THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 18, Issue of April 30, pp. 18870–18877, 2004

Printed in U.S.A.

Directed *in Vitro* Evolution and Crystallographic Analysis of a Peptide-binding Single Chain Antibody Fragment (scFv) with Low Picomolar Affinity*

Received for publication, August 19, 2003, and in revised form, January 6, 2004
Published, JBC Papers in Press, January 30, 2004, DOI 10.1074/jbc.M309169200

Christian Zahnd‡, Silvia Spinelli§¶, Béatrice Luginbühl‡¶, Patrick Amstutz‡, Christian Cambillau§, and Andreas Plückthun‡‖

From the ‡Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland and §Architecture et Fonction des Macromolécules Biologiques, CNRS, 31 Chemin Joseph Aiguier, F-13402 Marseille Cedex 20, France

We generated a single chain Fv fragment of an antibody (scFv) with a binding affinity of about 5 pM to a short peptide by applying rigorous directed evolution. Starting from a high affinity peptide binder, originally obtained by ribosome display from a murine library, we generated libraries of mutants with error-prone PCR and DNA shuffling and applied off-rate selection by using ribosome display. Crystallographic analysis of the scFv in its antigen-bound and free state showed that only few mutations, which do not make direct contact to the antigen, lead to a 500-fold affinity improvement over its potential germ line precursor. These results suggest that the affinity optimization of very high affinity binders is achieved by modulating existing interactions via subtle changes in the framework rather than by introducing new contacts. Off-rate selection in combination with ribosome display can evolve binders to the low picomolar affinity range even for peptide targets.

Directed evolution of proteins *in vitro* has become a widely applied strategy to generate proteins with a desired property (1). Especially in the generation of high affinity binders, the iterative succession of randomization and selection was shown to efficiently mimic natural affinity maturation. The success of this approach is dependent on the size and the quality of the library and the power of the selection method used. Technologies that work entirely *in vitro* such as ribosome display (2) and mRNA display (3) are more favorable than methods that work partially *in vivo* such as phage display (4) or fully *in vivo* such as the yeast two-hybrid system (5) or the protein complementation assay (6) as the *in vitro* technologies do not need transformation of cells after each new round of randomization. Therefore, they allow the handling of much larger libraries and more cycles of randomization, and the experimental work is accelerated greatly.

A prerequisite of any evolution is randomization between different selection rounds. In ribosome display this is facilitated by the use of linear DNA. The randomization occurs at a low rate by the intrinsic error rate of the polymerase used but can be enhanced by error-prone PCR (7), by DNA shuffling (8), or both, thereby generating highly diverse pools.

While the generation of binders having binding constants in the subnanomolar range can be achieved, *e.g.* with synthetic antibody libraries and established techniques (9), the generation of very high affinity binders with binding constants in the lowest picomolar affinity range is difficult for several reasons. First, a very stringent selection pressure must be applied to separate improved binders from the already very high affinity precursors. Second, selected binders must be eluted efficiently, which may become very difficult for binders with very slow off-rates. Here ribosome display offers a significant advantage since bound binders must not be eluted, but the ribosomally bound mRNA can be recovered by the addition of chelating agents, which destabilize the ribosomal complex (1). In particular the generation of very high affinity peptide binders is made difficult by the relatively high flexibility of the peptide in the unbound state and the corresponding loss of entropy upon binding. This is less of a problem for comparatively rigid antigens such as hydrophobic small molecular weight compounds for which subpicomolar binders are known (10).

We applied a competitive selection for increased off-rates to affinity mature a high affinity peptide binder previously selected with ribosome display from a murine library (11). The peptide was derived from the yeast transcription factor GCN4. We constructed different mutants of a high affinity binder and generated from them second generation libraries using DNA shuffling and error-prone PCR. From these libraries we successfully isolated binders in the low picomolar affinity range. By determining the crystal structure both in the free and antigen-bound state we could show that the gain in affinity of 500-fold, compared with its likely germ line precursor, was almost exclusively a result of second sphere mutations not being in direct contact with the antigen. These findings may have great impact on future library design and affinity maturation strategies.

EXPERIMENTAL PROCEDURES

Expression and Purification of Single Chain Fv Fragments—The scFv[1] genes were cloned into the periplasmic expression vector pAK400 (12) introducing a His_6 tag, expressed in *Escherichia coli* SB536 (13),

* This work was supported by Schweizerische Nationalfonds Grant 31-65344.01. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1P4I and 1P4B) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ Both authors contributed equally to this study.

‖ To whom correspondence should be addressed. Tel.: 41-1-6355570; Fax: 41-1-6355712; E-mail: plueckthun@bioc.unizh.ch.

[1] The abbreviations used are: scFv, single chain Fv fragment; V_L, variable domain of the light chain; V_H, variable domain of the heavy chain; 8-oxo-dGTP, 8-oxo-2′-deoxyguanosine 5′-triphosphate; dPTP, 6-(deoxy-β-D-*erythro*-pentofuranosyl)-3,4-dihydro-8*H*-pyrimido-[4,-5*c*][1,2]oxazine-7-one-5′-triphosphate; RIA, radioimmunoassay; CDR, complementarity-determining region.

TABLE I
Crystal parameters and refinement statistics

	Free scFv	Complex
Data collection		
Space group	$MP2_12_12_1$	$P2_1$
Total/unique number of reflections	51,966/6,870	129,374/10,027
Percentage of data $>1\sigma$ (overall/last shell)[a]	96.0/96.0	97.2/97.6
Overall I/σ (I) (last shell)	11.9/2.6	8.5/4.1
Resolution limits	24.6/2.8	20.0/2.3
R_{merge} (%) (overall/last shell)[a]	5.9/28.1	6.2/18.4
Refinement		
Number of protein/solvent atoms	1,658	17,648/66
Number of reflections	6,796	9,940
Resolution limits (Å)	12.0–2.8	15.0–2.35
R/R_{free} value (%)	20.05/26.4	18.0/21.5
r.m.s.d.[b] on bonds (Å)/angles (°)	0.012/2.0	0.025/2.1
r.m.s.d. on impropers/dihedrals (°)	1.14/27.8	0.86/25.2
Mean B-factor (Å^2) (main/side chain)	35.4	23.8/27.0
Peptide (main/side chain)		25.0/28.0
Water		33.0

[a] Last shell, 2.9–2.8 Å (free scFv); 2.43–2.35 Å (complex).
[b] r.m.s.d., root mean square deviation.

and purified by immobilized metal ion affinity chromatography and subsequent antigen affinity chromatography as described previously (11). For structure determination, a seleno-Met-containing variant of clone H6 was grown in 1 liter of M9 minimal medium to an OD_{600} of 0.6. An amino acid mixture containing 100 mg/liter each of Lys, Thr, and Phe and 50 mg/liter each of Leu, Ile, Val, and seleno-Met was added. After 1 h, isopropyl-1-thio-β-D-galactopyranoside was added to a final concentration of 1 mM for expression overnight. The protein was purified as described above.

