Shuffled antibody libraries created by \textit{in vivo} homologous recombination and yeast surface display

Jeffrey S. Swers, Brenda A. Kellogg and K. Dane Wittrup*

Department of Chemical Engineering and Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA, USA

Received December 31, 2003; Revised and Accepted January 20, 2004

**ABSTRACT**

Homologous recombination in yeast can be exploited to reliably generate libraries of $>10^7$ transformants from a pool of PCR products and a linearized plasmid vector. Homology in the PCR insertion products drives shuffling of these genes \textit{in vivo} by yeast homologous recombination. Two scFvs that share 89.8\% homology were shuffled \textit{in vivo} by homologous recombination, and chimeric genes were generated regardless of whether or not one of the scFv PCR products lacked 5\’ homology to the cut vector. A majority of the chimeras had single crossovers; however, double and triple crossovers were isolated. Crossover points were evenly distributed in the hybrids created and homology of as little as two nucleotides was able to produce a chimeric clone. The numbers of clones isolated with a given number of crossovers was approximated well by a Poisson distribution. Transformation efficiencies for the chimeric libraries were of the order of $10^4$–$10^5$ transformants per microgram of insert, which is the same order of magnitude as when a single PCR product is inserted alone into the display vector by homologous recombination. This method eliminates ligation and \textit{Escherichia coli} transformation steps of previous methods for generating yeast-displayed libraries, requires fewer PCR cycles than \textit{in vitro} DNA shuffling and, unlike site-specific recombination methods, allows for recombination anywhere that homology exists between the genes to be recombed. This simple technique should prove useful for protein engineering in general and antibody engineering, specifically in yeast.

**INTRODUCTION**

Various methods have been developed for the creation of diversity within protein libraries, including random mutagenesis (1–3), \textit{in vitro} DNA shuffling (4,5) and site-specific recombination (6–9). Random mutagenesis techniques utilize either a non-proof-reading DNA polymerase in the presence of MnCl$_2$ (10), mutator \textit{Escherichia coli} strains (11) or nucleotide analogs that cannot be correctly read by the DNA polymerase (12). Random mutagenesis has the advantage of allowing for the isolation of beneficial mutations anywhere within the gene that may not be obvious a priori. However, point mutation methods do not allow for a radical restructuring of the contact regions and therefore are restrictive in the sequence space that can be probed. DNA shuffling methods have the advantage of being able to generate hybrid genes that contain portions of sequence space that have already proven to be functional. DNA shuffling consists of four steps: DNase I digestion of the genes to be recombined, PCR reassembly without primers, amplification of recombined gene products of the correct size from the primerless PCR pool, and ligation of the reassembled product into an acceptor vector. Although DNA shuffling has proven to be a highly effective method, numerous PCR and associated purification steps are required. Site-specific recombination has the advantage of utilizing portions of known sequence space that have already proven functional, but it is often an impractical method because it requires the engineering of restriction enzyme sites into the genes to be combined. Often finding unique sites is difficult and the process can become tedious when multiple chimeric products are desired. Other forms of site-specific recombination have been used to make libraries such as the n-CoDer antibody library in which CDRs were shuffled using specific primers and PCR reassembly (13), chain-shuffled scFv libraries in which the shuffling was performed by PCR utilizing homology in the linker region between the heavy and light chains (14), and chain-shuffled antibody libraries using Cre-catalyzed recombination of antibody heavy and light chains that are flanked by \textit{lox} sites (8,9). In all of these libraries, the points at which gene exchange occur are fixed and previous knowledge of the gene sequences for the creation of PCR primer sets or engineering of specific sites is required. A simple method for creating hybrid genes that does not require the extensive PCR steps of DNA shuffling, the pre-engineering of site-specific recombination methods, or ligation would be desirable. In this paper, we present a technique for creating large, chimeric antibody libraries using plasmid reconstruction by homologous recombination in the yeast \textit{Saccharomyces cerevisiae} (15,16). This method allows for the coupling of diversity generation and protein production within the host organism, rather than having a separate \textit{in vitro}
in vivo recombination in yeast has been used in the past to create hybrid genes (17–19) and has been successfully used to generate a library for the directed evolution of a heme peroxidase enzyme (19). However, this technique has not been exploited for the generation of antibody libraries. This absence can be explained by the fact that most in vitro antibody engineering is performed using phage, which relies upon the machinery of E. coli for gene propagation and protein production. Although intra- (Fig. 1A) and intermolecular recombination (Fig. 1B and C) have both been performed in E. coli to make hybrid genes, intermolecular recombination has generally been used for plasmid construction and not recombination (20). Intermolecular recombination is inherently more flexible than intramolecular recombination because only one (Fig. 1B and C) or none (Fig. 1D) of the genes to be recombined must be within a vector, while intramolecular recombination requires both genes to be within the same vector.

