Evolving Affinity Enhanced Antibodies:
High Mutational Rates for Discovery of Coupled Mutations

Introduction

The expanding market for antibody therapeutics demonstrates the importance of protein engineering. As of 2004, there were 140 protein therapeutics in clinical use within the US and Europe and about 500 more in clinical trials.[1] The success of protein engineering can often be attributed to directed evolutionary approaches resulting in improved or enhanced proteins. However, the primary limitation of optimized directed evolution methods is a bottleneck in the evolutionary process of protein engineering. In order to prevent the loss of function and stability of a protein library, the error-prone evolutionary processes have traditionally been limited to low mutational rates. Previous studies show that an increasing rate of mutagenesis is correlated with an exponential decline in library functionality.[2] However, it has also been demonstrated that high mutational rate libraries do not follow this trend and may outperform their low mutational rate counterparts.[3–5] In this study, we seek to demonstrate that the high functionality of high mutational rate libraries is due to the introduction of synergistic mutations: high mutagenesis rates reveal combinations of mutations that interact to produce enhanced proteins. Evidence supporting this hypothesis would demonstrate enhanced library functionality and decrease the number of rounds of library panning to identify most improved proteins due to combinations of synergistic mutations that would remain hidden in low error-rate libraries. Ultimately, the outcome of this study may contribute to future strategies of *in vitro* directed evolution: greater use of high error rate mutagenesis for protein evolution and a faster and more efficient lab to market therapeutic development process.

Background

General Method

There are three steps to the general method of *in vitro* directed evolution. First, error-prone PCR is used to introduce random mutations across the gene of interest. The resulting array of mutated genes is commonly called a library. Second, the corresponding mutated genes are expressed in a cellular host, allowing for the coupling of individual transcribing library genes

to corresponding translated, functional proteins. Third, clones that display enhanced function are isolated through a screening process.

Characterization of Mutational Rates

Historically, library mutagenesis strategies have employed techniques that generate low mutational rates because fewer mutations are thought to allow the protein to retain function. This idea suggests that the best method for protein evolution would slowly and steadily improve protein activity exploration of a gene sequence through stepwise, repeated, rounds of low mutational frequency.[6,7] Previous studies demonstrated that overall library functionality decreases exponentially with an increase in mean mutations per gene.[2] This implies that a high percentage of a cell population of a low mutational rate library is isolated due to the high retention of library functionality (Fig. 1). Furthermore, it has been hypothesized that the ability of high error rate libraries to sample all 20 amino acids at each mutated position of the sequence decreases as the mean number of mutations increases. This effective decrease in sampling of sequence space in libraries with high mutation frequencies is due to a finite library size. Therefore, rare improved mutants may not be represented within the reduced number of functional clones in high error-rate libraries.

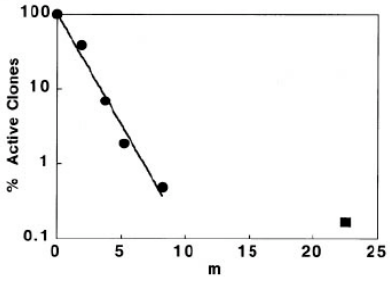

Figure 1: Mutational rate trend of library functionality[2]
Y axis: percent active/functional clones
X axis: mean mutations per gene

Despite evidence to the contrary, high mutation rate libraries have been used to enhance protein function.[2,8,9,10] In our research group, Daugherty *et. al* attempted to quantify the effect of the mutation rate on library functionality and protein improvement by analyzing the effect of mutations ranging from 1.7 to 22.5 mutations per gene.[2] The main conclusion of Daugherty's study is that high error-rate libraries yield more functional clones than would be predicted based on and extrapolation of the functionality of lower error-rate libraries: the library of 22.5 mutations per gene produced a 772-fold greater functionality than predicted by the trend of low error rate libraries. Isolated clones showed that libraries of both high and low error rate produced improvement in proteins. Specifically, Daugherty *in vitro* evolved an

anti-digoxin single chain antibody using a high error rate mutagenesis strategy that demonstrated comparable affinity improvements compared to using a low error rate mutagenesis strategy.[2]

