4422–4426 *Nucleic Acids Research, 1997, Vol. 25, No. 21*

Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA

Peter M. Warnecke[1,2], Clare Stirzaker[1], John R. Melki[1], Douglas S. Millar[1], Cheryl L. Paul[1] and Susan J. Clark[1,3,*]

[1]Kanematsu Laboratories, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia, [2]School of Biological Sciences, A12 University of Sydney, NSW 2006, Australia and [3]CSIRO Division of Molecular Science, Sydney Laboratory, PO Box 184, North Ryde, NSW 2113, Australia

Received June 2, 1997; Revised and Accepted September 9, 1997

ABSTRACT

Methylation analysis of individual cytosines in genomic DNA can be determined quantitatively by bisulphite treatment and PCR amplification of the target DNA sequence, followed by restriction enzyme digestion or sequencing. Methylated and unmethylated molecules, however, have different sequences after bisulphite conversion. For some sequences this can result in bias during the PCR amplification leading to an inaccurate estimate of methylation. PCR bias is sequence dependent and often strand-specific. This study presents a simple method for detection and measurement of PCR bias for any set of primers, and investigates parameters for overcoming PCR bias.

Bisulphite genomic sequencing provides a positive and efficient method for detecting and quantitating the level of methylation of individual cytosines in genomic DNA (1–4). One common application of bisulphite sequencing is the quantitation of relative amounts of methylated and unmethylated DNA at particular loci, such as studies in X-inactivation (5), genomic imprinting (6) and tumourigenesis (7). The technique involves bisulphite conversion of DNA whereby cytosine is converted to uracil but 5-methylcytosine is non-reactive. The target sequence is PCR amplified using specific primers to yield fragments in which all uracil and thymine residues are amplified as thymine and only 5-methylcytosine is amplified as cytosine. However, if the DNA region to be amplified contains molecules with vastly different methylation states, then the bisulphite-converted DNAs will differ substantially in their cytosine content; highly methylated molecules will give rise to relatively C-rich DNA while unmethylated DNA will give rise to T-rich sequences. Therefore, it is possible that one population of sequences may amplify preferentially, leading to a PCR bias and an inaccurate estimate of methylation. Potential PCR bias was highlighted during our studies on the methylation pattern of the retinoblastoma (Rb) tumour suppressor gene CpG island promoter in retinoblastoma tumours (7). Southern blot data indicated that the Rb tumour DNA was fully methylated, while normal (leukocyte) DNA was unmethylated. However, using bisulphite sequencing we found that the top and bottom strand of the Rb promoter, which are no longer complementary after bisulphite conversion, displayed different methylation levels in some tumour samples; methylation of the top strand was reduced (~50%) relative to the bottom strand (~100%) (7). This raised the possibility that unmethylated T-rich DNA from normal cells, present in the tumour samples, was being amplified preferentially in the top strand PCR reaction.

To address the problem of potential PCR bias in quantitative bisulphite methylation analysis, we have devised a simple test assay for the detection and measurement of PCR bias for any set of primers and corresponding target sequence. We define PCR bias as the number of methylated molecules amplified for each unmethylated molecule amplified at the conclusion of the PCR reaction. The assay involves performing PCR reactions on a panel of DNA samples with defined methylation states. Genomic DNA that is both unmethylated and extensively methylated at the region of interest is required. As a source of unmethylated DNA we have used either DNA from cells which are known to be unmethylated in the target sequence such as human leukocyte DNA, or DNA from murine ES cells that are deficient in the enzyme DNA methyltransferase ($Dnmt^{c/c}$) (8). As a source of methylated DNA we have used DNA from tumour cells which is known to be extensively methylated or SssI methylated genomic DNA. The methylated and unmethylated DNAs are mixed in defined proportions, bisulphite-treated (1) and PCR amplified with the primer sets to be tested. The amount of methylated DNA in each sample, at the conclusion of the PCR reaction, is quantitated by restriction enzyme digest or Genescan analysis (4). PCR bias is measured by plotting the percent of methylated DNA (y) recovered after the PCR reaction for each sample mix as a function of the percent input methylated DNA (x). A value for the bias (b) that gives the line of best fit for Equation 1 $[y = (100bx)/(bx – x + 100)]$ or $b = [y(100 – x)]/[x(100 – y)]$ is derived using regression analysis.

