DTU Library

SAMURAI (Solid-phase Assisted Mutagenesis by Uracil Restriction for Accurate Integration) for antibody affinity maturation and paratope mapping

Hu, Francis Jingxin; Lundqvist, Magnus; Uhlén, Mathias; Rockberg, Johan

Link back to DTU Orbit

SAMURAI (Solid-phase Assisted Mutagenesis by Uracil Restriction for Accurate Integration) for antibody affinity maturation and paratope mapping

Francis Jingxin Hu[1], Magnus Lundqvist[1], Mathias Uhlén[1,2,3] and Johan Rockberg[1,*]

[1]KTH - Royal Institute of Technology, Department of Protein Science, 106 91 Stockholm, Sweden, [2]KTH - Royal Institute of Technology, Science for Life Laboratory, Solna 171 65, Sweden and [3]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2970 Hørsholm, Denmark

ABSTRACT

Mutagenesis libraries are essential for combinatorial protein engineering. Despite improvements in gene synthesis and directed mutagenesis, current methodologies still have limitations regarding the synthesis of complete antibody single-chain variable fragment (scFv) genes and simultaneous diversification of all six CDRs. Here, we describe the generation of mutagenesis libraries for antibody affinity maturation using a cell-free solid-phase technique for annealing of single-strand mutagenic oligonucleotides. The procedure consists of PCR-based incorporation of uracil into a wild-type template, bead-based capture, elution of single-strand DNA, and *in vitro* uracil excision enzyme based degradation of the template DNA. Our approach enabled rapid (8 hours) mutagenesis and automated cloning of 50 position-specific alanine mutants for mapping of a scFv antibody paratope. We further exemplify our method by generating affinity maturation libraries with diversity introduced in critical, nonessential, or all CDR positions randomly. Assessment with Illumina deep sequencing showed less than 1% wild-type in two libraries and the ability to diversify all CDR positions simultaneously. Selections of the libraries with bacterial display and deep sequencing evaluation of the selection output showed that diversity introduced in non-essential positions allowed for a more effective enrichment of improved binders compared to the other two diversification strategies.

INTRODUCTION

Monoclonal antibodies (mAbs) have become indispensable tools in therapeutics and diagnostics, and according to some estimates, their worldwide market can reach ∼$125 billion in the next few years (1). Several approaches have been developed in the last few decades for the generation of mAbs, such as hybridoma technology (2), transgenic mice (3,4) and various display technologies for selection of recombinant binders (5–9). Although antibody engineering can involve improvement of properties with regards to effector function and stability (10,11), perhaps the most commonly modified property is binding affinity (12–14). This factor is of particular importance for therapeutic antibodies for which high affinity toward the antigen is critical for increasing efficacy, reducing dosages and reducing side effects (15,16).

Numerous procedures have been used for antibody affinity maturation which can roughly be divided into structure-based approaches, where antibody-antigen complexes are analyzed in detail, and modeling aids the engineering of a refined structure or combinatorial methods which are needed when little or no structural information is available. *In vitro* display techniques are attractive and widely implemented because highly diverse libraries can be generated and screened by panning or cell sorting (13). Error-prone PCR can be used to diversify the antibody gene and create libraries swiftly. This approach, however, suffers from the introduction of mutations randomly throughout the gene, which can lead to detrimental mutations in the framework regions of the antibody. Therefore, targeted mutagenesis focusing only on the complementarity-determining regions (CDRs) can be advantageous. The structures of antibody-antigen complexes indicate that the majority, if not all of the six CDRs, may contribute to antigen-binding in most cases (17,18). Thus, it would be attractive to mutate multiple CDRs simultaneously for a proper *in vitro* antibody affinity maturation, to allow for possible additive or synergistic effects from all loops. Furthermore, mutating all six CDRs simultaneously mimics nature's somatic hypermutation, which tends to spread diversity to also the more peripheral residues, not necessarily in direct contact with the antigen (19,20).

*To whom correspondence should be addressed. Tel: +46 8 790 9988; Fax: +46 8 5537 8481; Email: johanr@biotech.kth.se

The various approaches that can be used to diversify specific selected CDRs include CDR walking (21), hot-spot mutagenesis (22), look-through mutagenesis (23), massive mutagenesis (24). More examples include Kunkel based annealing of oligonucleotides (25), Pfunkel (26) and nicking mutagenesis (27). However, in these reported methods, either one CDR or site is mutated at a time, or a maximum up to four CDRs/sites have been mutated simultaneously (18,21,25–27). To successfully mutate all six CDRs, typically multiple mutagenesis rounds are necessary. Furthermore, some of these methods also require preparation of uracil-containing dsDNA from *E. coli*, which can also be problematic as the yield, and uracil-incorporation efficiency *in vivo* can vary.

In this paper, we describe a novel method addressing the need for swift construction of antibody scFv affinity maturation libraries using solid-phase mutagenesis. This mutagenesis method, SAMURAI (Solid-phase Assisted Mutagenesis by Uracil Restriction for Accurate Integration), is based on the *in vitro* creation of single-stranded uracil-containing DNA wild-type template, using an uracil-incorporating high-fidelity polymerase and biotinylated oligonucleotides for solid-phase elusion (i.e. displacement) of complementary strands and straightforward purification between steps. Subsequent site-directed mutagenesis of the CDR regions, by annealing and extension of mutagenic oligonucleotides, and following enzymatic degradation of the uracil template, allowed for mutation of up to all six CDRs in a single step while minimizing the risk of wild-type gene after mutagenesis.

To identify the scFv residues that were crucial for binding, SAMURAI was used to generate a complete set of point mutations for an alanine scan of all CDR residues. This scan was done in combination with automated high-throughput solid-phase cloning (28,29) for rapid (8 h) mutagenesis and cloning to expression vectors of all 50 variants. SAMURAI was subsequently used to quickly create three combinatorial libraries based on the crucial positions that were identified with the alanine scan. Affinity selection using bacterial display and flow cytometry was then used to isolate affinity-matured clones with increased affinity. Deep sequencing was applied to both analyze the combinatorial libraries as well as the affinity maturation process.

Our presented methodology shows a convenient and versatile approach to create specific user-defined site-directed mutagenic libraries. It allows for multiple mutations in a single step with minuscule levels of wild-type clones and exemplifies an effective screening/selection rationale using flow cytometry. Finally, in addition to creating tailored antibody affinity maturation libraries, SAMURAI can also be envisioned to be used to develop other used-defined mutagenesis libraries employed in mutational studies, such as deep mutational scanning (30).

MATERIALS AND METHODS

Solid-phase mutagenesis and cloning for libraries and single mutants

The preparation of wild-type template (31) (VH-linker-VL) used a forward biotinylated primer, which annealed at the start of framework 1 (FR1) of VH and a reverse primer, which annealed at FR3 in VL, upstream of CDRL3, omitting CDRL3. Primers used for amplification were '2600 TegBio' and '2602 Alexa647'. See supplementary information for sequences of all primers that were used in the study. The template gene was amplified with Phusion U (ThermoFisher Scientific, Waltham, MA, USA), which is able to incorporate dUTP and read through uracil present in DNA templates, with the nucleotides A, G, C and U using a standard Phusion protocol. The following cycling conditions were used: 98°C initial denaturation; 30 cycles of 98°C denaturation (8 s), 68°C annealing (25 s), 72°C extension (30 s); 72°C final extension. The biotinylated uracil-containing wild-type gene was immobilized using DynaBeads m280 Streptavidin (ThermoFisher Scientific, Waltham, MA, USA); 20 μl of bead suspension (with bead concentration 10 mg/ml) was used. These beads were washed twice with 2× binding/wash (B/W) buffer (10 mM Tris–HCl (pH 7.5), 1 mM EDTA, 2 M NaCl, 0.1% Tween 20). The beads were re-suspended in 50 μl 2× B/W buffer and added to 50 μl of PCR product containing uracil incorporated wild-type scFv gene. The immobilization process took place on a rotamixer for 30 min and beads were washed three times with 1× TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). The antisense strand was eluted by incubation of the beads with 50 μl 0.15 M NaOH for 10 min and the eluate was discarded.

