A new PCR based method for the generation of nested deletions

Jeannette M. Whitcomb, Ayoub Rashtchian¹ and Stephen H. Hughes*

ABL – Basic Research Program, NCI – Frederick Cancer Research and Development Center, PO Box B, Frederick, MD 21702-1201 and ¹Life Technologies Inc., BRL, Box 6009, Gaithersburg, MD 20884-9980, USA

Received February 22, 1993; Revised and Accepted June 28, 1993

ABSTRACT

We have developed a simple, PCR-based protocol, random primed/anchored-PCR (RPA-PCR), that allows the selective amplification and efficient cloning of segments that are adjacent to any known sequence. We demonstrate that RPA-PCR can be used to prepare a nested set of evenly spaced deletions suitable for DNA sequencing. However, it should also be possible to use this technique for a number of other purposes: generating deletions for the analysis of eukaryotic promoters, extending cDNA clones in the 5' direction, cloning the insertion sites of retroviral proviruses and transposons, and analyzing intron/exon boundaries.

INTRODUCTION

The construction of subclones that represent a nested set of deletions is often a key step in the sequence analysis of cloned DNA. We have developed a novel method to generate such subclones from regions for which sequence information is unavailable. The strategy, which we call random-primed/anchored-PCR (RPA-PCR) consists of three steps: 1) random-primed DNA synthesis (1), 2) PCR amplification (2), and 3) uracil DNA glycosylase (UDG)-cloning (3,4). We have used this method to generate a set of nested deletion clones from a cloned DNA insert. One end of each subclone generated by RPA-PCR is determined by an anchored PCR primer and, as a consequence, this end is shared by all subclones. The other end of each subclone is unique since it is determined by the sequences from which random-primed DNA synthesis commenced. The random primers that we use also contain a specified sequence at their 5' end that can serve as a primer recognition sequence in a subsequent PCR amplification. The other commonly used methods for generating nested deletion clones are dependant on enzymatic digestion (by *Exo III* or *Bal 31*) of the cloned insert (5,6). These methods have a number of disadvantages. Specifically, they require that *CsCl*-banded DNA is used as the starting material and that unique restriction sites are present in the vector and absent from the insert. In addition, a relatively large amount of DNA is required for these analyses; therefore, the application of these methods

to large DNA segments cloned into cosmids or yeast artificial chromosomes (YACs) is difficult. By contrast, RPA-PCR can be used on either crude or purified preparations of DNA; requires no knowledge of the restriction enzyme map of the insert; and is quick, simple, and efficient.

MATERIALS AND METHODS

Random primed/anchored-PCR (RPA-PCR)

1) Plasmid DNA was random primed as follows: 2 μ l of DNA (~ 50 ng) was added to 5 μ l of random primer (1 OD/ml), boiled for 10 minutes, and placed on ice. 5.5 μ l of a reaction mixture containing 2.5 μ l 5x Sequenase buffer (final concentration: 40 mM Tris-HCl, pH 7.5, 20 mM MgCl₂, 50 mM NaCl), 1.25 μ l of 0.5 mM dNTPs, 1.25 μ l of 0.1 M DTT, and 0.5 μ l of Sequenase Version 2.0 T7 DNA polymerase (diluted 1:4) was then added. This reaction was incubated for 30 minutes at 37°C.

2) Sephacryl S-400 spin column: Unextended random primers were removed by passing the reaction over a Sephacryl S-400 spin column. Spin columns were prepared in 1.5 ml centrifuge tubes. A small hole was melted in the bottom of a 1.5 ml centrifuge tube with a hot 25-gauge needle. Approximately 50 μ l (packed volume) of acid-washed glass beads was added to the tube followed by 1.5 ml of Sephacryl S-400 slurry (Sigma Chemical Co). The Sephacryl was allowed to settle briefly by gravity and the column was packed by centrifugation at 1400 $\times$ g for 3 minutes. The random-primed DNA synthesis reaction was loaded onto the column and the column was spun for 3 minutes at 1400 $\times$ g. The material that passed through the column was collected and used in the PCR.

3) PCR amplification: The single-stranded, random-primed DNAs were used as template in the following reaction: 5 μ l of 10 $\times$ PCR buffer (final concentration: 50 mM Tris-Cl, pH 8.3, 50 mM KCl, 7 mM MgCl₂), 5 μ l of 1.7 mg/ml BSA, 2 μ l of 25 mM each dNTPS, 2 μ l of anchored primer with tail 'B' (5 OD/ml), 2 μ l of 'A'-tail primer (5 OD/ml), 12.5 μ l of random primed DNA reaction, 21 μ l of water and 0.5 μ l of Taq DNA polymerase. Twenty cycles of PCR were performed as follows: 94°C, 1 minute; 55°C, 1 minute; 72°C, 3 minutes.

