Protein Engineering vol.10 no.4 pp.435–444, 1997

Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved *in vivo* folding and physical characterization of an engineered scFv fragment

Lars Nieba, Annemarie Honegger, Claus Krebber and Andreas Pluckthun[1]

Biochemisches Institut der Universitat Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[1]To whom correspondence should be addressed

By constructing Fv and single-chain Fv (scFv) fragments of antibodies, the variable domains are taken out of their natural context in the Fab fragment, where they are associated with the constant domains of the light (C_L) and heavy chain (C_H1). As a consequence, all residues of the former variable/constant domain interface become solvent exposed. In an analysis of 30 non-redundant Fab structures it was found that at the former variable/constant domain interface of the Fv fragment the frequency of exposed hydrophobic residues is much higher than in the rest of the Fv fragment surface. We investigated the importance of these residues for different properties such as folding *in vivo* and *in vitro*, thermodynamic stability, solubility of the native protein and antigen affinity. The experimental model system was the scFv fragment of the anti-fluorescein antibody 4-4-20, of which only 2% is native when expressed in the periplasm of *Escherichia coli*. To improve its *in vivo* folding, a mutagenesis study of three newly exposed interfacial residues in various combinations was carried out. The replacement of one of the residues (V84D in V_H) led to a 25-fold increase of the functional periplasmic expression yield of the scFv fragment of the antibody 4-4-20. With the purified scFv fragment it was shown that the thermodynamic stability and the antigen binding constant are not influenced by these mutations, but the rate of the thermally induced aggregation reaction is decreased. Only a minor effect on the solubility of the native protein was observed, demonstrating that the mutations prevent aggregation during folding and not of the native protein. Since the construction of all scFv fragments leads to the exposure of these residues at the former variable/constant domain interface, this strategy should be generally applicable for improving the *in vivo* folding of scFv fragments and, by analogy, also the *in vivo* folding of other engineered protein domains.

Keywords: antibody engineering/domain interface/hydrophobic patches/protein aggregation/protein engineering/single-chain Fv fragment

Introduction

Small antibody fragments show exciting promise for use as therapeutics, diagnostic reagents and for biochemical research (Winter and Milstein, 1991; Pluckthun, 1994; Huston *et al.*, 1995). For most of these applications they have to be prepared in large amounts. The functional expression of antibody fragments in the periplasm of *Escherichia coli* (Pluckthun, 1992), especially Fv or single-chain Fv (scFv) fragments, is now used in many laboratories. Functional expression yields of these fragments vary widely, however, over several orders of magnitude even when the cell density is accounted for or when fragments in the identical host–vector system are compared (Carter *et al.*, 1992; Pluckthun *et al.*, 1996). Despite numerous studies (reviewed by Pluckthun *et al.*, 1996), the factors influencing antibody expression levels are still only poorly understood. Initial efforts had focused on transcription and translation efficiency, but with present vector systems these problems have been satisfactorily solved. The differences in expression yield between different antibody sequences show that protein sequence-related factors are the major remaining challenge. Folding efficiency and stability of the antibody fragments, as well as protease lability and toxicity of the expressed antibody fragments to the host cells, often severely limit actual production levels, but the rational understanding of these problems is only beginning.

Knappik and Pluckthun (1995) showed that the primary sequence of a particular antibody emerges as the most decisive factor determining the yield of functional protein. Based on a sequence comparison of different antibodies, back-engineering pin-pointed the effect to just a few point mutations. While these particular mutations are only applicable to a few antibodies and thus constitute a tailor-made solution, they do show that very minor sequence changes can have a dramatic effect on the *in vitro* aggregation properties of these molecules and the *in vivo* functional expression of antibody fragments in the periplasm of *E.coli*. Similarly, Ullrich *et al.* (1995) found that point mutations in the complementarity-determining regions (CDRs) can increase the yields in periplasmic antibody fragment expression.

The observations of Knappik and Pluckthun (1995) indicate that optimizing those parts of the antibody fragment which are not directly involved in antigen recognition can significantly improve folding properties and production yields of recombinant Fv and scFv constructs. The causes of the improved expression behavior lie in the decreased aggregation behavior of these molecules. The understanding of how specific sequence modifications change these properties is still very limited and currently under active investigation.

Normally, the majority of the side chain residues exposed at the surface of a protein are hydrophilic. Extended hydrophobic patches on the surface often indicate protein–protein interaction sites. If individual protein domains are taken out of their natural biological context, surfaces which are normally buried can become exposed. In the case of the antigen-binding Fv domain of an antibody, the Fv domain is taken away from the constant domains C_L and C_H1. While the resulting loss of V_H/V_L interaction energy can be overcome by engineering scFv or disulfide-linked Fv (dsFv) fragments (Huston *et al.*, 1988; Bird *et al.*, 1988; Glockshuber *et al.*, 1990; Brinkmann *et al.*, 1993; Huston *et al.*, 1995; Young *et al.*, 1995), the residues at the former variable/constant (v/c) domain interface, which are usually buried in this interface, remain exposed to

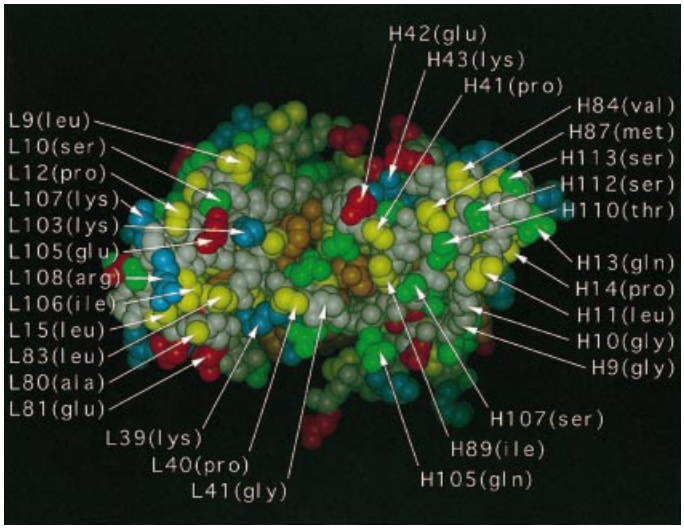

Fig. 1. Space-filling representation of the Fv fragment, seen from the v/c interface, of the antibody 4-4-20, color coded for residue types. Orange, aromatic side chains (Tyr, Phe, Trp); yellow, aliphatic side chains (Leu, Ile, Val, Pro, Ala); sulfur containing side-chains (Met, Cys); green, uncharged, hydrophilic side chains (Thr, Ser, Asn, Gln); red, acidic side chains (Glu, Asp); blue, basic side chains (His, Arg, Lys); white, main-chain (hydrophobicity color code would be yellow–green)

the solvent and form an extended hydrophobic patch (Figure 1). These exposed hydrophobic residues may influence the *in vivo* folding pathway by stabilizing misfolded structures and promoting the aggregation of folding intermediates.

In this study, we investigated a series of mutations introduced into an scFv fragment of the anti-fluorescein antibody 4-4-20 (Bedzyk *et al.*, 1990). In its original form, this fragment is very poorly produced in functional form. Nearly all of the expression product harvested from the periplasm of *E.coli* consists of insoluble, non-functional material (Bedzyk *et al.*, 1990; Denzin *et al.*, 1991). Based on the X-ray stucture of the Fab fragment of this antibody (Whitlow *et al.*, 1995), we identified the hydrophobic residues at the newly exposed v/c domain interface. Substitution of one particular newly exposed hydrophobic residue in this area increases the functional product 25-fold and we characterized the effect on the physical properties of this protein.

Material and methods

Calculation of solvent accessibility

Solvent-accessible surface areas for 30 non-redundant Fab fragments and the Fv fragments derived from these by deleting the constant domain coordinates from the PDB file were calculated using the latest version of the program NACCESS (http: // www.biochem.ucl.ac.uk / ~roman / naccess / naccess) based on the algorithm described by Lee and Richards (1971).

scFv gene synthesis

The scFv fragment in the orientation V_L–linker–V_H of the antibody 4-4-20 (Bedzyk *et al.*, 1990) was obtained by gene synthesis (Prodromou and Pearl, 1992). The V_L domain carries a three amino acid long FLAG tag (Knappik and Pluckthun, 1994) and the His_5-tag at the end of V_H (Ge *et al.*, 1995). We have used two different linkers of length 15 $(Gly_4Ser)_3$ and 30 amino acids $(Gly_4Ser)_6$, respectively. The gene so obtained was cloned into a derivative of the vector pIG6 (Ge *et al.*, 1995). The mutant antibody fragments were constructed by site-directed mutagenesis (Kunkel *et al.*, 1987) using single-stranded DNA and up to three oligonucleotides per reaction.