The most flexible scheme, presented in Figure 1D, allows for recombination anywhere in the genes to be shuffled and is most applicable to antibody engineering because none of the genes must be within the acceptor vector. Thus, pools of antibodies can be shuffled with no bias towards any individual acceptor antibody. To our knowledge, the crossover scheme depicted in Figure 1D has not been used to generate antibody libraries in E. coli. This fact can be explained by the proficiency of S. cerevisiae in homologous recombination in comparison with the inherent problems in the endogenous E. coli homologous recombination mechanism (21). For example, the E. coli endogenous homologous recombination mechanism is initiated by the cooperation between RecA and the enzyme RecBCD, which impedes the use of linear DNA because RecBCD is a vigorous exonuclease (21). In the cases where E. coli has been used to create hybrid genes (Fig. 1B and C), mutant strains must be used, which are not advantageous because they can exhibit genetic instabilities during library transformation and propagation. Although it is possible to exploit the yeast recombination method to make chimeric genes and then recover the plasmids containing the hybrid gene products for transfer to E. coli, this method is hampered by both the inefficient recovery of plasmids from yeast minipreps and the added difficulty of having to switch hosts. Not until the advent of yeast surface display did it become practical to exploit yeast in vivo homologous recombination methods for making chimeric antibody libraries.

This paper requires fewer PCR cycles than in vitro DNA shuffling, does not require a ligation step, and, unlike site-specific recombination methods, it allows for recombination anywhere that homology exists between the genes to be recombined. Moreover, the diversity generated is not limited to single crossovers; hybrid antibodies created by double and triple crossovers were easily obtained. This technique should prove to be an important addition to the aforementioned methods for the isolation of affinity-improved antibodies.

**MATERIALS AND METHODS**

**DNA preparation for homologous recombination**

The two scFvs used in the shuffling experiments were isolated from a previously constructed yeast-surface-displayed non-immune library (22). To test homologous recombination frequencies of a mutagenic PCR product, a third scFv from this library was amplified in a PCR reaction using a nucleotide analog mutagenesis procedure that has been described previously (12). All PCR products were inserted into the vector PCTCON (22), which had been restriction digested from Nhel to BamHI (New England Biolabs) and gel purified using a gel purification kit (Qiagen) according to the manufacturer’s instructions. As described in the study by Raymond et al. (16), to obtain the best transformation efficiency, homologous recombination primers were designed so that the inserts would have an ~50 bp overlap at each end with the cut acceptor vector. The primer used to make inserts with 5′ homology to the cut vector was 5′-CGACGCATTG-
AAGGTAGATAACCCTACGACTAACGCTC-TGCAG-3', and the primer used to make inserts with 3' homology to the cut vector was 5'-CAGATCTCGAGCTA-TTACAAGTCTCTTCAAAATAAGCTTTGTC-3'. To make either scFv 1 or scFv 2, which lacked 5' homology to the cut vector, the primer 5'-GCTAGCCAGGTACAGCTGC-AGC-3' was used. To make either scFv 1 or scFv 2, which lacked 3' homology to the cut vector, the primer 5'-AACAGGTAGAGCCGGTACAGCTGC-3' was used. All oligonucleotides were obtained from MWG-Biotech. PCR reactions were carried out on a Perkin Elmer DNA Thermal Cycler 480. A typical PCR reaction was carried out in a 100 μl volume using 10–100 ng of template, 1 μM primers, 0.2 mM of each dNTP, 6.25 U Taq (Invitrogen) and 2 mM MgCl2. Cycling conditions used were: one cycle at 94°C for 1 min followed by 35 cycles at 94°C for 1 min, 50°C for 1 min and 72°C for 2 min, followed by one cycle at 72°C for 10 min. PCR products were gel purified using a Qiagen kit. Insert fragments were concentrated with Pellet Paint (Novagen) to a concentration of 5 μg/μl according to the manufacturer’s instructions and cut backbone was likewise concentrated to a concentration of 1 μg/μl.

