Bioengineered Bugs 1:5, 337-340; September/October 2010; © 2010 Landes Bioscience

Rapid generation of random mutant libraries

Mary Abou-Nader and Michael J. Benedik*

Department of Biology; Texas A&M University; College Station, TX USA

Key words: in vivo cloning, lambda red proteins, error prone PCR, mutant libraries, directed evolution

A simple and efficient method utilizing in vivo recombination to create recombinant libraries incorporating the products of PCR amplification is described. This will be especially useful for generating large pools of randomly mutagenized clones after error-prone PCR mutagenesis. Here we investigate various parameters to optimize this approach and we demonstrate that as little as 1 pmole of PCR fragment can generate a library with greater than 10^4 clones in a single transformation without ligation.

Introduction

Directed evolution is often used in protein engineering to improve the properties of several enzymes.[1,2] However, its success hinges on the ability to generate libraries of significant size and diversity. A variety of methods for mutagenesis and subsequent library construction have been previously described along with appropriate methods for screening or selection. Herein, we will describe a simple and streamlined approach to rapidly generate diverse randomly mutagenized libraries.

Error Prone PCR (EP-PCR) is a method of choice to generate random mutations throughout a gene or gene region without being limited by a size constraint (within the limits of conventional PCR). It relies on the use of *Taq* or similar DNA polymerase lacking proofreading activity so that mismatched bases are not removed.[3] To further increase this mutation rate the addition of Mn^{2+} ions and/or changes in the Mg^{2+} ion concentration along with unbalancing the available nucleotide pool have all been shown effective.[4-6] Amplification of the gene under these conditions leads to a population of molecules carrying random point mutations. The degree of mutagenesis can be controlled by varying those conditions or adjusting the number of amplification cycles allowing an easily manipulable choice of mutation rate.

The rate limiting step is often cloning the DNA products generated from EP-PCR into a vector that will allow appropriate expression for screening. Screening a large population of mutants increases the likelihood of finding the one with the desired properties. However the ease of construction and the size of the library depend on the methods employed. Most common is the use of restriction digestion and ligation with T4 DNA ligase, either using internal sites from the gene or by the use of primers with internal restriction sites,[7] with subsequent ligation to an appropriately prepared vector. Alternately TA and TOPO cloning utilize a vector with 5'-T overhangs that compliment the 3'-A overhangs generated by the *Taq* polymerase nontemplate-dependent terminal transferase activity.[8]

Other non-ligase dependent approaches to build libraries involve in vitro site specific recombination. The Invitrogen Gateway system of recombinational cloning uses primers flanked by λ*attB* sites[9] incubated in vitro with a vector containing the target *attB* recombination site in the presence of bacteriophage λ integrase. Elledge et al.[10] have proposed a different but elegant in vivo method to transfer DNA fragments between vectors.

©2010 Landes Bioscience.
Do not distribute.

Transformation of *E. coli* with linear DNA is problematic due to its degradation by the RecBCD nuclease complex.[11] The use of *recBCD* mutants has been exploited for allelic replacement in *E. coli*[12,13] although these mutants are dramatically reduced in their overall rate of recombination. However the use of recombinational genes from bacteriophage λ has proven even more effective. Upon infection of *E. coli,* the phage expresses Gam (γ), a protein that specifically inhibits RecBCD and allows the survival of rolling circle replicative DNA.[14] Supplementing this with the Red α and β subunits a functional recombinase system is reconstituted that works effectively with linear DNA. This system has been exploited both for allelic replacement[15] and for in vivo cloning.[16] The latter uses linearized vector to co-transform *E. coli* along with DNA fragments sharing flanking homology; the fragments can be generated either by a restriction digest or amplified by PCR. By using *E. coli* cells that transiently express the Red recombinase and Gam gene products allow recombination to generate plasmids carrying the co-transformed insert.

Our work extends that method to the generation of highly complex libraries after EP-PCR. We have optimized in vivo recombinational cloning, and to further reduce background have developed a "universal" positive selection vector carrying the F plasmid toxin gene *ccdB* that is replaced after recombination, which can readily be used for any gene cloned in any of the myriad blue/white screening vectors in use.

