

# Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies

Patrick S. Daugherty\*<sup>†</sup>, Gang Chen<sup>‡</sup>, Brent L. Iverson<sup>§¶</sup>, and George Georgiou\*\*<sup>¶</sup>

Departments of \*Chemical Engineering and <sup>§</sup>Chemistry and Biochemistry, and <sup>†</sup>Institute for Cell and Molecular Biology, University of Texas, Austin, TX 78712

Communicated by Edwin N. Lightfoot, Jr., University of Wisconsin, Madison, WI, December 3, 1999 (received for review August 24, 1999)

**Random mutagenesis and selection using phage or cell surface display provides an efficient method for affinity maturation of single chain Fv (scFv) antibodies, thereby improving function in various applications. To investigate the effects of mutation frequency on affinity maturation, error-prone PCR was used to generate libraries containing an average ( $m$ ) of between 1.7 and 22.5 base substitutions per gene in a high affinity scFv antibody that binds to the cardiac glycoside digoxigenin. The scFv antibody libraries were displayed on *Escherichia coli*, and mutant populations were analyzed by flow cytometry. At low to moderate mutation frequencies with an average mutation rate of  $m \leq 8$ , the fraction of clones exhibiting binding to a fluorescently labeled conjugate of digoxigenin decreased exponentially ( $r^2 = 0.99$ ), but the most highly mutated library ( $m = 22.5$ ) had significantly more active clones than expected relative to this trend. A library with a low error rate ( $m = 1.7$ ), one with moderate error rate ( $m = 3.8$ ), and the one with high error rate ( $m = 22.5$ ) were screened for high affinity clones under conditions of identical stringency using fluorescence-activated cell sorting. After several rounds of enrichment, each of the three libraries yielded clones with improved affinity for the hapten. The moderate and high error rate libraries gave rise to clones exhibiting the greatest affinity improvement. Taken together, our results indicate that (i) functional clones occur at an unexpectedly high frequency in hypermutated libraries, (ii) gain-of-function mutants are well represented in such libraries, and (iii) the majority of the scFv mutations leading to higher affinity correspond to residues distant from the binding site.**

**I**n *vitro* evolution has led to a number of remarkable successes in the engineering of proteins with improved function or stability (1–5). Notable recent examples of improved enzymes obtained by *in vitro* evolution include the generation of a mutant aspartate aminotransferase exhibiting over a  $10^6$ -fold change in substrate specificity, the engineering of fungal peroxidase variants with greatly increased thermal and oxidative stability, and the isolation of thymidine kinase mutants with markedly enhanced specificity for AZT (6–8). Also, *in vitro* evolution has been used to improve the binding affinity of a number of antibodies (5, 9–13). The experimental design of *in vitro* evolution studies is straightforward: First, random mutations are introduced in the DNA by *in vitro* or *in vivo* techniques (2, 9, 14, 15). Subsequently, the corresponding genes are expressed in a microbial host, and clones producing proteins with improved function are isolated by screening or, in special cases, using a biological selection.

It has been suggested that, because the size of the library that can be screened with existing technologies is rather small (typically fewer than  $10^5$  clones can be screened for enzymatic catalysis), it is sensible to seek improved enzyme fitness by iterative mutagenesis and screening of small libraries designed to have 1–2 amino acid substitutions (16, 17). With such a low mutation frequency, a large fraction of the sequence space corresponding to all possible 1–2 amino acid substitutions should be represented within a library of moderate size. The best mutants from such a library are isolated and subjected to

additional rounds of mutagenesis and screening. Similarly, mutant libraries with a theoretical diversity smaller than or comparable to the number of library clones are overwhelmingly used for the isolation of high affinity peptides and antibodies by phage display (9, 12, 18).

The iterative screening of libraries with low mutation frequency has been broadly adopted for *in vitro* evolution studies for several reasons: First, natural evolution is thought to involve the slow and stepwise accumulation of genetic changes (16, 17). Second, the overwhelming majority of clones in libraries with high mutation loads are expected to be nonfunctional. Third, sequence space is vastly expanded in libraries with high mutation frequencies, making the sampling of that space extremely sparse. Therefore, rare, gain-of-function mutants may not be represented within the relatively small number of functional clones in the library. However, a handful of experimental observations in recent years indicates that *in vitro* evolution strategies other than iterative low frequency mutagenesis deserve further consideration. Christians and Loeb (19) succeeded in isolating alkyltransferases with higher activity from  $4.1 \times 10^5$  transformants in a library with an average of 7.4 nucleotide changes per gene. Martinez *et al.* (20) showed that functional mutants of dihydrofolate reductase with as many as 7 amino acid substitutions (of a total of 78 amino acids) can be isolated from moderately sized libraries. Finally, Zacco and Gherardi (15) isolated TEM 1  $\beta$ -lactamase variants with activity against cefotaxime by screening small pools of clones ( $<10^5$ ) from libraries containing up to 27 nucleotide substitutions per gene.

In this work we have sought to analyze quantitatively the effect of mutation frequency by error-prone PCR on (i) the frequency of clones that remain functional and (ii) the likelihood of isolating gain-of-function mutants of a high affinity single chain Fv (scFv) antibody. Such an analysis was made possible by displaying the scFv antibodies on *Escherichia coli* (21), coupled with the use of fluorescence-activated cell sorting (FACS) for population analysis and for the isolation of clones with improved affinity. We and others have demonstrated that high affinity binding proteins can be readily isolated from libraries displayed on the surface of microorganisms and screened by FACS (11, 22–24). Here, we show that higher affinity scFv clones could be isolated from libraries with an average mutation rate ( $m$ ) of 1.7, 3.8, and 22.5. For the  $m = 22.5$  library, only about 10,000 clones or 0.17% of the total library exhibited hapten binding activity.

