Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4298–4302, July 1982
Biochemistry

"Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method

(BAL-31 digestion/deletion mutant library/nuclease S1 insert sizing)

M. PONCZ, D. SOLOWIEJCZYK, M. BALLANTINE, E. SCHWARTZ, AND S. SURREY

Division of Hematology, The Children's Hospital of Philadelphia; and Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Communicated by Werner Henle, April 22, 1982

ABSTRACT **We describe a rapid "nonrandom" DNA sequence analysis procedure that facilitates the nucleotide sequence determination of large contiguous regions of DNA. The method consists of cloning a restriction endonuclease fragment of interest into bacteriophage M13 followed by construction of a series of nuclease BAL-31 deletion mutants originating from a single site in M13 that is close to the DNA insert. Determination of the size of the deletion mutant is accomplished by hybridization to a complementary single-stranded probe derived from M13 containing the total insert followed by nuclease S1 treatment. Single-stranded M13-insert DNAs of progressively smaller sizes are isolated and analyzed by using a site-specific M13 DNA primer and the dideoxy chain-termination method. In this way, analysis of the DNA sequence proceeds from one end of the total insert to the other in a nonrandom fashion due to generation of a controlled overlapping set of deletion mutants.**

Advances in DNA sequence analysis techniques have revolutionized the study of cloned gene structure (1–3). The most commonly used method for DNA sequence analysis has been the chemical degradation procedure (1). There are several disadvantages to this technique because prior knowledge of the restriction endonuclease map of the fragment is required for detailed formulation of an analysis strategy. In addition, end labeling and sequence analysis from a single 5′ or 3′ end necessitates the use of relatively large amounts of purified fragment, a radioisotope of high specific activity, and relatively long exposure times for sequence readings.

The use of bacteriophage M13 for subcloning and DNA sequence analysis offers distinct advantages in overcoming some of these difficulties. A number of recent reports describe the use of M13 for "random" DNA sequence analysis by the dideoxy chain-termination method using an M13 site-specific primer (4, 5). This bacteriophage is well suited for DNA sequence analysis by the dideoxy chain-termination method, because cloning and isolation of recombinants is rapid and single-stranded phage DNA of the "+" strand covalently linked to single-stranded insert DNA is easily isolated from the culture medium. DNA sequence analysis is then accomplished by the dideoxy chain-termination method using the single-stranded M13 (+) insert DNA and a site-specific M13 primer that hybridizes close to the insert. One limitation of this technique is that only several hundred nucleotides from the primer site can be reliably read from gels. Therefore random analysis of a large DNA fragment involves the use of several restriction enzymes to develop a series of small overlapping fragments that are subcloned into bacteriophage M13 and then analyzed. The entire DNA sequence is then assembled by matching of overlapping sequences, frequently with the aid of a computer. Another disadvantage of this random method is unnecessary redundancy in analysis; some regions may be analyzed several times before the entire DNA sequence can be assembled. Potential difficulties also arise in analyzing regions of eukaryotic DNA containing a variety of interspersed repetitive DNA elements (6–12) or regions that have internal secondary structure (i.e., "snapback" loops).

We now describe a "nonrandom" technique for DNA sequence determination that facilitates the analysis of both DNA strands. The method is based on generation of a progressive set of deletion mutants using exonuclease BAL-31. These overlapping variable-length inserts are then subcloned into phage M13 so that the deletion region is immediately next to the M13 primer site. A complete DNA sequence can be analyzed by selecting subclones of progressively smaller insert size. This method is described below using a 3.6-kilobase (kb) DNA fragment from the human genome as an example.