Preparation of electrocompotent yeast for homologous recombination

The method of yeast preparation closely follows that described by Meilhoc et al. (23). First, 50 ml of YPD was inoculated with the S.cerevisiae strain EBY100 (11) to an optical density (OD) of 0.1 from an overnight culture of EBY100 in YPD. Next, the cells were grown with shaking at 30°C to an OD of 1.3–1.5 (~6 h of growth). Cells were harvested by centrifugation and resuspended in 50 ml of freshly prepared 10 mM Tris pH 8.0, 25 mM dithiothreitol (DTT) in YPD and shaken for 20 min at 30°C. The cells were washed once with 25 ml buffer E (10 mM Tris pH 7.5, 270 mM sucrose, 1 mM MgCl2) and again with 5 ml of E buffer. Finally, cells were suspended in buffer E to give 2 × 108 cells per 50 μl aliquot.

Homologous recombination protocol

The ratio of total insert fragment to cut acceptor vector was maintained at 10:1 for all transformations. For shuffling experiments where both scFvs were transformed together, half of the insert fragment pool consisted of one scFv and half of the insert fragment pool consisted of the other. One microgram (1 μl) of cut acceptor vector and 10 μg (2 μl) of insert were added to a 50 μl aliquot of electrocompotent yeast and incubated on ice for 5 min. Electroporation was carried out using a Bio-Rad Gene Pulser with a 0.2 cm cuvette (voltage 0.54 kV, capacitance 25 μF) giving a time constant of ~18 ms. After pulsing, the cell aliquots were transferred to 1 ml of YPD media and incubated for 1 h at 30°C. Cells were then harvested at 3500 r.p.m. for 4 min and resuspended in SDCAA selective media (~Ura, ~Trp). A small aliquot of cells was removed and plated on SDCAA plates to determine transformation efficiency.

DNA isolation and sequencing

Colonies from the SDCAA plates were grown in 5 ml of SDCAA overnight and the DNA was isolated using a Zymoprep kit (Zymo Research) according to the manufacturer’s protocol. Two microliters of Zymoprep DNA were used in an XL1-Blue (Stratagene) E.coli transformation according to the manufacturer’s instructions. Cells were plated on selective Luria–Bertani (LB) plates supplemented with 100 mg/l ampicillin. Colonies from these plates were grown overnight at 37°C in LB media plus 100 mg/l ampicillin and DNA was isolated using a Qiagen miniprep kit according to the manufacturer’s instructions. DNA was sequenced on an Applied Biosystems model 3730 DNA sequencer using version 3.0 Big Dye chemistry.

RESULTS

Creation of large yeast-surface-displayed libraries by homologous recombination

As has previously been reported, the yeast S.cerevisiae is highly efficient at reconstructing plasmids from a linearized plasmid and a PCR product that contains sufficient homology at the 3' and 5' ends (15,16). Figure 2 outlines the homologous recombination strategy. The first step involves the linearization of a yeast surface display plasmid, which contains an irrelevant gene flanked by a 5' NheI site and a 3' BamHI site, in a restriction digest. The second step involves the creation of a pool of PCR fragments that share 5' and 3' homology to the cut display vector. Because Raymond et al. have reported that overlaps of ≥50 base pairs on each end of the PCR product yield the highest number of recombinants (16), we made PCR fragments of mutagenic antibody library DNA that have ~50 base pairs of homology to the display vector at their 5' and 3' ends. Co-transformation of these fragments with linearized plasmid allowed for the creation of large yeast-displayed libraries. Table 1 shows the results of two different transformations of mutagenic antibody library DNA using homologous recombination. Both experiments yielded a transformation efficiency of an order of magnitude 105/μg of insert DNA. Homologous recombination by electroporation simplifies library generation by completely eliminating the need for ligation of PCR products into a cut vector and the subsequent cloning into E.coli. Library sizes of 107 are routinely achieved without great effort.

