Cytosine Methylation: Quantitation by Automated Genomic Sequencing and GENESCAN™ Analysis

BioTechniques 21:126-133 (July 1996)

Cheryl L. Paul¹ and Susan J. Clark^{1,2}

¹Kanematsu Laboratories, Royal Prince Alfred Hospital, Camperdown, and ²CSIRO, Division of Biomolecular Engineering, North Ryde, Australia

ABSTRACT

Bisulfite treatment and PCR amplification of genomic DNA permits the methylation analysis of any cytosine residue in a target sequence. By cloning and sequencing the PCR product, the methylation of individual molecules can be determined, whereas direct sequencing of the PCR product can provide an average of the methylation status in the population of molecules. Reliable quantitation of cytosine methylation by direct sequencing, however, has not been possible with current methods. In this paper we describe an accurate and innovative protocol to directly quantitate the methylation of any cytosine residue in the target sequence by fluorescence-based automated genomic sequencing. Only the cytosine and thymine residues of bisulfite-treated and amplified genomic DNA are sequenced. The degree of methylation is obtained by direct comparison of the cytosine and thymine signals, which have been labeled with the same fluorescent dyes. GENESCAN analysis is employed to achieve a fast and accurate estimate of methylation at every cytosine in the target sequence. Combining direct bisulfite genomic sequencing and GENESCAN analysis permits the rapid survey of detailed DNA methylation profiles. Using this approach we have found the unexpected result that multicopy plasmid DNA grown in a Dcm host is not always fully methylated as suggested by restriction enzyme data.

INTRODUCTION

Bisulfite genomic sequencing provides a positive and efficient method for detecting individual 5-methylcytosine (5-MeC) residues in genomic DNA (1,3–5). The method relies on the ability of sodium bisulfite to efficiently deaminate cytosine residues to uracil in single-stranded DNA, under conditions whereby 5-MeC remains nonreactive. The target DNA sequence can be amplified by polymerase chain reaction (PCR) with two sets of strand-specific primers to yield a pair of fragments, one derived from each strand, in which all uracil and thymine residues have been amplified as thymine (T) and only 5-MeC residues have been amplified as cytosine (C) (3).

Analysis of cytosine methylation from the bisulfite-treated and PCR-amplified DNA can be determined by either cloning and sequencing individual molecules or by direct PCR sequencing. The cloning strategy is very useful in determining the pattern of mosaic methylation within molecules and in determining any methylation differences between alleles. This approach can be used for quantitation of cytosine methylation, but this can be tedious because it often requires sequencing of 50–100 individual clones to provide an accurate estimate of methylation in the population of molecules. It would be more convenient to estimate the degree of methylation at each cytosine by direct sequencing of the PCR product and subsequent quantitation of the thymine signal (unmethylated cytosine) relative to the cytosine signal (methylated cytosine). However, although currently used direct sequencing and analysis protocols can provide a qualitative indi-

cation of methylation, we have found accurate quantitation to be difficult. For example, densitometric scanning of radiolabeled sequencing gels can be unreliable due to high background bands often observed on direct PCR sequencing gels, whereas automated direct sequencing of PCR products, using AmpliTaq® DNA polymerase with either dye-labeled dideoxy chain terminators or dye-primers, results in traces with nonquantitative peak heights (6,7). Therefore, a new strategy is required to further exploit the benefits of bisulfite genomic sequencing and allow accurate quantitation of cytosine methylation. In order to understand the mechanics of methylation within the cell, it will be important to have access to both cloning and direct sequencing protocols in future methylation studies.

In this study, we have devised a strategy that enables efficient and accurate quantitation of methylation at cytosine residues by direct PCR sequencing of amplified bisulfite-treated genomic DNA. Since analysis of automated fluorescence-based sequencing data is not sufficiently quantitative, we have employed GENESCAN™ analysis for accurate quantitation. We have developed a protocol whereby C and T sequencing reactions are carried out using the same fluorescent dye to minimize any spectral differences. The sequencing raw data are then analyzed using the GENESCAN Analysis program, which quantitates the C fluorescence signal relative to the T fluorescence signal at each cytosine residue in the target sequence independently of neighboring bases. The accuracy of this protocol was initially defined on mixtures of methylated plasmid DNA and then used to directly quantitate methylated genomic DNA.

