Short Technical Reports

SHORT TECHNICAL REPORTS

Manuscripts published in the Short Technical Reports section are shorter and less comprehensive in scope than full Research Reports.

Simple Strategy for Sequencing cDNA Clones

BioTechniques 21:446-452 (September 1996)

ABSTRACT

*We describe a simple method for constructing subclones containing overlapping nested deletions from cDNA clones (both λ phage clones and plasmid clones). A PCR-amplified insert is partially digested with 4-cutter restriction enzyme(s). Complete digestion of this DNA with two restriction enzymes, having unique cutting sites at one or the other end of the amplified DNA, creates two sets of overlapping nested subfragments. When recloned into each of two doubly cut pBluescript[®] plasmid vectors, only the two sets of nested subfragments are produced. Minimal nested sets can be constructed by screening subclones using colony PCR, and this set can then be used to determine the entire sequence of the cDNA clone. This method requires only a single cloning step and can be generated from an insert that is amplified directly from a λ phage clone. This procedure eliminates the sequencing redundancy problem inherent in shotgun cloning, allows large clones to be sequenced using universal primers only and is well-suited for automated DNA sequencing. Using this method, we successfully sequenced five cDNA clones of five *Drosophila* subobscura genes.*

INTRODUCTION

High-throughput sequencing of complete cDNA clones in λ phage vectors, such as Stratagene's Lambda-ZAP[®], presents several challenges. Typically, a phage clone must first be converted into a plasmid before carrying out subcloning procedures. For inserts greater than 500–600 bases in length, the entire sequence cannot be obtained using only universal primers. Complete sequences must be obtained either by primer walking, a relatively slow and sometimes laborious process, or by subcloning pieces of the insert. The latter involves either the creation of shotgun clones (1,3–7,10) or the creation of ordered nested deletions (8,9,11). In addition, oligo(dT) primed reverse transcription (first-strand cDNA synthesis) produces a long homopoly-

mer run at the 3' end of each cDNA clone. Such a run promotes DNA polymerase skipping, thereby preventing the determination of the sequence from universal primer sites flanking the 3' ends of the clone.

Nested deletion and shotgun cloning strategies involve trade-offs between efficiency and convenience. Both procedures require a number of DNA manipulations. Shotgun cloning involves several low-efficiency steps: size fractionation of sheared or sonicated DNA, end repairing and blunt-end ligation. Obtaining full coverage of an insert requires redundancy of sequencing effort since the position of each shotgun clone relative to the complete sequence cannot be determined before sequencing. In a typical 5-fold or 6-fold redundancy shotgun sequencing project, some regions will have at least 10-fold coverage (3,4). The nested deletion strategy reduces the sequencing effort required to obtain a complete sequence. However, nested subclones cannot be generated in a λ phage vector, which necessitates recloning the cDNA insert into an appropriate plasmid vector. In the recloning step, the insert cannot come from a polymerase chain reaction (PCR)-generated product without introducing the possibility of *Taq* DNA polymerase-generated mutations.

Here we describe two enzymatic procedures (using blunt ends or 5' CG overhangs) for generating overlapping nested deletions from λ phage clones. Both procedures use the same basic strategy (illustrated in Figure 1). The insert DNA is obtained by PCR amplification of a cored plaque from universal primers sites flanking the insert. The DNA is partially digested with one or more 4-cutter restriction enzymes to produce either blunt ends or 5' CG overhangs at numerous locations throughout the insert. The sequences at the ends of the amplified DNA contain recognition sites for the restriction enzymes *NotI* or *ApaI*. Digestion with the two rare cutting enzymes allows subfragments to be directional cloning into appropriately prepared plasmid vectors. For inserts with a blunt end, two double-cut plasmid vectors are produced by digesting pBluescript[®] with the restriction enzyme *SacI*, which is adjacent to a good universal sequencing primer site,

blunting the ends with *Pfu* DNA polymerase and digesting with either *Not*I or *Apa*I. Alternative vectors for cloning inserts with 5' CG-overhangs are produced by substituting *Cla*I for *Sac*I and not filling in the ends. Following dephosphorylation of the insert DNA, ligation into the prepared vectors and transformation into *E. coli*, colony PCR with universal primers are used to screen for nested subclones. Sequencing is also carried out using universal primers. We have simplified the procedure as much as possible so that most of the steps are carried out in a single buffer and tube.