In vivo shuffling of antibody DNA by homologous recombination

It is our desire to extend the technique of homologous recombination in yeast to allow for the easy construction of chimeric scFv antibody libraries. To test if chimeras can be created by homologous recombination, a series of four different transformation experiments (Fig. 3) were performed using two different antibodies, termed scFv1 and scFv2, which shared 89.8% homology between them. These antibodies, which were isolated from a non-immune S.cerevisiae-displayed library (22), were chosen not only because they shared significant patches of homology, but also because they had numerous differences spread throughout the entire length of their sequences. These differences are critical for the determination of crossover points. In the experiments depicted in Figure 3A and B, the two genes to be recombined had homology with the linearized display vector at one end only. In principle, neither fragment alone could recreate a whole plasmid because each fragment was missing critical homology at one end, therefore only recombinants would be isolated...
after transformation. To determine if forcing recombination is necessary for the creation of hybrid genes, a third experiment was performed (Fig. 3C) using scFv1 and scFv2, where both genes contained 5′ and 3′ homology to the display vector. In a fourth experiment (Fig. 3D), scFv1 was transformed alone to verify that both 5′ and 3′ homology are necessary for plasmid reconstruction.

Figure 4 shows the results of the in vivo homologous recombination shuffling experiments. For all of the results depicted in Figure 4, the crossover point in the chimeric antibody sequence was determined to be the last nucleotide of homology 5′ to a difference between the chimeric antibody sequence and whichever scFv sequence to which the chimeric antibody had homology up until that point. Because scFv1 and scFv2 share 60 base pairs of homology at the beginning of their sequences and 44 base pairs of homology at the end of their sequences, it is possible for crossovers in these regions to generate genes that have 100% sequence identity to one of the two original scFs. It was our initial hypothesis that for the insertion into the display vector to be successful, the insert fragment needs both 5′ and 3′ homology to the cut plasmid. However, the results depicted in Figure 4D indicate that this hypothesis is false because 10 of the 12 clones sequenced were 100% homologous to scFv1, although scFv1 lacked the 5′ homology necessary for in vivo homologous recombination to occur. Apparently, the single-stranded end of the NheI cut plasmid can recombine with the blunt, double-stranded DNA of the scFv to be inserted. It is likely that the BamHI-restricted end of the vector can also recombine with blunt end DNA in a similar manner. This integration phenomena might be a form of illegitimate integration, defined as recombination involving little or no sequence homology, that has been previously reported to occur in S.cerevisiae (24). Thus, it is impossible to tell if a gene that has 100% homology to either scFv1 or scFv2 is a hybrid or if it was generated through illegitimate integration. In this paper, only genes that clearly contain portions of scFv1 and scFv2 will be considered chimeric.

Table 1. Transformation efficiencies of two different homologous recombination experiments

<table>
<thead>
<tr>
<th>Trial</th>
<th>Insert to vector ratio</th>
<th>Micrograms of insert transformed</th>
<th>Number of transformed clones</th>
<th>Transformants per microgram of insert</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10:1</td>
<td>31</td>
<td>$1.8 \times 10^7$</td>
<td>$5.81 \times 10^6$</td>
</tr>
<tr>
<td>2</td>
<td>10:1</td>
<td>60</td>
<td>$2.3 \times 10^7$</td>
<td>$3.83 \times 10^6$</td>
</tr>
</tbody>
</table>

