PCRless library mutagenesis via oligonucleotide recombination in yeast

Nathan Pirakitikulr,[†,‡] Nili Ostrov,[†] Pamela Peralta-Yahya,[§] and Virginia W. Cornish*

Department of Chemistry, Columbia University, New York, New York 10027

Received 20 January 2010; Revised 11 September 2010; Accepted 13 September 2010
DOI: 10.1002/pro.513
Published online 8 October 2010 proteinscience.org

Abstract: The directed evolution of biomolecules with new functions is largely performed *in vitro*, with PCR mutagenesis followed by high-throughput assays for desired activities. As synthetic biology creates impetus for generating biomolecules that function in living cells, new technologies are needed for performing mutagenesis and selection for directed evolution *in vivo*. Homologous recombination, routinely exploited for targeted gene alteration, is an attractive tool for *in vivo* library mutagenesis, yet surprisingly is not routinely used for this purpose. Here, we report the design and characterization of a yeast-based system for library mutagenesis of protein loops via oligonucleotide recombination. In this system, a linear vector is co-transformed with single-stranded mutagenic oligonucleotides. Using repair of nonsense codons engineered in three different active-site loops in the selectable marker *TRP1* as a model system, we first optimized the recombination efficiency. Single-loop recombination was highly efficient, averaging 5%, or 4.0×10^5 recombinants. Multiple loops could be simultaneously mutagenized, although the efficiencies dropped to 0.2%, or 6.0×10^3 recombinants, for two loops and 0.01% efficiency, or 1.5×10^2 recombinants, for three loops. Finally, the utility of this system for directed evolution was tested explicitly by selecting functional variants from a mock library of $1{:}10^6$ wild-type:nonsense codons. Sequencing showed that oligonucleotide recombination readily covered this large library, mutating not only the target codon but also encoded silent mutations on either side of the library cassette. Together these results establish oligonucleotide recombination as a simple and powerful library mutagenesis technique and advance efforts to engineer the cell for fully *in vivo* directed evolution.

Keywords: homologous recombination; *in vivo* mutagenesis; oligonucleotide recombination; protein libraries; *TRP1*

Abbreviations: bp, base pairs; ds, double strand; DSB, double strand break; HR, homologous recombination; ss, single strand; SSA, single strand annealing.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institutes of Health; Grant number: GM62867; Grant sponsors: Pfizer Undergraduate Research Fellowship, Pfizer/Mazur Research Fellowship.

[†]Nathan Pirakitikulr and Nili Ostrov contributed equally to this work.

[‡]Nathan Pirakitikulr's current address is School of Medicine, Yale University, New Haven, CT 06511.

[§]Pamela Peralta-Yahya's current address is California Institute of Quantitative Biomedical Research, University of California, Berkeley, California 94720.

**Correspondence to:* Virginia W. Cornish, Department of Chemistry, Columbia University, Havemeyer Hall, MC 3111, 3000 Broadway, New York, NY 10027. E-mail: vc114@columbia.edu

Introduction

As advances in synthetic biology improve our cell engineering capabilities and increasing research efforts are aimed at generating networks of molecules that function *in vivo*, technologies are needed to allow directed evolution to be carried out directly in the cell. Here, we adapt homologous recombination (HR) techniques, now routinely used for targeted gene alteration, for cassette library mutagenesis of protein loops. Inspiration for this notion comes both from the longstanding use of bacterial mutator strains to generate libraries of random DNA mutations and the widespread use of HR for making gene deletions. We designed a recombination-based system specifically to meet the needs of protein directed evolution and challenged this system to search large

Published by Wiley-Blackwell. © 2010 The Protein Society

libraries of mutations. This library-oriented approach expands the use of HR for DNA manipulation in yeast, offering a straightforward method for library mutagenesis *in vivo* that could ultimately replace current PCR-based *in vitro* methods.

In vitro DNA mutagenesis is now used routinely to create libraries of protein and nucleic acid molecules for directed evolution.[1] *In vitro* mutagenesis techniques for library generation have been developed extensively, and robust methods for both random library mutagenesis, such as error prone PCR and DNA shuffling, and targeted library mutagenesis, such as cassette mutagenesis, are available today.[2,3] The most popular *in vitro* library methods have been analyzed extensively for their mutagenesis rate, mutational bias, dependence on DNA sequence and length and reproducibility.[4,5] This comprehensive development and characterization of *in vitro* library mutagenesis methods have made these approaches generally accessible.

By contrast, *in vivo* library mutagenesis approaches are less developed. Random mutagenesis can be carried out *in vivo* using bacterial mutator strains, but the chromosomal DNA is also mutated at a higher rate, limiting the potential for carrying out multiple rounds of mutagenesis and selection without removing the target gene of interest from the host strain. Furthermore, the mutagenesis cannot be targeted to residues of particular interest.[4] While toxic to the host cell in its current inception, an error-prone DNA polymerase engineered by Loeb and coworkers offers a clever strategy for carrying out *in vivo* mutagenesis without chromosomal modification.[6] Using multiple rounds of random *in vivo* mutagenesis by somatic hypermutation, a mechanism unique to B cell lines, Tsien and coworkers outperformed traditional *in vitro* mutagenesis techniques obtaining substantial improvement of a red fluorescent protein variant.[7] Reported optimization and characterization of these newer *in vivo* mutagenesis methods is limited. Looking forward, there is significant potential for the development of diverse approaches to *in vivo* DNA library mutagenesis.

HR holds the possibility for a simple and powerful library mutagenesis technique. The highly efficient HR machinery of *S. cerevisiae* is now routinely used, both with single-stand (ss) and double-stranded (ds) DNA, to insert and knockout genes for strain and plasmid construction.[5,8–12] Significantly, high recombination efficiency (>10%) was achieved in yeast using *Delitto Perfetto*, an *in vivo* targeted mutagenesis technique where ssDNA replaces a counter-selectable marker.[13] The Court laboratory has shown that recombination machinery from bacteriophage λ can be used to support similarly efficient HR in *E. coli* in a technique they call recombineering,[14] and others have used recET proteins in *E. coli* to demonstrate recombination between linear and circular DNA.[15] A handful of reports exploit HR for library mutagenesis: already in 1995, Sherman and coworkers randomized a single position in the chromosomal *CYC1* gene in yeast,[16] while Novo Nordisk and others reported the use of HR in yeast to shuffle beneficial mutations obtained from previous rounds of directed evolution.[17–19] Moreover, Wittrup and coworkers recently applied *in vivo* loop shuffling to engineer a fibronectin variant with picomolar affinity to lysozyme,[20] and the Church laboratory used automation for repeated cycles of oligonucleotide recombination via the λ recombineering system in *E. coli* to optimize strain background for lycopene production.[21] Of note, all of this prior work has focused on recombination of individual cassettes, such that the efficiency of simultaneously mutating, for example, multiple loops in a protein or multiple distant genes is yet to be explored. Surprisingly, however, HR is yet to become a mainstay technology for library generation and directed evolution.

