doi:10.1006/jmbi.2001.4665 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **309**, 701–716

JMB

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

Sabine Jung[1], Silvia Spinelli[2], Bernhard Schimmele[1] Annemarie Honegger[1], Luisa Pugliese[2], Christian Cambillau[2] and Andreas Plückthun[1]*

[1]*Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190 CH-8057 Zürich Switzerland*

[2]*Architecture et Fonction des Macromolécules Biologiques UMR 6098-CNRS 31 Chemin Joseph Aiguier F-13402 Marseille Cedex 20 France*

**Corresponding author*

The N-terminal segment (FR-H1) of the heavy chain (V_H) of antibodies shows significant conformational variability correlating with the nature of the amino acids H6, H7 and H10 (Kabat H9). In this study, we have established a causal relationship between the local sequence and the structure of this framework region and linked this relationship to important biophysical properties such as affinity, folding yield and stability. We have generated six mutants of the scFv fragment aL2, covering some of the most abundant amino acid combinations in positions H6, H7 and H10 (according to a new consensus nomenclature, Kabat H9). For the aL2 wild-type (w.t.) with the sequence $6_Q7_P10_A$ and for two of the mutants, the X-ray structures have been determined. The structure of the triple mutant aL2-$6_E7_S10_G$ shows the FR-H1 backbone conformations predicted for this amino acid combination, which is distinctly different from the structure of the w.t, thus supporting our hypothesis that these residues determine the conformation of this segment. The mutant aL2-$6_E7_P10_G$ represents a residue combination not occurring in natural antibody sequences. It shows a completely different, unique structure in the first β-strand of V_H, not observed in natural Fv fragments and forms a novel type of diabody. Two V_H domains of the mutant associate by swapping the first β-strand. Concentration-dependent changes in Trp fluorescence indicate that this dimerization also occurs in solution. The mutations in amino acids H6, H7 and H10 (Kabat H9) influence the dimerization behavior of the scFv and its thermodynamic stability. All the observations reported here have practical implications for the cloning of Fv fragments with degenerate primers, as well as for the design of new antibodies by CDR grafting or synthetic libraries.

Keywords: immunoglobulin variable domains; framework structure; scFv; ampicillin; antibody engineering

E-mail address of the corresponding author: plueckthun@biocfebs.unizh.ch

Present addresses: S. Jung, Bayer AG, Central Research, Biotechnology/Molecular Biology, Build. Q18, D-51368 Leverkusen, Germany; L. Pugliese, IRCC Institute for Cancer Research, University of Torino Medical School, Strada Provinciale 142 ,I-10060 Candiolo (Torino) Italy.

Abbreviations used: ABG, antibody germline database; CDR, complementarity-determining region; ELISA, enzyme-linked immunosorbent assay; FR, framework region; IMAC, immobilized metal ion affinity chromatography; MPD, 2-methyl-pentane-2,4-diol; NTA, nitrilotriacetic acid; PDB, Protein Data Bank; scFv, single-chain variable fragment of an antibody; V_H, variable domain of antibody heavy chain; V_L, variable domain of antibody light chain; w.t., wild-type.

0022-2836/01/030701–16 $35.00/0

Introduction

The overall fold of the antibody variable domains is strongly conserved, as can be deduced from the large number of X-ray structures available in the Brookhaven Protein Data Bank (PDB). Currently (June 2000), the PDB† contains the 3D structures of 281 V_H domains representing 181 non-identical sequences, 269 V_κ domains representing 183 non-identical sequences and 56 V_λ-domains representing 28 non-identical sequences. In studying the structural variability of these domains, most of the attention has been focused on the three hypervariable regions that form the majority of the antigen contacts and were thus termed complementarity-determining regions (CDR 1-3).[1,2]

A subset of the framework residues, consisting of the residues that are characterized by comparatively low sequence variability and are well embedded in the hydrogen bonding network of the two beta-sheets defining the immunoglobulin structure, shows particularely low structural variability between different molecules (low deviation of 3D aligned structures from the average structure[3]) and low structural flexibility within individual molecules (low B-factors and comparison of different structures of the same antibody). Forming a highly conserved core structure (average rms deviation from the average C^α position for all the the V_H domains in the PDB database, 0.40 Å), these C^α positions can be used as a reference system, relative to which conformational difference between the frameworks can be analyzed and described.[3] Besides these highly conserved framework residues, there are other parts of the framework that show a significant structural divergence between the different germline clans defining the immune repertoire of man and mouse.

In this study, we investigate the sequence differences causing the marked structural divergence observed in the framework 1 region of the antibody V_H domain (Figure 2 of Honegger & Plückthun,[4] and see Figure 3(b)), and their influence on stability and production yield. Saul and Poljak[5] described the correlation between three of the characteristic main-chain conformations occuring in human V_H domains and the type of amino acid found in positions H10 (H9)‡ and H74 (H67) as well as the side-chain orientation of the residues in positions H10 (H9), H19 (H18), H74 (H67) and H93 (H82). Honegger & Plückthun[4] confirmed this correlation and extended it to include residues H6 and H7. Based on the conformation of the polypeptide backbone between residues H6 and H12 (H11), they grouped human V_H domain into three and murine V_H domains into four distinct structural subtypes (type I to IV) that strongly correlate with the type of amino acids at positions H6, H7 and H10 (H9) in the published structures. The subtypes can be characterized by the most prevalent amino acids at positions H6, H7 and H10 (H9): $6_E7_S10_P$ (type I), $6_E7_S10_G$ (type II), $6_Q7_S9_A$ (type III) and $6_Q7_P10_A$ (type IV).

Several researchers observed drastic effects of mutations affecting residue H6 on antibody functionality and stability.[6–9] H6 is either a glutamine or a glutamate residue in almost all V_H sequences. Kipriyanov *et al.*[7] found that repair of a primer-induced Gln to Glu substitution markedly improved expression yield and shelf-life. Langedijk *et al.*[9] could compensate the stability loss in a V_H lacking the conserved disulfide bridge by a glutamate to glutamine exchange at position H6, reverting a similar substitution. Brégérère *et al.*[6] observed a severe decrease in stability caused by a primer-induced Gln H6 to Glu exchange, and de Haard *et al.*[8] even found a complete loss of binding and severe *in vivo* folding problems caused by an amino acid exchange at position H6 towards glutamate. In all four cases, the nature of the amino acid at position H6 had been changed accidentally by degenerate primers during the cloning of the gene for the expression of a recombinant protein. Langedijk *et al.*[9] first pointed out that the different H-bonding requirements of the two amino acids could be satisfied only by different side-chain conformations of Glu and Gln.

To address the question of whether the amino acids in positions H6, H7 and H10 (H9) indeed cause a particular framework conformation, as postulated by Honegger & Plückthun,[4] or whether the observed correlation is just an effect of both local sequence and global structure being dependent on the overall sequence context determined by the germline clan, we introduced different residue combinations into a scFv fragment of known structure to evaluate their influence on antigen affinity, production yield, stability and structure. The scFv fragment aL2 served as a model system for this study, as its crystal structure was available.[10] This scFv is one of several anti-β-lactam scFv fragments that had been isolated from a murine immune library by phage display[11] after immunization with ampicillin. In addition to the wild-type (w.t.) aL2 (type IV $6_Q7_P10_A$), we determined the X-ray structures of a mutant in which H6, H7 and H10 (H9) had been changed to the consensus pattern of a different subtype (type II $6_E7_S10_G$) and of a mutant representing a combination not observed in natural antibodies ($6_E7_P10_G$), and analyzed the effects of these and other mutations on biophysical properties such as affinity, thermodynamic stability and folding efficiency.

† http://www.rcsb.org/

‡ For residue numbering, the unified numbering scheme introduced by Honegger & Plückthun[3] is used throughout. In addition, the traditional Kabat numbering is indicated in parentheses.

Table 1. Allowed combinations of residues

Type	H6	H7	H9	H10
I	Glu	Not Pro	Gly	Pro
II	Glu	Not Pro	Gly	Gly
III	Gln	Not Pro	Gly	Any
IV	Gln	Pro	Gly	Any

Results

The aL2 model system

The ampicillin-binding scFv aL2 was derived from a phage display library prepared from immunized mice and was characterized in great detail, including the determination of its X-ray structure.[10] In cloning this antibody, a series of closely related scFvs with V_H domain sequences belonging to either structural subtype III ($6_Q7_S10_A$) or the less prevalent subtype IV ($6_Q7_P10_A$) were isolated. The 25 most closely related murine germline sequences found in the ABG database† all showed the sequence pattern characteristic for subtype IV ($6_Q7_P10_A$), and analysis of the primers used for cloning[11] revealed that the sequence diversity in this position could well have been introduced by the primer mix used for PCR, although a somatic mutation introduced before cloning cannot be excluded.

The aL2 scFv was chosen as a model to study the relationship between local sequence changes and the FR-H1 structure. To test the hypothesis that residues H6, H7 and H10 together determine the structural subtype of V_H, independent of the global sequence context defined by the germline clan, these residues were mutated by PCR mutagenesis to the consensus amino acids diagnostic for the four structural subtypes I to IV found in natural antibody sequences (Table 1), as well as to some combination not normally found in antibodies. The mutants listed in Table 2 were expressed in the periplasm of *Escherichia coli* and purified either by hapten affinity chromatography as described by Burmester *et al*.[10] or by a combination of Ni^{2+}-NTA chromatography and cation-exchange chromatography (see Materials and Methods). The w.t. aL2 ($6_Q7_P10_A$) was produced with reasonable yields in the periplasm of *E. coli* (200-700 μg per absorbance unit), while the production yields of the mutants were significantly lower (Table 2).

