ANALYTICAL BIOCHEMISTRY 175, 521–524 (1988)

A DNA Sequencing Strategy

YU-MAY LEE* AND SHENG-CHUNG LEE*,†,[1]

*Institute of Biological Chemistry, Academia Sinica, and †Graduate Institute of Clinical Medicine, School of Medicine, National Taiwan University, Taipei, Taiwan, Republic of China

Received April 28, 1988

A modification of Lin's systematic DNA sequencing strategy is described. A method based on the religation of compatible cohesive ends generated by *Sau*3AI and *Bam*HI was developed. The original procedure has been simplified and the yield of transfectant has been greatly improved. After complete digestion with *Bam*HI and limited cleavage with *Sau*3AI, the single-cut linear DNA does not have to be separated from the supercoil or the open circular DNA on an agarose gel. After ligation, the DNA is digested with the restriction enzyme between the cloning site and *Bam*HI site again. The original intact DNA is linearized, whereas the deleted subclone is not. Therefore the background is decreased to an undetectable level. This DNA sequencing strategy was tested on a 1.4-kb cDNA fragment containing the haptoglobin-related sequences. It is not necessary to purify large amounts of RF DNA (500 ng is enough) to get enough subclones. A set of subclones was produced in 1 day and the yield of plaques was about sixfold higher than that published.

KEY WORDS: subcloning; sequencing strategy; M13; *Sau*3AI–*Bam*HI; self-ligation.

DNA sequencing has become a major tool for the analysis of genes. Reliable DNA sequencing systems (1,2) that allow the determination of a fragment of DNA up to about 400 nucleotides (3) are available. The major problem encountered in sequencing is the generation of a subclone bank. Several improved methods are currently used for obtaining overlapping clones such as overlapping restriction fragments, the DNase I method (4–6) or the sonication method (7), and the exonuclease III method (8–10) which increase the speed of generating DNA sequence data by Sanger's dideoxy chain-termination method using single-stranded bacteriophage M13 (11). The shotgun approach (12), as well as other random subcloning procedures (7), initially accumulates DNA sequence information rapidly; however, it is difficult to finish small gaps within continuous sequences. Several systematic subcloning strategies have been reported. Some require time-consuming gel fractionation (5,6,13,14), some require special restriction sites and more manipulations (8,15), while some require delicate quantitation of DNase I and reduction of yield by blunt end ligation (4,5). In this paper, we describe a systematic DNA sequencing strategy derived from Lin's procedure (4) that eliminates the fill in and blunt end ligation step, increases the transfectant yield, and reduces the time required to generate subclones.

MATERIALS AND METHODS

Construction of Original Clones

A 1.8-kb cDNA containing the haptoglobin-related gene from rat liver (Lou strain) was cloned into the *Eco*RI site of pGEM4 (16). DNA from minilysate was prepared as described (17). Plasmid DNA was digested with *Eco*RI and *Bam*HI, and two fragments of 1.4 and 0.4 kb were obtained. The larger 1.4-kb fragment was ligated to *Eco*RI- and *Bam*HI-digested bacteriophage M13mp18

[1] To whom all correspondence should be addressed.

0003-2697/88 $3.00

(18) RF DNA and transfected into *Escherichia coli* strain JM101 (11,19). RF DNA was prepared from a 5-ml culture.

Construction of Systematic Deletion Subclones

(a) BamHI and Sau3AI treatments. RF DNA (500 ng) was suspended in 25 μl of buffer (150 mM NaCl, 10 mM Tris, pH 7.5, 10 mM $MgCl_2$) and digested with *Bam*HI (20 units) for 2 h. The reaction was stopped by incubating at 65°C for 10 min. Twenty-four microliters of H_2O and 1 μl of 50× buffer (250 mM Tris, pH 7.5, 250 mM $MgCl_2$, 0.025 mM DTT[2]) were added to the vial as DNA mixture. *Sau*3AI (Toyobo) was serially diluted to concentrations of 0.03, 0.06, 0.125, and 0.25 U/ul with medium salt (50 mM NaCl, 10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mm DTT). Ten microliters of DNA mixture was added to each vial with 2 μl of diluted *Sau*3AI (final *Sau*3AI concentrations: 0.003, 0.005, 0.01, and 0.02 U/μl) and incubated at 37°C for 5 min. The reaction was stopped by incubating at 65°C for 10 min.

