Offset recombinant PCR: a simple but effective method for shuffling compact heterologous domains

David A. Rozak* and Philip N. Bryan

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA

Received January 18, 2005; Revised March 18, 2005; Accepted May 2, 2005

ABSTRACT

DNA shuffling and other in vitro recombination strategies have proven highly effective at generating complex libraries for mutagenesis studies. While most recombination techniques employ DNA polymerases in part of a multi-step process, few seek to exploit the natural recombinogenic tendencies and exponential amplification rates of PCR. Here, we characterize a simple but effective method for using standard PCR to promote high recombination frequencies among compact heterologous domains by locating the domains near one end of the template. In a typical amplification reaction, Pfu polymerase generated chimeric crossover events in 13% of the population when markers were separated by only 70 nt. The fraction of recombinant sequences reached 42% after six consecutive rounds of PCR, a value close to the 50% expected from a fully shuffled population. When homology within the recombinant region was reduced to 82%, the recombination frequency dropped by nearly half for a single amplification reaction and crossover events were clustered toward one end of the domain. Surprisingly, recombination frequencies for template populations with high and low sequence homologies converged after just four rounds of PCR, suggesting that the exponential accumulation of chimeric molecules in the PCR mixture serves to promote recombination within heterologous domains.

INTRODUCTION

The intrinsic ability of PCR to generate recombinant products from mixed homologous template populations was recognized as early as 1988 when researchers observed the appearance of chimeric products during the amplification of two alleles with the Klenow fragment of DNA polymerase I (1). Similar results were later described for amplification reactions involving Taq and Vent polymerases (2–4). In each of these cases, recombination occurred at a relatively low frequency, making the phenomena more of an inconvenience for researchers interested in amplifying allelic DNA than an effective mechanism for in vitro recombination. Despite the frequent reliance of DNA shuffling strategies on in vitro polymerization reactions as part of a multi-step process, researchers have yet to embrace PCR itself as an effective recombination technique.

PCR-based recombination is thought to occur when a primer is extended first on one template and then on another to form a chimeric molecule with a distribution of genetic markers that differs from either of the parent templates. In order for this process to play an appreciable role in the amplification reaction, primers must regularly be terminated between template polymorphisms during one PCR cycle and reanneal to a template with a different assortment of genetic markers during a later cycle. A successful in vitro recombination technique known as the staggered extension process (StEP) uses a highly abbreviated annealing/elongation phase to generate nested primers and promote crossover events along the full length of the template (5). However, because the process requires multiple cycles to fully extend a single primer, StEP fails to achieve the exponential product growth that is characteristic of PCR.

The efficient reannealing of partially extended primers to new templates is another important factor in the formation of chimeric PCR products. During each annealing phase, unextended primers, partially extended primers and full-length templates compete with one another to form DNA duplexes. PCR amplification mixes are generally saturated with excess amounts of unextended primers to efficiently prime the exponentially growing template populations. In contrast, partially extended primers, which are needed to promote recombination, are relatively few—especially in the early stages of PCR—and unable to effectively compete with unextended primers for a limited number of templates. Given these unfavorable conditions, partially extended primers are more likely to accumulate in the reaction mixture than contribute to the formation of chimeric products. By adding a separate high-temperature annealing phase to the standard PCR

*To whom correspondence should be addressed. Tel: +1 301 738 6291; Fax: +1 301 738 6225; Email: rozak@umbi.umd.edu

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org
cycle, Judo (6) was able to preferentially promote the reannealing of partially extended primers and achieve a noticeable increase in the recombination frequency for Vent and Taq polymerases. Significantly, Judo’s modified PCR protocol used a 45 s elongation phase as opposed to the highly abbreviated one employed by StEP. While Judo’s longer extension time supported exponential growth, it is unclear how it affected the distribution of crossover points within the amplicon.

While Judo relied on altered polymerase cycling conditions to promote PCR recombination, it should be possible to achieve similar results by simply positioning the recombinant region toward one end of the amplicon. Careful consideration of PCR chemistry suggests that primers extended significantly beyond the amplicon’s midpoint should be unaffected by competition from large concentrations of unextended primers. Rather, nearly extended forward and reverse primers can freely anneal to one another without competing with unextended primers for full-length templates (Figure 1). Furthermore, by locating the recombinant region on one end of the amplicon, it may be possible to generate an adequate distribution of crossover events in the targeted region without resorting to highly abbreviated elongation times. In short, standard PCR could be used as an effective in vitro recombination technique for offset regions.