Biophysical characterisation of the aL2 w.t. and its mutants

The thermodynamic stability of the mutated scFvs was determined by urea denaturation. The shift of the fluorescence emission wavelength maxima was plotted against the urea concentration for all mutants (Figure 1). The urea denaturation was completely reversible for all constructs, as indicated by identical curves measured in denaturation and renaturation experiments (data not shown). Antibody scFvs frequently show complex urea denaturation curves.[12–14] Dependent on the differences in the intrinsic stabilities of the V_L and V_H domains, as well as on the contribution of the dimer interface to the overall stability, they can show anything from a fully cooperative unfolding transition, best described by a simple two-state model, to complex multi-state unfolding transitions. As a consequence, the ΔG values derived from such curves have to be treated with caution.[12–14] Interpretation of the measured denaturation curves for aL2 and its mutants by a three-state model did not yield a significantly better fit than a two-state model. Therefore the ΔG and m values listed in Table 2 were derived from a fit to a two-state model.

The w.t. aL2 ($6_Q7_P10_A$) was the most stable of the aL2 variants tested and with a denaturation midpoint of 5.6 M urea and a ΔG of 78 kJ/mol, its stability compares favorably with other antibody single-chain fragments. The mutant aL2-$6_Q7_P10_P$ was only slightly less stable than the w.t.. Structurally, it should fall into the same subtype (type IV) as the w.t. aL2-$6_Q7_P10_A$. Replacement of glutamine residue H6 by glutamate led to a significant destabilization. The non-natural combination aL2-$6_E7_P10_P$ was the least stable of the mutants tested, with a denaturation midpoint of 4.0 M, while the consensus combinations $6_E7_S10_P$ (type I) and $6_E7_S10_G$ (type II), and the non-natural combination $6_E7_P10_G$ showed intermediate stabilities with denaturation midpoints of 4.4 M to 4.7 M urea. In all combinations tested, the variant with a proline residue in position H10 (Figure 1, open symbols) was somewhat less stable than a similar construct with a more flexible Ala or Gly residue in this position (Figure 1, filled symbols).

Hapten binding constants

The aL2 scFv contains seven tryptophan residues. Of these, two are located in highly conserved positions (L43 (L35) and H43 (H36)) buried in the core of the domains. The fluorescence of these two residues is highly quenched in the native structure. The other five tryptophan residues are all located within or in close proximity to the antigen-binding pocket. The unliganded w.t. aL2 scFv showed a

† http://www.ibt.unam.mx/vir/V_mice.html

Table 2. Biophysical characterization of scFv aL2 and its V_H mutants

aL2-mutant	Expected structural subclass	Yield (μg/l *A*)	K_D (M)	Midpoint urea denaturation (M) (λ_{max}-plot)	ΔG (kJ/mol)	*m* (kJ/mol M)
$6_Q7_P10_A$	IV	200-700	1.6×10^{-6}	5.6	78.1	14.0
$6_Q7_P10_P$	IV	60	1.4×10^{-6}	5.3	44.7	8.4
$6_E7_S10_G$	II	50	-[a]	4.7	37.5	8.0
$6_E7_S10_P$	I	n.d.	3.3×10^{-6}	4.5	28.9	6.4
$6_E7_P10_G$	?	10-40	-[a]	4.4	33.0	7.5
$6_E7_P10_P$	?	30	-[a]	4.0	25.4	6.3

n.d., Not determined.
[a] Not evaluated quantitatively from fluorescence titration experiments due to a biphasic intensity curve.

strong tryptophan fluorescence, which was quenched upon binding of the hapten ampicillin (Figure 2(a)). This allowed the direct determination of antigen dissociation constants by fluorescence titration. In independent preparations, the dissociation constant of ampicillin was determined to be between 1.5×10^{-6} M[10] and 1.6×10^{-6} M (Table 2). The K_D of the mutant aL2-$6_Q7_P10_P$ was found to be 1.4×10^{-6} M and is thus indistinguishable from the affinity of the w.t.

The K_D titration curves of the constructs carrying glutamate in position H6 showed a different behavior. Initially, upon addition of small amounts of ampicillin, the fluorescence intensity, corrected numerically for the change in scFv concentration, decreased with a similar dependence on hapten concentration as for the w.t., consistent with a K_D very similar to that of the w.t. aL2 (Figure 2(a)-(c)). Upon addition of larger amounts of antigen, however, the fluorescence increased again, in some cases to values significantly higher than the starting value (Figure 2(b)). For the mutant aL2-$6_E7_S10_P$, a K_D of 3.3×10^{-6} M was determined, for the constructs aL2-$6_E7_S10_G$, aL2-$6_E7_P10_G$ and aL2-$6_E7_P10_P$, a reliable fit was not possible. To determine the cause of this unexpected fluorescence change, β-lactamase, which hydrolyzes ampicillin to a product no longer recognized by the antibody, was added at the end of the titration (Figure 2(c)). While the fluorescence quenching by the hapten could be reverted by addition of β-lactamase, the fluorescence increase was not affected by the addition of β-lactamase, indicating that it is an effect of the antibody dilution, even though the data are already normalized for the volume increase. Indeed, this fluorescence increase did not depend on the addition of ampicillin. ''Titration'' of aL2 with a buffer solution not containing any ampicillin, i.e. a simple dilution, led to a similar fluorescence increase (Figure 2(c)). This effect, first observed in aL2-$6_E7_S10_G$, was even more marked in mutants aL2-$6_E7_P10_G$ and aL2-$6_E7_P10_P$.

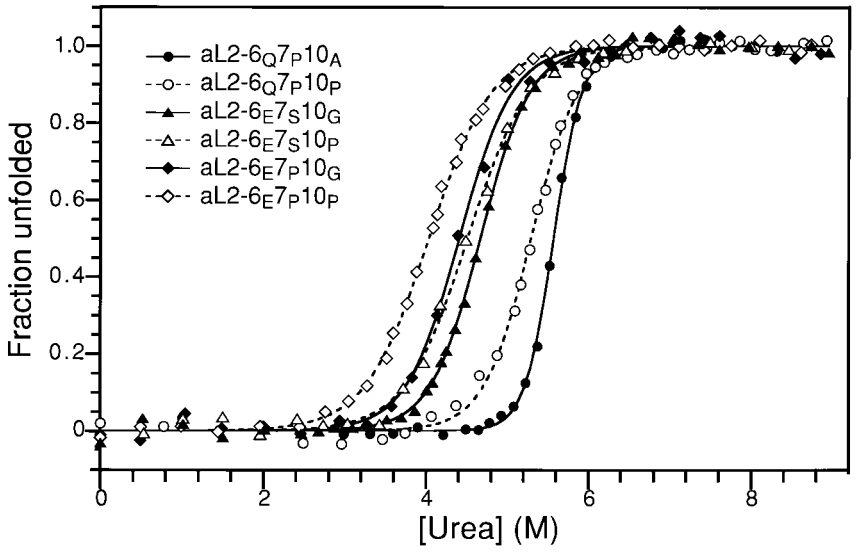

Figure 1. Comparison of the intrinsic stabilities of aL2 mutants by urea equilibrium denaturation curves of w.t. aL2-$6_Q7_P10_A$ (●) and mutants aL2-$6_Q7_P10_P$ (○), aL2-$6_E7_S10_G$ (▲), aL2-$6_E7_S10_P$ (△), aL2-$6_E7_P10_G$ (◆) and aL2-$6_E7_P10_P$ (◇) monitored by the shift of the emission maximum wavelength of the intrinsic protein fluorescence as a function of urea concentration and fitted according to Pace.[31] Numeric values for ΔG and *m* derived from these curves are listed in Table 2.

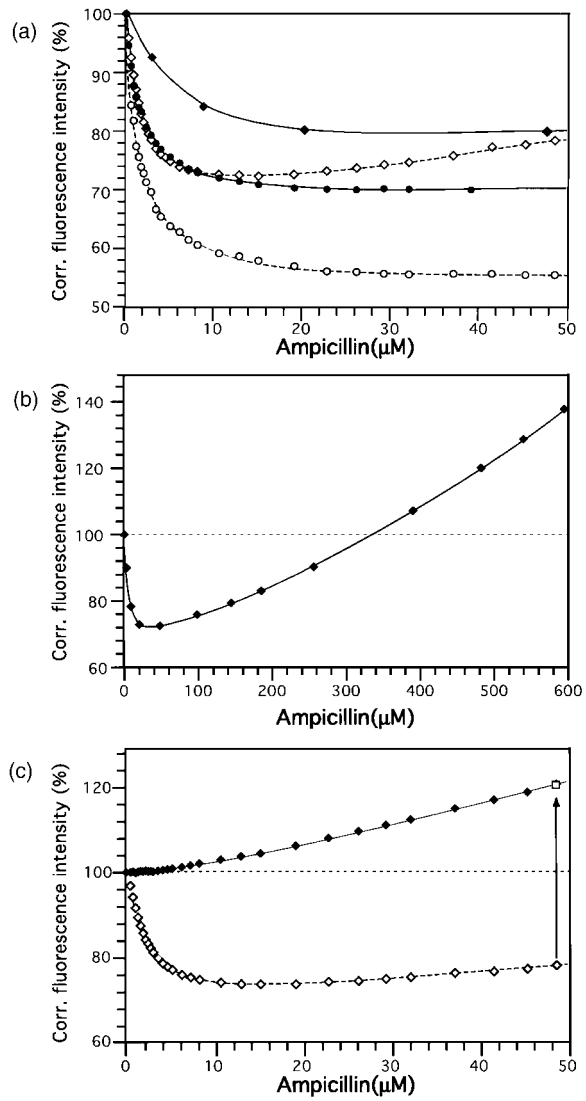

Figure 2. Ampicillin titration curves of the different aL2 mutants monitored by the change of the intrinsic protein fluorescence. The fluorescence values are corrected for dilution and are normalized to the fluorescence value measured in the absence of any ampicillin (100 %). (a) Binding of ampicillin to w.t. aL2-$6_Q7_P10_A$ (●) and mutants aL2-$6_Q7_P10_P$ (○), aL2-$6_E7_P10_G$ (◆) and aL2-$6_E7_P10_P$ (◇). (b) Titration of mutant aL2-$6_E7_P10_P$ with ampicillin. (c) Titration of aL2-$6_E7_P10_G$ with ampicillin (◇) or with buffer not containing any hapten(◆). Fluorescence intensity measured after addition of β-lactamase (□) to the sample with the highest ampicillin concentration.