MATERIALS AND METHODS

Construction and Isolation of an M13mp7 Clone Containing a 3.6-kb Human Genomic 3′-β-Globin *Eco*RI Fragment. We described previously the isolation and partial characterization of a human genomic clone containing the linked δ- and β-globin genes in bacteriophage Charon 4A (11). DNA from this clone was digested with *Eco*RI and subcloned into the *Eco*RI site of M13mp7. The procedures for subcloning, growth of phage in *Escherichia coli* strain 71.18 *traD*, and isolation of recombinant clear plaques were as described (4, 5). Clear plaques were screened for the presence of the 3.6-kb *Eco*RI fragment in viral double-stranded DNA following restriction endonuclease digestion (13). One of the resultant subclones contained the 3.6-kb *Eco*RI fragment that includes the 3′-β-globin gene coding, noncoding, and flanking DNA regions. This recombinant virus (M13β3.6) was used to demonstrate the general applicability of the nonrandom DNA sequence analysis technique.

Construction of a Deletion Mutant Library with BAL-31. M13β3.6 double-stranded DNA was isolated from YT broth culture after 6 hr growth at 37°C. Viral DNA was extracted as described (13) and purified on a CsCl gradient. Double-stranded DNA was linearized with *Bgl* I, an enzyme that cleaves M13mp7 once, near the M13-insert boundary, but does not cleave the 3.6-kb insert. The 10.8-kb linearized DNA (25 μg) was digested at 37°C with 6 units of BAL-31 in 0.3 ml of 12 mM $CaCl_2$/12 mM $MgCl_2$/600 mM NaCl/20 mM Tris·HCl, pH 8.1/1 mM EDTA. Aliquots were removed each minute and placed on ice, and portions were analyzed by electrophoresis on 0.7% agarose gels to monitor the rate of BAL-31 digestion. These conditions resulted in removal of 100–200 base pairs (bp)/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: kb, kilobase(s); bp, base pair(s).

Aliquots containing deletions of the desired size ranges were pooled and extracted with phenol, and the DNAs were precipitated with ethanol, digested with *Eco*RI, and analyzed by electrophoresis on a low-melting agarose gel. Fragments of ≈3.6 kb and smaller were isolated from the gel (14).

Construction of M13 Variant (M13dR). The digested DNA fragments isolated above have one blunt end due to BAL-31 digestion and one cohesive end due to *Eco*RI digestion. A variant of M13mp7 was constructed to allow cloning of these fragments. The variant virus (M13dR) was constructed by partial *Eco*RI digestion of parental M13mp7 followed by nuclease S1 treatment to blunt the resulting *Eco*RI ends. The digest was extracted with phenol, and the virus was precipitated with ethanol, ligated, and used to transfect *E. coli*. Elimination of one or both *Eco*RI sites in M13mp7 by this treatment results in clear plaques due to a reading frame shift in the viral β-galactosidase gene caused by the nuclease S1 treatment. Several clear plaques were selected and grown. Single-stranded DNA was isolated and its sequence was determined by using the site-specific M13 primer; direct confirmation for loss of either or both *Eco*RI sites was found. The variant (M13dR) containing an intact *Eco*RI site that was originally distal to the primer site was isolated and used for subsequent cloning.

Subcloning in M13dR. M13dR was digested with *Hin*cII and *Eco*RI, and the 7-kb viral fragment was isolated after electrophoresis on a 0.7% low-melting agarose gel. This DNA fragment has one blunt end due to *Hin*cII digestion and one cohesive end due to *Eco*RI digestion. A library of recombinant viruses was constructed by ligating this fragment to the deletion fragments resulting from digestion with BAL-31 and *Eco*RI. The recombinants contain differently sized BAL-31-induced deletion inserts oriented with the blunt end immediately adjacent to the M13 primer site. After ligation, ≈75% of the resultant clear plaques contained inserts.

Assay for Insert Size. Individual recombinants from the BAL-31-induced deletion mutant library were grown in 10 ml of YT broth, and single-stranded viral DNA was isolated as described (15, 16). The yield of single-stranded DNA was adequate for insert sizing and subsequent DNA sequence analysis using the dideoxy chain-termination method.