* To whom correspondence should be addressed

The PCR products were size selected on a 1% low-melting-point agarose-Tris acetate-EDTA gel. Gel slices from the desired size range were cut out and the DNA was recovered using GeneClean (Bio101).

Cloning PCR products using the CloneAmp System (Gibco BRL). 5 μ l of the melted gel slice or 5 μ l of DNA extracted from the gel slice was added to 2 μ l of pAMP1 vector DNA, 1 μ l of uracil DNA glycosylase, 2 μ l of 10x PCR buffer and 10 μ l of water. pAMP1 and UDG are components of the CloneAmp kit. The reaction was incubated at 37°C for 30 minutes and 5 μ l was used to transform 50 μ l of competent *E. coli* DH5 α cells (GibcoBRL).

Screening colonies by hybridization. Bacterial colonies were transferred from agar plates onto nitrocellulose filters and lysed *in situ*. The DNA was fixed to the filters by UV-crosslinking in a Strata-linker (Stratagene) using the auto-crosslink setting. The filters were prehybridized for 1 hr and hybridized with an insert-specific 5' end-labeled oligonucleotide probe for either 3 hrs or overnight. The probe was labeled with polynucleotide kinase and [γ -³²P] dATP (Amersham). The filters were then washed twice for 30 minutes in 6 $\times$ SSC, 0.1% SDS at room temperature. If necessary, an additional wash was done for 30 minutes in 6 $\times$ SSC, 0.1% SDS at 37°C. The filters were wrapped while still damp with clear plastic wrap then exposed to film either for 3 hrs or overnight at -70°C with an intensifying screen. Positive colonies were transferred from the original plate to broth culture. Plasmid DNA was prepared using a standard alkaline lysis mini-prep procedure (5).

RESULTS AND DISCUSSION

The RPA-PCR strategy is outlined in Figure 1. The first step in this strategy is a random-primed, DNA synthesis reaction. The random primer (5' CUA CUA CUA CUA NNN NNN NNN N 3') consists of 10 'N' residues (this portion of the primer is synthesized using equal amounts of the four nucleotides at each cycle) and a defined 5' tail ('A' in Figure 1) that contains deoxyuridine residues.

To determine which DNA polymerase to use in the random priming step, we tested a variety of polymerases including *E. coli* DNA polymerase I (New England Biolabs), Klenow fragment of *E. coli* DNA polymerase I (New England Biolabs), Klenow fragment (exo-) (USB), T7 DNA polymerase (USB), Sequenase version 2.0 modified T7 DNA polymerase (USB), Taq DNA polymerase (Promega), Pfu DNA polymerase (Stratagene), Vent DNA polymerase and Vent (exo-) DNA polymerase (New England Biolabs). We had expected that enzymes with strand displacement activity would yield longer products than those without; however, this is not what we observed. We have determined empirically that the length and quantity of products following PCR amplification are increased when T7 DNA polymerase and Sequenase modified T7 DNA polymerase are used in the random-priming reaction compared to the other enzymes tested (data not shown). The random primed DNA synthesis generates single stranded DNAs (ssDNAs) that are complementary to both strands of the original DNA substrate. Because of the defined 5' tail of the primer, these ssDNAs share a common 5' sequence which can serve as a primer sequence for the subsequent PCR. Prior to the PCR, the excess random primers and short ssDNAs are removed from the reaction mixture using a Sephacryl S-400 spin column (see Materials and

Figure 1. Schematic diagram of the RPA-PCR procedure for generating nested deletion clones sets. Details are given in the text.

Methods). When a Sephacryl S-200 spin column was used, instead of an S-400 spin column, there was an increase in short PCR products and a concomitant decrease in the longer PCR products seen on an ethidium bromide-stained agarose gel. This suggests that PCR amplification of smaller fragments is favored relative to the amplification of longer products.

The longer ssDNAs, that are not retained by the Sephacryl S-400 spin column, are used as the template for the PCR. Two new primers are used in this PCR. The sense primer (5' CUA CUA CUA CUA 3') corresponds to the defined 5' tail of the random primer ('A' in Figure 1) and the antisense primer (5' CAU CAU CAU CAU GTA AAA CGA CGG CCA GTG 3') is complementary to a defined anchor sequence that fixes the constant end of the subclones. In these experiments, we designed the 3' end of the antisense primer to be complementary to the universal primer sequence (UPS, Figure 1) which is present flanking the polylinker region in a number of commercially available vectors, including M13, pUC, and

Figure 2. Schematic diagram of the RSV genome and the *EcoRI* fragment that was subcloned into the vector pBLUESCRIPT II KS. The individual flags designate the start sites of the various nested subclones obtained by the RPA-PCR using the universal primer sequence (UPS) as the anchored primer.

pBLUESCRIPT based vectors. The antisense primer also contains deoxyuridine at its 5' end (shown as 'B' in Figure 1).

