Selective Protection of Restriction Endonuclease Sites

BioTechniques 34:512-522 (March 2003)

ABSTRACT

A difficulty that is encountered when attempting to insert a PCR-amplified product or DNA fragment of interest into a particular vector is the presence within the insert of one or more internal restriction endonuclease (RE) sites identical to those selected for the flanks of the insert. Our method circumvents this problem by partially protecting internal RE sites while flanking sites for the same RE are cleaved. The amplified product is first heat denatured in the presence of excess amounts of perfectly complementary oligonucleotides that can anneal to the flanks of the insert. The mixture is allowed to anneal and is subsequently digested with the appropriate endonucleases. This results in the cleavage of the flanking RE sites while digestion at the internal RE site is not efficient. The mixture is subsequently heat denatured and column purified to remove the oligonucleotides. The product is then allowed to anneal and can be used directly in a ligation reaction with the plasmid vector. This method facilitates the construction of recombinant molecules by creating desired flanks while preserving internal RE sites.

INTRODUCTION

A variety of methods have been developed to clone PCR-amplified products into plasmid vectors of choice (1–5). A major obstacle is the presence of one or more “internal” restriction endonuclease (RE) sites identical to those selected for the flanks of the insert. In these situations, most current methods rely on multi-molecular ligation (4,6) or on stepwise insertion of sub-fragments. These methods have drawbacks in that a multi-molecular ligation, for instance, can give rise to a large number of improperly oriented recombinant clones. In addition to being time consuming, stepwise insertion can also result in a large number of incorrect recombinant

clones, depending on the number of common internal and flanking restriction sites. An alternative method that relies on the PCR amplification and use of the type IIS RE *Eam*1104 I, an enzyme that cleaves outside its recognition sequence, has been described (2). In this method, primers that contain the *Eam*1104 I recognition site (5'-CTCTTC-3') are used to amplify the DNA fragments of interest and the vector sequence. The drawback with this method is that both vector and insert must be amplified, regardless of size. In addition, the amplification has to be performed in the presence of 5-methyldeoxycytosine to prevent the endonuclease from recognizing naturally occurring internal *Eam*1104 I sites in the vector and the insert. Another method involves a strategy in which blunt-ended overlapping sequences on a vector and PCR-amplified insert are subjected to limited λ exonuclease (λ exo) digestion, thus generating complementary cohesive ends that are subsequently annealed and transformed in the absence of ligase (5). The primary difficulty with this method is synchronizing the 5'→3' digestion of the insert and vector by λ exo. Yet another method that uses synthetic pseudocomplementary peptide nucleic acids (pcPNAs) has been shown to shield RE sites from the action of the endonucleases (1). However, pcPNAs hybridize less efficiently with linear DNA compared to supercoiled DNA, which makes them less attractive when one is attempting to clone a PCR product with one or more sites to be protected and others to be digested. In addition, salt concentrations required for the annealing of the pcPNA to the site to be protected can interfere with RE digestion at other selected sites.

We have developed a simpler method that relies mainly on the ability of short oligonucleotides that are in molar excess to anneal to a complementary sequence within a long stretch of DNA faster than the equimolar amount of a full-length complementary strand. The feasibility of our strategy was confirmed by the insertion of the *rep* gene of adeno-associated virus type 2 (AAV-2) in place of the bacterial chloramphenicol acetyltransferase (*cat*) gene of plasmid pNFA-CAT, which contains a modified version of the human immunodeficiency virus type 1 (HIV) long terminal repeat (LTR) promoter.

MATERIALS AND METHODS

Plasmids

Plasmid pAV2 carries the full-length AAV-2 genome (7). Plasmid pNFA-CAT (8) (a gift from Arnold Rabson, Rutgers University, Piscataway, NJ, USA) contains a mutated HIV-LTR promoter, with a deletion of the NF κ B binding site, ligated to the *cat* gene. The *cat* gene within pNFA-CAT is flanked by an upstream *Hind*III site and a downstream *Bam*HI site (Figure 1B). These are suitable sites for the insertion of the *rep* gene. Plasmid pBlueScript® II SK⁺ was obtained from Stratagene (La Jolla, CA, USA).

