A new method for specific cleavage of megabase-size chromosomal DNA by λ -terminase

Yixin Wang and Ray Wu*

Section of Biochemistry, Molecular and Cellular Biology, Cornell University, Ithaca, NY 14853, USA

Received January 19, 1993; Revised and Accepted March 22, 1993

ABSTRACT

The development of methods for cleavage of DNA at specific site(s) that are widely spaced would facilitate physical mapping of large genomes. Several methods for rare and specific cleavage of chromosomal DNAs require a nearly complete methylation of a given type of restriction site except the one that is specifically protected. It is expected that as the target DNA increases in length, it will become less likely to achieve nearly complete methylation. The intron-encoded endonucleases may also provide a capability to cleave megabase-sized DNA segments due to their very large recognition sequences. However, there are endogenous cleavage sites in the chromosomes of most organisms. We present here a new method to specifically cleave intact chromosomal DNA using λ -terminase. A plasmid containing two specific cleavage sites (cohesive-end sites) for λ -terminase was specifically introduced into the *E. coli* genome and into chromosome V of *S. cerevisiae*. Chromosomal DNA was prepared from the resulting strains, and then cleaved with λ -terminase. The results showed that the 4.7-megabase pair (Mb) circular *E. coli* chromosome and the 0.58-Mb linear yeast chromosome V were specifically cleaved at the desired sites with very high efficiencies. The approach of using the λ -terminase cleavage reaction is a simple one-step procedure with a high specificity which is particularly suitable for mapping very large genomes of eucaryotes.

INTRODUCTION

Construction of a contiguous, low density physical map of large genomes requires new methods for cleaving megabase-size chromosomal DNA specifically into two large fragments. Two enzymatic approaches for specific cleavage of large DNA molecules have been reported. The first approach involves the use of enzymes which have large recognition sites (1). Several intron-encoded endonucleases have been characterized and tested with this respect (2, 3, 4, 5). The hallmark of these endonucleases is the large size of their recognition sequences, which range from 15 to 39 base pairs (1). Some of them have been produced in either *E. coli* or a yeast expression system, and have been used

to cut large chromosomal DNA. For example, Dujon and co-workers have been able to cleave yeast and bacteriophage T7 genomes by using the endonuclease I-Sce I (6, 7). In the second approach, two enzymes are required: a DNA methyltransferase (MTase) that methylates all the sites of a given type of restriction enzyme in an entire genome, and a corresponding restriction endonuclease to cleave the single restriction site left unmethylated by protection with either a sequence-specific DNA-binding protein or a formation of DNA triple helix. The success of the method depends on the completeness of the methylation reaction followed by specific cleavage of genomic DNA. This approach was first introduced by Koob and Szybalski (8, 9), in which a *lac* operator sequence that contained a Hae II site was integrated into one of the *S. cerevisiae* chromosomes and the *E. coli* chromosome. The chromosomal DNA was incubated in the presence of *lac* repressor, and then enzymatically methylated. After removal of *lac* repressor, Hae II was used to cleave the DNA at the single *lac* repressor protected site to produce large fragments of predicted size. Using this method, they were able to cleave the 4.7-megabase pair (Mb) *E. coli* chromosome and the 0.58 Mb chromosome V of *S. cerevisiae*. Using a similar strategy, Strobel and Dervan (10) made use of a synthetic homopyrimidine-homopurine tract to form triple helical structures after incorporating the sequence, together with an EcoRI restriction site, into a 0.34 Mb yeast chromosome. Triple helix formation, followed by MTase treatment, triple helix disruption and EcoRI digestion of the chromosomal DNA, generated two expected fragments of 0.11 Mb and 0.23 Mb. They also liberated a segment from the tip of human chromosome IV (11). More recently, Ferrin and Camerini-Otero, Szybalski and colleagues improved the method by avoiding genetic intervention of the genome required previously (12, 13, 14). They took advantage of the ability of *E. coli* RecA protein to pair a single-stranded oligonucleotide to its homologous sequence in duplex genomic DNA to form a three-stranded complex. This complex protected a specific adjacent restriction site from methylation by the MTase. After methylation of all other restriction sites in the genome and removal of RecA protein, the restriction enzyme cleavage was limited to the site previously protected. Szybalski and colleagues also reported a similar method by using integration host factor (IHF) to mediate protection against methylation (15). It is expected that the success of this second approach requires the

* To whom correspondence should be addressed

nearly complete methylation of a given type of restriction site except at the site that is protected by *lac* repressor, triple helical structure, RecA-mediated complex or IHF, respectively.

