Mutagenic PCR.

R C Cadwell and G F Joyce

Genome Res. 1994 3: S136-S140

Mutagenic PCR

R. Craig Cadwell[1] and Gerald F. Joyce

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Most practitioners of PCR prefer to carry out DNA amplification in an accurate manner, introducing as few base substitutions as possible. This is especially critical when one is studying clonal isolates and must distinguish natural variation from artifactual variation that is introduced by polymerase error. Fortunately, thermostable DNA polymerases are available that operate with high fidelity because of an intrinsic 3′ → 5′ exonuclease activity (for review, see ref. 1). Manipulation of PCR conditions can lead to further improvement of copying accuracy.

Here, we consider the other side of the fidelity issue—those instances where promiscuity is a virtue. Oftentimes, in probing the structure or function of a protein or nucleic acid, one wishes to generate a library of mutants and apply a screening method to isolate individuals that exhibit a particular property. For mutations over a short stretch of nucleotides within a cloned gene, it is appropriate to replace a portion of the gene with a synthetic DNA fragment that contains random or partially randomized nucleotides.[2–5] For mutations over a longer segment, up to the size of an entire gene, it may be preferable to scatter random mutations over the entire sequence, typically at a frequency of one or a few mutations per molecule. In such cases, it is most convenient to introduce random mutations through inaccurate copying by a DNA polymerase, especially if the polymerase is a thermostable enzyme that can operate in the context of the PCR. Each pass of the polymerase during the PCR allows for the possibility of mutation, so that the cumulative error rate can become substantial.

The error rate of *Taq* polymerase is the highest of the known thermostable DNA polymerases, in the range of 0.1×10^{-4} to 2×10^{-4} per nucleotide per pass of the polymerase, depending on reaction conditions.[6–9] Over the course of the PCR, in which the polymerase makes an average of 20–25 passes, the cumulative error rate is $\sim 10^{-3}$ per nucleotide. In most cases this is insufficient to generate a diverse library of variant sequences, especially over a region shorter than 1000 nucleotides. A further drawback is that the errors made by *Taq* polymerase under standard PCR conditions are heavily biased toward A·T → G·C changes.[6] We have devised a mutagenic PCR that has an overall error rate of $\sim 7 \times 10^{-3}$ per nucleotide and does not exhibit substantial sequence bias.[10]

MUTAGENESIS PROCEDURE

The top priority of mutagenic PCR is to introduce the various types of mutations in an unbiased fashion rather than to achieve a high overall level of amplification. The DNA input in a 100-μl reaction mixture consists of 10^{10} molecules (20 fmoles), which are amplified ~1000-fold to yield 10^{13} molecules (20 pmoles). This modest amplification requires an average of 10 passes of the polymerase. However, 30 cycles of the PCR are carried out to ensure that mismatched termini have ample opportunity to become extended to produce complete copies. The large input prevents the PCR products from being influenced by the effects of clonal expansion. Even if a mutation occurs in the first pass of the polymerase and is passed along to all of the descendent molecules, there is very little chance that any two molecules isolated from the final population will carry the same mutation as a consequence of their being derived from a common ancestor.

Protocol

The protocol for mutagenic PCR is derived from "standard" PCR condi-

[1]Present address: Department of Biological Sciences, University of California, Santa Barbara, California 93106

3:S136–S140©1994 by Cold Spring Harbor Laboratory ISSN 1054-9805/94 $1.00

tions:[11] 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris (pH 8.3 at 25°C), 0.2 mM each dNTP, 0.3 μM each primer, and 2.5 units of *Taq* polymerase in a 100-μl volume, incubated for 30 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min in a conventional thermal cycler. The following changes are made to enhance the mutation rate:

1. The $MgCl_2$ concentration is increased to 7 mM to stabilize noncomplementary pairs.[8,9]
2. 0.5 mM $MnCl_2$ is added to diminish the template specificity of the polymerase.[12,13]
3. The concentration of dCTP and TTP is increased to 1 mM to promote misincorporation.[10,13]
4. The amount of *Taq* polymerase is increased to 5 units to promote chain extension beyond positions of base mismatch.[14]

The experimental protocol is as follows:

1. Prepare a 10× mutagenic PCR buffer containing 70 mM $MgCl_2$, 500 mM KCl, 100 mM Tris (pH 8.3 at 25°C), and 0.1% (wt/vol) gelatin.
2. Prepare a 10× dNTP mix containing 2 mM dGTP, 2 mM dATP, 10 mM dCTP, and 10 mM TTP.
3. Prepare a solution of 5 mM $MnCl_2$. DO NOT combine with the 10× PCR buffer, which would result in formation of a precipitate that disrupts PCR amplification.
4. Combine 10 μl of 10× mutagenic PCR buffer, 10 μl of 10× dNTP mix, 30 pmoles of each primer, 20 fmoles of input DNA, and an amount of H_2O that brings the total volume to 88 μl. Mix well.
5. Add 10 μl of 5 mM $MnCl_2$. Mix well and confirm that a precipitate has not formed.
6. Add 5 units (2 μl) of *Taq* polymerase (Cetus or licensed supplier), bringing the final volume to 100 μl. Mix gently. Cover with mineral oil or a wax bead, if desired.
7. Incubate for 30 cycles of 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min. Do not employ a "hot start" procedure or a prolonged extension time at the end of the last cycle.
8. Purify the reaction products by extraction with chloroform/isoamyl alcohol [24:1 (vol/vol)] and subsequent ethanol precipitation.
9. Run a small portion of the purified products on an agarose gel stained with ethidium bromide to confirm a satisfactory yield of full-length material. Mutagenic PCR should be carried out in parallel with standard PCR (omitting the four changes listed above); the yields of full-length DNA should be comparable.