Here, we describe the use of a lacZ reporter system to characterize PCR-induced recombination between markers that are located at one end of the amplicon. By varying the reaction conditions, we were able to explore the effects of cycle number, extension time and sequence homology on recombination frequencies near product ends. Our results suggest that this strategy—referred to here as offset recombinant PCR (OR-PCR)—offers a simple but effective approach for generating recombinant libraries of compact heterologous domains.

**Figure 1.** The anticipated impact of DNA duplexes on PCR-mediated recombination in centered and offset regions. Primers and templates are depicted as half arrows pointing toward their 3’ ends. The complexes favored by high primer and template concentrations are rendered in bold. Assuming incomplete elongation during earlier PCR cycles, an annealing phase can result in the formation of DNA duplexes between unextended primers, partially extended primers and completed templates. Owing to competition for full-length templates from excess amounts of unextended primers, duplexes are less likely to involve template–template pairs and templates paired with partially extended primers. However, those primers that have been sufficiently extended beyond the product’s midpoint during an earlier elongation cycle (indicated by asterisks) are more likely to form duplexes with their counterparts in the reverse direction. This suggests that primers terminated between markers near the product end (Region A) are more likely to reanneal and extend to form chimeras than primers terminated at or before the product’s center (Region B).

**MATERIALS AND METHODS**

**pUC19 mutants**

The QuickChange™ Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) was used to insert ochre codons and silent point mutation into the N-terminal region of the lacZ (open reading frame) ORF on the pUC19 plasmid (GenBank accession no. L09137 X02514). Where necessary, we followed the protocol outlined by Wang and Malcolm (7) for QuickChange™ mutagenesis reactions involving primers that exceeded the 40 nt limit recommended by Stratagene. XL-10 Gold Super Competent Cells (Stratagene) were transformed with pUC19 mutants and spread along with 1.2 mg 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 5 μmol isopropyl β-D-thiogalactopyranoside (IPTG) on Luria–Bertani (LB) broth agar plates containing 100 μg/ml ampicillin to confirm the absence of blue colonies containing lacZ’ genes. pUC19 mutagenesis was also confirmed by DNA sequencing.

**Polymerase chain reaction**

All PCR recombination experiments were performed under the same general reaction conditions with variations noted in the Results section. Reaction mixtures consisted of 2.5 U cloned Pfu polymerase (Stratagene), 200 μM each dNTP, 0.5 μmol each primer and a 100 ng equal mixture of pUC19 mutants in 50 μl of the recommended reaction buffer. Each of the amplification reactions used the P1 primer combined with either P2 or P3 to create 329 or 511 bp products, respectively (P1: 5′-TAA CTA TGC GGC ATC AGA GC-3′; P2: 5′-GAC CAT GAT TAC GCC AAC C-3′; P3: 5′-GCG TTG GCC GAT TCA TTA-3′). Thermocycling began with 30 s at 95°C followed by 30 cycles of 30 s at 95°C, 30 s at 55°C and 1 min at 75°C. PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen, Valencia, CA), and concentrations were determined via ultraviolet (UV) absorption at 260 nm.

**Transformation and screening**

*Escherichia coli* transformation and screening began with the restriction digest of recovered PCR products and subsequent ligation back into the pUC19 vector. Seventy-five nanograms of the purified PCR product was cut with NdeI and HindIII before being repurified with the QIAquick® PCR Kit or QIAquick® Gel Extraction Kit (Qiagen). When the pUC19 mutants had been amplified using P1 and P2, the QIAquick® PCR Kit was used to efficiently remove the short terminal fragments cleaved from the PCR product during the digest reaction. However, when pUC19 mutants had been amplified using the P1 and P3 primers, gel filtration was necessary to isolate the 264 bp recombinant restriction fragment for ligation into pUC19. In this case, the corresponding band was excised from the gel and cleaned up with the QIAquick® Gel Extraction Kit.

