JMB

Reproducing the Natural Evolution of Protein Structural Features with the Selectively Infective Phage (SIP) Technology. The Kink in the First Strand of Antibody kappa Domains

Stefania Spada, Annemarie Honegger and Andreas Plückthun*

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

The β-sandwich structure of immunoglobulin variable domains is characterized by a typical kink in the first strand, which allows the first part of the strand to hydrogen bond to the outer β-sheet (away from the V_H-V_L interface) and the second part to the inner β-sheet. This kink differs in length and sequence between the V_κ, V_λ and V_H domains and yet is involved in several almost perfectly conserved interactions with framework residues. We have used the selectively infective phage (SIP) system to select the optimal kink region from several defined libraries, using an anti-hemagglutinin single-chain Fv (scFv) fragment as a model system. Both for the kink with the V_κ domain length and that with the V_λ length, a sequence distribution was selected that coincides remarkably well with the sequence distribution of natural antibodies. The selected scFv fragments were purified and characterized, and thermodynamic stability was found to be the prime factor responsible for selection. These data show that the SIP technology can be used for optimizing protein structural features by evolutionary approaches.

© 1998 Academic Press

Keywords: selectively infective phage; phage display; libraries; single-chain Fv fragment; *cis*-proline

*Corresponding author

Introduction

Over the past few years, directed molecular evolution has emerged as a new tool in protein engineering that may complement computer-aided approaches in studies where the forces and effects are too subtle to be predicted by current methods.

Present address: S. Spada, Department of Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland.

Abbreviations used: BSA, bovine serum albumin; CDR, complementarity-determining region; ELISA, enzyme-linked immunosorbent assay; gIII, geneIII; IPTG, isopropyl β-D-thiogalactopyranoside; N1 and N2, N-terminal domains of geneIII protein; Ni-IDA, nickel-iminodiacetic acid; PBS, phosphate-buffered saline; PVDF, polyvinyldifluoride; SB, super broth; scFv, single-chain variable fragment of an antibody; SIP, selectively infective phage; V_H, heavy chain variable domain; V_L, light chain variable domain; V_κ, light chain variable domain type kappa; V_λ, light chain variable domain type lambda; wt, wild-type.

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

Representative examples in several application areas illustrate the power of this approach (Hall & Knowles, 1976; Stemmer, 1994; Low *et al.*, 1996; Baca *et al.*, 1997; Hanes & Plückthun, 1997; Moore *et al.*, 1997). Like the natural evolution of proteins, directed evolution alternates between creation of diversity and selection. Depending on the problem at hand, diversity may be focused on certain regions in a protein or the whole sequence may be sprinkled with mutations.

Out of this diversity, superior molecules need to be selected. In most non-enzyme cases, the only directly selectable function is binding to a ligand. However, many structural features that affect the amount of the protein molecule produced or its stability indirectly affect selection efficiency, as long as ligand-binding is limited to the correctly folded molecule. It is this coupling of function to the native state that we exploit in the present study to optimize protein structure.

A whole array of techniques has been developed to select for molecular interactions, such as phage display (Smith, 1985; Dunn, 1996), the yeast two-hybrid system (Bai & Elledge, 1996; Warbrick,

1997), the peptides-on-plasmids system (Cull *et al.*, 1992), ribosome display (Mattheakis *et al.*, 1994; Hanes & Plückthun, 1997), yeast surface display (Boder & Wittrup, 1997) and bacterial display (Georgiou *et al.*, 1997).

A selection for structural parameters by using phage display has been observed to occur e.g. in the optimization of human growth hormone. A variant in which 15 residues were mutated was selected, resulting in a 400-fold improved receptor-binding activity, but residues involved in the packing were left unchanged (Ultsch *et al.*, 1994). Similarly, in selecting variants of peptostreptococcal protein L for binding on IgG, the buried positions, important in the formation of the small hydrophobic core, remained highly conserved (Gu *et al.*, 1995). Minimized two-helix variants of the staphylococcal protein A were selected for binding on IgG, which compensated the loss of the third helix by several changes within the domain, increasing stability (Braisted & Wells, 1996; Starovasnik *et al.*, 1997). Antibodies without disulfides, which still could specifically bind the antigen, were evolved by selecting for globally stabilizing mutations, compensating the loss of the disulfides (Proba *et al.*, 1998).

To select for structural features, the selection has to be quite resistant to enriching "sticky" molecules and must be convenient enough to be applied over multiple rounds, in order to enhance very small selective advantages. We therefore chose to use the selectively infective phage (SIP) system (reviewed by Spada & Plückthun, 1997; Spada *et al.*, 1997). The SIP technology allows selection of interacting partners in solution, without any solid-phase requirement, and directly couples recognition to phage propagation. A protein of interest is fused to the C-terminal domain of the filamentous phage gene III (gIII) protein in all of its copies and displayed on the tip of the phage. In contrast to phage display, however, the resulting phage is non-infective. Only upon specific interaction with an adapter molecule, in which the N-terminal domains of the gIII protein (N1 and N2) are connected to the ligand, is the gIII protein, essential for infection, restored and does the phage become infective (Figure 1). Selection is a result of the ability to infect: only phages that carry a protein recognizing the ligand will infect bacteria and the genetic information for this protein variant will be propagated. Since this procedure is very simple, it can be repeated multiple times and used to amplify even small selective advantages. This directness makes the SIP system very powerful and extremely useful in molecular optimization, eliminating one intrinsic limitation of traditional phage display technology: in phage display, all phages are infective, and selection occurs upon panning on a ligand. Binders are collected and used to infect cells; the problem resides in the unwanted amplification of non-specific as well as specific binders (Adey *et al.*, 1995).
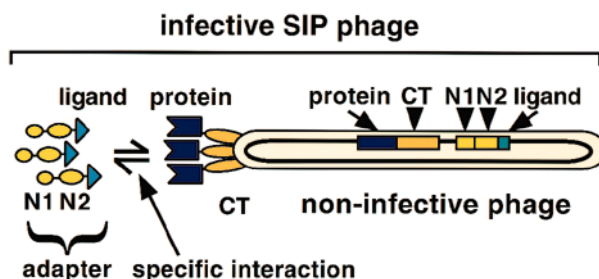

Figure 1. The selectively infective phage system. Only upon specific protein-ligand interaction is gIII protein functionality restored and does the phage becomes infective. N1 and N2 indicate the N-terminal and CT the C-terminal domains of the phage gIII protein (reviewed by Spada *et al.*, 1997; Spada & Plückthun, 1997).

We have applied this technology to study a conserved motif of immunoglobulin structure by molecular evolution. Antibodies contain two heavy and two light chains, each consisting of variable and constant domains with the so-called immunoglobulin fold, a β-sandwich with characteristic topology. The variable domains use this conserved scaffold to hold the complementarity-determining regions (CDRs) or hypervariable loops (Poljak, 1991). While in constant domains each chain segment forms a strand that clearly belongs to either one or the other of the two sheets of the β-sandwich, in variable domains the N-terminal segment switches in a sharp kink from hydrogen bonding to the "outer" sheet (relative to the dimer interface) to hydrogen bonding to the inner sheet. This kink is present in immunoglobulin and T-cell receptor variable domains, but not in constant domains (Figure 2).