Results

By employing a DNA of ordinary nucleotide composition, the mutagenic PCR introduces errors at a frequency of 0.66%±0.13% per position over the course of the PCR [95% confidence interval (C.I.)].[10] Nearly all of these changes are base substitutions. The combined frequency of insertions and deletions is <0.05% (one-tailed test, 95% C.I.). The number of mutations per DNA copy follows a Poisson distribution. The probability of mutating each of the 4 bases is approximately equal except for a 1.5-fold enhanced probability of mutating T residues, which is significant at the 99% confidence level. The most common specific mutations are A → T and T → A changes, which we attribute to T·T mismatches that manifest as either A → T changes in the same strand or T → A changes in the opposing strand. The least common mutations are G → C and C → G changes, which presumably reflects the difficulty in forming and extending G·G and C·C mismatches. Summing up all types of muta-

tions and correcting for the base composition of the mutated gene, the ratio of A·T → G·C to G·C → A·T changes is 1.0 (0.6–1.7, 95% C.I.).

Troubleshooting

The most common difficulty with the mutagenic PCR stems from the fact that 30 temperature cycles are employed even though *Taq* polymerase makes an average of only 10 passes along the DNA. As noted above, this provides ample opportunity for extension of mismatched termini, which is necessary to lock in mutations. However, it also favors the occurrence of amplification artifacts.[15] Compounding the problem is the markedly elevated $MgCl_2$ concentration, which lowers the stringency of primer hybridization, thereby promoting the formation of nonspecific amplification products. As a general rule, one should begin the mutagenic PCR with either cDNA or a double-stranded DNA fragment that encompasses only the region of interest. It is risky to employ plasmid DNA and hopeless to begin with a genomic library. We limit the use of mutagenic PCR to DNAs no longer than ~1000 nucleotides. For longer target sequences, the DNA can be divided into two or more fragments that are mutagenized separately.

Because the mutagenic PCR enhances primer mishybridization, there will be certain combinations of primers and target sequence that inevitably give rise to short amplification products that outcompete the full-length DNA. These artifacts are best seen by carrying out the reaction with a radiolabeled primer and separating the products on a nondenaturing polyacrylamide gel. On the basis of their size and (if necessary) sequence, it should be possible to discern the site of primer mishybridization and redesign the primers accordingly. Alternatively, it may be preferable to attach well-chosen primer-binding sites to the ends of the DNA and carry out PCR amplification using these handles.

Another source of difficulty is the tendency to make slight modifications of the protocol without evaluating their consequences. If, for example, C → G changes are less frequent than T → A changes, then why not double the concentration of dGTP to 0.4 mM? Doing so, it turns out, results in a fourfold increase in the ratio of A·T → G·C to G·C → A·T changes.[10] We encourage others to explore alternative reaction conditions that may lead to an improved PCR mutagenesis procedure. However, in view of the extreme sensitivity of *Taq* polymerase to dNTP concentrations and other aspects of the reaction conditions, general users are encouraged to follow the protocol to the letter.

DISCUSSION

The distribution of variants that results from mutagenic PCR depends on the error rate and the length of the sequence that is being randomized. The probability P of having k mutations in a sequence of length n is given by $P(k,n,\epsilon) = (n! / [(n-k)!\ k!])\ \epsilon^k\ (1-\epsilon)^{n-k}$, where ϵ is the error rate per position. In the present case, the error rate is 0.66% per position over the course of the PCR ($\epsilon = 0.0066$). Thus, for a target sequence of 500 nucleotides, the resulting population of variants would consist of ~4% wild-type, 12% one-error mutants, 20% two-error mutants, 22% three-error mutants, 18% four-error mutants, 12% five-error mutants, and 12% mutants with six or more errors. The number of distinct sequences with k errors, N_k, increases exponentially with increasing k: $N_k = (n! / [(n-k)!\ k!])\ 3^k$. Thus, 20 pmoles of material resulting from the mutagenesis of a 500-nucleotide target sequence would contain all possible one-, two-, three-, and four-error mutants but only ~2% of the possible five-error mutants and a progressively sparser sampling of the ever higher-error mutants. These calculations refer to the composition of the DNA;

for the corresponding protein, they must be modified to take into account the degeneracy of the genetic code.