In light of these results, this research aims to increase the understanding of the molecular basis of this phenomenon by investigating whether high error rate libraries recover paired or synergistic mutations. In other words, the improved performance of a high error rate library may be due to an increased number of coupled and favorable DNA mutations selected in favor of the protein function. These conserved, paired mutations would be clearly identified through a high-throughput FACS-based screening method. The FACS based screening method has demonstrated its ability to isolate high affinity binding proteins displayed on the surface of microorganisms.[11,12]

Hypothesis

The exponential decline in library function as mean mutations per gene increases, results from individual mutations that work independently of one another; however, a deviation from this trend may indicate that mutations are capable of synergy. This phenomenon may be supported by the notion that high error rate libraries have the capacity to produce combinations of mutations that are not as likely to exist in low error rate libraries. Similarly, synergistic combinations are not likely to be found in libraries created by low mutational rates because key mutation combinations are not able to form.

Experimental Approach

In order to test the above hypothesis, the following approach was employed: Three high error-rate libraries of an anti-digoxin single chain antibody (scFv) gene were constructed. This scFv antibody fragment was derived from the 26-10 monoclonal antibody.[13] The effects of high mutational rate by screening for function are evaluated: here, as established by Daugherty,[2] high error rate libraries should demonstrate a better than expected library functionality and therefore functionality should not decrease exponentially with increase in mean mutations per gene. We perform successive rounds of screening for high activity/functionality of clones and analyze sequences for apparent consensus mutations. If consensus mutations exist, these mutations are introduced onto the wild type sequence to determine whether combinations of mutations can work synergistically to improve protein

function. Synergistic effects can be denoted when combinations of mutations enhance function of the protein more than additive effects of single mutations.

Results

Three high error rate libraries of the anti-digoxin scAb were constructed and produced a mean number of mutations per gene for each of the three libraries ranging from 15–20. The ratio of observed to expected active (isolated) clones yielded an average of 366.33 (Table 1). This demonstrates that the constructed high mutation rate libraries deviated from the expected exponential functionality decline of lower mutation rate libraries: the high mutagenesis rate libraries did not exponentially decline in function.

Table One: Observed and expected fraction of isolated active clones from each library (L)

Presorted Library	**Observed fraction active**	**Expected fraction active**	**Ratio of observed to expected**
L1	1.7×10^{-3}	2.2×10^{-6}	772
L2	1.2×10^{-3}	1.7×10^{-5}	71
L3	4.1×10^{-4}	1.6×10^{-6}	256

Furthermore, the mutation distribution does not follow a Poisson distribution and has been overlaid around the mean of each library for visual confirmation (Fig. 2). The deviation of mutation distribution from Poisson statistics also predicts the observed deviation of higher mutational rate library functionality from the expected exponential decline.[14]

Distribution of Numbers of Mutations Per Gene

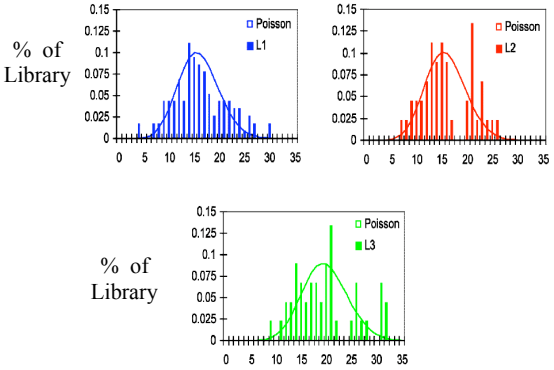

Figure 2: Mean mutations per gene for three libraries ranges from 15-20

Consensus Mutations

Pairs of mutations were found in each library at astonishingly high rates. Table Two denotes the percentages of the most commonly occurring mutations. Amino acid substitution mutations occurring often in pairs at positions 24 and 29 or 24 and 34 were found and are considered consensus paired mutations. The most mutated residue in all 3 libraries is at position AA 24, occurring in 51.66 % of the isolated clones.

Table Two: Consensus mutations of libraries

Library	Percentage mutated in SerH24	Second mutated residue	Percentage mutated in both
L1	58	PheH29	24
		MetH34	18
L2	62	GlnH6	19
		SerH25	24
		ThrH30	29
		PheH32	33
L3	35	SerH25	19
		MetH34	15

Consensus mutation pairs were incorporated into the wild type sequence to determine whether combinations of mutations can work synergistically to improve protein function.