MATERIALS AND METHODS

PCR Primers

The 5' primers were biotinylated at the 5' end during synthesis of the forward primer using Biotin Amidite from Applied Biosystems (Perkin-Elmer/Applied Biosystems Division [PE/ABI], Foster City, CA, USA), according to manufacturer's directions. The 3' primers were tailed with the sequence for the M13 (-21) universal primer at the 5' end of the reverse primer.

pBluescript primers. The specific pBluescript primers used to amplify bisulfite-converted top-strand DNA were as follows: 5' GTGGAATTCTT-ATTGGTTTATAATTTTATA, position 825–854; and 5' AACCGAATCCC-ACCTTTAAATAAATACTAA, position 1090–1115 (SK sequence; Stratagene, La Jolla, CA, USA). The corresponding M13 primer and Biotin primer pairs were as follows: 5'-TGTAACACGACGGCCAGTGTGGAATTCTTGGTT; and 5'-BiotinAACCGAATCCCACC-TTTAAATAAATACTAA.

Thy-1 DNA primers. The specific Thy-1 primers used to amplify bisulfite-converted top-strand DNA were as follows: 5' AATTGAGGGTTCTAG-ATTGGGTGTAGTAAT, position 497–526; 5' CCAAAAAGTAGTCTTAA-TATCCTTCCTCATAA, position 837–806 (mouse Thy-1 1.2 glycoprotein gene, Accession No. M12379, GenBank®). The corresponding M13 primer and Biotin primer pairs were as follows: 5' TGTAACACGACGGCC-AGTAATTGAGGGTTCTAGATTGG; and 5' Biotin CCAAAAAGTAGTCT-TAATATCCTTCCTCATAA.

Bisulfite Treatment and PCR Amplification

A 2- μ g aliquot of unmethylated pBluescript (SK) plasmid DNA isolated from *E. coli* DH5 α , *Hpa*II methylated plasmid DNA (2 μ g) and mouse L cell genomic DNA (2 μ g) were linearized with *Eco*RI and denatured with a tenth volume of 3 M NaOH for 15 min at 37°C. Bisulfite treatment of each DNA sample was carried out for 16 h at 55°C under conditions described in Clark et al. (3). The samples were purified using Magic™ DNA Clean-Up System desalting columns (Promega, Madison, WI, USA), eluted in 50 μ L of

H₂O. DNA buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) was added to bring the volume to 50–200 μ L and incubated with a tenth volume of 3 M NaOH for 15 min at 37°C. The solutions were neutralized by the addition of NH₄OAc, pH 7.0, to 3 M and the DNA was ethanol-precipitated, dried, resuspended in DNA buffer (100 μ L) and stored at -20°C.

PCR amplifications were performed in 100- μ L reaction mixtures containing 1–5 μ L of bisulfite-treated plasmid or genomic DNA, 200 μ M dNTPs, 1 μ M primers, 3 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl pH 8.3 and 0.5 μ L (2.5 units) AmpliTaq DNA polymerase (Perkin-Elmer, Norwalk, CT, USA) in a Corbett Research FTS-IC Fast Thermal Sequencer (Sydney, Australia) under the following conditions: 94°C for 2 min $\times$ 1 cycle; 94°C for 1 min, 50°C for 2 min, 72°C for 3 min, $\times$ 5 cycles; 94°C for 0.5 min, 50°C for 2 min, 72°C for 1.5 min, $\times$ 25 cycles; and 72°C for 6 min $\times$ 1 cycle.