*Correspondence to: Michael J. Benedik; Email: benedik@tamu.edu
Submitted: 05/20/10; Revised: 07/02/10; Accepted: 07/09/10
Previously published online: www.landesbioscience.com/journals/biobugs/article/12942
DOI: 10.4161/bbug.1.5.12942

Table 1. Effect of length of homology region on recombination rate

PCR primers	Vector (pmole)	PCR fragment (pmole)	30°C			37°C		
			Blue colonies	White colonies	Recombination rate	Blue colonies	White colonies	Recombination rate
	0.25	0				4×10^4	0	
	0	1				0	0	
-20M13	0.25	1	1.3×10^4	2.2×10^3	14 ± 3%	1.7×10^4	5.4×10^3	24 ± 2%
-60M13	0.25	1	8.8×10^3	1.6×10^4	64 ± 6%	1.6×10^4	3.4×10^4	68 ± 2%
-200M13	0.25	1	1.5×10^4	4×10^4	73 ± 4%	2×10^4	6.5×10^4	76 ± 4%

PCR fragments were generated using three primer pairs that confer different lengths of homology with the recipient plasmid. After transformation, cells were left to recover at 30°C or 37°C. The recombination rate reflects the number of white colonies divided by the total number of colonies. Values are the average of 3 experiments.

Table 2. Effect of vector treatment

Vector treatment	Blue colonies	White colonies	Recombination rate
Phosphorylated vector	3.3×10^4	2.7×10^4	45 ± 2%
Dephosphorylated vector	1.6×10^4	3.4×10^4	68 ± 2%
Uncut vector	$>1 \times 10^5$	0	0

0.25 pmole of linear phosphorylated and dephosphorylated plasmid or 0.125 pmole of double stranded linearized plasmid were co-transformed into MB4091 cells with PCR products generated using -60M13 primers. The recombination rate reflects the number of white colonies divided by the total number of colonies. Values are the average of 3 experiments.

Results

©2010 Landes Bioscience.
Do not distribute.

Preliminary experiments confirmed published results[15] that in vivo recombination using the λRed system was effective at creating recombinant clones. Vector DNA linearized by restriction digestion could be co-transformed with insert DNA containing terminal homology and recombinants generated with reasonable efficiency. Here we report experiments designed to investigate and optimize the parameters for this approach.

Effect of length of homology. One parameter likely to affect the efficiency with which recombinants are obtained is the length of the homology region. As a means of varying the homology region to test this, PCR products were generated using different primer pairs that yield different lengths of homologous regions. The donor and recipient plasmid both carry the *lacOP-lacZα* region commonly found in blue/white screening vectors allowing us to utilize "universal" sequencing primers of varying distance from the polylinker, but within the *lac* region. MB4091 cells were co-transformed with linearized pBC plasmid as the recipient and DNA fragments generated by PCR from pMB3504, a pBluescript II (KS^+) clone carrying the *cynD* gene of *B. pumilus*. Three different primer pairs were used to generate the PCR products: the -20M13 primer pair generates a fragment with 81 and 89 bp of homology, the -60M13 primer pair generates a fragment with 121 and 127 bp of homology, and the -200M13 primer pair generates a fragment with 194 and 269 bp of homology from each end of the 1 kb insert respectively.

The percent recombinants were determined from the numbers of blue (nonrecombinant) and white (recombinant) colonies. All white colonies tested were shown to be true recombinants both by plasmid digestion profiling of about 20 colonies, as well as by measuring the ability to express CynD activity from about 95 colonies.

As shown in **Table 1**, efficiency of recombination improved by nearly 3-fold when the length of homology between the plasmid and the PCR fragment increased from 81 bp to 121 bp. There was a further small increase with the -200 primer pairs. In both cases there was a concomitant increase both in the number of white colonies as well their percent of the total. The number of blue colonies, which represent non-recombinant plasmids, remained relatively similar. The presence or absence of the *red* genes post-transformation did not affect the recombination rate significantly as shown by similar values when transformants were recovered at the permissive (30°C) or non-permissive (37°C) temperatures for the pKD46 replication.

Vector preparation. To analyze what vector treatment was best for efficient recombination, the recombination rate was determined using undigested plasmid, digested plasmid and digested and dephosphorylated plasmid. The latter would be expected to decrease potential plasmid self-religation and possibly increase the opportunity for recombination with the PCR fragments.

