing individual domains were reassembled into Fv, liganded Fv, Fab and liganded Fab fragments, as far as the original data allowed. Renumbered and aligned PDB files of immunoglobulin domains will be made accessible on our Web site‡.

Structural alignment of immunoglobulin domains

To achieve a reproducible structural alignment, the structurally least variable C^α positions within the immunoglobulin variable domains were determined by an iterative structural alignment using the program InsightII (MSI/Biosym, San Diego, USA). The C^α coordinates of residues L3-L6, L20-L24, L41-L47, L51-L57, L78-L82, L89-L93, L102-L108 and L138-L144 in the V_L domain or H3-H6, H20-H24, H41-H47, H51-H57, H78-H82, H89-H93, H102-H108 and H138-H144 in the V_H domain (consensus residue numbering[1]), corresponding to residues L3-L6, L20-L24, L33-L39, L43-L49, L62-L66, L71-L75, L84-L90 and L97-L103 in the V_L domain and H3-H6, H19-H23, H35a-H40, H44-H50, H67-H71, H78-H82, H88-H94 and H102-H108 in the V_H domain according to the Kabat numbering scheme, were used for least-squares superposition of the domains. From these aligned sets of domains, lists of torsion angles, main-chain and side-

chain hydrogen bonds and residue and atom solvent-accessible surface areas were generated and analyzed.

Computer programs

Automated sequence alignments and similarity searches were performed using the appropriate modules of the GCG program package (Wisconsin Package V.9.0, Genetics Computer Group, Madison, Wisc. USA). Modeling and structural alignments were performed using the program InsightII versions 98 and 2000 (MSI/Biosym, San Diego, USA). To determine residue solvent-accessibility, the program NACCESS§ was used. EXCEL (Microsoft) Visual Basic utilities were written to combine, evaluate and visualize the data from the different analyses and to generate scripts to automate the comparative analysis of a large number of X-ray structures in InsightII.

References

1. Honegger, A. & Plückthun, A. (2001). Yet another numbering scheme for immunoglobulin variable domains: An automatic modeling and analysis tool. *J. Mol. Biol.* **309**, 657-670.
2. Winter, G. & Milstein, C. (1991). Man-made antibodies. *Nature,* **349**, 293-299.
3. Huston, J. S., Tai, M. S., McCartney, J., Keck, P. & Oppermann, H. (1993). Antigen recognition and targeted delivery by the single-chain Fv. *Cell Biophys.* **22**, 189-224.
4. Plückthun, A., Krebber, A., Krebber, C., Horn, U., Knüpfer, U., Wenderoth, R., Nieba, L., Proba, K. & Riesenberg, D. (1996). Producing antibodies in *Escherichia coli*: from PCR to fermentation. In *Antibody Engineering: A Practical Approach* (McCafferty, J. & Hoogenboom, H. R., eds), pp. 203-252, IRL Press, Oxford.
5. Wörn, A. & Plückthun, A. (2001). Stability engineeering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989-1010.
6. Young, N. M., MacKenzie, C. R., Narang, S. A., Oomen, R. P. & Baenziger, J. E. (1995). Thermal stabilization of a single-chain Fv antibody fragment by introduction of a disulphide bond. *FEBS Letters,* **377**, 135-139.
7. Jung, S. & Plückthun, A. (1997). Improving *in vivo* folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng.* **10**, 959-966.
8. Willuda, J., Honegger, A., Waibel, R., Schubiger, P. A., Stahel, R., Zangemeister-Wittke, U. & Plückthun, A. (1999). High thermal stability is essential for tumor targeting of antibody fragments: engineering of a humanized anti-epithelial glycoprotein-2 (epithelial cell adhesion molecule) single-chain Fv fragment. *Cancer Res.* **59**, 5758-5767.
9. Proba, K., Wörn, A., Honegger, A. & Plückthun, A. (1998). Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.* **275**, 245-253.

---

† http://immuno.bme.nwu.edu/
‡ http://www.unizh.ch/biochem/Databases
§ http://wolf.bms.umist.ac.uk/naccess/

10. Jung, S., Honegger, A. & Plückthun, A. (1999). Selection for improved protein stability by phage display. *J. Mol. Biol.* **294**, 163-180.