MATERIALS AND METHODS

Vector Preparation

Blunt-end. Five micrograms of pBluescript II SK(+) (Stratagene, La Jolla, CA, USA) plasmid DNA were digested with *Sac*I (20 U enzyme, 3 h to overnight at 37°C), verified on a 1% agarose gel and purified with a QIAquick® Spin PCR Purification Column (Qiagen, Chatsworth, CA, USA) in 50 µL distilled water. To make the *Sac*I ends blunt, the digest was treated with 10 U *Pfu* DNA Polymerase (Stratagene) in reaction buffer provided by the company [20 mM Tris-HCl (pH 8.75), 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1% Triton® X-100, 0.1 mg/mL bovine serum albumin (BSA)] for 30 min at 72°C. The reaction was terminated by purifying the DNA for a second time with a QIAquick column. The DNA was divided equally into two tubes, to which 10–25 U of *Not*I or *Apa*I and appropriate 10× restriction buffers were added. After complete digestion (overnight at 37°C for *Not*I and 25°C for *Apa*I), the reactions were terminated with a third QIAquick column purification. The volume was adjusted to yield a DNA concentration of 20 ng per µL in 1/10 TE buffer (1 mM Tris, 0.1 mM EDTA, pH 8.0).

5' CG-overhang. Five micrograms of plasmid DNA were digested with 20 U of *Cla*I instead of *Sac*I. Except for the fill-in step, which was omitted, the remaining steps in the vector preparation were identical to the procedure described above.

Insert Preparation

Blunt-end. PCR amplification of 2 µL of a cored plaque eluted in 200 µL of SM buffer (9) was carried out in a final volume of 80 µL containing 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl₂, 50 mM KCl, 1.6 U *Taq* DNA Polymerase (Boehringer Mannheim, Indianapolis, IN, USA), 200 µM deoxyribonucleoside triphosphates and 150 ng each of oligonucleotide primers. The universal primers are M13 reverse sequencing primer (5' GGAAACAGCTATGACCATG 3') and M13-20 (5' GTAAAACGACGGCCAGT 3'). The phage DNA

was amplified using the following program: 2 min for the initial denaturation at 95°C; 30 rounds of 30 s for denaturation (94°C), 30 s for annealing (56°C) and a 1.5-min extension (72°C); and a final extension for 5 min at 72°C. The PCR product was checked on a 1% agarose gel and purified with a QIAquick column yielding approximately 1 µg DNA in 50 µL water. Partial digestion of the DNA was carried out with one or more of the following enzymes to create blunt-ends: *Alu*I (AG↓CT), *Cvi*JI ([A/G]G↓G[T/C]), *Cvi*RI (TG↓CA), *Dpn*I (GA↓TC), *Hae*III (GG↓CC), *Mvn*I (GC↓GC) and *Rsa*I (GT↓AC).

Figure 1. Strategy for producing nested deletions from the PCR-generated inserts of λ phage cDNA clones. A single partial digestion of the PCR product with a blunt-end-producing 4-cutter enzyme (e.g., *Cvi*JI), followed by complete digestion with *Not*I and *Apa*I, yields nested sets of deletions from both ends of the insert DNA. Each of the two prepared vectors allows for one of the nested sets to be subcloned. An alternate procedure that produces 5' CG overhangs is described in the text.

Short Technical Reports

The digestion, containing 0.05–0.1 U enzyme (1–2 U for *Cvi*II), was carried out for 10–20 min at the recommended temperature. In general, the reaction is better controlled by adjusting the amount of enzyme than by adjusting the digestion time. NEB buffer No. 3 (50 mM Tris-HCl, 10 mM MgCl₂, 100 mM NaCl and 1 mM dithiothreitol [DTT], pH 7.9 and 0.1 µg/µL BSA; New England Biolabs, Beverly, MA, USA) was used for the partial digestions, so that the subsequent complete digestion can be carried out in the same buffer.

Aliquots of the partial digestion (15 µL) were taken at 10, 15 and 20 min and then heat-inactivated. One half of each sample was run on a 1% agarose gel, and the best partial digestion was chosen for subsequent digestion. The best results were obtained when approximately 25%–50% of the PCR product remained undigested. The selected partial digests were pooled and digested with 5 U each of *Not*I and

*Apa*I (1 h, 37°C). One unit of shrimp alkaline phosphatase (1 U/µL; United States Biochemical, Cleveland, OH, USA) was also added to the reaction to dephosphorylate the ends. The DNA was purified with a QIAquick column and resuspended in 50 µL of water or 1/10 TE buffer.

5' CG-overhang. To generate fragments with 5' CG overhangs, *Ac*II (C↓CG↑C), *Hin*PII (G↓CGC), *Hpa*II or *Msp*I (C↓CGG), *Mae*II (A↓CGT) or *Taq*I (T↓CGA) can be used. *Mae*II and *Taq*I cannot be heat-inactivated; these partial digests must be column-purified before dephosphorylation.