Effect of combination and single consensus mutations:

Paired mutations demonstrate the highest relative affinity to digoxin. Single mutations demonstrate higher affinity than wild type (Fig. 3). For more precise affinity measurements toward digoxin antigen, dissociation constants, K_d must be verified through higher resolution kinetic experiments.

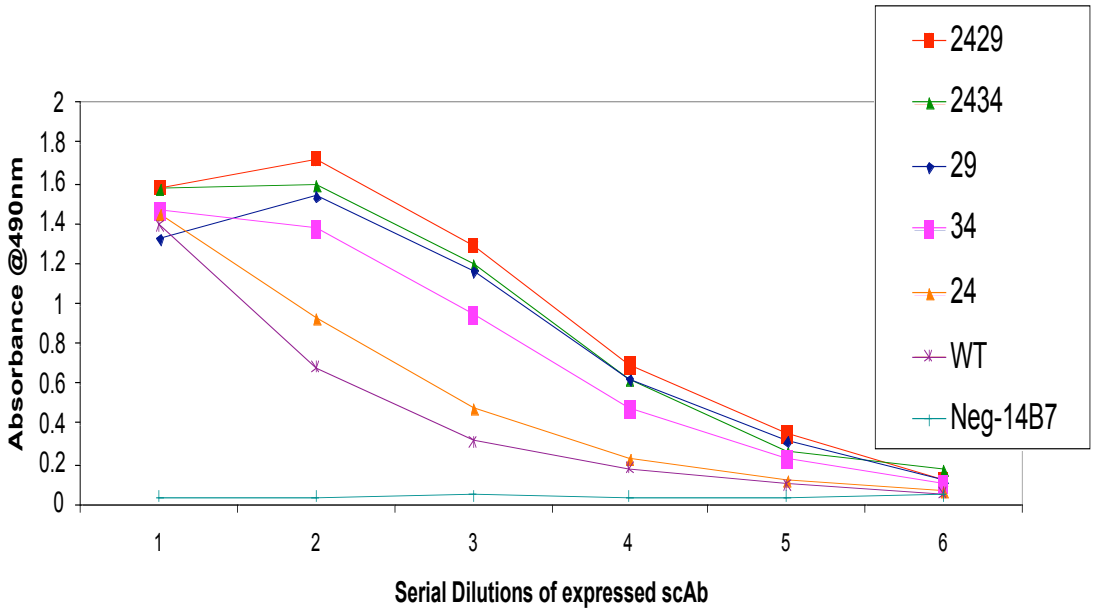

Figure 3: Preliminary ELISA data, demonstrating relative affinity to digoxin and compared to the relative affinity of the wild type protein.

Normalized amounts of protein expression from all mutants are greater than the wild type (Fig. 4). This is expected because the method used, screening of surface anchored protein libraries by FACS, selects in favor of well-expressed protein as well as affinity.[11] Combination mutants do not demonstrate a notable increase in expression over the single mutants. This implies that the coupled mutations do not work positively to improve expression.

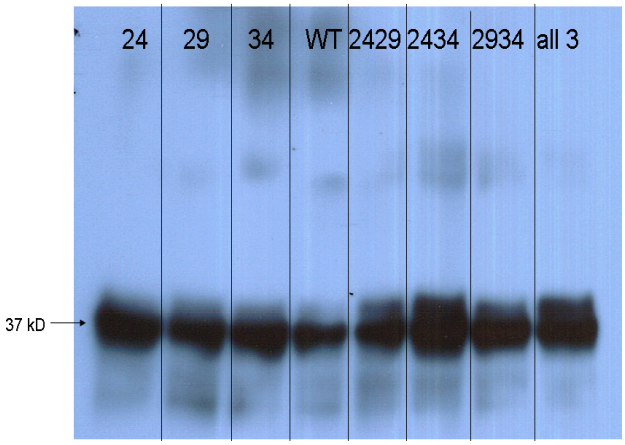

Figure 4: Normalized amounts of protein expression from each combination mutant, visualized using the western blot analysis technique.

Conclusions

The acquired data supports previous trends that high error rate libraries deviate from exponential decline in functionality. Our high error rate libraries also demonstrate higher functionality than expected. Preliminary results support the hypothesis that coupled mutations synergize to produce enhanced affinity. These results must be verified through higher resolution kinetic experiments using surface plasmon resonance (SPR) technology to determine the absolute dissociation constants (K_d). Preliminary results also demonstrate that high error rate libraries can be used to identify better expressing mutant proteins. However mutation coupling does not synergistically improve expression when compared to single mutations. This work may support future use of high error-rate library construction for enhanced *in vitro* protein evolution and the identification of mutations that, upon coupling, produce positive, gain-of-function proteins.