The Library Construction—The scFv fragments C11L34, L24, L107, L135, L107–135, H6, and H67 in the vector pAK400 were PCR-amplified using primers SDAla+ (5′-AGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATCCATGGCGGACTACAAAGAT) and Sfi_rescue (5′-GCCCTCGGCCCCCGAGGC). A total of 1 μg of PCR product of an equimolar mixture of all clones was used for DNase I shuffling (14) as described previously (15). Some of the reassembled PCR products were further randomized by error-prone PCR using primers SDAla+ and Sfi_rescue. Error-prone PCR was performed using the dNTP analogues 8-oxo-dGTP and dPTP according to Ref. 7 with small modifications. Twenty-six cycles of error-prone PCR were performed in the presence of 85 μM dPTP, 85 μM 8-oxo-dGTP, and 50 μM dGTP, dATP, dTTP, and dCTP each. The final mutation rate after DNase I shuffling and error-prone PCR and the distribution of the mutations were determined by sequencing about 2000 bp. A gene III linker was fused to the library as described earlier (11).

Selection for Improved Affinities—The library was transcribed and translated *in vitro* as described in Ref. 11. The ternary complexes of ribosome, mRNA, and displayed proteins were equilibrated with 1 nM biotinylated peptide GCN4(7P14P) (16) at 4 °C overnight. Every sample was split into two aliquots, and only to one, 1 μM non-biotinylated GCN4(7P14P) was added. The aliquots were incubated for a defined time span, which was increased from round to round, starting with 2 h in the first round and going up to 10 days after the fourth round in a rollover shaker at 4 °C. The complexes were recovered by binding to streptavidin-coated magnetic beads (Roche Applied Science) for 30 min. The beads were washed five times, and the RNA was eluted and purified as described in Ref. 11.

Affinity Comparison of Pools and Single Clones by Radioimmunoassay (RIA) and Inhibition BIACORE—For the analysis of single clones, the selected pools were cloned into pTFT74 (17). RNA of single clones was transcribed directly from plasmid pTFT74, whereas the pools were transcribed from a PCR product. RIAs were performed as described previously (18). Of some clones, absolute affinities were measured using the inhibition method on a BIACORE 3000 (19). The purified protein was diluted to 1 nM and incubated with different concentrations of antigen overnight at 4 °C. The samples were injected over a chip, which was coated to maximal density with the antigen used for selection. The slope of the association curves in the linear phase was plotted against the concentration of soluble antigen. K_D was determined from at least three independent curves as described previously (11).

Crystallization—Crystals of clone H6 in complex with the antigen and clone C11L34 in the absence of the peptide were obtained using the hanging drop method. Drops of 1 μl (protein concentration, 8 mg/ml) of the unliganded Fv GCN4 were mixed with 1 μl of the well solution (1.1 M ammonium sulfate, 150 mM sodium citrate, pH 4.8). Crystals appeared after 2 days at 20 °C belonging to space group $P2_12_12_1$ (a = 35.08 Å, b = 60.53 Å, and c = 123.05 Å) and contained one molecule per asymmetric unit (V_m = 2.29 Å^3/Da, 46% solvent (20)). Crystals of the complex of H6 and the peptide YHLENEVARLKK were obtained by mixing the scFv and the peptide in a 1:2 molar ratio. Drops of 1 μl of the complex (protein concentration, 7.4 mg/ml) were mixed with 1 μl of a solution containing 32–28% (w/v) polyethylene glycol 4000, 0.1 M Tris/HCl, pH 7.5. The space group was $P2_1$ (a = 37.24 Å, b = 36.29 Å, c = 84.46 Å, β = 90.5°) with one complex per asymmetric unit and a specific protein volume of 2.03 Å^3/Da.

Data Collection and Processing—Data of the free scFv fragment were collected at 100 K on a MAR-Research Imaging Plate (MAR, Hamburg, Germany) placed on a Rigaku RU200 rotating anode using the CuK_α radiation. Crystals were frozen using 25% glycerol, and they diffracted to 2.6 Å. Data integration and reduction were performed using DENZO (21) and SCALA (22). A crystal of the complex was frozen with 16% glycerol. Data were collected at 100 K in the beam-line ID14-EH4 at the European Synchrotron Radiation Facility (Grenoble, France), λ = 0.988 Å. The data reduction was performed with SCALA (22). The data of the complex were merged with TRUNCATE (22) in CCP4i using the anisotropic correction option. The R_{Sym} for the free scFv and the complex were 5.9 and 6.2%, respectively. The statistics of the data sets are shown in Table I.

Structure Determination and Refinement—Both structures were solved by molecular replacement (23) using the program AMoRe (24). Rotation and translation searches were performed using the murine Fv SE155-4 (Protein Data Bank entry 1MFA) for the unliganded Fv. The structure of the unliganded Fv was used to solve the structure of the complex. The molecular replacement solution for the unliganded Fv gave an initial correlation coefficient of 45% and R-factor of 47%, while the values for the complex were 46 and 48%, respectively, between 8.0 and 3.0 Å. Refinement was performed first with CNS (25) using standard protocols followed by REFMAC (26) using maximum likelihood, incorporating bulk solvent corrections and translation-libration-screw (TLS) anisotropic refinement. After each refinement cycle, a new map was calculated, and the model was fitted with Turbo-Frodo (27). Final refinement data are summarized in Table I. The coordinates and structure factors of the free and peptide-bound Fv have been deposited in the Protein Data Bank at Research Collaboratory for Structural Bioinformatics as entries 1P4I and 1P4B, respectively.