Aliquots of single-stranded DNA from individual cultures were screened for DNA insert size after nuclease S1 digestion of hybrids formed with single-stranded DNA from M13β3.6, which contains the entire 3.6-kb insert. Since the original double-stranded 3.6-kb fragment could insert in two possible orientations in M13mp7 and only the M13 (+) strand is secreted (15–17), two complementary full-length single-stranded probes would result, each of importance as a probe in further studies. Isolation of full-length complementary probes was accomplished after hybridization and nuclease S1 treatment of single-stranded DNAs from several different plaques containing the full-length insert. After the two full-length complementary single-stranded probes were identified, the appropriate full-length single-stranded probe was used for analyzing a given deletion mutant library.

Approximately 1 μg each of appropriate single-stranded probe and deletion mutant DNA from virus in a single plaque were incubated in 10 μl of 10 mM Tris·HCl, pH 7.9/10 mM $MgCl_2$/50 mM NaCl/10 mM 2-mercaptoethanol; the contents were flame sealed in a small capillary tube. The tube was immersed in a boiling water bath for 2 min and then allowed to cool slowly to room temperature over ≈45 min. The mixture was diluted to 50 μl to contain 50 mM NaOAc (pH 4.5), 150 mM NaCl, and 0.5 mM $ZnSO_4$, and this mixture was incubated at 37°C for 30 min with 0.3 units of nuclease S1. The resultant double-stranded hybrids were sized after electrophoresis on agarose or acrylamide gel.

Fragments of the original 3.6-kb DNA arising from BAL-31 and *Eco*RI digestion can insert in only one orientation relative to the M13dR (+) strand. Controlled sequence analysis of both DNA strands can be accomplished by the generation of two mutant libraries. Two different procedures can be used to generate these complementary libraries: (*i*) each deletion mutant library can be derived from the original 3.6-kb fragment inserted in opposite orientations relative to the M13mp7 (+) strand, with linearization at the same site in M13mp7 or (*ii*) the original 3.6-kb fragment can be inserted in one orientation, and the two deletion mutant libraries can be generated by linearization and BAL-31 digestion at each end of the insert.

DNA Sequence Analysis. Single-stranded DNA from plaques whose insert sizes differed by 100–200 bp were hybridized to a 15-bp synthetic M13 primer, and the sequence of the DNA was analyzed by the dideoxy chain-termination method as described (3, 4).

Enzymes. Restriction endonucleases, exonuclease BAL-31, and nuclease S1 were from Bethesda Research Laboratories; the 15-bp M13 primer, T4 ligase, and the Klenow fragment from *E. coli* DNA polymerase I were from New England BioLabs. Buffer conditions were as described by the supplier, unless otherwise specified.

Containment. All experiments involving recombinant DNA were performed initially under P3/EK2 conditions and subsequently under conditions prescribed in the revised National