Expression for measuring in vivo solubility

Growth curves were obtained as follows: 20 ml of 2×YT medium containing 100 μg/ml ampicillin and 25 μg/ml streptomycin were inoculated with 250 μl of an overnight culture of *E.coli* JM83 harboring the plasmid encoding the respective antibody fragment and incubated at 24.5°C until an OD_{550} of 0.5 was reached. Isopropyl-β-D-thiogalactopyranoside (Biomol Feinchemikalien) was added to a final concentration of 1 mM and incubation was continued for 3 h. The OD_{550} was measured every hour, as was the β-lactamase activity in the culture supernantant to quantify the degree of cell leakiness. Three hours after induction an aliquot of the culture was removed and the cells were lysed by urea lysis followed by lysozyme treatment exactly as described by Knappik and Pluckthun (1995). The β-lactamase activity, a marker of the periplasmic fraction, was measured in the culture supernatant, in the insoluble and in the soluble cellular fraction to verify the quality of fractionation and the extent of periplasmic leakiness. Since the leakiness was found to be very low, the results from normalizing to OD_{550} (number of cells) or amount of β-lactamase (accounting for potential loss of periplasmic content to the medium) were identical. The fractions were assayed for antibody fragments by reducing SDS–PAGE, with the samples normalized to OD_{550}. The gels were blotted and immunostained using the FLAG antibody M1 (Prickett *et al.*, 1989) as the first antibody, an Fc-specific anti-mouse antiserum conjugated to horseradish peroxidase (Pierce) as second antibody, using a chemiluminescent detection assay described elsewhere (Ge *et al.*, 1995).

Purification

Mutant scFv fragments were purified by a two-column procedure. After French press lysis of the cells, the crude *E.coli* extract was first purified by immobilized metal ion affinity chromatography (IMAC) [Ni-nitrilotriacetic acid (NTA) Superflow, Qiagen] (20 mM HEPES, 500 mM NaCl, pH 6.9; step gradient of imidazole 10, 50 and 200 mM) (Lindner *et al.*, 1992) and, after dialyzing the IMAC eluate against 20 mM MES, pH 6.0, finally purified by cation-exchange chromatography (S-Sepharose fast flow column, Pharmacia) (20 mM MES, pH 6.0; salt gradient 0–500 mM NaCl). Purity was controlled by Coomassie-stained SDS–PAGE. The functionality of the scFv was tested by competition ELISA.

The wild-type (wt) 4-4-20 was expressed as cytoplasmic inclusion bodies in the T7-based system (Studier and Moffatt, 1986; Ge *et al.*, 1995). The refolding procedure was carried out as described elsewhere (Ge *et al.*, 1995). For purification, the refolding solution (2 l) was loaded over 10 h without prior dialysis on to a fluorescein affinity column, followed by a washing step with 20 mM HEPES, 150 mM NaCl, pH 7.5. Two column volumes of 1 mM fluorescein (sodium salt, Sigma), pH 7.5, were used to elute all functional scFv fragments. Extensive dialysis (7 days with 12 buffer changes) was necessary to remove all fluorescein. All purified scFv fragments were tested by gel filtration (Superose-12 column, Pharmacia SMART-System, 20 mM HEPES, 150 mM NaCl, pH 7.5).

K_D determination by fluorescence titration

Protein concentrations were determined photometrically using an extinction coefficient calculated according to Gill and von

Hippel (1989). Fluorescence titration experiments were carried out by taking advantage of the intensive fluorescence of fluorescein. A 2 ml volume of 20 mM HEPES, 150 mM NaCl, pH 7.5, containing 10 or 20 nM fluorescein was placed in a cuvette with an integrated stirrer. The excitation wavelength was 485 nm and emission spectra were recorded from 490 to 530 nm. Purified scFv (in 20 mM HEPES, 150 mM NaCl, pH 7.5) was added in 5–100 µl aliquots and after 3 min of equilibration a spectrum was recorded. All spectra were recorded at 20°C. The maximum of the emission at 510 nM was used for determining the degree of complexation of scFv to fluorescein, seen as quenching, as a function of the concentration of the antibody fragment. The K_D value was determined by Scatchard analysis.

Equilibrium denaturation measurement

Equilibrium denaturation curves were obtained by denaturation of 0.2 µM protein in HEPES-buffered saline (HBS) (20 mM HEPES, 150 mM NaCl, 1 mM EDTA, pH 7.5) and increasing amounts of urea (1.0–7.5 M; 20 mM HEPES, 150 mM NaCl, pH 7.4; 0.25 M steps) in a total volume of 1.7 ml. After incubating the samples for 12 h at 10°C and an additional 3 h at 20°C prior to measurements, the fluorescence spectra were recorded at 20°C from 320 to 360 nm with an excitation wavelength of 280 nm. The emission wavelength of the fluorescence peak shifted from 341 to 347 nm during denaturation and was used for determining the fraction of unfolded molecules. Curves were fitted according to Pace (1990).

Thermal denaturation

For measuring the thermal denaturation rates, purified scFv was dissolved in 2 ml of HBS buffer to a final concentration of 0.5 µM. The aggregation was followed for 2.5 h at 40°C and at 44°C by light scattering at 400 nm.

Solubility measurements

For the solubility determination in HBS buffer (Stevenson and Hageman, 1995), the purified native proteins were concentrated to 1 mg/ml. A stock solution of 40% (w/v) PEG 6000 (Fluka) was prepared in this buffer. A 20 µl volume of protein was incubated for 1 h at room temperature in PEG of various concentrations (5, 8, 10, 15 and 20%) and HBS buffer was added to a final volume of 40 µl. The precipitated scFv fragment was removed by centrifugation at 14 000 r.p.m. (14 500 *g*) in an Eppendorf tube. Coomassie-stained SDS gels of the supernantant were scanned densitometically to estimate the amount of soluble protein as a function of PEG concentration and extrapolated to zero PEG concentration.

Results

Comparison of known antibody sequences

Compared with other domain/domain interfaces in proteins, the interface between immunoglobulin variable and constant domains (v/c interface) is not very tightly packed. A comparison of 30 non-redundant Fab structures in the PDB database showed that between the light chain variable and constant domain an area of 410 ± 90 Å^2 per domain is buried, while the heavy chain variable and constant domains interact over an area of 710 ± 180 Å^2. Some, but not all, of the v/c interface residues are hydrophobic, predominantly aliphatic. Generally, sequence conservation of the residues contributing to the v/c domain interface is not particularly high. Still, the v/c domain interface shows up as a marked hydrophobic patch on the surface of an Fv fragment (Figure 1).

Solvent-accessible surface areas for 30 non-redundant Fab fragments and their corresponding Fv fragments (derived from the Fab fragment by deleting the constant domain coordinates from the PDB file) were calculated using the program NACCESS (Lee and Richards, 1971). Residues participating in the v/c domain interface were identified by comparing the solvent-accessible surface area of each amino acid side chain in the context of an Fv fragment with its accessible surface in the context of a Fab fragment. Figure 2 shows a plot of the relative change in side-chain accessibility upon deletion of the constant domains as a function of sequence position. Residues which show a significant reduction in side-chain accessibility are also highlighted in the sequence alignment. To assess sequence variability in the positions identified in Figure 2, the variable domain sequences collected in the Kabat database (status March 1996) were analyzed (Table I). Of the 15 interface residues identified in the V_L domain of the antibody 4-4-20 (Figure 1 and Table I), L9(Leu), L12(Pro), L15(Leu), L40(Pro), L83(Leu) and L106(Ile) are hydrophobic and therefore candidates for replacement. Of the 16 interface residues in the V_H domain, H11(Leu), H14(Pro), H41(Pro), H84(Val), H87(Met) and H89(Ile) were identified as possible candidates for substitution by hydrophilic residues in the scFv fragment of the antibody 4-4-20 (Figure 1 and Table I).

Not all of these hydrophobic residues are equally good candidates for replacements, however. While residues which are hydrophobic in one particular sequence but hydrophilic in many other sequences may appear most attractive, the conserved hydrophobic residues listed in Table I have also been investigated, since the evolutionary pressure which kept these conserved residues acted on the Fab fragment within the whole antibody, but not the isolated Fv portion. In this study, we did not consider the proline residues since Pro L40 and Pro H41 form the hairpin turns at the bottom of the framework II region, while the conserved V_L *cis*-proline L8 and proline residues H9 and H14 determine the shape of framework I of the immunoglobulin variable domains.