To test if there was bias in PCR amplification in the top or bottom strand sequence of the human Rb gene promoter we prepared mixtures of unmethylated (leukocyte) DNA, and increasing proportions (0, 50, 80, 90, 95, 99, 100%) of methylated (Rb tumour) DNA (7). The mixed DNA samples were bisulphite treated and PCR amplified with AmpliTaq polymerase (Perkin-Elmer,

*To whom correspondence should be addressed. CSIRO Division of Molecular Science, Sydney Laboratory, PO Box 184, North Ryde, NSW 2113, Australia. Tel: +612 9490 5148; Fax: +612 9490 5005; Email:susan.clark@molsci.csiro.au

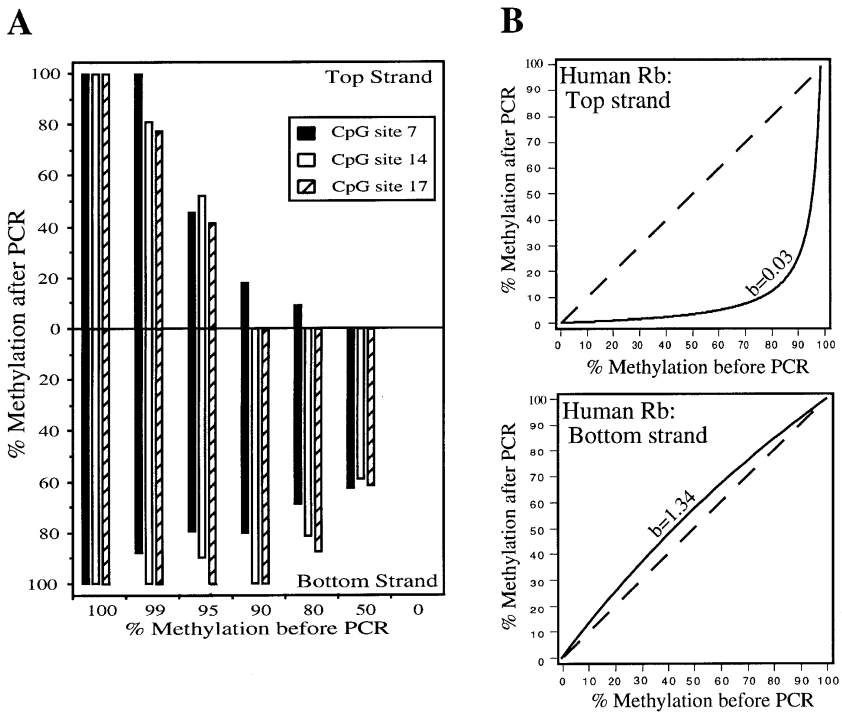

Figure 1. PCR bias in quantitative methylation analysis of top and bottom strands of the human retinoblastoma gene promoter. (**A**) Genescan analysis was used to quantitate percentage methylation at each CpG dinucleotide (4). CpG sites 7, 14 and 17 from the Rb promoter (7) are shown for the top and bottom strands. PCR primers and conditions are as previously described (7). Percent (%) of methylated retinoblastoma tumour (384T) DNA amplified is shown on the x-axis. (**B**) Bias plots (solid line) for Rb top and bottom strand, with unbiased b = 1 plot (dotted line) shown for comparison. To obtain a numeric value for PCR bias, percent of methylated DNA (y axis) quantitated for each sample mix was plotted as a function of the input methylated DNA (x axis) and a best-fit value for bias derived by regression (MacCurveFit). The value b for PCR bias represents the difference between the proportion of methylated DNA before (x) and after (y) the PCR, i.e. $[x / (100 – x)].b = [y/(100 – y)]$, which is algebraically equivalent to $y = (100bx)/(bx – x + 100)$ i.e. Equation 1.