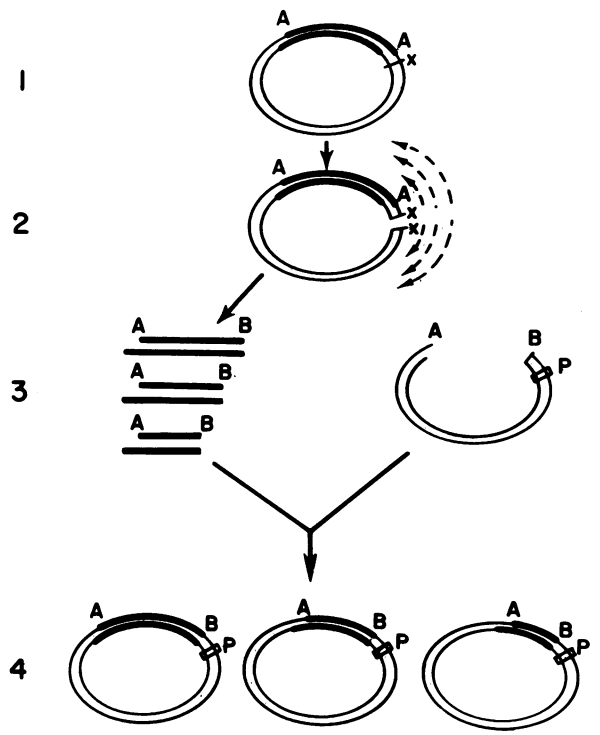

FIG. 1. Construction of the deletion mutant library. Steps: 1, the insert (thick lines) is cloned into the cohesive end A site in M13; 2, after linearization at site X, the virus is digested with BAL-31 for various times [broken lines, extent of BAL-31 digestion into insert (thick lines) and M13 (thin lines)]; 3, the digest is cleaved at site A, and the BAL-31-induced continuum of inserts is isolated, resulting in a family of differently sized fragments each of which has a BAL-31-induced blunt end B and a cohesive end A; 4, the fragments are subcloned into M13 so that blunt end B is proximal to the primer site P used for DNA sequence analysis.

Institutes of Health guidelines for recombinant DNA research (November 1980).

RESULTS

The general strategy for creating and cloning an ordered series of deletion mutants suitable for nonrandom DNA sequence analysis in M13 by the dideoxy chain-termination method is shown in Fig. 1.

The recombinant virus M13β3.6 was constructed as described above and contained the 3.6-kb *Eco*RI 3′-β-globin gene region cloned into the *Eco*RI site in M13mp7. This 10.8-kb virus was used to demonstrate the specific applicability of the nonrandom DNA sequence analysis technique. The 10.8-kb fragment was first cleaved with *Bgl* I and then digested with BAL-31, and separately timed aliquots were collected based on the predetermined rate of BAL-31 digestion (Fig. 2*a*). If BAL-31 digests in both directions (i.e., into the 3.6-kb insert and into M13mp7) with the same kinetics, this continuum of fragment sizes should include molecules that have the total 3.6-kb insert (i.e., 10.8 kb) and molecules that completely lack the insert (i.e., 10.8 − 7.2 = 3.6 kb).

Evidence supporting this conclusion is shown in Fig. 2*b*, where two distinct populations of DNA size ranges are discernable after *Eco*RI digestion of the BAL-31 digest. These two *Eco*RI-generated populations represent intact M13mp7 (7.2 kb) and its BAL-31-induced continuum of fragments, followed by the intact 3.6-kb insert and its spectrum of fragments. This procedure results in DNA fragments that have a blunt end due to BAL-31 digestion at one end and an *Eco*RI end at the other. DNA fragments of 3.6 kb and smaller were preparatively isolated from low-melting agarose and cloned into M13dR.

Aliquots of single-stranded DNA prepared from cultures derived from individual plaques were separately screened for DNA insert size after nuclease S1 digestion of hybrids formed with complementary single-stranded M13β3.6 total insert DNA (Fig. 3*a*). Nuclease S1-resistant hybrids of various sizes were found, allowing selection of sets of overlapping deletion mutants for complete (Fig. 3*B*) or targeted (Fig. 3*C*) DNA sequence analysis.

The exact extent of the deletion in a particular insert was determined directly by overlap sequence comparison with larger fragments. A direct correlation is apparent between the size of the insert measured by the nuclease S1 assay and by direct sequence analysis (Fig. 4).

The generation of an overlapping set of deletion mutants originating at the same site, whose insert size can be readily determined by the nuclease S1 assay, facilitates rapid DNA sequence analysis. We have used this methodology to determine the sequence of the entire 3.6-kb *Eco*RI insert from one DNA strand.

DISCUSSION

We have developed a nonrandom DNA sequence analysis technique in phage M13, using nuclease BAL-31 to prepare an overlapping set of deletion mutants and the dideoxy chain-termination method to determine the sequence. The technique is rapid, involving a minimal number of digestions, gel extractions, ligations, and transfections. Also, there is no need for ligation to restriction endonuclease linkers as previously described for nonrandom sequence analysis in pBR322 by the chemical degradation procedure (18).