The PCR products are cloned using the CloneAmp System (GibcoBRL). The inclusion of deoxyuridine in the PCR primers provides a means for cloning the PCR product without an intervening ligation step. The uracil bases are specifically excised from DNA by the enzyme uracil DNA glycosylase. The removal of the uracil bases destabilizes base pairing which produces, on the complementary strand, single-stranded regions that are long enough to allow cloning of the PCR products without ligation into a vector (pAMP1) which has complementary single-stranded ends (see Figure 1). The deoxyuridine-containing tail sequences 'A' and 'B' were chosen to be neither identical nor complementary which results in the efficient unidirectional cloning of the PCR products. The UDG cloning step has several advantages: it is unidirectional, ligation independent, simple, and highly efficient (4). Only those PCR products containing one 'A'-tail and one 'B'-tail can be cloned in this step providing a stringent selection for cloning only the PCR products with one random end and one defined end.

To test the efficiency of this technique, we used RPA-PCR to generate a series of nested deletion clones from a cloned *EcoRI* fragment of Rous sarcoma virus (RSV) subcloned into the vector pBLUESCRIPT II KS (Figure 2). The endpoints of the nested subclones generated by RPA-PCR are shown by the individual flags in Figure 2. The random-primed starts of the subclones are evenly spaced over a 2-kb segment of the insert. The length of individual inserts can be estimated by restriction enzyme digestion or by PCR using primers that flank the insert. Figure 3 shows a restriction enzyme digestion of some of the clones. The clones contain inserts that range in size from slightly more than 2 kb to about 500 bp. The nested deletion clones were sequenced using a primer complementary to the pAMP1 vector adjacent to the variable end of the nested subclones. In the example shown, an SP-6 sequencing primer was used (see Figure 1).

Figure 3. Restriction enzyme digestion of nested deletion subclones generated by the RPA-PCR. Miniprep DNAs from recombinant plasmids were digested with both *HindIII* and *Asp* 718 to excise the inserts. Fragments were separated on a 1% Tris Phosphate EDTA-agarose gel containing 1 μ g/ml ethidium bromide. The inserts greater than 1.3 kb contain an internal *HindIII* site and thus generate two insert-specific fragments.

With RPA-PCR, a complex mixture of DNA fragments which have identical ends but are of various lengths are amplified. Not surprisingly, we have found that the smaller fragments are amplified preferentially. We have used low-stringency PCR conditions and long cycle times to help correct this size bias; however, these conditions can lead to other PCR artifacts. In the experiments presented we have found two major classes of 'incorrect' clones. The first is composed of primer concatemers. We believe that these clones are generated through interaction of the repetitive sequences in the primer tail regions. It should be possible to avoid this particular artifact by using a non-repetitive sequence in the primer (7,8). The second class of 'incorrect' clones are caused by mispriming by the anchored primer. In our analysis of the clones generated by RPA-PCR from a known target sequence, mispriming occurred at regions that share homology with the 3' end of the anchored primer. Both of these 'incorrect' clones can be eliminated from the final population of subclones by including a hybridization selection step prior to further analysis. After cloning, membrane bound colony lysates can be probed with a labeled oligonucleotide complementary to sequences adjacent to the anchored primer (shown in Figure 1 by *). Only those clones that are recognized by the hybridization of the complementary oligo will be chosen for further analysis. Alternatively, the products generated by mispriming of the anchored primer can be eliminated by use of a nested primer designed to anneal 3' of the anchored primer. This approach has been used successfully to increase the specificity of the PCR in several cases (9,10). Inclusion of these additional screening steps may be especially important when more complex DNAs are used as the starting material for the RPA-PCR method.

Taq polymerase can introduce errors during PCR. However, a major advantage of the RPA-PCR is that the nested subclones are, by definition, generated by independent random priming and PCR amplification events. It is unlikely that an identical PCR-introduced error would be found in two independent subclones. Therefore, simply reading the same sequence from overlapping nested subclones can be used to confirm the sequence. In addition, other thermostable polymerases have been described that are less error prone than Taq polymerase which will help to circumvent this problem.

