2614–2615 *Nucleic Acids Research, 1996, Vol. 24, No. 13*

DOP-vector PCR: a method for rapid isolation and sequencing of insert termini from PAC clones

Chenyan Wu, Shigui Zhu, Stacey Simpson and Pieter J. de Jong*

Department of Human Genetics, Roswell Park Cancer Institute, Buffalo, NY 14263, USA

Received April 1, 1996; Accepted May 21, 1996

To develop a physical map, one often uses a large-insert library to establish a contiguous set of overlapping genomic clones ('contig'). One of the most efficient ways to prepare contigs in limited chromosomal regions is chromosome-walking. For this approach, it is important to retrieve DNA sequences from the termini of the insert fragment to screen for overlapping clones. Many methods for the isolation of DNA insert termini have been developed: subcloning (1), inverse PCR (2), Alu-vector PCR (3), bubble-vector PCR (4), adapter-vector PCR (5), junction trapping (6), capture-PCR (7), TAIL-PCR (8) and direct sequencing of the end fragments (9). Many of these methods work well but are difficult to scale up to large numbers of end-rescue reactions. Our goal was to improve the TAIL-PCR method, such that nearly every insert end from P1-derived artificial chromosome (PAC) clones (10) can be amplified directly from bacterial suspension, thus avoiding the cost of plasmid DNA extraction.

Here we present an efficient method, termed DOP-vector PCR, to amplify unknown terminal sequences from PAC clones. This technique should also be applicable to other systems, e.g. BAC or YAC clones. The basic concept is to employ a partially-degenerate oligonucleotide primer (DOP) in combination with a vector primer (Fig. 1). The partially degenerate primer 6-MW (CCGACTCGAG-NNNNNNATGTGG) has been described previously (11). It has six bases specified at the 3′-end, six completely degenerate bases in the middle, and an arbitrary 10 nt sequence at the 5′-end. A single cycle of DNA synthesis is carried out at a low stringency annealing temperature in the presence of the 6-MW primer. Primer extension is catalyzed by the modified T7 DNA polymerase (Sequenase version 2.0, USB) as has been reported previously (12). Under these low stringency conditions, priming and subsequent DNA elongation probably occurs from numerous discrete sites in the PAC DNA (and the *Escherichia coli* DNA), which match for at least 5 or 6 nt with the 3′ terminal nucleotides of the DOP primer. For our application, it is essential to permit random priming while preventing random amplification. Therefore, only a single cycle of low stringency DNA synthesis is allowed. To selectively amplify the vector/insert junction sequences, high-stringency PCR conditions are imposed using a vector primer in combination with the DOP primer and using *Taq* polymerase for DNA synthesis. During the first PCR cycle, a subset of the nascent DNA strands (from the Sequenase synthesis) are used as templates for DNA synthesis primed by the vector primer. The products include, for the first time, the complementary sequence for the DOP primer, which can serve as a high stringency priming site for the 6-MW primer in subsequent PCR cycles. The specificity for junction fragments can be further enhanced by a second PCR employing a nested vector primer in combination with the DOP primer (Fig. 1D). Please note that the anticipated junction PCR fragments generated from the first PCR reaction will be truncated in the second PCR reaction with nested primers. This will result in a shifting of the PCR mobility of 73 bp for the T7 junction and 78 bp for the SP6 junction.

We have employed DOP-vector PCR to attempt rescuing both ends of 125 PAC clones. From the 250 ends, 246 ends were amplified on the first attempt. After a second try for the failed ends, we succeeded in amplifying the remaining junctions using conditions identical to the first attempt. The PCR product sizes generated with *Taq* polymerase range from a few hundred base pairs to ~1 kb. Among 250 PAC end fragments we used for sequencing, 72.8% of them are >600 bp; 8.8% between 500 and 599 bp; 6%, 400–499 bp; 8.8%, 300–399 bp; 2.4%, 200–299 bp; and 1.2%, 100–199 bp. Most PCR reactions result in several (two to five) distinct PCR products per end. The long-range PCR preparation 'Expand' from Boehringer Mannheim has also been used and results in mixtures of PCR fragments which also contain slightly larger products sizes (≤3 kb) and lacking the over-representation of the smallest PCR fragments. A sample of PCR products generated for four different PAC clones is shown in Figure 2. As can be seen, all four PAC clones generate at least one major PCR product and often more fragments of lower intensity are detectable. Most of the major and minor PCR products shift in mobility after the second PCR with the nested vector primers, consistent with the expectation that the products are derived from the vector/insert junctions. Frequently, a weak smear can be observed and most likely consist of non-specific products generated from many template locations with the partially degenerate primer. If it is the purpose to sequence the junction fragments, then the mild contamination with non-junction sequences is of no relevance since these background products will not have the capability to bind the sequencing primer.