The beads with immobilized single-stranded wild-type gene were dissolved in 50 μl annealing/extension/ligase mix containing: 32μl H$_2$O, 10 μl 5× HF Phusion buffer (final conc. 1× HF), 5 μl 2mM dAGCT (final conc. 200 μM dAGCT), 0.5 μl Phusion U Hot-Start (2 U/μl), 3 μl 10 mM ATP (final conc. 600 μM ATP), 1 μl 9N ligase (40 U/μl, New England Biolabs, Ipswich, MA, USA) and 5 pmol of the 5′ phosphorylated mutagenic oligonucleotide mix. For simultaneous annealing, all mutagenic oligonucleotides corresponding to all six CDRs were added at once (5 pmol was the total amount of primer and primers were mixed in equimolar proportions). The annealing/extension/ligase mix was then subjected to the following cycling conditions for the synthesis of the mutagenic antisense strand: 98°C initial denaturation (30 s) and 10 cycles of 98°C denaturation (8 s), 56 or 63°C annealing (1 min), 72°C extension (1 min); 72°C final extension (10 min). For sequential annealing, only oligonucleotides corresponding to CDRH1 were initially added and more oligonucleotides for subsequent CDRs were added to the mix after every 2 cycles. Four additional cycles were used after the addition of the CDRL3 primer mix. Final total amount of primers was 5 pmol and primers were mixed in equimolar proportions.

In the alanine scan, 50 mutagenesis reactions were carried out in parallel. In each reaction, 5 pmol of a single mutagenic oligo carrying the mismatch alanine codon was added and five PCR cycles were used. To complete the synthesis of the mutagenic anti-sense strand, 5 pmol of an oligonucleotide corresponding to CDRL3 wild-type (primer 'sc1_L3_WT') was added to the mix followed by five cycles additional and a final extension. Note that no CDRL3 wild-type oligo was added if the mutagenesis targeted a position in L3. See supplementary information regarding primer and template sequences.

To degrade the remaining uracil-containing template, 1 U USER enzyme (New England Biolabs, Ipswich, MA, USA) was added directly to the enzyme mix and left to incubate at 37°C for 30 min. The mixture was then used as the template for a regular PCR (A, C, G and T nucleotides) using the standard high fidelity Phusion DNA polymerase (New England Biolabs, Ipswich, MA, USA) with standard nucleotides, incapable of incorporating uracil. The same PCR program was used as for the initial amplification of the wild-type gene. Primer '2600 TegBio' and '2601' were used in this amplification. The libraries were then cloned into the Gram-positive Staphylococcal display vector pSCEM2 (32) with the restriction enzymes NotI and SgsI (ThermoFisher Scientific, Waltham, MA, USA). Vector sequence for pSCEM2 can be found in the supplementary information. The generation of individual single alanine mutants was done by using the same restriction enzymes, but cloning into the display vector used automated solid-phase cloning (28) with Magnatrix 8000 (NorDiag AS, Oslo, Norway). To create the libraries, transformation of *Staphylococcus carnosus* TM300 was carried out with a previously described protocol (32) and 100 electroporations were used for each of the three libraries.

Flow cytometry sorting of *S. carnosus* libraries

An aliquot (50 µl) of the library corresponding to a number of cells that was ten times the size of the library was inoculated in TSB with 10 µg/ml chloramphenicol and cultivated at 37°C for 16 h at 150 <rpm. An aliquot (50 µl) of the overnight culture was washed with 1000 µl PBS-P (PBS with 0.1% Pluronic acid, F-127 Sigma-Aldrich) and pelleted by microcentrifugation (4000g, 4°C, 7 min) and re-suspended in 500 µl PBS-P containing biotinylated extracellular domain of HER2 human recombinant protein with his-tag, hereafter referred to as the antigen (SinoBiological 10004-H08H), and incubated at room temperature under gentle mixing for 3 h. HER2 was biotinylated using a FluoReporter Mini-biotin-XX Protein Labeling Kit (Invitrogen) following the manufacturer's instructions. The cells were thereafter washed with 1000 µl PBS-P and re-suspended in 200 µl PBS-P containing streptavidin – R-phycoerythrin Conjugate (rPE) (1 ng/ml; ThermoFischer) and HSA–Alexa 647 (40 nM), followed by incubation on ice in the dark for 30 min. HSA was labeled in-house with Alexa Fluor 647 succinimidyl ester (Invitrogen, Carlsbad, CA, USA) according to the supplier's recommendations. HSA-Alexa 647 binds to the ABP (albumin binding protein) which is fused to the C-terminal of the surface expressed scFv. A positive signal from HSA-Alexa 647 indicates surface expression of the scFv-ABP fusion protein. The cells were finally washed with 1000 µl PBS-P and re-suspended in 1000 µl PBS-P. The library was sorted using MoFlo Astrios (Beckman Coulter Inc., Brea, CA, USA). The gated cells were sorted into 200 µl TSB and incubated at 37°C for 2 h under gentle shaking and thereafter spread onto agar plates containing 10 µg/ml chloramphenicol.

Flow cytometric analysis of single *S. carnosus* clones

Staphylococcal cells expressing scFvs were cultivated in TSB medium with 10 µg/ml chloramphenicol for 16 h at 37°C and 150 rpm. Cells were washed with 200 µl PBS-P (PBS with 0.1% Pluronic acid, F-127 Sigma-Aldrich) and pelleted by microcentrifugation (4000g, 4°C, 7 min) and re-suspended in 100 µl PBS-P containing biotinylated antigen and incubated at room temperature under gentle mixing for 2 h. After incubation, the cells were washed with 200 µl PBS-P and re-suspended in 100 µl PBS-P containing streptavidin—rPE (1 ng/ml; Life Technologies, Carlsbad, CA, USA) and HSA–Alexa 647 (40 nM), followed by incubation on ice in the dark for 30 min. The cells were then washed with 200 µl PBS-P and re-suspended in 200 µl PBS-P, and analyzed by flow cytometry using a Gallios Flow Cytometer (Beckman Coulter Inc., Brea, CA, USA).

Production of scFv binders as soluble molecules in *E. coli* and Biacore affinity measurement

The best binder from each library was re-cloned and expressed as soluble scFvs in *Escherichia coli* as previously described (31). Chosen clones were amplified with PCR primers containing SfiI and AvrII restriction sites and digested with the corresponding enzymes, cloned into an *E. coli* expression vector pHP2-15 (33) and transformed into TOP10 *E. coli* cells for expression of soluble molecules. Colonies were inoculated in Terrific Broth medium containing 0.25 mM isopropyl thiogalactoside (IPTG) and soluble scFv produced overnight at 30°C. The proteins were purified from the periplasmic fraction making use of the polyhistidine tag and HisPur™ Ni-NTA Magnetic Beads (Thermo Fisher) following the manufacturer's recommendation. Sample elution was performed in PBS containing 500 mM imidazole at pH 8. The eluted scFvs were buffer exchanged using PD-10 columns (GE Healthcare). For Biacore affinity measurement, Her2 extracellular domain was diluted to 4 µg/ml in PBS-T (PBS with 0.05% Tween20) and immobilized onto a Biacore CM5 sensor chip (GE Healthcare). The soluble expressed scFvs were diluted in PBS-T to the following concentrations: 100, 50, 25, 12.5, 6.25 nM (wildtype); 50, 25, 12.5, 6.25, 3.125 nM (binder from random library); 25, 12.5, 6.25, 3.125, 1.56 nM (binder from enhancing library) and subjected to an association time of 300 sec and dissociation time of 700 sec. The dissociation equilibrium (K_D), the association rate (k_a), and the dissociation rate (k_d) constants were calculated with the Biacore T200 Evaluation Software 2.0, assuming a Langmuir 1:1 interaction model.