(b) PEG precipitation. The DNA was precipitated with an equal volume of PEG solution (20% PEG 6000, 1.5 M NaCl) at 4°C for 1.5 h. After centrifugation at 4°C for 10 min (top speed, microcentrifuge), the DNA was washed with 70% ethanol twice and dried.

(c) Self-ligation and redigestion. The DNA pellet was resuspended in 15 μl ligation buffer (50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) with 2 U of T4 DNA ligase and incubated for 4–16 h at 14°C. The reaction was stopped by heating at 68°C for 10 min and digested with the enzyme which was between the cloning site and *Bam*HI site in cloning vector in 50 μl reaction volume for 2 h at 37°C. Four microliters of the sample was transfected into 0.2 ml of competent *E. coli* strain JM101 and plated as described (11,19), while IPTG and X-gal were not needed.

[2] Abbreviations used: DTT, dithiothreitol; PEG, polyethylene glycol.

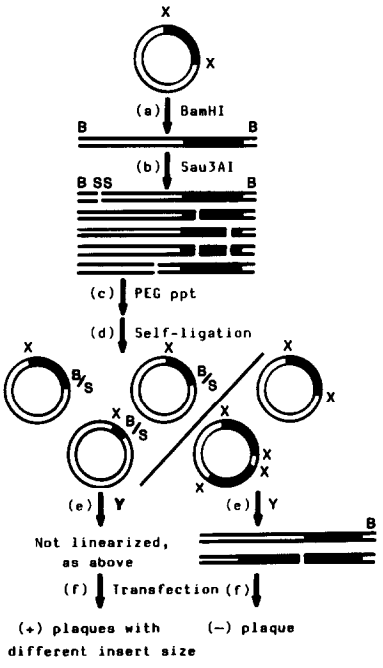

FIG. 1. Outline of sequencing strategy. X represents different restriction endonuclease recognition sites. Solid lines indicate the insert DNA; open lines indicate the M13 DNA. B, *Bam*HI; S, *Sau*3AI; S/B, altered *Sau*3AI/*Bam*HI site. Y represents the restriction site located between the cloning site and *Bam*HI site in the polylinker of the cloning vector.

(d) Characterization of subclones. Approximately 150–200 plaques per plate were obtained. Plaque was picked and grown in 3 ml of 2YT broth supplemented with JM109 bacteria for 6–8 h as described (11). Bacteriophages were harvested and RF DNA was prepared as previously described. The size of each deletion was determined on a 1.5% agarose gel after digestion with *Eco*RI plus *Pst*I or *Hin*dIII, or the culture supernatant was determined on a 0.7% agarose gel by direct gel electrophoresis (21) for 5 h at 100 V.

RESULTS AND DISCUSSION

Sau3AI Digestion

The frequent existence of *Sau*3AI sites in DNA was employed for the development of this methodology. The strategy of systematic subcloning is outlined in Fig. 1. The appro-

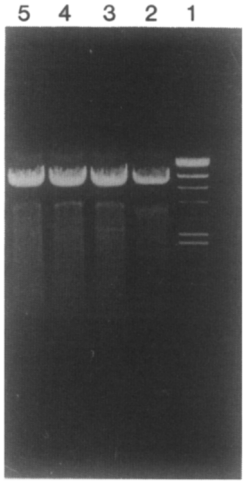

FIG. 2. *Sau*3AI treatment of *Bam*HI-digested M13 RF DNA. Lane 1, *Hin*dIII marker. Lane 2, *Bam*HI-digested DNA. Lanes 3–5, DNA treated with 0.005, 0.01, and 0.02 U/μl of *Sau*3AI. Approximately 0.3 μg of each DNA was electrophoresed on a 0.75% agarose gel.