Norwalk, CT) using nested primer sets directed to the top and bottom strands of the Rb promoter (7). To determine if the methylated DNA is amplified in proportion, the PCR product from each sample set was sequenced and the methylation status quantitated using Genescan analysis (4) (Fig. 1). The methylated (tumour) sample shows substantial methylation at all CpG sites in both the top and bottom strands. In the sample containing 5% unmethylated and 95% methylated DNA, the amplified top strand appears to be only 50% methylated (Fig. 1A). Whereas, in the sample with 10% unmethylated and 90% methylated DNA, little or no methylation was detected in the top strand DNA sequence, indicating that only T-rich unmethylated molecules are being amplified. In contrast, all the bottom strand amplifications are consistent with the methylation content prior to the PCR. It is clear from these results that there is a strong bias towards the amplification of unmethylated DNA for the top strand primers.

To obtain a quantitative measurement of PCR bias, the amount of methylated DNA recovered after PCR for each sample mix was plotted as a function of the input methylated DNA, and a value for bias that gave the curve of best fit for Equation 1 was derived using regression analysis. Figure 1B show plots of bias for the top and bottom strand primers. The bias value of b = 0.03 derived for the top strand primers represents a 33-fold preference in amplification of unmethylated DNA versus methylated DNA. In contrast, the bottom strand primers were calculated as having a bias value of b = 1.34 or a 1.34-fold preference in amplification of methylated DNA, indicating that methylated and unmethylated DNAs are amplified with almost equal efficiency.

To determine if PCR bias is a significant problem in the bisulphite methylation analysis of other CpG-rich regions we used the test assay system described in Figure 2 to analyse several other primer sets and corresponding target DNA for possible PCR bias. For each set of primers used, the amount of methylated DNA following PCR, as determined by restriction enzyme analysis, is plotted as a function of the input DNA and a value for bias obtained by regression analysis. Substantial bias was observed for a number of different primer sets including mouse Rb (b = 0.12- or 8.3-fold) and human p16 (b = 0.05- or 20-fold); whereas other primer sets showed little bias such as mouse H19A (b = 1.4- or 1.4-fold) and mouse H19B (b = 0.38- or 2.6-fold). In most cases the bias was towards the preferential PCR amplification of unmethylated DNA. In fact we have not detected any instances of a substantial bias towards amplification of methylated DNA. Methylated DNA gives rise to a bisulphite-treated derivative with a higher (G+C) content than unmethylated DNA, and it is possible that this higher (G+C) content may raise the melting temperature of the DNA and increase the likelihood of secondary structure formation for some sequences, resulting in a lower PCR efficiency when compared to unmethylated sequences. Therefore, the proportion of unmethylated DNA will be increased after PCR of a sample containing a mixture of methylated and unmethylated bisulphite-treated DNA.

We have analysed the amplified bisulphite-converted DNA sequence corresponding to the different primer sets to determine if there is a direct correlation of PCR bias and (G+C) content (Table 1). The primers listed have been constructed in accordance with guidelines previously described (1), and are designed to bind