Library preparations for deep sequencing

The pSCEM2 plasmid carrying the scFv gene was extracted from 1 ml of overnight cultures of *S. carnosus* using a Miniprep kit (Qiagen) and following the manufacturer's instructions but with one important remark: P2 buffer lysis was carried out for 5 min. A region of the scFv (643 bases in length), covering all the CDRs, was PCR amplified with primers including adapters for Illumina sequencing using a Phusion Hot-Start polymerase (Thermo Fisher Scientific) in 50 µl reactions and the following PCR protocol: initial denaturation 98°C 30 s, 15 cycles of denaturation 98°C 8 s, annealing 61°C 25 s, extension 72°C 30 s and final extension 72°C 7 min; 17 ng of template plasmid was used. PCR

products were gel purified and extracted using a MinElute kit (Qiagen). See supplementary information for primer and template sequences.

Deep sequencing

Libraries were sequenced using Illumina MiSeq at the National Genomics Infrastructure in Stockholm. Three 2×300 sequencing kits were used. One kit was used for the unsorted random library (*random_library_unsorted* in the supplementary material*)* and one kit was used for the unsorted paratope library (*paratope_library_unsorted)*. The unsorted enhancing library and all of the sorted libraries were multiplexed on one chip (*random_library_sort1, random_library_sort2, paratope_library_sort1, paratope_library_sort2, enhancing_library_unsorted, enhancing_library_sort1* and *enhancing_library_sort2*). DNA concentrations of the PCR products were determined on a Qubit and mixed to a final concentration of 4 nM.

Sequencing data analysis

Sequencing reads were processed and analyzed using in-house Python scripts. Paired-ends were joined, and CDR regions were identified by detecting seven bases before and after respective CDR and verifying its expected sequence length. A Phred quality score (Q score) threshold was set to exclude the most dubious base callings; a score threshold >30 was used for the heat map analyses while a score >24 was used to analyze the mutation frequencies in the 'Gene overview' sheet in the supplementary. Codons were translated into their corresponding amino acid if all their nucleotides were above the quality threshold. For the heat map analyses, all translated amino acids were used to calculate the variation while for the 'gene overview' frequencies, only reads that had all 50 CDR positions translated were used for analysis.

RESULTS

### SAMURAI—a method for generation of antibody libraries and paratope mapping

In SAMURAI (Figure 1A), the wild-type gene is first amplified by PCR using the nucleotides A, G, C and U (no T). A biotinylated primer allows the amplified gene to be immobilized on streptavidin-coated magnetic beads. The second step involves elution of the antisense strand by incubating the beads at a high pH with NaOH, and the beads are subsequently washed with a neutral buffer before phosphorylated mutagenic oligonucleotides are annealed to the sense strand. Next, the mutated antisense strand is extended using standard nucleotides (A, G, C, T) with DNA polymerase and ligated in a single step with a thermostable ligase. Subsequent addition of uracil excision enzymes (USER) directly to the reaction tube completely degrades the original sense strand, and the new complementary sense strand is synthesized using standard PCR amplification and the resulting product ligated into a vector.

Since the presence of the wild-type gene is undesirable in libraries, we have implemented three levels of stringency to address this issue. Firstly, the original gene is amplified without any thymine (T) so that full incorporation of uracil (U) is achieved. The complete digestion of the original uracil sense strand with USER enzyme mixture removes all traces of the wild-type gene (Supplementary Figure S1). Secondly, the original gene is amplified without the L3 region and therefore only one digestion site is present. The subsequent annealing of the L3 oligonucleotide ensures that strands with failed L3 annealing cannot be ligated into the vector as the second digestion site is present in the L3 oligonucleotide. Thirdly, a standard high-fidelity polymerase, deficient in the capability of reading uracil positions in DNA template, is used for the re-synthesis of the final strands.

Critical paratope residues and enhancing residues revealed with alanine scanning

Although all six CDR loops are involved in antigen recognition, earlier studies have shown that that only 20–30% of the residues within the CDRs actually form contacts with the antigen (34). Furthermore, residues that are directly involved in the interaction with the antigen are also, in general, the most variable ones (35). Therefore, we decided to map the paratope residues that are crucial for binding by alanine scanning. For the paratope mapping, residues in the CDRs of an anti-HER2 scFv that are crucial for binding to the target protein were determined. Fifty individual mutants were created in an automated and high-throughput manner. Here, each mutant had one CDR residue replaced with alanine by annealing and extension of site-specific oligonucleotides onto immobilized single-stranded uracil-containing wild-type template, as illustrated in (Figure 1A) and the resulting extension products were subsequently cloned into the Gram-positive Staphylococcal display vector. Of 199 sequenced colonies, 170 carried the correct mutation to alanine, resulting in an overall mutation efficiency of 85%. The alanine scan was done in two rounds as 11 positions, the majority located in CDR H2, proved to be difficult, presumably due to secondary structures with high melting points. For this reason, the annealing temperature in the second attempt was increased from 55 to 63°C, which resulted in successful mutagenesis for all difficult positions, i.e. at least one of the four colonies that were screened had the desired mutation. If excluding the tricky positions, 92% of the screened clones had the correct mutation in the first attempt and 77% of the screened clones had the correct mutation in the second attempt that only included the more challenging positions. Characterization of binding to the target antigen was done using flow cytometry (Figure 1B). The scFv is fused to an albumin binding protein (ABP) at its C-terminal and a pro-peptide (PP) at the N-terminal for efficient secretion. The PP–scFv–ABP complex is anchored to the peptidoglycan cell wall via the XM domain (36,37). Flow cytometric characterization or selection uses labeled antigen that binds to the displayed scFv. Labeled human serum albumin (HSA) binds to ABP and is used to normalize surface expression levels. The binding strength of each alanine mutant was compared to that of the wild-type scFv (Figure 2). The alanine scanning experiment of CDR residues successfully revealed several crucial paratope
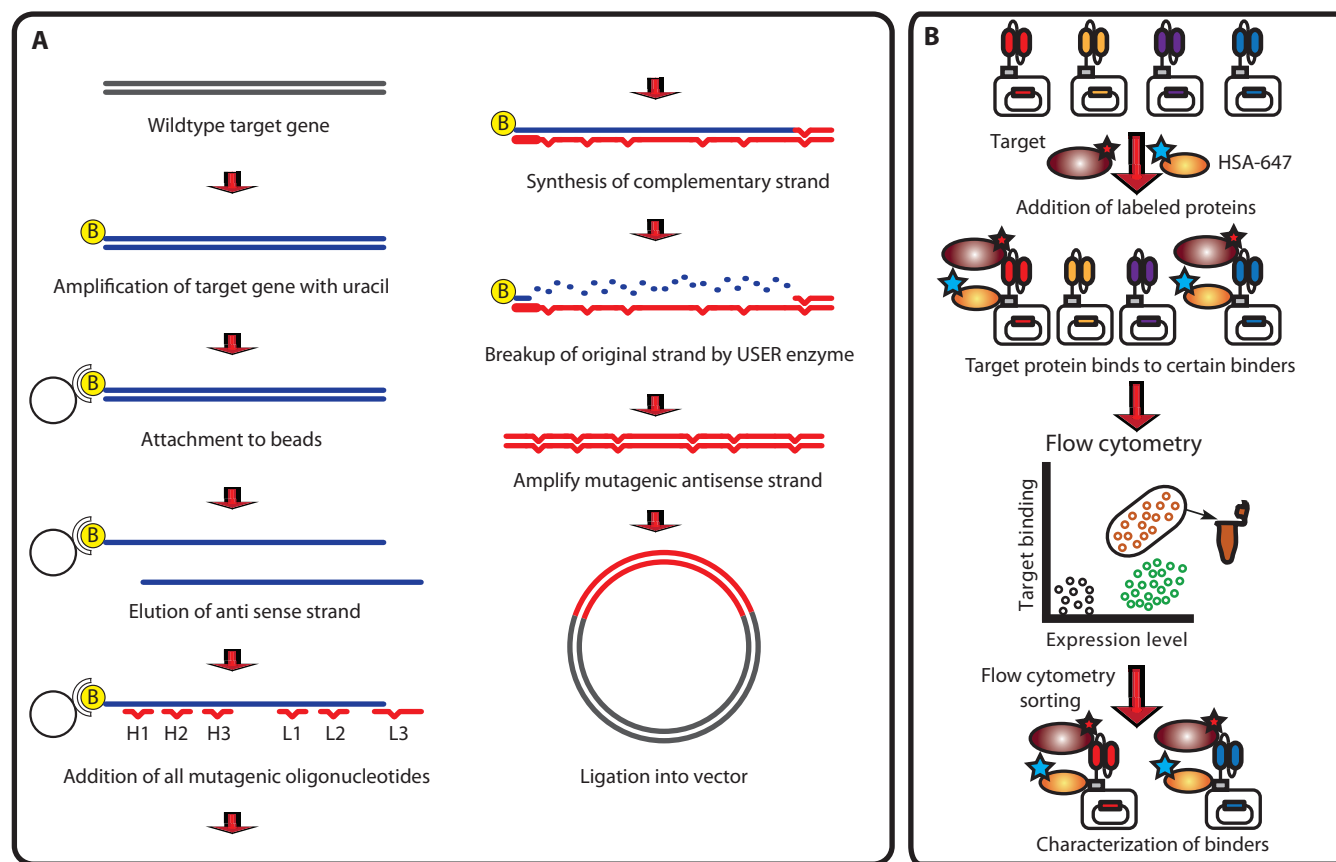
Figure 1. Solid-phase antibody library generation and screening using Gram-positive display. (**A**) Solid-phase mutagenesis of an antibody library. A wild-type scFv (gray) is amplified using dAGCU (0% T) nucleotides and the PCR product (blue) is immobilized on magnetic beads. Following elution of the antisense strand by sodium hydroxide, mutagenic oligonucleotides are annealed, extended and ligated in a single step, yielding a mutated antisense strand (red). Uracil excision degrades the original sense strand, and a new sense strand is synthesized, yielding a double-stranded mutated gene. (**B**) The Staphylococcal display platform was used for both screenings of individual clones as well as library screening/selection. Cells were incubated with labeled antigen target and labeled human serum albumin (HSA-Alexa 647) for detection of surface expression. HSA-Alexa 647 binds to the ABP (albumin binding protein) which is fused to the C-terminal of the surface expressed scFv. A positive signal from HSA-Alexa 647 indicates surface expression of the scFv-ABP fusion protein. Clones were sorted (or screened in the case of individual clones) using flow cytometry, and the strongest binders (upper right corner, circled) were collected for subsequent rounds or characterization.