Excluding prolines, this leaves residues L9 (Leu in 4-4-20, Ser in most κ chains), L15 (Leu, usually hydrophobic), L83 (Leu, usually Val or Phe) and L106 (Ile, as in 86% of all κ chains) in the V_L domain and H11 (Leu, as in 60% of all heavy chains), H84 (Val, in other V_H domains frequently Ala or Ser), H87 (met, usually Ser) and H89 (Ile, most frequently Val) in V_H as possible candidates for replacement in the 4-4-20 scFv fragment.

Point mutations in the 4-4-20 scFv

For the particular case of the 4-4-20 scFv fragment, some of the v/c interface residues are already hydrophilic, but nine residues are of hydrophobic nature (including Pro12 in the light chain) (Table I). However, since all permutations of these nine residues with other residues will generate a prohibitive number of experiments, we chose three residues for closer analysis by mutations to test the general concept and biophysical consequences and to define which hydrophobic patch is most important for the aggregation behavior of the wt scFv.

Leu15 in V_L is a hydrophobic amino acid in 98% of all κ chains (Table I). Leu11 is conserved in V_H (Table I) and is involved in v/c interdomain contacts (Lesk and Chothia, 1988). In contrast, valine occurs very infrequently at position H84; mainly found at this position are threonine, serine and alanine (Table I). As can be seen in Figure 1, Val84 contributes to a large hydrophobic patch at the newly exposed surface of V_H.

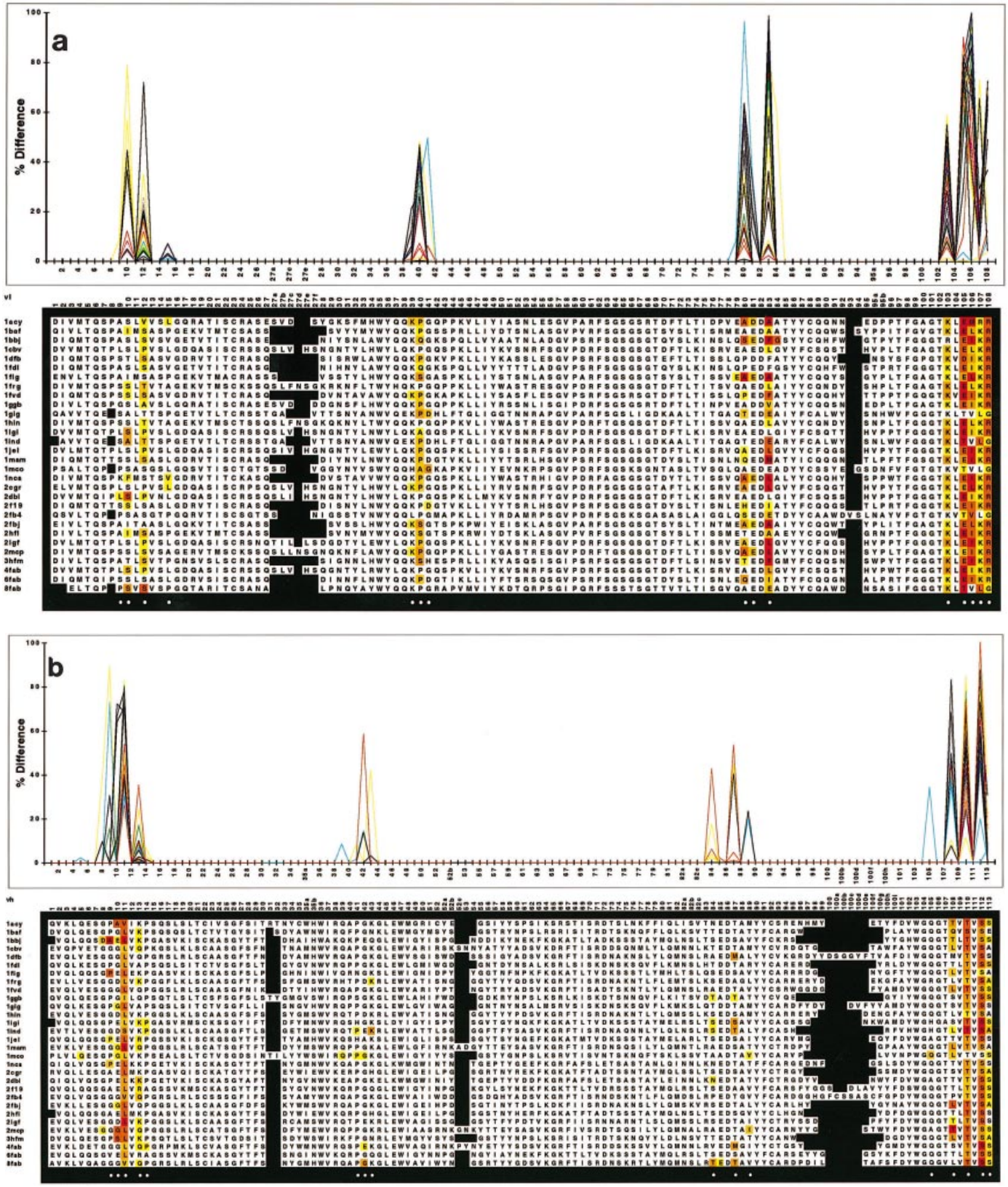

Fig. 2. Variable/constant domain interface residues for (**a**) V_L and (**b**) V_H. For 30 non-redundant Fab fragments taken from the Brookhaven Databank, the solvent-accessible surface of the amino acid side chains was calculated in the context of an Fv and of an Fab fragment. The plot shows the relative reduction in accessible surface upon contact with the constant domains (color code: 30 different colors for the 30 Fv fragments). In the sequence alignment, residues contributing to the v/c interface are highlighted. The shading from white (<1%) to red (⩾80%) reflects the relative reduction of solvent accessible surface upon removing the constant domains (color code: white, <1%; yellow, <20%; yellow–orange, <40%; orange, <60%; red–orange, <80%; red, ⩾80%). Circles indicate those positions which are further analyzed in Table I.

Table I. Sequence variability of residues contributing to the v/c interface

Pos.		L9				L10		L12				L15				L39				L40				L41				L80			
% exp. (FAB)		91				59		41				48				37				64				114				57			
% exp. (ind.)		91				65		47				49				37				82				121				74			
% buried v/c		0				10		12				0				1				22				3				21			
		kappa		lambda		kappa		kappa		lambda		kappa		lambda		kappa		lambda		kappa		lambda		kappa		lambda		kappa		lambda	
Species		hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu
Seq. 4-4-20		Leu				Ser		Pro				Leu				Lys				Pro				Gly				Ala			
Cons.		Ser	Ala	Ser	Ala	Thr	Ser	Ser	Ser	Ser	Thr	Pro	Leu	Pro	Pro	Lys	Lys	Lys	Lys	Pro	Pro	Pro	Pro	Gly	Gly	Gly	Asp	Pro	Ala	Ala	Thr
Dist.	Asp	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	3	96	0	0	0	0
	Glu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	4	0	0	0	8	0	0
	Lys	0	6	0	0	0	0	0	0	0	0	0	0	0	0	83	92	30	96	0	0	0	0	0	0	0	0	0	0	0	0
	Arg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	7	11	0	1	0	0	0	0	1	2	0	0	0	0	0
	His	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	2	3	0	0	0	0	0
	Thr	0	4	0	1	49	6	0	3	0	93	1	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	3	17	94
	Ser	43	24	90	4	48	60	84	50	99	6	0	1	0	0	1	0	0	0	2	19	5	0	0	1	0	0	9	6	19	0
	Asn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	Gln	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	3	0	0	0	0	0	0	0	6	0	0
	Gly	25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	99	80	90	3	0	0	0	0
	Ala	11	41	7	90	0	0	4	23	0	0	0	8	1	0	0	0	0	0	1	1	1	1	0	0	2	0	16	64	56	4
	Cys	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Pro	0	2	0	0	0	0	11	20	0	0	45	28	84	96	0	0	0	0	94	77	92	96	0	0	0	0	74	9	2	1
	Val	1	0	2	1	0	0	0	0	0	0	41	10	0	0	0	0	3	0	1	0	0	0	0	0	1	0	0	2	5	0
	Ile	0	0	0	3	0	24	0	0	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Leu	11	19	0	0	0	2	0	0	0	0	11	49	14	4	0	0	24	0	0	0	0	0	0	0	0	1	0	0	0	0
	Met	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
	Phe	0	1	0	0	2	7	0	0	0	0	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	Tyr	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	0	2	3	0	0	0	0	0	0	0	0	0	2	0	0
	Trp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0

