Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1021–1025, February 1992
Biochemistry

Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting

(methylation/Vent DNA polymerase/terminal transferase/DNase I/transcription)

PAUL A. GARRITY AND BARBARA J. WOLD*

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Communicated by Norman Davidson, October 8, 1991 (received for review September 4, 1991)

ABSTRACT **We have developed a simplified procedure for the ligation-mediated polymerase chain reaction (LMPCR) using *Thermococcus litoralis* DNA polymerase (Vent DNA polymerase). We show that Vent DNA polymerase produces correct, blunt-ended primer extension products with substantially higher efficiency than *Thermus aquaticus* (*Taq*) DNA polymerase or modified T7 DNA polymerase (Sequenase). This difference leads to significantly improved genomic sequencing, methylation analysis, and *in vivo* footprinting with LMPCR. These improvements include representation of all bands with more uniform intensity, clear visualization of previously difficult regions of sequence, and reduction in the occurrence of spurious bands. It also simplifies the use of DNase I cut DNA for LMPCR footprinting.**

Footprinting experiments are commonly and productively used to study protein–DNA interactions and DNA configuration *in vitro*. Analogous *in vivo* experiments done on genes in the living cell can bring a different and useful data set to the problem of gene expression, but they require special methods for visualizing the result. Direct genomic sequencing techniques, which permit the examination of single-copy genes in large genomes, are being used increasingly for this purpose (1–4). Ligation-mediated PCR (LMPCR) is a recently introduced method that substantially increases the absolute signal and the signal-to-noise ratio obtained for genomic sequencing (2, 5, 6). It does so by coupling PCR with genomic sequencing to provide specific amplification of a sequence "ladder," while preserving the identity and relative quantitative representation of each rung in the original cleaved genomic DNA preparation. Its application has made *in vivo* footprinting (2) and chromosomal methylation analysis (6) more readily accessible for organisms with large genomes (e.g., mammals).

While LMPCR has been used successfully by a number of investigators to obtain high quality *in vivo* footprint and methylation information (2, 6, 7, 8), it has had two problems that can significantly compromise data quality. These effects are minor in some regions of sequence but can be problematic in others. First, certain bands are consistently weak or missing in the genomic ladders. Second, "extra" bands occasionally appear in the genomic ladders. These bands, which aren't predicted from the sequence as independently determined from cloned DNA, are usually adjacent to expected bands and therefore convert some triplets into quartets, some doublets into triplets, and so on. We present here a solution for these problems that also permits simplification of the LMPCR procedure. These improvements stem from the use of *Thermococcus litoralis* DNA polymerase (Vent polymerase). This thermostable polymerase possesses no detectable terminal deoxynucleotidyltransferase activity under our conditions, and this characteristic dramatically improves LMPCR genomic sequencing. For *in vivo* footprinting and genomic sequencing applications, Vent polymerase yields substantially superior results, improving overall signal and, most importantly, the quality of sequence in difficult regions. We also show that starting material possessing 3′-hydroxyl ends (in this case DNase I-cut DNA), which had required modification of template ends with dideoxynucleotides in the older form of LMPCR (9), can now be used for *in vivo* footprinting purposes without modification.

MATERIALS AND METHODS

Cell Culture and DNA Preparation. L cells were grown in Dulbecco's modified Eagle's medium with 10% undialyzed calf serum (Irvine) and 2 mM glutamine. Naked and *in vivo* dimethyl sulfate (DMS)-treated MM14 DNA was provided by P. Mueller. DNA samples for genomic sequencing and DMS footprinting were prepared as in refs. 10 and 11. *In vivo* DNase I treatment was as in ref. 12, except that cells were permeabilized on ice with lysolecithin (0.25 mg/ml) for 60 sec. Addition of dideoxynucleotides prior to LMPCR where noted was as in ref. 9.