residues (marked in red in Figure 2A and B; hereby referred to as critical residues), which had a very unfavorable impact on antigen binding when mutated to alanine. While mutation of the majority of the CDR residues to alanine had a small effect on binding, there were a few residues which showed improved binding to the antigen when mutated to alanine (marked in blue in Figure 2A and B; hereby referred to as enhancing residues). Unsurprisingly, most of the critical residues are located in CDRH3 and CDRL3 as these two CDRs occupy the central binding site and contribute most to antigen binding (Figure 2B). On the other hand, the four enhancing residues are distributed across three different CDRs.

Creation of directed libraries by SAMURAI

Randomizing all CDR residues is a very straightforward approach, which does not require any information regarding the epitope–paratope interaction. However, a disadvantage with randomizing all CDR residues is that a relatively small percentage of the library will contain beneficial mutations and the resulting library is usually too big to be prac-

tically screened with display methods. Therefore, two additional directed libraries were created based on the information obtained from alanine paratope scanning. Thus, in the first library, the random library, all fifty CDR residues were targeted for mutagenesis. In the second library, a directed library with the four residues which had the most beneficial effect on target binding in the alanine scan mutated; we'll refer to this as the enhancing library. In the third library, the six residues with the most detrimental effect on binding in the alanine scan were mutated; this library will be referred to as the paratope library. All libraries were created with SAMURAI and each randomized position was allowed to mutate to all 20 different amino acids with degenerate oligonucleotides containing randomized 'NNK' codons. All primers were added simultaneously and the total amount of primer was five pmol; primers were mixed in equimolar proportions. Since the six residues in the paratope library are located in three different CDRs (2 in H3, 1 in L1 and 3 in L3), a mixture of oligonucleotides were used to mutate H3 and L3. For example, oligonucleotides containing one, two or three mutations were used to mu-
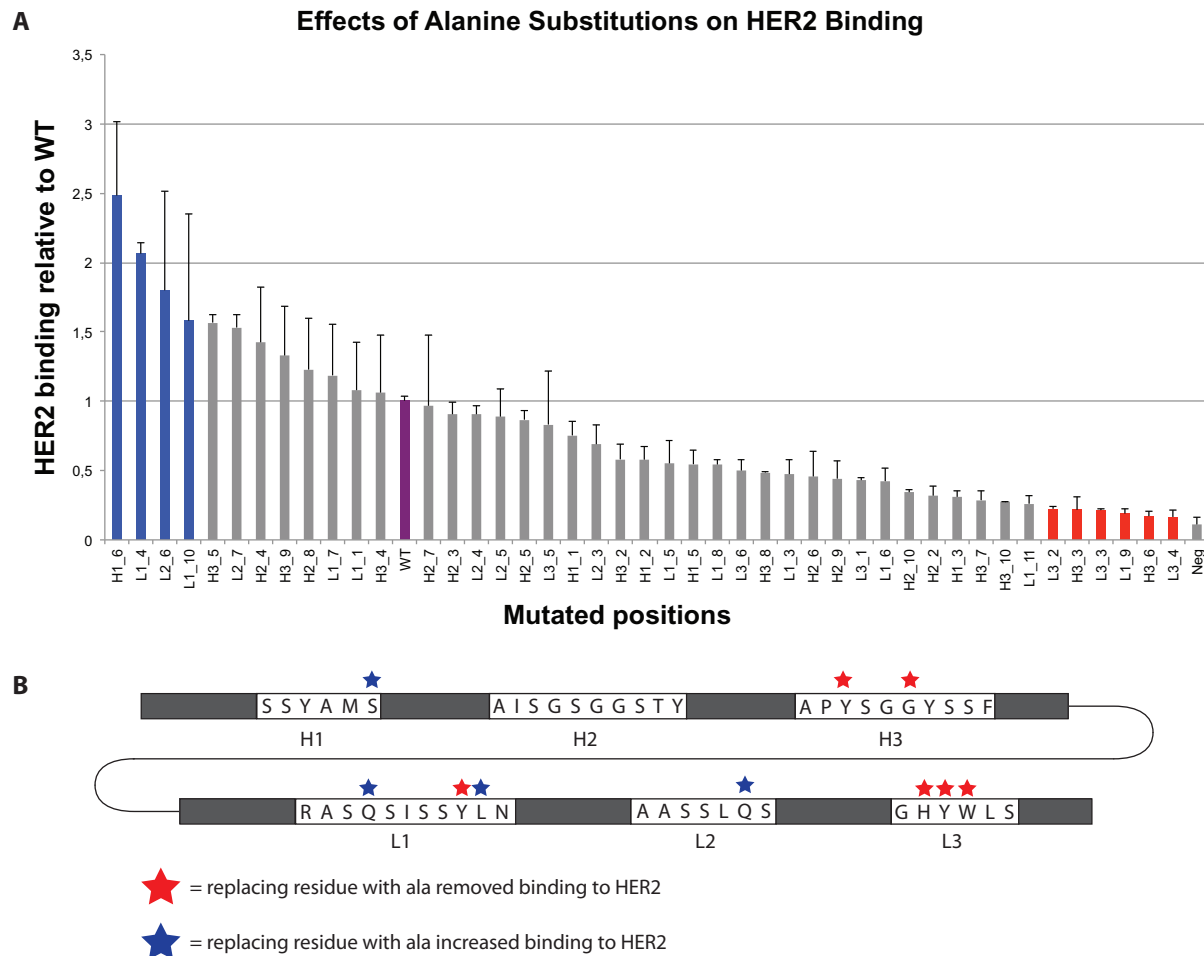
Figure 2. Paratope mapping of the wild-type scFv revealed enhancing and diminishing residues. (**A**) Forty-four single alanine mutants were created, each corresponding a non-alanine CDR residue in the scFv. Each surface displayed mutant scFv clone was incubated with labeled HER2 antigen and labeled HSA (for measuring surface expression) followed by analysis with a flow cytometer. The signal ratio between antigen binding and surface expression (i.e. antigen binding/surface expression) was then compared to that of the wild-type scFv clone. The experiment was done in duplicates. The binding screening assay revealed enhancing and diminishing residues compared to the wild-type (purple). (**B**) Location of enhancing (blue) and diminishing (red) residues on the scFv gene. Most of the critical residues were located on H3/L3 whereas the enhancing residues were found on other CDRs.

