Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 9208–9212, December 1989
Biochemistry

ExoMeth sequencing of DNA: Eliminating the need for subcloning and oligonucleotide primers

(genome project/dideoxynucleotides/exonuclease/polymerase/methylated DNA)

JOSEPH A. SORGE AND LAURA A. BLINDERMAN

Stratagene, 11099 North Torrey Pines Road, La Jolla, CA 92037

Communicated by Ernest Beutler, September 5, 1989

ABSTRACT **A method is reported for sequencing DNA based on exonuclease III digestion and strand protection by using modified nucleoside triphosphates. Up to 10 kilobases of sequence information may be obtained from each strand of a given template without subcloning. Prior knowledge of the restriction map is not important; prior knowledge of any of the sequence is not required. Nor are oligonucleotide primers needed. Double-stranded cosmids, plasmids, λ phage, or linear molecules (including amplified molecules) may be used as starting material. The method creates a single-stranded template from these starting molecules, thus generating high-quality sequence ladders. Most commonly used DNA polymerases may be utilized, including reverse transcriptase and T7 DNA polymerase. The approach is "ordered", so little time is wasted on redundant sequencing.**

The sequencing of nucleic acids can be divided into three main tasks: generation of templates, enzymatic/chemical reactions, and identification of nucleic acid fragments that have been separated according to their length. The identification step has been automated (1, 2). The process of automating the enzymatic/chemical reaction step has begun (3, 4). However, the generation of templates currently requires numerous steps that are difficult to automate, including restriction mapping, preparing subfragments for subcloning, identifying subclones, growing bacterial cultures, and purifying nucleic acids. Such template generation steps account for >80% of the effort expended in the sequencing of nucleic acids. Unfortunately, current sequencing paradigms require the generation of a new template for each 300–500 nucleotides (nt) sequenced. Under the assumption of no overlapping sequence between templates, the sequencing of both strands of an entire mammalian genome would require at least 20 million templates 300 nt each in length. A nonordered approach, such as shotgun sequencing (5), would require the generation of 100–200 million templates.

We have developed a procedure that circumvents the need for primer-binding sites and also eliminates the need to know the restriction map (Fig. 1). We call this procedure ExoMeth sequencing. ExoMeth sequencing applies the fact that exonuclease III (Exo III) can be used to remove one strand of double-stranded DNA from only one end of the molecules (6). By properly timing the extent to which Exo III is allowed to act, one obtains a relatively homogeneous DNA preparation that is double-stranded at one end but contains single-stranded DNA in that portion of the molecule where Exo III has digested. The shorter strand of the double-stranded region then serves as a primer for the synthesis of a labeled polynucleotide. With ExoMeth sequencing, the partially single-stranded DNA from each time point is exposed to chain-termination sequencing reactions in the presence of a labeled nucleoside triphosphate and a modified nucleotide, such as 5-methyldeoxycytidine triphosphate (5-Me-dCTP). The DNA is digested with a frequently cutting restriction enzyme that cannot digest the DNA that has incorporated the modified nucleotide. Thus, the labeled strands are all truncated at a fixed location, producing fragments that yield sequence information by electrophoresis (Fig. 1).

The method greatly extends the amount of sequence information that can be obtained from one template, thus substantially reducing the number of templates needed to sequence a given nucleic acid segment. Moreover, the sequence information can be obtained from many regions of the template in parallel. Lastly, the sequence information obtained simultaneously from many regions is "ordered" rather than "shotgun", thus allowing one to return directly to uninformative regions for further clarification.

MATERIALS AND METHODS

Enzymes and Nucleotides. The Klenow fragment of *E. coli* DNA polymerase I, avian reverse transcriptase, *Thermus aquaticus* (*Taq*) DNA polymerase, restriction enzymes, Exo III, and pBluescript I SK$^-$ were all obtained from Stratagene stocks that are commercially available. T7 DNA polymerase was purchased from Pharmacia. All nucleotides and buffers were used directly from Stratagene ExoMeth sequencing kit. Radionucleotides were obtained from New England Nuclear.