Ligation

Twenty nanograms (1 µL) of prepared vector DNA and 16 µL of insert DNA were ligated overnight at 16°C in a final volume of 20 µL containing 50 mM Tris-HCl (pH 7.8), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, 25 µg/mL BSA and 1 µL T4 DNA ligase (400 U/

µL; New England Biolabs). XL1 Blue (Stratagene) competent cells were transformed with 3 µL of the ligation mixture and were plated on antibiotic plates (LB with 100 µg/mL ampicillin) spread with 40 µL 5% 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and 20 µL of 100 mM isopropyl-β-D-thiogalactopyranoside (IPTG).

Colony Screening

Colony PCR was used to screen for inserts. White colonies were touched with pipet tips (200 µL) and washed off into 80 µL PCR premixture containing M13 reverse and M13-20 primers, as described above. The insert sizes were visualized by agarose gel electrophoresis and ethidium bromide staining of 1/10 of the PCR products. Clones with inserts representing a range of sizes and differing by 200–400 bp were chosen for subsequent sequencing. Two clones of each size class were sequenced.

DNA Sequencing

The insert DNA samples amplified by colony PCR were purified with QIAquick columns and resuspended in 50 μ L water, of which 9.5 μ L (ca. 200 ng) were used in a cycle sequencing reaction. For blunt-ended clones, P51 (5' AAGGGAACAAAAGCTGG 3') (2) was used to sequence all the clones (both strands). For 5' CG overhang clones, P51 was used to sequence inserts cloned in an *Apa*I-digested vector (the + strand sequencing), and P172 (5' TATAGGGCGAATTGGGT 3') (2) was used to sequence inserts cloned in *Not*I-digested vector (the - strand sequencing). For cycle sequencing of all the clones, 9.5 μ L (ca. 200 ng) and 1 μ L (50 ng) of one of the above universal primers were mixed with 9.5 μ L of cycle sequencing premixture containing dye terminators, dNTP and DNA polymerase (PRISMTM; Perkin-Elmer/Applied Biosystems Division, Foster City, CA, USA). The cycle sequencing conditions are 15 s for denaturation at 95°C, 30 s for annealing at 56°C and a 4-min extension at 60°C for 25 rounds. After cycle sequencing, samples were extracted twice in an equal volume of phenol/chloroform (1:1), followed by ethanol-precipitation (300 μ L ethanol, 15 μ L 2.0 M Na acetate, pH 4.5). Samples were washed with 1 mL 70% ethanol, dried and resuspended in a formamide:EDTA buffer (5:1) (10 mM EDTA) containing blue dextran resin (Catalog No. D 5751; Sigma Chemical, St. Louis, MO, USA). The latter aids in

the visualization of samples during gel loading. Samples were heated at 94°C for 2 min before loading on a 4.5% DNA sequencing gel. Gels were run at 35 W for 12–14 h.

RESULTS AND DISCUSSION

We determined the sequences of five cDNA clones, screened from a *Drosophila subobscura* adult cDNA library. The identities of the five clones were determined by comparison of their partial sequences with *Drosophila melanogaster* sequences in GenBank[®]. The clones are homologous to ADP/ATP translocase, ATP synthase β subunit (ATPsyn- β), enolase (Eno), glutamic acid decarboxylase (GAD) and triose-phosphate isomerase (Tpi). Insert lengths ranged from 0.7–3.0 kb. GAD, a 3-kb clone, was prepared for sequencing using *Cvi*JI to generate blunt-ended nested subfragments. An example of nested clones produced by this method is shown in Figure 2. The average length of our sequences is approximately 400 bases and contained fewer than 3 ambiguous bases. A single 40-base gap in the sequence remained after screening and sequencing 33 subclones. This gap was filled with a specific sequencing primer.

Nested deletions for the other four genes were generated by pooling the four clones and digesting the combined sample with three 4-cutter enzymes (*Hin*p1I, *Hpa*II and *Taq*I) producing 5' CG overhangs. The best partial digestion among the three time points (10, 15 and 20 min) was selected for each enzyme, and the digests were pooled for subsequent cloning and sequencing. A total of 83 subclones were sequenced, and the sequences were assembled using the SeqManTM Assembling Program (DNASTAR, Madison, WI, USA). The program assembled the sequences into four contigs, corresponding to each of the four genes. All four clones were completely covered by at least two independent sequences. Most regions contained sequences from both strands. As expected, the region corresponding to the interval between the poly(A) homonucleotide run at the 3' end of each clone, and the first internal restriction site had only single-

Figure 2. A nested deletion set generated from a 3-kb clone using *Cvi*JI to generate partial digestions. Lanes 2–11 are nested subclones selected for sequencing. Lane 1 is the *Dra*I-digested λ DNA size marker.