Apparently, the mutant aL2 molecules showing this non-linear dependence of fluorescence intensity on protein concentration interact in a fast dimerization or multimerization equilibrium, which affects their intrinsic fluorescence. Dilution, leading to an equilibrium shift towards the monomeric species, leads to an increase in the molar fluorescence intensity (data are always corrected for dilution). Unfortunately, other methods to verify a fast monomer-dimer equilibration did not give precise results. Light-scattering measurements confirmed that the protein solution was not monodisperse, indicating the presence of dimers and higher aggregates (data not shown). Gel-filtration chromatography of the aL2 scFv and its mutants did not yield any information on the size distribution of the aggregates, since the protein bound to the Superose and Sephadex column materials and eluted at an apparent molecular mass less than the expected monomeric molecular mass.

Crystallization and structure determination

The wild-type aL2 and two of the mutants yielded crystals of sufficient quality to determine their structure by X-ray crystallography. Other mutants either failed to crystallize under the conditions tested or yielded crystals of insufficent size and quality. Details of the successful crystallization and structure determination of w.t. aL2-$6_Q7_P10_A$ and of the two mutants aL2-$6_E7_P10_G$ and aL2-$6_E7_S10_G$ are described in an accompanying paper.[10] Data collection and refinement statistics for the mutants are summarized in Table 3. The mutant aL2-$6_E7_P10_P$ did not crystallize under any of the conditions tested, and aL2-$6_Q7_P10_P$ yielded only small crystals of insufficent quality for crystallographic analysis.

Structure

The w.t. aL2 scFv has an extremly short CDR H3 (only three residues) as well as a short CDR L1, resulting in a compact structure with little main-chain flexibility. The structures of w.t. aL2-$6_Q7_P10_A$ crystal forms 1 and 2, aL2-$6_E7_S10_G$ and aL2-$6_E7_P10_G$ could be superimposed by a least-squares fit of the C^α positions of residues L1-L147 and H11-H148 (Kabat L1-L106 and H10-H112) with rms deviations of 0.39 Å (w.t. forms 1 and 2), 0.50 Å (w.t. form 1 and aL2-$6_E7_S10_G$) and 0.57 Å/0.77 Å (w.t. form 1 and aL2-$6_E7_P10_G$ first and second molecule in asymmetric unit) (Figure 3(a)). Residues H1-H10 (H9) were not used for superpositions, since, as expected, the conformation in this region differed between the w.t. and the two mutants.

In the crystals of w.t. aL2-$6_Q7_P10_A$, crystal forms 1 and 2, and in the mutant aL2-$6_E7_S10_G$, crystal contacts are obstructing the antigen-binding site. FR-H1 (residues H14-H20, Kabat H13-H19) and FR-H3 (residues H75-H81 and H92-H99, Kabat H64-H70 and H81-H85) of the V_H domain of a symmetry-related molecule packed against the antigen-binding site in such a way that all six CDRs were involved in crystal contacts (depicted in Figure 4 of Burmester *et al.*[10]). In the w.t. aL2-$6_Q7_P10_A$ crystals of form 1, grown in the presence or absence of hapten, a molecule of methylpentane diol (MPD) occupied what remained of the antigen-binding pocket. In w.t. aL2-$6_Q7_P10_A$

Table 3. Data collection and final refinement statistics

	aL2-mutant $6_E7_P10_G$ form 3-free	aL2-mutant $6_E7_P10_G$ form 3-complex	aL2-mutant $6_E7_S10_G$ form 4-free
Space group	$P2_12_12_1$	$P2_12_12_1$	$P4_12_12$
Molecules/a.u.	2	2	1
Data collection			
V_M (Å^3/Da)	2.40	2.40	2.43
Data resolution	13-2.75	13-2.40	20-1.85
Observations	81,293	93,952	303,033
Unique	14,189	20,445	21,449
Completeness (%)[a]	97.5/90.7	98.0/82.8	98.8/98.8
R_{sym} (%)[a]	11.0/40	8.2/22.5	8.2/27.5
Refinement			
Reflections	13,978	19,137	20,522
Resolution	13-2.75	13-2.40	18-1.85
R-value/*R*-free (%)	17.7/25.7	20.5/27.7	23.0/27.0
Atoms	3400	3400	1812
Solvent atoms	110	298	104
B-factors (Å^2)			
Main-chains	31.3	20.9	31.0
Side-chains	35.0	23.8	35.0
Solvent	34.5	29.5	46.0
rms deviations			
Bonds (Å)	0.010	0.011	0.013
Angles (deg.)	1.60	1.58	1.80
Dihedral (deg.)	26.7	26.8	27.7
Improper (deg.)	0.90	1.16	1.10
Ramachandran regions[b]			
Most favored (%)	84.0	85.0	88.2
Additional (%)	15.5	14.5	11.2
Generously allowed (%)	0	0	0
Disallowed (%)	0.5 (Ala L67)	0.5 (Ala L67)	0.5 (Ala L67)

[a] Overall/last shell.
[b] With PROCHECK.[40]

crystals of form 2 and in the crystals of aL2-$6_E7_S10_G$, the pocket contained disordered solvent. This crystal packing was clearly not compatible with antigen-binding, as an ampicillin molecule in the binding pocket would overlap with the V_H domain obstructing the binding pocket. In the crystals of mutant aL2-$6_E7_P10_G$, different crystal contacts dominated, leading to a packing that was compatible with hapten binding. Ampicillin could be soaked into preformed crystals, and the structure of the complex could be determined. The hapten bound into a deep pocket along the pseudo 2-fold axis of the V_L/V_H heterodimer in such a way that the phenyl ring of ampicillin bound into the same position as the MPD molecule in the w.t. form 1 crystals (Figure 3(a)).

As predicted, the main-chain conformation of FR-H1 w.t. aL2-$6_Q7_P10_A$ and mutant aL2-$6_E7_S10_G$ differed. Figure 3(b) summarizes the predicted conformation for the four consensus patterns, Figure 3(c) shows the actual conformations observed in the structures of aL2 and its mutants. While in the two wild-type structures (indicated in green and yellow), FR-H1 clearly assumed a type IV conformation, FR-H1 of the mutant aL2-$6_E7_S10_G$ showed the predicted type II conformation. In the mutant aL2-$6_E7_P10_G$, with a combination of amino acids in positions H6, H7 and H10 not observed in natural antibodies, a novel FR-H1 conformation was observed. Residues H1-H10 did not pack into their own V_H domain, but pointed away from the domain (Figure 3(a)). Instead, the corresponding residues of the second scFv in the asymmetric unit (H1′-H10′) packed into their w.t. position (Figure 3(c)), while H1-H10 of the first molecule packed into the second molecule, resulting in a strand-swap (Figure 4) and the association of the two scFvs in the asymmetric unit into a novel form of diabody.

Residues H1′ to H4′ of the swapped strand assumed positions very similar to residues H1 to H4 in the w.t. structure. Residues H2′ and H4′ pointed into the core, while residues H1′ and H3′ were exposed. For residue H5′, the conformation in the swapped strand of the mutant was different from that of residue H5 in the cognate strand of the w.t. In the first β-strand of the w.t. domain, which is part of the outer β-sheet of the immunoglobulin domain in a normal type IV conformation, glutamine H5 was fully exposed on the surface of