Biotinylated-PCR Amplification

Biotinylated-PCR amplifications were performed on approximately 5 ng of first-round, amplified, bisulfite-treated DNA (3) that had been eluted in H₂O from a 4% polyacrylamide gel. Amplifications were performed in 100 μ L final volume containing PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 3 mM MgCl₂), 200 μ M dNTPs, 12 pmol biotinylated primer, 40 pmol M13-tailed primer and 3 U of AmpliTaq DNA polymerase. Thermal cycling was performed under the same conditions as described above, except annealing was performed at 55°C and the PCR products were purified using Magic or Wizard™ PCR purification columns (Promega).

Magnetic Bead Purification of Biotinylated PCR Products

The method described has been modified from the ABI User Bulletin No. 21 (PE/ABI). A 15- μ L aliquot of Dynabeads® M-280 Streptavidin magnetic beads (10 mg/mL) (DynaL AS, Oslo, Norway) was transferred to 1.5-mL centrifuge tubes and washed twice with 200 μ L TTL buffer (6 M LiCl, 300 mM Tris-HCl pH 8.0, 0.3% Tween® 20) applying a Dynal Magnetic Particle

Concentrator (DynaL MPC®-E). The beads were resuspended in 20 μ L of TTL buffer with gentle pipetting and transferred to the 45- μ L PCR containing 200–400 ng DNA, incubated at 43°C for 12–15 min for binding of biotinylated PCR products to the Dynabeads. Following binding, the magnet was applied and the supernatant discarded. The beads were resuspended in 200 μ L of freshly prepared 0.15 M NaOH/0.075% Triton® X-100 solution (prepared by dissolving 0.3 g NaOH in 50 mL Milli-Q® water (Millipore, Bedford, MA, USA) with 37.5 μ L Triton X-100) and incubated at room temperature for 4 min to denature the DNA. The magnet was applied, supernatant discarded and the beads rinsed with 200 μ L of NaOH/Triton solution, followed by two washes with 200 μ L TT buffer (10 mM Tris-HCl pH 8.0, 0.1% Tween 20). All traces of NaOH/ Triton rinse and TT buffer washes were removed, as any remaining liquid may affect sequencing reactions. The bead pellet was resuspended in 7.2 μ L in Milli-Q water and stored at 4°C until used for sequencing. NB: All washes and resuspensions were done by gentle pumping with a pipet, and the non-biotinylated complementary strand was not used.

Sequenase® Dye Primer Sequencing

Sequencing reactions were performed using a PRISM™ Sequenase® Dye Primer Sequencing Kit (PE/ABI) according to the manufacturer's directions except for the following modifications. The master mixture for A and C reactions per dye primer (for a single volume) contained the following: 0.5 μ L 10 $\times$ MOPS buffer; 0.5 μ L 10 $\times$ Mn²⁺/isocitrate buffer; 2 μ L M13-21 dye primer (JOE for A and FAM for C); and 1 μ L α thio-d/ddN mixture. The master mixture for G and T reactions per dye primer (for a single volume) contained the following: 1.5 μ L 10 $\times$ MOPS buffer, 1.5 μ L 10 $\times$ Mn²⁺/isocitrate buffer, 4 μ L M13-21 dye primer (TAMRA for G and ROX for T), 3 μ L α thio-d/ddN-mixture. Annealing reactions were performed in 1.5-mL microcentrifuge tubes, with either 1 μ L of bead-bound template and 4 μ L of master mixture for A and C, or 3 μ L of bead-bound template and 10 μ L of master mixture, for G and T and incu-

Research Reports

bated at 65°C for 2 min, then at room temperature for 15 min to ensure complete primer annealing. The tubes were placed on ice and 2 μ L of Sequenase/pyrophosphatase enzyme were added to the A and C reactions and 4 μ L to the G and T reactions. The tubes were incubated at 37°C for 15 min, placed on ice and the four individual bead-bound reactions pooled into 200 μ L TT buffer. A magnet was applied, the supernatant removed and the beads washed once with 200 μ L of TT buffer, taking care to remove all traces of buffer. The bead-bound reactions were resuspended in 4 μ L of loading buffer (prepared from 5 μ L recrystallized deionized formamide, 1 μ L 50 mM Na₂EDTA pH 8.0/blue dextran [30 mg/mL]) heated at 100°C for 2 min, placed on ice immediately and loaded onto a 6% polyacrylamide urea gel using a Model 373A DNA Sequencer (PE/ABI). Gel preparation, electrophoresis and sequencing data collection were performed as described in the ABI 373A DNA Sequencer User Manual (PE/ABI).