A 50 ng sample of a pUC19 lacZ- mutant was cut with NdeI and HindIII and treated with calf intestinal alkaline phosphatase before purification with the QIAquick® PCR Kit. A pUC19 lacZ knockout mutant was used as a cloning vector to avoid the possibility that small amounts of undigested plasmids could contribute to the observed phenotype rescue frequency. Digested PCR products and pUC19 vectors were...
mixed and incubated for 2 h at 25°C with T4 DNA ligase before a 1 μl aliquot was used to transform 20 μl XL-10 Gold Super Competent Cells with a 30 s heat shock at 42°C. After 1 h at 37°C, the cells were spread along with X-Gal and IPTG on LB plates containing ampicillin and grown overnight at 37°C to observe the fraction of blue colonies on the plate—described throughout this paper as the phenotype rescue frequency. A minimum of 1000 colonies were sampled for each data point. Experimental errors reported throughout this paper represent the statistical uncertainty inherent in the size of the sampled population.

**DNA sequencing**

DNA samples were prepared for sequencing by growing selected colonies overnight at 37°C in LB with 100 μg/ml ampicillin. Plasmid DNA was extracted from the cell cultures using the Wizard® Plus SV Minipreps (Promega Corporation, Madison, WI). The P2 primer was used to amplify target DNA using Perkin-Elmer/Applied Biosystem’s AmpliTaq-FS DNA polymerase and Big Dye terminators with dITP. Dye-terminated products were then run on an Applied Biosystems model 3100 DNA sequencer to produce sequence chromatographs.

**RESULTS**

**The lacZ reporter system**

We first sought to create a simple reporter system in which phenotype rescue frequencies could be used to monitor recombination among lacZ alleles with knockout mutations. Site-directed mutagenesis was used to eliminate the β-galactosidase phenotype by placing pairs of adjacent ochre stop codons at two different points within the lacZ ORF. As expected, cells transformed with the pUC19-03 (ochre mutations at β-galactosidase positions 35 and 36) and pUC19-05 (ochre mutations at positions 9 and 11) mutants failed to produce blue colonies. These pUC19 constructs are diagrammed in Figure 2A and collectively referred to as set A mutants.

Two variants of pUC19-03 and pUC19-05, which contain a total of eight silent point mutations in the 82 nt region, were also created to explore PCR recombination among heterologous stretches of DNA. These variants, which were designated pUC19-06 and pUC19-07 and are collectively referred to as set B mutants, exhibit an 82% DNA sequence homology to one another within the recombinant region (Figure 2A).

Three primers were designed to amplify mixed populations of the pUC19 mutants. As shown in Figure 2B, when the pUC19 mutants are amplified with P1 and P2, the 82 nt recombinant region is located toward one end of a 329 bp product. However, when pUC19 populations are amplified using P1 and P3, the recombinant region is centered on a 511 bp product. These two primer combination were used in this study to compare recombination frequencies between offset and centered markers.

Paired stop codons were designed into each of the pUC19 mutants to reduce the chance of phenotype recovery by random point mutations. To validate this approach, pUC19-03 and pUC19-05 constructs were amplified separately using P1 and P2 primers and 1 min elongation times. The PCR products were then ligated back into a pUC19 lacZ0 construct and used to transform E.coli. Fewer than 1 in 1000 colonies transformed with pUC19-03, or pUC19-05 PCR products recovered the blue lacZ+ phenotype through simultaneous point mutations in both of the paired stop codons.

---

**Figure 2.** Design of the lacZ reporter system. (A) Point mutations for each of the pUC19 constructs used in our PCR recombination experiments. Although all nucleotide sequences in this paper are presented in terms of the lacZ coding strand, they are numbered in accordance with their compliments for consistency with the published pUC19 sequence. Nucleic acid mutations that lead to the creation of ochre stop codons in the lacZ ORF are shown in bold. All other mutations have no impact on the wild-type β-galactosidase amino acid sequence, which is aligned with the nucleotide sequences at the bottom of the frame. (B) Locations of primers, restriction sites and the 82 bp recombinant region (shaded box) with respect to the lacZ ORF on the pUC19 vector. Dashed lines mark center points for the 329 and 511 bp PCR products. Half arrows represent the arrangements of forward and reverse primers (P1, P2 and P3) on the two amplicons.
Phenotype rescue as a function of OR-PCR cycle