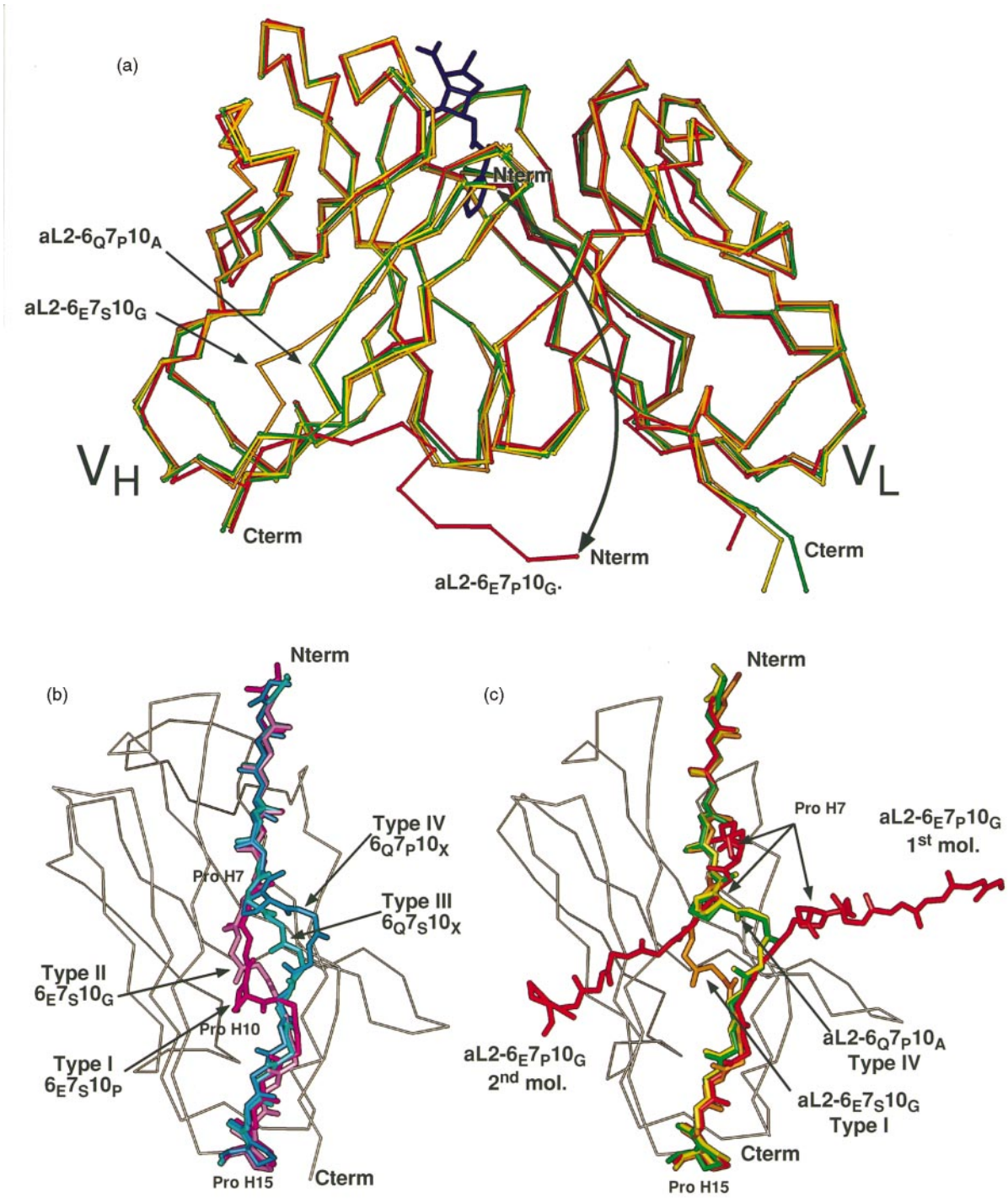

Figure 3. Overview of aL2 w.t. and mutant structures. (a) Superposition of the structures of w.t. aL2-$6_Q7_P10_A$ crystal form 1 (yellow) and form 2 (green), aL2-$6_E7_S10_G$ (orange) and aL2-$6_E7_P10_G$ (red) by a least-sqares fit of the C^α positions of residues L1-L147 and H11-H148 (Kabat L1-L106 and H10-H112). A double-headed arrow links the position of the N terminus of the V_H domains of aL2-$6_Q7_P10_A$ and aL2-$6_E7_S10_G$, and the N terminus of the V_H domain of aL2-$6_E7_P10_G$, emphasizing the conformational change. (b) Structural diversity of the FR-H1 segment in natural antibodies: type I (magenta), murine Fv fragment of antibody D1.3 (C. Marks *et al.*, unpublished results, PDB entry 1A7N, 2.0 Å resolution); type II (pink), human Fab B7-15A2[41] (PDB entry 1AQK, 1.84 Å resolution); type III (cyan), murine Fab 2e8[42] (PDB entry 12E8, 1.9 Å resolution); type IV (blue), first Fab fragment (chain B) in a murine idiotype-anti-idiotype complex[43] (PDB entry 1CIC, 2.5 Å resolution). (c) Structural diversity of the FR-H1 segment in the aL2 constructs: w.t. aL2-$6_Q7_P10_A$ crystal form 1 (yellow) and 2 (green) correspond to the type IV consensus structure, aL2-$6_E7_S10_G$ (orange) to the type II consensus structure and aL2-$6_E7_P10_G$ (red) shows a novel diabody structure in which the N-terminal segment is provided by a neighboring molecule.

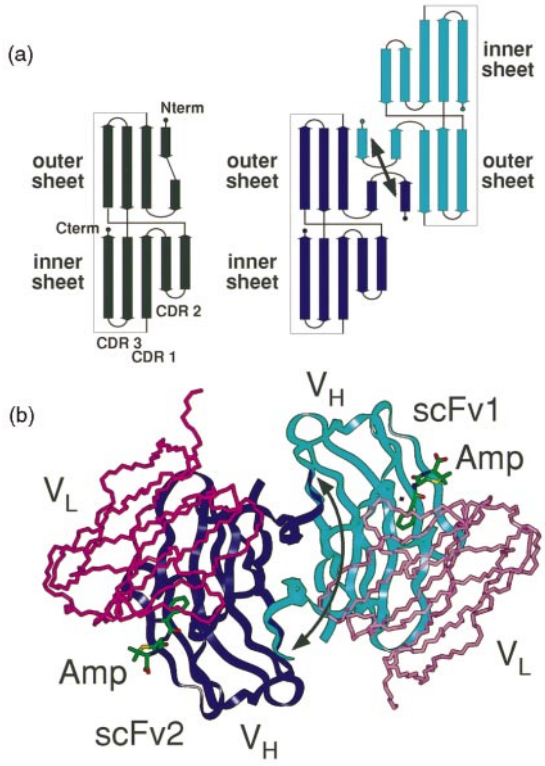

Figure 4. (a) A representation of the strand-swap in aL2-$6_E7_P10_G$ resulting in a novel diabody structure: while in w.t. aL2-$6_Q7_P10_A$, as in all normal V_H domains, the N-terminal segment (residues H2-H7) forms the first strand of the outer β-sheet of the domain, in the crystal structure of aL2-$6_E7_P10_G$, the V_H domains of the two scFvs in the asymmetric unit have exchanged their N-terminal segments in such a way that the N-terminal segment of one V_H domain participates in the outer β-sheet of the other V_H domain. (b) Interaction of the two scFvs in the asymmetric unit of the crystals of mutant aL2-$6_E7_P10_G$. The two V_H domains are indicated as cyan and blue ribbons, the V_L domains as pink and magenta backbone traces. Amp indicates the position of the hapten ampicillin. The double-headed arrow indicates the two N termini that have swapped their positions.

the molecule, while glutamine H6 was buried in the core of the domain, forming hydrogen bonds to main-chain atoms of the inner β-sheet and to the side-chain OH group of Thr H143 (H107) (Figure 5(a)). In the mutant aL2-$6_E7_P10_G$ structure, glutamine H5′ of the dimer partner was rotated in such a way that its side-chain amide group took over the position of the w.t. glutamine H6 side-chain amide group in the core of the domain (Figure 5(b)), while the glutamate H6′ side-chain was exposed. Residues H6′ to H10′ (H9′) of aL2-$6_E7_P10_G$ spanned the space between the two neighboring V_H domains, whereas in the regular V_H domain structure of the w.t. residues H7-H10 (H9) formed the transition region of FR-H1 from the outer to the inner β-sheet of the immunoglobulin domain.

Hence, the design of a combination of amino acid residues in positions H6, H7 and H10 (H9) that do not occur in nature yielded a novel scFv structure not seen in natural antibodies. The structure of the framework region in question was heavily affected, adopting a very unusual, non-canonical framework structure. The two mutants aL2-aL2-$6_E7_S10_G$ and aL2-$6_E7_P10_G$ differ by only a single residue. This finding reveals a key role of residue H7 in the subtype division. The fact that in natural sequences Pro H7 never occurs in combination with Glu H6 could thus be explained by the X-ray structure of aL2-$6_E7_P10_G$, which indicates that this residue combination cannot pack in the canonical V_H-fold.

Discussion

Antibody V_H domains can be classified into four distinct structural subtypes (Figure 3 of Honegger & Plückthun[4]), based on their distinct backbone conformation in the framework region 1 (FR-H1). The authors demonstrated, by an analysis of the known V_H structures, the existence of a striking correlation between the nature of the amino acid residues H6, H7 and H10 (H9 according to Kabat) of V_H and the divergent conformations of the polypeptide backbone between residues H6 and H11 (H10) (Table 1). While in natural antibodies the local sequence in these positions, the global sequence and, as a consequence, the tertiary structure of the domain all depend on the evolutionary history of the particular V_H germline clan coding for the domain and therefore do not vary independently of each other, the nature of the amino acids found in positions H6, H7 and H10 (H9) led us to the conclusion that these residues determine the local framework structure, independent of the global sequence context.

In this study, we have confirmed experimentally that the nature of residues H6, H7 and H10 (H9) indeed determines the FR-H1 structure independent of the global sequence context. Mutants of the antibody scFv aL2 were designed in which residues H6, H7 and H10 (H9) were changed to conform to the sequence pattern defining the four different structural subtypes (Table 1), as well as to some combinations not seen in natural antibody sequences (Table 2). The thermodynamically very stable single-chain fragment aL2-$6_Q7_P10_A$ was classified as type IV by sequence and showed the expected type IV FR-H1 structure as determined by X-ray crystallography (Figure 3(c)). Variants of aL2 belonging to type III ($6_Q7_S10_A$) arose spontaneously either by somatic mutation or during cloning and amplification and persisted throughout the phage display selection of the scFv.[10] Variants aL2-$6_E7_S10_P$ (type I) and aL2-$6_E7_S10_G$ (type II) were produced by PCR mutagenesis. The mutant aL2-$6_E7_S10_G$ could be crystallized and its structure determined at 1.85 Å resolution.