Figure 3. In vivo homologous recombination experiments. (A) Homologous recombination when scFv 1 has 3′ homology to the display vector and scFv 2 has 5′ homology to the display vector. (B) Homologous recombination when scFv 1 has 5′ homology to the display vector and scFv 2 has 3′ homology to the display vector. (C) Homologous recombination when both scFv 1 and scFv 2 have 5′ and 3′ homology to the display vector. Solid lines indicate a crossover producing chimeric clones that have 5′ homology to scFv 1 and 3′ homology to scFv 2. Dashed lines indicate a crossover to produce chimeric clones with the opposite arrangement. A mixed path of one solid line crossover and one dashed line crossover without crossover between the two scFs would produce a plasmid without a chimeric gene. (D) Homologous recombination when scFv 1 has 3′ homology to the display vector and scFv 2 is not present. This recombination occurs between the (gly4-ser)3 linker that joins the scFv heavy and light chains and the (gly4-ser)3 linker located immediately 5′ to the the ultimate location of the gene in the display plasmid.
Figure 4A (corresponding to the experiment depicted in Fig. 3A) shows that of the 48 clones sequenced, 12 were chimeric, and Figure 4B (corresponding to the experiment depicted in Fig. 3B) shows that of the 46 clones sequenced, 17 were chimeric. It is not surprising that all of the hybrids in Figure 4A shared both homology 5' to the crossover point with scFv2 and homology 3' to the crossover point with scFv1, because in this experiment scFv2 had 5' homology to the cut plasmid and scFv1 had 3' homology to the cut plasmid. As expected, the reverse is generally true for the experiment.
Table 2. Expected and experimentally determined number of clones with a given number of crossovers as predicted by a Poisson distribution

<table>
<thead>
<tr>
<th>Number of crossovers per sequence (x)</th>
<th>Number of sequences with x crossovers in a 1 × 10⁷ library with λ = 0.13</th>
<th>Poisson calculated number of sequences with x crossovers</th>
<th>Actual number of sequences with x crossovers</th>
<th>Poisson calculated number of sequences with x crossovers</th>
<th>Actual number of sequences with x crossovers</th>
<th>Poisson calculated number of sequences with x crossovers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8 777 137</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7 144 844</td>
<td>40</td>
<td>120</td>
<td>240</td>
<td>198 000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>74 664</td>
<td>380</td>
<td>2280</td>
<td>6840</td>
<td>186 219 000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5 324 64</td>
<td>2280</td>
<td>27 360</td>
<td>123 120</td>
<td>1.1 × 10¹¹</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1 × 10⁷</td>
<td>2702</td>
<td>29 763</td>
<td>130 204</td>
<td>1.1 × 10¹¹</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Expected number of total clones and unique clones with a given number of crossovers as predicted by Poisson distribution for a 1 × 10⁷ library

<table>
<thead>
<tr>
<th>Number of crossovers per sequence (x)</th>
<th>Number of unique sequences assuming 20 distinct crossover points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recombine 2 genes</td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Double (clone 24 and 32) and triple (clone 11 and 25) crossovers greatly increase the diversity potential of libraries generated by *in vivo* homologous recombination. To estimate this diversity, we applied the Poisson distribution to estimate the probability of obtaining a clone with *x* = 0, 1, 2… crossovers in a given sequence by using the following equation:

\[ p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \]

where \( \lambda \) is defined as the average number of crossovers per sequence in a given experiment. The value of \( \lambda \) was 0.29, 0.43 and 0.13 in the experiments depicted in Figure 3A–C, respectively. For each of these experiments the expected and actual number of clones obtained are listed in Table 2. As can clearly be seen, the number of clones experimentally obtained for *x* = 0 or *x* = 1 is well approximated by the Poisson distribution. The limited sampling size of between 45 and 50 clones per experiment explains the lack of agreement for *x* = 2 and *x* = 3. Table 3 gives the expected number of clones containing a given number of crossovers for a 1 × 10⁷ library. Although in the second column this table gives an idea of the distribution of crossovers, it does not take into account the number of unique sequences, which is the true measure of library diversity. The following equation can be used to determine the number of unique sequences (**u**) obtained with a given number of crossovers (**x**) when **y** genes are crossed:

\[ u = y(y-1)^x \frac{z!}{x!(z-x)!} \]