Here, as a first step towards the long-term goal of developing HR as a robust technology for *in vivo* library mutagenesis, we designed and characterized a yeast HR-based system for the straightforward library mutagenesis of multiple protein loops using synthetic oligonucleotides. The mutagenesis is effected by simple co-transformation of vector linearized downstream of the target gene and ss oligonucleotide(s) encoding the library mutations under standard electroporation conditions. Repair of nonsense codons engineered in three different active site loops of the selectable *TRP1* gene was used as the model system for development of the technology. First, the efficiency of oligonucleotide recombination was optimized at a single loop. Then, these optimized conditions were used to evaluate the feasibility of simultaneously mutating two or three loops. Finally, the utility of oligonucleotide recombination for directed evolution was challenged by carrying out a mock library selection.

Results

Design of oligonucleotide recombination system for library mutagenesis

While there are now standard protocols for performing gene knockouts using HR, these protocols cannot be simply translated to library mutagenesis. For library mutagenesis, it is important to be able to mutate multiple regions simultaneously, to have a simple protocol that can be implemented readily and rapidly during the iterative steps of mutation and assay, and to have high efficiency HR to cover large libraries of mutations. In fact, the handful of papers that have exploited HR for library mutagenesis have used different strategies for recombination.[22–24] Thus, we designed an oligonucleotide recombination

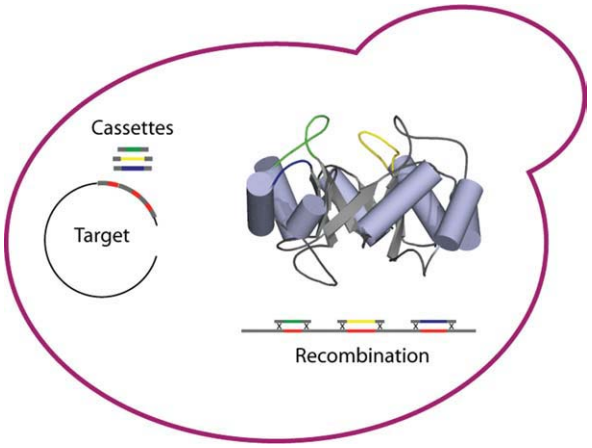


Figure 1. Oligonucleotide recombination via yeast HR. Oligonucleotide recombination provides a general method for generating targeted libraries of DNA mutants *in vivo*. A linearized vector expressing the target gene (gray) and linear DNA oligonucleotides (green, yellow, blue) are co-transformed into yeast. HR between the target gene and DNA oligonucleotides yields libraries of the mutated target gene.

system explicitly for the needs of directed evolution with these criteria in mind.

As illustrated in Figure 1, in this system, a double-strand break (DSB) was introduced just downstream of the 3′ terminus of the gene of interest and mutations were introduced via recombination with sense ss oligonucleotides. The library was constructed simply by co-transformation of the linearized vector encoding the target gene and the mutagenic ss oligonucleotides. A DSB is introduced for high efficiency HR[13] to enable coverage of large libraries, and since the cut site is located outside of the coding gene, as opposed to internally as in most HR technologies, it allows for the possibility of mutating multiple regions simultaneously. It is important to distinguish the two mutagenesis schemes addressed in this work: (1) mutagenesis of multiple codons within a single loop using a single oligonucleotide and (2) mutagenesis of multiple loops simultaneously using multiple oligonucleotides. Assuming typical protein loop sizes and homology regions, the ss oligonucleotide can simply be synthesized, making the technique very straightforward to implement.

The mutagenic ss oligonucleotides were designed to have sufficient homology for high efficiency HR, but at the same time to be short enough for commercial synthesis. Specifically, as depicted in Figure 3A, each oligonucleotide consists of 30 bp upstream homology, the codon(s) to be mutagenized, and then 30 bp of downstream homology. Published studies have established that 30 bp is the minimal homology required for high-efficiency HR.[25] Conveniently, 30 bp is shorter than a typical protein β-strand or α-helix,[26] allowing multiple mutagenic oligonucleotides to be used simultaneously. Assuming a typical protein loop size of 10 amino acids or 30 bp, the total oligonucleotide size would be 90 bp. However, further investigation is required to optimize the number of residues to be mutagenized in each protein loop. Nevertheless, sequences of this length scale can be directly made by solid-phase synthesis,[27] eliminating the need for further enzymatic manipulation of the mutagenic oligonucleotide.

To optimize the likelihood of high efficiency mutation, particularly at multiple loops, the transformation was carried out at high concentrations of plasmid and oligonucleotides. While standard plasmid transformation protocols are expected to yield on average a single circular plasmid per yeast cell, it is now established that oligonucleotide transformation results in multiple oligonucleotides per cell. This is supported by multiple studies of gene targeting[22] and further demonstrated in a recent report by Venter and colleagues, showing the average yeast cell likely takes up substantially more than 25 (DNA) pieces in a single transformation experiment.[28] Furthermore, early experiments demonstrated that co-transformation of large DNA fragments enhances oligonucleotide transformation efficiency.[29] Thus, our working model is that each cell receives the plasmid and multiple copies of mutagenic oligonucleotide(s).

TRP1 *model system*

Repair of nonsense codons in the classic yeast selection marker *TRP1* presented a convenient model to develop our oligonucleotide recombination system. The gene product *N*-(5′-phosphoribosyl)-anthranilate isomerase (yPRAI) is a TIM-barrel enzyme. Approximately 10% of all enzymes share the TIM-barrel fold,[26] and the active-site loops of TIM-barrels are positioned between α-helices and β-sheets, making them particularly well placed for oligonucleotide recombination. The enzyme yPRAI catalyzes an essential step in the tryptophan biosynthesis pathway of *S. cerevisiae* and thus offers a convenient selection for growth in the absence of tryptophan.[30] However, oligonucleotide recombination as characterized in this work is not limited to mutagenesis of protein loops, but rather is a general technology readily applicable for library mutagenesis of any multiple-component system.