Vector DNA prepared by each treatment was co-transformed into MB4091 with PCR fragments generated using -60M13 primer pairs. In a comparison of digested vector with digested and dephosphorylated vector the number of blue colonies decreased by about two-fold after dephosphorylation while the number of white colonies increased slightly, resulting in a higher recombination rate (**Table 2**). Dephosphorylation of the cloning vector decreased the background but had no significant affect on the recovery of recombinants. On the other hand, the use of uncut vector yielded a very high rate of transformation with non-recombinant plasmid as might be expected and recombinants were not detected, likely masked by the high background.

Effect of induction time. The expression of λRed recombinase from pKD46 is regulated by the arabinose inducible *pBAD* promoter. An analysis of optimal induction time by arabinose is shown in **Table 3**. The highest recombination rate and the highest yield of recombinants were observed when cells were induced for 30 to 60 minutes with 1% arabinose. The lack of recombinants without induction shows that Red expression was necessary.

Table 3. Effect of induction time

Induction time (minutes)	Blue colonies	White colonies	Recombination rate
0	2.4 x 10^4	1.8 x 10^2	0.7 ± 1%
30	1.3 x 10^4	6.1 x 10^4	82 ± 3%
60	1.1 x 10^4	5 x 10^4	82 ± 2%
90	2.6 x 10^4	4.8 x 10^4	65 ± 2%

Cells were induced with 1% arabinose for the times listed before making competent. The cells were harvested at similar OD_{600} of around 0.3 and were similarly competent. 1 pmole of PCR fragment amplified using the primer pairs -60M13 and 0.25 pmole of digested vector were used in all the transformations. Values are the average of 3 experiments.

Table 4. The ratio of fragment to vector

Ratio F/V	Blue colonies	White colonies	Recombination rate
1:1	2 x 10^4	1.6 x 10^4	44 ± 4%
2:1	1.8 x 10^4	2.7 x 10^4	60 ± 5%
3:1	1.5 x 10^4	2.1 x 10^4	58 ± 5%
4:1	1.4 x 10^4	3 x 10^4	68 ± 2%
6:1	1.5 x 10^4	3.7 x 10^4	71 ± 6%
8:1	1.2 x 10^4	2.4 x 10^4	66 ± 5%

The vector concentration was kept constant at 0.25 pmole of linearized vector co-transformed into MB4091 along with 0.25, 0.5, 0.75, 1.5 and 2 pmoles of PCR fragments to confer a ratio of 1:1, 2:1, 3:1, 4:1, 6:1, 8:1 respectively. Values are the average of 3 experiments.

Ratio of fragment to vector. In **Table 4** recombination rates are shown for a variety of different vector/insert ratios. The vector concentration was kept constant for these experiments with the quantity of insert as the variable. The recombination rate increased with increasing insert amount up to about a 4 to 1 ratio. Beyond that there was no significant further increase in recombination rate.

Construction of a positive selection vector. Although we routinely achieved recombination rates above 60% we wanted to see if we could improve that to nearly 100% with the use of a positive selection vector where a toxic gene would be replaced through recombination by the insert of choice, yet still maintain the ability to use the universal primer pairs.

©2010 Landes Bioscience.
Do not distribute.

The *ccdA* and *ccdB* (cell controlled death) genes comprise a plasmid toxin/antitoxin system carried by the F plasmid to ensure the maintenance of the plasmid in *E. coli* daughter cells.[17] The toxin CcdB (11.7 KDa) selectively targets DNA gyrase[18,19] and the antitoxin CcdA (8.7 KDa) both represses the promoter of *ccd* operon inhibiting the expression of the toxin CcdB as well as forming a tight complex with CcdB thereby inhibiting its activity. The half-life of CcdA is shorter than that of CcdB,[20] consequently daughter cells losing the F plasmid do not survive.

The *ccdB* gene was cloned into pBC SK$^+$ to generate plasmid pMB4105 and the plasmid maintained in the F' strain MB1547. Co-transformation of this vector with insert DNA into a strain lacking F, in this case we used the same MB4091 strain, eliminated the background of non-recombinant plasmids to less than 1% while maintaining comparable numbers of recombinant white clones.

Vector digestion was also shown crucial for recombination even when using pMB4105. If the plasmid was not digested, no recombinant plasmids were observed hence no colonies.