Restriction Digests. CsCl-purified plasmid DNA was digested to obtain an entry point for Exo III on one DNA strand only. Typically, unique *Eco*RI and *Sac* I, or *Xho* I and *Kpn* I sites in the polylinker of pBluescript SK$^-$ were used because each pair is located to one side of genes cloned unidirectionally from *Eco*RI to *Xho* I in pBluescript (pBluescript instruction manual). To determine the amount of DNA to digest, we divided the length of the insert in nt by 250, multiplied that number by 5 μg, and then added 10 μg to the product. Because it is crucial that the restriction enzyme being used to create a 3′ overhang digests the DNA to completion and thereby protects one strand of DNA from digestion by Exo III (enzyme A in Fig. 1), the DNA was digested with this enzyme first. Electrophoresis was used to confirm complete digestion. The DNA was then digested with the second restriction enzyme (enzyme B in Fig. 1). The reaction mixture was then heated to 65°C for 20 min. The DNA was ethanol precipitated, and the pellet was washed with 70% (vol/vol) ethanol, lyophilized, and resuspended in H_2O to a concentration of 1 μg/μl.

Exo III Deletions. DNA was subjected to Exo III digestion in Exo III buffer (50 mM Tris, pH 8/10 mM $MgCl_2$/10 mM 2-mercaptoethanol) at a DNA concentration of 200 μg/ml. One microgram of DNA was removed before addition of Exo III and saved for subsequent gel analysis. The reaction tube was equilibrated for 5 min at either 22°C (60 bases per min

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Exo III, *Escherichia coli* exonuclease III; 5-Me-dCTP, 5-methyldeoxycytidine triphosphate; nt, nucleotide(s).