We identified three target residues, Arg44, Arg78, and Ser201, in yPRAI for engineering of ochre (TAA) nonsense codons (Fig. 1). The residues were selected based on inspection of a homology model of PRAI from *S. cerevisiae* constructed using Swiss Model[31] and inspected by VMD.[32] The three target residues Arg44, Arg78, and Ser201 lie in catalytic loops 2, 3, and 8 in yPRAI, respectively. The residues vary in their distance from the vector cut site at the 3′ of *TRP1* gene to control for the

Table I. *Vectors and Oligonucleotides (5′–3′) Used in This Study*

Vectors			
Vector	Mutation	Genes	Strain
pHT2150		See Ref. 34	
p424MET25			ATCC 87321
pNP2278		*URA3*, *TRP1*	NP2278
pNP2279	*trp1-R44**	*URA3*, *trp1*	NP2279
pNP2284	*trp1-R78**	*URA3*, *trp1*	NP2284
pNP2282	*trp1-S201**	*URA3*, *trp1*	NP2282
pNP2283	*trp1-R44*R78**	*URA3*, *trp1*	NP2283
pNP2280	*trp1-R44*S201**	*URA3*, *trp1*	NP2280
pNP2281	*trp1-R44*R78*S201**	*URA3*, *trp1*	NP2281

Oligonucleotides		
Name	Primer	Sequence
ARG44Fix	VWC2041	CTGGGTATTATATGTGTGCCCAATAGAAAGAGAACAATTGACCCGGTT ATTGCAAGGAAAATT
ARG44Fix_ds	VWC2042	AATTTTCCTTCGAATAACCGGGTCAATTGTTCTCTTTCTATTGGGCACACA TATAATACCCAG
ARG78Fix	VWC2043	GGCACTCCAAAATACTTGGTTGGCGTGTTTCGTAATCAACCT AAGGAGGATGTTTTGGCTCTG
SER201Fix	VWC2044	AGATTAAATGGCGTTATTGGTGTTGATGTAAGCGGAGGTGTGGA GACAAATGGTGTAAAAGAC
SER201LibraryFix	VWC2218	TAGATTAAATGGCGTTATTGGTGTTGATGTGTCCGGAGGCGTGG AGACAAATGGTGTAAAAGAC
SER201LibraryOpal	VWC2220	TAGATTAAATGGCGTTATTGGTGTTGATGTCTGAGGAGGAGTGG AGACAAATGGTGTAAAAGAC

Target codons for mutagenesis are indicated in red.
Silent mutations are indicated in blue.
Restriction sites (BsaW1 cut site for VWC2218, BseR1 cut site for VWC2220) are indicated in gray.

dependence of the HR efficiency on proximity to the DSB. Additionally, the loops vary in their distance from one another, to explicitly test whether simultaneous mutagenesis of two loops is distance dependent. The reversion rate of target plasmids carrying single, double, or triple ochre codons was tested and found to be $<3 \times 10^6$. (See Supporting Information).

The yPRAI gene was encoded on a 2-μ vector carrying a *URA3* marker such that the efficiency of oligonucleotide recombination upon co-transformation of the linearized vector and oligonucleotide could be scored simply as the ratio of colonies that survive on tryptophan and uracil deficient plates versus uracil deficient plates. We consider the recombination efficiency to be a working, rather than a theoretical, definition, because formally the vector DSB must be repaired to survive in uracil deficient media. Interestingly, in control experiments where the number of viable URA^+ transformants was compared, transformation of linearized vector resulted in an order of magnitude less transformants than circular vector. However, no difference in the number of transformants was observed when linear vector was co-transformed with oligonucleotides (data not shown). To minimize false positives, all experiments were carried out in yeast strain ATCC4017202, which has a complete deletion of the *TRP1* gene.[33]

While quantification of oligonucleotide recombination efficiency was carried out by simply targeting an ochre codon, the mutagenic oligonucleotides used for Ser201 were designed with additional silent mutations to provide a record of the recombination event. Thus, mutagenic oligonucleotides were designed with two additional silent mutations, one upstream and the other downstream of the mutated target codon, so that three codons were effectively mutagenized by each oligonucleotide. Incorporation of these additional silent mutations both provides markers that the recombination event had occurred and, in combination with the target codon mutations, adds a unique restriction site (Table I). Importantly, control experiments using a single oligonucleotide with or without silent mutations showed the efficiency of recombination to be very similar. Hence, successful repair of nonsense Ser201 codon in *trp1* by HR with a fixing oligonucleotide generated functional yPRAI carrying silent mutations and led to cell survival on media lacking tryptophan.

Oligonucleotide recombination at a single loop

First, we examined a broad range of conditions to optimize the efficiency of oligonucleotide recombination at a single loop. Specifically, we sought conditions that result in the greatest number of recombinants possible while maintaining a large number of