Abbreviations: scFv, single chain Fv; FACS, fluorescence-activated cell sorting; SPR, surface plasmon resonance.

<sup>†</sup>Present Address: Fred Hutchinson Cancer Research Center, Division of Human Biology, 1100 Fairview Avenue North, Seattle, WA 98109-1024.

<sup>¶</sup>To whom reprint requests should be addressed. E-mail: biverson@utxvms.cc.utexas.edu or ggmjd@mail.utexas.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.030527599. Article and publication date are at [www.pnas.org/cgi/doi/10.1073/pnas.030527599](http://www.pnas.org/cgi/doi/10.1073/pnas.030527599)

Nevertheless, mutants with significantly higher affinity than the wild type were well represented within the active fraction of the population. These results suggest that it is possible to explore effectively sequence space and to improve fitness by taking relatively large mutational leaps, affecting several residues simultaneously. On a more practical level, the screening of libraries containing a moderate to high frequency of random mutations appears to be an effective strategy for *in vitro* affinity maturation.

## Materials and Methods

**Strains, Plasmids, and Reagents.** Plasmids pSD192 and pB30D have been described elsewhere (23, 24). The vector pB30DN was constructed by deleting the wild-type scFv(dig) gene from pB30D by *EcoRI* digestion and religation. *E. coli* strain LMG194 [*F*<sup>-</sup>  $\Delta$ *lacX74 galE galK thi rpsL  $\Delta$ phoA (PvuII)  $\Delta$ ara714 leu::Tn10] was used for all library screening experiments. Cultures were grown in 10 ml of LB medium supplemented with chloramphenicol (15  $\mu$ g/ml) and ampicillin (75  $\mu$ g/ml) in 125-ml shake flasks at 37°C. Oligonucleotide primers were from Genosys (The Woodlands, TX), and restriction enzymes were from Promega. BODIPY-FL-EDA was from Molecular Probes. The synthesis of digoxigenin-BODIPY-FL has been described (24).*

**Random Mutant scFv(dig) Libraries.** The gene encoding the wild-type scFv(dig) was randomized by using error-prone PCR to create libraries with 1.7–22.5 mutations per gene. For low frequency mutagenesis, scFv(dig) and a short fragment of chloramphenicol acetyltransferase were randomized by using primers 1 (5'-TGGACCAACAACATCGGT-3') and 2 (5'-AGGGCAGCATGCACTGCCTTA-3') with pSD192 as template, essentially as described (25), with the exception that MgCl<sub>2</sub> was used at a final concentration of 1.5 mM. For high frequency mutagenesis, error-prone PCR was performed exactly as described (26) with primers 1 and 3 (5'-TATTCGAAGACGTTAGCTCTCGGGTC-3'). Error-prone PCR was performed using a Perkin-Elmer 9600 thermocycler, as follows: 1 cycle, 3 min at 94°C; 30 cycles, 1 min at 94°C, 2 min at 50°C, 3 min at 72°C; 1 cycle, 5 min at 72°C. To avoid excessive mutagenesis of the *cat* gene, it was amplified separately by non-error-prone PCR with primers 2 and 4 (5'-GACCCCGAGGACTAACGTCTTCGAATAAATAC-3'), using a high fidelity polymerase and pSD192E1 as the template. The latter differs from pSD192 only by the absence of a second *EcoRI* site within the *cat* gene (23). The amplified *cat* gene was then joined with the error-prone PCR product of the scFv by overlap extension PCR with primers 1 and 2. PCR products were precipitated in ethanol, were resuspended in 10 mM Tris-HCl (pH 8.5), were digested with *EcoRI* (for low error frequency libraries) or *EcoRI* and *SphI* (for high frequency libraries) for 3 hr, and were gel purified by using a Qiagen (Santa Clarita, CA) gel extraction spin kit. The purified DNA (10  $\mu$ g) was ligated with *EcoRI*-digested pB30DN (15  $\mu$ g) or *EcoRI/SphI* digested pB30DN in a total volume of 400  $\mu$ l for 24 hr at 16°C. Ligation products were precipitated in ethanol and were resuspended in 40  $\mu$ l of 10 mM Tris-HCl (pH 8.5). Two 20- $\mu$ l aliquots were transformed separately by electroporation into 200  $\mu$ l of *E. coli* LMG194. Transformed cells were pooled and incubated for 1 hr at 37°C in 15 ml of SOC medium with gentle shaking. Serial dilutions were plated onto LB plates supplemented with 1% glucose, 50  $\mu$ g/ml carbenicillin, and 15  $\mu$ g/ml chloramphenicol to determine the library size. The transformed cells were cultured for 10 hr in a total volume of 200 ml of LB medium (supplemented as above), and plasmid library DNA was isolated. The error rate was estimated by sequencing 400–700 bp from each of 4–20 randomly picked colonies for each library. To verify the sequencing accuracy, both sense and antisense strands were sequenced.

**Flow Cytometric Analysis and Sorting of Random Mutant Libraries.** Randomized scFv libraries were analyzed and sorted as described (24). Plasmid libraries were transformed into LMG194 by electroporation, and cultures were grown overnight at 37°C.

Cells were subcultured 1:100 and were regrown at 25°C. At an OD<sub>600</sub> of 0.4–0.6, cultures were induced with 0.2% arabinose and were incubated for an additional 6 hr at 25°C with shaking. Cells (40  $\mu$ l) were harvested and placed into PBS buffer (960  $\mu$ l) with BODIPY-digoxigenin at a final concentration of 100 nM and then were incubated for 45 min. The cells were pelleted by centrifugation and were resuspended in PBS to yield a FACS throughput rate of  $\approx$ 2,000 events/sec. For flow cytometric analyses, the cells were analyzed 5–10 min after resuspension.