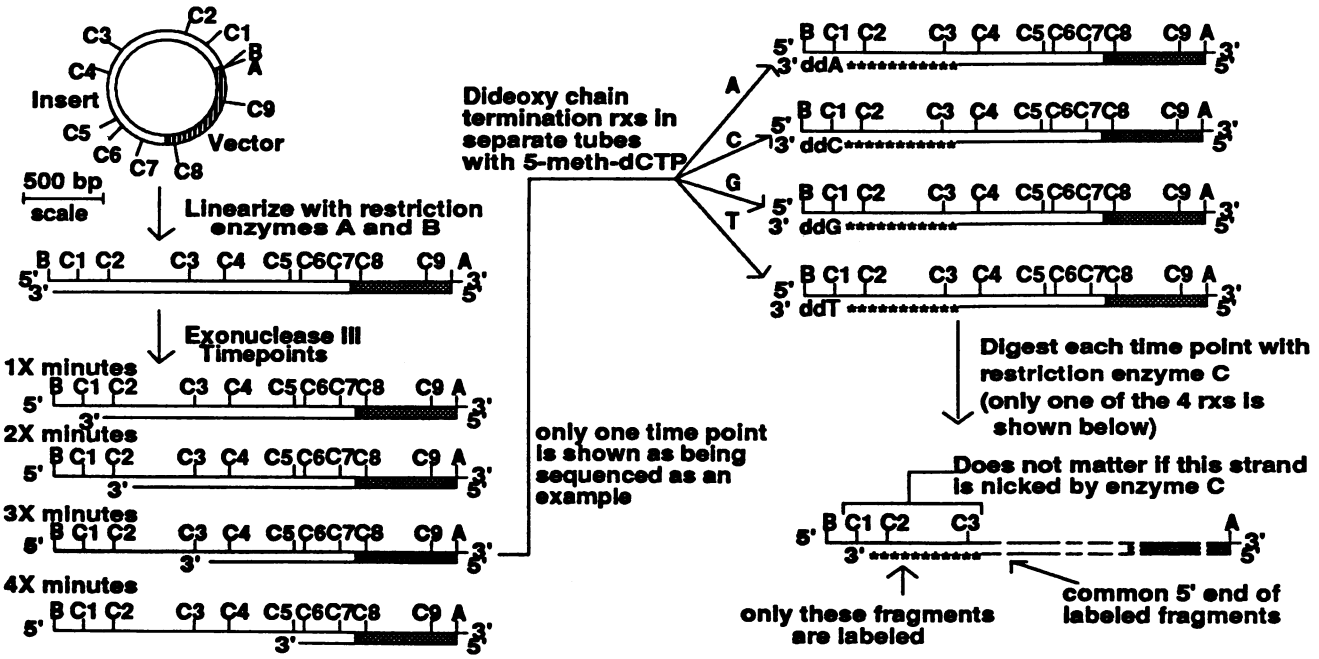

FIG. 1. A schematic diagram of the ExoMeth sequencing strategy. A, B, and C1–C9 represent restriction sites. Site A is protected from Exo III digestion, whereas site B allows Exo III to digest from the 3′ end. Asterisks represent newly synthesized DNA in the presence of at least one labeled nucleotide, 5-Me-dCTP, and one of four dideoxynucleoside triphosphates. rxs, reactions. Restriction sites C1–C9 are not cleaved by the corresponding restriction endonuclease when 5-Me-dCMP has been incorporated into the DNA.

deleted) or 30°C (200 bases per min deleted). Twenty units of Exo III per μg of DNA was added, and the contents were mixed well. Incubation was continued at the chosen temperature, a 25-μl aliquot (5 μg of DNA) was removed at each time point, and each aliquot was made 20 mM in EDTA. The Exo III in each aliquot was further inactivated by incubation at 65°C for 10 min. One microgram (5 μl) was removed from each time point for gel analysis. The remaining DNA was phenol/chloroform extracted, ethanol precipitated, and resuspended to 0.5 μg/μl in H_2O.

Alkaline Agarose Gel Analysis. One microgram of DNA from each time point was electrophoresed on a vertical 1.5-mm-thick 0.8% alkaline agarose gel (7). The samples were electrophoresed for 18 hr at 20 mA. The gel was stained with ethidium bromide and destained in water; the bands were then visualized under UV light.

DNA Sequencing. Nucleotide sequence was determined using the dideoxynucleotide chain-termination approach with either reverse transcriptase or T7 DNA polymerase. For reverse transcriptase reactions, 4 μg of DNA from each time point was mixed with 4 μl of [α-^{35}S]dATP (1350 Ci/mmol, 12.5 mCi/ml; 1 Ci = 37 GBq) and 10 units of avian myeloblastosis virus reverse transcriptase in a total reaction volume of 30 μl containing 12 mM Tris·Cl (pH 8.3), 46 mM NaCl, 12 mM $MgCl_2$, and 8.3 mM dithiothreitol. Aliquots, 6 μl each, were immediately added to microcentrifuge tubes containing 4 μl of the appropriate adenine, cytosine, guanine, or thymine mix. Each mix contained 0.5 mM dGTP, dTTP, and 5-Me-dCTP. The adenine, cytosine, guanine, and thymine mixes also contained, respectively, 0.001 mM ddATP, 0.25 mM ddCTP, 0.125 mM ddGTP, or 0.5 mM ddTTP. The reaction mixtures were incubated for 20 min at 42°C. Then 2 μl of chase mixture (0.25 mM dGTP, nonradioactive [α-S]dATP, dTTP, and 5-Me-dCTP) was added to each tube, and the mixtures were incubated for an additional 15 min at 42°C and placed on ice.

For the T7 DNA polymerase reactions, 4 μg of DNA from each time point was incubated with 2 μl of [α-^{35}S]dATP (1350 Ci/mmol, 12.5 mCi/ml), 4 μl of labeling mix (2 μM dGTP, dTTP, and 5-Me-dCTP), and 6 units of unmodified T7 DNA polymerase in a 40-μl total vol containing 28 mM Tris·Cl (pH 7.5), 10 mM $MgCl_2$, 35 mM NaCl, and 15 mM dithiothreitol for 5 min at 22°C. Aliquots, 9 μl each, of this reaction mixture were then added to 5 μl of the adenine, cytosine, guanine, and thymine termination mixes that were prewarmed to 37°C. Each termination mix contained 150 μM nonradioactive (α-thio)dATP, 5-Me-dCTP, dGTP, and dTTP. The adenine, cytosine, guanine, and thymine mixes also contained, respectively, 15 μM ddATP, ddCTP, ddGTP, or ddTTP. After incubation at 37°C for 5 min, the tubes were placed on ice.