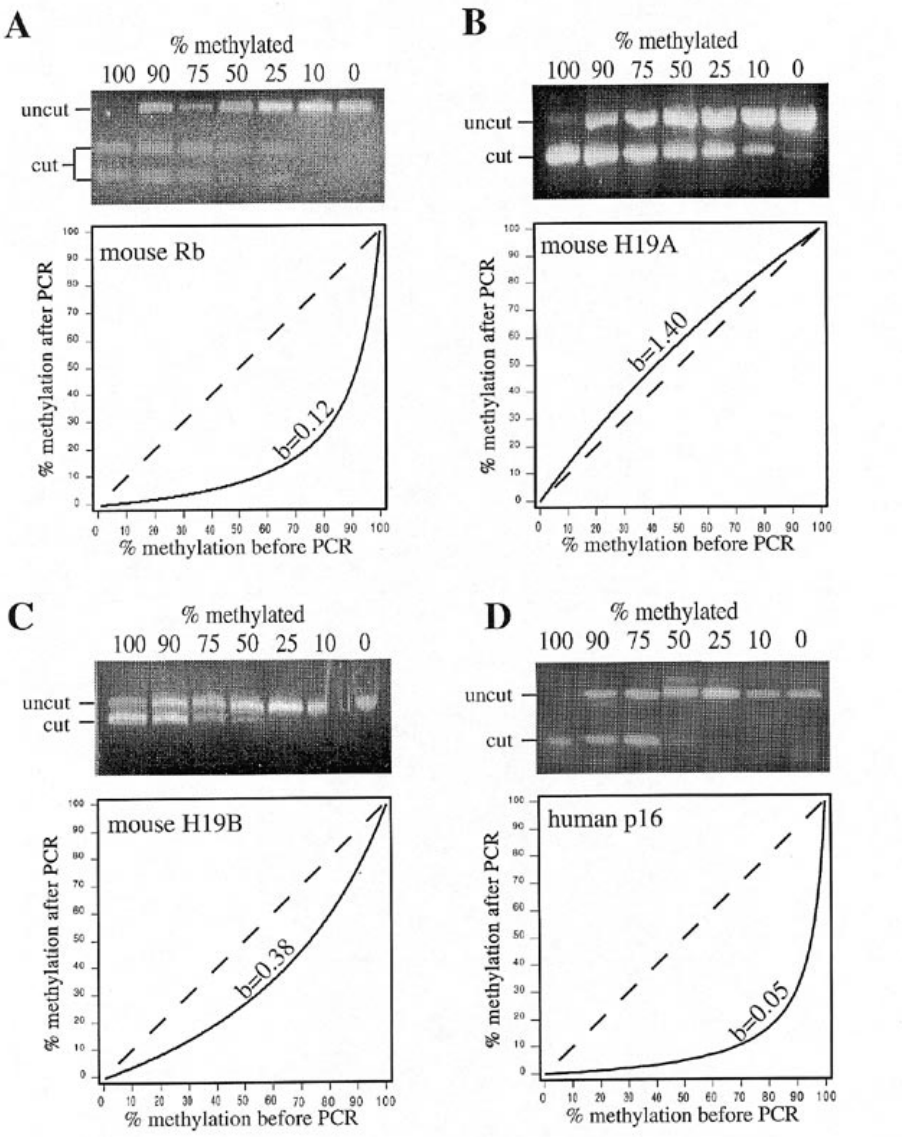

Figure 2. Bias Measurement Test Assay. Unmethylated genomic DNA (~10 µg), either human (leukocyte DNA) or murine (Dnmt$^{c/c}$ ES DNA; 8), was divided into two aliquots; one aliquot was treated with SssI methylase (New England Biolabs) as recommended by the manufacturer, and the second aliquot was left unmethylated. The SssI-methylated and unmethylated genomic DNAs were used to prepare a series of mixed samples containing 100, 90, 75, 50, 25, 10 and 0% SssI-methylated DNA (1 µg total) and were bisulphite treated as previously described (1,2). To test for bias, a series of PCR reactions using the primers to be tested, were performed in 50 µl reaction volumes using 40 ng of the mixed genomic DNA, as previously described (11). The amount of methylated DNA amplified, at the conclusion of the PCR, was determined by digestion with informative restriction enzymes such as *Bst*UI (CGCG) or *Hha*I (GCGC); these restriction sites remain after bisulphite treatment only in methylated DNA. The proportions of cut and uncut DNA were quantitated either by densitometer scanning of an agarose gel following ethidium bromide staining or phosphorimager analysis (Molecular Dynamics) of radiolabelled PCR products by addition of [^{32}P]dATP to the PCR reaction. The 0 and 100% methylated DNAs were used to control for background gel fluorescence, complete restriction and complete SssI methylation. After correcting for the size difference between the cut and uncut bands, the proportion