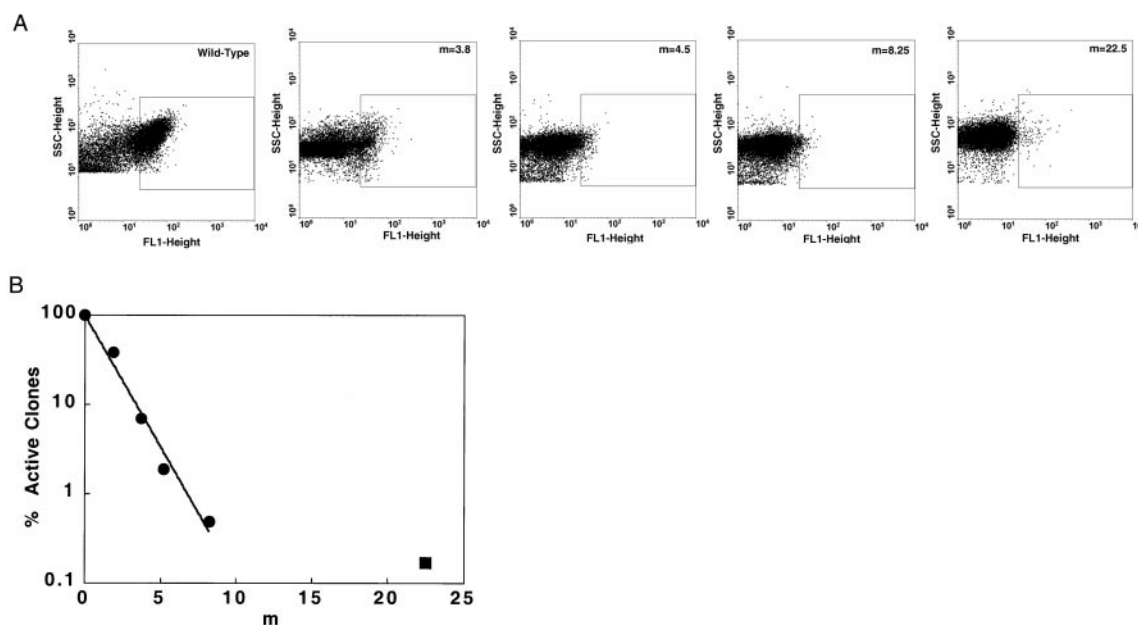
Each library was subjected to four or five rounds of selection with the first two rounds in Recovery mode followed by two or three rounds of selection in Exclusion mode. For the first round, the number of sorted cells was at least 5-fold greater than the number of independent transformants. Optimal screening parameters were chosen to minimize the probability that desired clones will be lost (27). During the first round of sorting,  $9 \times 10^6$  to  $3 \times 10^7$  cells were sorted in Recovery mode. A gate was set to recover 0.18–0.67% of the most fluorescent cells. Sorted cells were recovered as 15-ml fractions and were diluted into an equal volume of SOC medium in separate shake flasks; each fraction was regrown in a total volume of 30 ml of SOC/PBS. After overnight growth at 37°C, 100- $\mu$ l aliquots from each sort fraction were pooled in a 1.5-ml tube, were mixed thoroughly by inversion, and were subcultured 1:100 into 20 ml of LB medium for induction of expression. In round two,  $1.6 \times 10^6$  to  $1.8 \times 10^6$  cells were sorted in Recovery mode. The most fluorescent 0.6% of the population ( $\approx 10^4$  events) were recovered into SOC medium and were regrown. In the third round, the pooled cell population was washed twice with PBS and was incubated for 30 min with 2  $\mu$ M digoxin as a competitor, and  $0.91 \times 10^6$  to  $1.2 \times 10^6$  cells were sorted in Exclusion mode at a sort rate of  $\approx$ 1,000 cells/sec and were collected in 5- to 15-ml aliquots of SOC medium. For round three, the sort gate was set to include  $\approx$ 0.2% of the population. Round four was performed as in round three, except fractions containing the most fluorescent 0.2% were collected after 1 hr of incubation with digoxin (as an unlabeled competitor). A final round of sorting was performed by collecting the upper 0.1% of the population after incubation with competitor for 4 hr.

**Measurement of Hapten Dissociation Kinetics.** The hapten dissociation rates for the populations enriched after round four (for the  $m = 1.7$  library) or five (for the  $m = 3.8$  or  $m = 22.5$  libraries) were estimated by labeling the cells with 100 nM digoxigenin-BODIPY-FL and measuring the population fluorescence at 15-min intervals in the presence of 2  $\mu$ M digoxin. The dissociation rate constants for individual clones selected at random for the above populations were determined by flow cytometry as described (23).

**Characterization of Isolated Clones.** ScFv antibodies from clones isolated after library screening ( $m = 3.8, 22.5, 22.5$ ) were produced in soluble form as follows. The scFv genes were amplified by PCR, were digested with *NcoI* and *NheI*, and were ligated into a modified pET-25b vector containing a *pelB* signal sequence with a 3' *NcoI* restriction site for in-frame cloning of scFv antibody fragments. The resulting expression vectors were transformed into BL21(DE3) cells. Cultures were grown 4 hr in LB medium at 25°C and were induced with 0.05 mM IPTG. Cells were harvested 4 hr after induction and were subjected to osmotic shock to release periplasmic proteins. ScFv antibodies were purified from the osmotic shock fractions by immobilized metal affinity chromatography (IMAC) and subsequent gel filtration by FPLC to remove any higher molecular weight species. The binding kinetics of the purified, monomeric scFv proteins to digoxin were determined by surface plasmon resonance using a BIAcore 1000 instrument exactly as described (23, 28).