After the sequencing reactions were cooled on ice, a 7-μl aliquot was frozen for later use, and 2–4 units of restriction enzyme C were added directly to the remaining mixture. Restriction enzyme C, as described in *Results*, must be incapable of digesting DNA containing 5-Me-dCTP. Moreover, its recognition sequence is typically 4 to 5 nt in length and preferably close to 50% guanine:cytosine and 50% adenine:thymine. If necessary, the NaCl concentration was adjusted to suit the restriction enzyme. Incubation was continued for 45 min at 37°C. Formamide loading dye, 7 μl, was added, samples were heated at 90°C for 2 min, and samples were then cooled on ice, and 3.5 μl was electrophoresed through a 6% acrylamide/urea sequencing gel at 55 W. The dried gel was exposed to Kodak XAR-2 film at room temperature overnight.

RESULTS

We used Exo III for differing lengths of time to produce a set of single-stranded deletions in a pBluescript plasmid carrying a mouse actin cDNA between the unique *Eco*RI and *Xho* I sites. The plasmid was digested with *Eco*RI and *Sac* I or *Xho* I and *Kpn* I because these two sets of restriction enzymes each cleaved on one side of the insert (Fig. 1). Enzyme A (*Sac* I/*Kpn* I) was chosen to leave a 3′ overhang. Enzyme B (*Eco*RI/*Xho* I) was chosen to leave a 5′ overhang. (The cleavage site of enzyme B must lie between the cleavage site

of enzyme A and the gene to be sequenced so that the Exo III deletions proceed into the gene.)

The doubly digested DNA was then subjected to Exo III digestion for different lengths of time by removing aliquots from the reaction vessel periodically. The resulting DNA fragments are shown in Fig. 2 in which the lower bands in lanes 2–6 represent the deleted strand. Note that even in the 20-min time point (lane 6), in which ≈4000 nt had been deleted, the thickness of the band representing the deleted strand was <300 nt. We have seen that most of the deletion lengths are ± 4% of the mean deletion size. We will refer to the deviation in deletion sizes, as observed on a gel, at any given time point as the *range* of endpoints.

The deleted DNA from the various time points was then subjected to modified dideoxynucleotide sequencing reactions and digested to completion with restriction enzyme C. Because we did not have knowledge of the restriction map of the insert, enzyme C was selected on the following basis: it must be inhibited from cutting the modified strand by 5-Me-dCTP, and it should have a recognition sequence that digests the target DNA approximately once every 250–500 nt. This typically means an enzyme with a 4- or 5-base recognition sequence. The CpG and TpA content of the target DNA can vary significantly and dramatically influence the frequency of recognition sites for enzymes the recognition sequences of which are either all guanine:cytosine or all adenine:thymine (8). Therefore, enzymes having recognition sites with an equal representation of adenine, cytosine, guanine, and thymine are preferable, such as *Sau*3AI or *Dde* I.

Because the purpose of adding 5-Me-dCTP to the reaction mixture was to inhibit the digestion of the labeled region of the labeled strand by restriction enzyme C, we tested various enzymes for inability to digest DNA modified by incorporation of 5-Me-dCTP. To date, we have found 26 enzymes suitable for this purpose: *Acc* I, *Ava* II, *Bal* I, *Bam*HI, *Ban* II, *Bsp*1286I, *Bst*XI, *Dde* I, *Eco*RII, *Hae* II, *Hae* III, *Hgi*AI, *Hin*dIII, *Hpa* II, *Hph* I, *Nar* I, *Nci* I, *Not* I, *Pst* I, *Pvu* II, *Sac* I, *Sau*3AI, *Sau*96I, *Scr*FI, *Sma* I, and *Xho* I. As a primary enzyme, when no map information is known, we prefer *Sau*3A, *Dde* I, or *Nci* I because we have obtained reliable sequence information with these enzymes. *Nci* I cuts less frequently than *Sau*3A and *Dde* I and is used when the Exo III range begins to increase in size (see below). We have tried other modified nucleotides in place of or in conjunction with 5-Me-dCTP. For example, nonradioactive (α-thio)nucleotides can be used in the sequencing reactions. The use of such nucleotides somewhat increases the number of restriction enzymes C that can be used with this method. However, we have found that use of all four thionucleotides is somewhat inhibitory to the polymerase reaction, and the added expense of using the thionucleotides is not justified by the relatively small number of additional restriction enzymes C that can be used.