Equal mixtures of set A mutants were amplified with P1 and P2 to observe phenotype rescue via the recombination of offset markers as a function of OR-PCR cycle number. Reaction mixtures and cycling conditions were selected to reproduce a typical amplification reaction. These cycling conditions included a 1 min elongation phase to promote complete extension of the 329 bp PCR products based on a previously reported Pfu elongation rate of 25 bases/s (8). Identical amounts of PCR product, as determined by OD$_{260}$, were taken from the thermocycler on odd cycles and ligated into pUC19 for transformation and screening. Absorbance readings confirmed an exponential growth in PCR products. As shown in Figure 3, phenotype rescue was undetected in sampled colonies until the 11th cycle and reached a frequency of 0.11 ± 0.02 blue colonies per sampled population after 30 cycles. The higher ratios of blue colonies generated during the second half of the cycling reaction are consistent with the observations of PCR-induced recombination reported elsewhere (3,6).

Phenotype rescue as a function of elongation time

Amplification reactions were run with elongation times ranging from 1 to 120 s in order to observe the effect on OR-PCR. After 30 cycles, equal concentrations of each PCR product were ligated into pUC19 lacZ vectors and transformed into E.coli to observe phenotype rescue. The results, which are plotted in Figure 4, suggest that elongation times of <40 s dramatically increase the frequency of phenotype rescue to as much as 18%. Within the limits of experimental error, no significant difference was observed in phenotype rescue frequencies for elongation times between 1 and 40 s. This broad range of effective polymerase extension times suggests that the hyper-attenuated 5 s 55°C annealing/elongation phase employed by StEP (5) is not necessary to promote distributed recombination events among offset markers. Unless otherwise stated, reactions described in the rest of this paper were conducted with 15 s elongation times.

Comparing phenotype rescue frequencies for centered and offset markers

In order to compare phenotype rescue frequencies for centered and offset markers, the amplification reaction described above was run with 15 s elongation times using P1 and P3 primers to generate a 511 bp product with the 82 nt recombinant region centered ~255 bp from each end (Figure 2B).

Because P1 and P3 are equidistant from the recombinant region, the 15 s elongation phase should be equally effective in terminating both primers within the 82 nt region to promote observable crossover events. However, primers terminated within this region will also be within 40 nt of the amplicon’s center, reducing the chance they can avoid competition from unextended primers by annealing to one another. These less favorable reannealing conditions should have an observable effect on the phenotype rescue frequency.

Indeed, when PCR products were ligated into the pUC19 lacZ vector and expressed in E.coli, they exhibited a phenotype rescue frequency of 0.14 ± 0.1 compared with the frequency of 0.17 ± 0.1 observed for the P1- and P2-primed templates with an offset recombinant region. Because elongation times have not changed and P1 is just as likely to be terminated within the 82 nt recombinant region as in previous experiments, the drop in phenotype rescue frequency is likely attributable to the reduced ability of center-terminated P1 and P3 primers to form extendable complexes.

The negative impact of the centered markers on primer-mediated recombination is even more pronounced when one considers that both P1 and P3 are equally likely to be terminated in the centered recombinant region. Therefore, the phenotype rescue frequency observed for centered markers is twice that attributable to the reannealing and extension of P1 or P3 alone. This should be compared with experiments involving offset markers where elongation conditions made it highly unlikely that P2 would be terminated within the neighboring recombinant region and phenotype rescue frequencies were almost exclusively attributable to the action of P1. This analysis suggests that P1 is nearly two-and-a-half times as effective at generating recombinants among offset markers than centered markers.
Similar results were observed when this experiment was repeated with 1 min elongation times. Under these more typical amplification conditions, centered markers generated a phenotype rescue frequency of 0.11 ± 0.02, which fails to surpass frequencies generated by offset markers under identical reaction conditions, despite the advantage of both forward and reverse primers having equal chances of terminating between centered markers and contributing to recombination.