Oligonucleotides

Oligonucleotides were synthesized by Invitrogen (Carlsbad, CA, USA). Table 1 provides these sequences. The *rep* gene of AAV-2 encodes four proteins (Rep78, Rep68, Rep52, and Rep40) from overlapping ORFs. Primers P1 and P2 were used to amplify the 1929-bp

Figure 1. Relevant portions of plasmids used in this study. (A) Plasmid pAV2. The Rep68 ORF is shown as gray rectangles. Numbers indicate base positions within the AAV-2 genome. (B) Plasmid pNFA-CAT. The modified HIV-LTR is shown as an open rectangle. The *cat* gene is shown as a gray rectangle. Drawings are not precisely to scale.

Short Technical Reports

Table 1. Oligonucleotides Used in this Study

Name	Description	Sequence ^a (5'→3')	RE Site
P1	rep ORF upstream PCR primer	agtaagcTTGAACGCGCAGCCGCCATG	<i>Hind</i> III
P2	rep ORF downstream PCR primer	ggaagatctttaTTTGAGCTTCCACCACTGTCTTA	<i>Bgl</i> II
CP1	Complement of P1	CATGGCGGCTGCGCGTTCAAgcttact	<i>Hind</i> III
CP2	Complement of P2	TAAGACAGTGGTGAAGCTCAAAtaaaagatcttcc	<i>Bgl</i> II
HG1	Complement ^b of internal <i>Hind</i> III site	GTCAGACGCGGtAGCTTCGATCAACTACGC	None

^aRestriction enzyme sites are underlined. Sequences within the *rep* gene are capitalized. There is a box around the mutated *Hind*III site in HG1.

^bSingle base mismatches are indicated by lowercase letters.

portion of the *rep* gene that contains all four ORFs (Figure 1A) from plasmid pAV2 (7). Oligonucleotides CP1 and CP2 are complementary to P1 and P2, respectively. A *Hind*III site was incorporated into primers P1 and CP1, and a *Bgl*II site was included in primers P2

and CP2. *Hind*III and *Bam*HI were used to remove the *cat* gene from pNFA-CAT, which enabled us to insert the *rep* gene coding sequence. The *Bgl*II recognition sequence was incorporated into the primer pair P2 and CP2 because the *rep* gene coding sequence contains an inter-

nal *Bam*HI site (Figure 1A), and *Bgl*II overhangs are compatible with *Bam*HI overhangs. Primers P2 and CP2 also contain the AAV-2 polyadenylation signal (9). HG1 is a 30-bp oligonucleotide that is complementary to the region of the internal *Hind*III site that is 1560–1565 bp downstream from the initiation ATG of the Rep68 and Rep78 ORFs (9,10), except for an A→T change within the *Hind*III recognition sequence.

PCR Conditions

The amplification of the *rep* gene coding sequence was performed as follows. The reaction contained 1× cloned *pfu* buffer (Stratagene), 250 μM dNTPs (New England Biolabs, Beverly, MA, USA), 100 ng template (pAV2), 250 ng each of primers P1 and P2, 5 U *Pfu-Turbo*® DNA polymerase (Stratagene), and water for a total volume of 100 μL. Amplification was performed by 30 cycles of denaturation at 94°C for 40 s, annealing at 52°C for 45 s, and extension at 65°C for 2 min. For analysis, 8 μL PCR product were resolved on a 1% agarose gel containing 1× TBE buffer. The rest of the product was used for the partially protected RE digestion experiments after column purification using a QIAquick™ PCR purification kit (Qiagen, Valencia, CA, USA) to remove the bulk of the oligonucleotides.

Partially Protected RE Digestion

In our initial procedure (schematic shown in Figure 2), the purified PCR product (5–10 pmol) and 100 pmol each of primers HG1, P1, P2, CP1, and CP2 (20 μL final volume in 10 mM Tris-Cl, pH 8.5) were incubated at 90°C–95°C on a heat block for 5 min.

Figure 2. Summary of cloning strategy. Depicted is the amplification and partially protected RE digestion of the *rep* ORF, followed by its insertion into the large fragment of *Hind*III/*Bam*HI-digested pNFA-CAT. *Bgl*II/*Bam*HI indicates the ligation of *Bgl*II and *Bam*HI overhangs. Drawings are not to scale. Oligos, oligonucleotides.

