Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity

Eric T. Boder*, Katarina S. Midelfort†, and K. Dane Wittrup**†‡

Department of *Chemical Engineering, and †Biophysics, University of Illinois, Urbana, IL 61801

Communicated by Herman N. Eisen, Massachusetts Institute of Technology, Cambridge, MA, June 27, 2000 (received for review May 16, 2000)

Single-chain antibody mutants have been evolved in vitro with antigen-binding equilibrium dissociation constant $K_d = 48 \text{ fM}$ and slower dissociation kinetics (half-time > 5 days) than those for the streptavidin–biotin complex. These mutants possess the highest monovalent ligand-binding affinity yet reported for an engineered protein by over two orders of magnitude. Optimal kinetic screening of randomly mutagenized libraries of $10^{5} \sim 10^{7}$ yeast surface-displayed antibodies enabled a >1,000-fold decrease in the rate of dissociation after four cycles of affinity mutagenesis and screening. The consensus mutations are generally nonconservative by comparison with those already occurring mouse Fv sequences and with residues that do not contact the fluorescein antigen in the wild-type complex. The existence of these mutants demonstrates that the antibody Fv architecture is not intrinsically responsible for an antigen-binding affinity ceiling during in vivo affinity maturation.

Over 100 antibody therapeutics are currently in clinical trials for cancer, viral, autoimmune, and other diseases (1). High-affinity antigen recognition is at the heart of all such therapies and is generally attained by affinity maturation in vivo in the mammalian immune system. It has been observed that, during in vivo affinity maturation, the B cell response exhibits an apparent affinity ceiling near $K_d > 0.1 \text{ nM}$, because of the inability to discriminate slower dissociation kinetics relative to intrinsic B cell receptor internalization rates (2, 3). Because affinity is a critical variable for therapeutic applications such as antibody tumor targeting (4–7), extensive efforts have been made to generate higher affinity antibodies by directed evolution. The highest antibody affinities reported to date are $K_d \approx 10^{-20} \text{ pM}$ (8–10).

Antibody/hapten recognition in the 4-4-20/fluorescein model system used in this study has been characterized extensively by structural (11, 12), thermodynamic (13), kinetic (14), computational (15), spectroscopic (16), and mutagenic (17) means. The 4-4-20 affinity for the fluorescein–biotin (FL-bio) hapten ($K_d = 0.7 \pm 0.3 \text{ nM}$ in PBS) is near the affinity ceiling of the tertiary antigen-binding equilibrium dissociation constant. The 4-4-20 scFv antibody fragment can be displayed on the surface of yeast, and mutants with increased affinity can be isolated by flow cytometric cell sorting (18).

Herein, we report that optimal kinetic screening of randomly mutagenized libraries of $10^{5} \sim 10^{7}$ yeast surface-displayed 4-4-20 antibodies enabled a >10,000-fold decrease in the rate of dissociation (in PBS) after four cycles of affinity maturation and screening. These in vitro evolved single-chain antibody mutants have an antigen-binding equilibrium dissociation constant $K_d = 48 \text{ fM}$ and slower dissociation kinetics (half-time > 5 days) than those for the streptavidin–biotin complex, representing the highest reported monovalent ligand-binding affinity for an engineered protein by over two orders of magnitude. The existence of these mutants demonstrates that the antibody Fv architecture does not possess an intrinsic affinity ceiling for antigen binding.

Materials and Methods

Construction of Randomly Mutated scFv Libraries. A single point mutant of the 4-4-20 scFv (mutant 4 M1.1) with 3-fold higher affinity was isolated previously from a random library generated in a mutator E. coli strain (18). A random scFv library based on this clone was created by adapting the sexual PCR method of Stemmer (19). The expression cassettes of 4 M1.1 and the unselected mutator strain-generated library were amplified by PCR with T3 and T7 promoter standard primers. An equimolar mixture of amplified 4 M1.1 and library PCR products (approximately 30 µg) were fragmentated to <200 bp by DNase I, purified by gel-filtration with Centri-Sep columns (Princeton Separations, Princeton), and recombined essentially by following the method of Stemmer (19) replacing Taq polymerase with Pfu polymerase and performing 45 cycles of PCR. Final PCR amplification was performed with Taq polymerase in the presence of 2.25 mM MgCl$_2$ and 0.375 mM MnCl$_2$ to introduce further mutations. Primers for final amplification included sequences of $\approx$100 bp 5’ and 50 bp 3’ flanking the scFv ORF to allow efficient enzymatic digestion. PCR products were purified by agarose gel electrophoresis and electroelution, digested with NheI and XhoI, gel purified, and ligated into digested pCT302 backbone (18). Ligation reactions were exchanged into distilled water, concentrated with Centricon-30 and Microcon-50 filters (Amicon), and transformed into DH10B Electrocompetent cells (Life Technologies, Grand Island, NY). Transformants were pooled, and aliquots were plated to determine library diversity before amplification and purification of plasmid DNA with the Qiagen (Chatsworth, CA) Maxiprep kit. Library DNA was transformed subsequently into yeast strain EBY100 (11) by using the method of Gietz and coworkers (http://to.trends.com).