Two-Base Sequenase Sequencing Reactions for Quantitation

For quantitative sequencing reactions, the FAM M13(-21) dye primer (blue) is used for both C and T sequencing reactions (or G and A reactions). Master mixtures and sequencing reactions were performed as described above for the C reactions. The reactions were terminated with 100 μ L of TT buffer, the magnet was applied, supernatant removed and beads washed once with 100 μ L of TT buffer. To the beads, 0.8 μ L of Applied Biosystems GENESCAN-350 TAMRA Size Standard (yellow) from PE/ABI and 3.7 μ L of loading buffer were added to each C and T reaction, which were loaded separately into alternate lanes on the sequencing gel.

Quantitative Analysis

A modified version of GENESCAN 672 analysis software (UNLOCK 1.0, supplied by PE/ABI) was used to preprocess the sequence gel file into GENESCAN format (note: GENESCAN Analysis Version 1.2–2.1 can also be used to preprocess gel file). GENESCAN Analysis Version 1.2.0

was used to analyze the results data. The electrophoretograms show C and T tracks superimposed, and the blue color of T tracks is altered to red to distinguish the peaks. The percent methylation is calculated using peak height of C vs. peak height of C plus peak height of T for each position.

RESULTS AND DISCUSSION

Cytosine Methylation Quantitation Strategy

Sequencing analysis software usually includes normalization routines to compensate for different loadings of each reaction; and, in four-color sequencing, differences in the spectral properties of the four different fluorescent dyes are also used. Thus, in samples such as those encountered in our method, where the base composition is highly skewed (T or A rich), the sequencing analysis software routines may not be sufficiently quantitative for our purposes. We have therefore chosen to use ABI GENESCAN analysis, which explicitly does not contain assumptions about the nature of the DNA ladder being analyzed.

Figure 1 shows the strategy we devised to enable accurate quantitation of methylated cytosine residues, by direct PCR automated sequencing together with GENESCAN analysis of bisulfite-treated and amplified genomic DNA. Genomic DNA is bisulfite-treated, and the target sequence of interest is amplified with specific primers designed to the converted sequence. The PCR product generated is reamplified with another pair of specific primers, one of which is biotinylated and the other includes a 5' M13 (-21) sequence in addition to the bisulfite-specific converted sequence. The biotinylated PCR amplification is a second PCR step that is necessary not only to generate single-stranded DNA for reliable sequencing with Sequenase, but also for the creation of an M13 sequencing priming site that can be used with commercial dye-labeled primers.

Obtaining sequence data in both directions requires PCR amplification with two sets of alternate biotin and M13 (-21) tailed primer pairs. The biotin-labeled PCR fragments are purified

using Dynabeads and alkali denaturation. The Sequenase reaction is performed on the biotinylated single strand (still attached to the Dynabeads) in four separate tubes, each with a different dye-labeled M13 primer. The sequencing reaction is only required for checking the sequence of the PCR product. For quantitation of cytosine methylation in the target sequence, only the cytosine (C) and thymine (T) bases need to be sequenced. The C and T Sequenase reactions are performed in separate tubes using the same dye-labeled primer. A size standard is added, after the reaction, to each tube to permit the automated alignment of extension reactions run in different lanes without concern for lane-to-lane variability and gel idiosyncracies. As such, the data can be displayed as overlapping electrophoretograms for ease of analysis (as shown in Figure 2). The level of methylation is then calculated from the peak height of the C signal over the peak height of the C+T signals for each site.