For some purposes, an error rate of 0.66% per position will be insufficient. It is possible to carry out successive rounds of mutagenic PCR to double or even triple the overall error rate. However, two potential pitfalls must be avoided. First, if a small aliquot of one reaction mixture is used to seed the next, there is an increased chance that molecules isolated from the final pool will be related by descent. Taking one-thousandth of the products from a first mutagenic PCR to seed a second should not be a problem, but taking one-thousandth of the second to seed a third would reduce diversity to an unacceptably low level. This problem could be remedied by scaling up the third reaction mixture to 10-ml volume, preferably in multiple reaction vessels containing 100 μl each. A second potential pitfall is the risk of generating nonspecific amplification products, made more likely by the increased number of temperature cycles. It may be necessary to gel-purify full-length DNA after the first mutagenic PCR before proceeding with the second.

Until a more error-prone thermostable DNA polymerase is found in nature or developed through enzyme engineering, *Taq* polymerase provides the most effective way to generate a library of DNAs that contain random mutations over a stretch of 100–1000 nucleotides. If one is interested in a library of RNAs, then a promoter sequence for T7 RNA polymerase can be included near the 5′ end of the appropriate PCR primer, allowing the DNA products to serve as templates in an in vitro transcription reaction.[16] If one is interested in a library of proteins, then the PCR primers can be designed to include either restriction sites for cloning into a suitable expression vector or a ribosome-binding site and start codon for in vitro translation.

An important advantage of mutagenic PCR is that it allows repeated randomization of a population of nucleic acids without isolating clones and obtaining sequence information. After one has generated a library of mutants and applied a screening method to obtain individuals that exhibit a particular property, the selected individuals can then be used directly as input for a second mutagenic PCR. Repeating the cycle of selection and mutagenic amplification allows one to carry out in vitro evolution of nucleic acids, including those that have catalytic function.[17–19] Similarly, a population of protein-encoding DNAs, harvested from a selected subset of cells or viral particles, can be treated as an ensemble and subjected to mutagenic PCR to produce variants of the selected variants.

REFERENCES

1. Cha, R.S. and W.G. Thilly. 1993. Specificity, efficiency, and fidelity of PCR. *PCR Methods Applic.* **3:** S18–S29.
2. Matteucci, M.D. and H.L. Heyneker. 1983. Targeted random mutagenesis: The use of ambiguously synthesized oligonucleotides to mutagenize sequences immediately 5′ of an ATG initiation codon. *Nucleic Acids Res.* **11:** 3113–3121.
3. Wells, J.A., M. Vasser, and D.B. Powers. 1985. Cassette mutagenesis: An efficient method for generation of multiple mutations at defined sites. *Gene* **34:** 315–323.
4. Oliphant, A.R., A.L. Nussbaum, and K. Struhl. 1986. Cloning of random-sequence oligodeoxynucleotides. *Gene* **44:** 177–183.
5. Ner, S.S., D.B. Goodin, and M. Smith. 1988. A simple and efficient procedure for generating random point mutations and for codon replacements using mixed oligonucleotides. *DNA* **7:** 127–134.
6. Keohavong, P. and W.G. Thilly. 1989. Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci.* **86:** 9253–9257.
7. Eckert, K.A. and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18:** 3739–3744.
8. Eckert, K.A. and T.A. Kunkel. 1991. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Applic.* **1:** 17–24.
9. Ling, L.L., P. Keohavong, C. Dias, and W.G. Thilly. 1991. Optimization of the polymerase

chain reaction with regard to fidelity: Modified T7, *Taq*, and *Vent* polymerases. *PCR Methods Applic.* **1:** 63–69.
10. Cadwell, C. and G.F. Joyce. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Applic.* **2:** 28–33.
11. Coen, D.M. 1991. The polymerase chain reaction. In *Current protocols in molecular biology*. Wiley Interscience, New York.
12. Beckman, R.A., A.S. Mildvan, and L.A. Loeb. 1985. On the fidelity of DNA replication: Manganese mutagenesis *in vitro*. *Biochemistry* **24:** 5810–5817.
13. Leung, D.W., E. Chen, and D.V. Goeddel. 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1:** 11–15.
14. Gelfand, D.H. and T.J. White. 1990. Thermostable DNA polymerases. In *PCR protocols: A guide to methods and applications* (ed. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White), pp. 129–141. Academic Press, San Diego, CA.
15. Mullis, K.B. 1991. The polymerase chain reaction in an anemic mode: How to avoid cold oligodeoxyribonuclear fusion. *PCR Methods Applic.* **1:** 1–4.
16. Chamberlin, M. and T. Ryan. 1982. Bacteriophage DNA-dependent RNA polymerases. In *The enzymes* (ed. P.D. Boyer), vol. 15, pp. 85–108. Academic Press, New York.
17. Beaudry, A.A. and G.F. Joyce. 1992. Directed evolution of an RNA enzyme. *Science* **257:** 635–641.
18. Lehman, N. and G.F. Joyce. 1993. Evolution *in vitro* of an RNA enzyme with altered metal dependence. *Nature* **361:** 182–185.
19. Bartel D.P. and J.W. Szostak. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science* **261:** 1411–1417.