Transformants were pooled in SD-CAA medium [6.7% (vol/vol) yeast nitrogen base/2% (vol/vol) glucose/5% (vol/vol) casamino acids/100 mM sodium phosphate, pH 6.0], and aliquots were plated to determine library diversity. Subsequent libraries were constructed with similar methods, but these methods began with plasmid DNA recovered from a mixed yeast population screened for improved binders in the prior round and the unselected library from the prior round. PCR-amplified DNA from these sources was mixed at a 9:1 molar ratio before DNase I fragmentation. Incorporation of 10% unselected library DNA and error-prone final PCR amplification allowed introduction of further mutations as well as recombination of selected mutations in the resulting library. Individual libraries ranged from $3 \times 10^6$ to $2 \times 10^7$ clones.

Screening of Yeast-Displayed scFv Libraries. Yeast cells [$10^8$ in 2.5 ml of TBS (20 mM Tris base/137 mM sodium chloride, pH 7.6)] displaying mutagenized 4-4-20 scFvs were first incubated with 1 μM FL-bio to saturate surface binding, washed, and then incubated in 1 μM 5-aminofluorescein competitor for a fixed period at 25°C. Competition times were calculated from a mathematical model (20). Labeling with streptavidin-R-
phycoerythrin after competition allowed isolation by flow cytometric sorting of those cells retaining the highest levels of bound FL-bio. Cells were co-tainted with 12CA5 mAb (Roche Molecular Biochemicals) to normalize for cell-to-cell variation in expression level as described (18). Cells were sorted with gate settings determined by using a mathematical analysis (20); four rounds of sorting and regrowth were performed to isolate highly enriched (>80%) populations of improved mutants.

**Kinetic and Sequence Analysis of Isolated Clones.** Individual clones chosen at random were assayed for dissociation rate constants ($k_{diss}$). Cells were saturated with FL-bio (1.0 μM; room temperature; 30 min), washed, labeled with streptavidin-R-phycoerythrin [1:100 streptavidin-phycoerythrin (PharMingen); 30 min on ice], and incubated at 25°C with at least a 100-fold molar excess of 5-aminofluorescein (10 μM vs. ≈1 nM surface-displayed antibody in the reaction). Aliquots were removed and analyzed by flow cytometry at various times; dissociation rate constants were determined by fitting fluorescence data to a first-order kinetic model. Plasmid DNA was recovered from yeast cells by glass bead lysis and phenol:chloroform extraction (21), followed by purification with the Wizard DNA Cleanup kit (Promega) and transformation into competent DH5α, DH10B, or XL1-Blue cells. scFv ORF sequences were determined by the University of Illinois Biotechnology Center by dideoxy terminator sequencing with an ABI Prism 377. Sequenced clones were chosen based on differing values of $k_{diss}$ to avoid sequencing multiple isogenic clones.

**Fluorescence Quenching of Antigen by Soluble scFv Protein.** scFvs were subcloned and expressed solubly in *Saccharomyces cerevisiae* under conditions as described (22). Protein was recovered from raw culture supernatants by affinity chromatography with FITC-coupled BSA. scFv-containing fractions were purified further by FPLC on a Hi-TrapQ anion exchange column (Amersham Pharmacia). Protein concentration, binding constants, and maximal fluorescence quenching constants were determined by direct titration with FL-bio in an SLM Aminco SPF500 spectrofluorometer. Excitation wavelength was 492 nm, and emission was collected at 510 nm; all experiments were conducted in quartz cuvettes thermostated to 25°C. Dissociation rate constants were measured by addition of a 500-fold molar excess of 5-aminofluorescein (nonfluorescent competitor) at time 0 to an equilibrated mixture of scFv and FL-bio in PBS (150 mM NaCl/10 mM sodium phosphate buffer, pH 8) or low salt buffer (LSB; 1 mM sodium phosphate, pH 8). Association rate constants were determined by injection of FL-bio into equilibrated scFv samples and fitting quenching data to a model with $k_{diss}$ values determined from competitive dissociation experiments. Control experiments lacking 4 M5.3 scFv, FL-bio, or 5-aminofluorescein were performed in parallel to ensure the absence of artifacts from adsorption, evaporation, photobleaching, or other processes over the course of the experiment.