Nature has found a number of different solutions to maintain this overall structure. In V_κ domains a *cis*-proline residue in position L8 (sequence numbers according to Kabat *et al.*, 1991) is strongly conserved. In V_λ domains, the kink is one amino acid residue shorter, and usually contains at least one *trans*-proline residue (Figure 3). V_H domains, which correspond to V_λ in length at this part of the chain, contain glycine residues in positions H8 and/or H10, frequently with positive ϕ angles, to make the kink.

In this study we wanted to examine whether directed molecular evolution can be used to rationalize the natural selection of the kink residues and whether it would reproduce the selection nature has carried out. We chose a kappa chain as a model system. As discussed in more detail later, 100% of the human and 85% of the mouse kappa chains listed in the Kabat database (http://immuno.bme.nwu.edu) contain proline in position L8. Crystal structures show that the peptide bond of this proline residue is in the *cis* conformation, while in the few cases of those murine antibodies that do not contain proline at position L8 the replacement amino acid is *in trans*. The lambda chain is
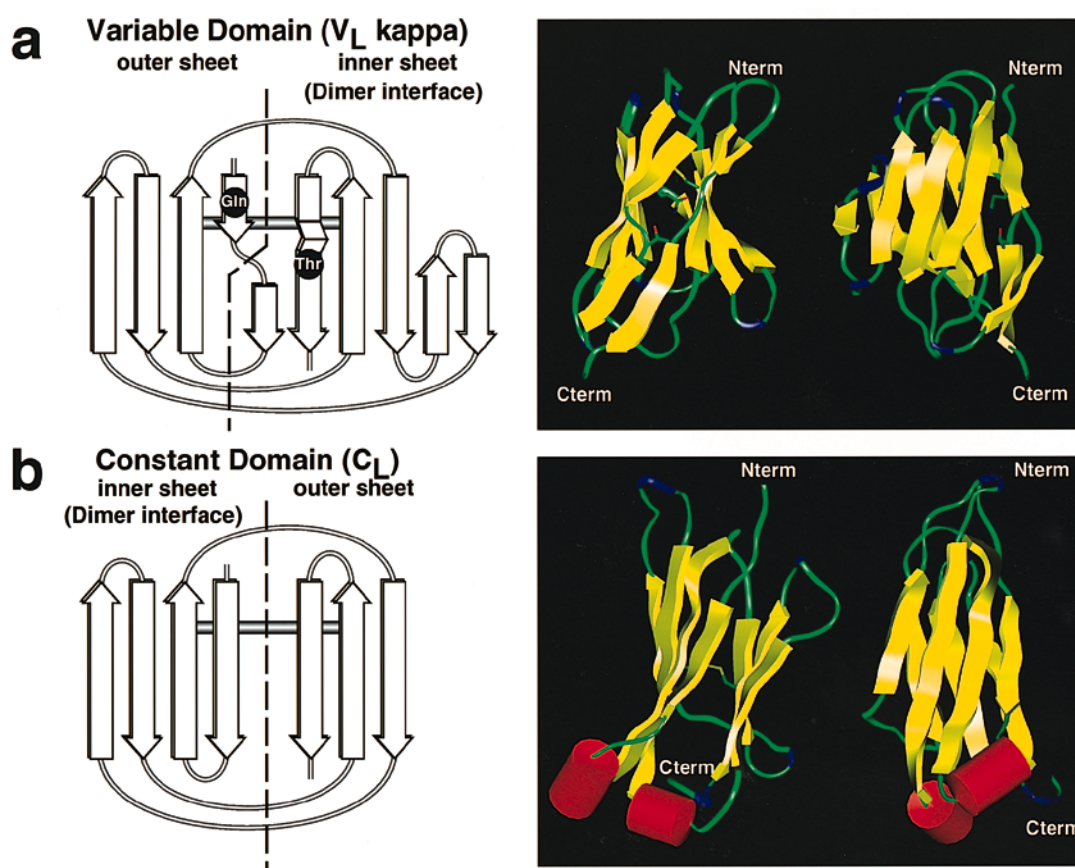
Figure 2. A representation of the immunoglobulin variable domain (a) and constant domain (b) folding topology. The N-terminal segment of the variable domain (a) switches from hydrogen bonding to the outer (relative to the dimer interface) β-sheet to hydrogen bonding to the inner β-sheet. This switch is visible as a marked kink in the N-terminal strand, juxtaposed to the G-X-G β-bulge in the most C-terminal strand of variable domains (corresponding to positions L99 to L101 in V_L domains). The disulfide bonds are indicated as rods, the β-bulge as a pleat, and the kink as an interruption of the first β-strand, and the location of the perfectly conserved Gln and Thr residue are indicated, whose sides-chains stabilize this topology. In constant domains (b), the two β-sheets making up the immunoglobulin β-sandwich are completely separated. Note that in C_L and C_H1 domains the dimer interface involves the β-sheet homologous to the outer sheet in variable domains. Thus, the location of dimer interface is not conserved between different types of immunoglobulin domains.

one amino acid residue shorter and Kabat *et al.* (1991) place this sequence gap at position L10. As Figures 3 and 4 show, however, the length variability is accommodated fully between positions L6 and L9 of V_κ, and we therefore placed the gap in position L7. In human V_λ, the sequence Pro-Pro is preferred in positions L8 and L9 (56%), all other human V_λ contain at least one of the two proline residues. The few mouse V_λ chains described predominantly contain the sequence Glu-Ser in this position, and all seem to be derived from the $V_{\lambda 1}$ gene, even though the mouse contains three V_λ genes (Sanchez *et al.*, 1987).

In this study we randomized the *cis*-proline residue at position L8 and its two flanking residues in the kappa chain of the anti-hemagglutinin 17/9 antibody scFv fragment. We included in the library the possibility of a lambda-like kink by having only two random amino acid residues and one deletion, and we investigated what sequence the SIP system selects for. The molecules obtained were characterized to enable us to understand better the

parameters that guided the selection. A striking reproduction of the natural sequence distribution was found. The physical properties of the selected clones suggest that stability is the prime criterion of selection.

Results

Design and construction of the library

The behavior of the scFv fragment of the anti-hemagglutinin 17/9 antibody in SIP experiments (Figure 1) was previously investigated, and it was demonstrated that the infectivity of the phage is restored by the specific interaction between the scFv and its antigen (Krebber *et al.*, 1995). We made use of this selection for function as a selection for native molecules, and exploited the very low background of the selection system.

