

SURVEY AND SUMMARY

Identifying 5-methylcytosine and related modifications in DNA genomes

Theo Rein, Melvin L. DePamphilis and Haralabos Zorbas^{1,*}

National Institute of Child Health and Human Development, Building 6, Room 416, National Institutes of Health, Bethesda, MD 20892-2753, USA and ¹Institute of Biochemistry, Ludwig-Maximilian University of Munich, Feodor-Lynen-Strasse 25, D-81377 Munich, Germany

Received January 7, 1998; Revised and Accepted March 30, 1998

ABSTRACT

Intense interest in the biological roles of DNA methylation, particularly in eukaryotes, has produced at least eight different methods for identifying 5-methylcytosine and related modifications in DNA genomes. However, the utility of each method depends not only on its simplicity but on its specificity, resolution, sensitivity and potential artifacts. Since these parameters affect the interpretation of data, they should be considered in any application. Therefore, we have outlined the principles and applications of each method, quantitatively evaluated their specificity, resolution and sensitivity, identified potential artifacts and suggested solutions, and discussed a paradox in the distribution of m⁵C in mammalian genomes that illustrates how methodological limitations can affect interpretation of data. Hopefully, the information and analysis provided here will guide new investigators entering this exciting field.

INTRODUCTION

At least seven different covalent base modifications have been identified at significant levels in the DNA genomes of prokaryotic and eukaryotic cells, bacteriophage and viruses, resulting in an intense effort to identify their biological significance (1–4). For example, 5-methylcytosine (m⁵C), the most abundant covalently modified base in the genomes of eukaryotic cells, plays a role in the regulation of gene transcription, X chromosome inactivation, genomic imprinting, cell differentiation and tumorigenesis. Unicellular eukaryotes also contain N⁶-methyladenine (m⁶A) and 5-hydroxymethyluracil (hm⁵U) (5,6). m⁶A and N⁷-methylguanine (m⁷G) have been reported in insects and humans (7), but their biological significance is not yet evident. In prokaryotic cells, the most prevalent covalent modifications are m⁶A and m⁵C, although N⁴-methylcytosine (m⁴C) also has been detected (8). All of these modifications are involved in restriction and modification of DNA, and m⁶A is involved specifically in the regulation of DNA replication and in DNA repair.

These and other modified bases can constitute a large part of the genome. For example, all of the cytosines in bacteriophage T4

DNA are either 5-hydroxymethylcytosine (hm⁵C) or glucosylated hm⁵C, and up to 50% of the cytosines in plant DNA are m⁵C (9,10). Modified DNA bases also can represent a minor fraction of the genome, but still exhibit strong biological effects. For example, only 3–10% (11) of cytosines in mammalian genomes are m⁵C, but they generally repress transcription when clustered at the 5'-ends of genes (3,12). In general, m⁵C is nearly universal among eukaryotes with genomes >10⁸ bp (e.g. mammals and plants), but rare among eukaryotes with smaller genomes (e.g. yeast, flies and nematodes) (13–15).

The goal of this review is to provide a guide for identifying m⁵C and related covalent base modifications in DNA genomes. At least eight different methods, along with several variations, have been developed over the past three decades for characterizing covalently modified bases in DNA genomes. Each method has advantages and disadvantages, most of which become evident by examining their specificity, resolution, sensitivity and potential artifacts. Specificity is the ability to distinguish a particular covalent base modification either from another modification or from the unmodified base. Horizontal resolution is the ability to identify the position of a modified base along a DNA strand. Some methods, such as total base composition and nearest neighbor analysis, can determine the occurrence and abundance of virtually any modified base within an entire genome. Other methods, such as differential base modification by bisulfite, hydrazine or permanganate followed by DNA sequencing, can detect only m⁵C, but these methods can map m⁵C to a precise nucleotide position in virtually any stretch of DNA. Thus, the two methods that identify the widest range of specific base modifications also have the lowest horizontal resolution, while the three methods that have the highest horizontal resolution are specific only for m⁵C. This is not a problem when analyzing the genomes of animals or plants where the exclusive presence of m⁵C has been established by analyses of total base composition. However, for genomes with other modifications, these methods are of limited use. These methods also are too laborious for screening large amounts of DNA sequence. Both problems can be addressed to some extent using modification-sensitive restriction endonucleases (MSREs). MSREs provide a relatively simple method for mapping methylated cytosines and adenines to specific DNA sites, and this strategy has been adapted to large scale screening procedures that can map the genomic locations of modified bases in specific restriction endonuclease recognition sequences. However, horizontal resolution is

*To whom correspondence should be addressed. Tel: +49 89 74017450; Fax: +49 89 74017448; Email: zorbas@lmb.uni-muenchen.de

reduced, because the number of sites that can be examined is limited by the number of sequences recognized by MSREs and by the availability of the appropriate enzyme. Another method, immunochemical analysis, may also help address these problems, because in principle, antibodies can be developed against most, if not all, covalently modified bases. However, horizontal resolution is reduced even further.

Three other important parameters in selecting a method are the minimum amount of DNA required to detect a particular base modification (sensitivity), the ability to identify a base modification when it is contained in only a fraction of the population (vertical resolution), and potential artifacts. For example, the hydrazine and permanganate methods compare the relative strengths of cytosine bands in a genomic sequence with those from an unmodified control sequence. While methods such as ligation-mediated PCR can improve their sensitivity (less DNA is required to see the genomic sequence), their vertical resolution (the ability to measure the relative amount of two signals) remains essentially the same, because the signal to noise ratio remains essentially the same. Some methods are subject to potential artifacts that affect interpretation of data. For example, failure to completely denature the DNA and improper design of PCR primers in the bisulfite method can lead to false identification of m⁵C, and resistance of DNA to cleavage by a MSRE can result from factors other than methylated cytosines and adenines.

In an effort to help the new investigator select the methods most appropriate for a particular application, we have listed them in the order they are frequently applied to a genome with unknown DNA modifications. The experimental protocols are found in the literature cited. Wherever possible, we have provided quantitative evaluations of the sensitivity and resolution of each method, based on our own experience and on published information. In addition, we have described the principle of each method, its potential artifacts, and its simplicity of application. Together, these parameters determine a method's utility and accuracy. Finally, we have attempted to illustrate problems that may be encountered when applying these methods by examining the distribution of m⁵C in mammalian genomes.

IDENTIFYING COVALENTLY MODIFIED BASES AND THEIR DINUCLEOTIDE COMPOSITION

Identification of the covalently modified bases present in a particular genome and determining their abundance is prerequisite to mapping their locations in specific DNA sequences. The methods available for sequence analyses can identify only a limited number of modified bases, and with the exception of MSREs, are too labor intensive to screen more than a few kilobases of sequence. To obtain the total base composition of a genome, ~10 µg of DNA has to be hydrolyzed to completion. Originally, this was achieved by chemical means, but this method produced a complicated array of products that made detection and quantification of specific adducts difficult. These problems can be avoided by enzymatic hydrolysis using calf spleen phosphodiesterase and micrococcal nuclease to produce 3'-phosphorylated mononucleotides, or pancreatic DNase I and snake venom phosphodiesterase to produce 5'-phosphorylated mononucleotides (16,17). Following hydrolysis, 3'- and 5'-phosphates are removed with alkaline phosphatase (18), and the products are fractionated using standard chromatographic and electrophoretic

techniques along with external standards of known modified bases (19–23). The occurrence of m⁵C and m⁶A has been quantified using high pressure liquid chromatography (HPLC) (24,25). This method has a sensitivity of ≤10 µg DNA (24) and a vertical resolution of 0.04% (24,25) to 0.005% (26). Mass spectrometry also has been applied to quantify the content of m⁵C in DNA (27), detecting as little as one m⁵C in 10 kb using 1–10 µg of DNA. However, total genome composition is devoid of sequence information.