To prove that the PCR products are, indeed, derived from the vector/insert junctions, we purified the main PCR fragment from each of the 250 reactions using electrophoresis in 1% low melting point agarose. The desired bands (which demonstrated in all cases the predicted size shift) were cut out and purified with the Wizard™ PCR purification kit (Promega) or, in some cases, using the gelase method (Epicentre Technologies). The sequencing reactions were performed using the *Taq* DyeDeoxy Terminator Cycle Sequencing kit (ABI) with the nested vector primers as the sequencing primers. The results obtained with the automated DNA sequencer 373A (ABI) indicate that in all cases the vector sequences immediately adjacent to the cloning site are present in

* To whom correspondence should be addressed

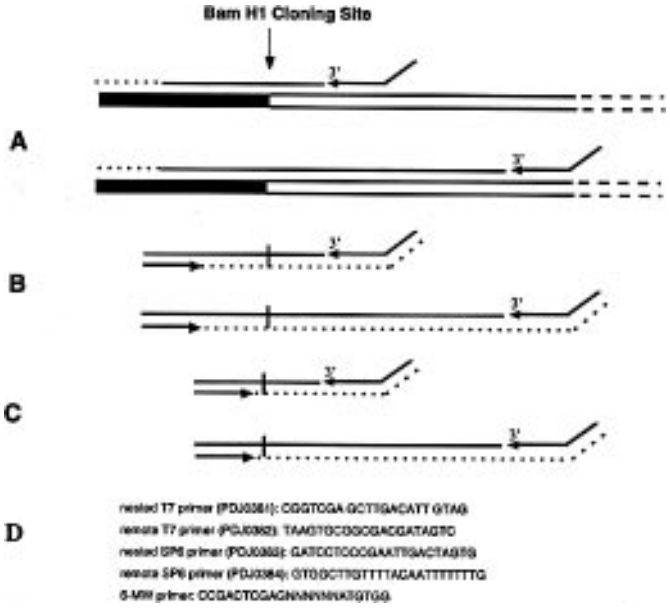

Figure 1. Isolation of vector/insert junction sequences from PAC clones. Junction fragments are isolated after three subsequent reactions: (**A**) Low stringency priming with a partially degenerate primer (6-MW) and DNA synthesis using Sequenase. For simplicity, only two priming sites towards the vector are shown. (**B**) Preferential PCR amplification of junction sequences using a remote vector primer and the 6-MW primer. (**C**) Further enrichment for junction sequences using a nested vector primer and the 6-MW primer. (**D**) Primers used in this method. Prior to the Sequenase reaction, 0.5µl of bacterial cell suspension [overnight culture (LB media), 10-fold diluted with 1× PCR buffer] is added to 3.5 µl premixed stock solution [1 µl 5× Sequenase buffer (100 mM potassium phosphate, pH 7.4, 5 mM DTT, 0.5 mM EDTA), 5 pmol 6-MW primer and 1.25 nmol of each dNTPs]. The reaction is preheated at 96°C for 3 min, cooled to 30°C, then Sequenase version 2.0 (USB) (1 µl, 0.8 U) is added. The temperature is ramped to 37°C over a 1 min interval and the reaction is incubated at 37°C (3 min). After incubation, the temperature is increased to 72°C immediately and held at this temperature (~10 min) pending the addition of 45 µl pre-warmed mixture, which contains 1× PCR buffer (50 mM KCl, 10 mM Tris–HCl pH 8.3, 0.001% gelatin) (Perkin-Elmer), 2.0 mM $MgCl_2$, 0.2 mM each of dNTP, 0.5 µM remote vector primer [PDJ0382 or PDJ0384, see (D)], 0.1 µM 6-MW primer and 1 U Ampli *Taq* polymerase (Perkin-Elmer). PCR is executed for 35 cycles at 95°C for 1 min, 58°C for 2 min and 72°C for 2 min with a final extension at 72°C for 5 min. In the secondary PCR, 0.5 µl primary PCR products are used in 50 µl reactions under the same conditions as in the first PCR, except that the nested vector primer (PDJ0381 or PDJ0383) (0.5 µM) is used and that 6-MW has a higher concentration (0.5 µM). The distance between the remote and nested vector primers is 73 bp for the T7 end and 78 bp for the SP6 end.