This nonrandom technique offers other advantages over both the random sequence analysis and chemical degradation procedures. Controlled sequence analysis of both DNA strands is readily accomplished by construction of two deletion mutant libraries. This method eliminates the need to repeatedly analyze regions already analyzed and allows the analysis of long contiguous regions from eukaryotic genomes that, because of repetitive DNA content, might not be entirely suitable for random techniques involving computer-assisted sequence overlap assembly. It also eliminates the requirement for formulation of a detailed DNA sequence analysis strategy based on prior

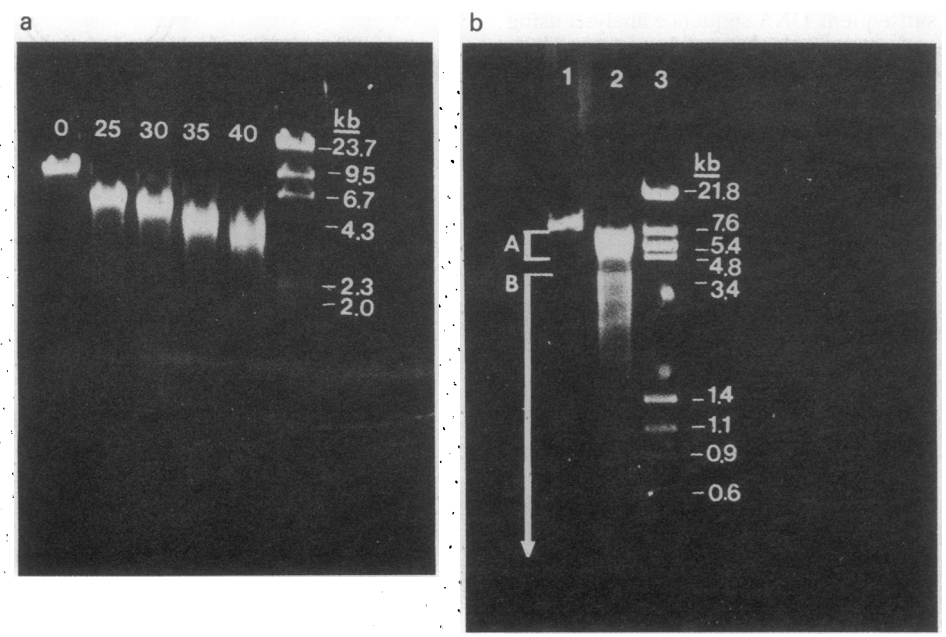

FIG. 2. Nuclease BAL-31 digestion of linearized virus. (*a*) M13β3.6 was linearized at the *Bgl* I site in M13mp7 and then digested with BAL-31. Aliquots were removed at 0, 25, 30, 35, and 40 min, and fragments were sized on a 0.7% agarose gel. Markers: λ DNA was digested with *Hin*dIII. (*b*) Aliquots of the BAL-31 digest were pooled and digested with *Eco*RI, and fragments were sized on a 0.7% agarose gel (lane 2); two fragment size ranges are present—A represents M13mp7-derived fragments and B represents insert fragments. Lane 1: *Bgl* I-linearized M13β3.6 (10.8 kb). Markers: λ DNA was digested with *Eco*RI and φX174 was digested with *Hae* III (lane 3).