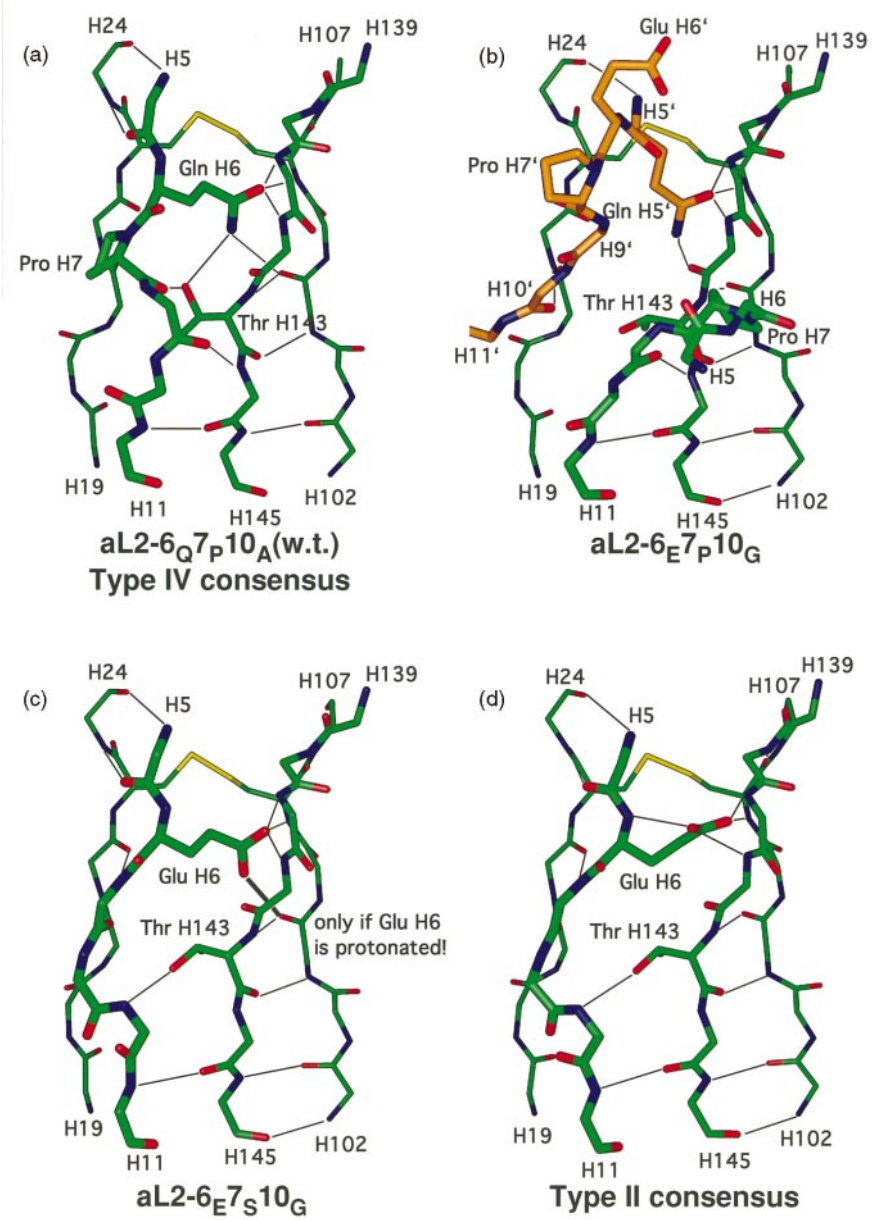

Figure 5. Detailed view of the hydrogen bonding pattern involving the side-chain of residue H6. (a) w.t. aL2-$6_Q7_P10_A$: the core hydrogen bonding pattern around Gln H6 corresponds to the consensus pattern described for the structural subtype IV.[4] The oxygen atom of the side-chain amide group can accept a hydrogen bond from two of the three main-chain NH groups of residues Gly H142 (H106) and Gly H140 (H104), which form the conserved β-bulge in FR-H4, and from Cys H106 (H92). The side-chain amide nitrogen atom of Gln H6 donates a hydrogen bond to the main-chain carbonyl group of Tyr H104 (H90) and to the side-chain hydroxyl group of Thr H143 (H107), which in turn donates a hydrogen bond to the main-chain carbonyl group of Pro H7. (b) Mutant aL2-$6_E7_P10_G$: residues H1-H10 point away from the domain. The N-terminal segment of a second molecule replaces the N-terminal strand. Gln H5′ binds into the position of Gln H6 of the w.t. structure in such a way that its side-chain oxygen and nitrogen atoms assume almost exactly the same positions as the side-chain oxygen and nitrogen atoms of Gln H6 of the w.t. structure. The side-chain of Thr H143 (H107) is rotated into the conformation normally seen in type I and type II structures. (c) Mutant aL2-$6_E7_S10_G$: glutamate H6 has the same side-chain conformation as glutamine H6 in the w.t. aL2. The hydrogen bond to the main-chain carbonyl group of Tyr H104 (H90) can be formed only if the Glu side-chain is protonated, but the short distance between the two oxygen atoms (2.67 Å, Figure 4) suggests strongly that it is indeed protonated. (d) Type II consensus hydrogen bonding pattern ($6_E7_S10_G$). One oxygen atom of the side-chain carboxylate group accepts two hydrogen bonds from the main-chain NH groups of Glu H6 and of Gly H142 (H106), the other from the main-chain NH groups of Cys H106 (H92) and of Gly H140 (H104). The side-chain of Thr H143 (H107) is rotated relative to the conformation seen in the type III and IV structures.

The kink conformation is determined locally by residues H6, H7 and H10

Although H6 is always hydrophilic, being either a glutamine or glutamate residue, its side-chain is fully buried in the core of the V_H domain. The two residue types satisfy their different hydrogen bonding requirements by assuming different side-chain conformations (Figure 5(a) and (d)). In both cases, this results in the termination of the first β-strand. In the segment between H7 and H10 (H9), the peptide chain switches from the outer β-sheet to the inner (dimer interface) β-sheet of the immunoglobulin domain β-sandwich. The exact shape of the kink in the chain is determined by the flexibility of the residues in positions H7 to H10 (H9) and by the presence or absence of a hydrogen bond from the hydroxyl group in the side-chain of Thr H143 (H107), whose orientation depends on the nature of residue H6, to the main-chain carbonyl group of H7.

If the different hydrogen bonding requirements of glutamate compared to glutamine, modulated by the steric preferences of the neighboring amino acids, were indeed the determining factor for the structural diversity observed in the framework 1 region, we would expect the structure of the mutant V_H to change from the type IV to the type II consensus pattern. The Glu side-chain would accept hydrogen bonds from the peptide NH groups of H6, H106 (H92), H141 (H105) and H142 (H106) (Figure 5(d)) and thus fully satisfy its hydrogen bonding requirements and compensate, to some extent, for the deleterious effects of burying a negative charge in the core of the domain. If, instead, the steric constraints of the core packing determined by the global sequence context dominated, in particular the steric relay from H74 to H78 to H93 to H19 to H10 (H63, H67, H82, H18, H9) across the β-sheet observed by Saul & Poljak[5] (Figure 6), we would expect a main-chain conformation of type III (as expected for Gln H6, Ser H7, Ala H10) or IV (as in the w.t. scFv), combined with a Glu-type side-chain conformation of residue H6.

In the X-ray structure of aL2-$6_E7_S10_G$, the main-chain conformation is shifted, as predicted[4], from a type IV to a type II conformation (Figure 3(c)). However, the electron density for the H6 side-chain is very clear and shows a Gln-typical side-chain conformation, juxtaposing the second oxygen atom to the main-chain carbonyl group of H104 (H90) (Figure 7). Nevertheless, the Thr H143 (H107) side-chain is rotated with respect to the consensus conformation found in type I and type II structures (Figure 5(d)), with the CH_3 group oriented towards the domain core and the OH group oriented towards the surface of the molecule. In this conformation, there is no hydrogen