## Results

**Library Construction and Analysis.** As a model protein, we used an scFv antibody fragment derived from the 26-10 monoclonal



**Fig. 1.** (A) Flow cytometric analysis of different error-rate libraries labeled with 100 nM digoxigenin-BODIPY-FL. (B) Plot of the percentage of positive clones versus the average number of mutations per single chain Fv gene ( $m$ ) in different error-rate libraries.

antibody (21). This scFv binds with high affinity to digoxin and other cardiac glycosides, including digoxigenin. Digoxin and digoxigenin have an identical steroid ring structure; however, digoxin contains a tridigitoxose trisaccharide that is absent in digoxigenin. In the crystal structure of the 26-10 F<sub>AB</sub>, all contacts between digoxin and the antibody protein are with the steroid portion of the molecule that is essentially buried within the binding pocket of the antibody (29). For this reason, the digoxin and digoxigenin haptens are recognized by the intact antibody and by the scFv with similar affinities. Specifically, the  $K_D$  values for the purified scFv binding to digoxin and digoxigenin are  $0.9 \pm 0.2 \times 10^{-9} \text{ M}^{-1}$  and  $2.4 \times 10^{-9} \text{ M}^{-1}$ , respectively (28).

For display on *E. coli*, the scFv gene was fused to Lpp-OmpA' and was expressed from the low copy number vector pB30D (24). The Lpp-OmpA'-scFv fusion is transcribed from the arabinose inducible P<sub>BAD</sub> promoter. In the *ara*<sup>-</sup> *E. coli* strain LMG194 transformed with pB30D, on induction with 0.2% arabinose, >95% of the cell population expressed a high level of Lpp-OmpA'-scFv, as evidenced by the fluorescence distribution of cells incubated with digoxigenin conjugated to the fluorescent dye BODIPY-FL (Fig. 1). In the experiments described below, cultures were propagated under conditions in which the synthesis of the scFv fusion protein was repressed by the presence of glucose. Expression was induced before FACS analysis or sorting and then was turned off again by incubating under repressed conditions. In this way, Darwinian selection favoring faster growing clones in the library was minimized (24). Consequently, in library screening experiments, cells were isolated on the basis of the hapten binding properties of their displayed scFv antibodies.

Six scFv gene libraries with differing mutation rates were constructed by error-prone PCR. PCR amplification was carried out

with primers designed to amplify a DNA segment consisting of the entire scFv gene followed by a short, 5' segment of the chloramphenicol acetyltransferase (*cat*) gene (23). Ligation of the PCR product into pB30DN, which contains the remaining 3' portion of the *cat* gene, results in a plasmid that can confer resistance to chloramphenicol as well as ampicillin. Only cells transformed with plasmid DNA containing an insert in the correct orientation are able to grow in the presence of chloramphenicol. This allows for the elimination of background transformants not containing inserts and leads to the unequivocal determination of the number of scFv-containing clones. However, because both the scFv gene and the 5' portion of the *cat* gene are amplified under error-prone conditions, mutations that inactivate the latter also occur, decreasing the effective number of *cat*<sup>+</sup> transformants. Inactivation of the *cat* gene severely affected the library size for moderate to high error frequency libraries. To obtain a satisfactory library size for the three libraries with the highest mutagenesis rates, the scFv gene without the 5' portion of the *cat* gene was first subjected to error-prone PCR. Subsequently, the 5' piece of the *cat* gene was attached to the scFv gene fragments by overlap extension PCR. Using this strategy, for example, the number of Amp, chloramphenicol-resistant transformants was  $6 \times 10^6$  clones for the  $m = 22.5$  library.

The mutation rate in various libraries was determined by sequencing, in both directions, the scFv gene from 4–20 randomly picked clones (4,000–14,000 base pairs per library). For the  $m = 22.5$  library, 20 clones were analyzed to obtain a more accurate estimate of the error rate. Sequence analysis revealed that the fraction of nucleotide substitutions leading to transitions was  $\approx 70\%$ . Six libraries having  $m = 1.7$ – $22.5$  were analyzed further for hapten binding. In brief, cultures were incubated at 25°C and were induced with 0.2% wt/vol arabinose. These

**Table 1. Summary of libraries subjected to enrichment for high affinity clones by FACS**

Library	Size of library	Mutation rate, %	Percent of positive clones	No. of positive clones	$k_{diss}$ of final round sorting, $\times 10^3$ sec	$k_{diss}$ of best clone by FACS, $\times 10^3$ sec	$k_{diss}$ of best clone by BIAcore, $\times 10^3$ sec
$m = 1.7$	$3 \times 10^5$	0.22	40	$1.4 \times 10^5$	1.3	0.7	0.7
$m = 3.8$	$1 \times 10^6$	0.5	6.7	$6.7 \times 10^4$	0.3	0.4	0.55
$m = 22.5$	$6 \times 10^6$	3	0.17	$1 \times 10^4$	0.3	0.23	0.44