Pos.		L81				L83				L103				L105				L106				L106A		L107				L108			
% exp. (FAB)		70				17				42				19				14				56		56				47			
% exp. (ind.)		74				38				58				57				28				65		65				103			
% buried v/c		5				46				26				66				56				15		15				55			
		kappa		lambda		kappa		lambda		kappa		lambda		kappa		lambda		kappa		lambda		lambda		kappa		lambda		kappa		lambda	
Species		hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu
Seq. 4-4-20		Glu				Leu				Lys				Glu				Ile						Lys				Arg			
Cons.		Glu	Glu	Glu	Glu	Phe	Leu	Glu	Glu	Lys	Lys	Lys	Lys	Glu	Glu	Thr	Thr	Ile	Ile	Val	Val	Leu	Leu	Lys	Lys	Gly	Gly	Arg	Arg	Gln	Gln
Dist.	Asp	10	7	4	0	0	0	0	15	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Glu	86	93	69	82	0	0	99	68	0	0	2	0	84	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Lys	0	0	0	0	0	0	0	0	88	98	83	100	0	0	0	0	0	0	0	0	0	0	94	97	0	0	0	0	0	0
	Arg	0	0	0	0	0	0	0	0	8	1	6	0	0	0	0	0	1	0	0	0	0	0	2	1	8	3	97	100	0	0
	His	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
	Thr	0	0	0	0	0	3	0	0	2	0	2	0	0	0	97	100	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	Ser	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	5	5	0	0	0	0
	Asn	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0	2	1	0	0	0	0	0	0
	Gln	0	0	1	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	54
	Gly	1	0	17	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83	92	1	0	2	0
	Ala	1	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	Cys	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	Pro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6
	Val	0	0	1	0	20	12	0	0	0	0	0	0	1	0	0	0	3	1	100	100	2	2	0	0	0	0	0	0	0	0
	Ile	0	0	0	0	6	11	0	0	0	0	0	0	0	0	1	0	88	73	0	0	0	0	0	0	0	0	0	0	0	0
	Leu	0	0	0	0	1	32	0	0	0	0	0	0	1	0	0	0	1	25	0	0	97	97	0	0	3	0	1	0	0	0
	Met	0	0	6	0	0	2	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	Phe	0	0	0	0	72	7	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	Tyr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Trp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Pos.		H9		H10		H11		H13		H14		H41		H42		H43		H84		H87		H89		H105		H108		H110		H112		H113	
% exp. (FAB)		29		65		31		69		36		72		111		65		76		28		34		72		52		26		25		72	
% exp. (ind.)		33		72		71		73		36		72		115		78		77		31		37		82		64		55		68		78	
% buried v/c		7		8		58		5		0		0		3		2		2		8		1		1		18		52		62		7	
		hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu	hu	mu
Seq. 4-4-20		Gly		Gly		Leu		Gln		Pro		Pro		Glu		Lys		Val		Met		Ile		Gln		Ser		Thr		Ser		Ser	
Cons.		Gly	Pro	Gly	Glu	Leu	Leu	Lys	Lys	Pro	Pro	Ala	Arg	Pro	Pro	Gly	Gly	Ala	Ser	Thr	Ser	Val	Val	Gln	Gln	Leu	Thr	Thr	Thr	Ser	Ser	Ser	Ser
Dist.	Asp	0	0	5	2	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Glu	2	0	27	54	0	0	2	0	0	0	0	0	0	0	1	19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	Lys	0	0	0	0	0	0	59	54	0	0	0	3	0	0	0	0	0	0	0	0	0	0	10	0	1	0	0	0	0	0	0	0
	Arg	0	0	1	0	0	0	3	18	0	0	1	36	1	0	0	1	0	0	0	0	0	1	5	0	3	0	0	0	0	0	0	0
	His	0	0	0	0	0	0	1	0	0	0	1	0	0	11	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
	Thr	1	2	1	0	0	0	0	0	0	1	1	13	1	1	0	0	4	12	96	46	4	15	0	4	23	51	89	99	0	0	0	0
	Ser	1	1	0	2	0	0	0	0	2	3	10	18	2	3	8	0	18	70	2	51	0	0	0	0	0	23	2	0	98	100	97	77
	Asn	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Gln	0	0	0	0	0	1	34	22	0	0	0	0	1	0	0	0	0	0	0	0	0	0	80	83	0	0	0	0	0	0	0	0
	Gly	42	29	61	38	0	0	0	0	0	0	0	0	0	0	90	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Ala	33	31	3	1	0	0	0	5	1	8	59	7	1	0	1	2	55	12	0	1	0	0	0	11	0	0	0	0	0	0	0	22
	Cys	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Pro	21	36	0	0	0	0	0	0	96	85	18	14	94	83	1	0	14	1	0	0	0	0	2	0	1	0	0	0	1	0	2	0
	Val	0	1	0	1	38	1	0	0	0	0	1	1	0	0	0	0	4	0	0	0	76	50	0	0	0	0	1	0	1	0	0	0
	Ile	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	1	1	0	0	5	13	0	0	0	0	7	0	0	0	0	0
	Leu	0	0	0	0	60	95	0	0	0	3	1	0	0	0	0	0	0	0	0	0	6	7	0	0	63	25	0	0	0	0	0	0
	Met	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	1	7	13	0	0	9	0	0	0	0	0	0	0
	Phe	0	0	0	0	2	0	0	0	0	0	0	6	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Tyr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Trp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Residue statistics are based on the variable domain sequences in the Kabat database (March 1996). Sequences which were <90% complete were excluded from the analysis. Number of sequences analyzed: human VL κ, 404 of 881; murine VL κ, 1061 of 2239; human VL λ, 223 of 409; murine VL λ, 71 of 206; human VH, 663 of 1756; murine VH, 1294 of 3849.

Pos., sequence position according to Kabat *et al.* (1991); % exp (FAB), relative side-chain accessibility (average of 30 non-redundant PDB files) in a Fab fragment as calculated by the program NACCESS [v2.0 by S.Hubbard (http://www.biochem.ucl.ac.uk/~roman/naccess/naccess.html)]; % exp (ind.), relative side-chain accessibility in the isolated VL or VH domain; % buried, relative difference in side-chain accessibility between Fv and Fab fragments; Cons., sequence consensus; Dist., distribution of residue types.

Table II. Mutations introduced in the scFv fragment of the antibody 4-4-20

	L15E (V_L)	L11N (V_H)	L11D (V_H)	V84D (V_H)
Flu 1	●			
Flu 2		●		
Flu 3			●	
Flu 4				●
Flu 5		●		●
Flu 6			●	●
Flu 7	●	●		
Flu 8	●		●	
Flu 9	●	●		●
Flu 4 short				●

Each line represents a different protein carrying the mutations indicated. The residues are numbered according to Kabat *et al.* (1991).

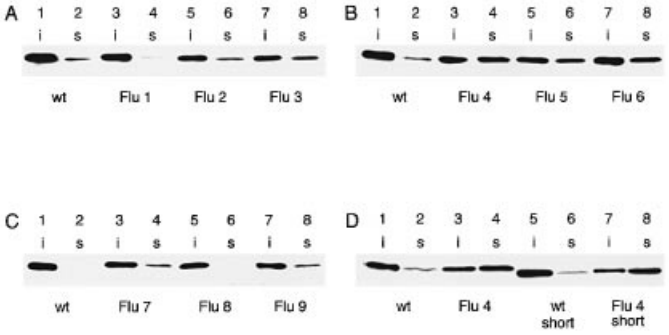

Fig. 3. Western blots showing the insoluble (i) and soluble (s) fractions of cell extracts, prepared as described in Materials and methods, expressing the mutated scFv fragments of the antibody 4-4-20. The amino acids substituted in the various mutants are given in Table II.

All three positions were mutated into acidic residues and L11 was also changed to asparagine (Table II).