**Results and Discussion**

The complete 4-4-20 scFv ORF was mutagenized by error-prone DNA shuffling (19). Fluorescently labeled clones exhibiting slowed antibody-hapten dissociation kinetic constants were identified and isolated by flow cytometry with optimal screening and sorting conditions calculated from a mathematical model (20). Up to 20 improved clones were selected randomly for individual measurement of the dissociation rate constant $k_{diss}$, and 10 improved clones exhibiting the widest range of $k_{diss}$ values were selected for further analysis and are represented in Fig. 1A. The complete collection of isolated mutants was then recombined by modified DNA shuffling, together with further error-prone PCR mutagenesis. This cycle of mutagenesis and screening was repeated three times, resulting in mutant scFvs with

**Fig. 1.** Dissociation kinetics of mutant scFv at 25°C. (A) Yeast displaying mutant scFv isolated from a random library was assayed for antigen dissociation rate. The 4-4-20 and VHS95A values were calculated from fluorescence quenching data with soluble scFvs. Dissociation rates for barnase/barstar (45), streptavidin/biotin (46), avidin/biotin (46), and trypsin/bovine pancreatic trypsin inhibitor (BPTI) (47) are indicated for comparison. (B) Dissociation kinetics of purified, soluble 4 M5.3 and 4-4-20 scFvs were assayed by fluorescence quenching. Triplicate traces for 4-4-20 scFv in PBS (150 mM NaCl/10 mM sodium phosphate buffer, pH 8) are shown, and 12 independent replicates are shown for 4 M5.3 in PBS or LSB. $\tau_{1/2}$, half-time for dissociation.
dissociation rate constants over four orders of magnitude slower than the $k_{\text{diss}}$ of the 4-4-20 scFv (Fig. 1). As a basis for comparison, the dissociation rates for other high-affinity protein–ligand complexes are also shown in Fig. 1. The half-time for dissociation of the 4 M5.3 mutant is over 5 days, such that optimal screening of a fifth library would require kinetic competition times longer than 3 weeks (20) for each of four subsequent enrichment sorting steps.

A clone from the fourth library screen (4 M5.3) was solubly expressed, and the association and dissociation constants were measured in solution (Table 1; Fig. 1B). These kinetic values are in quantitative agreement with those measured on the cell surface (Fig. 1A). Electrostatic steering of fluorescein into the binding pocket (23, 24) is evidenced by the salt dependence of $k_{\text{ass}}$. The association rate constant of 4 M5.3 is $14\times$ lower than that of 4-4-20 scFv in both PBS and LSB. This decrease somewhat offsets the 16,000-fold improvement in $k_{\text{diss}}$ (3,000-fold in LSB), resulting in $K_d$ = 270 fM in PBS and $K_d$ = 48 fM in LSB. To our knowledge, the 4 M5.3/FL-bio interaction is the highest affinity noncovalent protein–ligand complex yet engineered.

Sequences were obtained of each of the mutants represented in Fig. 1, and mutated residues are shown in Table 2. With the exception of 4 M4.15, 16, and 19 (which are identical isolates),

### Table 1. Binding and dissociation rate constants for soluble scFvs

<table>
<thead>
<tr>
<th>scFv</th>
<th>Buffer</th>
<th>$k_{\text{ass}}$ (× $10^{-2}$ M$^{-1}$s$^{-1}$)</th>
<th>$k_{\text{diss}}$ (× $10^6$ s$^{-1}$)</th>
<th>$K_d$ fM</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-4-20</td>
<td>PBS</td>
<td>8.0 ± 3.5</td>
<td>25,100 ± 100</td>
<td>310,000</td>
</tr>
<tr>
<td></td>
<td>LSB</td>
<td>41.0 ± 8.0</td>
<td>4,270 ± 100</td>
<td>10,000</td>
</tr>
<tr>
<td>4 M5.3</td>
<td>PBS</td>
<td>0.6 ± 0.03</td>
<td>1.6 ± 0.3</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>LSB</td>
<td>2.9 ± 0.9</td>
<td>1.4 ± 0.4</td>
<td>48</td>
</tr>
</tbody>
</table>

ND, not determined.