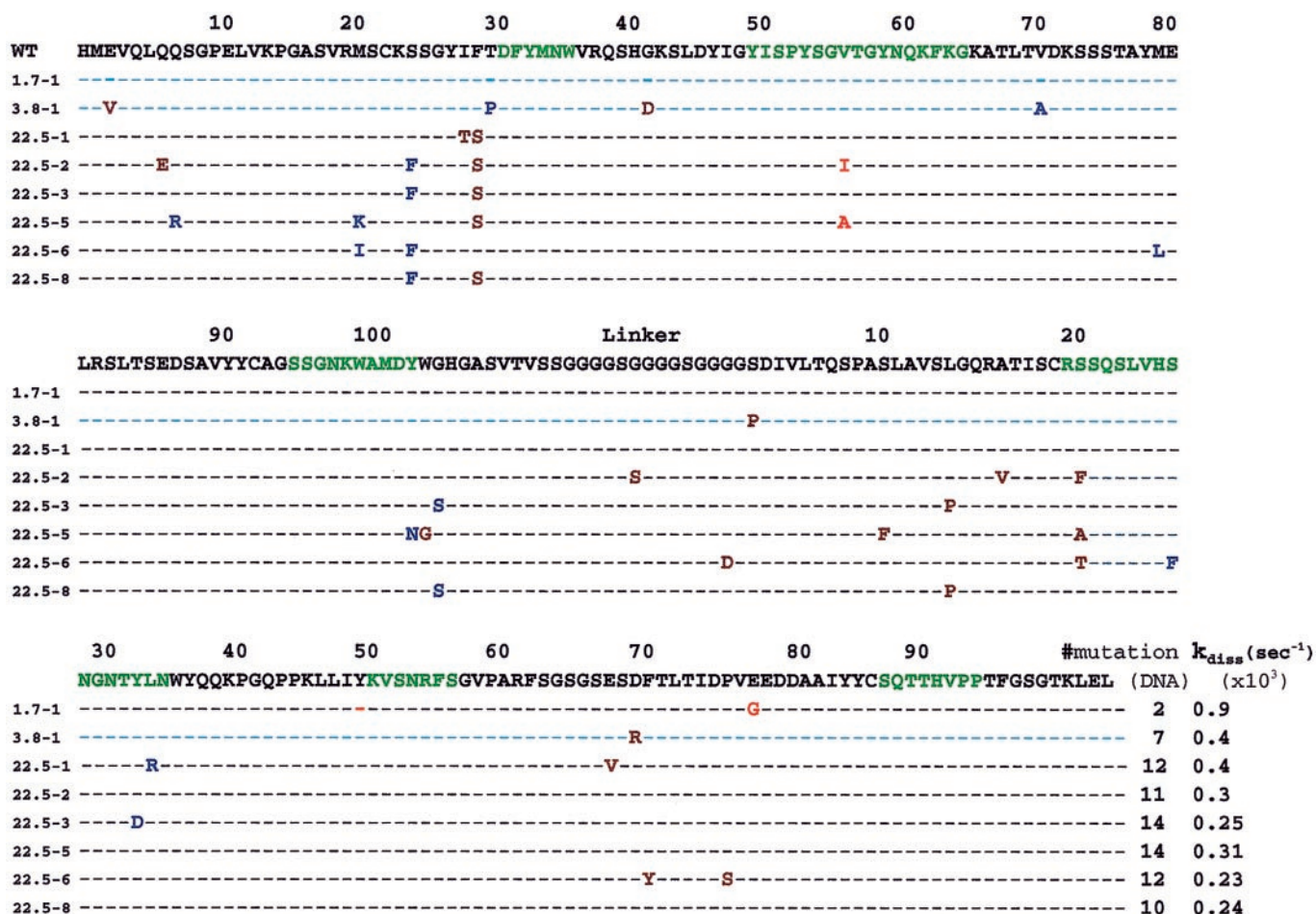


Fig. 2. Amino acid sequences of affinity-improved clones enriched from three libraries with different mutation rates ( $m = 1.7, 3.8,$  or  $22.5$ ). The total number of nucleotide substitutions and the dissociation rate constants measured by flow cytometry are shown. Green, complementarity-determining region residue; red, mutations in residues within 3 Å from the hapten; blue, mutations in residues located between 3 and 9 Å from the hapten; brown, mutations in residues  $>9$  Å from the hapten.

conditions have been determined to result in optimal display of Lpp-OmpA'-scFv fusions in *E. coli* (24). Forward scatter, which is indicative of cell size, and fluorescence caused by propidium iodide and digoxigenin-BODIPY-FL were measured by flow cytometry. Staining with propidium iodide was used to gate on intact cells. Using propidium iodide staining, nonviable and damaged cells that can bind nonspecifically to the fluorescent conjugate are excluded. Hapten binding was determined by labeling with 100 nM digoxigenin-BODIPY-FL conjugate. A FACS fluorescence threshold for BODIPY-FL fluorescence was defined such that 95% of all cells expressing the wild-type scFv had fluorescence values above the threshold (Fig. 1A). In the analysis of library populations, cells exhibiting fluorescence values above the threshold value were defined as "active" for hapten binding. For comparison, typically  $<0.01\%$  of *E. coli* that do not display the scFv antibody fall above the defined fluorescence threshold. Clearly, the above is an operational definition of hapten binding "activity"; more stringent or more relaxed criteria can be readily used by adjusting the concentration of the fluorescently tagged hapten or adjusting the threshold.

The fraction of library clones exhibiting hapten binding activity in flow cytometry assays was plotted as a function of  $m$  as shown in Fig. 1B. As expected, at low mutation rates ( $m = 1.7$ ), a sizable fraction of the library clones exhibited hapten binding. The fraction of active clones decreased exponentially with the number of mutations per gene ( $r^2 = 0.997$ ) for libraries with

$m \leq 8$ . However, as the mutation rate was increased from 8 to 22.5 mutations per gene, the fraction of active mutants dropped only slightly, from 0.3 to 0.17% of the cell population.