of methylated DNA for each PCR was determined and the best-fit value for bias derived by regression (MacCurveFit), as described in Figure 1. Examples of target DNAs tested for PCR bias are: (**A**) mouse retinoblastoma gene promoter (GenBank accession no. M86180 bases 814–1352, and M26391 bases 1–113) *Hha*I digested; (**B**) mouse H19A (GenBank accession no. U19619, bases 1089–1481) *Bst*UI digested; (**C**) mouse H19B (GenBank accession no. U19619, bases 2557–3235) *Hha*I digested and (**D**) human p16 exon I (GenBank accession no. U12818, bases 8–336) *Bst*UI digested.

methylated and unmethylated DNA with equal efficiency. DNA regions containing a large number of CpG dinucleotides, and hence a large difference in sequence and (G+C) content between methylated and unmethylated sequences following bisulphite treatment, tend generally to show more biased amplification. However, CpG density alone is not predictive of the extent or occurrence of PCR bias, for example the Rb bottom strand sequence is equally different in G+C content to the top strand sequence but shows little PCR bias. Bias may often be a consequence of secondary structure formation within the DNA molecule, and therefore, may be detectable only by experimental testing of primer sets using defined mixtures of methylated and unmethylated DNA as described.

In an attempt to overcome PCR bias, we have analysed a number of variables in the PCR conditions, including extension time, annealing and denaturation temperature, $MgCl_2$ concentra-