Quantitation of Cytosine Methylation of *HpaII*-Methylated Plasmid DNA

The accuracy of quantitation of cytosine methylation by fluorescence-based PCR sequencing in combination with GENESCAN Analysis was initially tested on mixtures of unmethylated and *HpaII*-methylated pBluescript (SK) plasmid DNA.

Controls

A number of controls were incorporated into the design of the experiment. First, a cytosine residue was incorporated into the primer sequence to act as an internal cytosine control signal. Second, the plasmid DNA sequence chosen for sequencing contained an *EcoRII* site that could be used as an internal methylation control signal. Since *EcoRII* sites are methylated in *Dcm+* hosts, this site should be methylated in both the *HpaII*-methylated and unmethylated plasmid DNA. Indeed, the *EcoRII* site was shown to be fully methylated in both plasmid DNA preparations, as determined by a complete lack of digestion after overnight incubation with the *EcoRII* restriction enzyme. Third, the *HpaII*-methylated plasmid DNA preparation was defined

to be 100% methylated at *HpaII* sites, since there was a complete lack of nicking at these sites after overnight incubation with *HpaII* enzyme.

Sequenase Vs. *Taq* DNA Polymerase

Direct sequencing profiles of the PCR products from bisulfite-treated plasmid DNA were initially compared using AmpliTaq polymerase and dye-labeled dideoxy chain terminators and dye-labeled primers, with Sequenase and dye labeled primers. Sequencing with AmpliTaq polymerase resulted in

traces with uneven peak heights, especially at strings of thymines (T) (data not shown). The peak height variations reflect the disproportionate rate of incorporation of deoxynucleotides, a phenomenon that is highly dependent on the neighboring DNA sequence (7). Therefore, accurate quantitation analysis based on peak heights is difficult and often misleading using AmpliTaq polymerase. Sequencing the same PCR products using Sequenase resulted in a more even peak height distribution, as observed in T stretches (Figure 2A).

The relatively even peak heights generated by Sequenase are due to the ability of this enzyme to show less bias in nucleotide incorporation (8). Therefore, it was decided to perform the subsequent analysis using Sequenase rather than AmpliTaq polymerase in the initial direct PCR sequencing reaction to ensure accurate quantitation. Sequences were analyzed using DNA Sequencing Analysis Version 1.2.0, as shown in Figure 2A; however, it should be noted that subsequent analysis using Version 2.1.0 improved peak height spread even more noticeably, especially for the adenine profile in the target DNA.

Fluorescence-Based PCR Sequencing

To determine the accuracy of quantitating cytosine methylation at single sites by fluorescence-based PCR sequencing and GENESCAN analysis, mixtures of amplified bisulfite-treated unmethylated and *HpaII*-methylated plasmid DNA were prepared containing 0%, 25%, 50%, 75% and 100% methylated DNA. First, the PCR product generated from the unmethylated (0%) and *HpaII*-methylated (100%) plasmid DNA preparations were sequenced using Sequenase and dye-labeled primers to demonstrate that the bisulfite reactions had gone to completion. For the *HpaII*-methylated plasmid (Figure 2A), the only cytosines not converted were in the primer sequence, *HpaII* site and *EcoRII* site, whereas the cytosine at the *HpaII* site in the unmethylated plasmid sequence was converted to a thymine (Figure 2G).

GENESCAN Analysis

For quantitation analysis, the *HpaII*-methylated mixtures (0%, 25%, 50%, 75% and 100%) were sequenced for C and T profiles only. Labeled size markers were included in each lane so that the data could be analyzed by GENESCAN analysis. Initially, the C reaction was labeled with blue dye (FAM) and the T reaction was labeled with red dye (ROX), which is usual for sequencing reactions. However, we found that there was a need to correct for spectral differences between the FAM and ROX dyes, which further complicated the analysis and was a source of additional inaccuracies. To minimize such problems, we

Figure 1. A schematic protocol for quantitation of methylation by fluorescence-based automated sequencing and GENESCAN analysis.