Finally, we would like to emphasize that RPA-PCR is by no means limited to the application we have described here. The anchored primer can be complementary to any known sequence and the random primers serve as a handle for any adjacent region, whether or not the sequence of this adjacent region is known.

The combined power of PCR and UDG-cloning should make it possible to selectively amplify and directly clone sequences not only from cloned DNAs in plasmid vectors, but also directly from lambda vectors, yeast artificial chromosomes, and possibly from more complex genomes. We foresee a number of additional applications including, but not limited to, deletion analysis of regulatory regions of eukaryotic promoters, cloning the integration sites of transposons, obtaining cDNAs for the 5' ends of mRNAs, determining 3' and 5' flanking sequences, obtaining genomic sequences using cDNA sequence information, and chromosome walking/running. There are other PCR-based techniques that can be used for some of these applications. For example, RACE (12), one-sided PCR (13), ligation-mediated PCR (14) or anchored PCR (15) have been designed primarily to amplify cDNA ends. These methods require the ligation of a primer recognition site onto the end of the nucleic acid to be amplified or homopolymeric tailing of the nucleic acid. In these procedures, if the end of the nucleic acid is far from the known sequence, the PCR amplification may be so inefficient that the method cannot be used successfully. The RPA-PCR can complement these procedures and serve as an intermediate step for obtaining sequence information from regions flanking known sequences. Since the method is based on random priming even if the end of the DNA is far away, the RPA-PCR can be used to reach the end of the DNA in a step-wise manner. Another method used to clone unknown segments adjacent to a known sequence is inverse PCR (16). This method depends on the presence of appropriate restriction enzyme recognition sites in the unknown region, requires a ligation step and demands that the ligated circles are small enough for efficient PCR amplification. Even when these methods do work, they do not yield a nested set of clones which can be used for rapid and efficient sequencing of the unknown sequence. Shen and Wayne (11) have devised a method for generation of unidirectional deletions in DNA fragments cloned into M13. Although this method is efficient it can only be used for DNA fragments cloned into a specific M13 derivative and can only be used with specific host cells. The RPA-PCR method is capable of generating nested deletions in virtually any nucleic acid regardless of the cloning vector. For these reasons we believe that RPA-PCR provides an attractive alternative to the available methods.

ACKNOWLEDGEMENTS

We are grateful to Marilyn Powers for the oligonucleotides, Gil Sharon for helpful discussions, Marjorie Strobel and Anne Arthur for critically editing the manuscript, and Hilda Marusiadis for her excellent secretarial assistance. Research sponsored in part by the National Cancer Institute, DHHS, under contract No. NO1-CO-74101 with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health & Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

REFERENCES

1. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
2. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi R., Horn G.T., Mullis K.B. and Ehrlich, H.A. (1988) *Science*, **239**, 487–491.
3. Nisson, P.C., Rashtchian, A. and Watkins, P.C. (1991) *PCR Meths. and Appl.*, **1**, 120–123.
4. Rashtchian, A., Buchman, G.W., Schuster, D.M. and Berninger, M.S. (1992) *Anal. Biochem.*, **206**, 91–97.
5. Maniatis, T., Fritsch, E.F. and Sambrook, J. (eds.) (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor, New York.
6. Slatko, B., Heinrich, P., Nixon, B.T. and Voytas, D. (1991) In Ausubel F.M., Brent, R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A. and Struhl K. (eds.) *Current Protocols in Molecular Biology*. John Wiley & Sons, New York., Vol 1, pp 7.2.1–7.2.20.
7. Rashtchian, A., Thornton, C., and Heidecker, G. (1992) *PCR Meth. & Appl.* **2**, 124–130.
8. Buchman, G., Booth, P., and Rashtchian, A. (1993) *Focus* **15**, 36–41.
9. Schuster, D.M., Buchman, G.W. and Rashtchian, A. (1992) *Focus*, **14**, 46–52.
10. Frohman, M.A. and Martin, G.R. (1989) *Technique*, **1**, 165–170.
11. Shen W. and Wayne, M.M.Y. (1988) *Gene*, **70**, 205–211.
12. Frohman, M.A., Dush, M.K. and Martin G.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002.
13. Ohara, O., Dorit, R.L. and Gilbert, W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5673–5677.
14. Mueller, P.R. and Wold, B. (1989) *Science*, **246**, 780–786.
15. Loh, E.Y., Elliot, J.T., Cwirla, S., Lanier, L.L. and Davis, M.M. (1989) *Science*, **243**, 217–220.
16. Triglia, T., Peterson M.G. and Kemp, D.J. (1988) *Nucl. Acids Res.*, **16**, 8186.