We chose as a model system the V_L domain in the scFv fragment of the anti-hemagglutinin antibody 17/9. Its crystal structure has been deter-
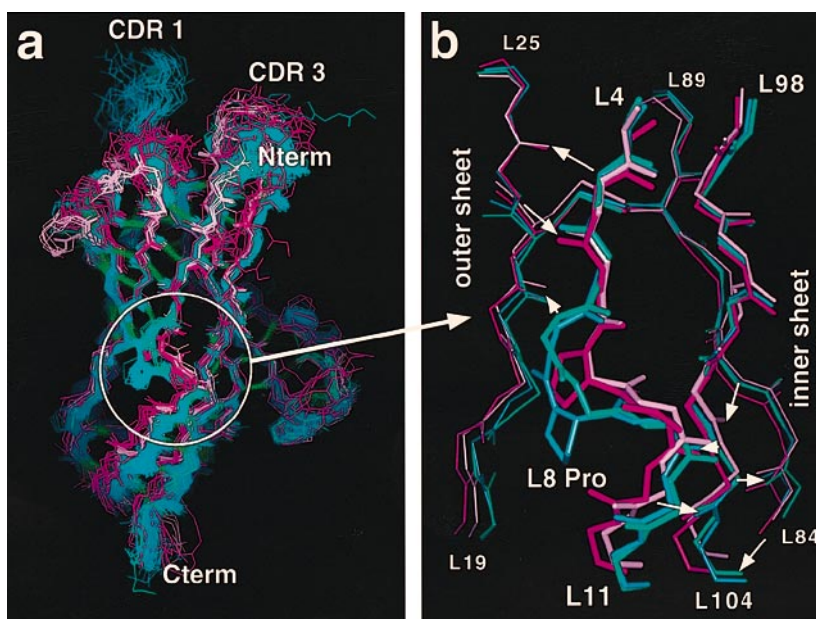
Figure 3. Comparison of the kink conformation in V_κ and V_λ domains. a, An overview of 99 aligned non-redundant mouse and human V_L domains. For the 3D alignment of the different V_L domains, a subset of the most highly conserved C^α-positions was defined. This subset consists of positions L3 to L7, L20 to L24, L33 to L39, L43 to L49, L62 to L66, L71 to L75, L84 to L90 and L97 to L103 (numbering according to Kabat *et al*, 1991). These 48 C^α-positions allow a least-squares superposition with a rms deviation of $\leqslant 0.5$ Å for almost all V_L domains listed in the Brookhaven Protein Data Bank. The aligned structures are color-coded according to their sequence subtype: blue, 75 V_κ domains (human and mouse) containing *cis*-Pro in position L8 (denoted V_κ *cis*-Pro in other Figures); cyan, seven mouse V_κ domains with L8 Thr or His (denoted V_κ no Pro); magenta, 12 human V_λ chains with L8/L9 Pro-Pro (denoted V_λ Pro-Pro); pink, 5 mouse V_λ domains with L8/L9 Glu-Ser (denoted V_λ no Pro). b, Detailed view of one representative sample of each subclass of V_L domains. V_κ *cis*-Pro (blue), murine anti-azophenyl arsenate antibody 36.71, PDB entry 6fab (1.9 Å resolution, Strong *et al*., 1991); V_κ no Pro (cyan), anti-cyclosporine antibody, PDB entry 1ikf (2.5 Å resolution, Altschuh *et al*., 1992); V_λ Pro-Pro (magenta), murine antibody SE155-4, PDB entry 1mfe (2.0 Å resolution, Cygler *et al*., 1991); V_λ no Pro (pink), human Bence-Jones protein rhe, PDB entry 2rhe (1.6 Å resolution, Furey *et al*., 1983). Arrows indicate hydrogen bonds, with the direction from donor to acceptor, and the size giving some depth-cueing.
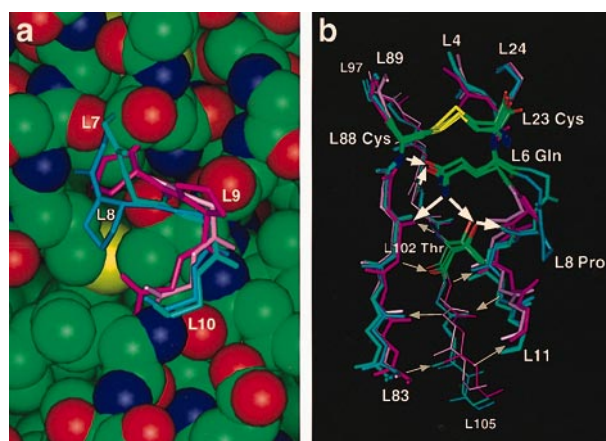

Figure 4. Environment of V_L residues L7-L10 in antibody 17/9. a, Residues L7 to L10 backbone and Pro side-chains in stick-mode represented against a CPK background of the remainder of antibody 17/9, residues L1 to L6 and L11 to L108 (PDB entry 1ifh, 2.8 Å resolution, Schulze-Gahmen *et al*., 1993). Stick representation of backbone and Pro side-chain heavy-atoms of residues L7 to L10 of the four different structural classes (see Figure 3a) are shown exemplified by 6fab (V_κ *cis*-Pro), 1ikf (V_κ no Pro), 1mfe (V_λ Pro-Pro), 2rhe (V_λ no Pro). The stick colors are the same as in Figure 3. b, Hydrogen bonding pattern of hydrophilic core residues Gln L6 and Thr L102. Arrows indicate hydrogen bonds, with the direction from donor to acceptor, and the size giving some depth-cueing. The stick colors are the same as in Figure 3.

mined (Schulze-Gahmen *et al*., 1993), and it has been shown to contain the sequence Ser-*cis*-Pro-Ser (positions L7 to L9). We randomized positions L7 to L9. For library construction, we used the degenerate codon NNS (N = A, T, C or G; S = C or G) that includes all the possible amino acids and reduces the library size while keeping the diversity. This strategy also allowed us to rule out any potential wild-type (wt) contamination of the selected molecules having the wt protein sequence, since their nucleotide sequence is different. Furthermore, we created a second library where the light chain was one amino acid residue shorter than the wt antibody light chain to mimic the situation in lambda chains, where we randomized the two positions frequently occupied by *trans*-proline (Table 1).

A SIP library was created with scFvs displayed on phages and the adapter molecule fused to the antigen encoded in the same phage vector (Figure 1). Both the genes for antigen (kept constant) and antibody (the library) are thus packaged in the same phage. The randomization of two or three positions at the kink around position L8 was carried out by site-directed mutagenesis with degenerate oligonucleotides, yielding 1.7×10^4 and 5×10^5 different clones for the length two (Lib2) and length three (Lib3) libraries around position L8, respectively (Table 1). The theoretical diversity of the libraries is 1.0×10^3 and 3.3×10^4, resulting

Table 1. Design of the libraries used for the randomization of amino acids at position 7, 8 and 9 in the 17/9 antibody light chain

Library	Amino acid position													
	1	2	3	4	5	6	**7**	**8**	**9**	10	11	12	13	14
17/9 antibody	D	I	V	M	T	Q	**S**	**P**	**S**	S	L	T	V	T
Lib2 (2 randomized positions)	D	I	V	M	T	Q	–	**X**	**X**	S	L	T	V	T
Lib3 (3 randomized positions)	D	I	V	M	T	Q	**X**	**X**	**X**	S	L	T	V	T

The sequence of the antibody 17/9 is shown as a reference. Positions 7, 8 and 9 are highlighted. The NNS (N = A, T, C or G; S = C or G) codon was used to introduce random amino acids (X). The theoretical diversity was 1.024×10^3 and 3.3×10^4 for the Lib2 and the Lib3 libraries, respectively, and the experimental libraries were 17 and 15 times this size.

in 17-fold and 15-fold oversampling, respectively. To avoid any bias for wt, a mutant in which a stop codon had been introduced at positions L8 and L9 was used as template for mutagenesis.

Random clones were analyzed by restriction digest and DNA sequencing before selection. All clones analyzed showed a correct open reading frame for the scFv fragment and no obvious bias in the distribution of amino acids in the randomized positions.