LMPCR. LMPCR using T7 DNA polymerase (Sequenase version 1.0; United States Biochemical) and *Thermus aquaticus* (*Taq*) DNA polymerase (AmpliTaq; Cetus) was done as in refs. 2 and 5. LMPCR using *Thermococcus litoralis* DNA polymerase (Vent; New England Biolabs) was done as below. All solutions were chilled and manipulations were performed on ice except as noted. The pH values given are for room temperature. To 5 μl (2 μg) of DNA in TE (10 mM Tris·HCl, pH 7.5/1 mM EDTA) was added 25 μl of first-strand mix [1.2× first-strand buffer (48 mM NaCl/12 mM Tris·HCl, pH 8.9/6 mM $MgSO_4$/0.012% gelatin) with 0.3 pmol of gene-specific primer 1, 240 μM each dNTP, and 1 unit of Vent polymerase]. First-strand synthesis used a thermal cycle of 5 min at 95°C, 30 min at 60°C, and 10 min at 76°C. The samples were immediately iced. (It is important to minimize Vent polymerase activity during the ligation step by keeping the sample cold.) Twenty microliters of dilution solution (110 mM Tris·HCl, pH 7.5/18 mM $MgCl_2$/50 mM dithiothreitol/0.0125% bovine serum albumin) and 25 μl of ligation solution [10 mM $MgCl_2$/20 mM dithiothreitol/3 mM ATP/0.005% bovine serum albumin with 100 pmol of unidirectional linker in 250 mM Tris·HCl (pH 7.7) (thawed and added on ice) and 4.5 units of T4 DNA ligase (Promega)] were added. After incubation for 12–16 hr at 17°C, samples were iced and 9.4 μl of precipitation solution (0.1% yeast tRNA/2.7 M sodium acetate, pH 7.0) and 220 μl of ethanol were added. The samples were placed at −20°C for ≥2 hr and then spun for 15 min at 4°C in a microcentrifuge. The pellets were washed with 75% ethanol and dried in a Speed-Vac rotary evaporator (Savant). Samples were resuspended in 70 μl of water at room

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: LMPCR, ligation-mediated polymerase chain reaction; DMS, dimethyl sulfate; MCK, muscle creatine kinase.

*To whom reprint requests should be addressed.

temperature and placed on ice. After addition of 30 μl of amplification mix [3.33× amplification buffer (133 mM NaCl/67 mM Tris, pH 8.9/17 mM $MgSO_4$/0.03% gelatin/0.3% Triton X-100) with 670 μM each dNTP, 10 pmol of gene-specific primer 2, and 10 pmol of linker primer LMPCR.1)] and 3 μl (3 units) of Vent polymerase, samples were overlaid with 90 μl mineral oil and subjected to PCR using 18 cycles of 1 min at 95°C, 2 min at 66°C, and 3 min at 76°C, with these modifications: (*i*) first-round denaturation was 3.5 min at 95°C; (*ii*) 5 sec was added to the 76°C step with each successive cycle (e.g., second round, 3 min 5 sec at 76°C); (*iii*) for cycle 18, the 76°C step was 10 min. Samples were then placed on ice and 5 μl of labeling mix [1× amplification buffer with 2 mM each dNTP, 2.3 pmol of gene-specific primer 3 (end-labeled as in refs. 2 and 5), and 1 unit of Vent polymerase] was added. The labeling cycle was 3.5 min at 95°C, 2 min at 69°C, 10 min at 76°C, 1 min at 95°C, 2 min at 69°C, and 10 min at 76°C. The reaction was stopped by placing the samples on ice and adding 300 μl of stop solution (10 mM Tris·HCl, pH 7.5/4 mM EDTA/260 mM sodium acetate, pH 7.0, containing tRNA at 67 μg/ml). Samples were shifted to room temperature and extracted with 400 μl of phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol). The aqueous phase was split into four aliquots of 94 μl, and 235 μl of ethanol was added to each. Before electrophoresis, samples were precipitated, washed, and dried as above. After resuspension in 7 μl of load dye (2, 5) and heating at 85°C for 5 min, samples were iced and then loaded on 6% polyacrylamide sequencing gel (2, 5). Loading one-fourth of an LMPCR mixture per lane yielded a strong signal on X-AR film (Kodak) after 3 hr with an intensifying screen at −80°C or 12 hr with no screen at −20°C. The unidirectional linker, linker primer, and muscle creatine kinase (MCK) oligonucleotides were as in ref. 2. The sequences (5′ to 3′) of the metallothionein I oligonucleotides were GAGTTCTCGTAAACTCCAGAGCAGC (primer 1), CAGAGCAGCGATAGGCCGTAATATC (primer 2), and AGCGATAGGCCGTAATATCGGGGAAAGC (primer 3).