Discussion

In this paper we show that λRed recombinase system is an efficient tool to use for the generation of complex recombinant libraries in lieu of in vitro reactions, when appropriate inserts can be generated carrying terminal homology. Combined with error-prone PCR mutagenesis, this method rapidly generates large pools of randomly mutagenized clones used in a selection or screening of proteins with improved or novel properties.

The length of the flanking homology was found to be important in the efficiency of generating recombinants. Increasing the homology from about 80 to 120 bps generated a three-fold increase in the number of recombinants. Further increasing the length of homology yielded smaller increases in efficiency. The generation of recombinants was dependent on the presence of Red recombinase enzymes and required the use of linearized vector, recombinants were not detected when the recipient vector remained circular. Dephosphorylation of the digested vector did decrease the number of background colonies but the use of a positive selection vector virtually eliminated any background. This cloning technique is very simple, fast and 1 pmole of PCR fragments is sufficient to generate large libraries in less than one day.

We anticipate this approach could find use for routine subcloning although there are clearly restrictions on the source and target DNA's stemming from the requirement to have homology. Our use of *lac* region homology extends the potential utility to a wide variety of vector systems, we routinely move inserts between blue/white screening plasmids in this manner. The method however is not limited to *lac* homology, any stretch of terminal homology (of about 100 bp) should be suitable.

Of perhaps greater interest is in the generation of random mutational libraries of high complexity generated by error prone PCR. Such libraries routinely find applications in directed evolution and protein engineering. We have applied this approach to build mutant libraries for a number of enzymes which commonly yield mutant alleles carrying multiple mutations. While no method can sample the entire sequence space of a protein, the complexity is sufficient to isolate mutants requiring multiple changes. The remarkable ease by which this methods generates large libraries makes it our favorite approach for protein engineering applications.

Materials and Methods

DNA fragment preparation. The plasmid template used for PCR was pMB3504, derived from pBluescript II KS$^+$ (Stratagene) containing the $cynD_{pum}$ gene encoding cyanide dihydratase from *Bacillus pumilus*[21] as a 1 kb insert cloned with XbaI and XhoI.

PCR reactions were performed as 50 microliters reactions containing 25 microliters of Taq 2X Master Mix (New England

Table 5. List of primers used in this work

Primer	Sequence
-20M13F	5'-CGC CAG GGT TTT CCC AGT CAC GAC
-20M13R	5'-GAG CGG ATA ACA ATT TCA CAC AGG A
-60M13F	5'-GCG AAA GGG GGA TGT GCT GCA AGG
-60M13R	5'-CAC TTT ATG CTT CCG GCT CGT ATG
-200M13F	5'-CCA TTC GCC ATT CAG GCT GCG CAA C
-200M13R	5'-CCC AAT ACG CAA ACC GCC TCT CCC

Biolabs). The standard reactions contained plasmid DNA concentration as listed and 100 ng of each PCR primer with the following reaction conditions: 30 cycles of 95°C for 30 seconds, 55°C for 1 minute and 72°C for 1 minute. The PCR primers used were the M13 universal sequencing primers that can be used with most blue/white screening plasmids and are listed in **Table 5**. The PCR products were ethanol precipitated, resuspended in water and stored at -20°C. The DNA concentration was determined by measuring the A_{260} using a Nanodrop spectrophotometer.

Vector preparation. The recipient plasmid used in this work was the chloramphenicol resistant pBC SK$^+$ from Stratagene. The DNA was digested using EcoRI and BamHI enzymes at 37°C for 2 hours. Dephosphorylation as indicated was done using Antarctic Shrimp Phosphatase followed by heat inactivation and ethanol precipitation. The DNA was resuspended in water and stored at -20°C and the DNA concentration was measured at A_{260}.

Positive selection vector. The *ccdB* gene was amplified from F-carrying *E. coli* DNA using ccdB-F (5'-CAG ACT GCA GGA AGG GAT GGC TGA) and ccdB-R (5'-CAC TGC CGG TAC CAT GAC TGC AGA) which introduced PstI sites at each end and thereby cloned into pBC SK$^+$. The resultant plasmid pMB4105 was transformed into the F' strain MB1547 [*supE thi-1* Δ *endA* Δ*(lac-proAB)* Δ*(mcrB-hsdSM)]5*/F' [*traD36 proAB$^+$ lacI^q lacZ*Δ*M15*]. For use in in vivo cloning, the vector was digested with EcoRI and BamHI, ethanol precipitated, resuspended in water and stored at -20°C.