SIP selection

The scFv fragments were selected for their ability to interact with the antigen, which in the adapter molecule is fused to the N-terminal domains of gIIIp (Figure 1) and co-expressed in the bacterial cell. Selected colonies were analyzed in order to exclude undesired genetic recombination (Krebber *et al.*, 1995) and sequenced. The numbers of input phages used for each round of selection, as well as the number of infected colonies obtained, are shown in Table 2.

The two libraries, Lib2 and Lib3, with two and three randomized positions, respectively, were kept separate for the first rounds of selection. In both cases, only clones that expressed an scFv fragment with a correct reading frame were found after one round of SIP, whereas clones that had stop codons at the randomized positions did not survive the selection, demonstrating the ability of the system to select for functional proteins.

In the case of the library with three randomized positions (Lib3), 50% of clones contained proline in position L8 after only one round, while after the second round all clones had proline in position L8. After three rounds of SIP we found a consensus for the flanking positions: 50% of the

clones showed the Ser-Pro-Leu (S-P-L) pattern (Figure 5).

In the Lib2 library, one amino acid residue shorter in the light chain and therefore of lambda-like length, we found 50% of clones with a proline residue (P-X) after one round of selection; after two rounds, 90% of the clones contained a proline residue, and the pattern P-P was found together with P-X and X-P; after three rounds, we observed an enrichment of the P-P pattern over any other X-P or P-X (Figure 5). These are the commonest motifs in natural lambda sequences (Figures 5 and 6). Proline is found to be 84% and 72% conserved in positions L8 and L9, respectively, in human lambda chains, while 56% of all human V_λ sequences contain proline in both positions.

We then combined clones isolated from the separate libraries after two rounds of selection and let them compete against each other. After two rounds, we found the Ser-Pro-Leu sequence to be enriched. This was observed also when the two libraries were pooled at the beginning and then subjected to SIP selection: the longer sequence with proline in position L8 was enriched over the short sequence (Figure 5).

We conclude that the SIP system has enriched the same molecules that nature selected for through evolution. When the kappa light chain of 17/9 antibody was randomized, the very conserved motif X-P-X found in the majority of the kappa chains was clearly selected. Moreover, the most prominent sequences Ser-Pro-Ser and Ser-Pro-Leu found in the SIP selection are also the commonest kappa sequences in the Kabat database of antibody sequences (Figure 5). Furthermore, we observed an enrichment of this motif over the shorter sequence found in lambda chains, confirming that SIP has found the same solution as nature.

Table 2. Number of selected colonies obtained after each round of SIP selection

Library	1st round	2nd round	3rd round
Lib2 (2 randomized positions) (A)	2.5×10^3 (10^{12})	10^3 (10^{12})	10^3 (5×10^{11})
Lib3 (3 randomized positions) (B)	9×10^2 (10^{12})	3.4×10^2 (5×10^{11})	5×10^3 (5×10^{11})
Lib2 + Lib3 combined at the beginning (C)	10^4 (10^{12})	10^3 (5×10^{11})	3×10^3 (10^{11})
Lib2 + Lib3 combined after 2 rounds each (D)	5×10^3 (10^{12})	10^3 (5×10^{11})	

The number of input phages for each selection experiment is shown in parentheses.
The letters after each library correspond to the experiments in Figure 5.

Figure 5 (*legend opposite*)

Characterization of selected molecules

Monitoring cell growth during expression

The scFvs from isolated clones were subcloned into the vector pIG6 for soluble, periplasmic expression (Ge *et al.*, 1995) for further characterization. We analyzed five clones in more detail: the wt scFv 17/9 (S-P-S), the clone enriched in the SIP selection (S-P-L), a ''kappa-length'' light chain sequence with no proline in position L8 (L-T-R), a ''lambda-length'' light chain sequence with prolines residues (P-P), and a ''lambda-length'' light chain sequence with no proline (S-S). Both proline-free sequences, L-T-R and S-S, were lost during the selection and served as a comparison for the selected sequences.

The expression of the different clones in *Escherichia coli* JM83 was compared, since it had been shown that folding defects in the antibody can influence the growth rate of *E. coli*, presumably because of toxic effects created by the unfolded protein (Knappik & Plückthun, 1995). The growth at 25°C and 37°C was monitored after induction of scFv production with IPTG. When expression was carried out at 25°C, the different clones followed identical growth curves (Figure 7a). At 37°C, we observed a quite different behavior (Figure 7b). The wt S-P-S sequence and the enriched clone S-P-L behaved very similarly. Cells grew for about three hours after induction, then stopped growing and after one more hour began to lyse. Surprisingly, in the case of the P-P mutant, cells were still growing after five hours of induction. On the other hand, cells producing the non-selected mutants stopped growing two hours after induction, showing that some properties of these mutant proteins can indeed affect the physiology of the cell (see below). This effect appeared more dramatic in the case of the L-T-R mutant than in the S-S mutant.

Soluble expression of scFv

The ratio of soluble scFv fragments to insoluble aggregates expressed in the bacterial cells was determined by separating the soluble and insoluble fractions of cell lysates after three hours of expression at 25°C. The fractions, normalized to the same absorbance, were analyzed by immunoblotting.

The amounts of soluble protein were found to be about the same for all the mutants, while we observed remarkable differences in the amount of insoluble protein (Figure 8). The wt S-P-S and the enriched mutant S-P-L showed the same ratio of soluble to insoluble protein. An even better ratio was observed for the P-P mutant, where almost no insoluble protein was detected. This is in agreement with the ability of the cells producing this mutant to continue growing for more than five hours after induction at 37°C, suggesting that it is due to the lower level of aggregates accumulated in the cell that, by a poorly understood mechanism, may cause the lysis response. Quite different results were observed for the non-selected mutants. More insoluble protein was detected for the S-S mutant, compared to the wt, but the result obtained for the L-T-R mutant was even more dramatic, with most protein detectable being insoluble. These results are in agreement with the analysis of purified proteins by analytical gel-filtration. For the partially purified L-T-R mutant, both monomers and aggregates were found, while the S-P-S, S-P-L and P-P mutants were found exclusively in monomeric form (data not shown) and could be purified.