The scFv fragment was tested and expressed with two different linkers, the frequently used 15-mer linker $(Gly_4Ser)_3$ (Huston *et al.*, 1995) and the same motif extended to 30 amino acids $(Gly_4Ser)_6$. All mutations were tested in both constructs. The *in vivo* expression results of the different mutations were identical and therefore only the results of the 30-mer linker are described in more detail. The periplasmic expression experiments were carried out at 24.5°C and the ratio of insoluble to soluble (i/s) protein was determined by immunoblotting for every mutant. In Figure 3A–D, lanes 1 and 2, the wt scFv is shown. Almost no soluble protein occurs in periplasmic expression, which is consistent with previous reports by Bedzyk *et al.* (1990) and Denzin *et al.* (1991), who described that the periplasmic expression of wt scFv leads mainly to periplasmic inclusion bodies.

The single-point mutation L15E in V_L (Flu1) shows no effect on the ratio i/s when compared with wt (Figure 3A, lanes 3 and 4). Mutating Leu at position 11 in the heavy chain to asparagine (Flu2) also shows nearly no effect compared with wt, whereas the substitution with aspartic acid (Flu3) changes the i/s ratio to more soluble protein, but still this effect is not very dramatic. In contrast, the point mutation at position 84 (Flu4, Figure 3B, lanes 3 and 4 and Figure 3D, lanes 3 and 4) had a very strong influence on the *in vivo* expression of the scFv fragment of the antibody 4-4-20. The i/s ratio is changed to about 1:1, resulting in a 25-fold increase in soluble expressed protein compared with wt.

The combination of V84D with L11N or L11D (Flu5, Flu6) also changes the i/s ratio compared with wt, but this ratio compared with V84D alone is not improved further (Figure 3B). Interestingly, the combination of Flu5 with the light chain mutation at position 15 (Flu9) leads to less protein in the soluble fraction (Figure 3C, lanes 7 and 8) than Flu5 itself (Figure 3B, lanes and 6). The negative influence of the L15E mutations can also be seen in Flu8 (Figure 3C, lanes 5 and 6) compared with Flu3 (Figure 3A, lanes 7 and 8). In Figure 3D the comparison of the wt (lanes 1, 2 and 5, 6) and Flu4 (lanes 3, 4 and 7, 8) is shown in both the 15-mer and 30-mer constructs. The single-point mutation V84D turned out to be the protein with the best i/s ratio in both constructs, with the 15-mer and the 30-mer linker scFv.

Functional expression and purification

The oligomerization of scFv fragments as a function of linker length has been investigated previously (Holliger *et al.*, 1993; Whitlow *et al.*, 1993, 1994; Desplancq *et al.*, 1994). A continuous decrease in the amount of dimer and multimer formation as a function of linker length has been reported (Desplancq *et al.*, 1994; Whitlow *et al.*, 1994). While the standard $(Gly_4Ser)_3$ linker has been shown to lead to monomeric scFvs in many cases in the V_H–V_L direction, this is often not the case in the V_L–V_H direction. This is caused by an asymmetry in the V_L/V_H arrangement, leading to a longer distance between the end of V_H and the N-terminus of V_L than between the C-terminus of V_L and N-terminus of V_H (Huston *et al.*, 1995). Consequently, a linker of identical length may lead to different properties of the resulting molecules.

Since we have chosen to use the minimal pertubation FLAG (Knappik and Pluckthun, 1994) at the N-terminus of V_L in our constructs and thus the V_L–linker–V_H orientation, we have investigated the use of longer linkers. In the periplasmic expression in *E.coli* no difference between the 15-mer and the 30-mer linker in the corresponding mutants is visible (Figure 3D), but when we attempted to purify the two Flu4 scFvs with long and short linker, a discrepancy between the two constructs was found. The purification of the Flu4 mutant (V84D) with the 15-mer linker leads to very small amounts of partially purified protein (~0.015 mg per liter and OD_{550}; estimated from SDS–PAGE after IMAC purification), whereas the 30-mer linker construct gives ~0.3 mg per liter and OD_{550} (~1 mg of highly pure functional protein after purification from 1 l normal shake-flask culture). All mutants with 30-mer linker were tested in gel filtration and found to be monomeric (data not shown).

For further *in vitro* characterization three mutants were purified with the 30-mer linker, V84D (Flu4), V84D/L11D (Flu6) and L11D (Flu3). A two-step chromatography, first using IMAC and then cation-exchange chromatography, led to homogeneous protein. The i/s ratio of the antibody fragments (Figure 3) is also reflected in the purification yield of functional protein under these conditions. It should be pointed out that these numbers are only relevant for a comparison of mutants; advances in fermentation technology allow an increase of several hundred-fold in volume yields (Horn *et al.*, 1996). To separate the question of yields per volume from yields per cell, we thus always normalize yields to liter and 1 OD_{550}. The mutant Flu4 (V84D) (Figure 3B, lanes 3 and 4) yields ~0.3 mg of purified and functional protein per liter and OD_{550} of cells, Flu6 (L11D/V84D) (Figure 3B, lanes 7 and 8) yields ~0.25 mg per liter and OD_{550} and Flu3 (less protein in the soluble fraction on the blot in Figure 3A, lanes 7 and 8) yields

Table III. K_D values of the different scFv mutants determined in fluorescence titration

	Flu wt	Flu 3	Flu 4	Flu 6	Flu wt[b]
K_D (nM)[a]	26 ± 2.3	20 ± 4	23 ± 3.3	25 ± 4.2	90

[a]The errors were calculated from the Scatchard analysis.
[b]Whole antibody, determined by Miklasz *et al.* (1995) by intrinsic protein fluorescence.

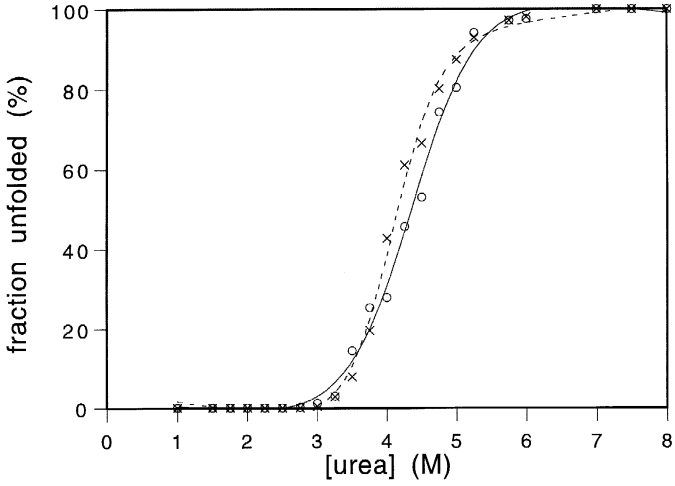

Fig. 4. An overlay plot of the urea denaturation curves is shown. (×) wt scFv; (○) Flu4.

0.05 mg per liter and OD_{550}. The wt scFv of the antibody 4-4-20 does not give any functional protein at all in periplasmic expression with either linker and it was therefore expressed as cytoplasmic inclusion bodies, followed by refolding *in vitro* and fluorescein affinity chromatography. The refolded wt scFv was shown by gel filtration to be monomeric with the 30-mer linker (data not shown).

Biophysical properties of the mutant scFvs

Since we changed amino acids which are conserved, it could not be excluded *a priori* that changes at these positions may be transmitted through the structure and have an effect on the binding constant, even though they are very far from the binding site (Chatellier *et al.*, 1996). To eliminate this possibility, we determined the binding constant of the mutants Flu3, Flu4 and Flu6 and the wt scFv. Fluorescence titration was used to determine K_D in solution by using the quenching of the intrinsic fluorescence of fluorescein when it binds to the antibody. The fluorescence quenching at 510 nm was measured as a function of added scFv. The K_D values (Table III) obtained for all three mutant scFvs and the wt scFv are very similar and correspond with the recently corrected K_D of the monoclonal antibody 4-4-20 (Miklasz *et al.*, 1995), determined by intrinsic protein fluorescence.