Transformation. The recipient strain MB4091 [DH10B (pKD46)] was used for in vivo recombination. The pKD46 plasmid[22] carries the λ *red* and *gam* genes expressed from the arabinose inducible *pBAD* promoter, confers resistance to ampicillin and is temperature sensitive for replication to allow plasmid loss after transformation.[22] The cells were grown in LB broth supplemented with 100 mg/L ampicillin at 30°C. Electroporation competent cells were made from cells grown in LB broth at 30°C to an O.D. (600 nm) = 0.3 with 0.1% arabinose added for specified times.

Electroporation was carried out using a Bio-Rad Micropulser and cells were allowed to recover in a 1 ml volume on a shaker at 37°C for 1 hour unless otherwise stated. Aliquots of 0.1 ml were spread on plates with chloramphenicol (25 mg/L) and Xgal (40 mg/L) and incubated at 37°C. The number of transformants per ml are reported and the recombination rate was calculated by dividing the number of white colonies by the number of total colonies on the plate. Experiments were all performed in triplicate.

©2010 Landes Bioscience. Do not distribute.

Acknowledgements

The financial support of The Welch Foundation (A1310) and the Texas Hazardous Substance Research Center are gratefully acknowledged.

References

1. Skandalis A, Encell LP, Loeb LA. Creating novel enzymes by applied molecular evolution. Chem Biol 1997; 4:889-98.
2. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol 2009; 10:866-76.
3. Tindall KR, Kunkel TA. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 1988; 27:6008-13.
4. Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. Genome Res 1992; 2:28-33.
5. Beckman RA, Mildvan AS, Loeb LA. On the fidelity of DNA replication: manganese mutagenesis in vitro. Biochemistry 1985; 24:5810-7.
6. El-Deiry WS, Downey KM, So AG. Molecular mechanisms of manganese mutagenesis. Proc Natl Acad Science USA 1984; 81:7378-82.
7. Scharf SJ, Horn GT, Erlich HA. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 1986; 233:1076-8.
8. Clark JM. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res 1988; 16:9677-86.
9. Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. Gen Research 2000; 10:1788-95.
10. Li MZ, Elledge SJ. MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. Nat Genet 2005; 37:311-9.
11. Smith GR. Mechanism and control of homologous recombination in *Escherichia coli*. Annu Rev Genet 1987; 21:179-201.
12. Winans SC, Elledge SJ, Krueger JH, Walker GC. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J Bacteriol 1985; 161:1219-21.
13. Russell CB, Thaler DS, Dahlquist FW. Chromosomal transformation of *Escherichia coli recD* strains with linearized plasmids. J Bacteriol 1989; 171:2609-13.
14. Unger RC, Clark AJ. Interaction of the recombination pathways of bacteriophage lambda and its host *Escherichia coli* K12: Effects on exonuclease V activity. J Mol Biol 1972; 70:539-48.
15. Murphy KC. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J Bacteriol 1998; 180:2063-71.
16. Oliner JD, Kinzler KW, Vogelstein B. In vivo cloning of PCR products in *E. coli*. Nucl Acids Res 1993; 21:5192-7.
17. Ogura T, Hiraga S. Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc Natl Acad Sci USA 1983; 80:4784-8.
18. Van Melderen L. Molecular interactions of the CcdB poison with its bacterial target, the DNA gyrase. Int J Med Microbiol 2002; 291:537-44.
19. Dao-Thi M-H, Van Melderen L, De Genst E, Afif H, Buts L, Wyns L, et al. Molecular basis of gyrase poisoning by the addiction toxin CcdB. J Mol Biol 2005; 348:1091-102.
20. Van Melderen L, Bernard P, Couturier M. Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. Mol Microbiol 1994; 11:1151-7.
21. Jandhyala D, Berman M, Meyers PR, Sewell BT, Willson RC, Benedik MJ. CynD, the Cyanide Dihydratase from *Bacillus pumilus*: Gene cloning and structural studies. Appl Environ Microbiol 2003; 69:4794-805.
22. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli K-12* using PCR products. Proc Natl Acad Sci USA 2000; 97:6640-5.