tate L3 in the paratope library. The same approach was also used to mutate L1 in the enhancing library. Which primers were used for each library can be seen in the supplementary information. The transformations yielded libraries of the size of 10^6, which according to GLUE-IT (38) correspond to a completeness of 0.91 for the enhancing library with four NNK codons, 0.013 for the paratope library with six NNK codons, and with 50 NNK codons the random library had too many randomized positions to function with GLUE-IT.

MiSeq deep sequencing of the three unsorted libraries after transformation into *S. carnosus* showed a low presence of wild-type in the random and paratope libraries: 0.8% and 0.6%, respectively (see Supplementary information—deep sequencing results.xlsx). For the paratope library, the six positions with expected mutations had mutation frequency between 21 to 56%, while the positions without expected mutations only showed a frequency between 0.03 and 0.3%. This low frequency is in line with the sequencing error rate considering the Phred quality threshold of >30. The ran-

dom library showed a mutation frequency of 0.4–12% for the CDR positions and an average of 4.3%. Although some residue positions had a lower degree of mutation in the random library (e.g. positions 38 and 39) whereas other positions had a higher degree of mutations (e.g. CDRL3), it should be possible to further fine-tune these positions by modifying the length of the oligonucleotides. See Supplementary Figure S2 for a heatmap of amino acid distributions. The enhancing library had a mutation frequency between 25 and 37% for the four positions with expected mutations and between 0.03 and 0.2% for the positions without expected mutations. This library had a higher frequency of wild-type gene (∼13%), which we believe is 're-created wild-type' instead of original 'template' wild type. Unlike the random or paratope libraries, a wild-type CDRL3 oligonucleotide was used when building the enhancing library, and annealing failure of the other three oligonucleotides would result in a re-creation of the wild-type gene. One solution to this issue would be to use sequential addition (Supplemen-

tary Figure S3) of mutagenic oligonucleotides when constructing the enhancing library.

Selections of the three affinity maturation libraries were carried out using flow cytometry sorting (Figure 3). First, the library-displaying cells were incubated with biotinylated-HER2 and allowed to reach equilibrium. The cells were then washed to remove the antigen and incubated with fluorescently labeled secondary reagent (streptavidin-rPE and HSA-Alexa 647) to monitor antigen binding and surface expression. In each case, the top binding population in the indicated gate in the unsorted library was sorted and expanded. The binding population was enriched following two rounds of subsequent sorting. Unsurprisingly, most of the mutations in the random library are detrimental and decreased binding to the antigen (Figure 3A). Additionally, mutating the six residues that contributed most to antigen binding was also disadvantageous in the paratope library (Figure 3C). However, mutating the four enhancing residues seemed to produce a higher percentage of functional binders in the enhancing library (Figure 3B).

Sequencing data shows mutation of enhancing residues is a viable approach for affinity maturation

After three rounds of flow cytometric library sorting 30 clones from each library were subjected to single clone analysis by flow cytometry. In a similar manner as during sorting, each clone was first incubated for two hours with biotinylated HER2, followed by washing and addition of streptavidin-rPE and HSA-Alexa 647. The clones were ranked based on antigen binding while taking surface expression into account. What is noteworthy is that no new functional binders were obtained from the paratope library. The seemingly enriched population from that library during the selection process consisted of the wild-type clone (0.6% occurrence in the unsorted library). In contrast, both the random and enhancing libraries were able to produce affinity-matured binders. The best binder from each library was re-cloned, expressed as soluble scFvs in *E. coli* and subjected to Biacore for affinity measurement (see Supplementary Figure S4 for Biacore results). The obtained KD values for the two candidates were 3 nM (random library) and 2 nM (enhancing library) respectively, which is an improvement over the wild-type clone (10 nM). This result was consistent with the clonal flow cytometric screening as both matured clones showed a noticeable increase in antigen binding signal (Figure 4). This comparison indicates that a smaller directed enhancing library is viable in producing desired binders. Although a 5-fold KD improvement may be considered minor, a harsher selection process (using a decreasing amount of antigen or increased washing) would possibly identify binders with higher affinity. Moreover, using less antigen in the clonal screening process could also distinguish between the two matured binders.
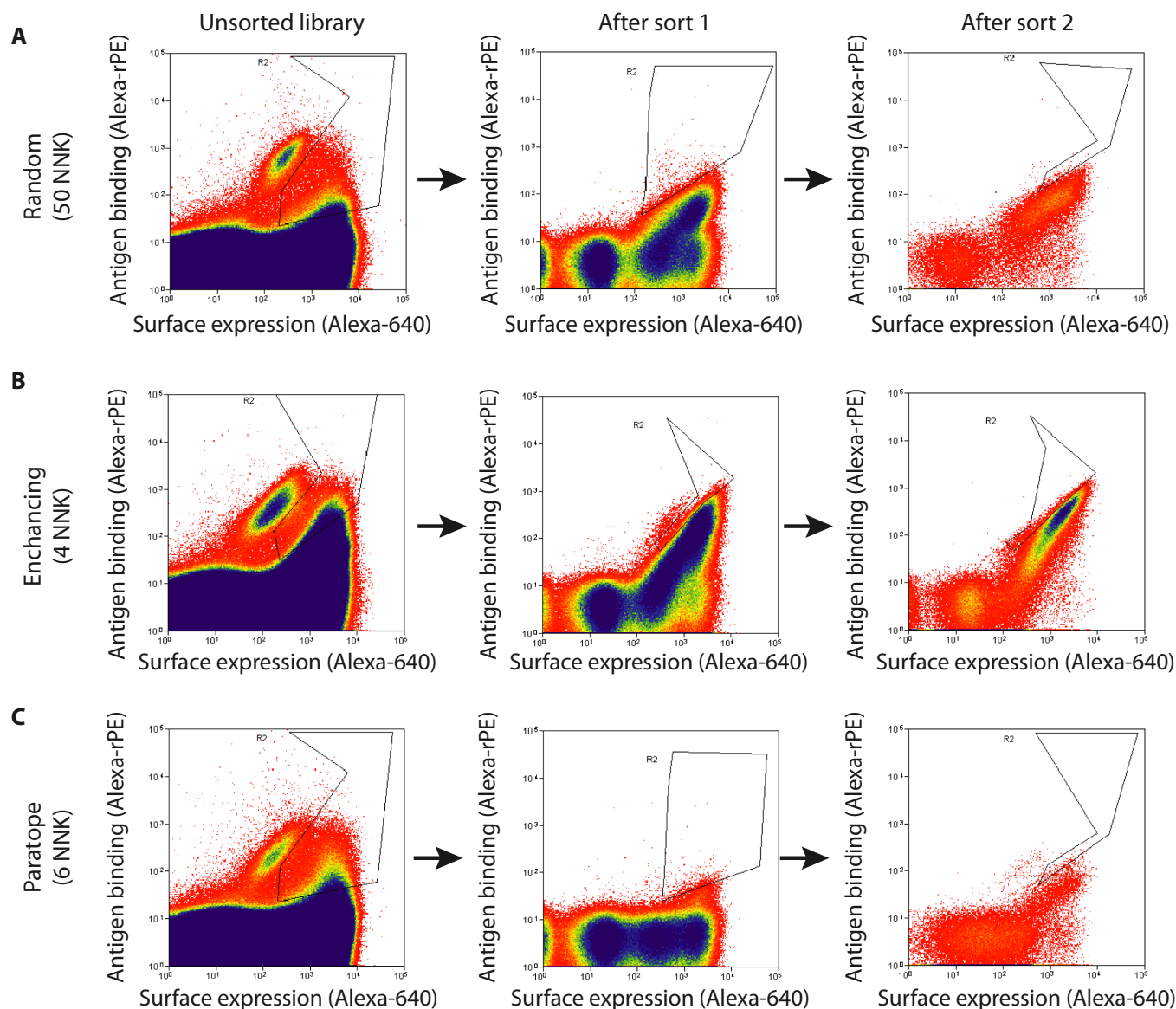
DNA from isolated cells following each round of library sorting was subjected to MiSeq deep sequencing to gain an insight into the changes in CDR amino acid composition throughout the maturation process. The sequence information obtained from MiSeq was used to generate heatmaps showing at which residue positions that enrichment had occurred (Figure 5). For each po-