### Table 2. Mutations in selected scFvs

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
<th>Column</th>
<th>Residue</th>
</tr>
</thead>
</table>
mutations cluster strikingly around VH CDR3; it has been
4-4-20 Fab crystal structure shown in Fig. 3. The consensus
arising from the mechanism of somatic hypermutation.
accessed by the
erally sampled areas of antibody sequence space infrequently
the VL domain is striking, with 6-fold, 7-fold, 7-fold, and 4-fold
6 of 10 are in CDR loops. The relative paucity of mutations in
in each of the four rounds, respectively. Comparison of frequency in the Kabat database
(25) for consensus mutations vs. 4-4-20 wild-type residues indi-
cates that 9 of 10 consensus changes are to rare amino acids at
those positions, as shown in Fig. 2. Most striking in this regard
is the VH103 mutation (fixed in the mutant population in round
1), which alters the nearly invariant VH103 tryptophan residue
(99.3% of mouse VH sequences in the Kabat database). Of the
10 consensus mutations after the fourth round of screening, 9 of 10 are in the VH domain, and
and 2 are in the fourth shell (Fig. 3). VH87 is on the scFv face
opposite the fluorescein-binding pocket. This spatial distribution
generally supports the observation that further affinity matura-
ion of antibodies with affinity in the low nanomolar range
occurs most effectively via changes in “vernier” or second-
sphere residues (31, 32) rather than contact residues (33). In
addition, 4 of 10 consensus mutations occur at VH–VL interfacial
sites, suggesting that improved stability and/or orientation of the
VH–VL domain pairing may be important in affinity improve-
ment for the scFv.
A molecular dynamics comparison of liganded and unliganded
4-4-20 demonstrates a significant increase in interresidue cor-
related motions on fluorescein binding, particularly in the VH
domain (15). The 4-4-20 VH domain possesses generally larger
motions than the VL domain (11), indicating greater
flexibility. Of the 10 consensus mutations, 9 lie within VH, which
could be consistent with the binding site preorganization mech-
anism for affinity improvement proposed previously for anti-
body affinity maturation from the germ-line sequence (34).

Definitions of contact were interatomic distances (in Å) equal to or less than
4.1 C-C, 3.3 O-O, 3.4 N-N and N-O, 3.8 C-C, and 3.7 C-O (48). Atomic coordinates
were from the high-resolution crystal structure of Whitlow, et al. (ref. 11; PDB
ID code 1FLR).

Fig. 3. Sites of consensus mutations in the 4-4-20 Fv. Backbone structure of
the first 118 heavy-chain residues (gray) and the first 112 light-chain residues
(blue) are shown. Fluorescein ligand (green) and mutated residues are de-
picted by space-filling models. Mutation sites are color-coded by distance from
the binding site: first-shell residues are magenta; second-shell residues are
yellow; third-shell residues are cyan; and fourth-shell residues are white. Residues in the first shell were defined as those with one or more atoms
directly contacting ligand; second-shell residues were defined as those with
one or more atoms directly contacting any residue in the first shell; third- and
fourth-shell residues contact second- and third-shell residues, respectively.
Definitions of contact were interatomic distances (in Å) equal to or less than
4.1 C-C, 3.3 O-O, 3.4 N-N and N-O, 3.8 C-C, and 3.7 C-O (48). Atomic coordinates
were from the high-resolution crystal structure of Whitlow, et al. (ref. 11; PDB
ID code 1FLR).