Limited sequence information in the form of dinucleotide composition is easily obtained by nearest-neighbor analysis. About 5 µg of DNA is labeled with one of the four [α-³²P]dNTP by nick translation at randomly generated single-stranded breaks or by DNA replication in a cell lysate (9,28). The DNA is digested completely to 3'-dNMPs using micrococcal endonuclease and calf spleen phosphodiesterase exonuclease in order to transfer ³²P from the 5'-position of the labeled nucleotide to the 3'-position of its neighbor. The resulting [3'-³²P]dNMPs are fractionated by chromatography or HPLC and quantified by comparison to internal standards (18). Vertical resolution is ~0.01% (29).

Alternatively, genomic DNA can be labeled extensively at m⁵C-positions *in vivo* with [³H-CH₃]methionine for detecting the nearest neighbors of m⁵C. Dinucleotides OH-NpN-OH are generated from the labeled DNA by limited digestion with DNase I and 5'-dephosphorylation with alkaline phosphatase. Dinucleotides containing ³H-labeled m⁵C are isolated by chromatography and paper electrophoresis, and the two isomers (OH-m⁵CpN-OH and OH-Np-m⁵C-OH) are distinguished by treatment with snake venom phosphodiesterase which generates OH-m⁵C-OH and pN-OH from the first isomer, and OH-N-OH and pm⁵C-OH from the second isomer. With this method, 2.5–5% m⁵C with a specific dinucleotide composition can be detected in 100 ng of DNA (30).

These two methods are specific for all known covalently modified bases, providing that a sample of the modified base is available as a standard for comparison. Although vertical resolution is excellent, horizontal resolution is limited to dinucleotide frequencies. The amount of DNA required for these methods limits their application to tissues or cell cultures that can provide ~2 million cells. Since neither method localizes a modified base within a specific DNA sequence, neither method can distinguish genomic DNA from contamination by either RNA or foreign DNA from viruses, mycoplasmas and other endoparasites. This potential artifact can lead to false identification of base modifications in small samples of genomic DNA.

MAPPING METHYLATED CYTOSINES AND ADENINES AT SPECIFIC DNA SITES

There are currently four methods available that can map specific covalent base modifications to specific DNA sites: modification-sensitive restriction endonucleases (MSRE) and differential base modification by either bisulfite (HSO₃⁻), hydrazine (N₂H₄) or permanganate (MnO₄⁻) followed by DNA sequence analysis. Each method has its own advantages and limitations such that a single method alone is often inadequate to address all problems. However, when used in conjunction with one another, these four methods can provide unambiguous identification of m⁵C, map its location with nucleotide resolution in any DNA sequence, and quantify the frequency it appears at a specific DNA site. Therefore, we have summarized their characteristics in Table 1, and outlined their protocols in Figures 1–3.

Table 1. Methods for mapping methylated bases at specific DNA sites in complex DNA genomes

Characteristic	MSRE	HSO ₃ ⁻	N ₂ H ₄	MnO ₄ ⁻ (pH 4.1)
Principle	cleavage sensitive to m ⁶ A, m ⁵ C, m ⁴ C, hm ⁵ C, glucosylated hm ⁵ C	m ⁵ C resistant to conversion to U	m ⁵ C resistant to hydrazinolysis	m ⁵ C sensitive to oxidation
Detection	Southern blotting/ hybridization or PCR	PCR and sequencing of total product or individual clones	direct genomic sequencing (LM-PCR)	direct genomic sequencing (LM-PCR)
Reactivity	cleavage site sequence	C	C + T	[m ⁵ C + T] > G
Substrate	dsDNA	ssDNA	dsDNA or ssDNA	ssDNA
Requirements	complete reaction	complete conversion of Cs	partial hydrazinolysis	partial oxidation
Horizontal resolution	limited to cleavage sites	nucleotide sequence	nucleotide sequence	nucleotide sequence
Vertical resolution	10% (Southern); 0.1% (PCR)	single molecule (DNA cloning) 0.1–10% (total PCR product) ^a	25%	10%
Sensitivity	10 µg (Southern); ≤10 ng (PCR)	≤10 ng	1–2 µg	1–2 µg
Advantages	rapid analysis of large DNA regions; high sensitivity and vertical resolution	highest sensitivity and vertical resolution; rapid genomic sequencing; positive display of m ⁵ C	no serious artifacts	m ⁵ C reactivity Positive display of m ⁵ C
Potential artifacts	incomplete cleavage	undenatured islands; cloning artifacts; incomplete C → U; incomplete m ⁵ C resistance	suppression of C bands	none described

Abbreviations are MRSE (modification-sensitive restriction endonucleases), HSO₃⁻ (bisulfite), N₂H₄ (hydrazine), and MnO₄⁻ (permanganate).

^a0.1% for methylation-specific PCR (60), 1% for COBRA (61) and methylation-specific single nucleotide primer extension (62), 10% for GENESCAN (58), and 5–10% for manual sequencing (51) of the total PCR product.

Modification-sensitive restriction endonucleases (MSREs)

MSREs provide the most convenient and rapid method for identifying modified bases at specific restriction endonuclease sites in virtually any genomic region. This method requires only a map of restriction sites in the region of interest, and can be used to survey large regions of DNA. While MRSEs provide a broader range of specificity than the differential base modification methods described below, their limitation to specific restriction sites reduces their horizontal resolution. Therefore, the MRSE method is frequently applied first in an effort to ascertain whether or not modified bases are likely to be found at biologically interesting sites such as transcriptionally active genes, replication origins and recombinational hot-spots.

More than 320 restriction endonucleases are sensitive to base modifications that lie within the enzyme's DNA recognition site (31). Most of these will not cut DNA if their cleavage site contains a methylated base, and in general, do not discriminate between m⁵C, m⁴C (32), hm⁵C (33) or glucosylated hm⁵C (33). However, there are some exceptions. *MvaI*, *BstNI*, *RsaI*, *KpnI* and *BstYI* (31,34) do not cut DNA if their recognition sequence contains m⁴C, but sites containing m⁵C remain sensitive (31). Similarly, *CviSIII* is inhibited by hm⁵C, but not by m⁵C. Alternatively, some enzymes cut their sites only when they contain m⁵C or hm⁵C. For example, *PvuRtsII* only cuts DNA containing hm⁵C (35).

McrBC cuts DNA at multiple sites within RC(N_{40–80})RC only when the outer cytosines are m⁵C, m⁴C or hm⁵C (36). Thus, the usefulness of *McrBC* decreases as the genomic density of m⁵C increases. Other enzymes either require m⁶A in order to cut DNA, or are inhibited by this modification (31). It should be noted that only a few restriction endonucleases have been tested for their sensitivity to modified bases that lie outside their recognition sites, and some of these are inhibited (31).