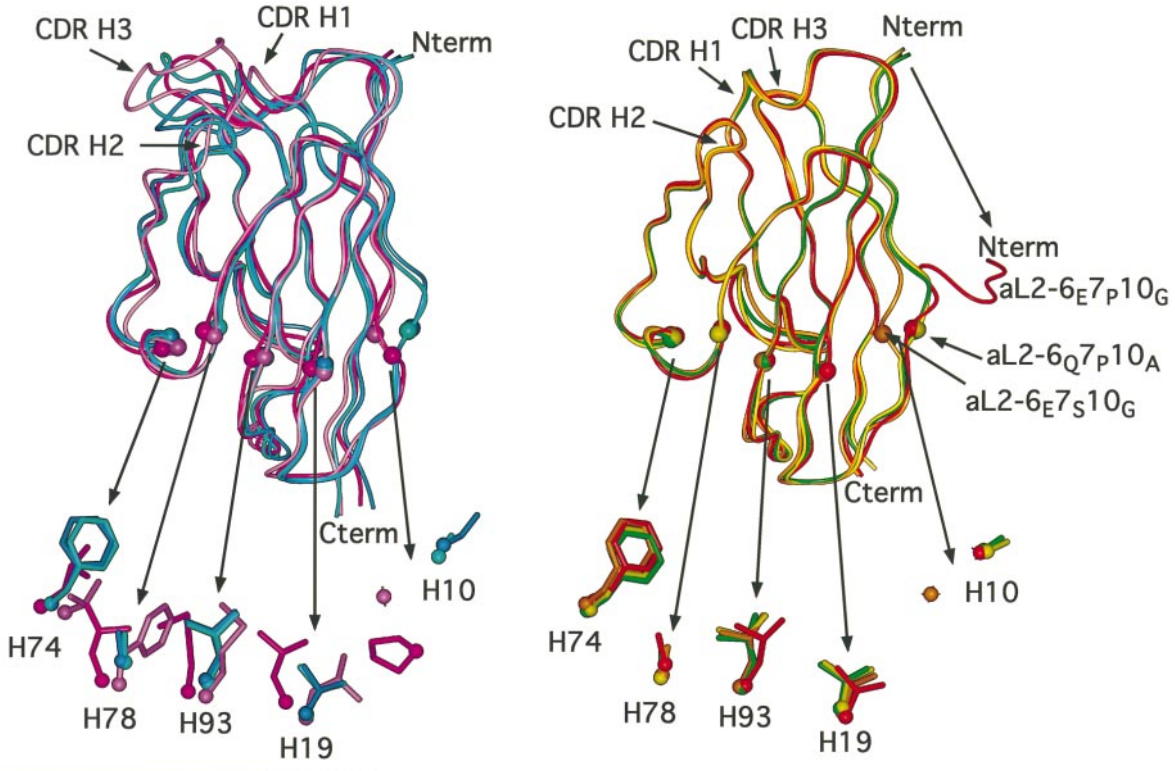

Figure 6. Correlation of FR-H1 conformation with the side-chain conformations and amino acid types across the core of the V_H domain.[4,5] (Left) Preferred amino acid types and side-chain conformations for the four structural subtypes: (type I, magenta; type II, pink; type III, cyan; and type IV, blue. (Right) Side-chain orientation of the different residues in w.t. aL2-$6_Q7_P10_A$ crystal form 1 (yellow) and form 2 (green), aL2-$6_E7_S10_G$ (orange) and aL2-$6_E7_P10_G$ (red).

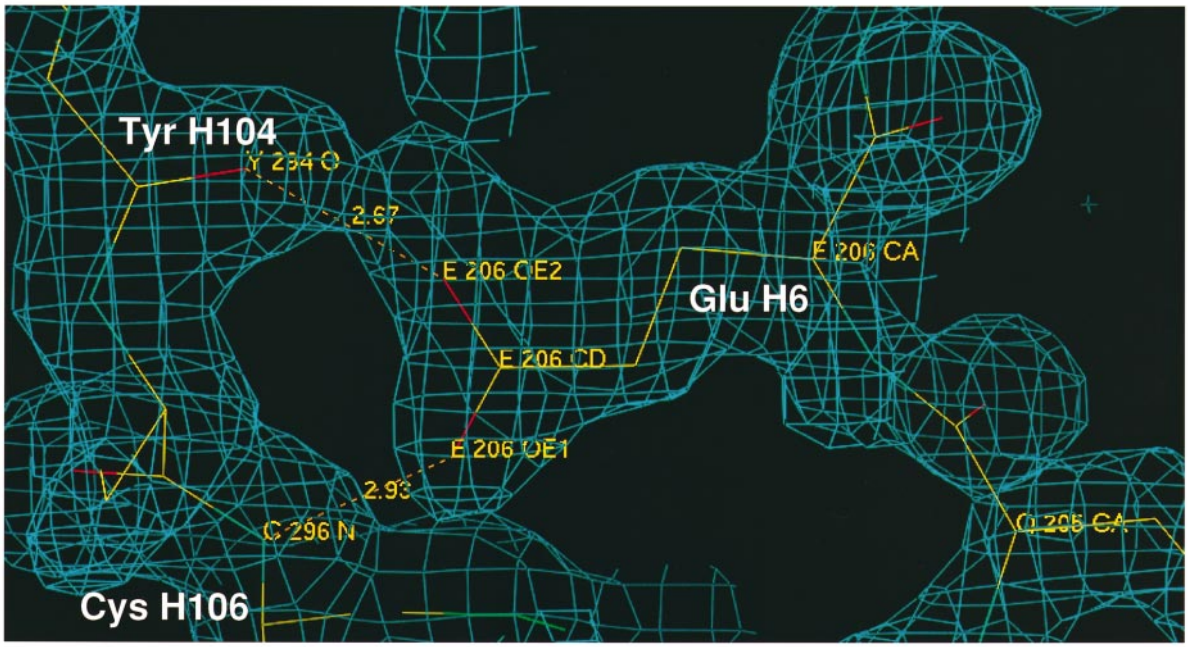

Figure 7. Side-chain conformation of glutamate H6 in the crystal structure of mutant aL2. The $2F_o - F_c$ electron density map was contoured at the 0.9 sigma level.

bond possible from the Thr side-chain to the main-chain carbonyl group of residue H10 (H9), and the main chain assumes the type II conformation defined by the Gly in position H10 (H9).

From the observed change of the FR-H1 main-chain conformation, we must conclude that this conformation depends mainly on the possibility of forming the hydrogen bond between the main-chain peptide carbonyl group of residue H7 and the side-chain OH of Thr H143 (H107), modulated by the steric requirements of the amino acid in position H10 (H9), rather than on the steric relay observed by Saul & Poljak[5], which would encourage a type III main-chain conformation. This view is substantiated by the fact that the side-chains of the residues involved in this steric relay remain largely in the type IV-typical conformation (Figure 6), indicating that the packing of the hydrophobic core is not disturbed significantly by the mutations. A deprotonated Glu H6 in the observed conformation would lose two potential hydrogen bonds compared to the type I and type II side-chain conformation, while a protonated Glu side-chain would lose one hydrogen bond. This loss of potential hydrogen bonds fits with the observation that the aL2-$6_E7_S10_G$ and the aL2-$6_E7_P10_G$ mutants are clearly destabilized compared to the w.t. aL2-$6_Q7_P10_A$, although they are significantly more stable than the aL2-$6_E7_P10_P$ mutant.

Protonation of Glu H6

None of the core residues surrounding the H6 side-chain correlates in conformation or identity with the different structural subtypes: Cys H23 (H22), Cys H106 (H92), Trp H43 (H36), Tyr H104 (H90), Gly H140 (H104) and Thr H143 (H107) represent the most conserved residues in the immunoglobulin variable domain structure, both in terms of sequence identity and of side-chain conformation. Only H21 (H20) (Leu, Ile or Val) shows some variability correlated with the germline clans. However, the aL2 wild-type contains Leu in this position, which is also the predominant amino acid found in the germline family prototypical for the type II structure. If this residue played any role in determining the H6 side-chain conformation in the aL2 mutants, it should encourage the Glu-specific side-chain conformation and not the Gln one. It is therefore very unlikely that the Gln-like side-chain conformation is forced on the Glu H6 side-chain by the packing of the surrounding residues.

The observed side-chain conformation can be explained if Glu H6 was protonated at the pH 5.8 used for crystallization. While this is higher than the pK_a of the γ-carboxylate group of a free glutamate (pK_a 4.07), the shielded environement of the fully buried Glu H6 side-chain will certainly influence its pK_a. A protonated Glu could donate a hydrogen bond to the main-chain carbonyl group of residue H104 (H90) and thus maintain a Gln-like side-chain conformation. However, it could not donate the second hydrogen bond to the side-chain OH group of Thr H143 (H107), which could therefore rotate into the type I/type II conformation observed. Since in this conformation the side-chain OH group of Thr 143 (H107) is not able to form a hydrogen bond to the H10 main-chain carbonyl group, the main-chain conformation of H7-H10 can switch to the type II conformation.

Eight out of 43 structures in the PDB carrying the Glu H6, Ser H7, Gly H10 combination in its natural context deviate from the type II consensus conformation of the H6 side-chain. Five of these (PDB entries 1FGV, 1FVC, 1IGM, 1CLO and 1AIF) show the same pattern as we observed for Glu H6, Ser H7, Gly H10, keeping the main chain and Thr H143 (H107) side-chain conformation. The protein in 1AIF was crystallized from a stock solution at pH 4[15] with no buffer added with the crystallization reagents, 1CLO at pH 5.6,[16] 1IGM at a pH between 6 and 8[17] and 1FGV at pH 6.5.[18] The antibody 1FVC (antibody 4D5 version 8) was crystallized at pH 7.5, the same pH as the closely related 1FVD and 1FVE, which show the consensus structure.[18] Most of these antibodies were thus crystallized at slightly acidic conditions, while for a reference sample of ten antibodies showing the consensus structure, neutral to basic conditions prevailed. Despite the unexpected Glu side-chain conformation, the aL2-$6_E7_S10_G$ X-ray structure showed the predicted change of main-chain conformation, confirming the hypothesis that the local rather than the global sequence determines the FR-H1 conformation.

Nature of the dimeric scFv

Mutant aL2-$6_E7_P10_G$ represents a sequence pattern not found in natural antibodies. A proline residue in position H7 is found only in combination with glutamine in position H6, resulting in a main-chain conformation (type IV) closely resembling that of a V_L lambda domain. Although the stability of this mutant was in the same range as those of aL2-$6_E7_S10_G$ and of aL2-$6_E7_S10_P$ (Figure 1, Table 2), structural analysis revealed an unconventional conformation of FR-H1. The residues H1-H7, which normally form the first β-strand of the outer β-sheet of the V_H domain, were turned out of the molecule. Instead, residues H1-H5 formed an anti-parallel β-sheet with residues H22′-H26′ of a neighboring V_H domain, whose N-terminal residues in turn replaced the first strand of the first molecule, forming a dimeric scFv (Figure 3). Although residues H1′-H4′ of the second molecule assume the same positions as the corresponding residues of the w.t. structure, Gln H5′ is turned around, pointing into the domain core, with its side-chain amide group group taking the place of and forming the same hydrogen bonds as the side-chain amide of Gln H6 of the w.t. structure, while Glu H6′ of the mutant is pointing outwards (Figure 5(b)).