In this report, we present a new method to specifically cleave intact chromosomal DNA into very large fragments. To test this method, we have introduced a sequence that contains the cohesive-end sites (*cos* sites) (Fig. 1) of bacteriophage λ into chromosome V of *S.cerevisiae* and the single *E.coli* chromosome. We then used λ -terminase to specifically cut the chromosomal DNA at the *cos* sites. This method allows a very specific and efficient cleavage because there are no natural λ -terminase cleavage sites present in eucaryotic genomes, and it involves a simple single-step reaction. The method also overcomes the difficulty of a nearly complete methylation of a large genome because modification of a given restriction site is unnecessary. In future experiments, plants or animals with much larger genome size will be tested.

MATERIALS AND METHODS

Construction and integration of plasmids for specific cleavage of chromosomal DNA

Plasmid pY1*cos* for use in yeast was made by inserting a 2.4 kb EcoRI–BamHI fragment from pcos2EMBL (16), which contains two *cos* sites, into the yeast integration vector pYIP56. The plasmid was then linearized at a unique NcoI (Fig. 2A) site located within the *URA 3* gene of the vector to target integration at *URA 3* locus, and transformed into a *URA 3*[−] *S.cerevisiae* strain 8534-8C (17) by the lithium acetate-treated method (18). *URA 3*⁺ transformants were selected to yield 8534-8C (pY1*cos*) strain. The presence of the *cos* sequence was confirmed by Southern blot analysis.

E.coli strain JC7623 (19) was the starting *E.coli* strain. The 2.4 kb EcoRI–BamHI fragment from pcos2EMBL (16) was cloned into plasmid pRS414 (20), which has a *lac* sequence, to generate pE1*cos*. After linearization by SstI (Fig. 2B) within the *lac* operon, pE1*cos* was transformed into JC7623 and integrated into *E.coli* chromosome at the *lac* site through homologous recombination (19). The desired transformants were selected by ampicillin resistance. Next, the plasmid pE2*cos* was introduced into the same *E.coli* genome using the same approach. The *cos* sequence and a kanamycin resistant gene were inserted into plasmid pVJS101 (21) which contains a *fdn* sequence to produce pE2*cos*. The resulting construct was linearized at a unique Bcl I site (Fig. 2B) within the *fdn* operon, and then transformed into JC7623 (pE1*cos*). Transformants that were resistant to both kanamycin and ampicillin were selected. In the resulting strains with either one or two copies of the *cos* sequence, the presence of *cos* sites was confirmed by Southern blot analysis.

Preparation of chromosomal DNA

Chromosome-size DNA preparation from *S.cerevisiae* and *E.coli* were performed as described previously (22).

Specific cleavage of chromosomal DNA by λ -terminase

λ -terminase was purchased from Takara Biochemical Inc.. For chromosomal DNA cleavage, DNA in agarose plugs was transferred from ESP solution (0.5 M EDTA, pH 8.0, 1% sarkosyl, 1 mg/ml proteinase K, which was used to release DNA from chromosomes), to TEP [10 mM Tris·HCl, pH 8.0, 10 mM EDTA, pH 8.0, and 1 mM phenylmethylsulfonyl fluoride (PMSF)] buffer, and incubated at 37°C for 2 hours in order to