RESULTS AND DISCUSSION

LMPCR (Fig. 1) relies on creation of a blunt end in the initial primer extension reaction to serve as a ligation substrate. Later, in the labeling reaction, precise blunt-end termination of the extension product is required. If the final labeling extensions stop short or add extra nontemplated bases, the result will be extraneous, inappropriate bands. In general, imperfect extension products may result from DNA polymerases adding a nontemplated additional base after creating a blunt end (referred to as terminal transferase activity) (13). Both polymerases commonly used for LMPCR display some terminal transferase activity. Sequenase, used in the first-strand synthesis reaction, adds an extra base to ≈50% of its products. *Taq*, used in the PCR amplification and labeling steps, adds an extra base to ≈95% of its products (P. Mueller and B.J.W., unpublished data). Such activity during the first-strand synthesis creates molecules unable to participate in the blunt-end ligation. Should this activity show sequence preference, it would lead to underrepresentation or even complete loss of specific bands in the final LMPCR product. Terminal transferase activity might also explain the origin of spurious "extra" bands in an LMPCR ladder. For example, a single band in a genomic sequencing ladder would appear as a doublet if some products of the labeling reaction acquired the extra base. We hypothesized that the terminal transferase activities of Sequenase and *Taq* were the major source of imperfect regions in LMPCR ladders and sought a DNA polymerase that lacks appreciable terminal transferase activity.

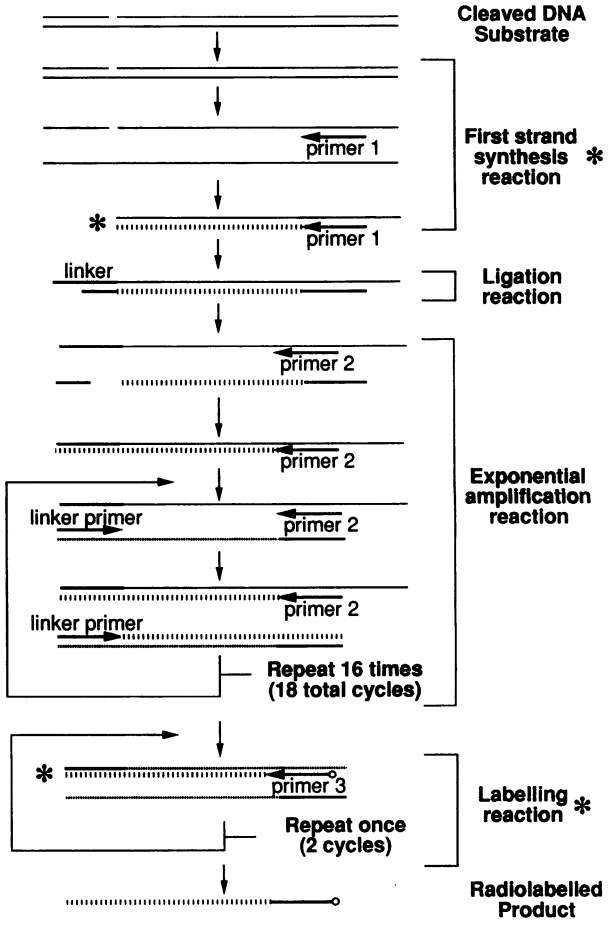

FIG. 1. LMPCR. Gene-specific primer 1 is hybridized to appropriately cleaved genomic DNA and extended using a DNA polymerase creating a blunt end (first-strand synthesis). This blunt end is ligated to a unidirectional linker of defined length and sequence with one blunt end (ligation reaction). This product is a substrate for PCR using gene-specific primer 2, which hybridizes internally, and a linker primer (LMPCR.1), which hybridizes to the ligated sequence (exponential amplification reaction). An end-labeled gene-specific primer, primer 3, is used to visualize the LMPCR product. Asterisks mark steps requiring efficient blunt-end generation. Open circle represents radioactive 5′-labeling of primer 3.