11. Martineau, P. & Betton, J. M. (1999). *In vitro* folding and thermodynamic stability of an antibody fragment selected in vivo for high expression levels in *Escherichia coli* cytoplasm. *J. Mol. Biol.* **292**, 921-929.

12. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A. & 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.

13. Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901-917.

14. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature,* **342**, 877-883.

15. Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). Structural repertoire of the human V_H segments. *J. Mol. Biol.* **227**, 799-817.

16. Barre, S., Greenberg, A. S., Flajnik, M. F. & Chothia, C. (1994). Structural conservation of hypervariable regions in immunoglobulins evolution. *Nature Struct. Biol.* **1**, 915-920.

17. Al-Lazikani, B., Lesk, A. M. & Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**, 927-948.

18. Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A. M. (1998). Conformations of the third hypervariable region in the V_H domain of immunoglobulins. *J. Mol. Biol.* **275**, 269-294.

19. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature,* **321**, 522-525.

20. Foote, J. & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224**, 487-499.

21. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, 5th edit., NIH Publication No. 91-3242, U.S. Department of Health and Human Services.

22. Kirkham, P. M., Mortari, F., Newton, J. A. & Schroeder, H. W., Jr (1992). Immunoglobulin V_H clan and family identity predicts variable domain structure and may influence antigen-binding. *EMBO J.* **11**, 603-609.

23. Saul, F. A. & Poljak, R. J. (1993). Structural patterns at residue positions 9, 18, 67 and 82 in the V_H framework regions of human and murine immunoglobulins. *J. Mol. Biol.* **230**, 15-20.

24. Saul, F. A. (1994). Structural implications of V_H sequence patterns. *Res. Immunol.* **145**, 61-66.

25. Brégégère, F., England, P., Djavadi-Ohaniance, L. & Bedouelle, H. (1997). Recognition of E. coli tryptophan synthase by single-chain Fv fragments: comparison of PCR-cloning variants with the parental antibodies. *J. Mol. Recogn.* **10**, 169-181.

26. Kipriyanov, S. M., Moldenhauer, G., Martin, A. C., Kupriyanova, O. A. & Little, M. (1997). Two amino acid mutations in an anti-human CD3 single chain Fv antibody fragment that affect the yield on bac-

27. de Haard, H. J., Kazemier, B., van der Bent, A., Oudshoorn, P., Boender, P., van Gemen, B., Arends, J. W. & Hoogenboom, H. R. (1998). Absolute conservation of residue 6 of immunoglobulin heavy chain variable regions of class IIA is required for correct folding. *Protein Eng.* **11**, 1267-1276.

28. Langedijk, A. C., Honegger, A., Maat, J., Planta, R. J., van Schaik, R. C. & Plückthun, A. (1998). The nature of antibody heavy chain residue H6 strongly influences the stability of a V_H domain lacking the disulfide bridge. *J. Mol. Biol.* **283**, 95-110.

29. Faber, C., Shan, L., Fan, Z., Guddat, L. W., Furebring, C., Ohlin, M., Borrebaeck, C. A. & Edmundson, A. B. (1998). Three-dimensional structure of a human Fab with high affinity for tetanus toxoid. *Immunotechnology,* **3**, 253-270.

30. Trakhanov, S., Parkin, S., Raffai, R., Milne, R., Newhouse, Y. M., Weisgraber, K. H. & Rupp, B. (1999). Structure of a monoclonal 2E8 Fab antibody fragment specific for the low-density lipoprotein-receptor binding region of apolipoprotein E refined at 1.9 Å. *Acta Crystallog. sect. D,* **55**, 122-128.

31. Bentley, G. A., Boulot, G., Riottot, M. M. & Poljak, R. J. (1990). Three-dimensional structure of an idiotope-anti-idiotope complex. *Nature,* **348**, 254-257.

32. Jung, S., Spinelli, S., Schimmele, S., Honegger, A., Pugliese, L., Cambilleau, C. & 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 V_H domains. *J. Mol. Biol.* **309**, 701-716.

33. Spada, S., Honegger, A. & Plückthun, A. (1998). Reproducing the natural evolution of protein structural features with the selectively infective phage (SIP) technology. The kink in the first strand of antibody kappa domains. *J. Mol. Biol.* **283**, 395-407.