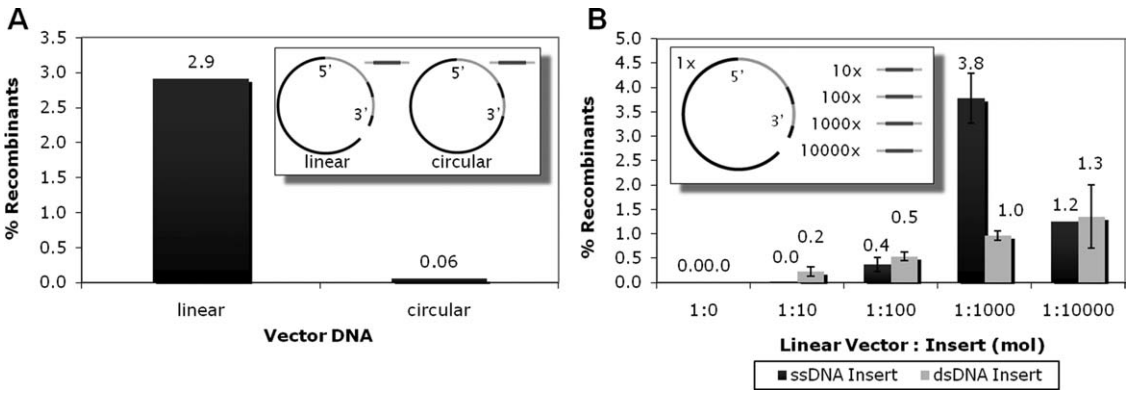


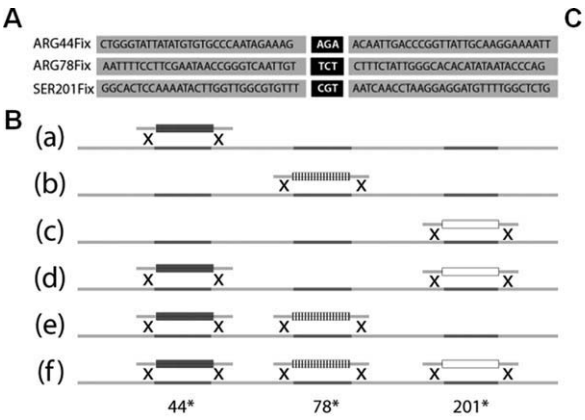
Figure 2. Optimization of experimental parameters for high rates of recombination. A: Linear or circular *trp1-Arg44** vector (1 μg) is co-transformed with 5 μg of the ss oligonucleotide ARG44Fix (1:1000 mol vector: oligonucleotide) into the Δ*trp1 S. cerevisiae* strain ATCC4017202. Co-transformation of linear vector and oligonucleotide yields the greatest percentage *TRP1* recombinants. B: Linear *trp1-Arg44** vector (1 μg) and varying amounts of ss and ds oligonucleotides ARG44Fix and ARG44Fix_ds, respectively, are co-transformed into the Δ*trp1 S. cerevisiae* strain ATCC4017202 and recombinant colonies are scored by plating on SC (Ura^-) and SC (Ura^-Trp^-) selective plates. Co-transformation of 1:1000 mol vector:oligonucleotide yields the greatest percentage *TRP1* recombinants. The data shown are the mean ± the standard error of at least three separate experiments. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

overall transformants. Based on the *S. cerevisiae* HR machinery literature, we considered three main variables: the nature of the vector DNA (circular vs. linear),[35] the nature of the oligonucleotide DNA (ss vs. ds),[24,36] and the ratio of vector to oligonucleotide.[29] We used a sense strand oligonucleotide with forward orientation for all single strand oligonucleotide experiments, unless otherwise indicated. All experiments were performed at least in triplicate, and data were only included from experiments in which at least 10^6 transformants were achieved.

The recombination efficiency was initially optimized based on repair of the nonsense codon in the *trp1-Arg44** gene encoded on vector pNP2279 by the oligonucleotide Arg44Fix [Fig. 3(A) and Table I]. To test whether co-transformation of circular or linearized vector yielded the largest number of recombinants, vector pNP2279 was linearized by digesting 30 bp downstream of the *trp1* gene using the ClaI restriction enzyme. As Figure 2(A) shows, oligonucleotide recombination using linearized vector gave 50-fold higher recombination efficiencies than that using circular vector. Next, recombination efficiencies via co-transformation of varying ratios of linearized vector and either ss or ds wt DNA oligonucleotides were compared. Figure 2(B) presents the recombination efficiencies using 1:10, 1:100, 1:1000, and 1:10,000 molar ratio of linearized vector to ss or ds oligonucleotide. A 1:1000 molar ratio of linearized vector to ss oligonucleotide yielded the highest recombination efficiency, 3.8 ± 0.5%. The efficiency at the 1:1000 molar ratio was 190-times that at the 1:10 molar ratio. The efficiency of ss oligonucleotide at the 1:1000 molar ratio was four times that of ds oligonucleotide at the same ratio.

To ensure that this measured efficiency for single-site recombination is general, the same optimized conditions were then used to repair nonsense codons in the *trp1-Arg78** and *trp1-Ser201** genes, encoded on vectors pNP2284 and pNP2282, respectively, by the oligonucleotides Arg78Fix and Ser201-Fix [Fig. 3(A) and Table I]. As Figure 3(C) shows, the recombination efficiencies determined at positions 44, 78 and 201 were 4.5 ± 0.9%, 1.4 ± 0.3% and 9.0 ± 0.4%, respectively. Thus, the average efficiency of oligonucleotide recombination at a single loop was 5 ± 2%, consistent with published efficiencies of HR at a DSB.[25,37] Interestingly, the measured efficiency was slightly higher at position 201, which is closer to the DSB, although there is not a correlation between the distance from the DSB and the efficiency of recombination. The small differences in measured efficiencies alternatively may arise from the permissiveness of the individual residue to amino acid or codon substitutions. Indeed in the mock selection experiment vide infra, we recovered not only the encoded TCC Ser codon, but also non-encoded codons (See Supporting Information).

Sequence analysis of recombinant colonies repaired at position Ser201 revealed that all colonies carried the oligonucleotide-encoded fixing codon (TCC). (See Supporting Information Fig. 1 for full sequencing data). Interestingly, 80% of tested colonies (16/20) were found to carry both downstream and upstream silent mutations, whereas 20% carried only the upstream silent mutation. This could be caused by either partial incorporation of the oligonucleotide or, alternatively, suggests a role for mismatch repair mechanism in oligonucleotide



C

	Oligonucleotides	% Recombinants	Std Error
(a)	Arg44	4.5	0.9
(b)	Arg78	1.4	0.3
(c)	Ser201	9.0	0.4
(d)	Arg44/ Ser201	0.32	0.15
(e)	Arg44/ Arg78	0.11	0.03
(f)	Arg44/ Arg78/ Ser201	0.010	0.001

Figure 3. Efficiency of oligonucleotide recombination at multiple loops. Linear vectors carrying *trp1-Arg44**, *trp1-Arg78**, *trp1-Ser201**, *trp1-Arg44*Arg78**, *trp1-Arg44*Ser201**, and *trp1-Arg44*Arg78*Ser201** were co-transformed into the Δ*trp1 S. cerevisiae* strain ATCC4017202 with the appropriate oligonucleotide (see Table I). A: Oligonucleotide used in this study to fix nonsense mutations in residues Arg44, Arg78, and Ser201 in yPRAI. Sequences of target codons are highlighted. B–C: Simultaneous mutagenesis of multiple loops exhibits multiplicative efficiency. (B) Schematic representation of one (a–c), two (d–e), or three (f) oligonucleotide mutagenesis. (C) The data shown are mean ± standard error for percent recombination of at least three separate experiments. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

recombination. Further investigation is required to determine the underlying recombination mechanism.