RESULTS

Library Construction—In a previous study, we selected a group of scFv fragments from a murine library by using ribosome display, a selection method that works entirely *in vitro* and therefore allows the selection from very large libraries (Fig. 1). The selected scFv fragments bound with very high affinity to the peptide GCN4(7P14P) derived from the yeast transcription factor GCN4 (11, 16). The highest affinity clone, named "C11L34," had an affinity of 40 pM and acquired a crucial amino acid substitution during ribosome display rounds

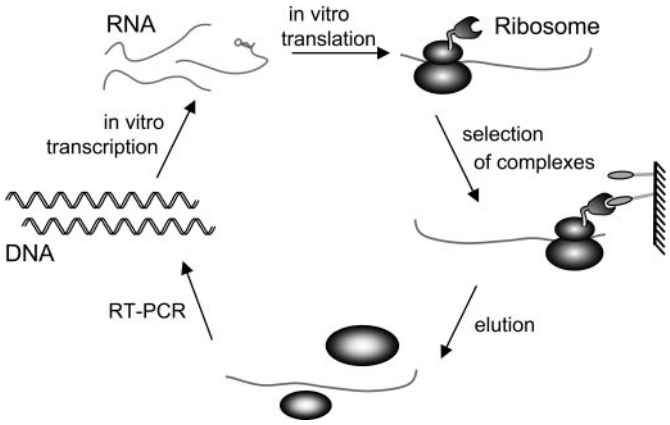

FIG. 1. **Ribosome display.** Ribosome display is a method to select binders from large libraries. A library of mRNA molecules, encoding for potential binders, is translated with an approximately stoichiometric amount of ribosomes *in vitro*. Since the mRNA contains no stop codon and the reaction is stopped with a buffer containing high concentrations of magnesium, the ribosome will not dissociate and form a stable complex with the translated protein and the mRNA. A spacer fused to the C terminus of the library takes up the place in the ribosomal tunnel and thereby ensures that the freshly translated protein can fold into its native structure even if it is displayed on the ribosome. The complexes are used for the selection of binders to immobilized antigen. After washing, the mRNA is eluted using chelating agents such as EDTA to dissociate the ribosome. Eluted mRNA is reverse transcribed, amplified by PCR, and potentially randomized to introduce more diversity. mRNA for the next selection round is transcribed directly from a PCR product. Note that at no step does the library have to be transformed into cells, which would otherwise drastically reduce the size of the library that is available for screening. The PCR can also be carried out under error-prone conditions thereby allowing the library to evolve (for details, see text). *RT*, reverse transcription.

that led to a 65-fold affinity improvement compared with its likely progenitor from the murine B-cells (11).

We constructed a second generation library in two steps. First, we generated five different point mutants of the high affinity clone and a double mutant, based on mutations that had been enriched in the first selection experiment, and measured their affinities (Table II and see Fig. 4). These mutants were taken as a template for the construction of three second generation libraries. The first pool, named "S," was generated by using DNA shuffling of all seven clones including C11L34, thereby recombining all mutations (8). Two other pools were generated by the use of error-prone PCR. Pool "R" was generated by error-prone amplification of C11L34, and pool "SR" was generated by the combination of the two approaches: the mutants were amplified by error-prone PCR and then subjected to DNA shuffling.

To enhance the mutation rate of the polymerases, we used high concentrations of the nucleotide analogues 8-oxo-dGTP and dPTP leading to mismatch incorporation (7). The final mutation rates of the pools were determined as 8.9 kbp^{-1} for pool S, 61 kbp^{-1} for pool R, and 78 kbp^{-1} for pool SR. Taking into account that many mutations are silent or have no effect on binding, *e.g.* because they are located in the linker or the lower framework, we estimated that the randomized pools will have two to three relevant amino acid substitutions per gene. As a consequence of the PCR-based randomization we found that mutations that were introduced during an early PCR cycle were more prominent in the final library. This clustering of mutations could be circumvented in future experiments by the use of a high template concentration and very high error rates, thereby reaching the final mutation rate within fewer amplification cycles.

Ribosome Display and Off-rate Selection—For ribosome display, all three libraries had to be fused to a protein spacer derived from gene III to allow the displayed protein to fold properly (15). Since panning followed by extensive washing would hardly be able to discriminate between binders with different affinities all lying in the picomolar range, we used a competitive selection for decreased off-rates. The ribosomal complexes formed after *in vitro* translation were equilibrated overnight with a 1 nM solution of biotinylated antigen. A 1000-fold excess of competitor antigen carrying no biotin label was then added, and the pools were incubated at 4 °C. All complexes dissociating from the biotinylated antigen, to which they were initially bound, will be captured by the competitor carrying no biotin label and thus cannot be bound to streptavidin-coated magnetic beads. Hence the duration of incubation with competitor is defining the stringency of the selection. The incubation was prolonged from round to round (2, 10, and 240 h), thereby increasing the selection pressure. After every round, selected mRNA was isolated and reverse transcribed, and the enriched pools were subjected to DNA shuffling. The pool SR was further randomized after the second round using the same conditions as in round 1.

Interestingly the mRNA of pool R, which had been generated by error-prone amplification of C11L34 only but which had not been subjected to DNA shuffling, could not be restored after the first round of selection. This indicates the importance of recombination in conjunction with high mutation rates to preserve a fraction of the pool in an active form. Furthermore the long off-rate selection times underscore the stability of the non-covalent ribosomal complex, which can survive more than 20 days at 4 °C.

Analysis of the Pools after Off-rate Selection—After every round of ribosome display, the pools were checked for improved binding by RIA (18). The pools were expressed *in vitro* in the presence of [^{35}S]Met, equilibrated with different amounts of free antigen, and allowed to bind to surface-immobilized antigen. The amount of competitor antigen needed to inhibit the binding of the scFv fragments to surface-immobilized antigen correlates with the mean affinity of binders found in the pool under investigation and decreased from round to round (Fig. 2). In the initial error-prone randomized pool even high concentrations of competitor did not affect binding of the pool to the plate. After 240 h of off-rate selection, however, 0.1 nM antigen was sufficient to prevent 50% of the pool from binding to the surface compared with the uninhibited signal, giving evidence that the mean affinity of the binders in the pool had improved affinities compared with clone C11L34. The total signal intensity of the pools decreased from round to round, indicating that the percentage of rescued binders decreased from round to round due to the very stringent selection pressure.