**OR-PCR recombination of heterologous DNA**

A major goal of our study is to probe the extent to which OR-PCR can promote recombination among heterologous stretches of DNA. We explored these limits by amplifying equal quantities of set B mutants, which exhibit 82% homology within the 82 nt recombinant region, with P1 and P2 primers and 15 s elongation times. Cells transformed with set B-derived OR-PCR products produced phenotype rescue frequencies of 0.11 ± 0.01 compared with 0.17 ± 0.01 observed when set A mutants were amplified under identical conditions.

In order to study the distribution of crossover events between the closely spaced nucleotide polymorphisms found among set B constructs, plasmid DNA was purified and sequenced from 16 colonies exhibiting the blue lacZ phenotype. The distribution of sequence polymorphisms (Figure 5) suggests that crossovers occurred primarily in the half of the recombinant region closest to the P1 primer. As per these results, Pfu polymerase is able to elongate partially extended P1 primers that reanneal to their heterolog in this region, despite the presence of mismatched bases within as few as 4 nt from the 3' terminal. A similar tolerance for 3' mismatches in heteroduplexed DNA has been reported elsewhere for Taq polymerase (9). However, Pfu polymerase appeared to be incapable of extending oligonucleotides that traversed more than half of the 82 nt heterologous region, possibly due to a decrease in local annealing temperature that results from an accumulation of 3' mismatches. If this is the case, the adverse effects of 3' heteroduplex instability on annealing and elongation might be mitigated by the use of lower annealing temperatures.

A careful review of DNA chromatographs revealed that half the samples represented in Figure 5 were derived from colonies containing two alleles of the lacZ gene. The mixed alleles most likely resulted from heteroduplex formation during PCR. In each of these cases, one allele represented the lacZ gene with its dominant blue phenotype while the other appeared to be a non-recombinant lacZ gene identical to that of the parental pUC19-07 construct. pUC19-06 derivatives were conspicuously absent from all of the sequenced samples. Heteroduplex formation is to be expected from PCR amplification of mixed templates. However, the presence of mixed alleles in only half of the sequenced samples and complete absence of stop codons at β-galactosidase positions 9 and 11 was unforeseen. Random assortment suggests that nearly all of the sequenced samples should show the less common recombinant lacZ gene paired with one of the more abundant lacZ mutants. Furthermore, the lacZ mutants should represent an equal distribution of the stop codon mutations found in both pUC19-06 and pUC19-07.

The low incidence of mixed alleles and complete absence of pUC19-06 derived N-terminal stop codons in the sequenced samples can be explained by the possibility that plasmids containing N-terminal stop codons are not well maintained in transformants. In this manner, the propagation of cells containing lacZ genes paired with lacZ genes possessing N-terminal stop codons would lead to the eventual loss of the unstable lacZ variant—whose truncated products may serve to exhaust cell resources and significantly impede growth—leaving only lacZ plasmids in the sampled population. Because cellular dynamics appear to alter the distribution of pUC19 mutants, observed phenotype rescue frequencies probably differ from actual OR-PCR recombination frequencies.

**Estimating OR-PCR recombination frequencies via serial amplification reactions**

Equal mixtures of set A mutants were subjected to consecutive rounds of OR-PCR in order to assess the degree to which observed phenotype rescue frequencies differ from OR-PCR recombination frequencies. We hoped that serial amplification reactions would lead to an accumulation of recombinant products and reveal an asymptotic approach to a maximum phenotype rescue frequency that could be used to estimate the recombination frequency.

Serial reactions were performed by taking a 2 µl aliquot from a completed 30-cycle OR-PCR and transferring it to a fresh reaction buffer for another round of thermocycling. A 75 ng sample of the product from each serial reaction was ligated into the pUC19 vector and transformed into E.coli to measure phenotype rescue frequency. The results of these experiments, which are reported as filled squares in Figure 6, show that more than half of the colonies exhibit the blue lacZ phenotype after being transformed with DNA derived from six consecutive reactions. This is significantly above the maximum phenotype rescue frequency of 43.75% expected from random duplexes of recessive lacZ and dominant lacZ alleles if the latter cannot exceed one quarter of a fully shuffled

---

**Figure 5.** Assortment of single nucleotide polymorphisms in 16 DNA sequences derived from blue colonies transformed with set B recombinants. pUC19-06 derived markers are indicated by upward ticks, while pUC19-07 derived markers are depicted with downward ticks. Crossover events (shaded boxes) are concentrated in the half of the recombinant region closest to the P1 primer. Asterisks indicate those sequences derived from colonies that also appeared to contain the non-recombinant pUC19-07 construct.
set A population. The higher than expected phenotype rescue frequency may be due to sequence-specific biases introduced during DNA recovery or post-transformational processing in *E. coli*. The absence of N-terminal stop codons observed above suggests that some sequence-specific selection is taking place. However, we felt that further efforts to explore the mechanisms behind these phenomena would exceed the scope of this study and have little impact on the conclusions presented here.