The data presented in Fig. 3 were generated using [α-^{35}S]dATP to label the DNA during polymerization. The sequencing reactions were performed with 5-Me-dCTP instead of dCTP in all lanes. We tried the Klenow fragment of *E. coli* DNA polymerase I, *Taq* DNA polymerase, avian reverse transcriptase, and T7 DNA polymerase for the sequencing reactions. The latter two enzymes yielded the highest quality sequence information (Fig. 3). Typically, up to 300–400 nt could be read from each informative time point. Moreover, we have found that the incorporation of 5-Me-dCTP does not adversely affect the quality of the sequence ladder and actually decreases the prevalence of some sequencing artifacts.

It should be emphasized that the highly interpretable sequence pattern seen in Fig. 3 was only found in about half the time points. In the other time points, two main situations were encountered. (*i*) If the closest enzyme C site is >400 nt in the 5′ direction of the 3′ deleted end of a given deletion endpoint (for example, the 2X-min time point in Fig. 1), all labeled DNA fragments will be >400 nt, and the sequence information will be too large to resolve well with a conventional gel.

(*ii*) If the range of Exo III deletion endpoints at a given time point encompasses a particular enzyme site C, which for argument purposes we shall refer to as site C1, then two populations of labeled DNA strands will be generated. A population of smaller strands will result from digestion by restriction enzyme C at site C1 in the subpopulation of molecules that lack methylcytosine due to Exo III having not proceeded beyond site C1. This population of labeled DNA strands will all share the same 5′ end at site C1. Another population of larger labeled strands will result from inhibition of digestion at the C1 sites that have incorporated methylcytosine. The common 5′ end of such molecules will be created by cleavage at site C2, which is 5′ to site C1 on the Exo III-deleted strand (Fig. 1). Because both populations of

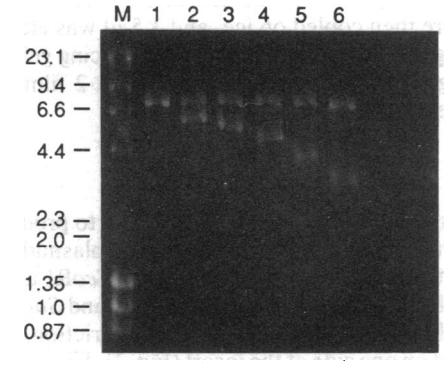

FIG. 2. Alkaline agarose gel stained with ethidium bromide of DNA fragments deleted with Exo III. M, kb markers; 1–6, time points.

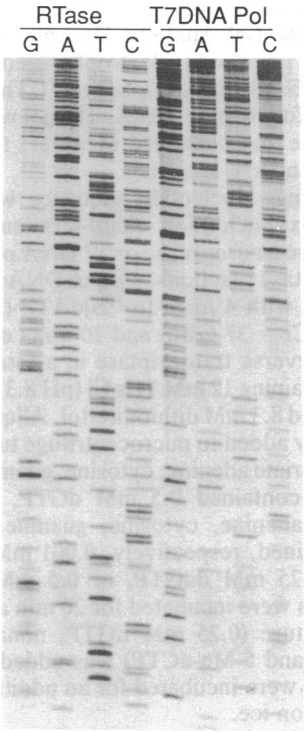

FIG. 3. Sequence ladders comparing reverse transcriptase (RTase) and T7 DNA polymerase.

labeled fragments will be in the same reaction mixture, electrophoresis may yield two superimposed sequence ladders depending on how far restriction enzyme site C2 lies from site C1. If site C2 is N nucleotides from site C1, then fragments from 1 to $\approx N$-range/2 nt in length, having their 5′ end created by cleavage at site C1, will be readable as a single sequence ladder. Fragments greater than $\approx N$-range/2 nt in length will appear as two superimposed sequences.