Vent DNA polymerase was tested in side by side comparisons with Sequenase and *Taq* DNA polymerase. The noncoding strand of the mouse MCK enhancer was used (14, 15) because it contains runs of G residues that have been problematic in Sequenase/Taq-based LMPCR. Mouse genomic DNA treated with DMS *in vitro* was used in the initial tests. Subsequent piperidine treatment gave G-specific cleavage (16), and LMPCR with MCK primers yielded the G-specific MCK sequence ladder. The activities of *Taq* DNA polymerase and Vent DNA polymerase were compared in the amplification and labeling stages of LMPCR. The products of Vent-catalyzed amplification and labeling consistently migrated more rapidly than those of *Taq*-catalyzed companion reactions by a one-base increment (Fig. 2). Since *Taq* is known to add an extra base to most of its products, we interpreted the migration shift as an indication that Vent lacked detectable terminal transferase activity and might therefore be an excellent candidate to replace both Sequenase and *Taq* in LMPCR.

When Vent was compared with Sequenase in the first-strand synthesis, the most obvious effect was that the yield of LMPCR product increased severalfold (Fig. 2). This is consistent with the creation of more blunt-ended molecules by Vent in the first-strand synthesis and confirms that this

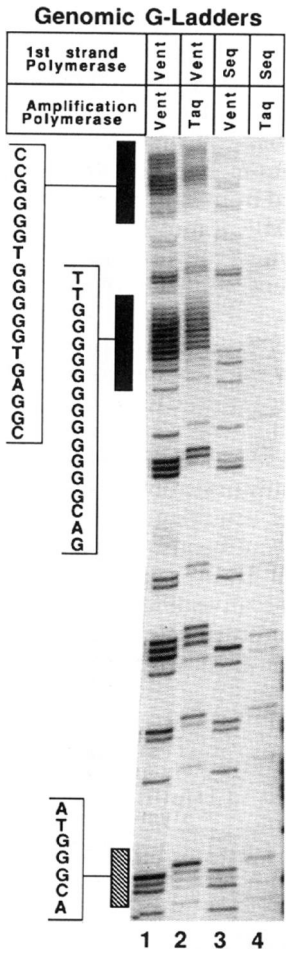

FIG. 2. LMPCR polymerase comparison. Each reaction used MCK primers and 2 μg of MM14 cell DNA treated with DMS *in vitro* and piperidine cleaved. Sequence at left is taken from Fig. 3. Hatched and solid bars are discussed in text. Seq, Sequenase.