Even in the absence of a clearly articulated mechanism for the post-recombinant fate of *lacZ* DNA, it is possible to arrive at an initial estimate of the rescue frequency by fitting experimental data with a statistical model that allows for the observed bias. The recombination frequency for a 30-cycle amplification reaction can be expressed as the probability $P$ that a single-stranded DNA product contains an assortment of polymer during a 30-cycle amplification reaction, one can use the binomial expansion to compute the chance a sequence undergoes $k$ chimeric recombination events during $n$ serial OR-PCR amplifications:

$$b(P, n, k) = \binom{n}{k} P^k (1-P)^{n-k}.$$  \hspace{1cm} 1

Only those sequences that have undergone an odd number of chimeric recombination events ($k = 1, 3, 5, ...$) during $n$ consecutive rounds of PCR will contain an assortment of stop codons different than those found in the original population. This subset of recombinant sequences will either contain knockout mutations from both constructs or none at all.

Since these two species should be relatively uniform in number, only half of the sequences that have undergone an odd number of chimeric recombination events will lack both sets of knockout mutations and represent functional *lacZ* genes. Therefore, the chance of a DNA strand encoding a functional $\beta$-galactosidase protein after $n$ serial rounds of PCR can be expressed as $Z(P, n)$ where:

$$Z(P, n) = \frac{1}{2} \sum_{k=1,3,5,\ldots} b(P, n, k).$$  \hspace{1cm} 2

Assuming the largely homologous PCR products randomly anneal to one another at the end of the amplification reaction, it is likely that each cell will be transformed with heteroduplexed pUC19 mutants representing two distinct alleles. The dominant blue phenotype will be observed in colonies if at least one of the two pUC19 strands from the original cell codes for a functional *lacZ* gene. If events $E_1$ and $E_2$ represent the incorporation of *lacZ* sense and antisense strands into a pUC19 heteroduplex, the probability $B$ of finding a blue colony on the plate can be expressed as:

$$B = P\{E_1 \cup E_2\} = P\{E_1\} + P\{E_2\} - P\{E_1E_2\}.  \hspace{1cm} 3$$

Assuming

$$P\{E_1\} = P\{E_2\} = Z(P, n),$$

this expression simplifies to

$$B(P, n) = 2Z(P, n) - Z(P, n)^2.$$

Finally, we multiplied Equation 5 by the constant $c$ to reflect the observed bias toward *lacZ*-encoding heteroduplexes. In the absence of further data on the cellular fate of recombinant DNA, $c$ provides a reasonable approximation of the impact post-recombinant factors have on the fraction of blue colonies, provided these factors act in a manner that is largely independent of $n$ and $P$.

$$B(c, P, n) = c \left[ 2Z(P, n) - Z(P, n)^2 \right].$$ \hspace{1cm} 6

By setting $B(c, P, n)$ equal to the experimentally determined fraction of blue colonies for $n$ serial offset recombination reactions, it is possible to fit the equation to the experimental data in Figure 6 by varying $c$ and $P$, which respectively impact the asymptotic height and rate of assent for the curve. Figure 6 shows a reasonably tight fit to experimental data from the serial recombination of pUC19-03 and pUC19-05 for $P = 0.13$ and $c = 1.36$. Since the error bars in Figure 6 are based solely on statistical uncertainty inherent in the sample size, it is understandable that some values for $B(c, P, n)$ fall slightly outside of these ranges, possibly due to procedural errors, which are not reflected in the error estimates.

While $P$ represents the fraction of the population containing a reassortment of terminal markers after one 30-cycle OR-PCR, the recombinant fraction is as high as 0.42 after six consecutive rounds of OR-PCR as computed by $2Z (P = 0.13, n = 6)$.