inactivate the proteinase K. After repeating this step, the plugs were washed with TE (10 mM Tris·HCl, pH 8.0, 1 mM EDTA) buffer twice at 37°C for 2 hours to remove PMSF. The agarose plugs were then cut into small pieces (10 mm×1 mm×2 mm), and equilibrated in the basal buffer (13 mM Tris·HCl, pH 8.0, 3 mM MgCl₂, 75 mM KCl, 5 mM spermidine, 2 mM ATP, 0.5 M EDTA, 0.01% Triton-X 100, and 6 mM mercaptoethanol) for 1 hour. Each cut-up agarose plug was digested by adding 10 units λ -terminase and incubating at 30°C for 4 hours. The cleavage products in the agarose plugs were subjected to PFGE using a CHEF-DR II apparatus (Bio-Rad). The gel was stained with ethidium bromide and visualized under UV light. Before transferred to a Nytran filter (Schleicher & Schuell) for Southern blotting, the gel was exposed to UV light (254–360 nm) for 5 minutes to break down large DNA fragments.

RESULTS

Previous studies have demonstrated that λ -terminase recognizes a sequence of about 100 bp at the *cos* site region of bacteriophage λ , and cleaves DNA at the *cos* site to generate a 5'-protruding end with 12 bases (23, 24, 25, 26) (Fig. 1). λ -terminase has been used in linearizing cosmid clones for restriction mapping (27) and in dissecting *E.coli* chromosome at endogenous cryptic *cos* sites (28). However, the cutting efficiency is variable because λ -terminase is a large protein complex that possesses several biochemical activities (23, 24, 25). For the work described in this report, we optimized the cleavage reaction of λ -terminase to obtain nearly complete cleavage of yeast and *E.coli* chromosome-size DNA after a 4-hr digestion.

To present experimental evidence of the usefulness of λ -terminase for physically mapping and precisely cleaving chromosome-size DNA, we constructed a plasmid, pY1*cos*, which contains the sequence of two *cos* sites (Fig. 2A). We then introduced the plasmid into the *URA 3* locus on the 0.58 Mb chromosome V of *S.cerevisiae*. The chromosomal DNA was then prepared from the resulting yeast strain 8534-8C-*cos* in agarose plugs and digested with λ -terminase. Finally, the cleavage

Figure 1. Structure of the *cos* region of bacteriophage λ . The top part represents a complete *cos* site, which is about 200 bp, predominantly from the left end of λ DNA. It contains three R sequences (R1 to R3) for gpNu1 binding, two I sequences (I1 and I2) for IHF binding and *cosN* site for λ -terminase cleavage (23). The bottom part shows the sequence at *cosN* region. The solid box indicates the λ -terminase recognition site, and the arrows indicate the normal cleavage position.

Figure 2. Construction of plasmids for specific cleavage of chromosomal DNA. (A) Construction of pY1cos. A sequence containing two *cos* sites was inserted into pYIP56, an integration vector containing *URA3* gene, at the polylinker region. A unique *NcoI* restriction site in the *URA3* gene (shown by an arrow) was used to linearize the plasmid before transformation. *Ap^R* represents the ampicillin resistant gene. (B) Construction of pE1cos and pE2cos. The sequence with two *cos* sites was cloned into pRS414 (20) at a position next to the *lac* operon, and into pVJS101 (21) at a position next to the *fdn* operon. A kanamycin resistant gene (*Km^R*) was then introduced into pVJS *cos* at an *EcoRI* restriction site to produce pE2cos. Before transformed into *E. coli*, pE1cos was linearized at a unique *SstI* site (shown by an arrow) within the *lac* operon, and pE2cos was linearized at a unique *BclI* site (shown by an arrow) within the *fdn* operon.

Figure 3. Cleavage of DNA of yeast chromosome V. (A) A simplified schematic drawing of *S. cerevisiae* chromosome V. The position of cleavage is shown by an arrow. (B) Agarose gel after PFGE. The yeast chromosomal DNA embedded in low-melting point agarose was digested by λ -terminase (23). The treated chromosomes were analyzed by PFGE on a CHEF-DR11 system (1% agarose, 200 V, 50 sec switch time for 24 hours). After electrophoresis, the gel was stained with ethidium bromide. Lane 1 shows the yeast chromosomal DNA without λ -terminase treatment. Lane 2 shows the complete cleavage with λ -terminase of the 580 kb yeast chromosome into a 450 kb and a 130 kb fragment. (C) Southern blot analysis using the *cos* sequence probe. The gel in (B) was treated by UV light (254–360 nm) for 5 minutes before the DNA was transferred to a nytran membrane and hybridized with a *cos* sequence probe. Lanes 1 and 2 represent the hybridization of the probe to the uncleaved (580 kb) and the cleaved chromosomal DNA (450 kb and 130 kb), respectively.

products were separated by pulsed field gel electrophoresis (PFGE), and the gel blot was hybridized to a *cos* sequence probe.