To determine whether the mutations had an influence on the thermodynamic stability of the protein, we determined the equilibrium unfolding curves by urea denaturation. The V84D mutant and the wt scFv were used for this analysis and in Figure 4 an overlay plot is shown. The midpoint of both curves is at 4.1 M urea. Both curves were fitted by an algorithm for a two-state model described by Pace (1990), but the apparent small difference between the V84D mutant and the wt scFv

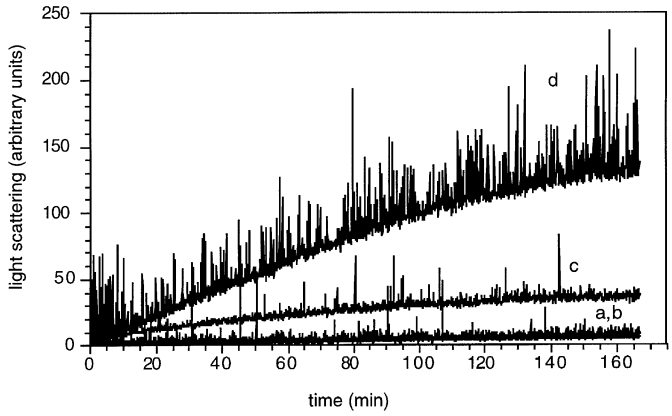

Fig. 5. Thermal denaturation time courses at 40 and 44°C for wt and Flu4 scFv fragments. (**a**) wt scFv at 40°C; (**b**) Flu4 at 40°C; (**c**) Flu4 at 44°C; (**d**) wt scFv at 44°C.

is not of statistical significance. Furthermore, it is unclear whether a two-state model describes this two-domain protein.

Aggregation of folding intermediates could be an explanation for the different *in vivo* results between the mutant scFvs and the wt scFv (Figure 3). In the periplasm of *E.coli*, the protein concentrations are assumed to be rather high (van Wielink and Duine, 1990) and the aggregation effects could thus be pronounced. In order to estimate the aggregation behavior *in vitro*, we measured the thermal aggregation rates at different temperatures. In Figure 5 it is clearly seen that the wt scFv significantly aggregates already at 44°C, whereas the mutant V84D tends to aggregate more slowly. The wt scFv is thus clearly more aggregation prone than the mutant scFv even though the urea denaturation curves are almost identical. This is very similar to the observations made with different mutations on the antibody McPC603 (Knappik and Pluckthun, 1995), where no correlation was found between equilibrium denaturation curves and expression behavior, but a good correlation was found with the thermal aggregation rates.

To determine whether the mutations act on increasing the solubility of the native protein or folding intermediates, we used the solubility in polyethylene glycol (PEG) solutions to determine the solubility of native proteins (Stevenson and Hageman, 1995). PEG can be used to increase the protein concentration above the solubility limit by its excluded volume effect (Middaugh *et al.*, 1979) and is frequently used to induce crystallization. When the concentration of saturated wt and mutant scFv solutions was determined in the presence of various PEG concentrations, after centrifuging precipitated scFv, very similar protein concentrations in the supernatant were obtained for wt and mutant protein (Figure 6). Upon quantification by gel scanning, the solubilities of the two proteins were found to be nearly identical within experimental error and they certainly do not differ by the factor of 25 found in expression yields. Furthermore, both the wt and the Flu4 mutant scFv could be concentrated at least to 10 mg/ml (data not shown). We conclude, therefore, that the mutation must predominantly act on the solubility of folding intermediates and not on that of the native protein.

Discussion

We have investigated whether there are factors in common to all scFv fragments of antibodies which might make their *in vivo* folding to functional, monomeric proteins inefficient,

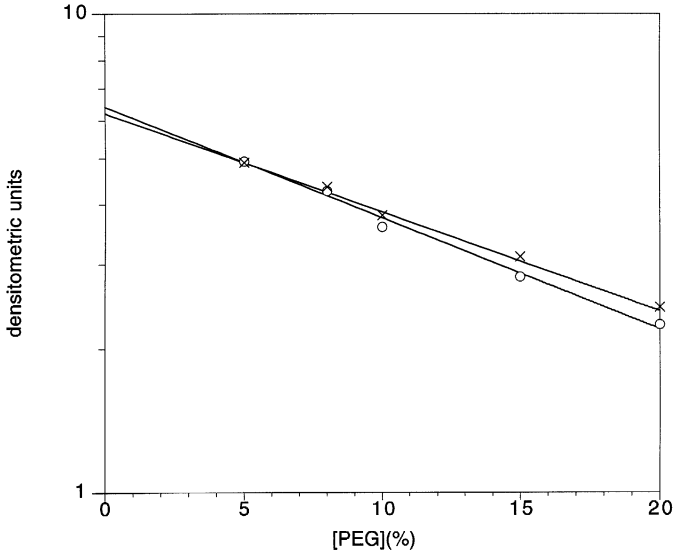

Fig. 6. Solubility of native scFv. (×) wt; (○) Flu4. PEG was used to increase the protein concentration above the solubility limit, precipitated scFv was centrifuged and scFv in the supernantant was analyzed by SDS–PAGE and densitrometric scanning.

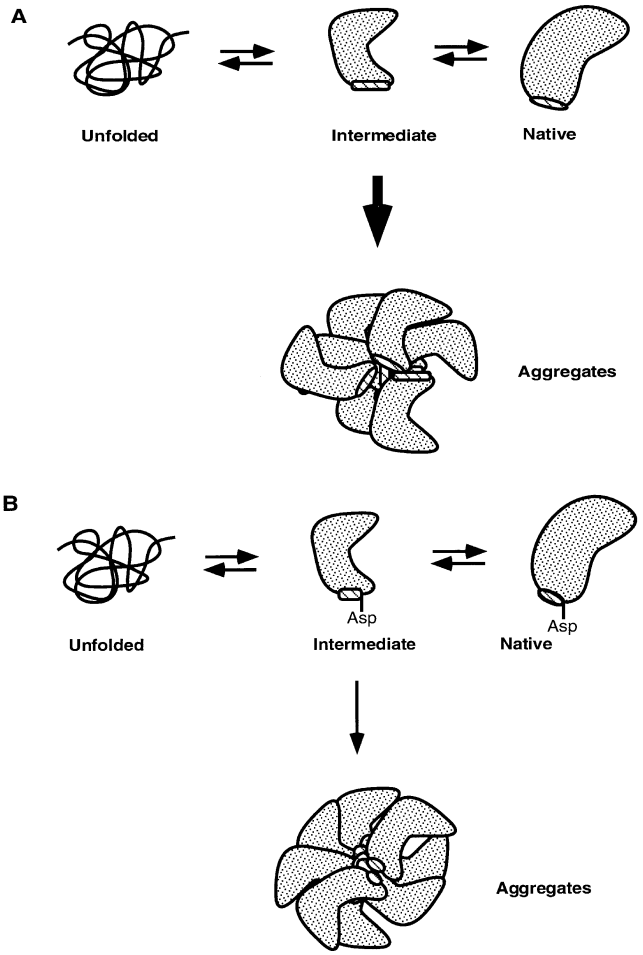

Fig. 7. Schematic folding pathway of the scFvs. (**A**) wt; (**B**) Flu4 mutant. The hydrophobic patch is indicating by hatching. There is no evidence of any changes along the productive pathway in rates or equilibrium (horizontally, from left to right). The mutation only seems to inhibit the aggregation reaction. During periplasmic folding, the reaction starts at the top left and proceeds to the top right, with protein being lost to aggregation, according to the thickness of the arrow. Upon thermal denaturation, the reaction starts at the top right and is faster because of a more favored aggregation step in the wt.

with the result of aggregation, but which can be corrected by protein engineering. An analysis of 30 known non-redundant crystal structures showed that there are hydrophobic patches at both former v/c domain interfaces, both in the light chain and in the heavy chain. Since this hydrophobic area has lost its function as a domain interface in an Fv or scFv fragment, the replacement of these hydrophobic surface residues by suitable hydrophilic residues should have no negative influence on folding and stability. Indeed, the disruption of the hydrophobic surface through the substitution of a few key hydrophobic residues by hydrophilic ones can significantly improve the expression and *in vivo* folding behavior of the recombinant Fv and scFv fragments. While the existence of the hydrophobic patches is preserved in all antibodies, their exact position and extent varies. It is thus plausible that the exact nature of these hydrophobic patches may be related to the aggregation properties of these molecules.

We can distinguish conserved residues which are almost always hydrophobic (such as residue 15 in V_L and 11 in V_H) and those which can be hydrophilic in a significant number of molecules in the database (such as residue 84 in V_H and 87 in V_H). From the present knowledge of aggregation we could not predict *a priori* which of these positions would have a significant influence on the aggregation reaction and should be substituted. Furthermore, the elbow angle (defining the angle between an axis going through V_H and one going through C_H, or V_L and C_L, respectively) varies widely between different antibody Fab fragments, leading to somewhat different v/c contact surfaces in different antibodies (Wilson and Stanfield, 1994). Therefore, we tested both conserved and less common residues and determined the biophysical properties of the purified scFv to understand the basis of these substitutions. Improvements were seen in both cases, even though the replacement of Val H84 to Asp showed a much stronger effect.