Short Technical Reports

The reaction mixture was cooled by removing the heat block containing the samples from the heater and placing it at room temperature for 5–10 min. The samples were then put on ice, and 2 μ L NEBuffer no. 2, 5 U *Hind*III, and 5 U *Bgl*II (all from New England Biolabs) were added to the reaction mixture for a final volume of 25 μ L and allowed to digest for 45 min at 37°C. EDTA was added to the reaction to a final concentration of 1 mM, and the reaction mixture was then heated for 5 min at 90°C to denature the DNA. Most of the oligonucleotides were selectively removed from the reaction mixture using a QIAquick PCR purification column (Qiagen). The DNA in the final eluent was allowed to re-anneal at room temperature for 10 min. The eluent was either fractionated on a 1% agarose gel (Figure 3) or used directly for ligation to the large fragment of the *Hind*III/*Bam*HI-digested pNFA-CAT plasmid.

Ligation and Transformation

The ligation of the digested *rep* gene PCR product with the large fragment of *Hind*III/*Bam*HI-digested pNFA-CAT or pBlueScript II SK⁺ (titration experiment) was performed using T4 DNA ligase (New England Biolabs) for 10 min at room temperature. Transformation was performed by the mixture of 1–2 μ L of the ligation reaction with 100 μ L *E. coli* strain DH5 α TM competent cells (Invitrogen), followed by a 30-min incubation on ice, a 45-s heat shock at 42°C, 2–5 min incubation on ice, the addition of 900 μ L Luria-Bertani broth (LB) (3), and a 1-h incubation at 37°C. The cell suspension (100 μ L) was spread on a LB/ampicillin (100 μ g/mL) plate. The plates were inverted and incubated at 37°C overnight. Plasmid minipreps were done on individual colonies using a mini plasmid prep kit (Qiagen). The re-

combinant plasmids were analyzed by restriction digestion pattern and/or by DNA sequencing (Genemed Synthesis, San Francisco, CA, USA).

RESULTS AND DISCUSSION

The AAV-2 *rep* gene's coding region (Figure 1A), which is involved in the AAV-2 DNA replication and gene regulation, was amplified from plasmid pAV2 and cloned into the large fragment of *Hind*III/*Bam*HI-digested plasmid pNFA-CAT (Figure 1B). Plasmid pNFA-CAT contains an HIV-LTR promoter with a deletion of the NF κ B enhancer element. A diagram of our cloning strategy is shown in Figure 2. Primers P1 (with a *Hind*III site near its 5' end) and P2 (with a *Bgl*II site near its 5' end) were used to amplify the *rep* gene's coding region. Thus, the PCR product contains a *Hind*III site at the upstream

Table 2. Increasing Primer-to-Template Ratio Increases Protection of Internal Restriction Site

pmol Oligo-nucleotides per PCR Product ^a	No. of Colonies with Full-Length Inserts	No. of Colonies Tested	% Colonies with Full-Length Inserts
1:1	0	10	0%
5:1	0	10	0%
25:1	0	12	0%
50:1	3	12	25%
100:1	2	6	33%
250:1	2	5	40%

^aAmounts shown are pmol of each of four oligonucleotides.

end and a *Bgl*II site at the downstream end of the coding region. These RE sites were chosen because they were most suitable for the insertion of the *rep* ORF downstream of the HIV-LTR promoter of the plasmid pNFA-CAT. However, the *rep* gene also contains a *Hind*III site within its coding sequence (Figure 1A).

To selectively preserve this internal *Hind*III site and digest the flanking *Hind*III site and *Bgl*II site, we used oligonucleotide pairs that can anneal to

their complementary sequences at the ends of each strand of the PCR product. P1 and CP1 encompass the *Hind*III site at the upstream end, while P2 and CP2 encompass a *Bgl*II site at the downstream end (Table 1 and Figure 2). In our initial experiments, we also added oligonucleotide HG1 (Table 1 and Figure 2), which was designed to anneal to

one strand of the internal *Hind*III site. HG1 should competitively inhibit re-annealing of the internal *Hind*III site. Because HG1 contains a single base change within the *Hind*III site, a duplex of HG1 with the internal *Hind*III site should not be cleavable by *Hind*III. The reaction mixture was heat denatured, annealed, and then digested with *Hind*III and *Bgl*II at 37°C for 45 min. Figure 3 shows the results of this digestion. Partial protection of the internal *Hind*III site was evidenced by the presence of an additional band (compared to the unprotected sample) that was equivalent in size to the full-length PCR product.