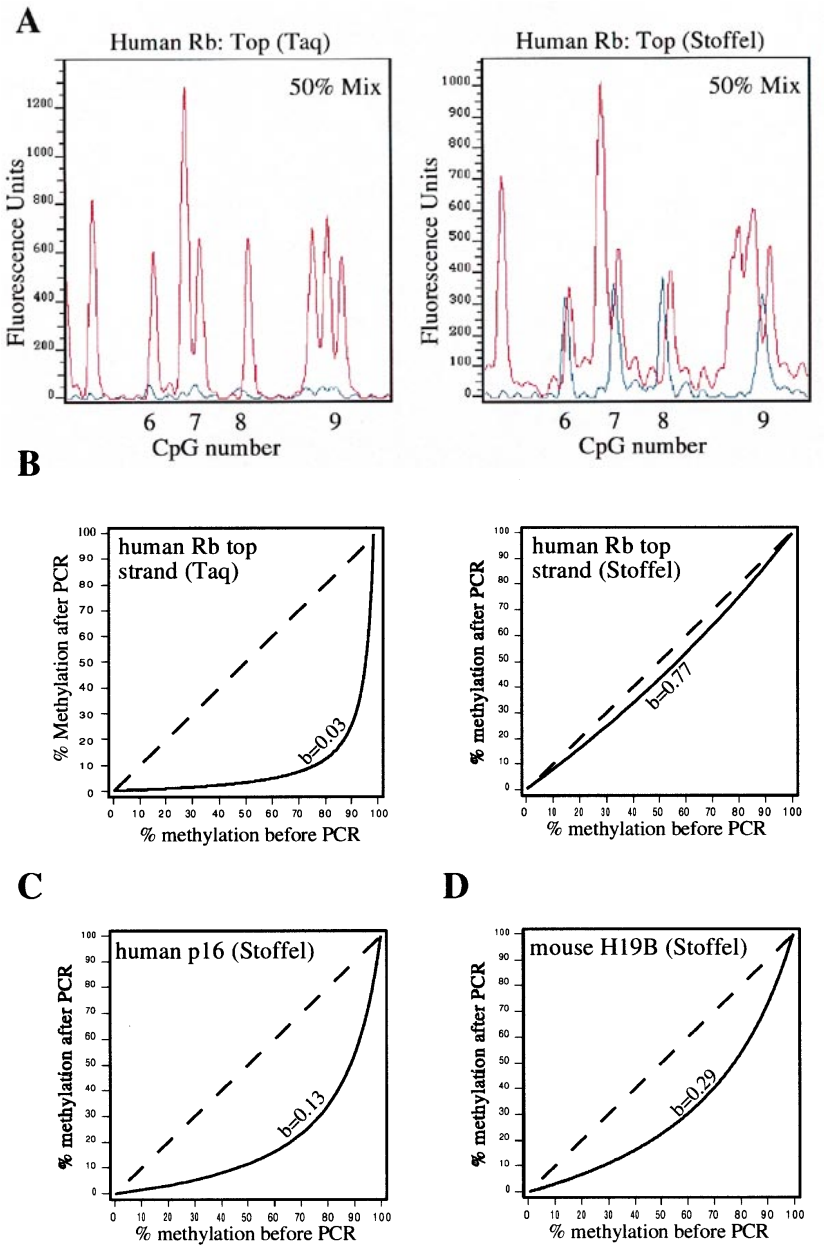

Figure 3. PCR bias comparing Amplitaq polymerase and the Stoffel fragment. (**A**) Genescan electrophoretograms of a 50% mix of methylated and unmethylated DNA amplified with human Rb top strand primers (as described in Fig. 2), using AmpliTaq polymerase (left) and Stoffel fragment (right). Both cytosine (C) and thymine (T) are sequenced using blue (FAM) dye-labelled primer. The profile of the T tracks is indicated by a red line and the C profile by a blue line to distinguish the peaks. The CpG dinucleotides in the sample sequence panel are numbered below the peaks. The PCR conditions for the Stoffel fragment were the same as those previously used, except the $MgCl_2$ concentration was increased to 3 mM. (**B**) Bias plot for human Rb top strand comparing amplification with AmpliTaq polymerase (left) and Stoffel fragment (right). (**C**) Bias plot for human p16 exon I top strand after amplification with Stoffel fragment in 2.5 mM $MgCl_2$. (**D**) Bias plot for mouse H19B primers after amplification with Stoffel fragment in 2.0 mM $MgCl_2$. Bias plots were derived as described in Figure 2.

tion and the addition of varying concentrations of the secondary structure inhibitors DMSO and formamide. None of these modifications resulted in a significant reduction of PCR bias using these primers (data not shown). Moreover, the use of different primers to amplify the same DNA region had no effect on the level of PCR bias, indicating that it is the body of the sequence being amplified and not the placement of individual primers that is causing PCR bias in this case. However, the substitution of Stoffel fragment (Perkin-Elmer, Norwalk, CT), a DNA polymerase noted for its ability to amplify different DNA sequences simultaneously (9), for Amplitaq polymerase resulted in a marked reduction of bias from b = 0.03 (33-fold) to b = 0.77 (1.3-fold) in the amplification of the human Rb top strand sequence (Fig. 3A and B). Whereas, the substitution of Stoffel fragment has not significantly resolved PCR bias for other DNA sequences as shown in Figure 3C and D. Stoffel fragment slightly reduced the bias in human p16 from b = 0.05 to b = 0.13 (Fig. 3C), but had little effect on mouse H19B (Fig. 3C). Therefore, Stoffel fragment is not a universal solution to all PCR bias but may help in some circumstances.