**Isolation of Affinity Improved Clones.** Libraries having low, moderate, and high error rates ( $m = 1.7, 3.8,$  and  $22.5$ , respectively) were screened for binding to digoxigenin. In the first and second rounds, positive clones were enriched after equilibrating the cell population with 100 nM digoxigenin-BODIPY-FL. Fluorescent cells were isolated by FACS using recovery mode in which cells that exhibit a strong fluorescence signal were collected together with any coincident cells that were present in the same volume of fluid. In rounds 3–5, cells were labeled with 100 nM digoxigenin-BODIPY-FL and were incubated for 30, 50, and 240 min, respectively, with 2  $\mu\text{M}$  digoxin to enrich preferentially those clones with slower hapten dissociation rates. For the  $m = 1.7$  library, enrichment was not observed between rounds three and four, and, therefore, no further enrichment rounds were attempted.

After the last round of screening, the mean hapten dissociation rates for the enriched population were determined by flow cytometry. As shown in Table 1, the enriched, polyclonal cell populations obtained from libraries with  $m = 1.7, 3.8,$  and  $22.5$  had dissociation rate constants ( $k_{diss}$ ) of  $1.3 \times 10^{-3} \text{ sec}^{-1}$ ,  $0.3 \times 10^{-3} \text{ sec}^{-1}$ , and  $0.3 \times 10^{-3} \text{ sec}^{-1}$ , respectively, compared with a value of  $1.0 \pm 0.1 \times 10^{-3} \text{ sec}^{-1}$  for the wild type. It should be noted that the  $k_{diss}$  value for the wild-type scFv determined by

flow cytometry is somewhat lower than the  $k_{diss} = 2.3 \times 10^{-3} \text{ sec}^{-1}$  measured by surface plasmon resonance (SPR) using purified scFv protein (28). This systematic discrepancy is presumably attributable to rebinding effects on the surface of *E. coli* (23). Nevertheless, the affinity ranking of different clones based on  $k_{diss}$  values measured by flow cytometry is identical to the ranking obtained from SPR data using purified scFv antibodies.

Between five and eight colonies from the final round of enrichment from each library were picked at random and were sequenced, and the corresponding hapten dissociation rate constants were determined by FACS. Results for the most improved clones are summarized in Fig. 2. The  $k_{diss}$  values for the individual clones from each library were consistent with the  $k_{diss}$  values measured for the enriched population (Table 1). For the  $m = 1.7$  library, the best clone ( $k_{diss} = 7 \times 10^{-4} \text{ sec}^{-1}$ ) had three base pair changes, corresponding to a single amino acid change. The remaining four clones from the  $m = 1.7$  library had  $k_{diss}$  values equal to or greater than the wild type ( $k_{diss} = 1 \times 10^{-3} \text{ sec}^{-1}$ ). For the  $m = 3.8$  library, five of five independent clones examined were found to have an identical nucleotide sequence, having seven nucleotide substitutions. In contrast, only two of seven clones analyzed from the  $m = 22.5$  library were identical. Clones enriched from the  $m = 22.5$  library contained between 10 and 14 mutations at the nucleotide level, but only 4–9 amino acid substitutions.

The wild-type scFv, the affinity improved clones from the  $m = 1.7$  and  $m = 3.8$  libraries, as well as two clones selected from the  $m = 22.5$  library were expressed in soluble form, without the Lpp-OmpA(46–159) sequence. All of the scFv antibodies could be expressed in the periplasm in relatively high yield. This was not unexpected because screening of surface displayed libraries by FACS selects on the basis of expression (higher overall signal) as well as affinity (23). ScFv proteins were purified from the osmotic shock fluid by immobilized metal affinity chromatography followed by size exclusion chromatography to eliminate any dimeric or higher molecular weight species. Freshly prepared, monomeric scFv proteins were used to determine the hapten binding kinetics by SPR. Measuring the binding kinetics on an SPR chip containing digoxigenin conjugated to BSA turned out to be complicated, presumably because of steric effects resulting from the steroid moiety being too close to the BSA surface. Therefore, hapten kinetics were determined by using the cardiac glycoside digoxin conjugated to BSA, rather than digoxigenin. Digoxin contains a carbohydrate “spacer” that, when bound to a protein such as BSA, should allow for better scFv access to the steroid portion of the molecule. All of the contacts between the parent 26–10 antibody and digoxin are with the steroid core of the molecule that is identical to the steroid core of digoxigenin. As a result, the relative ranking of kinetics of the scFv mutants measured by SPR using the digoxin-BSA conjugate are expected to be identical to those obtained by using digoxigenin conjugates. For wild-type scFv, the  $k_{assoc}$  and  $k_{diss}$  values were  $0.9 \times 10^6 \pm 0.1 \text{ M}^{-1}\text{sec}^{-1}$  and  $0.83 \times 10^{-3} \text{ sec}^{-1}$ , respectively, giving an equilibrium dissociation constant of  $0.92 \times 10^{-9} \text{ M}$ . The association rate constants for mutant antibodies were unchanged relative to the wild type. This was expected because mutants were selected on the basis of improved dissociation kinetics. For the 1.7-1, 3.8-1, 22.5-3, and 22.5-6 clones, the  $k_{diss}$  values were  $7 \times 10^{-4}$ ,  $5.5 \times 10^{-4}$ ,  $4.4 \times 10^{-4}$ , and  $4.4 \times 10^{-4} \text{ sec}^{-1}$ , respectively. In agreement with our earlier observations (23), FACS and surface plasmon resonance gave the same relative ranking of different clones with respect to hapten binding whereas the absolute  $k_{diss}$  values determined by the two techniques were somewhat different (see Fig. 2).