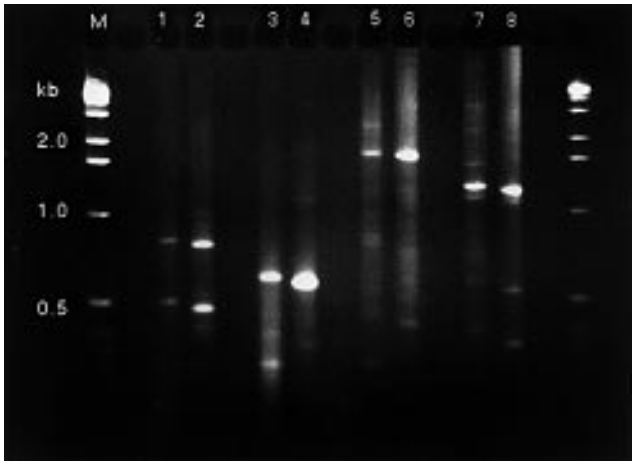

Figure 2. Gel Analysis of DOP-vector PCR products from PAC clones. PCR products obtained as described in Figure 1 were separated by electrophoresis in 1% agarose gels (TBE buffer). The left lane ('M') contains the 1 kb ladder (LifeTechnologies). Lanes 1, 3, 5 and 7 are the results from the first PCR reaction for four different anonymous PAC clones. Lanes 2, 4, 6 and 8 have the corresponding PCR products obtained after the subsequent PCR with nested vector primer. Lanes 3–6 were generated using primers from the T7 end of the vector and the remaining lanes from the SP6 end. Lanes 1–4 were PCR-amplified using *Taq* polymerase ('AmpliTaq', Perkin Elmer) whereas lanes 5–8 resulted from the use of long-range PCR enzyme ('Expand', Boehringer Mannheim).

the data. This demonstrates that the PCR products are, indeed, derived from the insert termini. As can be predicted, many of the sequence results indicate the presence of repetitive elements. Approximately 39% of the sequences contain at least one element of either an *Alu*, L1 or THE repeat (data not shown). Since it is often necessary to design a unique STS marker, the sequence can not always be used. In such cases, it is possible to sequence the gel purified PCR fragment from the other end, using the 6-MW primer as the sequencing primer. In these cases, it is desirable to obtain the PCR products using long-range PCR conditions, to generate fragments substantially >1 kb, and to have a good opportunity to have non-overlapping sequence data from the other end of the PCR fragment.

The DOP-vector method for the isolation of terminal insert fragments does not involve restriction enzyme digestion, ligation or cloning. It may be widely used in map construction. We are currently modifying the method to eliminate the gel electrophoresis requirement. For example, it may be possible to use biotinylated vector primer and streptavidin magnetic beads to purify the PCR products and automate this approach for large-scale end sequencing.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the US Department of Energy (#DE-FG02-94ER61883), NCHGR, NIH (#1R01RG-01165) and the NIH Eye Institute (EY10541).

REFERENCES

1 Traver,C.N., Klapholz,S., Hyman,R.W. and Davis,R.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5898–5902.
2 Ochman,H., Gerler,A.S. and Hartl,D.L. (1988) *Genetics*, **120**, 621–623.
3 Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7485–7489.
4 Riley,J.H., Butler,R., Ogilvie,D., Finniear,R., Jenner,D., Powell,S., Anand,R., Smith,J.C. and Markham,A.F. (1990) *Nucleic Acids Res.*, **18**, 2887–2890.
5 Kere,J., Nagaraja,R., Mumm,S., Ciccodicola,A., D'Urso,M. and Schlessinger,D. (1992) *Genomics*, **14**, 241–248.
6 Patel,K., Sheer,D. and Hampton,G.M. (1993) *GATA*, **10**, 42–48.
7 Lagerström,M., Parik,J., Malmgren,H., Stewart,J., Pettersson,U. and Landegren,U. (1991) *PCR Methods Applic.*, **1**, 111–119.
8 Liu,Y. and Whittier,R.F. (1995) *Genomics*, **25**, 674–681.
9 Kimmerly,W.J., Kyle,A., Lustre,V.M., Martin,C.H. and Palazzolo,M.J. (1994) *GATA*, **11**, 117–128.
10 Ioannou,P.A., Amemiya,C.T., Garnes,J., Kroisel,P.M., Shizuya,H., Chen,C., Batzer,M.A. and de Jong,P.J. (1994) *Nature Genet.*, **6**, 84–89.
11 Telenius,H., Carter,N.P., Bebb,C.E., Nordenskjold,M., Ponder,B.A. and Tunnacliffe,A. (1992) *Genomics*, **13**, 718–725.
12 Guan,X.-Y., Trent,J.M. and Meltzer,P.S. (1993) *Hum. Mol. Genet.*, *2*, 1117–1121.