This dimerized form is different from the usual form of diabodies. Normal diabodies result from a phenomenon termed domain swapping, where the V_H domain of one scFv is paired with the V_L domain of a partner scFv molecule and *vice versa*.[19] In contrast, the aL2 mutant aL2-$6_E7_P10_G$ shows dimerization by strand swapping, exchanging their first β-strands of FR-H1, which are mutually participating in the β-sheet of the partner molecule. This dimeric structure most likely is not simply a crystallization artefact, but present also in concentrated solutions. Although the protein concentration of the stock solution used for the K_D titration experiments (0.1-0.5 mg/ml) is well below that of the protein solution used for crystallization (3 mg/ml), the observed non-linear concentration dependence of the fluorescence intensity observed in the controls to the antigen titration experiments and the concentration dependence of the emission spectrum of the mutants strongly suggest interactions between the scFv molecules. Since the effect of quenching by the hapten was additive to the concentration effect and the two effects did not seem to compete, the atypical concentration dependence is most likely not due to the five tryptophan residues contributing to the antigen-binding site. However, two more tryptophan residues (L43 (L35) and H43 (H36)) are located in the cores of the V_L and the V_H domain. The dimer structure of the aL2-$6_E7_P10_G$ mutant offers a possible explanation of how the environment and thus the contribution to the fluorescence spectrum of one of these remaining tryptophan residues could be altered without denaturing the domain. The high rate with which the observed fluorescence increase occurs (solutions appeared to be in equilibrium directly after manual mixing), suggests that these novel diabodies may have much faster equilibration kinetics than the classical diabodies. This is plausible, since the dissociation of a single strand should require a smaller activation energy than the dissociation of the V_H and V_L domain.

The conserved tryptophan residue H43 (H36) in the core of the V_H domain is located close to Glu H6 and highly quenched in the w.t. structure. Upon denaturation of isolated V_H domains in guanidinium hydrochloride or urea, an increase of fluorescence intensity as well as a shift of the maximum to higher wavelengths is observed. In the case of mutant aL2-$6_E7_P10_G$, the dissociation of the N-terminal strand, provided by the second V_H domain, most probably would affect the side-chain packing around Trp H43 (H36), and thus reduce its quenching. Ampicillin binding reduces the fluorescence intensity without significantly affecting the wavelength of the maximum. In the mutants that show a fluorescence increase upon dilution, the wavelength of the fluorescence maximum at high scFv concentrations is higher than for the w.t., but shifts to a lower wavelength upon dilution, suggesting a more w.t.-like conformation at low concentrations.

Implications for antibody cloning and engineering

In most of the available primer mixtures used for the cloning of antibody fragments, position H6 is still covered by the primer, and primer-induced mutations at this position are seen frequently. This is underlined by the results of four other studies, in which degenerate primers exchanged Gln H6 by Glu.[6–9] The scFv fragments investigated in these

studies all belonged to the type III ($6_Q7_S10_A$), and the mutation of Gln H6 to Glu led to a residue combination not seen in natural antibodies ($6_E7_S10_A$), as in V_H domains of type I H10 (H9) is a proline residue, while for type II a glycine residue is required. The aL2-w.t. belongs to the newly defined type IV ($6_Q7_P10_A$).

The effects of a glutamine to glutamate exchange at residue H6 are not identical in all four examples. De Haard *et al.*[8] and Kipriyanov *et al.*[7] reported a loss of binding affinity, while in the anti-traseolide antibodies studied by Langedijk *et al.*[9] the hapten binding was not affected. In the aL2 case, the affinity could not be determined for the H6 glutamate mutants, although qualitative evaluation of the curves would suggest an affinity at least within the same order of magnitude as that of the w.t., consistent with the observation that the conformation of the antigen-binding pocket is not affected by the mutations.[10] In all cases, the *in vivo* folding yields were reduced for the Glu variant and reduced stabilities were implied. De Haard *et al.*[8] did not directly determine the stability of their mutants. However, they observed reduced protease resistence of their H6 glutamate mutants. The hypothesis presented by De Haard *et al.*,[8] that the first β-strand might be a folding nucleus, seems not very likely in light of the aL2-$6_E7_P10_G$ structure (Figure 4). Here, the first β-strand does not even participate in the framework structure, so that it is highly improbable that it has taken part in an intermediary tertiary structure. Moreover, the rest of the Fv fragment is folded correctly and is identical with the conventionally structured w.t., obviously without the need of the FR-H1 region. Therefore, this sequence region may be important for stabilizing the final structure, but not absolutely necessary for achieving or maintaining the fold, and therefore cannot be a folding nucleus.

It has been suggested that the burial of a charged glutamate residue would generally have a destabilizing effect, and the observed large shift of the side-chain pK_a of the glutamate residue in aL2-$6_E7_S10_G$ confirms this notion. However, if one compares the sequences of antibodies reputed to be particularly ''well behaved'' (good production yield, good refolding yields *in vitro*, high thermodynamic stability), a disproportionally large fraction of these seem to belong to type II. The humanized antibody 4D5,[18] the murine AB48,[20] the unpaired V_{HH} domains of camelid antibodies,[21,22] which have been suggested to be of high stability,[23] all belong to type II ($6_E7_S10_G$). In addition, the stability of the human V_H3 consensus construct encoded in the synthetic HuCal V_H3 framework[24] 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), indicating that in these antibodies the potentially deleterious effect of Glu H6 has been compensated in some way in the global sequence context.

Amongst the glutamate-containing mutants of aL2, the variant aL2-$6_E7_S10_G$, expressing the type II residue pattern, was the most stable, although still destabilised significantly compared to the w.t. aL2-$6_Q7_P10_A$. H6 glutamate definitely is more stringent in its requirement for the local sequence context: it is not compatible with a proline residue in position H7 and requires either a proline or glycine residue in position H10 (H9). A glutamine to glutamate substitution therefore almost invariably leads to a local sequence pattern not normally seen in natural antibodies, while the effects of a glutamate to glutamine substitution should be less drastic.

These structural findings stress how crucial a good sequence and primer design is for synthetic antibodies and antibody libraries. Degenerate primers should be designed taking this problem into account, and the amplification PCR must be carried out under stringent conditions.[25] For the cloning of individual antibody sequences and immune libraries, and in the humanization of antibodies by CDR grafting, the structural subclasses defined by residues H6, H7 and H10 (H9) and the interaction of these residues with the V_H core packing may play a crucial role. This was recently underlined by a graft of the CDRs of antibody MOC31 onto the 4D5 framework.[26] In this case, retention of the subtype-specific framework residues H6, H7 and H10, and of the domain core of the V_H domain of the CDR donor resulted in significantly enhanced thermodynamic stability compared to the classical graft.

Conclusions

Summarizing the rules for allowed combinations of amino acids at positions H6, H7 and H10 (H9) (Table 1), we can state that the amino acids in these positions define the distinct conformations of FR-H1. Glu in position H6 is not compatible with a proline-residue in position H7 and requires either a proline (type I) or a glycine residue (type II) in postion H10 (H9). Glutamine H6 is more permissive and compatible with any amino acid in positions H7 and H10, although Ser H7 and Ala H10 predominate in natural sequences. A proline residue in H7 leads to a type IV conformation of FR-H1, any other residue to a type III conformation. A concerted exchange of position H6, H7 and H10 (H9) from $6_Q7_P10_A$ to $6_E7_S10_G$ elicits a conformational change from type IV to type II, while the combination $6_E7_P10_G$ does not assume any of the four canonical conformations. Our study shows clearly that in both cases a destabilization occurs. This raises the question of whether Gln at position H6 is always more stable than Glu, and whether Glu is just tolerated in V_H domains that possess sufficient overall stability. Although all present data support this notion, this question waits for a conclusive answer from future investigations of scFv fragments.

The results obtained from the X-ray structure of the mutant aL2-$6_E7_S10_G$ prove the validity of the hypothesis that the combination of amino acids at positions H6, H7 and H10 (H9) determines the conformation of the FR-H1 region independent of the global V_H sequence context.[4] Out-of-context mutations, frequently induced by the primer mixtures commonly used for the cloning of individual antibodies and antibody libraries, can have severe effects on folding yields and stabilities of the antibodies thus obtained.

With the findings presented in this study, we have elucidated a new structural subclass division for Fv fragments with vital importance for such widely used applications as scFv/Fab library construction from natural sources, CDR-grafting or complete synthesis.

Materials and Methods

Plasmids and strains

The scFv fragments aL2 and its mutants were cloned into the plasmid pAK400"[11] in the orientation V_L-$(G_4S)_3$-V_H between the unique *Sfi*I restriction sites, resulting in a construct containing an N-teminal Flag-tag and a C-terminal His-tag. The mutant antibody fragments were constructed by site-directed mutagenesis using appropriate primers in a PCR amplification of the heavy chains with Vent polymerase (New England Biolabs). All PCR-amplified V_H fragments were sequenced after recloning into the expression vector with *Bam*HI and *Sfi*I. For expression studies of soluble scFvs, the *E. coli* strain JM83[27] was used. For large-scale production of scFv fragments, the *E. coli* strain SB536[28] was preferred.