Table 1. Primer sets analysed for PCR bias

Primer set	Size of region amplified (bp)	No. CpG	CpG density[a]	%(G+C) M/UM[b]	Δ(G+C)[c]	Bias
Human Rb (top)	420	55	0.13	47.9 / 34.8	13.1	0.03
Human Rb (bot)	317	38	0.12	43.2 / 31.2	12.0	1.34
Mouse H19A	393	12	0.031	29.3 / 26.2	3.1	1.40
Mouse H19B	682	25	0.037	29.5 / 25.8	3.7	0.38
Mouse Rb	550	90	0.16	51.5 / 35.1	16.4	0.12
Human p16	329	32	0.098	58.7 / 48.9	9.8	0.05

[a]CpG density equals the number of CpG dinucleotides divided by the size of the sequence.
[b]Percent (G+C) content of bisulphite treated sequence, for 100% methylated (M) and 100% unmethylated (UM).
[c]Difference in (G+C) content between methylated and unmethylated sequence.

Quantitative amplification of methylated and unmethylated DNA is not necessary for all applications involving bisulphite treatment of DNA. If only the methylated sequences are of interest, then the issue of PCR bias may be avoided altogether by designing primers that specifically amplify methylated DNA (10). Alternatively, if quantitative measurements are required and a moderate PCR bias is found to exist for a primer set, then the derived bias value may be used to correct data. As the exact degree of bias may vary depending on the reaction conditions used for some primer sets, we recommend as a simple check for unbiased amplification, the inclusion of a control containing a defined mix of methylated and unmethylated DNA in any PCR amplification of bisulphite treated DNA. This assay will be useful in ensuring the accurate quantitative analysis in samples where low levels of unmethylated sequences are present, such as in tumour samples with up to 5–10% normal cell contamination, or in samples with mixed methylation states, such as imprinted genes.

ACKNOWLEDGEMENTS

We thank Ted Dryja (Harvard Medical School) for providing retinoblastoma and leukocyte DNA samples and En Li for providing DNA from Dnmt$^{c/c}$ ES cells. This work was funded in part by an NH and MRC grant (# 970977) and NSW Cancer Council grant (RG/44/46) to S.J.C. P.M.W. and J.R.M. are supported by APRA and Anthony Rothe Memorial Trust scholarships.

REFERENCES

1 Clark, S. J., Harrison, J., Paul, C. L. and Frommer, M. (1994) *Nucleic Acids Res.*, **22**, 2990–2997.
2 Clark, S. J. and Frommer, M. (1997) In Taylor (ed.), *Laboratory Methods for the Detection of Mutations and Polymorphisms in DNA*. CRC Press, Boca Raton, pp. 151–162.
3 Clark, S. J. and Frommer, M. (1995) In Saluz and Wiebauer (eds), *DNA and Nucleoprotein Structure In Vivo*. R. G. Landes Company, Springer-Verlag, Heidelberg, Germany, pp. 123–136.
4 Paul, C. L. and Clark, S. J. (1996) *Biotechniques*, **21**, 126–133.
5 Park, J. G. and Chapman, V. M. (1994) *Mol. Cell. Biol.*, **14**, 7975–7983.
6 Zeschnigk, M., Schmitz, B., Dittrich, B., Buiting, K., Horsthemke, B. and Doerfler, W. (1997) *Hum. Mol. Genet.*, **6**, 387–395.
7 Stirzaker, C., Millar, D. S., Paul, C. L., Warnecke, P. M., Harrison, J., Vincent, P. C., Frommer, M. and Clark, S. J. (1997) *Cancer Res.*, **57**, 2229–2237.
8 Lei, H., Oh, S. P., Okano, M., Jüttermann, R., Goss, K. A., Jaenisch, R. and Li, E. (1996) *Development*, **122**, 3195–3205.
9 Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. -Y., Landre, P. A., Abramson, R. D. and Gelfand, D. H. (1993) *PCR Methods Appl.*, **2**, 275–287.
10 Herman, J. G., Graff, J. R., Myöhänen, S., Nelkin, B. D. and Baylin, S. B. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 9821–9826.
11 Clark, S. J., Harrison, J. and Frommer, M. (1995) *Nature Genet.*, **10**, 20–27.