Our initial attempt at sequence-specific cleavage of the *S. cerevisiae* chromosome at a single site is shown in Fig. 3. Chromosome V is 0.58 Mb in length, and the *URA3* gene is located at a position about 0.13 Mb from one end (Fig. 3A). Chromosome-size DNA from the yeast strain 8534-8C-*cos*, which had the *cos* sequence integrated at *URA3* through homologous recombination, was cleaved with λ -terminase. A complete digestion of chromosome V at the single site, and a 0.45 Mb and a 0.13 Mb fragment were observed as shown in an ethidium bromide-staining agarose gel (Fig. 3B, lane 2). Subsequent Southern blot analysis, using the *cos* sequence as the probe, confirmed that the cleavage was very specific and efficient (Fig. 3C, lane 2). To quantitate the cleavage efficiency, the X-ray film was traced with a densitometer, and the result showed that 100% of the DNA was cleaved into the desired two fragments.

To examine cleavage of a larger DNA by λ -terminase, we applied the same method to the 4.7 Mb *E. coli* genome. We constructed two plasmids, pE1cos and pE2cos, each of which contains the sequence of two *cos* sites (Fig. 2B). We then introduced the plasmids into the circular genome of *E. coli*. The *cos* sequence (pE1cos) was first integrated into the chromosome at the *lac* operon; then another copy (pE2cos) was integrated into the genome of the resulting strain at the *fdn* operon. Since the

Figure 4. Cleavage of *E. coli* chromosomal DNA. (A) A simplified schematic drawing of the *E. coli* genome. The *lac* and the *fdn* sites are shown by arrows. (B) Agarose gel after PFGE was stained with ethidium bromide that showed sequence-specific cleavage of the *E. coli* chromosome. Lane 1: The circular *E. coli* chromosome with one set of the *cos* sequences was cleaved into a 4.7 Mb linear molecule. Lane 2: The *E. coli* chromosome with two sets of the *cos* sequences was cleaved into two fragments of 3.6 Mb and 1.1 Mb in size. Lane 3: Chromosomal DNA of *S. pombe* which was used as size-markers, with their sizes indicated next to the gels. The λ -terminase manipulation was performed as described in Fig. 2. The treated chromosomal DNA were analyzed by PFGE (0.6% agarose, 50 V, 60 min switch time for 160 hours). (C) Southern blot analysis using the *cos* sequence probe. The experiment was performed as described in Fig. 2. The autoradiograph shows the hybridization of the probe to the cleaved *E. coli* chromosomal DNA.

two operons are 24 minutes apart on the genetic map, we expected that λ -terminase cleavage of the 4.7 Mb circular *E. coli* genome would yield two fragments of approximately 3.6 Mb and 1.1 Mb in length. The result is shown in Fig. 4. Two expected DNA fragments were observed with the strain containing two *cos* plasmids (Fig. 4B, lane 2), whereas a linearized *E. coli* chromosome was seen with the intermediate strain which has only a single copy of the *cos* plasmid (Fig. 4B, lane 1). Genomic Southern blotting showed that the cleavage in this reaction was nearly complete (Fig. 4C). A quantitation assay was carried out by densitometry which showed that approximately 90% of the DNA was cleaved into the desired fragments. A 3.4 Mb minor band appeared in both lanes 1 and 2, but was not observed in the agarose gel. This might represent a small amount of uncleaved superhelical *E. coli* DNA.