Oligonucleotide recombination at multiple loops

With optimized conditions for oligonucleotide recombination at a single loop, we proceeded to test the efficiency of simultaneous mutagenesis at two and three loops. These experiments were carried out essentially as for the single-site recombination, except that equal molar quantities of the appropriate combinations of two or three of the oligonucleotides Arg44Fix, Arg78Fix, or Ser201Fix were co-transformed with the linearized plasmids pNP2283, pNP2280, or pNP2281. Simultaneous mutagenesis at positions Arg44 and Arg78 had an efficiency of 0.11 ± 0.03%, and simultaneous mutagenesis of positions Arg44 and Ser201 had an efficiency of 0.32 ± 0.15% (Fig. 3). Oligonucleotide recombination at two loops simultaneously therefore had an average efficiency of 0.2 ± 0.1%. The efficiency may be slightly higher when the two loops are a greater distance from one another, but the difference is at most slight. Next, we measured the efficiency of simultaneous mutagenesis at all three positions, Arg44, Arg78, and Ser201; it was 0.010 ± 0.001% (Fig. 3). Thus, simultaneous oligonucleotide recombination at two loops was 25-fold less efficient than that at a single loop, and at three loops was 20-fold less efficient than that at two loops. Notably, the efficiencies at multiple loops were nearly multiplicative.

To increase the efficiency of simultaneous recombination at multiple loops, we tried two classic methods in yeast genetics.[5] First, an overlapping oligomer was used to link two mutagenic oligonucleotides in an attempt to improve the efficiency of multi-loop mutagenesis. In our system, such an approach did not lead to a significant increase in the number of recombinants (data not shown). Next, we aimed at improving the efficiency of vector re-circularization by co-transformation with an oligonucleotide that overlapped the DSB in the vector (Supporting Information Fig. 2 and Supporting Table II). Again, inclusion of this additional oligonucleotide did not improve the recombination efficiency. Thus, in its current form, our protocol for oligonucleotide recombination leads to drops in efficiency as additional loops are mutagenized.

Mock selection via oligonucleotide recombination

While the high efficiency of single-site recombination with our system suggests that it should allow for the generation of large libraries *in vivo*, and since HR is not routinely used for directed evolution, we explicitly challenged our system in a mock selection experiment. Given that 10^6 transformants can be readily obtained in *S. cerevisiae*, the mock selection experiment was designed to test the feasibility of enriching a functional *TRP1* gene obtained by oligonucleotide recombination from a pool of oligonucleotides encoding a mock library of 1 active to 10^6 inactive *trp1* variants. Specifically, linearized pNP2282, encoding the *trp1-Ser201** gene with an ochre codon at position 201, was co-transformed with a 1:10^6 mixture of oligonucleotide Ser201LibraryFix: Ser201Library*Opal* (Fig. 4 and Table I). Both the Ser201LibraryFix and Ser201Library*Opal* oligonucleotides were designed such that they not only introduced a unique restriction site but also

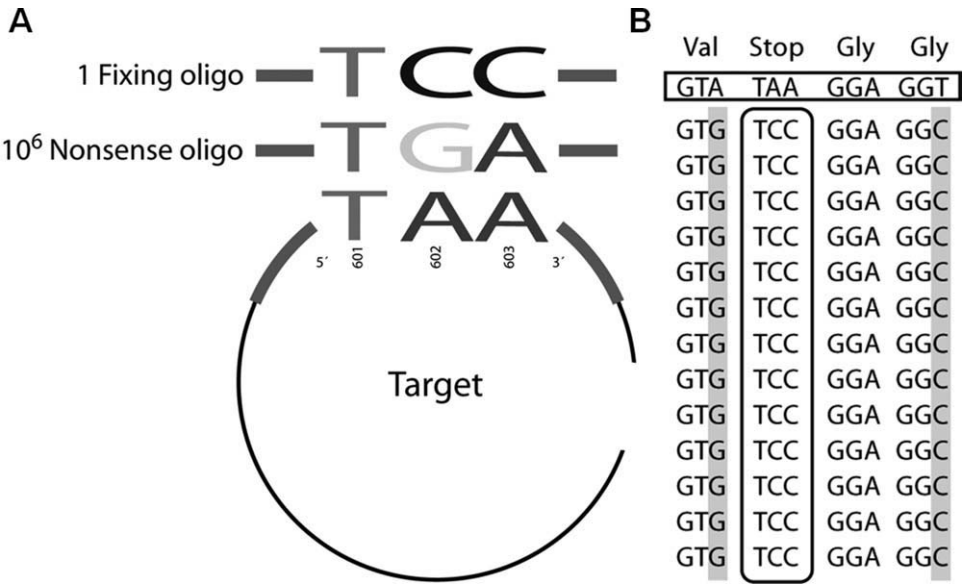


Figure 4. Mock library mutagenesis of *trp1-Ser201** vector. A wt codon was enriched from a mock library of 10^6 inactive variants in a single step. A: Vector NP2282, carrying a *trp1-Ser201** allele, is targeted with a $1{:}10^6$ mix of fixing to nonsense ssDNA oligonucleotides (oligonucleotides SER201LibraryFix and SER201Library*Opal*, respectively). B: Eighteen viable colonies were analyzed by sequencing (see text), and 13 high quality recombinant sequences were aligned using the clustalW server.[38] The vector sequence is shown at the top. All sequenced colonies carried the fixing codon (TCC) at position 201 (framed) as well as the encoded silent mutations upstream (A>G) and downstream (T>C) (highlighted). Library results suggest that the library size is fully covered and therefore may allow for larger library experiments. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

encoded unique silent mutations on either side of the 201 codon to mark the recombination event. Use of the opal codon in the mock library allows it to be readily distinguished from the vector ochre codon. The mock selection was carried out under the same optimized conditions used to measure the single-loop recombination efficiency (see Supporting Information for full experimental details). Thus, 10^8 cells were co-transformed with 3.5×10^{10} linear vector molecules and 3.6×10^{13} total oligonucleotides. Transformants were then selected on SC (Ura^-) plates to determine the library size, and recombinants were scored on SC (Ura^-Trp^-) plates to select for functional *TRP1*.