Figure 5. Sequence statistics in natural antibodies and SIP-selected libraries. a, Sequence statistics and selection results in the randomized positions L7, L8 and L9. b, Sequence statistics of residues L6, L10 and L102, which directly interact with the kink. For sequence statistics (denoted Kabat), kappa and lambda chains were considered separately, as were murine (mu) and human sequences (hu). In lambda sequences, in contrast to the standard alignment used by Kabat *et al.* (1991), the gap was introduced in position L7, which agrees more closely with the structural alignment. Human lambda sequences were either considered as a whole (denoted all) or further subdivided into those containing ProL8 and ProL9 (denoted PP), those containing Pro only at L8 (denoted PX), and those containing Pro only at L9 (denoted XP). Murine lambda sequences were not subdivided, because of the very small number of V_λ genes. Murine kappa sequences were either considered as a whole (denoted all) or subdivided into those sequences containing a Pro residue at position L8 (denoted P), or those not containing ProL8 (denoted X). Human kappa sequences were not further divided. Note that these divisions do not correlate with kappa or lambda subclasses. The most frequently found residue types at each position have colored backgrounds, with the darkness of the color indicating the proportion. Below each panel the most frequently found residue type is indicated (denoted consensus), as is the percentage of sequences carrying the consensus and the absolute numbers of sequences in each class. V_L entries with reasonably complete (>90%) sequence in the July 1997 release of the Kabat Database (http://immuno.bme.nwu.edu; 583 human V_κ sequences, 1388 mouse V_κ sequences, 428 human V_λ sequences and 96 mouse V_λ sequences) were the basis for the statistics. The right-hand panel of a (denoted SIP libraries) indicates the results of the selection experiments. For the randomized positions in the antibody 17/9-derived libraries, the theoretical amino acid distribution (indicated as round *, leftmost column) expected from a (A/T/C/G)-(A/T/C/G)-(C/G) library (31 coding triplets, one stop codon) was compared to the (position-independent) distribution found in the unselected library (denoted round 0) and the position-dependent distributions found in Libraries A, B, C and D (explained in Table 2) after one, two and three rounds of SIP selection (denoted as round). Below each panel the most frequently found residue type is indicated (denoted *consensus*), as is the percentage of sequences carrying the consensus and the absolute numbers of sequences in each class.
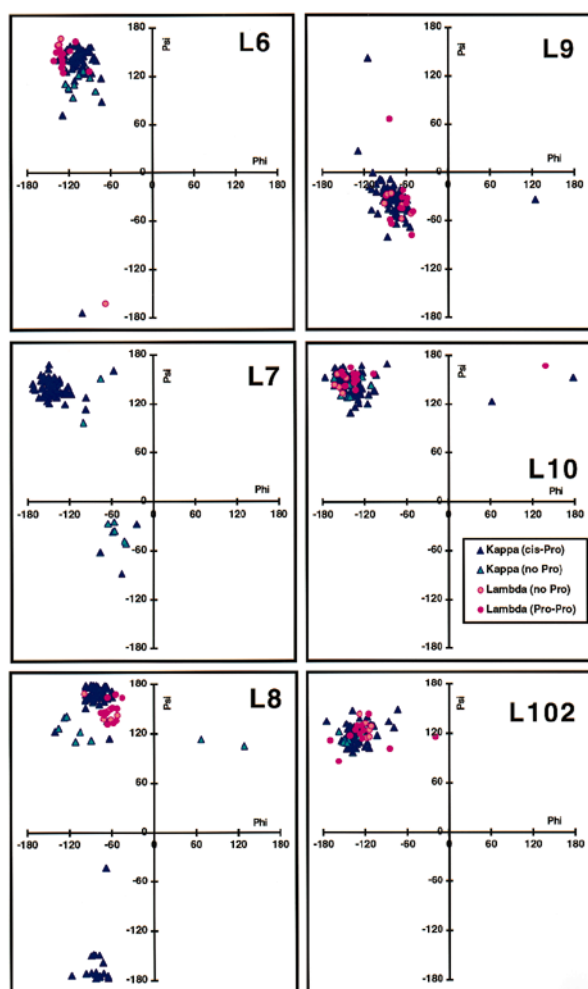
Figure 6. Main-chain torsion angles. From the 99 aligned non-redundant V_L structures in the PDB database (Figure 3a), main-chain torsion angles of residues L6, L7, L8, L9, L10 and L102 were extracted and plotted in separate Ramachandran plots for each sequence position.

Figure 7. Comparison of growth curves of *E. coli* JM83 cells expressing various scFv antibody fragments. a, Cells expressing the wt S-P-S, the S-P-L, P-P, S-S and L-T-R mutants were monitored at 25°C. Protein expression was induced with IPTG after three hours of cell growth, and cell density was followed by measuring absorbance at 550 nm. b, The same experiment was carried out at 37°C, except that protein expression was induced after two hours of cell growth.

All five mutants were tested for antigen-binding in ELISA, using soluble fractions of crude extract prepared after three hours of expression at 25°C. All were found to be able to bind antigen and specific binding was verified by the ability to be completely inhibited at an antigen concentration of 10^{-6} M in competition ELISA (data not shown).

Determination of scFv stability by urea denaturation curves

The stability of the scFv could be one of the possible parameters guiding the selection process, as it may influence the number of infective complexes formed on the phage. We therefore determined the stability of the wt S-P-S, and the mutants found after SIP selection, S-P-L and P-P. After purification the proteins were analyzed for stability by urea equilibrium-denaturation, which
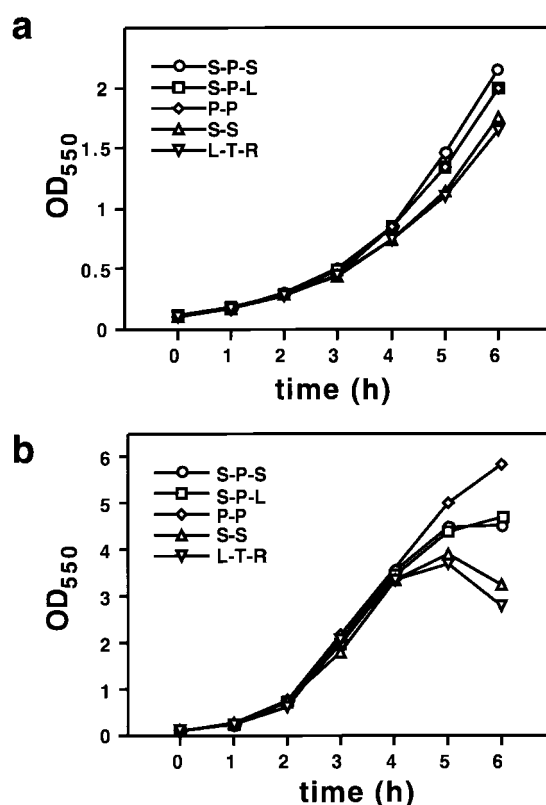
was followed by the shift of the emission maximum.

A midpoint of transition at 4 M urea was found for the wt S-L-S (Figure 9). The same value was found for the S-P-L mutant, whose curve could be superimposed upon the wt curve. A different behavior was found for the P-P mutant, with a midpoint of transition at only 3.3 M urea and a flatter curve at higher urea concentration. While this might suggest a less cooperative behavior than the other proteins analyzed, it is more likely that the curve does not reflect a two-state behavior and intermediate species with one unfolded domain or a disrupted dimer interface occur at an intermediate denaturant concentration (Figure 9). The scFv fragment with the kappa-length chain and proline in position L8 yielded a protein clearly more stable than any scFv with the lambda-length light chain, even though the latter was produced in bacterial cells with less insoluble material.
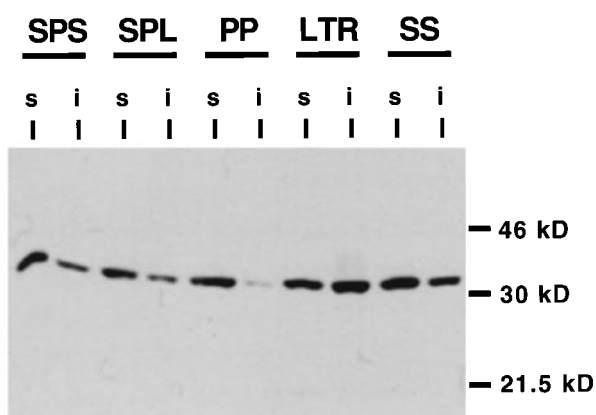
Figure 8. Western blot analysis of the soluble (s) and insoluble (i) fractions of the cell lysates, prepared as described in Materials and Methods. Cells expressing various scFv antibody fragments at 25°C for three hours were analyzed. Protein samples were normalized according to A_{550} of the cells before loading.