sition, a comparison was made between (i) the unsorted library and end of one round of sorting as well as (ii) the unsorted library and end of two rounds of sorting. The results were based on the following numbers of sequences: random library unsorted—21866840, random library sort 1—3911526, random library sort 2—824301, enhancing library unsorted—4987179, enhancing library sort 1—3807117, and enhancing library sort 2—688824. The heatmaps also show which of the 20 amino acids that the wild-type residue has mutated into in each position. In the random library (Figure 5A), the six critical paratope residues showed enrichment in the wild-type amino acid. The only exception is position 22 where cysteine was also enriched together with the wild-type glycine. The lack of enrichment in non-wild-type amino acids at these six critical paratope positions is consistent with the single clone analysis of the paratope library described earlier and illustrates that simultaneously mutating too many crucial paratope residues has a detrimental effect on antigen binding.

Sequential annealing—an alternative mutagenesis approach

In the maturation library assembly process, a few choices are possible during the oligonucleotide annealing step. One alternative is to add all mutagenic oligonucleotides to the reaction simultaneously, i.e. adding all mutagenic oligos at the same time. Another option involves sequential addition of the oligonucleotides for one CDR at a time (all primers targeting the same CDR were mixed before addition). Figure 1A depicts the simultaneous annealing, see Supplementary Figure S3 for an outline of the sequential annealing. The solid-phase mutagenesis for creating the three libraries in this study was done with simultaneous annealing. This method was picked to achieve the more conservative mutational rate, typically wanted for affinity maturation, with on average three mutated CDRs per clone in the random library (Supplementary Figure S5C). When we used sequential addition of 50 different mutagenic oligonucleotides to mutate 50 different CDR residues in all six CDRs the result was a very low percentage of wild-type gene and a high number of mutated residues (Supplementary Figure S5). The same set of primers were used for this sequential annealing as for the simultaneous annealing of the random library. More specifically, when 50 colonies were sequenced following sequential addition of oligonucleotides, the clone with the fewest number of mutations had three mutated CDRs per gene (three amino acid mutations per gene as the oligonucleotides were designed to mutate one residue per CDR). Approximately 23% of clones had six mutated CDRs, 30% had five mutated CDRs, and 32% had four mutated CDRs respectively (Supplementary Figure S5A). These numbers correspond to a higher site-directed mutation rate than previously reported examples (23,26,27). Furthermore, mutations had occurred in all six CDRs at a relatively even frequency with the only exception being CDRH2 (Supplementary Figure S5B). We believe the lower mutation rate in CDRH2 is due to differences in annealing temperature relative to the other oligonucleotides, which could be solved by modifying the length of the mutagenic CDRH2 oligonucleotide. All six CDRs in the random library had

A Random (50 NNK)

Unsorted library | After sort 1 | After sort 2

B Enchancing (4 NNK)

Unsorted library | After sort 1 | After sort 2

C Paratope (6 NNK)

Unsorted library | After sort 1 | After sort 2

Figure 3. Fluorescent cell sorting of three affinity maturation libraries. Three different designs were used to construct three libraries by either (**A**) mutating all CDR residues in a completely random manner, (**B**) mutating the four enhancing residues or (**C**) mutating the six diminishing residues. All three libraries were sorted under the same conditions. The cells from the gate in the unsorted library (left) were expanded and subjected to two more rounds of sorting (middle and right). The sorted cells from the gate in the final sorting round (right) were analyzed individually. Strongest binding signal and clearest enrichment of library was seen for the (**B**) track followed by (**A**), whereas (**C**) failed to generate any improved binders.

been mutated at a relatively even frequency but with higher mutation rate in L3 (Supplementary Figure S5D).

DISCUSSION

In this paper, we have presented an approach using solid-phase immobilization and enzymatic construction of specifically tailored mutagenesis libraries coupled with surface display and flow cytometric screening for affinity maturation. The creation of the target template containing uracil omits the need for plasmid preparation from *E. coli*, and the subsequent complete digestion of the original uracil-containing strand reduces 'carry-over' wild-type dramatically. Immobilization of the target template on streptavidin-coated magnetic beads allows for elution of the anti-