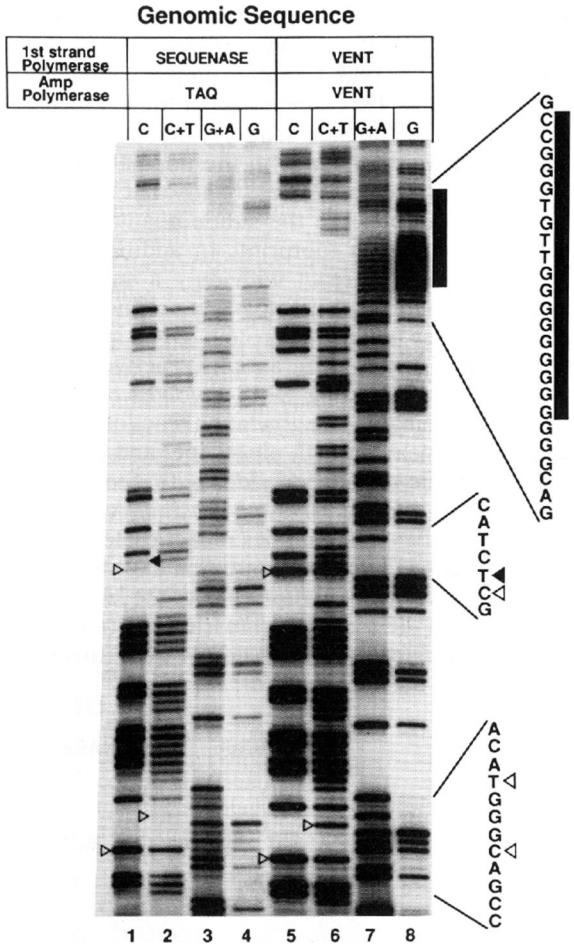

FIG. 3. LMPCR direct genomic sequencing method comparison. Each reaction used MCK primers and 2 μg of MM14 cell DNA treated with the appropriate Maxam and Gilbert sequencing reagent *in vitro* and then cleaved with piperidine. Black bar and triangles are discussed in text. Sequence at right was determined in ref. 15 from cloned DNA. An A residue reported in ref. 15 is here of near equal intensity in G and G+A lanes, so the sequence below the bar is given as GACGGG in agreement with determination from cloned DNA in ref. 14.

step is suboptimal with Sequenase. More important, Vent first-strand synthesis allowed visualization of bands that were underrepresented or absent in Sequenase-based LMPCR (note especially the several runs of G in the MCK enhancer denoted by the black bars in Fig. 2). Also, all bands in the ladder were more even in intensity. This is consistent with Vent creating blunt ends at many different sequences with similar efficiency and suggests that Sequenase shows sequence bias in blunt-end production. In addition to the shift in migration, the use of Vent instead of *Taq* in the amplification and labeling steps generated a sequencing ladder in which fewer spurious bands appeared. A good example is noted by the hatched bar in Fig. 2, where a known G triplet (15) appeared to be a quartet when *Taq* was used but resolved into the expected triplet when Vent was used. Thus, the use of Vent polymerase in LMPCR improved both the first-strand synthesis step and the amplification and labeling steps.

Sequencing and Methylation. In LMPCR-based genomic sequencing, total genomic DNA is cleaved by the Maxam and Gilbert base-specific reactions (11, 16) and this cleaved DNA serves as the substrate for LMPCR. Fig. 3 shows comparative genomic sequencing of the MCK enhancer done by either Sequenase/*Taq*-based or Vent-based LMPCR. Regions are noted where sequence determinations could not be made using Sequenase/*Taq*-based LMPCR (the published sequence is given at right). In the region noted at the bottom of Fig. 3, not only is a T residue missing (upper open triangle), but a C residue is rendered ambiguous by the presence of spurious bands in the G+A and G lanes (lower open triangle). In the middle region, the lower two bases of a CTC triplet are of equal intensity in the C lane (open triangle, correct band) due to a spurious band in the C lane (filled triangle). In each of these cases the Vent-based results are unambiguous. The extended G-rich regions (noted by the black bar) are even more dramatically improved by the use of Vent, though some ambiguity remains.

The analysis of chromosomal methylation [which has been implicated in such phenomena as chromosomal imprinting (17) and X chromosome inactivation (18)] is also improved. The Maxam and Gilbert C and C+T reactions require that C residues be unmethylated to participate. Because only unmethylated C residues appear in the sequence ladder, genomic sequencing provides both sequence and methylation information. The two adjacent cytosines in the G-rich region are represented well only with Vent-based LMPCR and the CTC triplet noted above appears to be a CCC triplet unless Vent-based LMPCR is used. Thus methylation information concerning these residues, which would have been either ambiguous or nonexistent using the older method, can now be clearly determined. It can be concluded from this experiment that the C residues of the MCK enhancer are not entirely methylated in MM14 cells, because they do appear in the ladder. However, the degree of partial methylation of any C residue can be determined only by a side-by-side comparison with unmethylated control DNA; hence, methylation of some copies of the MCK enhancer would not be detected in this experiment.