Research Reports

chose to label both the C and T reactions with the same FAM dye, which gives a strong signal.

Electrophoretograms of the GENESCAN analysis data from direct sequencing (C and T tracks) of PCR products containing the *Hpa*II-methylated mixtures are shown in Figure 2, B-F. Table 1 shows the peak heights for each C and corresponding T residue at the primer C, *Hpa*II and *Eco*RII sites. A number of important observations can be made from this data.

Peak heights between different DNA sequencing reactions cannot be directly compared because signal intensities vary slightly from one sequencing run to another. For example, as shown in Table 1, the absolute peak height for the primer C signal changes for each sequencing set and therefore cannot be used as a 100% cytosine control signal for comparison with other cytosine signals in different sequencing runs. In fact, it is also important not to compare signal intensity between sites even within the same sequencing run, because signal intensity at each site slowly decreases across the sequencing profile, as shown in Figure 2. However, if the C signal intensity is compared at the same site, with the sum of the C and T signal intensities from the same site in the same sequencing reaction, then the quantitation data is quite accurate and reliable. As shown in Table 1 and Figure 2, C/C+T ratios for the 75% *Hpa*II-methylated plasmid mixture indicated 67% methylation, 50% *Hpa*II-methylated plasmid mixture indicated 49% methylation and 25% *Hpa*II-methylated plasmid mixture indicated 22% methylation at the *Hpa*II site. As expected, the 100% *Hpa*II-methylated PCR product produced only a C signal, which indicated 100% methylation, and the unmethylated plasmid PCR product produced only a T signal, which indicated 0% methylation. These spiked PCR experiments were performed three times for each ratio set, and all resulted in quantitative data that were accurate within a 10% error margin. This error is extremely small, given the many possibilities for introduced technical errors. The same spiked methylated and unmethylated bisulfite-treated and amplified PCR products were quantitated by comparing G and A ratios in the

Figure 2. Sequencing and GENESCAN electrophoretogram of *Hpa*II methylated and unmethylated plasmid DNA. Panels A and G: Sequencing electrophoretograms of *Hpa*II methylated and unmethylated plasmid DNA, respectively, from 854–901 bases (Stratagene SK sequence). Panels B–F: GENESCAN electrophoretograms of *Hpa*II-methylated plasmid DNA mixtures ranging from 100%, 75%, 50%, 25% and 0%, respectively. Panel A has the pBluescript M13 primer sequence, *Hpa*II site and *Eco*RII site marked with a line over the corresponding sequence. Quantitation of the cytosine signal over the cytosine plus thymine signal is marked as a percentage under each GENESCAN panel. To aid presentation, the cytosine electrophoretogram is colored blue and the thymine electrophoretogram is superimposed in red, even though both are dye-labeled in the Sequenase reaction with blue (FAM) dye. In the sequencing reactions, cytosine (C) is shown in blue, thymine (T) in red, adenine (A) in green and guanine (G) in black.

Table 1. Quantitation Data Derived From GENESCAN Analysis of the Methylated Plasmid PCR Product Mixtures

PCR Product	Primer C		<i>Hpa</i> II site		% M C/C+T	<i>Eco</i> RII site		% M C/C+T
	C	T	C	T		C	T	
100% <i>Hpa</i> II M/ 0% <i>Hpa</i> II Non-M	2402	0	1919	0	100%	1203	1334	47%
75% <i>Hpa</i> II M/ 25% <i>Hpa</i> II Non-M	1717	0	1436	741	67%	1290	794	62%
50% <i>Hpa</i> II M/ 50% <i>Hpa</i> II Non-M	2589	0	1144	1210	49%	1448	663	69%
25% <i>Hpa</i> II M/ 75% <i>Hpa</i> II Non-M	1821	0	345	1232	22%	1166	380	75%
0% <i>Hpa</i> II M/ 100% <i>Hpa</i> II Non-M	2183	0	0	2709	0%	1368	501	73%

Peak height measurements of the cytosine base at primer C, *Hpa*II and *Eco*RII sites, from Figure 2, B–F, are shown for cytosine (C) and thymine (T) profiles. Percent cytosine methylation (%C) is calculated by peak height of C over the peak height of C+T.