**Effects of serial amplification reactions on heterologous recombination**

In order to observe the effects of serial OR-PCR on heterologous recombination, we repeated the preceding experiment using the set B constructs. The results, which are plotted as

---

Figure 6. Phenotype rescue frequencies observed after consecutive rounds of OR-PCR. Frequencies are derived from reactions involving set A (filled boxes) and set B (filled triangles) template mixes. Error bars express the statistical error inherent in the sample size. Phenotype rescue frequencies predicted by Equation 6 are shown for $c = 1.36, P = 0.13$ (open boxes) and $P = 0.08$ (open triangles). These values were selected to produce the best fit to experimental data.
filled triangles in Figure 6, show a distinct convergence of phenotype rescue frequencies for set A and B recombinants after only four rounds of PCR. Although one would expect phenotype rescue frequencies for each set to asymptotically approach a common maximum as the markers become evenly distributed among members of the population, the plots converge well before either show signs of leveling off at the maximum phenotype rescue rate of 0.6 predicted by Equation 6 when \( c = 1.36 \). When we attempt to fit Equation 6 to the data obtained from set B recombinants, experimentally derived rescue frequencies appear to undergo a clear transition during PCR rounds 3 and 4 between the datasets described by \( B(c = 1.36, P = 0.08, n) \) and \( B(c = 1.36, P = 0.13, n) \). One possible explanation for this phenomenon lies in the growing fraction of recombinant templates and partially extended primers carried over from one round to the next. During consecutive rounds of OR-PCR, set B-derived templates and primers increasingly represent a more homologous mixture of pUC19-06 and pUC19-07 markers. This may serve to increase the average homology of heteroduplexed pairs and make it more likely that a partially extended P1 primer will form an extendable heteroduplex with other members of the population.

Plasmid DNA was sequenced from 14 blue and 11 white colonies that had resulted from the six consecutive rounds of PCR performed on set B plasmids. To avoid sequencing mixed alleles, plasmid samples obtained from the original colonies were transformed back into E. coli and purified from individual plated colonies prior to sequencing. As a result, all of the sequences depicted in Figure 7 appear to have been derived from monoallelic samples.

As expected, sequences isolated entirely from blue colonies (Figure 7A) lacked the stop codons associated with pUC19-06 and pUC19-07 mutants and showed evidence of at least one crossover event. Two of the sequences revealed three distinct crossover events. Consistent with the limitations placed on \( k \) in Equation 2, even numbers of crossover events were not observed among these sequences.

Only half of the sequences isolated from white colonies showed signs of recombination (Figure 7B). Each of these appeared to have undergone two recombination events. Furthermore, none of the 12 sequences had a stop codon at the position 11 of the \( lacZ \) gene and only two had stop codons at position 9. The low incidence of N-terminal stop codons among white colonies reinforces the earlier hypothesis that pUC19 mutants with the N-terminal stop codons are poorly retained in the bacteria.

Even after six rounds of PCR, crossover events continue to show a distinct preference for the end of the recombinant region closest to P1. However, crossover events do occur further into the heterologous region and are found as far as the 4 bp region between the two \( lacZ \) N-terminal stop codons. As a testimony to the \( Pfu \) fidelity, none of the sequences exhibited point mutations within the recombinant region, even after six consecutive rounds of OR-PCR.

**DISCUSSION**

Our results indicate that the location of markers on an amplification product, polymerase elongation time and number of
Table 1. Recombination frequencies for several different recombination strategies