where \( \frac{z!}{x!(z-x)!} \) indicates the number of combinations with **z** crossover points. As shown in Table 3, for a cross of two genes with an
arbitrarily chosen number of 20 distinct crossover points, this equation yields 40 unique one-crossover clones, 380 unique two-crossover clones and 2280 unique three-crossover clones. The data in Table 3 reveal the trend that the number of unique clones with multiple crossovers will become greater than the number of clones predicted to have multiple crossovers by the Poisson distribution as the number of genes crossed is increased. For crosses of small numbers of antibodies, the chimeras in the library created would be dominated by many copies of each unique single crossover clone, but the added diversity created by double and triple crossovers would be non-zero. For a cross of 100 genes, the potential theoretical diversity would be dominated by unique multiple crossover clones and would greatly exceed the diversity of multiple crossover clones predicted to be in the library by the Poisson distribution; however, the unique single crossover clones shared homology to the cut display plasmid. Transformation of double and triple crossover clones as predicted by the Poisson distribution, 77 910, is 39% of the maximum number of unique single crossover clones, 198 000, indicating that double and triple crossovers are a non-trivial component of the total diversity. Of course, clones with four or more crossovers may exist, but because none were isolated in the limiting sampling of this experiment they were not included in the analysis. The possibility that forcing recombination reduced transformation efficiency below a level that was useful was considered. However, for all of the crosses performed here, the transformation efficiency was of an order of $10^4$ to $10^5$ per microgram of insert. Thus, forcing recombination does not appreciably affect transformation efficiency.

**DISCUSSION**

In this paper, we present a simple method for reliably producing large, yeast-surface-displayed chimeric antibody libraries. This method takes advantage of the homologous recombination pathway in yeast to reconstruct full plasmids from restriction-digested plasmids and PCR products that have 5' and 3' homology to that cut plasmid. Transformation of multiple PCR products that share homology can cause recombination events to yield chimeric gene products. Forcing recombination by making a PCR product of one gene with 5' homology to a cut vector and making a PCR product of a second gene with 3' homology to the cut vector yields greater numbers of chimeric gene products than when both genes are to be recombined share both 5' and 3' homology to the vector into which they are to be inserted. The transformation efficiency for forced recombination was not appreciably lower than that of a single PCR product, with both 5' and 3' homology to the cut display plasmid.

Although forced recombination yields the greatest number of hybrid genes, this method may prove impractical because one must know the sequence of the 5' or 3' ends of the genes that are to be shuffled. In the case presented in this paper, the two genes to be recombined had known sequences that facilitated primer design. However, one may wish to shuffle one scFv gene against an entire library of scFvs or shuffle a batch of scFv genes that all bind a given antigen. Generic primer sets for all antibody heavy and light chains are known and it would be possible to force recombination by creating PCR products that lack 5' or 3' homology to the vector into which the genes are to be inserted. This method is rather tedious and in our experience has proven unnecessary. As our experiments demonstrate, not forcing recombination allows for the creation of hybrid clones, albeit at a lower rate than forced recombination. The number of chimeric clones obtained is only 2- to 3-fold less than in forced recombination and is therefore not low enough to cause concern that forcing recombination is necessary for this technique to be practical. It is important to stress that the ability of as few as two nucleotides of homology to yield successful recombinants in conjunction with the presence of double and triple crossovers demonstrates that this technique has a vast potential for producing highly diverse libraries. Although it is true that most chimeras contain only one crossover, there will still be many unique clones in the library to ensure the necessary diversity for directed evolution experiments. It is likely that the unique diversity of libraries made by yeast in vivo homologous recombination will be sufficient for affinity improvements, even if the actual diversity of the libraries generated is orders of magnitude lower than the total number of transformants because the shuffling of antibodies is a rearrangement of diversity that has been maintained specifically because it is functional. The advantage of this technique is its simplicity and it is likely that many variations other than those presented here will emerge for the production of functional diversity in protein libraries.

**ACKNOWLEDGEMENTS**

The authors wish to thank the NIGMS/MIT Biotechnology Training Program (GM08334) for funding.

**REFERENCES**


