Multiplex evolution of antibody fragments utilizing a yeast surface display platform

Oh, Eun Joong; Liu, Rongming; Liang, Liya; Freed, Emily F.; Eckert, Carrie Ann; Gill, Ryan T.

Published in:
ACS Synthetic Biology

Link to article, DOI:
10.1021/acssynbio.0c00159

Publication date:
2020

Document Version
Peer reviewed version

Citation (APA):
Multiplex evolution of antibody fragments utilizing a yeast surface display platform

Eun Joong Oh¹, Rongming Liu¹, Liya Liang¹, Emily F. Freed¹, Carrie A. Eckert¹,², and Ryan T. Gill¹,³

¹Renewable and Sustainable Energy Institute (RASEI), University of Colorado Boulder, CO 80303, USA
²National Renewable Energy Laboratory (NREL), Bioscience Center, Golden, CO 80401, USA
³Novo Nordisk Foundation Center for Biosustainability, Danish Technical University, Lyngby, Denmark

Correspondence should be addressed to Ryan T. Gill (rtg@biosustain.dtu.dk).

Keywords CRISPR/Cas9, Saccharomyces cerevisiae, yeast display, fluorescent-activated cell sorting, antibody engineering.
Abstract

Advances in high-throughput synthetic biology technologies based on the CRISPR/Cas9 system have enabled a comprehensive assessment of mutations conferring desired phenotypes, as well as a better understanding of genotype-phenotype correlations in protein engineering. Engineering antibodies to enhance properties such as binding affinity and stability plays an essential role in therapeutic applications. Here we report a method, multiplex navigation of antibody structure (MINAS), that combines a CRISPR/Cas9-based trackable editing method and fluorescent-activated cell sorting (FACS) of yeast-displayed libraries. We designed mutations in all of the complementarity-determining and framework regions of a well-characterized scFv antibody and mapped the contribution of these regions to enhanced properties. We identified specific mutants that showed higher binding affinities up to 100-fold compared to the wild-type. This study expands the applicability of CRISPR/Cas9-based trackable protein engineering by combining it with a surface display platform.
The ability to engineer antibodies with enhanced properties such as binding affinity and stability has had a significant impact on therapeutic applications of antibodies. Antibody engineering is often performed by targeted mutagenesis or directed evolution using high-throughput screening of phage or yeast display libraries. Because strategies for generating sequence diversity by random mutagenesis can substitute only limited numbers of amino acids in the targeted position, these approaches might miss other regions that may also affect function that is not easily predicted. Generation of libraries for searching the mutational landscape has been demonstrated using a PCR-based approach with mutagenic oligos. However, such methods have limitations. Specifically, when a larger backbone plasmid was used to construct libraries, transformation efficiencies were much lower. In addition, increasing the size of the insert to be mutated may cause challenges with sequencing depth and cost. Also, if the backbone plasmid has multiple nicking sites (BbvCI restriction sites) with opposite orientations, both strands of the plasmid will be nicked during the first BbvCI digest, resulting in the degradation of the whole plasmid by exonuclease III.

Many of these limitations are overcome by using CREATE (CRISPR-EnAbled Trackable genome Engineering), an advanced approach based on CRISPR/Cas9 editing and barcode tracking that enables multiplex editing and mapping at the genome scale. Recent studies employed this approach for high-throughput genome-wide engineering in yeast. These studies have led to new insights into genotype-phenotype correlations in yeast as well as bacteria. Here, we report the development of a method, multiplex navigation of antibody structure (MINAS), that combines CRISPR/Cas9-based trackable editing and fluorescent-activated cell sorting (FACS) of yeast display libraries. Array-based synthesis of editing cassettes allows for the pooled cloning of plasmid libraries. Each cassette contains: 1) a gRNA targeting the antibody, and 2) a dsDNA repair template (homology arm) that introduces designed mutations at the target site (Fig. 1a).
Following editing with the libraries, the frequency of the designed mutations can be tracked by deep sequencing of the editing plasmids that act as genetic barcodes (Fig. 1b). We used this approach for multiplex editing of the complementarity-determining (CDR) and framework (FWR) regions of a single-chain variable fragment (scFv) antibody; however MINAS can be applied to any region in any antibody to introduce up to hundreds of thousands of mutations and map the effects of those mutations on desired phenotypes.