Two methods are commonly used to determine the extent of DNA digestion. The most direct method is to fractionate the DNA digestion products by gel electrophoresis (preferably agarose, because DNA may migrate aberrantly in polyacrylamide; 37) and then identify specific genomic sites by Southern blotting-hybridization using standard protocols (38; Fig. 1). In the case of an endonuclease that cannot cut a methylated site, absence of the expected DNA cleavage product indicates methylation at one or both of the endonuclease cleavage sites that mark the ends of the DNA fragment. The fraction of resistant DNA equals the fraction of cells that contain this modified site. Vertical resolution depends on the hybridization background (usually 10–20% of signal). Thus, if ≥10% of the DNA population is not modified at two consecutive restriction endonuclease sites, the corresponding DNA fragment will be detected. If ≥10% of the DNA population is modified at one or both sites, then a larger DNA fragment will appear as a result of cleavage events outside the region of interest

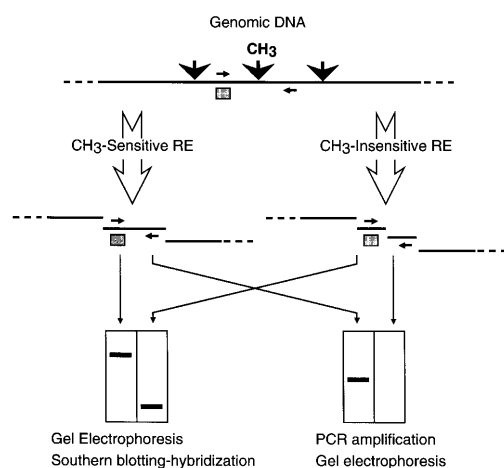


Figure 1. Methylation-sensitive restriction endonuclease protocol in combination with either Southern blotting-hybridization or PCR. Restriction enzyme (RE) sites (black downward arrows), one of which is methylated (CH₃), a hybridization probe (shaded square) and PCR primers (horizontal arrows) are indicated.

(Fig. 1). A single copy locus can be detected in ~10 µg of mammalian genomic DNA (~2 million cells) using ³²P-labeled oligonucleotide probes (39).

The sensitivity of this assay can be increased ~1000-fold and the vertical resolution ~100-fold using the DNA polymerase chain reaction (PCR) to amplify specific DNA segments before fractionating them by gel electrophoresis so that the DNA products can be visualized by ethidium bromide staining (40), but this requires sequence information in order to design PCR primers that flank the restriction site in question and thereby amplify the region of interest (Fig. 1). The PCR product is compared to a DNA fragment produced by amplification of an uncut control DNA. However, quantification (i.e. vertical resolution) now depends on obtaining PCR conditions that produce a linear response between input DNA and amplified product. This problem could be solved by using competitive PCR (41) where the same PCR primers amplify simultaneously both the genomic target and a homologous competitor. Alternatively, ligation-mediated PCR (LM-PCR; 42,43) could be used where the same PCR primers are employed to visualize simultaneously the relative extent of cleavage at both an MSRE site and a modification-insensitive restriction endonuclease site (44,45). This method requires only 0.6 ng of DNA (100 cells) for detection and 50 ng for a quantification (45).

When the DNA digestion products are fractionated by 2D gel electrophoresis, the MSRE method can be used to visualize simultaneously and quantitatively the methylation status of a large number of genomic loci ('restriction landmark genomic scanning of methylation sites'; 46). In fact, an entire genome can be screened for sites where DNA methylation patterns change as cells undergo differentiation, carcinogenesis and genomic imprinting (47,48), and then these sites can be cloned. A similar end-labeling method has been developed specifically for rare DNA modifications (49).

Potential artifacts in the MSRE method

The principle artifact that can affect the MSRE method is inefficient DNA cleavage, resulting in the false conclusion that some or all of

the restriction sites contain a covalently modified base. This problem generally results from impure DNA and can be corrected by its repurification. Careful extraction of genomic DNA in the presence of proteinase K and SDS followed by phenol extraction and extensive dialysis (no ethanol precipitation) usually result in complete DNA cleavage (39). If not, addition of 1 mM spermidine often alleviates the problem (39). Accessibility of DNA to cleavage can be checked using a modification-insensitive restriction endonuclease, ideally an isoschizomer of the modification-sensitive enzyme under identical reaction conditions. However, two enzymes can be affected differently by impurities in the DNA sample. For example, an unidentified, dialyzable inhibitor of the MSRE *AluI* can prevent this enzyme from completely cutting genomic DNA under conditions where the methylation-insensitive restriction enzyme control, *MboII*, can digest DNA to completion (39). Therefore, one should include in the same reaction mixture another DNA fragment, best radioactively labeled, with the same restriction site, but whose size or sequence allows it to be distinguished from the genomic target. In this way, the extent of cleavage by the diagnostic enzyme can be monitored in the same reaction simply by monitoring cleavage of the control fragment (39). Unfortunately, even this control is not foolproof. There have been reports of proteins that bind so tenaciously to DNA that they could not be removed during DNA purification, and the ubiquitous topoisomerases form covalent intermediates with DNA that can be trapped during DNA purification. Therefore, if the modified base is thought to be m⁵C, a prudent investigator will confirm its presence by one of the three methods described below.

Differential base modification by bisulfite

The bisulfite method has four advantages over the other two differential base modification methods (described below) for detection and mapping of m⁵C in any sequence of a complex genome (Table 1). First, the bisulfite method is the easiest to apply. Second, it is the most sensitive, requiring no more than 10 ng of genomic DNA (~2000 cells). In one case, 10 pg of genomic DNA (~2 cells) was sufficient (50), although the number of integrated target copies per cell in this transfection experiment was unknown. Third, both C and m⁵C residues appear during DNA sequencing (positive display of data). This makes it easy to quantify the fraction of m⁵C, which can be detected in as little as 5% of the total PCR product (51). In principle, vertical resolution is limited only by the number of cloned PCR products one is willing to sequence. Fourth, the methylation status of a specific genomic site can be determined both for the total cell population and for individual cells, depending on whether the total PCR DNA product is sequenced directly or individual cloned molecules from the PCR product are sequenced. This allows detection of subtle variations in cytosine methylation at specific genomic sites that may occur within a population of cells as they undergo differentiation and development (52), thus permitting detection of rare events, and analysis of partial methylation patterns. For example, if two cytosines at nearby positions are found to be methylated to an overall extent of 50% by sequencing the total PCR product, the following question arises: do both m⁵Cs sit on the same molecule (e.g. when the two m⁵Cs are only on one homologous chromosome of all cells, or on both homologous chromosomes in only 50% of the cells), or are they distributed on different molecules (e.g. when the two m⁵Cs are distributed on the two homologous chromosomes in all cells, or when one m⁵C is