Screening for Binders—After the third round, pools SR and S were cloned, plasmids of single colonies were isolated, and *in vitro* expressed protein was analyzed by RIA. Only 7 of 54 clones (13%) showed a significant binding signal to the surface-immobilized antigen in the absence of competitor. All of these clones could be completely inhibited with 10 nM antigen. Four of them were inhibited at even lower concentrations than C11L34 and were therefore ranked as affinity improved. In pool S, which was generated by DNA shuffling, 4 of 15 (25%) clones analyzed showed binding to the antigen after the third round of directed evolution, but only one had an improved inhibition signal (Table II).

Only a few molecules had reached improved affinities that let them survive the applied selection pressure, whereas the background signal, consisting of unspecific complexes and RNA sticking to the streptavidin-coated magnetic beads, remained constant. To improve the ratio of binders over background, a non-stringent enrichment round was performed. After this

TABLE II
Summary of the mutations and affinities of different GCN4 binders

Clone[a]	Mutations[b]		K_D[c]
			pM
C11L34[d]	L42(Asn → Ser)		40 ± 4
L24	L24(Arg → His)		32 ± 13
L107	L107(Ala → Val)		48 ± 6
L135	L135(Asn → Asp)		23 ± 4
H6	H6(Glu → Gln)		20 ± 2
H67	H67(Ile → Val)		47 ± 7
L107L135	L107(Ala → Val)	L135(Asn → Asp)	47 ± 5
L135H6	L135(Asn → Asp)	H6(Glu → Gln)	16 ± 13
52SR4	L13(Thr → Ser) L135(Asn → Asp)	L107(Ala → Val) H30(Ser → Leu)	5.2 ± 2.3
63S3	L145(Leu → Pro)		ND[e]
70SR4	L13(Thr → Ala) H96(Leu → Pro)	L135(Asn → Asp)	ND
82SR4	H70(Tyr → His)	H94(Asn → Ser)	ND
84SR4	H32(Thr → Ala) H94(Asn → Ser)	H70(Tyr → His)	ND

[a] The first eight clones listed were found in a previous selection (11), and their mutations compared with C11L34 were introduced by site-directed mutagenesis. Clones H6 and L135H6 were constructed because it was shown recently that the residue on position H6 is often critical for stability and affinity (32, 34). All these clones were used for the generation of different libraries by DNA shuffling and error-prone PCR with the exception of clone L112H6, which was constructed at a later point in the project. The last five clones listed were found during the directed evolution and showed promising signals in a competition RIA.

[b] The mutations are given in the AHo nomenclature (35).

[c] The affinities were determined by inhibition BIACORE. The affinity improvement during the directed evolution was monitored mainly by competition RIA (see text).

[d] Clone C11L34 was found to have a single point mutation compared with its likely progenitor (11). All other clones are derived from C11L34 and contain its mutation, L42(Asn → Ser). For historical reasons, its name is indicated in the Kabat nomenclature (36) and was retained.

[e] ND, not determined.

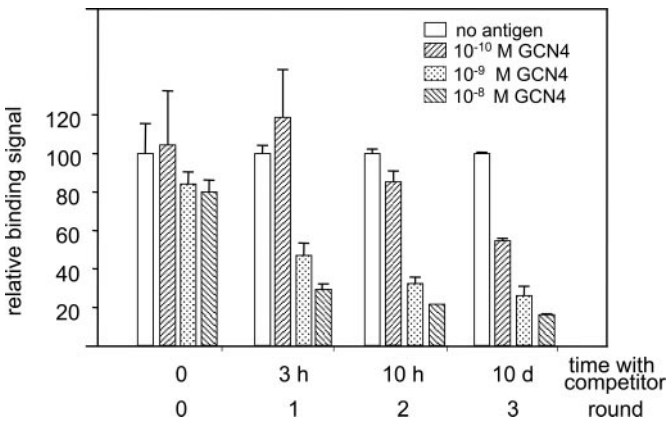

FIG. 2. **Inhibition RIA of pools after subsequent rounds of off-rate selection.** Three rounds of ribosome display were performed using selection for improved off-rates by competing biotinylated peptide with unlabeled peptide (see text). After each round, ^{35}S-labeled protein of the pools was expressed *in vitro* and equilibrated with different amounts of antigen. The equilibrated pools were then bound to surface-immobilized antigen. The binding signal of the pool is a measure of the amount of free scFv fragment present at equilibrium. High affinity binders are completely complexed by small concentrations of free antigen and thus do not bind to the immobilized antigen. The less free antigen is needed to prevent a pool from binding to the surface-immobilized antigen, the higher the mean affinity of the pool is. It can be seen that the concentration of free antigen needed to prevent the pool from binding to surface-immobilized antigen decreases from round to round. *d*, days.

enrichment, 14 of 16 (87%) randomly picked clones showed binding to the antigen, of which eight showed improved inhibition patterns compared with C11L34. Furthermore the signal intensity of the pool in the RIA increased by a factor of 120, indicating a strong enrichment of the binders in the pool over the nonspecific RNA. Thus, a non-selective enrichment round after extensive off-rate selection may be useful in general to amplify the selected clones.

Sequences of the Clones with Improved RIA Signal—The clones showing the most promising RIA signal were sequenced. They carried an average mutational load of one to four amino acid substitutions, whereas zero to two mutations derived from shuffled input DNA (Table II). The mutations were distributed over both domains, and some mutations showed up several times. It is likely that they were found independently since they had different codon usage. Interestingly the only mutation lying in CDRs, L135(Asn → Asp), originated from the clones used for DNA shuffling. All other mutations were located in framework positions.

BIACORE Measurements—The affinity of all clones generated by site-directed mutagenesis and used for the library construction was measured. In addition, the binding constant of the evolved clone showing the best RIA signal after off-rate selection was determined. All clones were expressed in the periplasm of *E. coli* and purified by immobilized metal ion affinity chromatography and antigen affinity chromatography (18). The dissociation constant K_D of the purified proteins was determined in solution by competition BIACORE analysis (19, 28, 29).