Materials and Methods

Specifically, the protein improved by methods of *in vitro* evolution is an anti-digoxin single chain antibody (scAb). The specific goal is to improve the original affinity of the scAb to digoxin. The method utilized for DNA randomization of the scAb is error prone PCR. Three high error rate libraries of the anti-digoxin scAb were constructed utilizing a high concentration of Mg^{2+} and a bias nucleotide distribution. The theoretical advantage of error PCR over other evolution methods is that it introduces mutations across the entire gene of interest. This error introduction is carried out via PCR amplification of the gene of interest

using a non-proofreading DNA polymerase and suboptimal conditions, i.e. variation of the mutation rate via Mg^{2+} concentration.

The error prone PCR product, or library, is ligated into an expression vector pBD16, with an ompA leader sequence ahead of the gene of interest under the control of an arabinose promoter. The leader sequence, lppOmpA encodes for an outer membrane protein and, when cells are induced to begin transcription, is responsible for the translocation and anchoring of the protein of interest to the outer membrane. The library was transformed into LGM194 *E. coli* cells. This library of cells was then labeled with the fluorescent digoxigenin-bodipy probe and screened flow cytometrically. Libraries were screened via successive rounds of fluorescent activated cell sorting (FACS) to identify and isolate improved clones. A total of 150 clones were selected from the isolated pool of cell libraries and sequenced. Sequences of each library were aligned to find consensus mutations, which were then analyzed for synergistic effects upon the wild type protein. Specifically, ELISAs and Westerns analysis were performed to examine affinity and expression changes due to mutations.

ELISA: Enzyme Linked Immuno-Sorbent Assay

Ninety-six well high affinity binding plates were coated with 200ng BSA-digoxin/well. FPLC purified and normalized mutant antibodies of 125ng/well were introduced to the antigen, detected using an HRP conjugated anti-His tag.

PAGE gel and Western Blot Protein Analysis

Whole cell extracts of constructed mutants induced to express single chain antibody, probed with an HRP conjugated anti-Kappa light chain antibody.

References

1. Walsh, G. Biopharmaceutical Benchmarks, 2003. *Nat. Biotechnol.* 2003; **21**: 865–870.
2. Daugherty, P.S., Gang, C., Iverson, B.L. & Georgiou G. *Proc. Natl Acad. Sci USA* 2000; **97**: 2029–2034.
3. Oue S, Okamoto A, Yano T, Kagamiyama H. *J Biol Chem.* 1999; **274**:2 344–2349.
4. Cherry J R, Lamsa M H, Schneider P, Vind J, Svendsen A, Jones A, Pedersen A H. *Nat Biotechnol.* 1999; **17**: 379–384.
5. Zaccolo M, Gherardi E. *J Mol Biol.* 1999; **285**: 775–783.
6. Arnold F H. *Nat Biotechnol*. 1998; **16**: 617–618
7. Arnold F H. *Proc Natl Acad Sci USA.* 1998; **95**: 2035–2036.
8 Christians F C, Scapozza L, Crameri A, Folkers G, Stemmer W P. *Nat Biotechnol.* 1999; **17**: 259–264.
9. Christians F C, Loeb L A. *Proc Natl Acad Sci USA.* 1996; **93**: 6124–6128.
10. Martinez M A, Pezo V, Marliere P, Wain-Hobson S. *EMBO J.* 1996; **15**: 1203–1210.
11. Daugherty P S, Chen G, Olsen M J, Iverson B L, Georgiou G. *Protein Eng.* 1998; **11**: 825–832.
12. Daugherty P S, Olsen M J, Iverson B L, Georgiou G. *Protein Eng.* 1999; **12**: 613–621.
13. Francisco J A, Campbell R, Iverson B L, Georgiou G. *Proc Natl Acad Sci USA.* 1993; **90**: 10444–10448.
14. Drummond DA, Iverson BL, Georgiou G, Arnold FH., *J Mol Biol.* 2005 Jul 22; **350(4)**: 806-16.