Encouragingly, with 2.6×10^6 successful transformants on SC (Ura^-) plates, 8328 colonies survived on SC (Ura^-Trp^-) plates (see Supporting Information for full experimental details), demonstrating that a library of $1{:}10^6$ was fully covered with this protocol. Interestingly, this number of colonies was higher than would be simply predicted based on a recombination efficiency of 5%. Thus, the mock selection from a library of $1{:}10^6$ was repeated three times. While the number of TRP^+ colonies varied in each experiment, the library was successfully covered each time (see Supporting Information for full experimental details). We speculate that the library size that can be covered is greater than that predicted by the recombination efficiency because there is a vast excess of oligonucleotide molecules to number of linearized vector molecules, and the vector molecules again are in excess of the number of transformed cells ($\sim 3 \times 10^5$ oligonucleotides and $\sim$300 plasmids per cell). However, this rationale has not yet been explicitly tested. The fact that the library size that can be covered cannot be simply calculated from the recombination efficiency demonstrates the importance of directly testing HR in the context of library construction. Furthermore, since the library of 10^6 was covered in excess, it may be possible to cover even larger libraries not only for single-loop but also multiple-loop recombination.

Eighteen individual colonies were first subjected to further selection in liquid media to cure the strain of what is presumed to be non-recombinant or recombinant non-viable pNP2282 vector and then analyzed by sequencing (see Supporting Information for full experimental details). Of the 18 colonies, 13 were successfully cured of non-recombinant pNP2282 and hence could be assigned to a *TRP1* gene and a 201 codon. The remaining five colonies that were not successfully cured were not analyzed further. As shown in Figure 4(B), all 13 recombinant colonies encoded functional *TRP1* using the TCC Ser codon encoded by the oligonucleotide Ser201LibraryFix and its silent mutations. Because of inherent variability in the mock selection experiment, we repeated the experiment multiple times. It should be noted that in some of these mock selections, we additionally observed a handful of viable recombinant carrying non encoded codons such as TGT (Cys) and TCA

(Ser) at position 201 (see Supporting Information for additional details). The mechanism by which this non-encoded codon arises is not known at this point. Regardless of the mechanism, clearly oligonucleotide recombination provides additional mutations, which may be advantageous in directed evolution experiments.

Discussion

Together, these results establish that oligonucleotide recombination can be used to construct large DNA libraries entirely *in vivo*. For library generation, as opposed to targeted gene alteration, the efficiency of recombination is critical because it directly limits the number of variants that can be tested. We were able to optimize the recombination efficiency at a single loop to ~5%. These optimized conditions allowed a wt codon to be enriched from a mock library of 10^6 inactive variants (a typical library size for directed evolution) in a single step. Furthermore, the method is particularly straightforward to implement. All that is required is co-transformation of commercial ss oligonucleotides and linearized vector using a standard electroporation protocol. Thus, oligonucleotide recombination is competitive with, and for *in vivo* selections easier to implement, than *in vitro* PCR library mutagenesis techniques.

The efficiency of oligonucleotide recombination at a single loop reported here is high and consistent with that reported in the recombination field with use of either a DSB or viral machinery.[14,37] The significant enhancement in recombination efficiency using a linearized vector is in agreement with previous studies demonstrating that DSB induction significantly enhances oligonucleotide recombination in yeast from ~0.03%[12] up to 20%.[37,39]

To our knowledge, this study is the first attempt to mutagenize multiple loops simultaneously by co-transformation of two or more oligonucleotides. Kmiec and coworkers demonstrated that multiple codons could be mutagenized simultaneously using a single oligonucleotide.[40,41] In MAGE, automation allowed for accumulation of multiple mutations by repetitive transformation.[21] But for directed evolution, it is advantageous to be able to mutate multiple positions at the same time. We find that simultaneous transformation of two or three targeting oligonucleotides gives multiplicative recombination efficiencies. Compared to 5% average efficiency for single oligonucleotide recombination, the efficiencies for mutation using two and three oligonucleotides simultaneously were 0.2% and 0.01%, respectively. The multiplicative efficiency is consistent with current fundamental understanding of crossover events, which are independent events whose frequency is proportional to the distance between potentially homologous regions.[42] Alternatively, the multiplicative efficiency could be explained simply as the probability of the two events occurring at the same time.

We tried to overcome the multiplicative effect of simultaneous oligonucleotide recombination, first by linking two oligonucleotides together with a third oligomer to minimize the number of necessary crossovers and next by using an oligomer to close the linearized vector. Neither of these strategies increased the recombination efficiency, significantly. However, it is possible that moving the vector cut site into the target gene will allow for increased recombination at multiple locations. Finally, if the multiplicative effect arises from probabilities, it is possible that for targets farther apart, such as different chromosomes, this effect could be eliminated.

Thus, oligonucleotide recombination should allow for the construction of large libraries at a single loop or moderate size libraries at two loops, but it is not yet sufficiently efficient to simultaneously mutate more than two loops. Assuming 10^7 transformants, an efficiency of 5% at a single loop predicts complete coverage of a library of 2×10^5. Interestingly, our actual coverage in the mock selection experiment was even larger than predicted by this estimation. We obtained ~8000 colonies from 2.6×10^6 unique transformants. This greater coverage may be attributed to the vast excess of oligonucleotides and plasmid molecules compared with the number of transformed cells, as was previously suggested by Truan and coworkers.[18] Thus, we may be able to generate libraries even higher than our recombination efficiencies would predict. Notably, this result demonstrates the significance of library-oriented experimental setups for testing novel mutagenesis techniques. The efficiency of nucleotide recombination at two loops of 0.2% predicts library coverage of 10^4. Therefore, multiple oligonucleotide recombination could be attractive for replacing iterative mutagenesis approaches where smaller, structure-based libraries are designed. For example, it has been shown that for directed evolution of enantioselective enzymes, simultaneous randomization is far more efficient than consecutive rounds of eror-prone-PCR.[43]

The recombination system presented here is engineered, and its mechanism is undefined at this point. The mutagenesis may occur via a combination of DSB repair and single-strand annealing (SSA) or during DNA replication.[44–46] Currently, we are investigating the mechanism using knockouts of yeast recombination machinery with the goal of improving the recombination efficiency, particularly at multiple loops. For example, deletion of *RAD51* could be used to establish that the mechanism involved is DSB repair,[44] while *RAD59* deletion may indicated SSA is taking place.[47,48] Furthermore, as previous studies suggest a strand bias exists that affects the efficiency of mutagenesis,[49] it will be interesting to see if such bias can be exploited to increase recombination efficiency at multiple loops.