DMS Footprinting. *In vivo* footprinting is an especially sensitive application of genomic sequencing because it focuses directly on the relative intensities of individual bands in different samples. This demands that the intensity of each band in the final sequence ladder reflect in a consistent manner its relative abundance in the population of starting material. Thus, spurious "extra" bands that comigrate with genuine bands can obscure the subtle quantitative changes that usually comprise a footprint. In addition, some very useful cleavage agents react with only a subset of bases. Visualization of all legitimate cleavage products is therefore important, because an area of protein–DNA interaction may contain only one or two differentially reactive sites.

In vivo footprinting using DMS involves exposing intact cells to DMS, terminating the alkylation reaction, purifying the alkylated DNA, cleaving with piperidine, and comparing the resulting G-specific sequence ladder with one generated by exposing purified, naked DNA to DMS *in vitro* (4). Band intensity changes between samples reflect protein binding and any other changes in DNA structure that alter reactivity with DMS. Fig. 4 shows *in vivo* footprinting of the MCK enhancer, which is active in differentiated muscle cells (myocytes) but not in undifferentiated muscle precursor cells (myoblasts) (2, 14, 15). G-specific sequence ladders were derived from *in vivo* DMS treatment of undifferentiated MM14 myoblasts and differentiated MM14 myocytes and from *in vitro* DMS treatment of naked MM14 DNA. LMPCR was performed with either Sequenase or Vent first-strand synthesis followed by either *Taq* or Vent amplification and labeling. The footprint information derived from this experiment was consistent with that of previous LMPCR footprints of the MCK enhancer (2). In the region shown, myocyte-specific footprints are noted at three regulatory elements that have previously been defined as important for function (reviewed in ref. 2). The biological implications of this pattern have been discussed (2), and we focus here on how the new methods affect *in vivo* footprint analysis. Vent-based LMPCR gives greater absolute signal, and interactions that are sometimes difficult to see using Sequenase/*Taq*-based LMPCR, such as those at MEF-2 and near MEF-1, are now more obvious. This improvement is the combined result of acquiring previously missing bands and eliminating extraneous bands. For footprinting purposes, it is vital that identical DNA samples yield identical LMPCR results so that quantitative differences between different DNA samples can be interpreted. Vent-based LMPCR yields highly reproducible results, as shown by the exact match between the