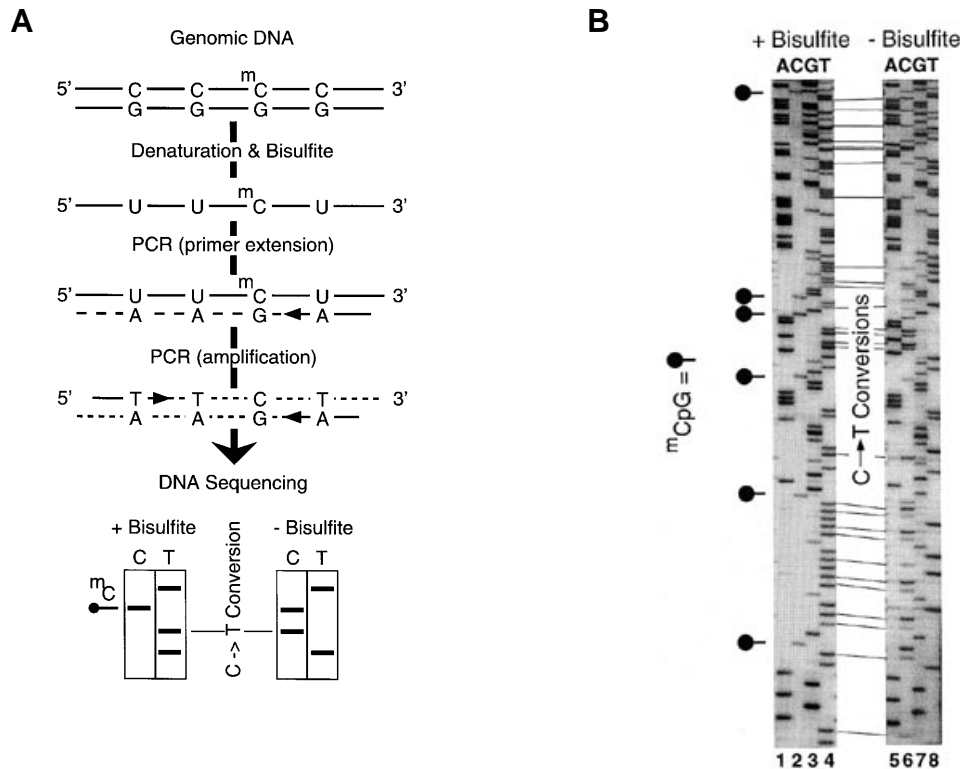


Figure 2. Bisulfite method. (A) Protocol. (B) Detection of $m^5\text{CpG}$ dinucleotides in genomic DNA from hamster cells as outlined in (A) and detailed in ref 51.

found on both homologous chromosomes in one half of the cells and on the other in the other half of the cells)? This question can be approached by the bisulfite technique, after cloning the PCR product. This information can be critical in determining whether or not $m^5\text{C}$ plays a significant role in regulating the activity of a specific promoter, replication origin, transposable element or imprinting element.

The bisulfite method is considered specific for identification of $m^5\text{C}$. Only cytosines that are present in single-stranded DNA or in 'distorted double-stranded DNA' regions can be converted to uracil by bisulfite; properly paired cytosines are not affected (53). Bisulfite sulfonation at C-6 of a susceptible cytosine facilitates spontaneous hydrolysis of the amino group at C-4 to produce sulfonated uracil (53). Uracil is recovered by removing unreacted bisulfite and then desulfonating under alkaline conditions. $m^5\text{C}$ is not converted. The reactivity of glucosylated $hm^5\text{C}$ and $m^4\text{C}$ with bisulfite remains to be determined. Probably $hm^5\text{C}$ cannot be detected by the bisulfite method, because bisulfite converts $hm^5\text{C}$ into a stable product (cytosine 5-methylenesulfonate; 54) that, by analogy to the inhibitory effect of pyrimidine adducts with permanganate or osmiumtetroxide on DNA polymerase elongation (e.g. 55), should inhibit PCR amplification.

In the standard method (Fig. 2A), genomic DNA is denatured and treated with bisulfite (56). The region of interest is then amplified by PCR to produce specific double-stranded DNA fragments. Since the DNA strands are no longer complementary after bisulfite conversion, each strand must be analyzed separately using appropriate PCR primers. PCR amplification results in conversion of U (previously C) to T, and of $m^5\text{C}$ to C. The former PCR amplified product will thus contain T:A base pairs in place

of C:G base pairs. The PCR product can be isolated by gel electrophoresis and the conversions detected by standard DNA sequencing protocols (51; Fig. 2). Unreacted C (presumed to be $m^5\text{C}$) is seen as a positive band in the C lane, while reacted C appears as a band in the T lane. These data represent the average methylated state of particular cytosines within the DNA population (Fig. 2B). Alternatively, the PCR products can be cloned into a plasmid vector and individual clones sequenced (56).

Innovations and variations in the bisulfite method

A number of technical innovations can improve the speed and versatility of the bisulfite method. Biotinylated primers can be used for PCR amplification so that PCR products can be purified using streptavidin coated magnetic beads and DNA sequencing can be automated (57). This procedure has been combined with the specialized sequencing analysis called GENESCANTM to quantify the extent of methylation at each cytosine position within a sequence (58). In addition, the methylation status can be rapidly screened at new restriction endonuclease sites that are created by conversion of cytosines to thymines (59).

Methylation-specific PCR can be used to determine rapidly the methylation status of CpG islands with a vertical resolution of 0.1% (60). The high density of CpG dinucleotides in these islands allows PCR primers to be designed that specifically amplify either $m^5\text{CpG}$ DNA or CpG DNA by making them complementary either to a sequence containing several C residues that are presumed to remain unconverted, or to a sequence in which several cytosines are presumed to have been converted to uracils (60). Primer specificity is verified by cutting the PCR amplification product

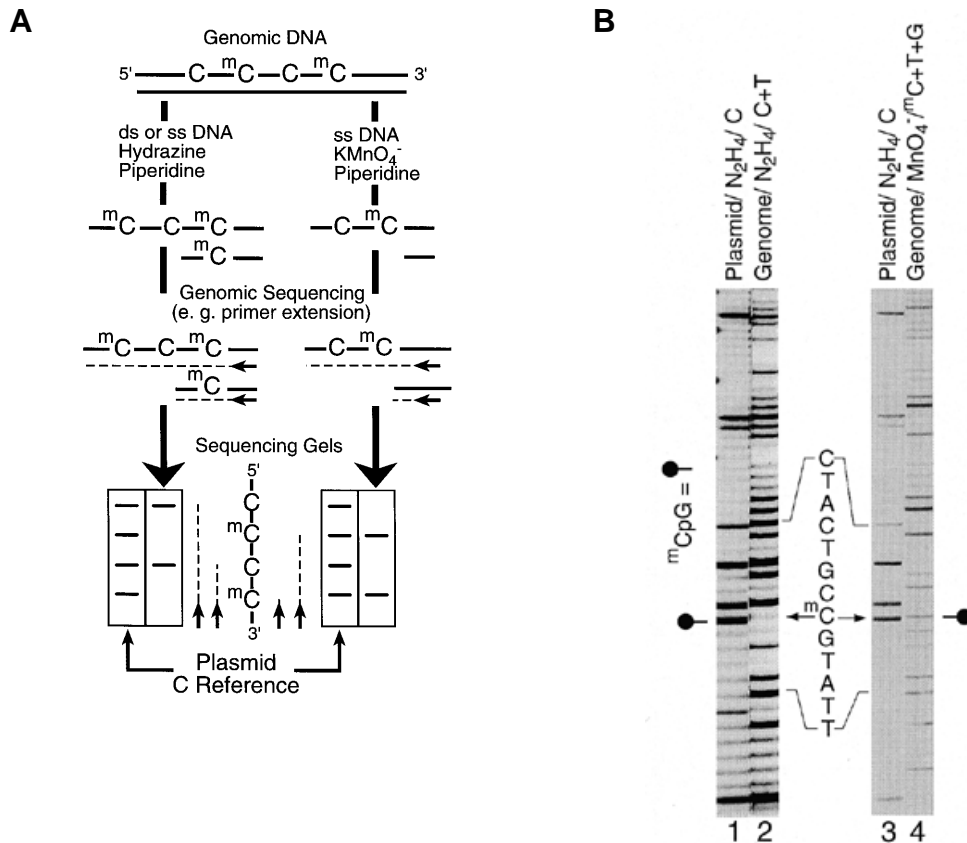


Figure 3. Hydrizine and permanganate methods. **(A)** Protocols. **(B)** Detection of a ^m5CpG dinucleotide in genomic DNA from hamster cells as outlined in **(A)** and detailed in ref 39. DNA was treated with hydrazine (lane 2) or permanganate (lane 4) followed by piperidine cleavage of the modified nucleotides. Cleaved sites were detected using LM-PCR. The same DNA sequence cloned into a plasmid (unmethylated DNA) was sequenced using hydrazine/piperidine and LM-PCR (lanes 1 and 3).

with restriction endonucleases, because the DNA cleavage products will differ for DNA products amplified from methylated versus unmethylated DNA. However, since PCR products are not sequenced, analysis is not at the nucleotide level.