priate *Sau*3AI treatment can easily be determined by starting with the conditions described under Materials and Methods. Once the conditions are established, the same conditions can be used for different DNAs. We tested five different clones of M13 RF DNA with satisfactory results. An example of *Sau*3AI treatment is shown in Fig. 2. The samples digested with 0.003, 0.005, and 0.01 U/μl of *Sau*3AI (final concn) were pooled. The decision as to which sample was used for further manipulation was based on how

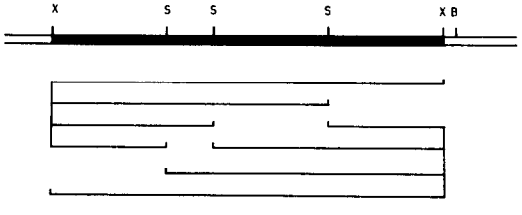

FIG. 3. Cartoon depicting the distribution of the deletion endpoints. The solid bar indicates the insert DNA. The open bar indicates the M13 DNA. The remaining portion of each subclone is indicated by the horizontal lines and the deletion endpoints are indicated by the vertical lines. X may be *Eco*RI, *Sst*I, *Kpn*I, or *Sma*I.

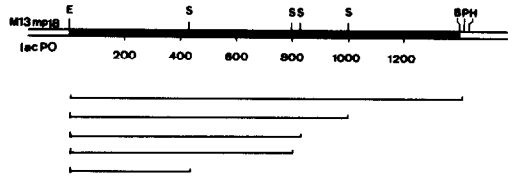

FIG. 4. Distribution of the deletion endpoints. The solid bar indicates the insert DNA. The open bar indicates the M13 DNA. The remaining portion of each subclone is indicated by the horizontal lines and the deletion endpoints are indicated by the vertical lines. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sau*3AI.

much DNA was removed by *Sau*3AI followed by *Bam*HI treatment. *Sau*3AI double digestion within the insert is satisfactory. Complete *Sau*3AI subclones can be obtained. In contrast, some erroneous interpretations could result from overdigestion of DNA by DNase I.

The DNase I method is based on proper linearization. Hong's protocol requires the separation of linearized DNA from the supercoiled and open circular DNAs by agarose gel electrophoresis. The M13mp series of bacteriophages (18) is larger than 7 kb. The efficiency of isolating DNA fragments larger than 7 kb from the agarose gel is low and the cross-contamination of different DNA fragments is a problem.

PEG Precipitation, Self-Ligation, and Redigestion

Clones containing progressively deleted insert DNA can be generated by *Sau*3AI digestion as shown in Fig. 1. M13 DNA containing insert can be precipitated by PEG (22), while small fragments removed by *Sau*3AI/*Bam*HI digestion are left in the supernatant. Another advantage of this method is the reduction of DNA loss during phenol extraction and ethanol precipitation for the small amount of DNA. It is conveniently religated by the compatibility of *Sau*3AI- and *Bam*HI-generated cohesive ends. The efficiency of cohesive end ligation by *Sau*3AI/*Bam*HI is higher than that by Klenow filled-in blunt end ligation

method described by Hong and Lin. Our yield of transfectants was also higher than that described in the previous method. The reduction of background relies on the restriction enzyme which was located between the cloning site and *Bam*HI site in the cloning vector, and that will not destroy the newly generated *Sau*3AI–*Bam*HI subclones.

Only two restriction enzymes, *Bam*HI and *Sau*3AI, are needed to obtain a set of subclones. The insert generated by *Eco*RI, *Sst*I, *Kpn*I, or *Sma*I digestion can be cloned into M13mp18, while the insert generated by *Pst*I, *Sph*I, *Hin*cII, *Acc*I, *Sal*I, *Xba*I, or *Hin*dIII digestion can be cloned into M13mp19. The *Bam*HI site of the M13 vector can be utilized to start subcloning. Two opposite directions of an insert cloned in mp18 will give rise to two sets of subclones (summarized in Fig. 3).