**Sequence Analysis of Affinity Improved Clones.** For the libraries analyzed, the number of nucleotide substitutions in the enriched clones deviated somewhat from the average for the library. For example, clone 3.8-1, isolated from the  $m = 3.8$  library, contained seven nucleotide substitutions. Assuming that the distribution of mutations follows the expected Poisson statistics, the frequency of clones

with seven nucleotide substitutions (5%) is expected to be 4-fold lower than the frequency of clones having four mutations, the approximate, experimentally determined average mutation frequency for the entire library. Similarly, for the  $m = 22.5$  library, the clones obtained from the final round of enrichment had between 10 (22.5-8) and 14 (clones 22.5-3, 22.5-5) mutations.

One potential advantage of libraries with a high frequency of mutations is that adjacent nucleotide substitutions can occur within a codon, in turn giving rise to a greater repertoire of amino acid substitutions. However, even with a mutation frequency of  $m = 22.5$ , such events were found to be quite rare; among all pre- and postsort clones analyzed, only one mutant (22.5-1) possessed an amino acid change resulting from two base changes in a single codon. The majority of mutations found in the affinity improved clones from the  $m = 22.5$  library were unique to each clone. Only a small number of mutations were shared by multiple clones. Specifically, all of the identical amino acid substitutions occurring in unique clones, namely H:28S (present in five clones), H:24F (four clones), and H:102S (two clones), were encoded by the same codons (Table 1). Most likely, common mutations found in unique clones arose early during error-prone PCR and were propagated within a large fraction of the population. This appears to be the case for the common mutations in clones 22.5-2, 22.5-3, and 22.5-8.

Six amino acid substitutions were present in clone 3.8-1 whereas clones isolated from the  $m = 22.5$  library contained between 4 and 9 amino acid changes. As can be seen from Fig. 2, the majority of the amino acid substitutions were nonconservative. Additionally, only 10 of 54 amino acid substitutions in all clones analyzed were in complementarity-determining region residues, indicating that substitutions in improved clones are not clustered within the complementarity-determining regions. Mutations in residues that make contact with the antigen were not found. Five mutations were found in residues making van der Waals contact with one or more of the antigen-contacting amino acids. Such residues were within a 3-Å radius from the hapten. Importantly, the majority of the mutations observed in the high affinity clones were in framework residues at a distance of 9 Å or more from the hapten. These results further illustrate that substitutions far removed from the binding site can exert profound effects on affinity (5, 9), and highlight the utility of random mutagenesis for affinity maturation.

## Discussion

For libraries with values of  $m \leq 8$ , a striking correlation was found between the average number of mutations per gene and the logarithm of the percentage of active clones in a library (Fig. 1). This result was based on the quantitative analysis of five large libraries, each containing at least  $10^5$  independent transformants. A similar exponential decrease in the fraction of active clones as a function of  $m$  was observed earlier from the analysis of smaller libraries of subtilisin mutants (31). Schellenberger and coworkers (31) noted that an exponential decrease in the fraction of active clones as a function of  $m$  can arise if (i) the distribution of mutations follows Poisson statistics (25), and (ii) the activity can be defined using a threshold such that clones above the threshold are fully active and, below, fully inactive. Accordingly, the fraction of active clones is given by the product of a Poisson distribution and a binomial distribution. Although the above model is based on these simplifying assumptions, it is striking that the effect of mutation frequency on the activity of two proteins of drastically different function and origin (i.e., subtilisin and an antibody) exhibit a similar behavior. The proportionality constant ( $q$ ) between the logarithm of the fraction of active clones and  $m$  is a quantitative measure of the plasticity of the protein or, in other words, its tolerance to mutations. Interestingly, the value of  $q$  for scFv(dig), 0.6, is greater than that measured for subtilisin [ $q = 0.27$  (ref. 31)], indicating a relatively low tolerance to mutation for scFv(dig). This low tolerance to mutation may be partially explained by the fact that the

parent monoclonal is of high affinity (23) and, therefore, had already undergone functional optimization by the immune system.

The library with a high mutation frequency ( $m = 22.5$ , corresponding to 3% nucleotide substitutions) resulted in a disproportionately high fraction of active clones relative to the trend identified for the low to medium mutation frequency libraries. Given the large size of the  $m = 22.5$  library ( $6 \times 10^6$  transformants), it is unlikely that the higher than expected fraction of active clones is a statistical artifact. In screening our scFv libraries with moderate ( $m = 3.8$ ) and high ( $m = 22.5$ ) error frequencies under conditions of identical stringency, affinity improved clones were recovered in both cases. As indicated in Table 1 and Fig. 2, the high error rate library contained only  $\approx 10,000$  active clones, yet several of these were gain-of-function mutants exhibiting hapten dissociation rates slower than the wild-type scFv. Given that gain-of-function mutants were isolated by exploring a minute fraction of the potential sequence space in this library, it is evident that, at least for this protein, there are numerous evolutionary paths to improved fitness. Correspondingly, the chemical problem of digoxigenin recognition within the context of the Ig fold apparently has multiple solutions. This hypothesis is supported by the following: (i) clones with comparable affinities but entirely different sequences were isolated from  $m = 3.8$  and  $m = 22.5$  libraries, and (ii) none of the mutations in contact residues that have previously been shown to result in higher hapten affinity (23, 32) was found in this work.