Protein production

Large-scale production of scFv fragments was carried out in 2YT medium in a 50 l fermenter (Laboratory Pilot Fermenter model LB351, Bioengineering) at 26 °C. Fifty liters of 2YT medium was inoculated with a 1 l overnight culture containing 35 μg/ml chloramphenicol (Cam) and 20 mM glucose, grown at 25 °C. Expression was induced at an $A_{550\ nm}$ of 4 by addition of IPTG to a final concentration of 1 mM. During the expression, the pH was adjusted automatically. The cells were harvested at a final $A_{550\ nm}$ of 10. The cells were resuspended in HBS (20 mM Hepes (pH 7.4), 150 mM NaCl) containing DNaseI and disrupted in a Gaulin homogenizer. The crude extracts were centrifuged, filtered and purified by IMAC (Ni-NTA, Qiagen) and subsequent cation-exchange chromatography (S-Sepharose, Pharmacia). The protein concentrations were determined by measuring the absorption at 280 nm using the method of Gill & von Hippel.[29]

Physicochemical characterization

The K_D was determined by fluorescence quenching of the purified scFv with stepwise addition of ampicillin in the same buffer (HBS containing 0.005 % (v/v) Tween-20). Five emission spectra from 320 to 370 nm per scFv concentration were recorded with an excitation wavelength of 280 nm. The averaged emission intensity at 342 nm was corrected for dilution and evaluated as a function of ampicillin concentration by a two-parameter fit.[30]

The thermodynamic stability was determined by equilibrium denaturation with urea as described.[30] Samples (1.7 ml) containing 5 μg of scFv in HBS (20 mM Hepes (pH 7.4), 150 mM NaCl) in different concentrations of urea (4 M to 9 M in 0.15 M steps) were incubated overnight at 10 °C and equilibrated to 20 °C prior to the measurements. Five fluorescence emission spectra of each sample were recorded from 325 nm to 365 nm at 20 °C with an excitation of the protein fluorescence at 280 nm, averaged, and the emission maximum determined by a Gaussian fit. The shift of the emission maximum with increasing urea concentration was used to calculate the fraction of unfolded scFv. The resulting curve was fitted according to Pace.[31]

Crystallization

Crystallization trials were performed at 20 °C employing the hanging-drop vapor-diffusion technique and an exhaustive set of conditions screened, comprising a commercially available bit,[32] an antibody screen designed by Stura *et al*.[33] and a local sparse matrix.[34] Two different forms of aL2 w.t crystals were obtained from a protein solution (16 mg/ml) added to 1.75 M $(NH_4)_2SO_4$, 50 mM sodium phosphate (pH 6.5) containing 5 % MPD (form 1) and from 1.75 M $(NH_4)_2SO_4$, 50 mM sodium phosphate (pH 5.5) containing 5 % (v/v) isopropanol (form 2). Addition of ampicillin to the protein solution delayed crystallization by two months and yielded crystals that did not contain any antigen.[10] The mutant aL2-$6_E7_S10_G$ (2.85 mg/ml) was crystallized in 1.9-2.0 M $(NH_4)_2SO_4$, 100 mM sodium acetate (pH 5.8), 1.5 % (v/v) 1,2,3-heptanetriol. Fv aL2-$6_E7_P10_G$ (3 mg/ml) was crystallized from 26 % (w/v) polyethylene glycol 2000 monomethyl-ether, 200 mM ammonium sulfate, 100 mM sodium acetate (pH 5.0). The structure of the mutant aL2-$6_E7_P10_G$ complexed with the hapten was obtained from crystals grown in the absence of the hapten and incubated repeatedly with mother liquor containing 33 mM ampicillin.

Data collection and processing

Data sets to 1.8 Å resolution were collected on the ID2 beamline at ESRF (Grenoble, France) using a Mar-300 imaging plate detector system (λ 0.988 Å) on frozen crystals. Crystals were cryo-protected by addition of glycerol to the mother liquor, 25 % (v/v) glycerol for the mutant aL2-$6_E7_S10_G$ and 14 % (v/v) glycerol for the mutant scFv aL2-$6_E7_P10_G$. The data were indexed and intergrated using DENZO,[35] and scaled and merged using SCALEPACK and CCP4.[36] Data collection parameters of all the crystal forms are summarized in Table 3.

Structure determination and refinement

The scFv mutant structures were determined with the program AMoRe.[37] The models were preliminarily refined with simulated annealing and torsion angle refinement to 2.5 Å using the program CNS version 0.5.[38] The position of each residue was confirmed with simulated annealing omit map calculations and model fitting with the molecular graphics program TURBO-FRODO.[39] Individual *B*-factors were refined and one sulfate ion, two glycerol molecules and 106 water molecules were added to the model of the mutant aL2-$6_E7_S10_G$. For

the mutant aL2-$6_E7_S10_G$, the N terminus of the heavy chain H1-H10 were built into omit maps. Two sulfate anions, the ampicillin and 298 water molecules were added to the model.

Protein Data Bank accession numbers

The four aL2 scFv structures have been deposited in the RCSB Protein Data Bank, entries 1I3G (aL2 w.t. form 1), 1H8N (aL2-$6_E7_S10_G$), 1H8O (aL2-$6_E7_P10_G$ unliganded) and 1H8S (aL2-$6_E7_P10_G$-ampicillin complex).

Acknowledgements

We thank Drs Jörg Burmester and Anke Krebber for helpful discussions. This work was supported by a grant from the Schweizerische Nationalfonds 3100-046624, the European Community BIO2-CT92-0367 and the CNRS.

References

1. Wu, T. T. & Kabat, E. A. (1970). An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Expt. Med.* **132**, 211-250.
2. Kabat, E. A. & Wu, T. T. (1971). Attempts to locate complementarity determining regions in the variable positions of light and heavy chains. *Ann. N.Y. Acad. Sci.* **190**, 382-393.
3. 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.
4. Honegger, A. & Plückthun, A. (2001). The influence of the buried glutamine or glutamate residue in position 6 on the structure of immunoglobulin variable domains. *J. Mol. Biol.* **309**, 687-699.
5. 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.
6. 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.
7. 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 bacterial secretion but not the affinity. *Protein Eng.* **10**, 445-453.
8. 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.
9. 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.
10. Burmester, J., Spinelli, S., Pugliese, L., Krebber, A., Honegger, A., Jung, S., Schimmele, B., Cambillau, C. & Plückthun, A. (2001). Selection, characterization and X-ray structure of anti-ampicillin single-chain Fv fragments from phage-displayed murine antibody libraries. *J. Mol. Biol.* **309**, 671-685.
11. Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R. & Plückthun, A. (1997). Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods,* **201**, 35-55.
12. Wörn, A. & Plückthun, A. (1998). Mutual stabilization of VL and VH in single-chain antibody fragments, investigated with mutants engineered for stability. *Biochemistry,* **37**, 13120-13127.
13. Wörn, A. & Plückthun, A. (1999). Different equilibrium stability behavior of scFv fragments: identification, classification, and improvement by protein engineering. *Biochemistry,* **38**, 8739-8750.
14. Wörn, A. & Plückthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989-1010.
15. Ban, N., Escobar, C. K. W. H., Day, J., Greenwood, A. & McPherson, A. (1995). Structure of an anti-idiotypic Fab against feline peritonitis virus-neutralizing antibody and a comparison with the complexed Fab. *FASEB J.* **9**, 107-114.
16. Banfield, M. J., King, D. J., Mountain, A. & Brady, R. L. (1997). V_L:V_H domain rotations in engineered antibodies: crystal structures of the Fab fragments from two murine antitumor antibodies and their engineered human constructs. *Proteins: Struct. Funct. Genet.* **29**, 161-171.
17. Fan, Z. C., Shan, L., Guddat, L. W., He, X. M., Gray, W. R., Raison, R. L. & Edmundson, A. B. (1992). Three-dimensional structure of an Fv from a human IgM immunoglobulin. *J. Mol. Biol.* **228**, 188-207.
18. 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.
19. Perisic, O., Webb, P. A., Holliger, P., Winter, G. & William, R. L. (1994). Crystal structure of a diabody, a bivalent antibody fragment. *Structure,* **2**, 1217-1226.
20. 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.
21. 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.
22. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A. & Hamers, R. (1994). Sequence and structure of V_H domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng.* **7**, 1129-1135.
23. 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.
24. 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.
25. Burmester, J. & Plückthun, A. (2000). Construction of scFv fragments from hybridoma or spleen cells by PCR assembly. In *Antibody Engineering* (Kontermann, R. & Dübel, S., eds), pp. 19-40, Springer Verlag, Heidelberg.
26. Willuda, J., Honegger, A., Waibel, R., Schubiger, P. A., Stahel, R., Zangenmeister-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.
27. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene,* **33**, 103-119.
28. Bass, S., Gu, Q. & Christen, A. (1996). Multicopy suppressors of prc mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated R1pA. *J. Bacteriol.* **178**, 1154-1161.
29. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.
30. 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.
31. Pace, C. N. (1990). Measuring and increasing protein stability. *Trends Biotechnol.* **8**, 93-98.
32. Jancarik, J. & Kim, S. H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallog.* **24**, 409-410.
33. Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). Strategies in the crystallization of glycoproteins and protein complexes. *J. Crystal Growth,* **122**, 273-285.
34. McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*, pp. 82-159, John Wiley and Sons, New York.
35. Otwinowski, Z. (1993). DENZO: oscillation data and reducing program. In *Data Collection and Processing* (Sawyer, L., Isaacs, N. W. & Bailey, S., eds), pp. 56-63, Daresbury Laboratory, Warrington, England.
36. CCP4, Number 4 Collaborate Computational Project (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.
37. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.
39. Roussel, A. & Cambillau, C. (1991). The Turbo-FRODO graphics package, Silicon Graphics, Mountain View, CA.
40. Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 91-97.
41. 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.
42. 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.
43. 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.

Edited by I. Wilson

(Received 4 December 2000; received in revised form 28 March 2001; accepted 29 March 2001)