Discussion

The final result of our selection was the enrichment of scFv fragments that contained a kappa-like chain sequence with a proline residue in position L8. Two different libraries were combined to represent both kappa-like and lambda-like sequences (the latter being one amino acid residue shorter). We conclude that a natural kappa light chain antibody, the anti-hemagglutinin 17/9, prefers the inherited kappa chain structure, which was more stable, despite giving rise to more insoluble protein than the shorter version. SIP has therefore selected for the same solution to the problem of maintaining the structure of the domain that nature has found through evolution. This observation makes SIP a very useful tool for the optimization of protein structures.
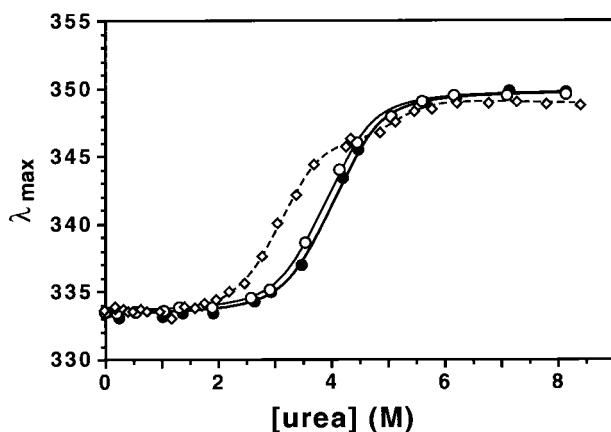

Figure 9. Equilibrium urea denaturation curves of the wild-type 17/9 scFv fragment (filled circles) and the S-P-L (open circles) and the P-P (open squares) scFv mutant proteins.

Structural difference between different kinks

What causes the kink in immunoglobulin variable domains, and why is it so remarkably conserved? Buried in the core of the V_L domain are a number of strictly conserved hydrophilic residues (Figure 4b). The residues Gln L6 and Thr L102 are as highly conserved as the core Trp residue L35 and the Cys residues L23 and L88 forming the disulfide bridge. They form a network of hydrogen bridges within the hydrophobic core of the domain, which also involves main-chain residues in the kink (C=O of L8) and the β-bulge formed by the conserved Gly-X-Gly motif (NH L101) in framework 4. The completely buried side-chain amide oxygen atom of Gln L6 accepts hydrogen bonds from the main chain NH of Cys L88 and Gly L101, while the side-chain amide nitrogen atom of Gln L6 donates hydrogen bonds to the main chain C=O of Tyr L86 and to the buried side-chain oxygen atom of Thr L102. Thr L102 in turn donates a hydrogen bond to the main-chain C=O of L8. Given the difference in the overall conformation of the different variants of kinks seen in different V_L domains, the placement and orientation of this C=O group of L8 in the middle of the kink is remarkably well conserved (Figures 3, 4 and 6).

The *cis*-Pro L8 of V_κ and *trans*-Pro L8 of V_λ (putting the sequence gap at position L7, as explained in the Introduction) are placed flat against the surface of the domain, thereby "capping" the hydrogen bonding network of the outer β-sheet (Figure 2) and orienting the C=O group of L8 towards the domain core to accept a hydrogen bond from the side-chain OH group of Thr L102 (Figure 4). The side-chains of L7 (V_κ) and L9 are pointing outwards without many interactions, explaining the tolerance for sequence variability in these positions. V_λ domains with and without proline in positions L8 and L9 show no significant difference in backbone conformation, while in V_κ domains lacking Pro L8 the *trans* peptide bond has to be accommodated by a change in main-chain torsion angles, which predominantly affects L7 and, to some extent, L8 (Figure 6). While in V_κ with *cis*-Pro in L8 a hydrogen bond can be formed between the main-chain NH of L21 and the C=O of L7, in V_κ without this *cis*-Pro the main-chain C=O of L7 is turned outwards and hydrogen-bonds with the solvent.

From analysis of the 17/9 Fab fragment structure (Schulze-Gahmen *et al.*, 1993) we did not observe any contact between either of the two flanking positions, L7 and L9, and any neighboring amino acid residues. This explains the weak consensus at these positions and the enrichment of more than one molecule from the library: no specific amino acid is needed in these positions and any small, uncharged amino acid residue does not disturb the structure of the antibody.

Comparison of sequences selected with the natural distribution

The most striking feature we found in all molecules enriched from the kappa-like library is the X-P-X pattern that was selected from the randomly distributed amino acid positions. After only one round, it constituted 50% of all molecules, and after two rounds 100% of the molecules contained it. Since all kappa domains of known structures with proline at this position have a *cis* bond, we reason that a *cis*-proline is very strongly selected. All human kappa chain sequences contain the same X-P-X pattern, with *cis*-proline 100% conserved in position L8, a quite conserved serine residue in position L7 (97%) and serine, glycine, alanine or leucine in position 9 (35%, 21%, 19% and 15%, respectively; Figure 5). In mouse kappa chains, *cis*-proline in position L8 is found in 85% of the sequences, with the rest being threonine (8%), alanine (3%) and histidine (2%). Similar to the human sequences, preferences but no strict rules are found for position L7 in which serine (71%) or threonine (23%) occur or position L9 in which alanine (45%), serine (24%) or leucine (17%) are found (Figure 5).

Remarkably, this corresponds quite well to the amino acids we obtained in those positions in our selection. One might caution, however, that the loss of variability after the first round of selection, where only 900 clones entered the second round, may have contributed to the enrichment of the consensus S-P-L in the Lib3 library.

In the case of the lambda-like library (one amino acid residue shorter than the kappa chains), the selected P-P pattern was only weakly enriched over the others, but proline was already very strongly enriched after the first round and the presence of at least one *trans*-proline residue must be advantageous for the protein. Again, a striking similarity to the distribution of sequences in the database can be found. Positions L8 and L9 in human lambda chains appear to be quite conserved: *trans*-proline occurs in position L8 in 84% of the sequences, and in 72% of the sequences in position L9 (Figure 5).

We clearly found a remarkable correlation between the molecules that SIP has selected in the two libraries and the molecules selected by nature through evolution. It thus appears that SIP imitates the natural selection parameters rather closely.

The selection parameters of SIP

Upon more detailed analysis of the parameters that guide the selection, the first observation is that SIP strongly selects for functional proteins: no clone with a stop codon in the open reading frame of the scFv fragment survived a single round of selection. This feature, already described in a study of an scFv-linker library (Hennecke *et al.*, 1997), emphasizes the advantage of this system for selection: infection is strictly dependent on the inter-

action between the two partners and no background is observed (Krebber *et al.*, 1995).

In SIP, selection is mediated by the interaction between the two partners that restores the gIII protein and therefore the infectivity of the phage; thus, a high affinity between the two partners is the essential requirement for a successful selection. As already described (Duenas *et al.*, 1996; Pedrazzi *et al.*, 1997), the system selects for high-affinity over low-affinity binders. Nevertheless, other molecular properties, such as folding efficiency and stability of the protein, influence the selection process.