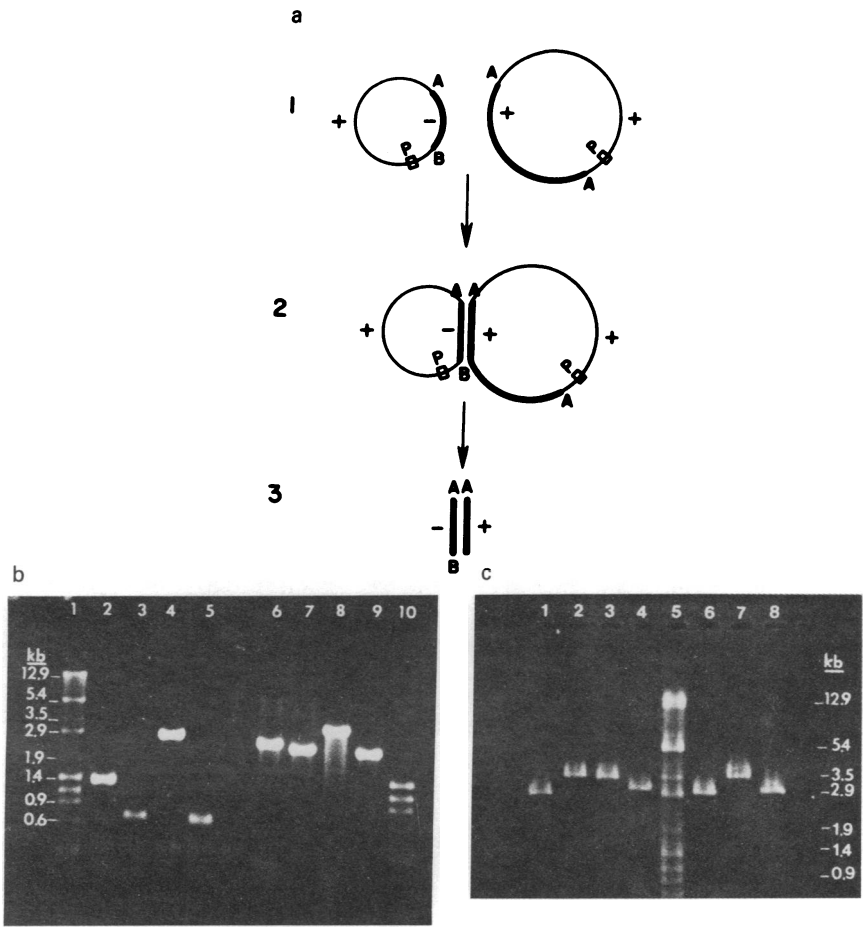

FIG. 3. Determination of insert size by nuclease S1 assay. (*a*) Nuclease S1 assay. Steps: 1, single-stranded DNA from different clones containing variable length inserts (thick line A—B) is hybridized to single-stranded DNA from a clone containing the entire complementary insert strand (thick line A—A); 2, if a complementary region exists between the two, then a hybrid forms: 3, M13 (+) strands (light lines) and unmatched insert regions are digested with nuclease S1. (*b*) Clones containing a wide range of insert sizes were processed, and nuclease S1-resistant hybrids were sized on a 0.7% agarose gel (lanes 2–9). Markers: λ DNA was digested with *Bam*HI/*Sma* I and φX174 was digested with *Hae* III (lane 1); φX174 was digested with *Hae* III (lane 10). (*c*) Clones containing a targeted range of insert sizes (i.e., a library was made from a specific time interval of BAL-31 digestion) were processed, and nuclease S1-resistant hybrids were sized (lanes 1–4 and 6–8). Markers: as in *b* (lane 5).

knowledge of the restriction endonuclease map of the insert and needs relatively short film exposure times (i.e., <12 hr) for DNA sequence reading. The deletion mutant library is easily constructed, and the nuclease S1 hybridization screening can be done quickly on a portion of the single-stranded deletion mutant DNAs. Overlapping deletions of the entire insert or of a specific region of interest can then be analyzed from aliquots of the remaining single-stranded deletion mutant DNA.

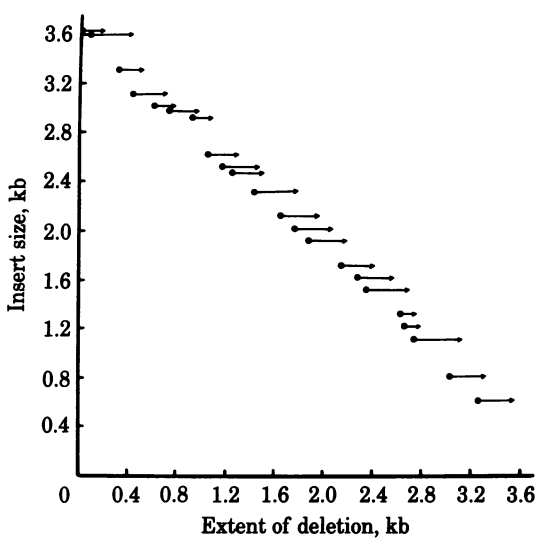

FIG. 4. Correspondence between insert size and extent of deletion. Insert size was estimated by the nuclease S1 assay described in Fig. 3. Extent of deletion was determined from the actual sequence analysis of overlapping clones. The linear relationship is indicated by the closed circles and the extent of sequence analysis for each clone is indicated by the length of the arrow. More than 70% of the total DNA sequence obtained was confirmed by overlap analysis.