DISCUSSION

To produce a highly-efficient cleavage, we tested different conditions for λ -terminase digestion. Since λ -terminase carries several biochemical activities, the optimal condition could be different depending on the purpose. For instance, ATP is normally substituted by an analog, adenylyl-imidodiphosphate (AMPPNP), in the reaction to avoid undesired nicking activity of λ -terminase at the *cos* site, in order to produce the uniform cohesive ends. For genome mapping, however, we are mainly interested in maximal cleavage with little concern about nicking activity at the *cos* site. Thus, the substitution of ATP by AMPPNP may not be necessary. On the other hand, ATP is essential for a disengagement of the cohesive ends after the cleavage (23). This function of ATP is particularly important for our method because the normal heating procedure at 65°C after cleavage cannot be used to denature the annealed cohesive ends of chromosomal DNA that are embedded in agarose plugs. Based on these considerations, we used 2 mM ATP in the reaction instead of AMPPNP. This modification improved the cleavage efficiency from 40% to over 90%. Our experiments also showed that the optimal reaction time is 2–4 hours, and that reactions longer than overnight might cause minor degradation of chromosomal DNA. This was probably due to a small amount of contaminating nuclease activity in the λ -terminase preparation used in the experiments.

The method described in this report allows very efficient and specific cleavage of chromosome-size DNA at pre-determined sites. Because the cleavage is highly efficient and specific, it is particularly valuable for large genomes. Because techniques for transformation and regeneration of more than 40 varieties of higher plants and a number of animal cell lines are available (29, 30, 31), the method can be applied to many organisms as a new tool for the specific cleavage of a very large genome.

Both the RecA protein-mediated method (12, 13, 14) and the λ -terminase-mediated method have advantages and disadvantages depending on the goal of the project. If the goal is to cleave the chromosomes of a higher plant (such as rice) in only a few locations, the RecA protein-mediated method is faster and simpler because no prior manipulation of the genome is needed. However, it remains to be tested whether the method can produce DNA fragments of 5–20 megabases in size with sufficiently high yield. On the other hand, if the goal is to cleave the 12 rice chromosomes in many places (such as 100) to obtain a number of large DNA fragments to overlap different fragments, 200 kb to 800 kb in size (such as YAC clones), in order to construct

a physical map of each chromosome, the RecA protein-mediated method would be very time consuming and expensive. This is because one would need to determine the DNA sequence in 100 suitably-spaced locations on the rice chromosomes, followed by synthesis of 100 oligodeoxynucleotides, each of 30–60 residues in length. In contrast, obtaining 100 specific cleavages of the rice genome using the λ -terminase approach is relatively simple and less expensive. Once a *cos* site-containing plasmid is transformed into rice, it is not too difficult to regenerate 100 independent transgenic plants (32), each containing one or several copies of integrated *cos* site plasmid. Since in higher plants most recombination events result from illegitimate recombination rather than homologous recombination (33), the 100 transgenic rice plants are likely to contain integrated *cos* site plasmids randomly distributed on the genome. Next, the *cos* site(s) in each transgenic plant can be cleaved by the λ -terminase, and the resulting large DNA fragments can be fractionated by PFGE, followed by hybridization using known RFLP probes. Each large DNA fragment can then be used to overlap different YAC clones or several RFLP probes. We have initiated work on *Arabidopsis thaliana* (genome size 100 Mb) and *Oryza sativa* (genome size 400 Mb) to test this new method. Individual chromosomes (size range of approximately 17 to 55 Mb) from these plants with integrated copies of the *cos* site-containing plasmid are expected to produce two fragments after cleavage with λ -terminase. These large DNA fragments can then be used to generate a contiguous physical map.

ACKNOWLEDGEMENTS

We thank V. Stewart for providing the plasmid pVJS101, H. Lehrach for providing the plasmid pcos2EMBL, K. Struhl for providing the plasmid pYIP56, and B. Tye for providing the *S. cerevisiae* strain 8534–8C. We also thank E. Keller, R. Last and C. Coster for critical reading of the manuscript. This research was supported by USDA grants 9101742 and 9202209. Y. Wang was supported by a predoctoral fellowship from the Cornell Plant Science Center, which is a unit of the USDA-DEO-NSF Plant Science Centers Program.