While used here for the mutagenesis of loops in the active site of an enzyme, the oligonucleotide recombination system reported here should be broadly useful for library mutagenesis not only of individual proteins but also of other biomolecules such as RNA, multi-component systems such as metabolic pathways, and regions other than loops such as gene promoters. For example, this mutagenesis strategy could be used to randomize promoter strengths of multiple genes in a biosynthetic pathway to maximize production of a natural product in a heterologous host. Alternatively, the technology could be used to randomize the strength of interactions among multiple ribozymes in an engineered, artificial circuit. As synthetic biologists seek to engineer complex systems with increasing numbers of components, the need for directed evolution tools that allow large numbers of variations to be tested will only increase.

While recombination is now widely employed as a tool for targeted gene alteration, surprisingly it is not yet routinely used to generate large libraries of DNA mutants. There are only a handful of papers over the last two decades where recombination has been used to construct DNA libraries. Oligonucleotide recombination has not been characterized and optimized specifically for library generation. Problems of how to mutagenize multiple positions simultaneously as well as the mechanism underlying chimeric codon formation have not been addressed. *In vivo* recombination is faster, cheaper, and more straightforward to implement than current *in vitro* PCR-based mutagenesis techniques, although the technology is less well developed for library mutagenesis at this time. To meet synthetic biology's goal of cell engineering, the cell's own synthetic machineries will be co-opted to construct new building blocks and pathways directly in living cells.

Materials and Methods

Plasmid construction

Standard protocols for molecular biology and yeast genetics were used.[50,51] The materials and primers used in this study are listed in Table I and in the Supporting Information. The *TRP1* gene was subcloned into pHT2150 under control of the *MET25* promoter using primers VWC2031 and VWC2032 to generate plasmid pNP2278. Plasmids for expressing *trp1* mutants were created using overlap extension PCR. Two modified strands incorporating the *trp1-Arg44** mutation were made from p424MET25 using the primers VWC2031 and VWC2036, and VWC2032 and VWC2035. Following fusion, this fragment was amplified using VWC2034 and VWC2033, digested with SfiI and inserted into the multiple cloning site of the pHT2150 vector to generate pNP2279. These steps were repeated on pNP2279 using primers VWC2031 and VWC2038, and VWC2032 and VWC2037 to generate pNP2280, *trp1-Arg44*-Ser201** mutations. These steps were then repeated on pNP2280 but instead using primers VWC2031 and VWC2040, and VWC2032 and VWC2039 to generate pNP2281, *trp1-Arg44*Arg78*Ser201** mutations. The remaining three plasmids pNP2282, *trp1-Ser201** mutation, pNP2283, *trp1-Arg44*Arg78** mutations, and pNP2284, *trp1-Arg78** mutation, were created using restriction enzymes and ligation. pNP2278, pNP2280, and pNP2281 were digested with BstXI and SfiI to generate fragments 303 bp and 403 bp in length. The 303-bp fragment from pNP2278, carrying no mutations, and the 403-bp fragment from pNP2280, carrying the *trp1-Ser201** mutation, were ligated to make pNP2282. The 303-bp fragment from pNP2281, carrying the *trp1-Arg44*Arg78** mutations, and the 403-bp fragment from pNP2278, carrying no mutations were ligated to make pNP2283. Finally, pNP2278 and pNP2283 were digested with MfeI and SfiI to generate fragments 145 bp and 561 bp in length. The 145-bp fragment from pNP2278, carrying no mutations, and the 561-bp fragment from pNP2281, carrying the *trp1-Arg78** mutation, were ligated to make pNP2284. Linearized plasmids were prepared by digesting with SalI or ClaI, which cut once at the 3′ of *TRP1* gene, within the multiple cloning site.

Yeast transformation

We used a high efficiency yeast electroporation protocol with slight modifications[52] (see Supporting Information Methods for detailed protocol). Cells were plated on SC (Ura^-) and SC (Ura^-Trp^-) selective plates, which give transformants and recombinants, respectively. Prior to analysis, recombinant colonies were further grown in SC (Ura^-Trp^-) liquid selective media for 4 days with multiple seedings using fresh selective media. The cells were then plated on SC (Ura^-Trp^-) selective plates, and the resulted colonies were sequenced (see Supporting Information for full experimental details). Double stranded oligonucleotides were prepared by annealing complementary ss oligonucleotides (see Supporting Information Methods). To generate a mock library, fixing and nonsense oligonucleotides (Ser201LibraryFix and Ser201Library*Opal*, respectively) were mixed to a final molar ratio of 1 to 10^6 Ser201LibraryFix to Ser201Library*Opal* and co-transformed with a linear *trp1*-Ser201* vector by electroporation. The same amount of DNA oligonucleotides was used as in previous experiments, as well as the same optimized molar ratio of vector to oligonucleotides (1:1000).

Acknowledgments

The authors thank Professors Aaron P. Mitchell and Lorraine S. Symington for helpful discussions both in

the design of this project and during the writing of this manuscript. They also like to acknowledge Prof. Jingyue Ju and the Columbia Genome Center for high-throughput DNA sequencing.