The ability to edit with high precision is one of the essential factors required in order to apply MINAS to antibody engineering. Thus, we measured editing efficiencies of MINAS cassettes with varying homology arm lengths (48-68 bp) and varying distances between the PAM and target site (13-31 bp). For these tests, we inactivated the ADE2 gene by introducing point mutations to convert codons 10 and 11 to TAATAA stop codons. The editing results were monitored by a white to red colony color change and confirmed by Sanger sequencing. The editing efficiency was 99.1% when the homology arm was longer (68 bp) and the distance between the PAM and target site was shorter (13 bp), consistent with previous studies. The effect of the distance between the PAM and target site on editing efficiency was more significant than the length of the homology arm (Fig. 1c). Also, we observed a significant increase in transformation efficiency when shorter distances between the PAM and target site were used, suggesting that high-efficiency editing and transformation are possible when cassettes are designed using these criteria (Fig. 1c).

We next tested the editing efficiencies in two different antibody expression systems when MINAS cassettes with various distances between the PAM and target site (13-66 bp) were used. We used a well-characterized antibody-antigen system, the 4-4-20 scFv antibody which binds fluorescein. First, the 4-4-20 antibody expression cassette was cloned into the yeast integration vector pRS404, and the resulting plasmid was integrated into the TRP1 locus of Saccharomyces cerevisiae strain EBY100. Second, a single-copy
plasmid expressing the 4-4-20 antibody, pCT302, was transformed into the EBY100 strain to express 4-4-20 from a plasmid rather than integrating it into the genome. Next, we synthesized editing cassettes targeting two different codons. After transforming the editing plasmids into yeast harboring each of the 4-4-20 antibody expression platforms, the editing efficiencies were assessed by sequencing both the plasmid barcodes and the targeting sites from 20 to 27 randomly chosen colonies. For the first editing cassette, which targeted codon 67 in the variable light chain (V\(_L\)) region, the editing efficiencies in the integration platform ranged from 7-74%, whereas those in the single-copy plasmid platform ranged from 4-100% depending on the distance between the PAM and target site. Although the editing efficiencies for the second editing cassette, which targeted codon 5 in the variable heavy chain (V\(_H\)) region, were lower than those for the first target, the single-copy plasmid platform exhibited higher editing efficiencies than the genome integration platform overall (Fig. 1d). Therefore, the single-copy plasmid system was employed for the expression of the 4-4-20 antibody.

To test the feasibility of MINAS as a tool for antibody engineering, we designed 4,580 mutations targeting all V\(_L\) (112 aa) and V\(_H\) (117 aa) domains in the 4-4-20 antibody (Supplementary Fig. 1). A full codon saturation mutagenesis library was introduced by replacing the target residue with the remaining 19 amino acids using the most frequent codons; a stop codon was also included as a negative control. The MINAS libraries were constructed in *Escherichia coli* as previously described\(^9\), and the accuracy of plasmid libraries was assessed by sequencing the PCR-amplified editing cassettes from randomly picked colonies. We observed that 55% (23/42) matched the cassette design with 100% identity, 12% contained a 1-2 bp insertion/deletion (indel) (5/42), and 33% (14/42) were chimeric cassettes. All the perfectly matched cassettes were unique, indicating that the libraries were effectively prepared with mutational diversity. Discrepancies in the plasmid libraries may have arisen during synthesis, PCR-amplification, or Gibson assembly. The
library plasmids were extracted and re-transformed into *S. cerevisiae* EBY100 harboring pCT302.

We first demonstrated the MINAS approach by mapping mutations that confer increased binding affinity. The resulting yeast libraries were incubated with biotinylated fluorescein and then exposed to a competitor, 5-aminofluorescein, causing a loss in antigen-binding activity. Cells displaying 4-4-20 mutants that bound fluorescein more strongly than wild-type cells after the binding competition were isolated by FACS (Fig. 1b). Antibody-bound fluorescein was visualized using PE dye, and the amount of surface-displayed antibody was separately visualized using Alexa Fluor 488 (c-Myc). We reasoned that designer mutations conferring improved affinity would outcompete the 5-aminofluorescein and thereby be enriched through repeated FACS sorting of the top 0.1–5% of antigen-binding c-Myc positive clones. Sequencing the barcode of the editing cassettes before and after sorting in the presence of the competitor enabled parallel tracking of mutants with enrichment and mapping of their contribution to desired phenotypes. To ensure high editing efficiency and adequate mutant identification, only reads that matched a designed cassette with an identity of greater than 99% were used. Deep sequencing of the initial yeast population that was transformed with MINAS cassettes prior to sorting showed 78–81% of the designs were contained in the library pools across two biological replicates. Sorting the pools with increasing stringency of sort gates resulted in sharply increased enriched populations of library cells (Fig. 2a). After five rounds of selection (Fig. 2b), plasmids in the enriched populations were isolated, and the regions containing the barcode of the editing cassettes were sequenced.