The improved clones isolated from the high mutation rate library ( $m = 22.5$ ) possessed only 10–14 DNA mutations (Fig. 2). However, it can be shown that the probability of randomly selecting six clones each having between 10 and 14 mutations would be  $\approx 0.02$ ,

assuming a Poisson approximation of error rates within the library. Thus, FACS has apparently selected for clones with fewer than average mutations for the  $m = 22.5$  library. Meanwhile, the proportion of silent mutations (39/73) represented among the isolated clones is also greater than among the unselected population. These data suggest a preferred mutation frequency for functional improvement of scFv(dig) may exist between  $m = 4$  and  $m = 22$  and that a library with such an error rate is even more likely to yield further improved mutants because of the presence of a larger fraction of active mutants. Alternatively, a non-Poisson distribution of mutation rates could be operating to bias our results with the high mutation rate library.

The data presented here have important implications for the directed evolution of antibodies and possibly other proteins. First, libraries with higher mutation rates can have a larger than expected number of active clones. Second, such libraries can yield mutants having improved function, despite the fact that a larger region of sequence space is being explored very sparsely. Third, the substitutions observed among affinity improved clones were distant from the binding site, further demonstrating that improvements in functional affinity result from diverse and poorly predictable mechanisms. Given the diversity of mutations found in hypermutated clones, subsequent *in vitro* recombination by DNA shuffling can be a productive way to achieve further, additive functional improvement. Finally, our results highlight the utility of cell surface display coupled with quantitative flow cytometric library analysis and screening.

We thank Jon Harmon for help with molecular modeling and Eileen Fehskens for assistance with DNA sequencing. This work was supported by a grant from the U.S. Army. Patrick Daugherty was supported in part by a National Institutes of Health Biotechnology Training Grant Fellowship.

1. Kuchner, O. & Arnold, F. H. (1997) *Trends Biotechnol.* **15**, 523–530.
2. Cramer, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. (1998) *Nature (London)* **391**, 288–291.
3. Yano, T., Oue, S. & Kagamiyama, H. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5511–5515.
4. Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12809–12813.
5. Saviranta, P., Pajunen, M., Jauria, P., Karp, M., Pettersson, K., Mantsala, P. & Lovgren, T. (1998) *Protein Eng.* **11**, 143–152.
6. Oue, S., Okamoto, A., Yano, T. & Kagamiyama, H. (1999) *J. Biol. Chem.* **274**, 2344–2349.
7. Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A. & Pedersen, A. H. (1999) *Nat. Biotechnol.* **17**, 379–384.
8. Christians, F. C., Scapozza, L., Cramer, A., Folkers, G. & Stemmer, W. P. (1999) *Nat. Biotechnol.* **17**, 259–264.
9. Low, N. M., Holliger, P. H. & Winter, G. (1996) *J. Mol. Biol.* **260**, 359–368.
10. Schier, R., McCall, A., Adams, G. P., Marshall, K. W., Merritt, H., Yim, M., Crawford, R. S., Weiner, L. M., Marks, C. & Marks, J. D. (1996) *J. Mol. Biol.* **263**, 551–567.
11. Boder, E. T. & Wittrup, K. D. (1997) *Nat. Biotechnol.* **15**, 553–557.
12. Hoogenboom, H. R., de Bruine, A. P., Hufon, S. E., Hoet, R. M., Arends, J. W. & Roovers, R. C. (1998) *Immunotechnology* **4**, 1–20.
13. Gram, H., Marconi, L. A., Barbas, C. F. D., Collet, T. A., Lerner, R. A. & Kang, A. S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3576–3580.
14. Cadwell, R. C. & Joyce, G. F. (1992) *PCR Methods Appl.* **2**, 28–33.
15. Zaccolo, M. & Gherardi, E. (1999) *J. Mol. Biol.* **285**, 775–783.
16. Arnold, F. H. (1998) *Nat. Biotechnol.* **16**, 617–618.
17. Arnold, F. H. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 2035–2036.
18. Yu, J. & Smith, G. P. (1996) *Methods Enzymol.* **267**, 3–27.
19. Christians, F. C. & Loeb, L. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6124–6128.
20. Martinez, M. A., Pezo, V., Marliere, P. & Wain-Hobson, S. (1996) *EMBO J.* **15**, 1203–1210.
21. Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10444–10448.
22. Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L. & Curtiss, R., III (1997) *Nat. Biotechnol.* **15**, 29–34.
23. Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L. & Georgiou, G. (1998) *Protein Eng.* **11**, 825–832.
24. Daugherty, P. S., Olsen, M. J., Iverson, B. L. & Georgiou, G. (1999) *Protein Eng.* **12**, 613–621.
25. Cadwell, R. C. & Joyce, G. F. (1992) *Genome Res.* **2** (1), 28–33.
26. Fromant, M., Blanquet, S. & Plateau, P. (1995) *Anal. Biochem.* **224** (1), 347–353.
27. Boder, E. T. & Wittrup, K. D. (1998) *Biotechnol. Prog.* **14** (1), 55–62.
28. Chen, G., Dubrawsky, I., Mendez, P., Georgiou, G. & Iverson, B. L. (1999) *Protein Eng.* **12**, 349–356.
29. Jeffrey, P. D., Strong, R. K., Sieker, L. C., Chang, C. Y., Campbell, R. L., Petsko, G. A., Haber, E., Margolies, M. N. & Sheriff, S. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10310–10314.
30. Zhao, H. & Arnold, F. H. (1997) *Nucleic Acids Res.* **25**, 1307–1308.
31. Shafikhani, S., Siegel, R. A., Ferrari, E. & Schellenberger, V. (1997) *BioTechniques* **23**, 304–310.
32. Short, M. K., Jeffrey, P. D., Kwong, R. F. & Margolies, M. N. (1995) *J. Biol. Chem.* **270**, 28541–28550.