We believe, therefore, that in the optimization of a given antibody fragment, a first inspection of the exposed residues should be carried out according to Table I. Notably those residues which are adjacent to another exposed hydrophobic residue in the structure and allow a common contiguous interface (such as Val H84 with Met H87 and Ile H89 and to some degree Pro H41; see Figure 1) are candidates for replacement. In contrast, single hydrophobic amino acids such as Leu H11 have a weaker influence, as can be seen in the mutants Flu2 and Flu3.

In the folding pathway of the wt scFv, only a small amount goes to the native form and more than 98% is diverted off the pathway to aggregates (Figure 7A). In the Flu4 scFv the off-pathway reaction is no longer the dominant pathway, when Asp at H84 is introduced (Figure 7B). Since the effect of the mutation is predominantly on folding intermediates and not on the the native protein (Figure 6), the region around H84 may be part of a much bigger hydrophobic surface, possibly involving V/L interface residues or possibly accessible internal ones. The weaker influence of the Leu H11 substitution can be explained by Leu H11 being a single hydrophobic residue on the surface, which may not become part of a contiguous hydrophobic surface in the intermediate and thus does not influence the off-pathway to the same degree. Therefore, also no benefit is visible when both mutations are combined. The difference between Flu2 and Flu3, which is detectable but

small (Figure 3B), may be due to differences in polarity between acidic residues and non-charged residues, as pointed out by Dale *et al.* (1994).

The question then arises to which amino acid types the hydrophobic amino acids should be changed. In general, it may be noted that the isoelectric point of most *E.coli* proteins is acidic (VanBogelen *et al.*, 1992). Thus, the introduction of Asp residues appears particularly attractive, since they also keep the charge close to the main chain. However, it cannot be ruled out that adjacent negative charges are severly destabilizing adjacent loops, as can be seen in the L15E mutation in the light chain (Figures 1 and 3C), where no solubility enhancing effect of the mutation is observed. In this particular case L15 is close to residue E82 in the light chain. Thus, new Asp residues have to be limited to those positions where no such repulsion may occur.

An investigation of the binding constants of the mutants showed that they are identical with the wt scFv within experimental error and thus no measurable conformational change is transmitted through the molecule. While these residues are on the opposite site of the antigen binding site, it is also known that V_L/V_H interface residues, which are about the same distance away from the binding site, can measurably change the binding constant (Chatellier *et al.*, 1996), probably by domain reorientation. The residues targeted here do not seem to have any long-range effects and are not expected to change the domain orientation.

The mutations described here do fall into the same class as those described earlier for an scFv fragment of the antibody McPC603 (Knappik and Pluckthun, 1995), green fluorescent protein (Crameri *et al.*, 1996) and glutathione reductase (Leistler and Perham, 1994) in that they do not affect the free energy of folding, as measured in urea denaturation experiments, but lower the aggregation rate *in vitro*. Thus, the large amount of insoluble wt protein *in vivo* does not result from an unfolding of an unstable native protein, because the mutants are not more stable (Figure 4), but give more soluble protein. We can also exclude the aggregation of native protein *in vivo* as the cause (Figure 6), since the solubility of the native mutant proteins is identical. Thus we propose that this aggregation occurs as an off-pathway in the folding reaction. The decisive factor is the aggregation rate itself, which differs between the mutants (Knappik and Pluckthun, 1995). These mutants are thus of a different nature from the mutants of REI (Chan *et al.*, 1996), where inclusion body formation and ΔG are inversely correlated.

Currently, there are three strategies available to improve the folding behavior of antibodies by protein engineering: (i) loop grafting to superior frameworks, (ii) back engineering from sequence comparisons and (iii) interface engineering. Clearly, these approaches are not mutually exclusive and their various implications will be briefly discussed.

The humanization of the anti-HER2 antibody 4D5 and thus the grafting to this human framework was found to be responsible for the superior expression properties of this Fab fragment, since the corresponding murine Fab fragment, in the identical *E.coli* host–vector system, gave rise to a 100-fold lower expression yield under identical fermentation conditions (Carter *et al.*, 1992). Loop grafting to superior frameworks must overcome several challenges, however, which again are only starting to be understood. It must be elucidated whether a framework is particularly robust against aggregation and how independent this is of the CDRs. Moreover, numerous contacts between CDRs and framework are important in mediating the exact conformation of all CDRs (Chothia *et al.*, 1989; Tramontano *et al.*, 1990), restricting the choice of some amino acids.

Back engineering from sequence comparisons (Knappik and Pluckthun, 1995) has been a valuable tool in establishing that indeed single residues can decide the fate of the protein and was originally inspired from the loop-grafting experiments mentioned above. However, many more experiments will be necessary before predictive rules will emerge. Since every given antibody may have different problems, these solutions will be tailor-made for a given antibody. The more important point is, however, that these rules or sequences emerging can directly be used for improving the general frameworks to be used in synthetic libraries.

While loop grafting and sequence back engineering can be carried out in all common antibody formats (Fab, Fv, scFv) and have been shown to have beneficial effects in all formats, the interface engineering will depend on the desired antibody format. The Fab fragment is a complete unit within the antibody, separated from the rest of the molecule by a flexible hinge region with no further protein–protein contacts, but any subdivision of the Fab fragment will expose a new interface. When using V_H domains by themselves (Ward *et al.*, 1989), the large V_H/V_L interface is exposed, frequently resulting in insoluble molecules (Kortt *et al.*, 1995). Camels (Desmyter *et al.*, 1996) and Llamas (Spinelli *et al.*, 1996) possess a subset of antibodies with unpaired soluble V_H domains and the crystal structures show that the residues normally forming the V_H/V_L interface are more hydrophilic. Indeed, it is possible to introduce these mutations in a human V_H domain (Davies and Riechmann, 1994, 1996) to increase solubility. However, while the domains become less 'sticky', they have lower melting points than the corresponding original human domains (Davies and Riechmann, 1996) even after adding stabilizing mutations. The V_H/V_L interface mutations are distinct from the v/c mutations investigated in the present study, which also do not cause a loss in stability.

The engineering of the v/c interface is in principle a general tool which can be applied both to general frameworks for libraries or individual antibody sequences. It was conceived from the notion that all Fv and scFv fragments have in common an unnatural former interface. A similar analysis has been carried out for scTCRs (Novotny *et al.*, 1991); however, different residues were mutated. We have identified all candidate residues at the former v/c domain interface and we have shown that this mutagenesis strategy can have a very strong effect. It will be very useful to avoid these hydrophobic patches in a synthetic scFv framework library, since they no longer serve any purpose. In such an optimized framework, adventitious exposed hydrophobic residues, which may arise in a random mutagenesis approach (Palzkill and Botstein, 1992; Stemmer, 1994) would be much less harmful since they would not likely be part of a contiguous hydrophobic surface. Thus, a short-term solution would be to use Table I and Figure 1 as a guide to spot rapidly hydrophobic patches in given sequences.

In the course of these studies we also reinvestigated the use of the standard $(Gly_4Ser)_3$ linker. A number of reports have shown that scFv fragments dimerize or multimerize with shorter linkers (Holliger *et al.*, 1993; Whitlow *et al.*, 1993, 1994; Desplancq *et al.*, 1994; Wu *et al.*, 1996). We compared a 15- and a 30-mer linker and found that the relative influence

of the mutations described is independent of the linker length. The difference in purification is due to oligomerization of the scFv fragment with the 15-mer linker (data not shown). The 30-mer linker we used seems to avoid diabody formation or multimerization. In general, we found 20-mer linkers to be sufficient to yield monomers reliably (Pluckthun *et al.*, 1996).

In conclusion, we have described a very general problem of scFv fragments in that they expose a former interface with hydrophobic patches and that strategic placement of negative charges can very significantly improve the *in vivo* folding properties of these molecules. Most scFv fragment used in biomedical research have to be produced in large amounts and thus the described strategy and the use of Figure 1 and Table I may serve as a guide to improve the expression level. If monoclonal antibodies will be replaced by synthetic libraries, such findings should be incorporated into the design of general frameworks, since the described hydrophobic residues no longer serve any purpose in the scFv fragments. Finally, this strategy may also be useful for other engineered protein domains, which are taken out of their natural context, to improve their folding properties and the *in vivo* expression yield.