M = Methylated; Non-M = Nonmethylated

Figure 3. Direct PCR sequence and GENESCAN electrophoretogram of a methylated region of the mouse Thy-1 gene. Panel A: Direct PCR sequence of a portion of PCR fragment amplified from bisulfite-treated L-cell DNA using Thy-1-specific primers. Panel B: GENESCAN electrophoretogram of the corresponding Thy-1 sequence. Panel C: Direct PCR manual sequence and autoradiograph of the same portion of Thy-1 PCR fragment. CpG dinucleotides are overlined in Panel A and underlined in Panel B. In Panels B and C, cytosine residues that correspond predominantly to methylated cytosines in the untreated genomic DNA are indicated by closed triangles (the percent methylation at these CpG sites is shown). The open triangle indicates a site where there is partial methylation in the population of molecules; that is, a mixture of methylated cytosines (C) and unmethylated cytosines (T). This is the same site that has previously shown partial methylation by cloning and sequencing (3). In the sequencing reactions, cytosine (C) is shown in blue, thymine (T) in red, adenine (A) in green and guanine (G) in black. To aid presentation, the cytosine electrophoretogram is colored blue and the thymine electrophoretogram is superimposed in red, even though both are dye-labeled in the Sequenase reaction with blue (FAM) dye.

Research Reports

sequence of the opposite strand. This, too, resulted in reliable quantitation of methylation and confirmed quantitation of methylation by C and T analysis. Therefore, a direct comparison C and T or G and A signal intensities in a GENESCAN analysis of bisulfite-treated DNA can be used to accurately determine the percent methylation at any site.

An unexpected result was observed when analyzing the GENESCAN data for methylation at the *Eco*RII site. Restriction data had indicated that the *Eco*RII site was fully methylated in each of the DNA preparations. However, the data presented in Table 1 and Figure 2, B and F, show that the *Hpa*II-methylated plasmid DNA was, in fact, only 47% methylated at *Eco*RII and the unmethylated plasmid DNA was 73% methylated at *Eco*RII. Therefore, GENESCAN analysis of the PCR mixing experiments indicated a change in the estimated methylation status of the *Eco*RII site. The increase in the apparent methylation of the *Eco*RII site was consistent with the decrease in concentration of the *Hpa*II-methylated plasmid DNA in the PCR mixture. We have subsequently determined, by sequencing cloned DNA from the PCR products, that the level of methylation at the *Eco*RII sites did in fact differ between the two plasmid preparations, even though the *Eco*RII digestion conditions used indicated complete methylation. By further restriction experiments, we demonstrated that a proportion of the *Eco*RII sites were in fact hemimethylated, which may explain the lack of initial restriction. We also found that the methylation status of *Eco*RII can vary significantly between multicopy pBlue-script plasmid DNA preparations; plasmid DNA isolated from less-aerated 2–3-mL cultures tended to be less methylated than plasmid DNA isolated from well-aerated 500-mL cultures.

Quantitation of Cytosine Methylation of Genomic DNA

The sensitivity of quantitation of cytosine methylation by fluorescence-based PCR sequencing and GENESCAN analysis was tested on genomic DNA. Mouse L-cell DNA was bisulfite-treated and amplified using specific primers to Thy-1 DNA (3), followed by

a second round of amplification using the corresponding biotinylated and M13-tailed primers. Biotinylated PCR product was sequenced directly with Sequenase and dye-labeled primers as shown in Figure 3A. The C and T reactions were sequenced separately using blue dye primers (FAM), and the percent methylation of each cytosine residue was calculated using GENESCAN analysis for each position (Figure 3B). This region of DNA was chosen to study quantitative methylation by fluorescence-based PCR sequencing and GENESCAN analysis, because we had previously shown the DNA to be extensively methylated by direct manual sequencing (Figure 3C), and sequencing of individual molecules demonstrated a partially methylated site (3). Comparison of C and T peak heights indicated that all the CpG dinucleotides were heavily methylated (86%–93%), except for one CpG site at base 129–130, which was 71% methylated. This methylation data confirmed the average methylation data obtained from sequencing 52 individually cloned molecules and visual analysis of the manual sequencing gel.