A similar strategy has been used to rapidly quantify methylation at specific sites in any DNA sample. 'Combined bisulfite restriction analysis' (COBRA; 61) uses restriction enzymes to cut the PCR product from bisulfite converted DNA. For example, bisulfite will convert the *Bst*UI cleavage site (CGCG) in unmethylated DNA to TGTG, but not in methylated DNA. Therefore, the percentage of PCR amplified molecules that are cut by *Bst*UI reflects the percentage of methylated molecules. COBRA is fast, sensitive and quantitative, but quantification relies on complete enzyme digestion and analysis is confined to restriction endonuclease cleavage sites in DNA.

An alternative approach is 'methylation-sensitive single nucleotide primer extension' (62). Genomic DNA is treated with bisulfite and then amplified by PCR. Methylation at a specific cytosine is analyzed by extension of an internal primer, located specifically 5-prime to that cytosine, with a single, radioactive nucleotide. The ratio of the extension product using dCTP (reflecting unconverted ^m5C) versus the extension product using dTTP (reflecting converted C) measures the ratio of methylation at a specific site. This method does not rely on restriction enzymes and, in principle, can be applied to any sequence. However, this method will be difficult to apply to regions rich in CpG

dinucleotides, because the primers will invariably contain CpGs of unknown or mixed methylation status.

Potential artifacts in the bisulfite method

There are four potential artifacts that can affect the reliability of the bisulfite method. The following information will help to avoid them.

Conversion of C to U can be incomplete. Since conversion of C to U requires that the DNA substrate be single-stranded, incomplete denaturation of genomic DNA or its partial renaturation during bisulfite treatment can result in C residues that fail to react with bisulfite (51). Unfortunately, the high salt molarity in the reaction favors renaturation. In addition, bisulfite treatment must be exhaustive in order to ensure complete conversion of C to U in single-stranded DNA. Incomplete conversion would appear as a partially methylated site. In practice, complete bisulfite conversion is not feasible—even if the DNA is kept single-stranded—because the DNA substrate undergoes degradation with time (see below). Note that in studies of secondary structure of nucleic acids, usually only a fraction of the C residues in a single-stranded region are converted to U due to incomplete deamination by bisulfite (63). Incomplete conversion is not a problem when the total PCR product is sequenced, because the occasionally unconverted Cs will not significantly affect detection of the average extent of methylation at a specific site. It can

become a major problem, however, if unconverted DNA is selected for amplification by a poor choice of PCR primers. Primers that anneal to sequences containing several C or m⁵C residues are selecting for regions of unconverted DNA. It can also be a problem when individual clones are sequenced, because an insignificant event may appear significant if it shows up in a relatively small number of molecules. Some plasmid vectors appear prone to this artifact (64). In fact, in three examples where all C residues were thought to be methylated on both strands, regardless of their dinucleotide composition (65,66), subsequent analyses by four independent methods demonstrated that all of the non-CpG methylation events in these regions represented cytosines that had failed to react with bisulfite (39,51).

Incomplete conversion can be dealt with in several ways (51). The extent of conversion of C to U can be checked by amplifying the target DNA with PCR primers designed to hybridize to a region devoid of C. These primers will not select for either converted or unconverted DNA. Alternatively, one can selectively amplify DNA that has reacted efficiently with bisulfite by designing PCR primers to anneal to sequences containing U in place of C, and by choosing sequences that originally contained several Cs. The presence of non-methylated cytosines interspersed with methylated cytosines indicates that this region of DNA was accessible to bisulfite during the reaction. The efficiency of the bisulfite reaction can be checked using DNA whose methylation status is known (e.g. plasmids or PCR generated fragments). Whenever possible, this control should be run in the same reaction tube, and the copy number of the control DNA should be comparable to that of test DNA sequences; otherwise its rate of renaturation could be artificially high. Finally, other methods that specifically rely upon partial modification of DNA, such as the hydrazine or permanganate methods discussed below can be independently applied to analyze the same region of DNA in order to confirm the presence of m⁵C.

To eliminate the problem of incomplete conversion a protocol has been developed that ensures ≥95% conversion of C to U with limited DNA degradation (51) by cleaving genomic DNA into ~1 kb fragments in order to reduce the likelihood of renaturation, and treating the DNA in bisulfite with repeated cycles of heating to 95°C in a thermocycler for only 5 h. Several changes of the original protocol have been reported (56) to reduce the chances of renaturation. For example, ethanol precipitation has been omitted after initial denaturation with NaOH (64,67). The DNA sample also has been diluted (67), embedded in agarose (68), incubated at 0°C instead at elevated temperature (67) during the bisulfite conversion, and the bisulfite concentration has been increased (69).

From 2 to 3% of m⁵Cs can be converted to T (56,70). This results in an underestimate of the number of m⁵Cs when individual DNA clones are analyzed, but assuming that this loss occurs randomly, it should not be significant when the total PCR product is sequenced directly.

Loss of DNA due to fragmentation. Incubation of DNA at a slightly acidic pH can generate apurinic sites (71). Since the bisulfite treatment must be performed for a long time at pH 5, apurinic sites will be introduced into the DNA substrate. These apurinic sites will then be broken in a classical Maxam and Gilbert reaction during desulfonation of deaminated cytosines at the required basic pH (71). Protocols have been developed that minimize this problem (51,67,68,72).

Biased PCR amplification. Quantification of methylated alleles can be jeopardized by biased PCR amplification (73). Surprisingly, the same primer pair can preferentially amplify either the methylated or unmethylated sequence, even though the sequence to which the primers anneal and the lengths of the PCR products are identical. The bias is less serious with the Stoffel fragment polymerase than with *Taq* polymerase, but the most reliable solution is to establish standard curves for each primer pair by measuring the ratio of products from known mixtures of methylated and unmethylated DNA standards in control reactions.

Differential base modification by hydrazine

With genomes of low sequence complexity such as viruses, the hydrazine and permanganate methods are easier to apply than the bisulfite method, because they employ simple primer extension or end-labeling to sequence directly the differentially modified DNA. In contrast, the hydrazine and permanganate methods are more difficult to apply to genomes of high sequence complexity such as mammals, because they require sequencing the differentially modified DNA using LM-PCR (74,75). This generally leads to lower sensitivity and higher background with its attendant lower vertical resolution (Table 1). Moreover, it often proves difficult to identify primers that are effective in LM-PCR. However, once LM-PCR is established at a sequence of interest, the same primers can also be used in methods designed to explore protein/DNA interactions such as *in vivo* footprinting. In addition, hydrazine or permanganate provide independent confirmation of results in cases where artifacts are suspected using the MSRE or bisulfite methods described above.