sense strand, and straightforward preparation of single-stranded template and annealing of diversity oligonucleotides for subsequent site-directed mutagenesis. Furthermore, by choosing simultaneous or sequential addition of the mutagenic oligonucleotides, it is possible to either conveniently mutate all six CDRs in a single step or more carefully control the number of mutations during the assembly. Furthermore, the use of magnetic beads for immobilization and cleanup allows for easiness of handling and creation of numerous individual mutants in an automated and high-throughput manner. Although we created single mutants containing only one mutation per oligonucleotide in this paper, we are certain that it is possible to create even more highly diverse libraries as well as single mutants us-
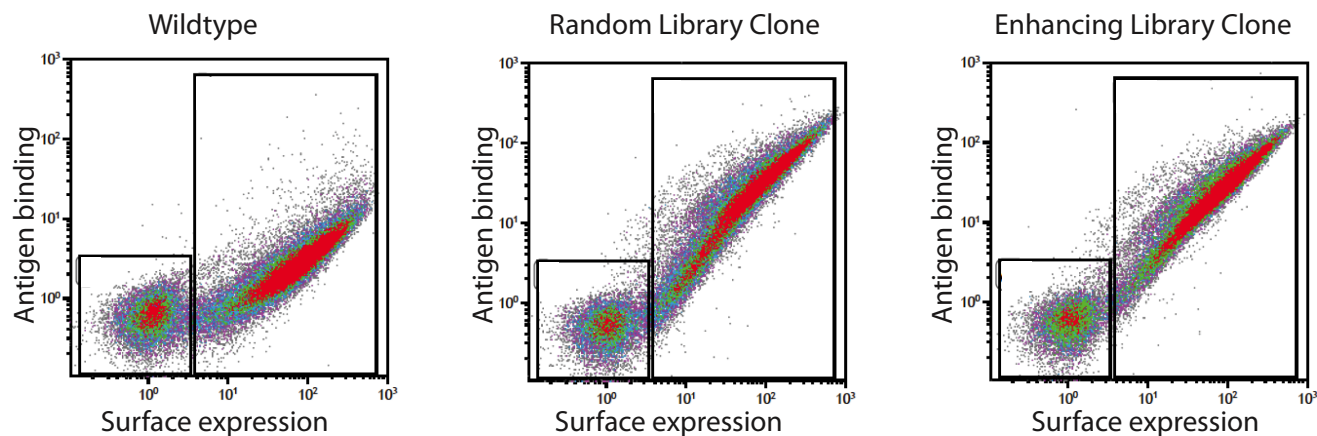
Figure 4. Analysis of single clones after FACS sorting showed that both random and enhancing libraries were able to generate binders with improved affinities. Thirty individual clones from each of the three libraries after the second FACS sorting were individually analyzed on flow cytometer (50 000 events for each clone) for improved binding to HER2. Following incubation, each clone was assessed based on antigen binding (y-axis) while considering surface expression (x-axis). The most promising clone from the random and enhancing libraries both showed increased binding to HER2 compared to the wild-type clone.
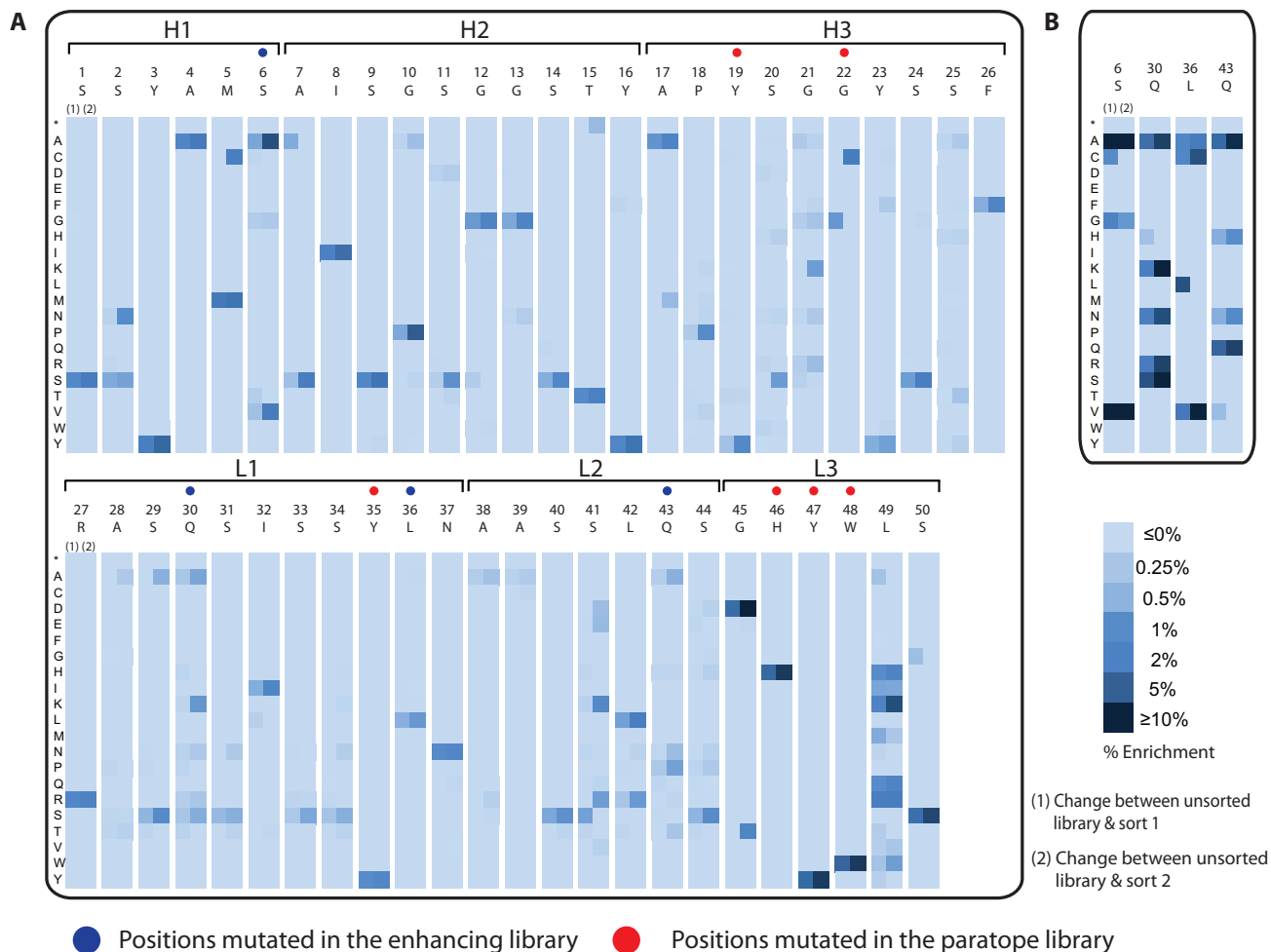
Figure 5. MiSeq analysis of two libraries during selection rounds indicate similar favored mutational positions. MiSeq was used to analyze the sequences in the unsorted library as well as after rounds 1 and 2 of sorting to identify which amino acids were enriched at each CDR position. Blue dots indicate positions mutated in the enhancing library and red dots indicate positions that were mutated in the paratope library. Both the random (**A**) and enhancing library (**B**) were subjected to sequence analysis. The enriched amino acid in each position is visualized with a color scheme, and the enrichment (or lack thereof) between the unsorted library and after one round of sorting is indicated by (1) whereas the enrichment between the unsorted library and after two rounds of sorting is indicated by (2).

ing oligonucleotides containing multiple mutations with the same approach.

Although complete random mutagenesis of all paratope residues would be the ideal way to find the best binder, such an approach often create a diversity too broad to be efficiently assessed by existing screening systems. It would thus be desirable to produce smaller and thought-out directed libraries rather than sampling the fraction of a library with high diversity. For this end, we generated two different directed libraries based on alanine scanning of the paratope in addition to a random library. Following selection, it was revealed that mutation of antigen contacting residues had a detrimental effect whereas mutation of the enhancing residues produced a small library containing a high percentage of functional binders. Deep sequencing of the isolated clones revealed that several non-critical residue positions in the random library were enriched for non-wild-type amino acids whereas wild-type amino acids were substantially enriched in the critical paratope residue positions. This enrichment shows that mutation of too many antigen contacting residues has a detrimental effect on antigen binding whereas creating a small directed library based on mutation of the enhancing positions is a viable approach for affinity maturation. Deep sequencing of the random library after each selection round also revealed that the enhancing positions in the random library were enriched for other amino acids than wild-type. The heat map for the random library revealed that the three best enhancing residue positions had undergone mutations to other amino acids, with a decrease in wild-type amino acid as a result. In addition, there were also several positions (mainly 45, 49, 21, 41) where a reduction in wild-type amino acids was seen, and other amino acids were enriched. We believe these are possible supporter positions that can enhance binding. Although these supporter positions did not distinguish themselves in the alanine scanning experiment, it should be noted that only one position was mutated at a time and that synergy effects were not taken into account. The three best enhancing residue positions had undergone mutations in the random library, which shows that making a small directed enhancing library containing a higher percentage of functional binders can be considered a viable approach for affinity maturation. Furthermore, two out of three mutations from the best isolated binder from the random library were in the enhancing positions, further indicating that identifying the critical paratope residues can aid the construction of directed libraries by circumventing complete random mutagenesis.