Acknowledgements

We thank Dr Karl Proba for helpful discussions. This work was financially supported by the Schweizerische Nationalfonds, grant 3100–037717.93/1.

References

Bedzyk,W.D., Weidner,K.M., Denzin,L.K., Johnson,L.S., Hardman,K.D., Pantoliano,M.W., Asel,E.D. and Voss,E.W.,Jr (1990) *J. Biol. Chem.*, **265**, 18615–18620.

Bird,R.E.., Hardman,K.D., Jacobson,J.W., Johnson,S., Kaufman,B.M., Lee,S.M., Lee,T., Pope,S.H., Riordan,G.S. and Whitlow,M. (1988) *Science*, **242**, 423–426.

Brinkmann,U., Reiter,Y., Jung,S., Lee,B. and Pastan,I. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7538–7542.

Carter,P. *et al.* (1992) *Bio/Technology*, **10**, 163–167.

Chan,W., Helms,L.R., Brooks,I., Lee,G., Ngola,S., McNulty,D., Maleeff,B., Hensley,P. and Wetzel,R. (1996) *Folding Des.*, **1**, 77–89.

Chatellier,J., van Regenmortel,M.H.V., Vernet,T. and Altschuh,D. (1996) *J. Mol. Biol.*, **264**, 1–6.

Chothia,C. *et al.* (1989) *Nature*, **342**, 877–883.

Crameri,A., Whitehorn,E.A., Tate,E. and Stemmer,W.P.C. (1996) *Nature Biotechnol.*, **14**, 315–319.

Dale,G.E., Broger,C., Langen,H., D'Arcy,A. and Stuber,D. (1994) *Protein Engng*, **7**, 933–939.

Davies,J. and Riechmann,L. (1994) *FEBS Lett.*, **339**, 285–290.

Davies,J. and Riechmann,L. (1996) *Protein Engng*, **9**, 531–537.

Denzin,L.K., Whitlow,M. and Voss,E.W.,Jr (1991) *J. Biol. Chem.*, **266**, 14095–14102.

Desmyter,A., Transue,T.R., Ghahroudi,M.A., Thi,M.D., Poortmans,F., Hamers,R., Muyldermans,S. and Wyns,L. (1996) *Nature Struct. Biol.*, **3**, 803–811.

Desplancq,D., King,D.J., Lawson,A.D. and Mountain,A. (1994) *Protein Engng*, **7**, 1027–1033.

Ge,L., Knappik,A., Pack,P., Freund,C. and Pluckthun,A. (1995) In Borrebaeck,C.A.K. (ed.), *Antibody Engineering. A Practical Approach*. IRL Press, Oxford, pp. 229–266.

Gill,S.C. and von Hippel,P.H. (1989) *Anal. Biochem.*, **182**, 319–326.

Glockshuber,R., Malia,M., Pfitzinger,I. and Pluckthun,A. (1990) *Biochemistry*, **29**, 1362–1366.

Holliger,P., Prospero,T. and Winter,G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6444–6448.

Horn,U. *et al.* (1996) *Appl. Microbiol. Biotechnol.*, **46**, 524–532.

Huston,J.S., Levinson,D., Mudgett-Hunter,M., Tai,M.S., Novotny,J., Margolies,M.N., Ridge,R.J., Bruccoleri,R.E., Haber,E., Crea,R. and Oppermann,H. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5879–5883.

Huston,J., George,A.J.T., Tai,M., McCartney,J.E., Jin,D., Segal,D.M., Keck,P. and Oppermann,H. (1995) In Borrebaeck,C.A.K. (ed.), *Antibody Engineering. A Practical Approach*. IRL Press, Oxford, pp. 185–228.

Kabat,E.A., Wu,T.T., Perry,H.M., Gottesmann,K.S. and Foeller,C. (1991) *NIH Publication 91-3242*. US Department of Health and Human Services, Public Health Service, National Institutes of Health, Washington, DC.

Knappik,A. and Pluckthun,A. (1994) *BioTechniques*, **17**, 754–761.

Knappik,A. and Pluckthun,A. (1995) *Protein Engng*, **8**, 81–89.

Kortt,A.A., Guthrie,R.E., Hinds,M.G., Power,B.E., Ivancic,N., Caldwell,J.B., Gruen,L.C., Norton,R.S. and Hudson,P.J. (1995) *J. Protein Chem.*, **14**, 167–178.

Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.

Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.

Lesk,A.M. and Chothia,C. (1988) *Nature*, **335**,188–190.

Leistler,B. and Perham,R.N. (1994) *Biochemistry*, **33**, 2773–2781.

Lindner,P., Guth,B., Wulfing,C., Krebber,C., Steipe,B., Muller,F. and Pluckthun,A. (1992) *Methods: Companion Methods Enzymol.*, **4**, 41–56.

Middaugh,C.R., Tisel,W.A., Haire,R.N. and Rosenberg,A. (1979) *J. Biol. Chem.*, **254**, 367–370.

Miklasz,S.D., Gulliver,G.A. and Voss,E.W.,Jr (1995) *J. Mol. Recogn.*, **8**, 258–269.

Novotny,J., Ganju,R.K., Smiley,S.T., Hussey,R.E., Luther,M.A., Recny,M.A., Siliciano,R.F. and Reinherz,E.L. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8646–8650.

Pace,C.N. (1990) *Trends Biotechnol.*, **8**, 93–98.

Palzkill,T. and Botstein,D. (1992) *Proteins*, **14**, 29–44.

Pluckthun,A. (1992) *Immunol. Rev.*, **130**,151–188.

Pluckthun,A. (1994) In van Oss,C.J. and van Regenmortel,M.H.V. (eds), *Immunochemistry*. Marcel Dekker, New York, pp. 201–235.

Pluckthun,A., Krebber,A., Krebber,C., Horn,U., Knupfer,U., Wenderoth,R., Nieba,L., Proba,K. and Riesenberg,D. (1996) In McCafferty,J. (ed.), *Antibody Engineering. A Practical Approach*. IRL Press, Oxford, pp. 203–252.

Prickett,K.S., Amberg,D.C. and Hopp,T.P. (1989) *BioTechniques*, **7**, 580–589.

Prodromou,C. and Pearl,L.H. (1992) *Protein Engng*, **5**, 827–829.

Spinelli,S., Frenken,L., Bourgeois,D., Ron,L., Bos,W., Verrips,T., Anguille,C., Cambillau,C. and Tegoni,M. (1996) *Nature Struct. Biol.*, **3**, 752–757.

Stemmer,W.P.C. (1994) *Nature*, **370**, 389–391.

Stevenson,C.L. and Hageman,M.J. (1995) *Pharm. Res.*, **12**, 1671–1676.

Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.

Tramontano,A., Chothia,C. and Lesk,A. M. (1990) *J. Mol. Biol.*, **215**, 175–182.

Ullrich,H.D., Patten,P.A., Yang,P.L., Romesberg,F.E. and Schultz,P.G. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 11907–11911.

VanBogelen,R.A., Sankar,P., Clark,R.L., Bogan,J.A. and Neidhardt,F.C. (1992) *Electrophoresis*, **13**, 1014–1054.

Van Wielink,J.E. and Duine,J.A. (1990) *Trends Biochem. Sci.*, **15**, 136–137.

Ward,E.S., Gussow,D., Griffiths,A.D., Jones,P.T. and Winter,G. (1989) *Nature*, **341**, 544–546.

Whitlow,M. *et al.* (1993) *Protein Engng*, **6**, 989–995.

Whitlow,M., Filupa,D., Rollence,M.L., Feng,S. and Wood,J.F. (1994) *Protein Engng*, **7**, 1017–1026.

Whitlow,M., Howard,A.J., Wood,J.F., Voss,E.W.,Jr and Hardman,K.D. (1995) *Protein Engng*, **8**, 749–761.

Wilson,I.A. and Stanfield,R.L. (1994) *Curr. Opin. Struct. Biol.*, **4**, 857–867.

Winter,G. and Milstein,C. (1991) *Nature*, **349**, 293–299.

Wu,A.M. *et al.* (1996) *Immunotechnology*, **2**, 21–36.

Young,N.M., MacKenzie,C.R., Narang,S.A., Oomen,R.P. and Baenziger,J.E. (1995) *FEBS Lett.*, **377**, 135–139.

Received September 12, 1996; revised November 8, 1996; accepted November 19, 1996