Hydrazine reacts specifically with C and T in either double-stranded or single-stranded DNA. Hydrazine also reacts with m⁴C (34) but not with m⁵C (76,77). Presumably hydrazine also does not react with hm⁵C or glucosylated hm⁵C, but it is not clear that this has been examined. Hydrazine modified nucleotides are usually cleaved by piperidine (71), and the cleavage products are detected by a variety of direct genomic sequencing techniques (discussed below). An unmethylated sequence, obtained either by DNA cloning or by PCR amplification of the genomic region of interest, is analyzed in parallel in order to locate the positions of all cytosines. m⁵C is identified by the absence of a cytosine from the genomic DNA sequence (Fig. 3). Sequencing the complementary DNA strand insures that absence of a band results from m⁵C rather than from some unidentified artifact. Hydrazine reaction with T is normally suppressed by addition of 1.5 M NaCl to facilitate identification of C (76). However, in regions where all of the cytosines appear to be methylated, T reactivity at low salt concentrations provides a useful internal standard for monitoring the efficiency of the hydrazine reaction (39).

Differential base modification by permanganate

The same direct DNA sequencing methods used to display hydrazine/piperidine cleavage sites are also used to display permanganate/piperidine cleavage sites (Fig. 3A). However, while hydrazine produces a negative data display [bands disappear from the sequencing gel at the positions of m⁵C (Fig. 3A and B, lane 2)], permanganate produces a positive data display [bands appear in the sequencing gel at the positions of m⁵C (Fig. 3A and B, lane 4)]. Therefore, the permanganate method provides the perfect complement to the hydrazine method for detection of m⁵C. At pH 4.1, permanganate reacts strongly with m⁵C and T in

single-stranded DNA (78), but not with C (79,39) (Fig. 3B, lane 4). Permanganate occasionally reacts with purines (G>>A) as a result of either direct oxidation or acidic depurination of DNA creating piperidine-labile sites (79). Permanganate, like bisulfite, reacts only with single-stranded DNA. However, this is not a problem, because oxidation of T by permanganate can serve as an internal control to insure that the sequences displayed were single-stranded.

LM-PCR analysis of genomic DNA cleaved by permanganate/piperidine resulted in nearly equivalent signals from m⁵C, T and G (Fig. 3B). This unexpected G reactivity relative to an end-labeled DNA fragment that was not subjected to LM-PCR (39) may have resulted from enhanced depurination caused by the initial heating at 95 °C that was necessary in LM-PCR. Nevertheless, permanganate did not react with any C residues in this region, but did react with all T and m⁵C residues (Fig. 3B).

Direct genomic sequencing employs either linear or exponential amplification to visualize both hydrazine and permanganate reacted bases. In general, linear amplification requires large amounts of DNA (~200 µg; 80–82). LM-PCR which employs exponential amplification (42,43) allows detection of m⁵C in 1–2 µg of genomic DNA (~3 × 10⁵ cells). When both C and m⁵C are present at the same site but on different DNA molecules, at least 25% of that site must be m⁵C to be detected by the hydrazine method. However, as a result of a positive data display, only 10% must be m⁵C for detection by the permanganate method (39). Both permanganate and hydrazine modifications of genomic DNA could be detected by genomic sequencing techniques that do not require piperidine cleavage of DNA (55,81–83), because they inhibit elongation of DNA polymerases.

No serious artifacts have been reported for the hydrazine and permanganate methods, although minor problems can lead to ambiguity. These include background cleavage events, closely spaced bands on the sequencing gel, and DNA concentration-dependent suppression of C-bands that have been reported for the hydrazine method (84). However, they can be eliminated by including sequencing controls to verify the positions of all cytosines.

METHODS WITH SPECIAL APPLICATIONS

Immunochemical analysis

Antibodies specific for methylated bases could be used as an independent method for confirming that resistance to cleavage by an MSRE is due to DNA methylation and not to artifactual resistance. For example, antibodies raised against m⁵C have been shown to identify that portion of the DNA that was resistant to *HpaII*, a MSRE that only cuts unmethylated CCGG sites (85). These antibodies can detect m⁵C with a vertical resolution of ≤0.008 mol% (85). They can react specifically with m⁵C in mammalian DNA bound to nitrocellulose paper (15,86), detecting ~1 mol% m⁵C in the human genome (15), consistent with estimates from total genome and nearest neighbor analyses. Similar experiments could be carried out with antibodies against m⁶A and m⁷G that have been reported to react with DNA from human, *Drosophila* and mealybugs (7). Immunofluorescence also has been used to determine chromosomal regions with a high frequency of m⁵C (87). Therefore, it might be possible to use antibodies to map the locations of clusters of m⁵CpG (clusters) in cellular chromosomes relative to the locations of specific chromatin associated

proteins, genes (88), replication origins, centromeres and telomeres using a second antibody or fluorescence *in situ* hybridization.

Differential base modification by UV radiation

UV radiation causes formation of pyrimidine (6-4) pyrimidone photo adducts between two adjacent pyrimidine bases (pyrimidine dimers) that are sensitive to cleavage with piperidine. In mammalian genomes, these cleavage sites have been mapped using LM-PCR to sequence the DNA (89). TpC and CpC dinucleotides produce a high frequency of pyrimidine dimers, but Tpm⁵C and Cpm⁵C dinucleotides do not (89). Therefore, wherever a m⁵CpG is preceded by a T or C, a pyrimidine dimer is not formed, leading to the absence of a band in the sequencing gel that indicates the presence of Tm⁵CG or Cm⁵CG, respectively. Thus, this method detects only a subset of the m⁵C sites in the genome. Sensitivity and vertical resolution have not been explored in detail, but since LM-PCR is used, they probably are the same as with the hydrazine method (Table 1). Although differential base modification by UV radiation has been used rarely so far and requires the more sophisticated LM-PCR technology, it should be quite useful in determining the effects of DNA methylation on UV induced DNA damage and repair.

HOW METHODOLOGICAL LIMITATIONS CAN AFFECT INTERPRETATION OF DATA

Since each method has its own limitations and potential artifacts, the results from different methods have not always led to the same conclusion. This problem is illustrated by analysis of the m⁵C distribution in mammalian genomes. The only known mammalian DNA methyltransferase is specific for maintaining m⁵CpG dinucleotides (90). Consistent with the properties of this enzyme, only two m⁵C residues have ever been reported in non-CpG dinucleotides in genomic DNA using either the hydrazine or permanganate methods (91). This amount is negligible compared to the thousands of nucleotides of genomic DNA on which the hydrazine method has been applied and so far has detected only m⁵CpG dinucleotides (84). However, some nearest neighbor analyses of genomic DNA suggest that up to 54.5% of all m⁵C are found at non-CpG sites (18,92,93). This conclusion is supported by some genomic sequence analyses where the bisulfite or MSRE method have detected m⁵C in both CpG and non-CpG dinucleotides (50,65,66,94,95). Moreover, mammalian cells have the ability to maintain m⁵CpNpGs sites integrated into the genome by transfection (50).

How might this paradox be resolved? First, non-CpG methylation activity has been detected in some mammalian extracts (30,92), but not in others (96–98), suggesting that the specificity of mammalian DNA methyltransferase may be altered by cofactors or limited proteolysis (99). In addition, some analyses may have overestimated the amounts of non-CpG methylation events. For example, reports of ‘densely methylated islands’ (DMIs) in which all cytosines within ~100 to ~500 bp regions were methylated, regardless of their dinucleotide composition (65,66), were later shown by stringent application of the bisulfite, MSRE, hydrazine and permanganate methods to be incorrect (39,51). Thus, other reports in which the bisulfite method produced similar results (100,101) should be considered with caution.