When mapped at single amino acid resolution, we observed significantly enriched mutations in both CDR and FWR regions across all targeted areas based on the average enrichment scores obtained in the two biological replicates (Fig. 2c and d). Of the two top mutations we identified, the K55W mutation in CDR2 of the *V<sub>L</sub>* domain was previously
unreported, and the W108M mutation in the framework region of the \( V_{H} \) domain was described previously\(^{10}\) (Fig. 2j). We individually reconstructed these two 4-4-20 variants and tested their binding affinities using the same competition process as described above. Both reconstructed variants exhibited more than 16-fold higher binding/display values, consistent with their respective enrichment profiles (Fig. 2e). Next, we determined \( K_D \) values for individual variants by measuring binding curves as described by Chao et al\(^{11}\). In addition to wild-type 4-4-20, K55W, and W108M, we assayed the 4 M5.3 variant from Boder et al\(^{10}\). As expected, the 4 M5.3 variant, with fourteen accumulated mutations, showed the highest affinity (\( K_D < 0.014 \) nM) among the variants, and the K55W (\( K_D = 0.030 \) nM) and W108M (\( K_D = 0.035 \) nM) mutants achieved roughly a 90- to 100-fold improvement in affinity as compared to the wild-type (\( K_D = 3.055 \) nM) (Supplementary Fig. 2, Supplementary Table 1, Fig. 2f). Amino acid substitutions in proteins can change multiple biochemical properties of the proteins\(^{12}\). For example, mutation(s) may lead to stabilization of intra-loop interactions or free energy differences, resulting in binding affinity improvements\(^{13-15}\). Here we show that both a residue in the CDR2 region, which constitutes a paratope, and a residue in the framework region adjacent to CDR3 were influential for affinity. These results suggest that although CDR3 typically contributes the most to antigen specificity\(^{16}\), broader mutational strategies including both CDR and FWR regions are critical in antibody engineering. Furthermore, these results demonstrate the potential ability of the MINAS approach to perform rapid mapping of target antibodies for the identification of novel mutations.

Exposure to low pH conditions is unavoidable for antibody purification and leads to beneficial effects on viral clearance in the production process. However, when exposed to acidic conditions, antibodies can undergo conformational changes\(^{17}\). Thus, enhanced low pH stability is an important trait for industrial antibody production. Therefore, we next used MINAS to improve the stability of 4-4-20 under acidic conditions. Yeast expressing
4-4-20 antibody libraries were incubated in buffers with a pH range of 2–3 for 1 hour at room temperature before antigen binding, and labeling reactions and FACS proceeded in the same manner as for binding competition. We found that there was no significant difference in low pH stability between pre-selection and after five rounds of selection across two biological replicates. Although we did not see an obvious difference in pre- and post-selection libraries, we nonetheless did enrichment analysis in an attempt to map mutations that might confer low pH stability (Fig. 2g and h). To confirm the enrichment analysis, we reconstructed the top three variants (Q27L, S48M, and G109Q) from the fifth-round of selection and examined their binding affinities under acidic conditions by measuring binding/display values. Although we did not directly measure stability with the reconstructed variants, it has been shown that measuring binding affinity by yeast display correlates very well to other measurements of stability done with soluble scFvs\(^1\). Among the three variants, the S48M and the Q27L mutants, which were the first and second most enriched in both replicates, showed 2- and 1.4-fold higher binding affinity compared to the wild-type, respectively, consistent with their enrichment scores (Fig. 2i and j). Although we could not observe a significant increase in enriched populations from the library cells during FACS, the dominant selection winners exhibited the desired phenotypes. These identified mutations might be involved in thermodynamic stability, perhaps through resistance to acid aggregation\(^1\).

Recent advances in next-generation sequencing and CRISPR/Cas9-based editing have allowed more thorough investigations of target proteins at single-nucleotide resolution\(^2\)–\(^4\). Yeast surface display enables quantitative library analysis and screening by flow cytometry and is a powerful platform for engineering antibodies\(^5\). In this study, we demonstrate a novel CRISPR/Cas9-based approach for rapid and trackable antibody engineering based on yeast surface display. By combining CRISPR/Cas9 gene editing with multiplexed oligo synthesis and deep sequencing, we were able to evaluate the
contribution of each mutation to the desired phenotypes of the antibody with trackable high-throughput mapping by a simple workflow.