The dissociation constants of the clones used for the library generation were between 20 and 50 pM (Table II). Clones L135(Asn → Asp) and H6(Glu → Gln) showed a significantly improved affinity of 23.3 and 20 pM, respectively. Therefore, we also constructed the double mutant L135H6. It had an affinity of 16 pM. However, this clone had not been available when the library was constructed. The affinities of the evolved clones were mainly monitored by RIA. The affinity of the clone with the best RIA signal, named 52SR4, was determined. It was found in pool SR after the third round, and its affinity in solution was determined to be 5.2 pM by the competitive BIACORE method (11). All measurements were repeated three times independently, resulting in uncertainties of 5–45%. Measuring affinities in this range is very difficult. The minimal concentration of the scFv that could be detected using BIACORE was about 1 nM. Due to the very high affinities, this is much above the K_D value at which concentration the measurement should have been performed ideally. However, even if the K_D of this highest affinity binder cannot be given with satisfy-

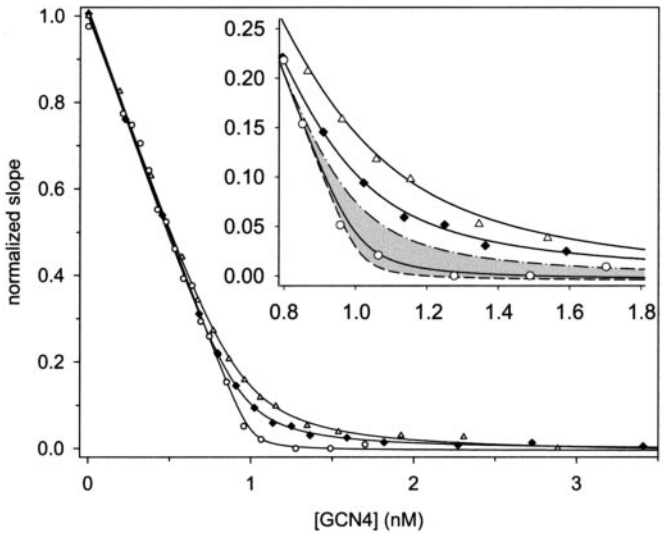

FIG. 3. **Affinity determination using inhibition BIACORE.** The dissociation constant K_D in solution of different single chain Fv fragments using inhibition BIACORE was determined at 20 °C. The *triangles* correspond to the starting clone C11L34, having a K_D of 40 pM, the *diamonds* correspond to clone H6 from which the crystal structure in complex with the antigen was determined to have a K_D of 20 pM, and the *circles* correspond to 52SR4, which was the evolved clone with the best RIA signal having a K_D of 5.2 pM. For better comparison, the theoretical curves for a 10 pM affinity (*dash-dotted line*) and a 1 pM affinity (*dashed line*) are indicated in the *inset*. Due to the very high affinities the measurements become close to the detection limit. The given value of 5.2 ± 2.3 pM was obtained in three independent measurements.

ing accuracy (5.2 ± 2.3 pM, determined from several independent protein preparations), a clear improvement of the scFv over the initial constructs can be observed (Fig. 3). The K_D values of C11L34 and 53SR4 were also determined by kinetic BIACORE analysis with very low coating density, and the equilibrium K_D data were confirmed (data not shown).

The Overall Structure of the scFv—The crystal structure of C11L34 was determined in the absence of the antigen to 2.35 Å, and the crystal structure of variant H6 in the presence of the antigen was solved to 2.8-Å resolution (Fig. 4*A*). The C-terminal His tag and the linker connecting the two variable domains were not defined in the electron density as is typical for other scFv fragments. The N-terminal part of V_H of the scFv in complex with the peptide was poorly defined. Initial attempts to crystallize the complex with the 33-amino acid peptide used for selection were not successful. Previous studies with CD had suggested that the peptide would adopt a random coil conformation (16) in the antibody-bound form, which implied that not more than 10–12 amino acids of the extended peptide would be recognized by the scFv. Therefore, we used a truncated peptide of 12 amino acids for a second crystallization trial. With this approach crystals were obtained. When the structure was solved, the peptide was found to bind in a three-turn helical conformation, thereby making its total contact area smaller than expected. Therefore, not all of the potentially interacting residues can be seen in the present structure. To better illustrate the complete potential binding interface, the peptide was elongated at its N terminus, and the corresponding residues, which we denote, M1 to M5, in contrast to the structurally resolved residues of the peptide, P1 to P12, were modeled in an α-helical conformation into the complex. We repeated the CD measurements, which led to the original assumption that the bound peptide would adopt a random coil conformation (16). From the spectra of the scFv alone compared with the scFv bound to the peptide we attempted to deduce the spectrum of the peptide when bound to the scFv. Whereas the peptide in solution was clearly random coil, the spectrum of the peptide in complex with the scFv, originally interpreted as random coil, was very noisy and did not allow a clear conclusion (data not shown). The antibody had originally been raised against a variant of the coiled-coil leucine zipper of the yeast transcription factor GCN4 in which two residues were changed to proline (16). The origin of the peptide explains that there is some residual tendency for α-helix formation. Indeed the almost perfect helical conformation of the peptide in complex with the scFv suggests that the helical conformation is further induced by binding to the antibody.

Conformational Changes Due to Antigen Binding—Superposition of the Fv structure in the antigen-bound and free state results in a root mean square deviation of only 0.75 Å. Deviations are located between residues H9–18 and H48–52, and they are mostly due to the poor definition of the N-terminal part of V_H in the structure of the complex. These changes located opposite to the binding site are probably not related to the complex formation. A major rearrangement, however, was seen in CDR H3. Specifically residues Gly-H109 and Leu-H110 deviate strongly by 3.4 and 1.4 Å, respectively. Since this segment is well defined in both electron density maps, it is clear that the conformational change is directly induced by the binding of the peptide to the scFv, namely by Arg-P9 (Fig. 4*B*). Beside these changes, only small structural changes in the CDRs are observed upon peptide binding, which are all less than 1 Å. Upon peptide binding, the two domains of the scFv undergo a small rotation relative to each other by 1.25°. This may allow a more optimized binding geometry (see below).