After the removal of the bulk of the oligonucleotides, the re-annealed digestion products were used directly for ligation into the large fragment of *Hind*III/*Bam*HI-digested pNFA-CAT, and *E. coli* was transformed with the resulting plasmids. Approximately 30% of the ampicillin-resistant colonies tested contained the desired plasmid, as deter-

Figure 3. Agarose gel fractionation of PCR product after digestion with *Hind*III and *Bgl*II, showing protection of internal RE site. The *rep* gene coding region was amplified with primers P1 and P2. The amplified product (PCR product) was column purified and either subjected directly to *Hind*III/*Bgl*II digestion (unprotected) or *Hind*III/*Bgl*II digestion was performed after the purified PCR product (5–10 pmol) was denatured and re-annealed in the presence of 100 pmol each of primers HG1, P1, P2, CP1, and CP2 (protected) (see Materials and Methods section). The samples were resolved on a 1% agarose gel containing ethidium bromide.

mined by restriction digestion. One of these plasmids was sequenced to confirm the correct orientation of the insert and to confirm that the *rep* gene's coding region was not mutated during amplification. We predict that the frequency of desired constructs would have been even higher if we had first gel purified the partially protected digestion product before ligation. However, gel purification would only improve the yield of the desired fragment if the protected full-length fragment is significantly larger than all other fragments that could compete for ligation into the vector.

Titration of Protection with Increasing Flanking Primers

As we refined our method, we determined that the cleavage of the internal *Hind*III site was inefficient, even if only oligonucleotides P1, P2, CP1, and CP2 were used in the pre-digestion annealing

mixture (i.e., HG1 was not necessary). Next, we wished to examine whether increasing the relative concentration of the flanking primers would afford increased protection. Partially protected RE digestion was performed as described earlier (see Materials and Methods section) except that 1, 5, 25, 50, 100, or 250 pmol of each flanking oligonucleotide (P1, CP1, P2, and CP2) were used in annealing reactions containing 1 pmol of the *rep* gene PCR product before digestion with *Hind*III and *Bgl*II. An aliquot of each digested sample was fractionated on a 1% agarose gel before the QIAquick PCR purification column step (Figure 4). Protection of the internal *Hind*III site was easily detected at a 25:1 ratio of oligonucleotides to PCR product, and protection was greatest at the highest ratios tested. However, even at a 250:1 ratio, the cleavage of the internal site was easily detected. This is not surprising because we expected the internal *Hind*III site to re-anneal at some frequency.

Another concern we had was that at high ratios of oligonucleotides to PCR product, the excess complementary oligonucleotides might self-anneal, creating RE sites that might competitively inhibit the digestion of the RE sites in the PCR product. This type of competition would have the undesired effect of blocking the digestion at the flanking RE sites. Since agarose gel electrophoresis cannot distinguish between undigested PCR products and PCR products that are only cleaved at the flanking RE sites, we had to test for proper digestion of the flanking sites by the ligation of the digested products into a vector plasmid (pBlueScript II SK⁺) that had been cleaved with *Hind*III and *Bam*HI. The ligation products were transformed into *E. coli*, and the RE digestion of the plasmid miniprep DNA was analyzed to show the frequency of full-length inserts (Table 2). The presence of full-length inserts in up to 40% of the clones analyzed indicated that even at the highest oligonucleotides to PCR product ratio tested, we are frequently able to digest the flanking RE sites of the PCR product without digesting the internal RE site. In contrast to our initial results, we did not easily detect plasmids with full-length inserts at a 25:1 ratio of oligonucleotides to PCR product; however, a ratio of 50:1 result-

Figure 4. Increasing ratio of oligonucleotides to PCR product. The amplified *rep* gene coding region (1 pmol) was either undigested, digested with *Hind*III and *Bgl*II (digested), or annealed with (as indicated above) 1, 5, 25, 50, 100, or 250 pmol each of oligonucleotides P1, P2, P3, and P4 before digestion with *Hind*III and *Bgl*II. Samples were resolved on a 1% agarose gel containing ethidium bromide. The bands in the lanes labeled "Undigested" and "Digested" show the positions of the full-length and internally cleaved fragments, respectively. The position of the free oligonucleotides is also indicated. Oligos, oligonucleotides.