The superimposed sequence problem becomes more prevalent as the size of the Exo III range increases. The size of the range increases as the size of the Exo III deletions increases. This is currently the limiting factor in applying the ExoMeth procedure to stretches of DNA >5–10 kilobases (kb). As size of the range increases, we generally use less frequently digesting enzymes for enzyme C. For example, *Nci* I digests approximately every 500 nt and yields fewer superimposed sequence ladders. The disadvantage in using *Nci* I, however, is that the sequence ladders are frequently in the higher molecular weight region of the gels. Thus, one must assess the data and use less frequent enzymes as the longer Exo III time points begin to yield superimposed sequence ladders >50% of the time.

It should be noted that even though in the first round of sequencing with restriction enzyme C only 50% of the time points yield substantial stretches of sequence information, the information gained is extremely useful for subsequently obtaining sequence information from the noninformative time points. For example, when sequence information is gained from a 10-min Exo III time point but no information is gained from an 8-min time point because the closest site C in the 5′ direction is >400 nt from the 8-min 3′ endpoints, the sequence information gained from the 10-min time point can be used to select a new restriction enzyme D more appropriate for the 8-min time point. In choosing enzyme D from the sequence information that was obtained from the 10-min time point, one looks for a site closest to the deletion endpoint of the 8-min time point. One also attempts to choose an enzyme D with a 5- or 6-base recognition sequence to decrease the probability that enzyme D will cleave again within the Exo III range of the 8-min deletion endpoints. Moreover, if enzyme D has a 6-base recognition sequence, its inhibition by methylcytosine is less important because the likelihood of it occurring again within the labeled DNA region is rather low.

A substantial amount of labor can be saved if one performs the original dideoxynucleotide reactions with a 2- or 3-fold excess of template and freezes the balance of the reaction mixture after removing an aliquot for digestion with enzyme C. Once some sequence information is obtained and secondary restriction enzymes D, E, F, G, etc. are identified for the secondary sequencing of noninformative time points, one merely needs to thaw the balance of the dideoxynucleotide reaction mixtures and add the appropriate restriction enzyme. After a 45-min restriction digest the mixture is applied to a sequencing gel. Thus, the work involved in obtaining sequence information from the originally uninformative time points is minimal.

To illustrate that the ExoMeth approach is useful and accurate in determining a novel nucleotide sequence, we compared the nucleotide sequence of the mouse smooth muscle actin cDNA with a previously published mouse skeletal muscle actin cDNA sequence (9). The only differences observed between ours and the published sequence were a thymine for a cytosine at nt 785 and a cytosine for a thymine at nt 938. We do not know whether these nucleotide differences are due to the fact that the cDNAs were isolated from mRNAs prepared from different tissues (smooth versus skeletal muscle) or whether they represent silent sequence polymorphisms (because there were no amino acid changes). However, we are reasonably certain that they do not represent errors in our sequence because we observed the appropriate complementary nucleotide on the complementary strand and examined each region several times.

DISCUSSION

Most existing sequencing strategies use a known primer-binding site, either for initiation of DNA synthesis (10) or for hybridization to different-length DNA fragments having a common end (11). Guo and Wu developed a method that eliminates the need for a primer-binding site using Exo III to remove varying amounts of the 3′ ends of a linear double-stranded DNA fragment (6, 12). However, the method requires knowledge of the restriction map of the DNA to be sequenced.

The ExoMeth procedure has several major advantages. (*i*) It eliminates the need to subclone DNA fragments adjacent to a primer (or hybridization probe)-binding site in the vector. This fact eliminates the need to generate a new subclone for each 300–500 nt to be sequenced. (*ii*) Sequence information can be obtained from many regions of the template in parallel. (*iii*) No prior knowledge of the restriction map is needed. Thus clones can be taken directly from a primary library and sequenced. (*iv*) The method allows one to know from which region of the template each stretch of sequence information is derived. Thus, if information is not obtained from certain regions initially, the "ordered" nature of the data allows one to go back to the troublesome regions and obtain additional information. (*v*) The size of the recombinant template molecules is not an issue, provided unique restriction sites are available for the initiation of unilateral Exo III digestion. We have obtained quality sequence information from templates as large as 45-kb λ and cosmid molecules. (*vi*) The method allows one to use double-stranded DNA preparations as a starting material; yet the method actually synthesizes labeled fragments from a single-stranded template. Sequence ladders produced from single-stranded templates, in general, are more clear and easy to interpret than ladders produced from double-stranded templates, as evidenced by the data obtained from the ExoMeth approach.