The Interaction of the scFv with the Antigen—The scFv forms a deep (6–8 Å) and broad (8 Å) cleft of 20 Å in length to which all six CDR sequences contribute. The bottom of the cleft is formed mainly by the relatively short CDR3 loops of the heavy and light chain. The borders are formed by the loops of CDR1 and CDR2 of both chains. The antigen lies as an almost ideal three-turn α-helix in the binding cleft. This engulfed binding of the antigen results in a buried surface of 700 Å^2, which corresponds to 45% of the total water-accessible surface of the peptide. Modeling the peptide in its full length into the binding groove by extending the helix yields in an interaction interface of about 1580 Å^2, which is bigger than the interface between the light chain and the heavy chain (1479 Å^2). This buried surface value is also larger than those observed between most Fab fragments/Fv fragments in complex with proteins (30).

The interaction is made up by a great number of specific contacts between the scFv and exclusively the side chains of the peptide. The helical peptide faces the scFv with the same side that is responsible for the leucine zipper formation in the natural protein (31). Almost 30% of the surface of the peptide is contributed by His-P2. Its aromatic plane forms a hydrophobic stacking interaction with the indole moiety of Trp-H59. His-P2 also establishes a hydrogen bond to Asp-H65 and is further stabilized by a hydrophobic contact to Ile-H67. The neighboring residue Leu-P3 is held between three aromatic residues (Tyr-L40, Trp-L109, and Trp-L137), and Arg-P9 on the adjacent helical turn makes a strong (double) ionic interaction with Asp-H137. The interaction with Arg-P9 enforces the conformation of CDR H3 (H109–137), especially that of Gly-H109. The conformation of Arg-P9 is most likely locked by a hydrogen bond to Glu-P6 (Fig. 4*B*). In addition, Glu-P6 establishes a contact to the backbone carboxyl oxygen of Gly-H40 via a water molecule. An ionic hydrogen bond is also made between Lys-P4 and Asp-H69. A great number of hydrophobic interactions, namely stacking of aromatic rings and packing of aliphatic side chains, are found. The pattern of charged and hydrophobic residues observed on the surface of the peptide is comple-