References

1. Boder ET, Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol 15:553–557.
2. Smith M (1985) In vitro mutagenesis. Annu Rev Genet 19:423–462.
3. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH (2005) Evolving strategies for enzyme engineering. Curr Opin Struct Biol 15:447–452.
4. Botstein D, Shortle D (1985) Strategies and applications of in vitro mutagenesis. Science 229:1193–1201.
5. Klinner U, Schafer B (2004) Genetic aspects of targeted insertion mutagenesis in yeasts. FEMS Microbiol Rev 28:201–223.
6. Camps M, Naukkarinen J, Johnson BP, Loeb LA (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. Proc Natl Acad Sci USA 100:9727–9732.
7. Wang L, Jackson WC, Steinbach PA, Tsien RY (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. Proc Natl Acad Sci USA 101: 16745–16749.
8. Rothstein RJ (1983) One-step gene disruption in yeast. Methods Enzymol 101:202–211.
9. Kunes S, Ma H, Overbye K, Fox MS, Botstein D (1987) Fine structure recombinational analysis of cloned genes using yeast transformation. Genetics 115:73–81.
10. Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. Nucleic Acids Res 21:3329–3330.
11. Raymond CK, Pownder TA, Sexson SL (1999) General method for plasmid construction using homologous recombination. Biotechniques 26:134–141.
12. Igoucheva O, Alexeev V, Yoon K (2004) Oligonucleotide-directed mutagenesis and targeted gene correction: a mechanistic point of view. Curr Mol Med 4:445–463.
13. Storici F, Durham CL, Gordenin DA, Resnick MA (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. Proc Natl Acad Sci USA 100:14994–14999.
14. Costantino N, Court DL (2003) Enhanced levels of lambda red-mediated recombinants in mismatch repair mutants. Proc Natl Acad Sci USA 100:15748–15753.
15. Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet 20:123–128.
16. Linske-O'Connell LI, Sherman F, McLendon G (1995) Stabilizing amino acid replacements at position 52 in yeast iso-1-cytochrome c: in vivo and in vitro effects. Biochemistry 34:7094–7102.
17. Cherry JR, Lamsa MH, Schneider P, Vind J, Svendsen A, Jones A, Pedersen AH (1999) Directed evolution of a fungal peroxidase. Nat Biotechnol 17:379–384.
18. Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic Acids Res 28:e88.
19. Swers JS, Kellogg BA, Wittrup KD (2004) Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. Nucleic Acids Res 32:e36.
20. Hackel BJ, Kapila A, Wittrup KD (2008) Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. J Mol Biol 381:1238–1252.
21. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894–898.
22. Wilson JH, Leung WY, Bosco G, Dieu D, Haber JE (1994) The frequency of gene targeting in yeast depends on the number of target copies. Proc Natl Acad Sci USA 91:177–181.
23. Negritto MT, Wu X, Kuo T, Chu S, Bailis AM (1997) Influence of DNA sequence identity on efficiency of targeted gene replacement. Mol Cell Biol 17: 278–286.
24. Liu L, Rice MC, Drury M, Cheng S, Gamper H, Kmiec EB (2002) Strand bias in targeted gene repair is influenced by transcriptional activity. Mol Cell Biol 22: 3852–3863.
25. Hua SB, Qiu M, Chan E, Zhu L, Luo Y (1997) Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. Plasmid 38:91–96.
26. Branden C, Tooze J (1999) Introduction to protein structure, 2nd ed. New York, NY: Garland Publishing Inc.
27. May M (2009) Engineering a new business. Nat Biotechnol 27:1112–1120.
28. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA, III (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc Natl Acad Sci USA 105:20404–20409.
29. Yamamoto T, Moerschell RP, Wakem LP, Ferguson D, Sherman F (1992) Parameters affecting the frequencies of transformation and co-transformation with synthetic oligonucleotides in yeast. Yeast 8:935–948.
30. Hommel U, Eberhard M, Kirschner K (1995) Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. Biochemistry 34:5429–5439.
31. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201.
32. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38, 27–28.
33. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285:901–906.
34. Tao H, Peralta-Yahya P, Decatur J, Cornish VW (2008) Characterization of a new glycosynthase cloned using chemical complementation. ChemBiochem 9:681–684.

35. Orr-Weaver TL, Szostak JW, Rothstein RJ (1981) Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci USA 78:6354–6358.
36. Wendland J (2003) PCR-based methods facilitate targeted gene manipulations and cloning procedures. Curr Genet 44:115–123.
37. Storici F, Resnick MA (2006) The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. Methods Enzymol 409:329–345.
38. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res 31:3497–3500.
39. Smith GR (2001) Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. Annu Rev Genet 35:243–274.
40. Alcalde M, Zumarraga M, Polaina J, Ballesteros A, Plou FJ (2006) Combinatorial saturation mutagenesis by in vivo overlap extension for the engineering of fungal laccases. Comb Chem High Throughput Screen 9:719–727.
41. Agarwal S, Gamper HB, Kmiec EB (2003) Nucleotide replacement at two sites can be directed by modified single-stranded oligonucleotides in vitro and in vivo. Biomol Eng 20:7–20.
42. Hartwell L (2004) Genetics: from genes to genomes, 2nd ed. Boston, MA: McGraw-Hill Higher Education.
43. Reetz MT, Kahakeaw D, Lohmer R (2008) Addressing the numbers problem in directed evolution. Chembiochem 9:1797–1804.
44. Bartsch S, Kang LE, Symington LS (2000) RAD51 is required for the repair of plasmid double-stranded DNA gaps from either plasmid or chromosomal templates. Mol Cell Biol 20:1194–1205.
45. Court DL, Sawitzke JA, Thomason LC (2002) Genetic engineering using homologous recombination. Annu Rev Genet 36:361–388.
46. Agmon N, Pur S, Liefshitz B, Kupiec M (2009) Analysis of repair mechanism choice during homologous recombination. Nucleic Acids Res 37:5081–5092.
47. Davis AP, Symington LS (2001) The yeast recombinational repair protein Rad59 interacts with Rad52 and stimulates single-strand annealing. Genetics 159: 515–525.
48. Krogh BO, Symington LS (2004) Recombination proteins in yeast. Annu Rev Genet 38:233–271.
49. Parekh-Olmedo H, Kmiec EB (2007) Progress and prospects: targeted gene alteration (TGA). Gene Ther 14: 1675–1680.
50. Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K (1995) Current Protocols in Molecular Biology. New York: Wiley.
51. Adams A (1998) Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual, 1997 ed. Plainview, NY: Cold Spring Harbor Laboratory Press.
52. Wittrup D (2007) Wittrup: yeast transformation. http://openwetware.org/wiki/Wittrup:_Yeast_Transformation. Accessed October 20, 2010.
53. Mezard C, Pompon D, Nicolas A (1992) Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell 70:659–670.
54. Hennig M, Sterner R, Kirschner K, Jansonius JN (1997) Crystal structure at 2.0 A resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. Biochemistry 36:6009–6016.