Consensus mutations accumulate in each round; once fixed in
an early round, they generally are present in subsequent rounds.
It is noteworthy that particular consensus mutations dominate
the screened population at each round. It is not possible to
ascertain from the available information whether mutations
added in subsequent rounds were introduced by recombination of
preexisting mutations or de novo mutagenesis during error-
prone PCR and DNA shuffling. Of 10 consensus mutations after
the fourth round of screening, 9 of 10 are in the VH domain, and
of 10 are in CDR loops. The relative paucity of mutations in
the VH domain is striking, with 6-fold, 7-fold, 7-fold, and 4-fold
excess mutations in VH relative to VL in each of the four rounds,
respectively. Comparison of frequency in the Kabat database
(25) for consensus mutations vs. 4-4-20 wild-type residues indi-
cates that 9 of 10 consensus changes are to rare amino acids at
those positions, as shown in Fig. 2. Most striking in this regard
is the VH103 mutation (fixed in the mutant population in round
1), which alters the nearly invariant VH103 tryptophan residue
(99.3% of mouse VH sequences in the Kabat database). Of the
10 consensus mutations identified in the present work are located
at 4-4-20 residues that were present in the loop-grafted 4D5 Flu
hybrid protein, indicating that they lie within the portion of the
scFv largely responsible for binding specificity. Of particular interest is the VH E55V mutation enriched in the fourth round of
this study, which was also identified independently as one of two
mutations that together improved the stability of 4D5 Flu by 4
kcal/mol, with an unexpected 20-fold increase in affinity (30).

A molecular dynamics comparison of liganded and unliganded
4-4-20 demonstrates a significant increase in interresidue cor-
related motions on fluorescein binding, particularly in the VH
domain (15). The 4-4-20 VH domain possesses generally larger
motions than the VL domain (11), indicating greater
flexibility. Of the 10 consensus mutations, 9 lie within VH, which
could be consistent with the binding site preorganization mech-
nanism for affinity improvement proposed previously for anti-
body affinity maturation from the germ-line sequence (34).
Dissection of the thermodynamic, kinetic, and structural mech-
anisms by which the mutations in 4 M5.3 increase binding free
energy by 3–4 kcal/mol should contribute to an improved
understanding of protein recognition. Information gleaned from
these gain-of-function mutations would complement loss-of-function mutational studies of strong protein–ligand interactions such as streptavidin–biotin (35).

The majority of current antibody engineering strategies focus mutagenesis on the antibody CDR loops (e.g., refs. 8, 10, 36, and 37), an approach that would not identify 4 of the 10 consensus mutations from this study. In fact, the first consensus mutation to become fixed (VH1W103L) occurred at a highly conserved framework position removed from the binding site; the possibility that 1 or more subsequently selected mutations depended on the context of the VH1W103 mutation must be considered. The experimentally simpler strategy of error-prone PCR of the whole scFv gene was therefore arguably more effective given these results. Concerns regarding potential immunogenicity of framework mutations in a therapeutic antibody might be addressed by judicious choice of buried residues from among the selected mutations.

The relative ease with which extremely high affinity has been attained in this study might be attributed to (i) quantitatively optimized screening methodology and (ii) minimization of expression bias by use of a eukaryotic expression host. In the first instance, labeling for optimal discrimination of improved clones (20) and reduced stochastic variation in dissociation kinetics because of assay of 10^4–10^5 scFv molecules per yeast cell enable fine discrimination of affinity (reproducibility of ±10% in k_{ass} assayed by flow cytometry for yeast-displayed scFvs and ±30% for K_d ref. 38). In the second instance, yeast’s secretory biosynthetic apparatus effectively folds and displays or secretes (22) the 4-4-20 scFv, a molecule that forms inclusion bodies and is 98% insoluble when expressed in the periplasm of wild-type E. coli (39). Thus, a greater proportion of protein shape space will be sampled by yeast display because of elimination of prokaryotic expression biases against scFvs such as 4-4-20.

Generation of antibodies that bind essentially irreversibly relative to the relevant physiological time scale could improve efficacy for cancer immunotherapy with noninternalizing tumor-associated antigens (4–7) and passive immunization against viral and microbial pathogens (40–43). The mutagenesis and screening methodology described herein has also been applied successfully to an antibody against hen egg lysozyme (data not shown) and an antibody against the T cell receptor (44). Beyond the implications of the general capability to engineer femtomolar affinity antibodies, more specifically, a molecule such as 4 M5.3 that binds and quenches fluorescein with an affinity similar to that of streptavidin–biotin may enable interesting approaches in analytical biochemistry when combined with the array of available fluorescence-detection methodologies.

The 4-4-20 scFv was obtained from D. M. Kranz, who also provided helpful comments on the manuscript. We thank G. Durack for assistance optimizing cell-sorting instrumentation. Funding was provided by the Whitaker Foundation and National Institutes of Health Grant GM55767.