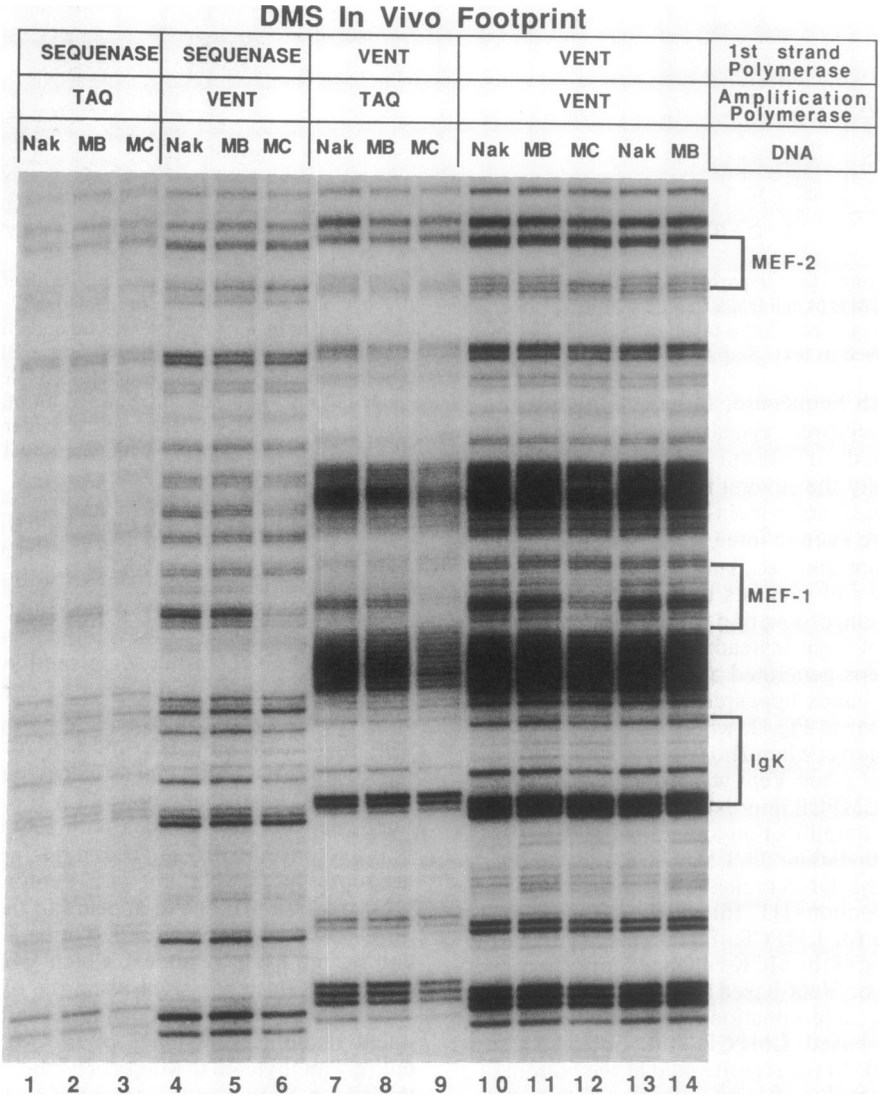

FIG. 4. LMPCR DMS *in vivo* footprinting method comparison. Each reaction used MCK primers and 2 μg of MM14 cell DNA treated with DMS *in vivo* or *in vitro* prior to piperidine cleavage. Naked (Nak) DNA was purified and then DMS-treated *in vitro*. Myoblast (MB) and myocyte (MC) DNA samples were from cultured cells treated with DMS. Binding sites indicated at right are described in ref. 2, where IgK is called κ and MEF-2 is called A-rich. Overexposure of lanes 10–14, required to see lanes 1–3, obscures footprints apparent at nonsaturating exposures.

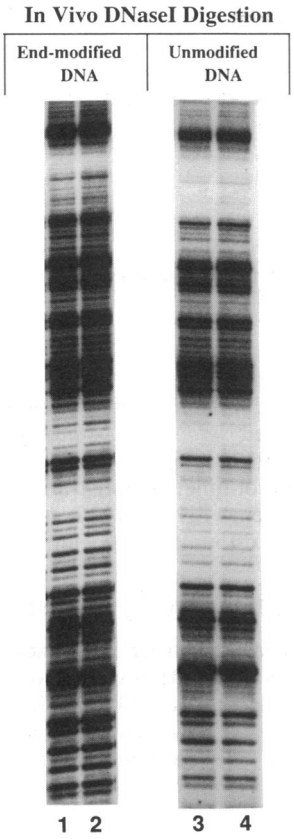

FIG. 5. Comparison of LMPCR of DNase I-digested DNA with (lanes 1 and 2) and without (lanes 3 and 4) prior dideoxynucleotide modification. Each reaction used metallothionein I primers and 2 μg of *in vivo* DNase I-digested L-cell DNA (lanes 1 and 3, DNase I at 37.5 μg/ml; lanes 2 and 4, 25 μg/ml). All lanes are from the same autoradiograph.

ladders from independent LMPCRs using duplicate DNA samples (Fig. 4, lanes 10 and 13, lanes 11 and 14). The LMPCR improvements observed were not restricted to the MCK enhancer region shown. Similar results were obtained for the mouse metallothionein I promoter (unpublished data).