In our study, we did not find a correlation between the amount of soluble protein produced and the stability of the molecule, as had been noted in earlier studies (Leister & Perham, 1994; Knappik & Plückthun, 1995; Nieba *et al.*, 1997). The S-P-L mutant appeared to be more stable than the P-P mutant, which, on the other hand, gives rise to less insoluble protein (Figure 8). When the two mutants S-P-L and P-P, representing the best molecules out of the kappa-like and the lambda-like libraries, respectively, competed against each other in a combined library, the more stable S-P-L was enriched over the P-P mutant by SIP selection. The longer, kappa-like sequence seems to be a better solution for the natural kappa-chain 17/9 antibody than the short lambda-like sequence. The presence of *cis*-proline at position L8 in kappa chains stabilizes the overall structure of the light domain and thus gives more stability to the whole protein, as discussed above.

Two *cis*-proline residues have been found to be highly conserved in kappa chains, the above-mentioned proline at position L8 and a second in the CDR3 at position L95. Recently, we obtained evidence that the rate-limiting step in the folding process of an scFv antibody with a kappa chain is the *trans* → *cis* isomerization of the Pro L95 peptide bond, but not that at Pro L8 (Jäger & Plückthun, 1997; Freund *et al.*, 1998). Nevertheless, a contribution of Pro L8 to the rate-limiting step can be detected in an *in vitro* folding experiment in the presence of high concentrations of peptidyl-prolyl-*cis/trans* isomerase, which efficiently catalyzes Pro L95 isomerization (Jäger & Plückthun, 1997), where an Ala L8 mutant folds slightly faster. This can explain the good *in vivo* folding behavior of the P-P (both *trans*-proline) mutant where no *cis*-proline is present, which results in only minor amounts of insoluble aggregates accumulating in the expressing cells and a prolonged bacterial growth. On the other hand, the P-P molecule is thermodynamically not as stable as the S-P-L mutant and this parameter seems to be determining the selection.

Conclusions

We conclude that in a context of lambda-like and kappa-like chain sequences SIP has selected

for the longer kappa-like chain with *cis*-proline at position L8, suggesting that for a kappa-based antibody the best solution is indeed that found in natural evolution. Using SIP, we enriched proteins according to the same criteria used by natural evolution. Thus, SIP appears to be a powerful tool for protein optimization.

Materials and Methods

Construction of the library

SIP experiments were carried out with the phage vector f17/9hag (Krebber *et al.*, 1995), encoding the anti-hemagglutinin scFv antibody fragment and the antigen fused to the N1-N2 domains. The randomization of amino acids in position L7, L8 and L9 in the antibody light chain was carried out by site-directed mutagenesis (Kunkel *et al.*, 1991) using the degenerate oligonucleotides Lib2 5′-CCAGCGGTAACGGTCAGGGASNNSN-NCTGC GTCATAACAATATCTTTGTAG-3′) and Lib3 5′-CCAGCGGTAACGGTCAGGGASNNSNNSNNCTGG-GTCATAACGATGTCTTTGTAGTC-3′ (N = A, T, C or G; S = C or G). An intermediate construct in which two stop codons replaced the three amino acid positions of interest was used as template for the mutagenesis. In the Lib3 library, before selection, 29% of scFvs were found to contain two stop codons derived from the template used for mutagenesis, which puts the efficiency of the mutagenesis at 71%.

E. coli XL1-Blue (Stratagene) (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI^q Z*ΔM15 Tn*10* (Tet*^r*)]*^c*) was electrotransformed with the mutagenesis mixtures and plated on 2 × YT agar containing 30 μg/ml chloramphenicol and 1% (w/v) glucose. All DNA manipulations were carried out using standard methods (Sambrook *et al.*, 1989).

SIP experiments

The SIP phages were produced in *E. coli* XL1-Blue cells. Cells were grown at 37°C in 2 × YT medium containing 30 μg/ml chloramphenicol and 1% glucose until an $A_{550} = 0.5$ was reached. The cells were centrifuged for ten minutes at 3500 *g* at 4°C, resuspended in 2 × YT medium containing 30 μg/ml chloramphenicol and 0.5 mM IPTG, and grown overnight at 25°C. SIP phages were enriched from the supernatant of the overnight culture by precipitation with polyethylene glycol and filtration, and quantified by spectrophotometry (Smith & Scott, 1993).

E. coli XL1-Blue cells were used as the recipient cells for the infection. Fixed amounts of phages were incubated with the cells at 37°C for one to two hours and then plated on 2 × YT agar plates containing 1% glucose and 30 μg/ml chloramphenicol for selection. Isolated clones were sequenced after each round of selection and a pool of all clones, scraped from plates, was used to produce SIP phages for the next round.

Expression and purification of scFv fragments

Selected clones were subcloned as *Xba*I-*Eco*RI cassettes into the vector pIG6 (Ge *et al.*, 1995) for soluble scFv expression. Following transformation of *E. coli* JM83 (Yanisch-Perron *et al.*, 1985), the cells were grown at 25°C in SB medium (20 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl, 50 mM K$_2$HPO$_4$) containing 100 μg/ml ampicillin and 25 μg/ml streptomycin until an $A_{550} = 0.5$ was reached. They were induced by the addition of 1 mM IPTG for three hours. After centrifugation, the cell pellets were resuspended in 20 mM Tris (pH 8.0), and passed through a French press twice. The suspension was centrifuged at 48000 *g* for 30 minutes at 4°C and the supernatant passed through a Poros20 Ni-IDA column (elution with 50 mM imidazole in 20 mM Tris/bis-propane buffer, pH 9.0) followed by a directly coupled perfusion chromatography Poros20 HQ column with a NaCl gradient (elution at about 45 mM NaCl) using a BioCAD 60 system (Perseptive Biosystems).

The purity of the protein was monitored by SDS-PAGE and gel-filtration chromatography (Superdex 75, Pharmacia SMART system) and spectrophotometry. The final protein concentration was determined from A_{280} according to Gill & von Hippel (1989).

Detection of scFv fragments in soluble and insoluble fractions of the cell lysate

The expression of scFvs was carried out as described above. After the French press treatment of the cells, soluble and insoluble fractions were normalized with respect to A_{550} of the cells and equal volumes loaded onto a reducing SDS-PAGE gel and blotted onto a PVDF membrane. The scFvs were detected with a recombinant anti-his tag scFv, fused to alkaline phosphatase (Lindner *et al.*, 1997).

Urea denaturation curves

Equal amounts of scFv fragment (final concentration 0.1 μM) were incubated overnight at 4°C with varying concentrations of urea (0 to 8.1 M) in 1.7 ml of 20 mM Tris (pH 8.0), 150 mM NaCl. The exact urea concentration in each sample was determined from its refractive index (Pace & Scholtz, 1997). Upon equilibration at 20°C, scFv unfolding was monitored by fluorescence spectroscopy ($\lambda_{ex} = 280$ nm, $\lambda_{em} = 320$ to 375 nm). The unfolding of the scFv was monitored as the wavelength shift of the emission maximum, determined by a Gaussian fit, at each urea concentration, according to Pace & Scholtz (1997).

References

Adey, N. B., Mataragnon, A. H., Rider, J. E., Carter, J. M. & Kay, B. K. (1995). Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene*, **156**, 27–31.

Altschuh, D., Vix, O., Rees, B. & Thierry, J.-C. (1992). A conformation of cyclosporin A in aqueous environment revealed by the X-ray structure of a cyclosporin-Fab complex. *Science*, **256**, 92–94.