Short Technical Reports

ed in three of 12 plasmids containing the full-length insert (Table 2). We attribute this discrepancy to the presence of oligonucleotide HG1 in our initial experiments (which probably provided additional protection for the internal *HindIII* site). Therefore, we recommend a molar ratio of each flanking oligonucleotide to the PCR product of 100:1 to compensate for the lack of an internally annealing oligonucleotide. We also noted a significant drop in the number of ampicillin-resistant colonies with the 250:1 ratio, as opposed to the lower ratios used. We attribute this to the aforementioned competition effect from excess oligonucleotides that might competitively inhibit digestion at the flanking RE sites.

A wide variety of vectors are available for routine cloning and/or expression purposes. However, the need to use a specific vector often arises when the vector has attributes that are required for a particular purpose, as in the case of a vector that contains a specific promoter. This, in turn, may engender difficulties in finding compatible RE

sites without compromising the integrity of the full-length vector and/or insert. Often, ideally placed RE sites within the multiple cloning site of a vector may not be used because identical sites may be found within the coding region of the insert. Our data suggest that it is possible to preferentially digest RE sites at nearly any position relative to identical recognition sites at other positions on the same stretch of DNA. It affords the use of simple components with no special modification of the nucleotides, and lengthy processes are avoided. It should also be possible to simultaneously protect multiple internal sites using this method.

The main restriction on the applicability of this technique is that the sequences surrounding the RE sites to be cleaved must be known. However, this method is timely because of the recent explosions in genomic sequencing (which provide information required for the use of this method), comparative genomics (which has identified human analogs of important genes from non-human sources), and gene therapy. In gene therapy vector development, it is often necessary to express a gene amplified by PCR from a heterologous, tissue-specific promoter. A T/A cloning vector that contains the desired promoter may not be available. This means that the ends would have to be cut with two REs to generate a substrate suitable for directional cloning. If there are one or more internal sites matching those chosen for the flanks, then our method provides an efficient way to generate the desired ends while sparing the internal sites.

ACKNOWLEDGMENTS

We thank Arnold Rabson for his kind gift of pNFA-CAT. We thank Kevin Gardner, Gary Conyers, and Jennifer Fain-Thornton for useful discussions. We also thank John Chiorini for his critical reading of the manuscript.

REFERENCES

1. Izvolsky, K.I., V.V. Demidov, P.E. Nielsen, and M.D. Frank-Kamenetskii. 2000. Sequence-specific protection of duplex DNA against restriction and methylation enzymes

- by pseudocomplementary PNAs. *Biochemistry* 39:10908-10913.
2. Padgett, K.A. and J.A. Sorge. 1996. Creating seamless junctions independent of restriction sites in PCR cloning. *Gene* 168:31-35.
3. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. CSH Laboratory Press, Cold Spring Harbor, NY.
4. Siegal, M.L., D.A. Petrov, and D. De Aguiar. 1996. Triple-ligation strategy with advantages over directional cloning. *BioTechniques* 21:614-619.
5. Tseng, H. 1999. DNA cloning without restriction enzyme and ligase. *BioTechniques* 27:1240-1244.
6. Higashinakagawa, T., H. Saiga, N. Shintani, M. Narushima-Iio, and T. Mita. 1981. Localization of putative transcription initiation site on the cloned rDNA fragment of *Tetrahymena pyriformis*. *Nucleic Acids Res.* 9:5905-5916.
7. Laughlin, C.A., J.D. Tratschin, H. Coon, and B.J. Carter. 1983. Cloning of infectious adeno-associated virus genomes in bacterial plasmids. *Gene* 23:65-73.
8. Leonard, J., C. Parrott, A. J. Buckler-White, W. Turner, E.K. Ross, M.A. Martin, and A.B. Rabson. 1989. The NF- κ B binding sites in the human immunodeficiency virus type 1 long terminal repeat are not required for virus infectivity. *J. Virol.* 63:4919-4924.
9. Srivastava, A., E.W. Lusby, and K.I. Berns. 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* 45:555-564.
10. Ruffing, M., H. Heid, and J.A. Kleinschmidt. 1994. Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif. *J. Gen. Virol.* 75:3385-3392.

Address correspondence to Dr. Roland A. Owens, Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bldg. 8, Rm. 310, National Institutes of Health, 8 Center Drive MSC 0840, Bethesda, MD 20892-0840, USA. e-mail: ro6n@nih.gov

Received 3 October 2001; accepted 6 January 2003.

Kagnew H. Gebreyesus and Roland A. Owens

*National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Department of Health and Human Services
Bethesda, MD, USA*