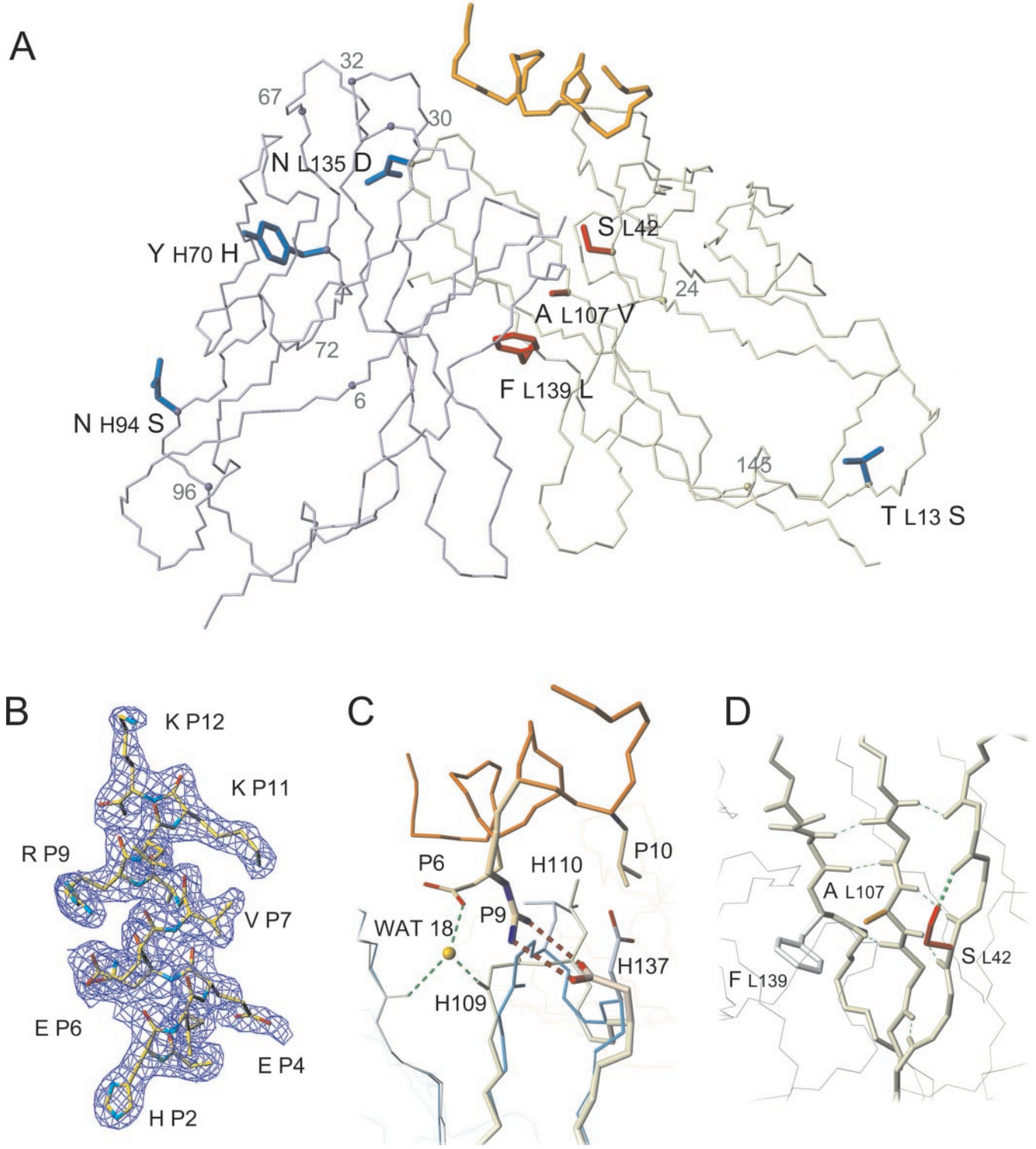

FIG. 4. **Overall structure of clone H6 in complex with the peptide and the location of the mutations.** In *A*, the backbone of the complex is shown. All mutations, which were found more than once in the directed evolution in any clone, are drawn with side chains including mutation L42(Asn → Ser), which led to clone C11L34 with 65-fold improved affinity compared with its likely murine progenitor. The Cα atoms of other mutations mentioned in the text are indicated by a *sphere* and a *number*. The light chain is drawn in *olive green*, the heavy chain is drawn in *light blue*, and the peptide is drawn in *orange*. The side chains of mutations lying in the interface of the two domains are shown in *red*. *B*, $2F_o - F_c$ electron density map contoured around the dodecapeptide. *C* shows the CDR H3 region overlaid in the antigen-bound and free state where the structural changes in the scFv due to antigen binding are significant. The antigen-bound form of the scFv fragment is drawn in *light orange*, and the free form is in *blue*, whereas the peptide is drawn in *orange*. Not all side chains are drawn. Arg-P9 establishes a strong interaction including two charged hydrogen bonds with Asp-H137 and thereby induces the conformational shift. The conformation of Arg-P9 may be locked by another weak hydrogen bond to Glu-P6. The greatest conformational change is undergone by Gly-H109, but there is a slight reorientation also by Leu-H110, which interacts hydrophobically with Leu-P10. The shift in the loop is stabilized by water molecule WAT18, which is hydrogen bonding to Glu-P6, Gly-H109, and Gly-H40. *D*, three mutations were lying close together in adjacent β-strands of the light chain facing the heavy chain. They may have an effect on the domain orientation and domain spacing. Residue Ser-L42 found in the previous selection is drawn in *red*. Ala-L107 (*orange*) and Phe-L139 (*gray*) were mutated to valine and leucine, respectively. The figures were prepared using the program MolMol (33).

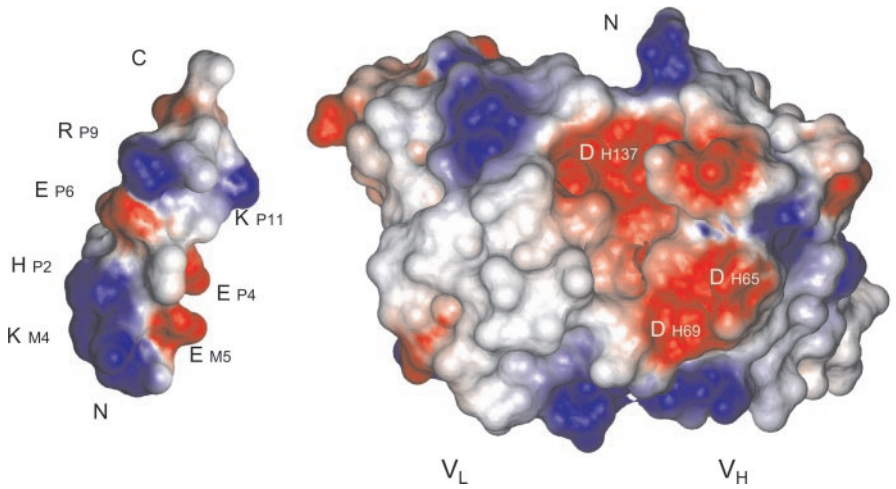

FIG. 5. **Electrostatic potential of the antibody with the antigen moved from the binding site.** A surface plot of the scFv and the antigen covering all amino acids potentially covering the binding pocket is shown. The six N-terminal amino acids DLPKQY of the peptide were modeled into the binding groove in a helical conformation. For better visualization, the antigen was removed from the scFv and rotated such that the binding face is visible. Folding the figure in the middle would restore the original orientation of the peptide.

mented electrostatically by the surface of the scFv (Fig. 5). Water molecules contribute strongly to the stabilization of the conformation of the peptide and are directly involved in interactions between the peptide and the antibody fragment. The conformational change in CDR H3 is probably favored by water molecule WAT18 (Fig. 4*B*).

A Structural Interpretation of the Evolution—Almost all the mutations that produced the high affinity binder are so-called "second sphere" mutations, which do not directly interact with the antigen and fulfill their beneficial effects via indirect interactions. The mutation L42(Asn → Ser), which leads to a 65-fold improved affinity of C11L34 relative to all its precursors and had been found in the original selection with ribosome display, is at least 6.6 Å away from the antigen. The light chain of C11L34 is of subtype λ. Interestingly murine germ line sequences of this subtype never carry a Ser at this position, whereas in murine light chains of κ-type, Asn and Ser side chains are common. In murine κ-type light chains, however, Asn side chains often make a hydrogen bond to residue H110 of V_H, stabilizing the interface, while Ser is rather seen in hapten binders, having a deep binding pocket, where the hydrogen bond is made directly to the hapten (*e.g.* Protein Data Bank entry 1C5C). In the case of C11L34, Ser-L42 plays an unusual role and forms a hydrogen bond to the backbone oxygen of Tyr-L40, stabilizing the conformation, which is typical for a backbone of λ-subtype. This stabilizing effect most likely reduces the flexibility of CDR L1, which is directly interacting with the peptide, through a hydrophobic contact between Tyr-L40 and Leu-P3. In addition, the space gained in the interface due to the smaller side chain may allow a more favorable domain orientation. A comparison of the scFv in the antigen-bound and unbound state reveals a domain rotation upon antigen binding and a reorientation of the CDR H3 loop, which brings CDR H3 1.5 Å closer to Ser-L42 than in the unbound state. In addition to Ser-L42, two other mutations, L107(Ala → Val) and L139(Phe → Leu), that showed up during the extensive off-rate selection lie in close proximity to Ser-L42 on adjacent β-strands of the light chain (Table II and Fig. 4*C*). These mutations are close to the pseudo-twofold axis of the scFv. It is likely that the interface mutations influence the relative domain orientation or domain spacing and thereby optimize the binding geometry.

The mutation H6(Glu → Gln) improved the affinity, by a factor of 2, to 20 pM when compared with clone C11L34. This residue is much too far away to interact with the antigen directly and must exert its beneficial effect by some long range interactions or "molecular shimming," influencing the orientation or flexibility of a loop or a domain. While the free scFv carries a glutamate at position H6, the structure of the scFv in complex with the peptide solved here is a variant carrying glutamine at this position. Position H6 was shown to define the conformation of the N-terminal part of the heavy chain (32). The conformation of the backbone in the free state is as it was expected for a glutamate. In the complex, however, this part of the structure is poorly resolved. Therefore, the occupancy of this area was set to 0. The poor density might indicate a conformational inhomogeneity or a higher than normal mobility, the reasons of which are unknown. The only mutation found during the selection that may interact directly with the antigen was introduced by site-directed mutagenesis during the library construction: L135(Asn → Asp). It most likely establishes a hydrogen bond to Lys-M4. However, this part of the peptide was not used for crystallization (see above). Most other mutations found during directed evolution are lying on the surface of the scFv fragment, and many may be neutral. However, some could also have indirect beneficial effects such as H30(Ser → Leu), which was found in the best affinity clone.