Another explanation is that the bulk of non-CpG methylation detected by total genome analysis is clustered at genomic sites that have not yet been examined by sequence analyses. Alternatively,

cell populations might be heterogeneous in their non-CpG methylation. If methylation at a unique non-CpG dinucleotide did not occur in more than 25% of the cells, it would not be detected by the hydrazine method, but would be detected by the bisulfite method because of the difference in their vertical resolutions (Table 1). In contrast, nearest-neighbor analysis collects all non-CpG dinucleotides of the same type into a single pool. Thus, nearest-neighbor analysis could observe significant non-CpG methylation events under conditions where methods with low vertical resolution (e.g. hydrazine) could not.

In conclusion, current methodology can map the distribution of m⁵C in any genome at nucleotide resolution, although a prudent investigator will apply more than one method to the same problem. New methods for mapping other covalent base modifications in higher eukaryotes will be necessary only when such modifications are discovered in eukaryotes. In prokaryotes where base modifications other than m⁵C are important, the MSRE method can map most, if not all of them, because they all appear to be involved in restriction and modification.

ACKNOWLEDGEMENTS

T.R. was a postdoctoral fellow of the Deutsche Forschungsgemeinschaft. This work was supported by research grant Zo 59-2/2 of the Deutsche Forschungsgemeinschaft awarded to H.Z. We thank Prof. E.-L. Winnacker for his support, and D.Natale, B.Howard, N.Martin and the reviewers for their helpful advice and information.

REFERENCES

- Doerfler, W., Schubert, R., Heller, H., Kammer, C., Hilger-Eversheim, K., Knoblauch, M. and Remus, R. (1997) *Trends Biotechnol.* **15**, 297–301.
- Jost, J.P. and Bruhat, A. (1997) *Prog. Nucleic Acid Res. Mol. Biol.* **57**, 217–248.
- Siegfried, Z. and Cedar, H. (1997) *Curr. Biol.* **7**, 305–307.
- Zingg, J.M. and Jones, P.A. (1997) *Carcinogenesis* **18**, 869–882.
- Rae, P.M. and Steele, R.E. (1978) *Biosystems* **10**, 37–53.
- Zhu, C.M. and Henney, H.J. (1990) *Biochem. Cell. Biol.* **68**, 944–948.
- Achwal, C.W., Iyer, C.A. and Chandra, H.S. (1983) *FEBS Lett.* **158**, 353–358.
- Klimasauskas, S., Steponaviciene, D., Maneliene, Z., Petrusyte, M., Butkus, V. and Janulaitis, A. (1990) *Nucleic Acids Res.* **18**, 6607–6609.
- Gruenbaum, Y., Naveh-Manly, T., Cedar, H. and Razin, A. (1981) *Nature* **292**, 860–862.
- Vanyushin, B.F. and Kirnos, M.D. (1988) *Gene* **74**, 117–121.
- Frostesjo, L., Holm, I., Grahn, B., Page, A.W., Bestor, T.H. and Heby, O. (1997) *J. Biol. Chem.* **272**, 4359–4366.
- Cedar, H. (1988) *Cell* **53**, 3–4.
- Bird, A. (1992) *Cell* **70**, 5–8.
- Bestor, T.H. (1990) *Philos. Trans. R. Soc. Lond. Biol.* **326**, 179–187.
- Achwal, C.W., Ganguly, P. and Chandra, H.S. (1984) *EMBO J.* **3**, 263–266.
- Cantoni, G.L. and Davis, D.R. (1966) *Procedures in Nucleic Acids Research*. Harper and Row, NY, London.
- Linn, S.M. and Roberts, R.J. (1982) *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Woodcock, D.M., Crowther, P.J. and Diver, W.P. (1987) *Biochem. Biophys. Res. Commun.* **145**, 888–894.
- Ehrlich, M., Gama, S.M., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C. and Gehrke, C.W. (1985) *Nucleic Acids Res.* **13**, 1399–1412.
- Annan, R.S., Kresbach, G.M., Giese, R.W. and Vouros, P. (1989) *J. Chromatogr.* **465**, 285–296.
- Clotten, R. and Clotten, A. (1962) *Hochspannungselektrophorese*, Georg Thieme Verlag, Stuttgart.
- Gommers, A.J., Lutgerink, J. and Borst, P. (1991) *Nucleic Acids Res.* **19**, 1745–1751.
- Randerath, K. (1961) *Angew. Chem.* **73**, 674–676.
- Eick, D., Fritz, H.J. and Doerfler, W. (1983) *Anal. Biochem.* **135**, 165–171.
- Woodcock, D.M., Crowther, P.J., Simmons, D.L. and Cooper, I.A. (1984) *Biochim. Biophys. Acta* **783**, 227–233.
- Patel, C.V. and Gopinathan, K.P. (1987) *Anal. Biochem.* **164**, 164–169.
- Razin, A. and Cedar, H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2725–2728.
- Gruenbaum, Y., Szyf, M., Cedar, H. and Razin, A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4919–4921.
- Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. and Razin, A. (1982) *FEBS Lett.* **146**, 148–152.
- Hubrich-Kühner, K., Buhk, H.-J., Wagner, H., Kröger, H. and Simon, D. (1989) *Biochem. Biophys. Res. Commun.* **160**, 1175–1182.
- McClelland, M., Nelson, M. and Raschke, E. (1994) *Nucleic Acids Res.* **22**, 3640–3659.
- Butkus, V., Klimasauskas, S., Petrauskienė, L., Maneliene, Z., Lebionka, A. and Janulaitis, A. (1987) *Biochim. Biophys. Acta* **909**, 201–207.
- Berkner, K.L. and Folk, W.R. (1979) *J. Biol. Chem.* **254**, 2551–2560.
- Butkus, V., Klimasauskas, S., Kersulyte, D., Vaitkevicius, D., Lebionka, A. and Janulaitis, A. (1985) *Nucleic Acids Res.* **13**, 5727–5746.
- Janosi, L., Yonemitsu, H., Hong, H. and Kaji, A. (1994) *J. Mol. Biol.* **242**, 45–61.
- Sutherland, E., Coe, L. and Raleigh, E.A. (1992) *J. Mol. Biol.* **225**, 327–348.
- Gärtner, U. and Zorbas, H. Manuscript in preparation.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Rein, T., Natale, D., Gärtner, U., Niggemann, M., DePamphilis, M.L. and Zorbas, H. (1997) *J. Biol. Chem.* **272**, 10021–10029.
- Singer-Sam, J., Grant, M., LeBon, J.M., Okuyama, K., Chapman, V., Monk, M. and Riggs, A.D. (1990) *Mol. Cell. Biol.* **10**, 4987–4989.
- Diviacco, S., Norio, P., Zentilin, L., Menzo, S., Clementi, M., G. Biamonti, Riva, S., Falaschi, A. and Giacca, M. (1992) *Gene* **122**, 313–320.
- Garrity, P.A. and Wold, B.J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1021–1025.
- Pfeifer, P.G., Steigerwald, S.D., Mueller, P.R., Wold, B. and Riggs, A.D. (1989) *Science* **246**, 810–813.
- McGrew, M.J. and Rosenthal, N. (1993) *Biotechniques* **15**, 722–729.
- Steigerwald, S.D., Pfeifer, G.P. and Riggs, A.D. (1990) *Nucleic Acids Res.* **18**, 1435–1439.
- Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S., Sugimoto, T., Kobayakawa, K., Kawara, T., Katsuki, M., Shibuya, T. and Mukai, T. (1993) *Electrophoresis* **14**, 251–258.
- Kuromitsu, J., Yamashita, H., Kataoka, H., Takahara, T., Muramatsu, M., Sekine, T., Okamoto, N., Furuichi, Y. and Hayashizaki, Y. (1997) *Mol. Cell. Biol.* **17**, 707–712.
- Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., Held, W.A., Hayashizaki, Y. and Chapman, V.M. (1996) *Nature Genet.* **14**, 106–109.
- Russell, D.W. and Hirata, R.K. (1989) *J. Biol. Chem.* **264**, 10787–10794.
- Clark, S.J., Harrison, J. and Frommer, M. (1995) *Nature Genet.* **10**, 20–27.
- Rein, T., Zorbas, H. and DePamphilis, M.L. (1997) *Mol. Cell. Biol.* **17**, 416–426.
- Razin, A. and Kafri, T. (1994) *Prog. Nucleic Acid Res. Mol. Biol.* **48**, 53–81.
- Shapiro, R., DiFate, V. and Welcher, M. (1974) *J. Am. Chem. Soc.* **96**, 906–912.
- Hayatsu, H. and Shiragami, M. (1979) *Biochemistry* **18**, 632–637.
- Sasse, D.S. and Gralla, J.D. (1991) *Methods Enzymol.* **208**, 146–168.
- Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831.
- Myöhänen, S., Wahlfors, J. and Jänne, J. (1994) *DNA Seq.* **5**, 1–8.
- Paul, C.L. and Clark, S.J. (1996) *Biotechniques* **21**, 126–133.
- Sadri, R. and Hornsby, P.J. (1996) *Nucleic Acids Res.* **24**, 5058–5059.
- Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. and Baylin, S.B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 9821–9826.
- Xiong, Z. and Laird, P.W. (1997) *Nucleic Acids Res.* **25**, 2532–2534.
- Gonzalzo, M.L. and Jones, P.A. (1997) *Nucleic Acids Res.* **25**, 2529–2531.
- Pieler, T., Digweed, M., Bartsch, M. and Erdmann, V.A. (1983) *Nucleic Acids Res.* **11**, 591–604.
- Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) *Nucleic Acids Res.* **22**, 2990–2997.
- Tasheva, E.S. and Roufa, D.J. (1994) *Mol. Cell. Biol.* **14**, 5636–5644.
- Tasheva, E.S. and Roufa, D.J. (1995) *Somatic Cell Mol. Genet.* **19**, 369–383.
- Feil, R., Charlton, J., Bird, A.P., Walter, J. and Reik, W. (1994) *Nucleic Acids Res.* **22**, 695–696.