| Strategy     | Size of recombinant region (nt) | Homology within region (%) | Chimeric recombination frequency | Polymerase | Thermocycles | Special cycling conditions |
|--------------|---------------------------------|-----------------------------|----------------------------------|------------|--------------|----------------------------|
| OR-PCR       | 82                              | 91                          | 0.42                             | Pfu        | 30 × 6       | Serial reactions with optimized elongation time |
|              | 82                              | 82                          | 0.42                             | Pfu        | 30 × 6       | Serial reactions with optimized elongation time |
|              | 82                              | 91                          | 0.13                             | Pfu        | 30           | Optimized elongation time               |
|              | 82                              | 82                          | 0.08                             | Pfu        | 30           | Optimized elongation time               |
|              | 82                              | 91                          | 0.08                             | Pfu        | 30           | None                                        |
| Centered PCR | 287                             | 99                          | 0.21                             | Taq        | 30           | High-temperature annealing phase (6)     |
|              | 287                             | 99                          | 0.19                             | Vent       | 25           | High-temperature annealing phase (6)     |
|              | 287                             | 99                          | 0.14                             | Taq        | 25           | High-temperature annealing phase (6)     |
|              | 287                             | 99                          | 0.07                             | Vent       | 25           | None (6)                                  |
|              | 287                             | 99                          | 0.01                             | Taq        | 25           | None (6)                                  |
| StEP         | 113                             | 96                          | 0.39                             | Taq        | 80           | Highly abbreviated elongation phase (5)  |
|              | 260                             | 95                          | 0.18                             | Vent       | 95           | Highly abbreviated elongation phase (10) |

*This region is inclusive of the two observable genetic markers at either end.

*The chance that a product will contain an assortment of genetic markers at the ends of the recombinant region different from those found in the original population.

at the primer’s 3′ end and a corresponding drop in the local melting temperature of the DNA heteroduplex. This result is mitigated by serial passage of the recombinant library through multiple OR-PCR amplifications. Given the low incidence of point mutations and other signs of template degradation after six consecutive rounds of amplification, serial OR-PCR may be an effective means of enhancing recombination among heterologous alleles.

Perhaps, the most intriguing observation to come out of this study is the apparent shift in recombination frequencies that resulted from serial amplifications of the heterologous pUC19-06 and pUC19-07 constructs. Sequence data suggest that this phenomenon may result from modest increases in the overall homogeneity of the template populations and highlights an underlying advantage of PCR-based recombination over competing techniques. The chain reaction phenomenon permits the rapid accumulation of recombinant templates in the mix, which provide a diverse supply of substrates to support the binding and extension heterologous primers. In fact, under the exponential amplification conditions, even the primer pool grows more diverse as primers are partially elongated on recombinant templates generated during earlier cycles. The net result is a homogenized population of heterologs, which have a better chance of forming extendable duplexes with one another from one cycle to the next.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM62154. The authors would like to thank Fenfong Song for providing the primers used in this study and are, as always, grateful for the ready guidance and advice of Kathryn Fisher, Patrick Alexander and Biao Ruan whose knowledge and experience helped overcome a number of procedural challenges. Funding to pay the Open Access publication charges for this article was provided by NIH Grant GM62154.

Conflict of interest statement. None declared.

REFERENCES

1. Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.
2. Bradley,R.D. and Hillis,D.M. (1997) Recombinant DNA sequences generated by PCR amplification. *Mol. Biol. Evol.*, **14**, 592–593.
3. Meyerhans,A., Vartanian,J.P. and Wain-Hobson,S. (1990) DNA recombination during PCR. *Nucleic Acids Res.*, **18**, 1687–1691.
4. Yang,Y.L., Wang,G., Dorman,K. and Kaplan,A.H. (1996) Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. *AIDS Res. Hum. Retroviruses*, **12**, 303–306.
5. Zhao,H., Giver,L., Shao,Z., Affholter,J.A. and Arnold,F.H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.*, **16**, 258–261.
6. Judo,M.S., Wedel,A.B. and Wilson,C. (1998) Simulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.*, **26**, 1819–1825.
7. Wang,W. and Malcolm,B.A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuickChange site-directed mutagenesis. *Biotechniques*, **26**, 680–682.
8. Takagi,M., Nishioka,M., Kakihara,H., Kitabayashi,M., Inoue,H., Kawakami,B., Oka,M. and Imanaka,T. (1997) Characterization of DNA polymerase from Pyrococcus sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.*, **63**, 4504–4510.
9. Kwok,S., Chang,S.-Y., Sninsky,J.J. and Wang,A. (1995) In *Dieffenbach,C.W. and Dveksler,G.S. (eds), PCR Primer: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 143–155.
10. Ninkovic,M., Dietrich,R., Arai,G. and Schwienhorst,A. (2001) High-fidelity in vitro recombination using a proofreading polymerase. *Biotechniques*, **30**, 530–536.