Short Technical Reports

strand coverage.

Our strategy for constructing nested overlapping deletions overcomes a problem with DNA sequences obtained from PCR-generated subclones: *Taq* DNA polymerase-generated mutations. Unlike procedures in which a single subclone of a PCR-generated product is the template for sequencing, every subclone in our procedure will be independently derived. A mutation in one subclone can, therefore, be revealed by comparing its sequences to that of other independently derived clones of the same region. To eliminate PCR-generated mutations, our goal was to sequence two clones of each size class and of each strand, for a total of 4-fold coverage. This procedure is expected to reduce sequencing error to a minimum.

Dephosphorylation of the insert DNA is necessary to avoid the production of chimeric subclones. In a trial experiment, we found that more than half

of the subclones were chimeras when this step was omitted. In contrast, we did not find any chimeric subclone out of a total of 116 subclones when the dephosphorylation step was included.

The 8-cutter site *NotI* is unlikely to be present in any cloned cDNA. This is not true for *ApaI*, which is a 6-cutter. For clones with internal *ApaI* cut sites, the procedure will generate a nested set for each *ApaI* subfragment. A complete sequence can be generated from the strand corresponding to the *NotI*-derived inserts.

PCR products obtained directly from λ phage clones can also be sequenced by a shotgun cloning procedure. Technically, this is a somewhat more difficult procedure to carry out because the ends of the PCR product must be polished initially to allow circularization before sonication or shearing. In addition, the cloning efficiency of shotgun-prepared DNA is lower than

that of restriction-digested DNA, especially with enzymes producing overhangs, as in our method.

Our method generates overlapping nested deletions of λ phage cDNA clones at high efficiency without the need for a plasmid intermediate. It combines the best features of shotgun cloning and nested deletion strategies. It can be adapted to be used for a number of different plasmid and phage cDNA clones. Although this procedure requires the preparation of two vectors to distinguish both nested deletion sets, all of the inserts are generated in one tube. Cloning efficiency of the prepared DNA is also high. Although the procedure relies on restriction sites for generating nested subclones, it is our experience that gaps are avoided with the use of several 4-cutter enzymes. We have found this method to be more efficient than either shotgun cloning or primer walking. The ability to use universal

--	--

primers for all sequencing allows optimization of the cycle-sequencing reaction, reliable production of high-quality sequences and high-throughput sequencing.

REFERENCES

1. **Anderson, S.** 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* 9:3015-3027.
2. **De Bellis, G., M. Manoni, R. Pergolizzi, M.E. Redolfi and M. Luzzana.** 1992. A more stringent choice of primers can improve the performance of fluorescent automated DNA sequencers. *BioTechniques* 13:892-897.
3. **Deininger, P.L.** 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* 129:216-223.
4. **Démolis, N., L. Mallet, F. Bussereau and M. Jacquet.** 1995. Improved strategy for large-scale DNA sequencing using DNaseI cleavage for generating random subclones. *BioTechniques* 18:453-457.
5. **Fitzgerald, M.C., P. Skowron, J.L. Van Eten, L.M. Smith and D.A. Mead.** 1992. Rapid shotgun cloning utilizing the two base recognition endonuclease *CviII*. *Nucleic Acids Res.* 20:3753-3762.
6. **Guo, L.-H., R.C.A. Yang and R. Wu.** 1983. An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid. *Nucleic Acids Res.* 11:5521-5540.
7. **Messing, J., R. Crea and P.H. Seeburg.** 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9:309-321.
8. **Ozkaynak, E. and S.O. Putney.** 1987. A unidirectional deletion technique for the generation of clones for sequencing. *BioTechniques* 5:770-773.
9. **Sambrook, J., E.F. Fritsch and T. Maniatis.** 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. **Schriefer, L.A., B.K. Gebauer, L.Q.Q. Qiu, R.H. Waterston and R.K. Wilson.** 1990. Low pressure DNA shearing: a method for random DNA sequence analysis. *Nucleic Acids Res.* 18:7455.
11. **Zhu, K.Y. and J.M. Clark.** 1995. Rapid construction of nested deletions of recombinant plasmid DNA for dideoxy sequencing. *BioTechniques* 18:222-224.

Address correspondence to Ling-Wen Zeng, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60636, USA. Internet: lzeng@midway.uchicago.edu

Received 31 August 1995; accepted 16 February 1996.

Ling-Wen Zeng and Martin Kreitman
University of Chicago
Chicago, IL, USA