Baca, M., Scanlan, T. S., Stephenson, R. C. & Wells, J. A. (1997). Phage display of a catalytic antibody to

optimize affinity for transition-state analog binding. *Proc. Natl Acad. Sci. USA*, **94**, 10063–10068.

Bai, C. & Elledge, S. J. (1996). Gene identification using the yeast two-hybrid system. *Methods Enzymol.* **273**, 331–347.

Boder, E. T. & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnol.* **15**, 553–557.

Braisted, A. C. & Wells, J. A. (1996). Minimizing a binding domain from protein A. *Proc. Natl Acad. Sci. USA*, **93**, 5688–5692.

Cygler, M., Rose, D. R. & Bundle, D. R. (1991). Recognition of a cell-surface oligo-saccharide of pathogenic *Salmonella* by an antibody Fab fragment. *Science*, **253**, 442–445.

Cull, M. G., Miller, J. F. & Schatz, P. J. (1992). Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc. Natl Acad. Sci. USA*, **89**, 1865–1869.

Duenas, M., Malmborg, A.-C., Casalvilla, R., Ohlin, M. & Borrebaeck, C. A. K. (1996). Selection of phage displayed antibodies based on kinetic constants. *Mol. Immunol.* **33**, 279–285.

Dunn, I. S. (1996). Phage display of proteins. *Curr. Opin. Biotechnol.* **7**, 547–553.

Freund, C., Gehrig, P., Baici, A., Holak, T. A. & Plückthun, A. (1998). Parallel pathways in the folding of short-term denatured scFv fragment of an antibody. *Folding Des.* **3**, 39–49.

Furey, W., Jr, Wang, B. C., Yoo, C. S. & Sax, M. (1983). Structure of a novel Bence-Jones protein (Rhe) fragment at 1.6 Å resolution. *J. Mol. Biol.* **167**, 661–692.

Ge, L., Knappik, A., Pack, P., Freund, C. & Plückthun, A. (1995). Expressing antibodies in *Escherichia coli*. In *Antibody Engineering* (Borrebaeck, C. A. K., ed.), 2nd edit., pp. 229–266, Oxford University Press, Oxford.

Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L. & Curtiss, R., III (1997). Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnol.* **15**, 29–34.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.

Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K. & Baker, D. (1995). A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* **4**, 1108–1117.

Hall, A. & Knowles, J. R. (1976). Directed selective pressure on a beta-lactamase to analyse molecular changes involved in development of enzyme function. *Nature*, **264**, 803–804.

Hanes, J. & Plückthun, A. (1997). *In vitro* selection and diversification of functional proteins using ribosome display. *Proc. Natl Acad. Sci. USA*, **94**, 4937–4942.

Hennecke, F., Krebber, C. & Plückthun, A. (1997). Non-repetitive single-chain Fv linkers selected by the selectively-infective phage (SIP) technology. *Protein Eng.* **11**, 405–410.

Jäger, M. & Plückthun, A. (1997). The rate-limiting steps for the folding of an antibody scFv fragment. *FEBS Letters*, **418**, 106–110.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, 5th edit., National Institutes of Health, Bethesda, MD.

Knappik, A. & Plückthun, A. (1995). Engineered turns of recombinant antibody improve its *in vivo* folding. *Protein Eng.* **8**, 81–89.

Krebber, C., Spada, S., Desplancq, D. & Plückthun, A. (1995). Co-selection of cognate antibody-antigen pairs by selectively-infective phages. *FEBS Letters*, **377**, 227–231.

Kunkel, T. A., Bebenek, K. & McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125–139.

Leister, B. & Perham, R. N. (1994). Solubilizing buried domains of proteins: a self-assembling interface domain from glutathione reductase. *Biochemistry*, **33**, 2773–2781.

Lindner, P., Bauer, K., Krebber, A., Nieba, L., Krebber, C., Honegger, A., Klinger, B., Mocikat, R. & Plückthun, A. (1997). Specific detection of his-tagged proteins with recombinant anti-his tag scFv-phosphatase or scFv-phage fusions. *BioTechniques*, **22**, 140–149.

Low, N. M., Holliger, P. H. & Winter, G. (1996). Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol.* **260**, 359–368.

Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl Acad. Sci. USA*, **91**, 9022–9026.

Moore, J. C., Jin, H. M., Kuchner, O. & Arnold, F. H. (1997). Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.* **272**, 336–347.

Nieba, L., Honegger, A., Krebber, C. & Plückthun, A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved *in vivo* folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10**, 435–444.

Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: A practical Approach* (Creighton, T. E., ed.), pp. 299–321, IRL Press, Oxford.

Pedrazzi, G., Schwesinger, F., Honegger, A., Krebber, C. & Plückthun, A. (1997). Affinity and folding properties both influence the selection of antibodies with the selectively infective phage (SIP) methodology. *FEBS Letters*, **415**, 289–293.

Poljak, R. J. (1991). Structures of antibodies and their complexes with antigens. *Mol. Immunol.* **28**, 1341–1345.

Proba, K., Wörn, A., Honegger, A. & Plückthun, A. (1998). Antibody fragments without disulfide bonds, made by molecular evolution. *J. Mol. Biol.* **275**, 245–253.

Sambrook, T., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanchez, P., Marche, P. N., LeGuern, C. & Cazenare, P. A. (1987). Structure of a third murine λ light chain variable region that is expressed in laboratory mice. *Proc. Natl Acad. Sci. USA*, **84**, 9185–9188.

Schulze-Gahmen, U., Rini, J. M. & Wilson, I. A. (1993). Detailed analysis of the free and bound conformations of an antibody. X-ray structures of Fab 17/9 and three different Fab-peptide complexes. *J. Mol. Biol.* **234**, 1098–1118.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science,* **228**, 1315–1317.

Smith, G. P. & Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol.* **211**, 127–140.

Spada, S. & Plückthun, A. (1997). Selectively infective phage (SIP) technology: a novel method for *in vivo* selection of interacting protein-ligand pairs. *Nature Med.* **3**, 694–696.

Spada, S., Krebber, C. & Plückthun, A. (1997). Selectively infective phages (SIP). *Biol. Chem.* **378**, 445–456.

Starovasnik, M. A., Braisted, A. C. & Wells, J. A. (1997). Structural mimicry of a native protein by a minimized binding domain. *Proc. Natl Acad. Sci. USA,* **94**, 10080–10085.

Stemmer, W. P. (1994). Rapid evolution of a protein *in vitro* by DNA. *Nature,* **370**, 389–391.

Strong, R. K., Campbell, R., Rose, D. R., Petsko, G. A., Sharon, J. & Margolies, M. N. (1991). Three-dimensional structure of murine anti-*p*-azo phenilarsonate Fab 36-71. 1. X-ray crystallography, site-directed mutagenesis, and modeling of the complex with hapten. *Biochemistry,* **30**, 3739–3748.

Ultsch, M. H., Somers, W., Kossiakoff, A. A. & de Vos, A. M. (1994). The crystal structure of affinity-matured human growth hormone at 2 Å resolution. *J. Mol. Biol.* **236**, 286–299.

Warbrick, E. (1997). Two's company, three's crowd: the yeast two hybrid system for mapping molecular interactions. *Structure,* **5**, 13–17.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. *Gene,* **33**, 103–119.