Although we were able to track the enrichment of library variants through multiplex fitness mapping for antibody engineering, some limitations were identified in our approach. It is necessary to overcome the low editing efficiency on a library scale and unintended mutations arising during library preparation by improving the CRISPR/Cas9 machinery\textsuperscript{24,25} and design/assembly process\textsuperscript{12,26}. Sorting the pools with increasing stringency of sort gates might result in the convergence of mutation. However, using a single sort, it is not easy to identify real positive variants because there are too many variants to test (Supplementary Table 2). To further enhance the MINAS approach, magnetic-activated cell sorting (MACS) can be used as the first step in the workflow to reduce false-positive mutants. Also, modulating parameters for selection pressures, including sort gates and competition times, could enable quantification with enhanced diversity. Despite these limitations, CREATE libraries can be built with >100,000 single nucleotide polymorphisms (SNPs) which can be mapped to identify mutations that confer phenotypes of interest in parallel\textsuperscript{6}. Using these larger libraries, MINAS would allow a full codon saturation mutagenesis of an entire antibody regardless of the size of the antibody. Extension of this technology might be adopted for high-throughput mapping of other antibody-antigen interactions of interest such as epitope mapping.

**Supplementary Information**

Contains the following: supplementary methods, yeast display construct, binding titration curves of yeast displayed 4-4-20 and mutant scFvs, $K_D$ measured using flow cytometry, the number of mutations with Log2 enrichment > 0 after selection, strains and plasmids used in this study, library plasmid construction primers, gBlock sequences used in this study.
Acknowledgments
We thank Karen Helm (Cancer Center Support Grant (P30CA046934) and the Skin Diseases Research Cores Grant (P30AR057212)) for assistance and support with FACS runs performed in this study. This work was supported by the US Department of Energy (Grant DE-SC0018368) and Inscripta, Inc.

Author contributions
E.J.O. and R.T.G. designed research. E.J.O. performed research and analyzed data. Sequencing data analysis was done by R.L. and E.F.F. L.L. helped with cloning. C.A.E. aided the design of experiments. E.J.O. and R.T.G. wrote the manuscript. E.F.F. and C.A.E. reviewed and edited the manuscript. All authors read and approved the paper.

Competing interests
The author R.T.G. has financial interest in the company Inscripta, Inc., which is commercializing the technology used in this manuscript. The remaining authors declare no competing interests.
References


**Figure legends**

**Figure 1.** MINAS Library construction and selection. (a) Design of the MINAS cassette. (b) The workflow of the strategy. A designed cassette is composed of a homology arm, including the target site mutation, a synonymous PAM mutation (SPM), and the corresponding gRNA. The editing cassettes are cloned in multiplex into a backbone plasmid harboring Cas9. Yeast strains expressing the target antibody are transformed with the library plasmids. Yeast displaying antibodies with various affinities to the antigen are isolated and enriched by iterative rounds of FACS. Following selection, the library plasmids are PCR-amplified, and the frequency of each plasmid before and after selection is determined by deep sequencing. Enrichment scores for each mutant caused by the corresponding cassette are calculated as the log₂ ratio of frequencies between post-selection and pre-selection. (c) Editing efficiency and the number of CFUs/µg DNA using editing cassettes with various homology arm lengths and distances between the PAM and target site. (d) Comparison of editing efficiency between genome integration and single-copy plasmid platforms for 4-4-20 expression.

**Figure 2.** Enrichment analysis of library cells for improved binding affinity (a-f) and low pH stability (g-i). (a) The median fluorescence intensity of the binding signal (MFI\text{Binding}) was normalized to the median fluorescence intensity of the display signal (MFI\text{Display}). The normalized median fluorescence intensity (binding/display = MFI\text{Binding}/MFI\text{Display}) showed progressive enrichment of mutants via FACS for improved binding affinities. (b) Histogram of library fluorescein binding, showing pre-selection (black) and after five rounds of selection for improved binding affinities (blue). (c, g) Mapping of enrichment across the
positions targeted in the 4-4-20 antibody. (d, h) Heatmap showing the enrichment of MINAS cassettes after five rounds of selection. All amino acid residues found in two biological replicates are shown with the average of the enrichment scores. (e) The reconstructed K55W and W108M variants exhibited higher normalized median fluorescence intensity (binding/display = MFI_{binding}/MFI_{display}) values with binding competition than the wild-type, corresponding to their respective $K_D$ values (f). (i) For the stability of the 4-4-20 antibody under acidic conditions, among the top three variants, the S48M and the Q27L mutants showed higher binding/display values compared to the wild-type. The values represent the means of three independent experiments. (j) Structure of the 4-4-20 antibody in complex with fluorescein (green). The highly enriched mutations for both binding affinity and low pH stability are shown in red (CDR) and blue (FWR). PDB code 1FLR.
For Table of Contents Use Only

Editing cassette → Bind to antigen → Selective enrichment

Target site mutation → gRNA

FACS
Figure 2.