There are some limitations of this method: (i) *Sau*3AI fragment must be within 0.4 kb to get a complete sequence. (ii) The clone of interest has one *Bam*HI site that gives rise to the double-digested fragment of X-*Bam*HI (i.e., *Eco*RI/*Bam*HI, *Sst*I/*Bam*HI, *Kpn*I/*Bam*HI, or *Sma*I/*Bam*HI) cloned into M13mp18. Then only one orientation will be obtained (Fig. 4). The first limitation can be overcome by using both Hong's method (5) and the method reported in this paper. If the fragment has more than one *Bam*HI site, other methods will be useful. However, in most cases the ease and speed of this procedure may make it simpler to get sequence data.

Characterization of Subclones

The most efficient method to screen the subclones is digestion with restriction endonuclease followed by analysis on agarose gels. We prepared RF DNAs by the alkaline lysis method from the same tubes used for the preparation of single-stranded DNAs. Twenty plaques were analyzed. The result (Fig. 4) demonstrated that the deletions are uniformly distributed within the insert.

Starting from 500 ng of DNA, 2500 plaques were obtained by the protocol described here, in contrast to 10,000 plaques obtained by Lin's method using 12 μg of DNA. The yield of transfectant from the improved method is approximately 6-fold higher than that from Lin's protocol and 360-fold higher than that from Hong's original method. The procedure described here is so simple that it takes only 1 day from the *Bam*HI digestion to the transfection and plating steps.

REFERENCES

1. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
2. Maxam, A. M., and Gilbert, W. (1980) *in* Methods in Enzymology (Grossman, L., and Moldave, K., Eds.), Vol. 65, pp. 499–560, Academic Press, New York.
3. Sanger, F., and Coulson, A. R. (1978) *FEBS Lett.* **87,** 107–110.
4. Hun-Chi, L., Lei, S. P., and Wilcox, G. (1985) *Anal. Biochem.* **147,** 114–119.
5. Hong, G. F. (1982) *J. Mol. Biol.* **158,** 539–549.
6. Frischauf, A. M., Garoff, and Lehrach, H. (1980) *Nucleic Acids Res.* **8,** 5541–5549.
7. Deininger, P. L. (1983) *Anal. Biochem.* **129,** 216–223.
8. Guo, L. H., and Wu, R. (1982) *Nucleic Acids Res.* **10,** 2065–2084.
9. Thoma, W. O. (1985) *Anal. Biochem.* **144,** 207–211.
10. Heniboff, S. (1987) *in* Methods in Enzymology (Wu, R., Ed.), Vol. 155, pp. 156–165, Academic Press, Orlando, FL.
11. Messing, J., Crea, R., and Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
12. Anderson, S. (1981) *Nucleic Acids Res.* **9,** 4087–4098.
13. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., and Surry, S. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4298–4302.
14. Laughon, A., and Scott, M. P. (1984) *Nature (London)* **310,** 25–31.
15. Henikoff, S. (1984) *Gene* **28,** 351–359.
16. Melton, D. A., *et al.* (1984) *Nucleic Acids Res.* **12,** 7035–7056.
17. Ish-Horowicz, D., and Burke, J. F. (1981) *Nucleic Acids Res.* **9,** 2989–2998.
18. Messing, J., and Vieira, J. (1982) *Gene* **19,** 269–276.
19. Yanisch-Perron, *et al.* (1985) *Gene* **33,** 103–119.
20. Messing, J. (1983) *in* Methods in Enzymology (Grossman, L., and Molkave, K., Eds.), Vol. 101, pp. 20–78, Academic Press, New York.
21. Lis, J. T. (1980) *in* Methods in Enzymology (Grossman, L., and Molkave, K., Eds.), Vol. 65, pp. 347–353, Academic Press, New York.