We use Sequenase enzyme in quantitation experiments to ensure a better peak height distribution. However, Sequenase can be susceptible to DNA secondary structure and this can lead to small stops in the sequence. An example can be seen at positions 107–110 and 140–146 on the sequence profile in Figure 3A, with consequent artefactual C peaks in the corresponding position of the GENESCAN profile in Figure 3B. These C peaks were indeed artefactual and not an indication of poor bisulfite conversion, as these peaks were not present when the same PCR product was sequenced with *Taq* DNA polymerase. Artefactual C peaks are only a problem if there is a major sequencing stop near the start of the sequence, as the subsequent signal intensities can be affected. Therefore, we found that it was very important to use biotin-labeled single-stranded DNA that had been purified with Dynabeads as the template for sequencing reactions, rather than heat-denatured PCR products. It was also important to reduce any primer dimers in the reaction by cleaning the biotin PCR product with a Wizard

clean-up column prior to sequencing.

Direct quantitation of methylation can now be used as an efficient means to estimate the average methylation status of each cytosine residue in genomic DNA. This will be important in future methylation studies to enable a more detailed profile of methylation patterns within the cell.

ACKNOWLEDGMENTS

We thank Mel Kronick from PE/ABI in Foster City, CA, USA, for his excellent advice and troubleshooting during development of this protocol, and Jane Shepherd and Tim Orpin from PE/ABI, Australia, for their help, particularly with automated sequencing and GENESCAN interpretations. We would also like to thank Marianne Frommer and Peter Molloy for their continued encouragement and for the careful reading of this manuscript. The work has been supported in part by an Anthony Rothe Memorial Grant.

REFERENCES

1. **Clark, S.J. and M. Frommer.** 1995. Deamination with NaHSO₃ in DNA methylation studies, p. 123-135. *In* H.P. Saluz and K. Weibauer (Eds.), DNA and Nucleoprotein Structure In Vivo. R.G. Landes, Austin.
2. **Clark, S.J., J. Harrison and M. Frommer.** 1995. CpNpG methylation in mammalian cells. *Nature Genet.* 10:20-28.
3. **Clark, S.J., J. Harrison, C.L. Paul and M. Frommer.** 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22:2990-2997.
4. **Frommer, M., L.E. McDonald, D.S. Millar, C.M. Collis, F. Watt, G.W. Grigg, P.L. Molloy and C.L. Paul.** 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* 89:1827-1831.
5. **Grigg, G.W. and S.J. Clark.** 1995. Sequencing 5-methylcytosine residues in genomic DNA. *Bioessays* 16:431-436.
6. **Myohanen, S., J. Wahlfors and J. Janne.** 1994. Automated fluorescent genomic sequencing as applied to the methylation analysis of the human ornithine decarboxylase gene. *DNA Seq.* 5:1-8.
7. **Parker, L.T., Q. Deng, H. Zakeri, C. Carlson, D.A. Nickerson and P.Y. Kwok.** 1995. Peak height variations in automated sequencing of PCR products using Taq dye-terminator chemistry. *BioTechniques* 19:116-121.
8. **Patel, S.S., I. Wong and K.A. Johnson.** 1991. Pre-steady-state kinetic analysis of processive DNA replication including characterisation of an exonuclease-deficient mutant. *Biochemistry* 30:511-525.

Received 10 November 1995; accepted 21 December 1995.

Address correspondence to:

Susan J. Clark
Kanematsu Institute
CSIRO, Division of Biomolecular Engineering
Sydney Laboratory, P.O. Box 184
North Ryde, NSW 2113, Australia
Internet: clark@pelican.dbe.csiro.au