The technique is versatile and with slight modification allows cloning and sequence analysis of fragments that have a variety of blunt or cohesive ends. In addition, the initial full-length insert need not be cloned into M13 for the linearization and BAL-31 digestion steps. However, M13 does offer several distinct advantages for use as a vector. Its complete sequence is known, and a number of unique sites (e.g., *Bgl* I, *Bgl* II, *Ava* I, and *Ava* II) in M13mp7 are available for use for linearization of the recombinant virus prior to the BAL-31 treatment step, as long as there is no corresponding site in the insert. The recent introduction of M13mp8 and M13mp9 further adds to the versatility of this technique, because unique *Sal* I, *Pst* I, *Hin*dIII, and *Bam*HI sites can serve for either insertion of fragments or for linearization prior to BAL-31 nuclease digestion (19). In addition, these modified viruses could be used in place of M13dR as cloning vectors for the blunt/cohesive-ended fragments after BAL-31 nuclease digestion, because they contain a variety of cohesive restriction sites distal to the primer as well as several unique sites that generate a blunt end close to the primer site.

We thank Dr. J. Messing for generously providing the bacterial strains and M13mp7. M.P. thanks Drs. W. Barnes, K. Alton, and T. Friedmann for patient teaching of the dideoxy chain-termination technique at a workshop on DNA sequence analysis (Given Institute of Pathobiology, Aspen, CO; 1980) and Dr. W. Barnes for continued help and advice during the course of this project. We also thank Eric Rappaport for help and Ms. Carol Way for assistance in the preparation of the manuscript and the photographs. This work was supported in part by National Institutes of Health Grants AM 16691 and HL 28157, a grant

from UNICO, and National Institutes of Health Research Service Award HL 07150.

1. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
2. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
3. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
4. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
5. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
6. Deininger, P. L. & Schmid, C. W. (1976) *J. Mol. Biol.* **106**, 773–790.
7. Coggins, L. W., Grindlay, G. J., Vass, J. K., Slater, A. A., Montague, P., Stinson, M. A. & Paul, J. (1980) *Nucleic Acids Res.* **8**, 3319–3333.
8. Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M. & Nienhuis, A. W. (1980) *Nucleic Acids Res.* **8**, 6113–6128.
9. Fritsch, E. F., Shen, C.-K. J., Lawn, R. M. & Maniatis, T. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 761–775.
10. Miesfeld, R., Krystal, M. & Arnheim, N. (1981) *Nucleic Acids Res.* **9**, 5931–5947.
11. Poncz, M., Solowiejczyk, D., Harpel, B., Mory, Y., Schwartz, E. & Surrey, S. (1982) *Hemoglobin* **6** (1) 27–36.
12. Miller, J., ed. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 443.
13. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
14. Cummings, I. W., Browne, J. K., Salser, W. A., Tyler, G. V., Snyder, R. L., Smolec, J. M. & Summers, J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1842–1846.
15. Messing, J., Gronenborn, B., Muller-Hill, B. & Hofschneider, P. H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3642–3646.
16. Heidecker, G., Messing, J. & Gronenborn, B. (1980) *Gene* **10**, 69–73.
17. Barnes, W. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4281–4285.
18. Frischauf, A. M., Garoff, H. & Lehrach, H. (1980) *Nucleic Acids Res.* **8**, 5541–5549.
19. Messing, J. (1982) in *Genetic Engineering, Principles and Methods*, eds. Hollaender, A. & Setlow, J. (Plenum, New York), Vol. 4, in press.