The ExoMeth approach currently has two limitations. Because the Exo III endpoints at any given time point vary by ±4% of the number of nucleotides deleted (the range), one cannot use a restriction enzyme C that digests the DNA substantially more frequently than range^{-1}. As one begins to delete >5000 nt with Exo III, the range exceeds 400 nt. Thus, restriction enzyme C can no longer have a 4-base recognition sequence. By choosing a restriction enzyme C that digests 1/500 nt, such as *Nci* I, one begins to find more and more time points in which the DNA sequence ladder is too large to resolve on an acrylamide gel. The method is still useful because 25–30% of time points will be informative in the initial round of sequencing, and the gaps can be filled in by using sequence information from adjacent time points or from sequencing time points from the opposite DNA strand. However, to extend ExoMeth sequencing to stretches of DNA exceeding 10,000 nt in each strand, one must control the size of the Exo III range as a function of deletion size. We currently are examining an approach that would use DNA-binding proteins having 6-base recognition sequences to block Exo III from passing the binding protein recognition site. This, in effect, would compress the range to zero at particular nodes in the template.

Even with the current practical limit of 10,000–20,000 new nt per template (5,000–10,000 per strand), the need for subcloning inserts from cosmids to plasmids can be avoided when the primary library contains clones with significant overlap. If each primary clone contains no more than 10,000 unique nt not represented in other clones in the library, the sequencing of 10,000 nt from each end of each clone will

produce sequence information from both strands of the cloned region. λ phage and cosmid recombinant libraries in which the primary library has a 5-fold overrepresentation of the target genome (such as a 1 million-member λ or a 500,000-member cosmid library of mammalian genomic DNA) have at least this much overlap. Calculations have shown that ≈20-fold overrepresentation will be needed in genomic libraries to ensure the linkage of large contiguous sequences (13). In such libraries, the sequencing of 2500 nt from each end of ordered, linked clones will likely produce a sequence of most of the cloned genome.

The second limitation with the ExoMeth approach is that the vector must be digested at two rather unique restriction sites, A and B, to allow Exo III to proceed from only one starting site and in only one direction. This requires plasmid and cosmid vectors that contain pairs of rare restriction sites on either side of the cloning site. However, it would be preferable to eliminate the restriction enzyme A and B steps entirely because a substantial amount of care and attention is required to ensure that enzyme A digests the template completely and that little or no DNA nicking occurs at "star sites." This can be accomplished by using DNA that has been amplified with the polymerase chain reaction, such that a DNA binding protein site incorporated into one of the amplification primers can be used to block Exo III digestion from one end of the DNA molecule.

We thank Connie Hansen for generously providing a mouse actin cDNA clone for use as a template and Ernest Beutler and William Huse for helpful discussions. This work was supported, in part, by National Institutes of Health Grant R01 GM40881-02.

1. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. D. & Hood, L. E. (1986) *Nature (London)* **321,** 674–679.
2. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. & Baumeister, K. (1987) *Science* **238,** 336–341.
3. Zimmerman, J., Voss, H., Schwager, C., Stegemann, J. & Ansorge, W. (1988) *FEBS Lett.* **233,** 432–436.
4. Frank, R., Bosserhoff, A., Boulin, C., Epstein, A., Heinrich, G. & Ashman, K. (1988) *Biotechnology* **6,** 1211–1213.
5. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
6. Guo, L.-H. & Wu, R. (1983) *Methods Enzymol.* **100,** 60–96.
7. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
8. Beutler, E., Gelbart, T., Han, J., Koziol, J. A. & Beutler, B. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 192–196.
9. Leader, D. P., Gall, I., Campbell, P. & Frischauf, A. M. (1986) *DNA* **5,** 235–238.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
11. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
12. Guo, L.-H. & Wu, R. (1982) *Nucleic Acids Res.* **10,** 2065–2084.
13. Lander, E. S. & Waterman, M. S. (1988) *Genomics* **2,** 231–239.