- 68 Olek, A., Oswald, J. and Walter, J. (1996) *Nucleic Acids Res.* **24**, 5064–5066.
- 69 Raizis, A.M., Schmitt, F. and Jost, J.-P. (1995) *Anal. Biochem.* **226**, 161–166.
- 70 Wang, R.Y.-H., Gehrke, C. and Ehrlich, M. (1980) *Nucleic Acids Res.* **8**, 4777–4790.
- 71 Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- 72 Reeben, M. and Prydz, H. (1994) *Biotechniques* **16**, 416–417.
- 73 Warnecke, P.M., Stirzaker, C., Melki, J.R., Millar, D.S., Paul, C.L. and Clark, S.J. (1997) *Nucleic Acids Res.* **25**, 4422–4426.
- 74 Hornstra, I.K. and Yang, T.P. (1993) *Anal. Biochem.* **213**, 179–193.
- 75 Pfeifer, G.P., Singer-Sam, J. and Riggs, A.D. (1993) *Methods Enzymol.* **225**, 567–583.
- 76 Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
- 77 Ohmori, H., Tomizawa, J.I. and Maxam, A.M. (1978) *Nucleic Acids Res.* **5**, 1479–1485.
- 78 Hayatsu, H. and Ukita, T. (1967) *Biochem. Biophys. Res. Commun.* **29**, 556–561.
- 79 Fritzsche, E., Hayatsu, H., Igloi, G.L., Iida, S. and Kössel, H. (1987) *Nucleic Acids Res.* **15**, 5517–5528.
- 80 Mirkovitch, J. and Darnell, J.E., Jr (1991) *Genes Dev.* **5**, 83–93.
- 81 Saluz, H. and Jost, J.P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2602–2606.
- 82 Saluz, H.P. and Jost, J.P. (1994) *Methods Mol. Biol.* **31**, 317–329.
- 83 Grimaldi, K.A., McAdam, S.R., Soulami, R.L. and Hartley, J.A. (1994) *Nucleic Acids Res.* **22**, 2311–2317.
- 84 Tommasi, S., LeBon, J.M., Riggs, A.D. and Singer, S.J. (1993) *Somat. Cell. Mol. Genet.* **19**, 529–541.
- 85 Achwal, C.W. and Chandra, H.S. (1982) *FEBS Lett.* **150**, 469–472.
- 86 Singer, J., Schnute, W.J., Shively, J.E., Todd, C.W. and Riggs, A.D. (1979) *Anal. Biochem.* **94**, 297–301.
- 87 Barbin, A., Montpellier, C., Kokalj, V.N., Gibaud, A., Niveleau, A., Malfoy, B., Dutrillaux, B. and Bourgeois, C.A. (1994) *Hum. Genet.* **94**, 684–692.
- 88 de Capoa, A., Aleixandre, C., Felli, M.P., Ravenna, L., Costantino, M.A., Giaccotti, P., Vicenti, O., Poggesi, I., Grappelli, C. and Miller, D.A. (1991) *Hum. Genet.* **88**, 146–152.
- 89 Pfeifer, G.P., Drouin, R., Riggs, A.D. and Holmquist, G.P. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1374–1378.
- 90 Yoder, J.A., Soman, N.S., Verdine, G.L. and Bestor, T.H. (1997) *J. Mol. Biol.* **270**, 385–395.
- 91 Toth, M., Müller, U. and Doerfler, W. (1990) *J. Mol. Biol.* **214**, 673–683.
- 92 Grafstrom, R.H., Yuan, R. and Hamilton, D.L. (1985) *Nucleic Acids Res.* **13**, 2827–2842.
- 93 Nyce, J., Liu, L. and Jones, P.A. (1986) *Nucleic Acids Res.* **14**, 4353–4367.
- 94 Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P. and Warren, W.D. (1997) *J. Biol. Chem.* **272**, 7810–7816.
- 95 Snibson, K.J., Woodcock, D., Orian, J.M., Brandon, M.R. and Adams, T.E. (1995) *Transgenic Res.* **4**, 114–122.
- 96 Gruenbaum, Y., Stein, R., Cedar, H. and Razin, A. (1981) *FEBS Lett.* **124**, 67–71.
- 97 Gruenbaum, Y., Cedar, H. and Razin, A. (1982) *Nature* **295**, 620–622.
- 98 Roy, P.H. and Weissbach, A. (1975) *Nucleic Acids Res.* **2**, 1669–1684.
- 99 Bestor, T.H. (1992) *EMBO J.* **11**, 2611–2618.
- 100 Bull, J. and Wootton, J. (1984) *Nature* **310**, 701–704.
- 101 Selker, E.U., Fritz, D.Y. and Singer, M.J. (1993) *Science* **262**, 1724–1728.