DISCUSSION

We have evolved a pool of single chain Fv antibody fragments against a peptide, which adopts a random coil in solution but a helical conformation in the complex, to achieve affinities of about 5 pM. We started from an already very tight peptide binder and generated a library by combining site-directed mutagenesis, DNA shuffling, and error-prone PCR. We applied directed evolution using ribosome display to improve the affinity a further 8-fold. Our result demonstrates that ribosome display is ideally suited for the identification of affinity-improving mutations and for the selection of binders under very stringent conditions even when already starting from picomolar affinities. In contradiction to previous concerns the stability of the ribosomal complex and the attached mRNA is not a limitation of the method.

One of the key results of the present study, analyzing the results from directed evolution crystallographically, is that only one mutation was making direct contact to the peptide. Three mutations were lying close together in the V_L/V_H interface. They may modulate the domain orientation, domain spacing, and the CDR loop flexibility. Many of the mutations found in the affinity-improved clones were found to lie on the surface

of the scFv not making direct contact to the antigen. Most likely these accumulated during the library generation and do not contribute to the improved affinities.

After the selection for a "first sphere" containing all important short range interactions, the overall geometry of the binding pocket was rearranged in a subtle manner. This was exclusively achieved by the mutation of key residues in the so-called second sphere, influencing the flexibility of binding loops and the orientation of domains rather than by changing interacting residues.

To our knowledge we have evolved the highest affinity reagent against a short unmodified peptide. Due to its very high affinity and the relatively small size of the antigen we expect the evolved scFv-peptide pair to be a powerful tool for biotechnological applications where tight binding to a tag is needed.

Acknowledgments—We thank Dr. Annemarie Honegger, Dr. Jozef Hanes, and Dr. Lutz Jermutus for helpful discussions and advice. We also thank Dr. David Zechel for critical reading of the manuscript.

REFERENCES

1. Amstutz, P., Forrer, P., Zahnd, C., and Plückthun, A. (2001) *Curr. Opin. Biotechnol.* **12,** 400–405
2. Hanes, J., and Plückthun, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4937–4942
3. Roberts, R. W., and Szostak, J. W. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 12297–12302
4. Smith, G. P. (1988) *Virology* **167,** 156–165
5. Fields, S., and Song, O. (1989) *Nature* **340,** 245–246
6. Mössner, E., Koch, H., and Plückthun, A. (2001) *J. Mol. Biol.* **308,** 115–122
7. Zaccolo, M., and Gherardi, E. (1999) *J. Mol. Biol.* **285,** 775–783
8. Stemmer, W. P. (1994) *Nature* **370,** 389–391
9. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A., and Virnekäs, B. (2000) *J. Mol. Biol.* **296,** 57–86
10. Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10701–10705
11. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., and Plückthun, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14130–14135
12. Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R., and Plückthun, A. (1997) *J. Immunol. Methods* **201,** 35–55
13. Bass, S., Gu, Q., and Christen, A. (1996) *J. Bacteriol.* **178,** 1154–1161
14. Adey, N. B., Stemmer, W. P. C., Kay, B. K. (1996) in *Phage Display of Peptides and Proteins* (Kay, B. K., Winter, J., and McCafferty, J., eds) pp. 280–292, Academic Press, Cambridge
15. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., and Plückthun, A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 75–80
16. Berger, C., Weber-Bornhauser, S., Eggenberger, J., Hanes, J., Plückthun, A., and Bosshard, H. R. (1999) *FEBS Lett.* **450,** 149–153
17. Ge, L., Knappik, A., Pack, P., Freund, C., and Plückthun, A. (1995) in *Antibody Engineering* (Borrebaeck, C. A. K., ed) pp. 229–266, Oxford University Press, New York
18. Auf der Maur, A., Zahnd, C., Fischer, F., Spinelli, S., Honegger, A., Cambillau, C., Escher, D., Plückthun, A., and Barberis, A. (2002) *J. Biol. Chem.* **277,** 45075–45085
19. Nieba, L., Krebber, A., and Plückthun, A. (1996) *Anal. Biochem.* **234,** 155–165
20. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497
21. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W., Jr., and Sweet, R. M., eds) Vol. 276, pp. 307–326, Academic Press, New York
22. Collaborative Computational Project No. 4 (CCP4) (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
23. Rossman, M. G., and Blow, D. M. (1962) *Acta Crystallogr.* **15,** 24–31
24. Navaza, J. (1992) in *Molecular Replacement: Proceedings of the CCP4 Study Weekend* (Dodson, E. J., Gover, S., and Wolf, W., eds) pp. 87–90, Science and Engineering Research Council (SERC) Daresbury Laboratory, Warington, UK
25. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
26. Murshudov, G. N., Alexei, A. V., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
27. Roussel, A., and Cambillau, C. (1991) *Silicon Graphics Geometry Partners Directory*, pp. 81, Silicon Graphics Corp., Mountain View, CA
28. Karlsson, R. (1994) *Anal. Biochem.* **221,** 142–151
29. Schuck, P. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26,** 541–566
30. MacCallum, R. M., Martin, A. C., and Thornton, J. M. (1996) *J. Mol. Biol.* **262,** 732–745
31. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science* **254,** 539–544
32. Honegger, A., and Plückthun, A. (2001) *J. Mol. Biol.* **309,** 687–699
33. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14,** 51–55, 29–32
34. Jung, S., Spinelli, S., Schimmele, B., Honegger, A., Pugliese, L., Cambillau, C., and Plückthun, A. (2001) *J. Mol. Biol.* **309,** 701–716
35. Honegger, A., and Plückthun, A. (2001) *J. Mol. Biol.* **309,** 657–670
36. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S., and Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, National Institutes of Health Publication No. 91-3242, 5th ed., United States Department of Health and Human Services, Bethesda, MD