34. Larrick, J. W., Danielsson, L., Brenner, C. A., Abrahamson, M., Fry, K. E. & Borrebaeck, C. A. (1989). Rapid cloning of rearranged immunoglobulin genes from human hybridoma cells using mixed primers and the polymerase chain reaction. *Biochem. Biophys. Res. Commun.* **160**, 1250-1256.

35. de Boer, M., Chang, S. Y., Eichinger, G. & Wong, H. C. (1994). Design and analysis of PCR primers for the amplification and cloning of human immunoglobulin Fab fragments. *Hum. Antibodies Hybridomas,* **5**, 57-64.

36. Zhou, H., Fisher, R. J. & Papas, T. S. (1994). Optimization of primer sequences for mouse scFv repertoire display library construction. *Nucl. Acids Res.* **22**, 888-889.

37. Welschof, M., Terness, P., Kolbinger, F., Zewe, M., Dübel, S., Dorsam, H., Hain, C., Finger, M., Jung, M. & Moldenhauer, G. *et al.* (1995). Amino acid sequence based PCR primers for amplification of rearranged human heavy and light chain immunoglobulin variable region genes. *J. Immunol. Methods,* **179**, 203-214.

38. Gilliland, L. K., Norris, N. A., Marquardt, H., Tsu, T. T., Hayden, M. S., Neubauer, M. G., Yelton, D. E., Mittler, R. S. & Ledbetter, J. A. (1996). Rapid and reliable cloning of antibody variable regions and generation of recombinant single chain antibody fragments. *Tissue Antigens,* **47**, 1-20.

39. Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R. & Plückthun, A. (1997). Reliable cloning of functional antibody vari-

able domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods,* **201**, 35-55.

40. Froyen, G., Hendrix, D., Ronsse, I., Fiten, P., Martens, E. & Billiau, A. (1995). Effect of V_H and V_L consensus sequence-specific primers on the binding and neutralizing potential of a single-chain Fv directed towards HuIFN-gamma. *Mol. Immunol.* **32**, 515-521.

41. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. (1993). X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.* **229**, 969-995.

42. Wörn, A. & Plückthun, A. (1998). An intrinsically stable antibody scFv fragment can tolerate the loss of both disulfide bonds and fold correctly. *FEBS Letters,* **427**, 357-361.

43. Wörn, A. & Plückthun, A. (1998). Mutual stabilization of V_L and V_H in single-chain antibody fragments, investigated with mutants engineered for stability. *Biochemistry,* **37**, 13120-13127.

44. Wörn, A. & Plückthun, A. (1999). Different equilibrium stability behavior of scFv fragments: identification, classification, and improvement by protein engineering. *Biochemistry,* **38**, 8739-8750.

45. Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N. & Hamers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature,* **363**, 446-448.

46. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A. & Hamers, R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng.* **7**, 1129-1135.

47. van der Linden, R. H., Frenken, L. G., de Geus, B., Harmsen, M. M., Ruuls, R. C., Stok, W., de Ron, L., Wilson, S., Davis, P. & Verrips, C. T. (1999). Comparison of physical chemical properties of llama V_{HH} antibody fragments and mouse monoclonal antibodies. *Biochim. Biophys. Acta,* **1431**, 37-46.

48. Johnson, G., Kabat, E. A. & Wu, T. T. (1996). The Kabat database of sequences of proteins of immunological interest. In *Weir's Handbook of Experimental Immunology. I. Immunochemisty and Molecular Immunology* (Herzenberg, L. A., Weir, W. M., Herzenberg, L. A. & Blackwell, C., eds), 5th edit., Blackwell Science Inc., Cambridge, MA.

49. Zdanov, A., Li, Y., Bundle, D. R., Deng, S. J., MacKenzie, C. R., Narang, S. A., Young, N. M. & Cygler, M. (1994). Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7 Å resolution. *Proc. Natl Acad. Sci. USA,* **91**, 6423-6427.

50. Li, H., Dunn, J. J., Luft, B. J. & Lawson, C. L. (1997). Crystal structure of Lyme disease antigen outer surface protein A complexed with a Fab. *Proc. Natl Acad. Sci. USA,* **94**, 3584-3589.

51. Whitlow, M., Howard, A. J., Wood, J. F., Voss, E. W., Jr & Hardman, K. D. (1995). 1.85 Ångstrom structure of anti-fluorescein 4-4-20 Fab. *Protein Eng.* **8**, 749.

Edited by I. Wilson