DNase I-Cleaved Substrates. Beyond the issues of ladder quality, a separate limitation of LMPCR had been an inability to use DNA possessing 3′-hydroxyl ends for *in vivo* footprinting from organisms with large genomes. Although DMS/piperidine cleavage does not leave 3′-hydroxyl ends, other useful footprinting agents, such as DNase I, do. Riggs and coworkers were able to circumvent this problem by the addition of a dideoxynucleotide to the 3′-hydroxyl ends of DNase I-digested material prior to Sequenase/*Taq*-based LMPCR (9) and have obtained *in vivo* footprints with DNase I (19). Though effective, the additional manipulations are time-consuming and in our hands have resulted in low recovery of input DNA. With the Vent-based protocol, unmodified *in vivo* DNase I-digested DNA samples yield ladders similar in clarity and intensity to those from dideoxynucleotide-modified DNA (Fig. 5). Thus the addition of a dideoxynucleotide prior to LMPCR of 3′-hydroxyl-containing DNA is no longer necessary. The basis for this is not certain, but increased temperature (76°C versus 47°C) in the Vent-catalyzed first-strand synthesis reaction may be important. It may inhibit the priming of DNA synthesis by the enormous numbers of genomic DNA 3′-hydroxyl ends present in the first-strand reaction, while still providing efficient extension from the hybridized, gene-specific oligonucleotide.

The more uniform and efficient use of starting material in Vent-based LMPCR should allow the use of less sample DNA while still obtaining statistically significant results. Fluctuations in band intensity due to sampling error occur when the population sampled is small (founder effect, as discussed in ref. 5). They can obscure sequence or be mistaken for a footprint. Improved efficiency in LMPCR reduces the potential for founder-based artifacts in formerly problematic sequences. Although the favorable effects of the Vent-based LMPCR procedure can be readily explained by an absence of terminal transferase activity, that is not formally proved here. Whatever their mechanistic origin, the properties of Vent extension reactions reported here suggest that Vent polymerase may also be superior for other applications in which blunt-ended products are desired.

We thank Paul Mueller for generous gifts of DNA, comments on the manuscript, and many helpful discussions; Linda Huang and Jeff Miner for advice and comments on the manuscript; and Joe Hacia and Pete Mathers for assistance and discussions. This work was supported by grants from the Muscular Dystrophy Association and National Institutes of Health to B.J.W. and by Predoctoral Training Award H600021.

1. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
2. Mueller, P. R. & Wold, B. (1989) *Science* **246**, 780–786.
3. Mirkovitch, J. M. & Darnell, J. E. (1991) *Genes Dev.* **5**, 83–93.
4. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. (1985) *Science* **227**, 134–140.
5. Mueller, P. R. & Wold, B. (1991) *Methods* **2**, 20–31.
6. Pfeifer, G. P., Steigerwald, S. D., Mueller, P. R., Wold, B. & Riggs, A. D. (1989) *Science* **246**, 810–813.
7. Kara, C. J. & Glimchēr, L. H. (1991) *Science* **252**, 709–712.
8. Rideout, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. (1990) *Science* **249**, 1288–1290.
9. Tanguay, R. L., Pfeifer, G. P. & Riggs, A. D. (1990) *Nucleic Acids Res.* **18**, 5902.
10. Mueller, P. R., Salser, S. J. & Wold, B. (1988) *Genes Dev.* **2**, 412–427.
11. Saluz, H. P. & Jost, J. P. (1987) *A Laboratory Guide to Genomic Sequencing* (Birkhauser, Boston).
12. Zhang, L. & Gralla, J. D. (1989) *Genes Dev.* **3**, 1814–1822.
13. Clark, J. M. (1988) *Nucleic Acids Res.* **16**, 9677–9686.
14. Jaynes, J. B., Johnson, J. E., Buskin, J. N., Gartside, C. L. & Hauschka, S. D. (1988) *Mol. Cell. Biol.* **8**, 62–70.
15. Sternberg, E. A., Spizz, G., Perry, W. M., Vizard, D., Weil, T. & Olsen, E. N. (1988) *Mol. Cell. Biol.* **8**, 2896–2909.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Solter, D. (1988) *Annu. Rev. Genet.* **22**, 127–146.
18. Lyon, M. F. (1991) *Trends Genet.* **7**, 69–70.
19. Pfeifer, G. P. & Riggs, A. D. (1991) *Genes Dev.* **5**, 1102–1113.