In summary, when compared to existing methods, SAMURAI is a versatile and convenient method that enables simultaneous site-specific mutagenesis of all six CDRs. When coupled with paratope scanning and assessment with deep sequencing, we illustrate that our approach for construction of libraries with site-directed mutagenesis to be a viable way for the generation of functional and diverse genetic libraries, particularly suitable for affinity maturation of antibodies. Finally, in addition to affinity maturation, we also envision the possibility to use SAMURAI for other purposes such as deep mutational scanning, especially when combined with deep sequencing.

DATA AVAILABILITY

Script used for analysis of sequencing data is available at: https://github.com/maglund/SPUX

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Gaughan,C.L. (2016) The present state of the art in expression, production and characterization of monoclonal antibodies. *Mol. Divers.*, **20**, 255–270.
2. Kohler,G. and Milstein,C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, **256**, 495–497.
3. Brüggemann,M., Osborn,M.J., Ma,B., Hayre,J., Avis,S., Lundstrom,B. and Buelow,R. (2015) Human antibody production in transgenic animals. *Arch. Immunol. Ther. Exp. (Warsz).*, **63**, 101–108.
4. Lonberg,N. (2005) Human antibodies from transgenic animals. *Nat. Biotechnol.*, **23**, 1117–1125.
5. Smith,G. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315–1317.
6. McCafferty,J., Griffiths,A.D., Winter,G. and Chiswell,D.J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, **348**, 552–554.
7. Hanes,J. and Plückthun,A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 4937–4942.
8. Mazor,Y., Van Blarcom,T., Mabry,R., Iverson,B.L. and Georgiou,G. (2007) Isolation of engineered, full-length antibodies from libraries expressed in Escherichia coli. *Nat. Biotechnol.*, **25**, 563–565.
9. Boder,E.T. and Wittrup,K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.*, **15**, 553–557.
10. Jefferis,R. (2009) Glycosylation as a strategy to improve antibody-based therapeutics. *Nat. Rev. Drug Discov.*, **8**, 226–234.
11. Wörn,A. and Plückthun,A. (2001) Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.*, **305**, 989–1010.
12. Daugherty,P.S., Chen,G., Olsen,M.J., Iverson,B.L. and Georgiou,G. (1998) Antibody affinity maturation using bacterial surface display. *Protein Eng.*, **11**, 825–832.
13. Hoogenboom,H.R. (2005) Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.*, **23**, 1105–1116.
14. Bowers,P.M., Horlick,R.A., Neben,T.Y., Toobian,R.M., Tomlinson,G.L., Dalton,J.L., Jones,H.A., Chen,A., Altobell,L., Zhang,X. *et al.* (2011) Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20455–20460.

15. Chiu,M.L. and Gilliland,G.L. (2016) Engineering antibody therapeutics. *Curr. Opin. Struct. Biol.*, **38**, 163–173.
16. Elgundi,Z., Reslan,M., Cruz,E., Sifniotis,V. and Kayser,V. (2017) The state-of-play and future of antibody therapeutics. *Adv. Drug Deliv. Rev.*, **122**, 2–19.
17. Kunik,V., Ashkenazi,S. and Ofran,Y. (2012) Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. *Nucleic Acids Res.*, **40**, 521–524.
18. Hu,D., Hu,S., Wan,W., Xu,M., Du,R., Zhao,W., Gao,X., Liu,J., Liu,H. and Hong,J. (2015) Effective optimization of antibody affinity by phage display integrated with high-throughput DNA synthesis and sequencing technologies. *PLoS One*, **10**, e0129125.
19. Raghunathan,G., Smart,J., Williams,J. and Almagro,J.C. (2012) Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens. *J. Mol. Recognit.*, **25**, 103–113.
20. Tomlinson,I.M., Walter,G., Jones,P.T., Dear,P.H., Sonnhammer,E.L. and Winter,G. (1996) The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.*, **256**, 813–817.
21. Yang,W.P., Green,K., Pinz-Sweeney,S., Briones,A.T., Burton,D.R. and Barbas,C.F. (1995) CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *J. Mol. Biol.*, **254**, 392–403.
22. Chowdhury,P.S. and Pastan,I. (1999) Improving antibody affinity by mimicking somatic hypermutation in vitro. *Nat. Biotechnol.*, **17**, 568–572.
23. Rajpal,A., Beyaz,N., Haber,L., Cappuccilli,G., Yee,H., Bhatt,R.R., Takeuchi,T., Lerner,R.A. and Crea,R. (2005) A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 8466–8471.
24. Laffly,E., Pelat,T., Cédrone,F., Blésa,S., Bedouelle,H. and Thullier,P. (2008) Improvement of an antibody neutralizing the anthrax toxin by simultaneous mutagenesis of its six hypervariable loops. *J. Mol. Biol.*, **378**, 1094–1103.
25. Persson,H., Ye,W., Wernimont,A., Adams,J.J., Koide,A., Koide,S., Lam,R. and Sidhu,S.S. (2013) CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J. Mol. Biol.*, **425**, 803–811.
26. Firnberg,E. and Ostermeier,M. (2012) PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One*, **7**, 1–10.
27. Wrenbeck,E.E., Klesmith,J.R., Stapleton,J.A., Adeniran,A., Tyo,K.E.J. and Whitehead,T.A. (2016) Plasmid-based one-pot saturation mutagenesis. *Nat. Methods*, **13**, 928–930.
28. Lundqvist,M., Edfors,F., Sivertsson,Å., Hallström,B.M., Hudson,E.P., Tegel,H., Holmberg,A., Uhlén,M. and Rockberg,J. (2015) Solid-phase cloning for high-throughput assembly of single and multiple DNA parts. *Nucleic Acids Res.*, **43**, e49.
29. Hudson,E.P., Nikoshkov,A., Uhlen,M. and Rockberg,J. (2012) Automated solid-phase subcloning based on beads brought into proximity by magnetic force. *PLoS One*, **7**, e37429.
30. Fowler,D.M. and Fields,S. (2014) Deep mutational scanning: a new style of protein science. *Nat. Methods*, **11**, 801–807.
31. Hu,F.J., Volk,A.L., Persson,H., Sall,A., Borrebaeck,C., Uhlen,M. and Rockberg,J. (2018) Combination of phage and Gram-positive bacterial display of human antibody repertoires enables isolation of functional high affinity binders. *New Biotechnol.*, **45**, 80–88.
32. Volk,A.-L., Hu,F.J. and Rockberg,J. (2014) Epitope Mapping of Monoclonal and Polyclonal Antibodies Using Bacterial Cell Surface Display. In: Ossipow,V. and Fischer,N. (eds). *Monoclonal Antibodies: Methods and Protocols*. Humana Press, Totowa, pp. 485–500.
33. Säll,A., Walle,M., Wingren,C., Müller,S., Nyman,T., Vala,A., Ohlin,M., Borrebaeck,C.A.K. and Persson,H. (2016) Generation and analyses of human synthetic antibody libraries and their application for protein microarrays. *Protein Eng. Des. Sel.*, **29**, 427–437.
34. Padlan,E.A. (1994) Anatomy of the antibody molecule. *Mol. Immunol.*, **31**, 169–217.
35. Padlan,E.A., Abergel,C. and Tipper,J.P. (1995) Identification of specificity-determining residues in antibodies. *FASEB J.*, **9**, 133–139.
36. Hansson,M., Stahl,S., Nguyen,T.N., Bachi,T., Robert,A., Binz,H., Sjolander,A. and Uhlenl,M. (1992) Expression of recombinant proteins on the surface of the Coagulase-Negative bacterium staphylococcus xylosus. *J. Bacteriol.*, **174**, 4239–4245.
37. Guss,B., Uhlen,M., Nilsson,B., Lindberg,M., Sjoquist,J. and Sjodahl,J. (1984) Region X, the cell-wall-attachment part of staphylococcal protein A. *Eur. J. Biochem.*, **420**, 413–420.
38. Firth,A.E. and Patrick,W.M. (2008) GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res.*, **36**, 281–285.