REFERENCES

1. P. S. Perlman and R. A. Butow. (1989) *Science* **246**, 1106.
2. D. E. Muscarella, E. Ellison, V. Vogt. (1990) *Molec. Cell. Biol.* **10**, 3386.
3. L. Colleaux, L. d'Auriol, M. Betermier, G. Cottarel, A. Jacquier, F. Galibert, B. Dujon. (1986) *Cell* **44**, 521.
4. A. Delahodde, V. Goguel, A. M. Becam, F. Creusot, J. Perea, J. Banroques, C. Jacq. (1989) *Cell* **56**, 431.
5. F. Durrenberger and J. D. Rochaix. (1991) *EMBO J.* **10**, 3495.
6. A. Thierry, A. Perrin, J. Boyer, C. Fairhead, B. Dujon, B. Frey, G. Schmitz. (1991) *Nucl. Acids Res.* **19**, 189.
7. A. Thierry and B. Dujon. (1992) *Nucl. Acids Res.* **20**, 5625.
8. M. Koob, E. Grimes, W. Szybalski. (1988) *Science* **241**, 1084.
9. M. Koob and W. Szybalski. (1990) *Science* **250**, 271.
10. S. Strobel and P. Dervan. (1991) *Nature* **350**, 172.
11. S. Strobel, L. Doucette-Stamm, L. Riba, D. Housman and P. Dervan. (1991) *Science* **254**, 1639.
12. L. J. Ferrin and R. D. Camerini-Otero. (1991) *Science* **254**, 1494.
13. M. Koob, A. Burkiewicz, J. Kur, W. Szybalski. (1992) *Nucl. Acids Res.* **20**, 5831.
14. M. Koob. (1992) *Methods in Enzymol.* **216**, 321.
15. J. Kur, M. Koob, A. Burkiewicz, W. Szybalski. (1992) *Gene* **110**, 1.
16. E. Ehrlich, A. Craig, A. Poustka, A. Frischauf, H. Lehrach. (1987) *Gene* **57**, 229.
17. A. Hinnen, J. B. Hicks, G. R. Fink. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1929.

18. F. M. Ausubel, R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith and K. Struhl. (1989) *Current Protocols in Mol. Biol.* Wiley Interscience, New York.
19. S. C. Winans, S. J. Elledge, J. H. Krueger, G. C. Walker. (1985) *J. Bacteriol.* **161**, 1219.
20. R. W. Simons, F. Houman, N. Kleckner. (1987) *Gene* **53**, 85.
21. B. L. Berg and V. J. Stewart. (1990) *Genetics* **125**, 691.
22. C. L. Smith, and C. R. Cantor. (1987) *Methods. Enzymol.* **155**, 449–467.
23. R. R. Higgins, H. J. Lucko, A. Becker. (1988) *Cell* **54**, 765.
24. M. Gold and A. Becker. (1983) *J. Biol. Chem.* **258**, 14619.
25. M. Feiss. (1986) *Trends. Genet.* **2**, 100.
26. R. Wu and E. Taylor. (1971) *J. Mol. Biol.* **57**, 491.
27. H. R. Rackwitz, G. Zehetner, H. Murialdo, H. Delius, J. H. Chai, A. Poustka, A. Frischauf and H. Lehrach. (1985) *Gene* **40**, 259.
28. H. Kotani, A. Kawamura, A. Takahashi, M. Nakatsuji, N. Hiraoka, K. Nakajima and M. Takanami. (1992) *Nucleic Acids Res.* **20**, 3357.
29. C. S. Gasser and R. T. Fraley. (1988) *Science* **244**, 1293.
30. K. Shigekawa and W. J. Dower. (1988) *BioTechniques* **6**, 742.
31. M. R. Capecchi. (1989) *Science* **244**, 1288.
32. W. Zhang, D. McElroy and R. Wu. (1991) *The Plant Cell* **3**, 1155–1165.
33. M. W. Saul and I. Potrykus. (1990) *Dev. Genet.* **11**, 176.