

Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase

Rachna Goyal¹, Richard Reinhardt² and Albert Jeltsch^{3,*}

¹Institut für Biochemie, FB 08, Heinrich-Buff-Ring 58, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany,

²Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, D-14195 Berlin-Dahlem, Germany and

³Biochemistry, International University Bremen, School of Engineering and Science, Campus Ring 1, 28759 Bremen, Germany

Received December 22, 2005; Revised January 23, 2006; Accepted February 9, 2006

ABSTRACT

DNA methyltransferase 1 (Dnmt1) has a central role in copying the pattern of DNA methylation after replication which is one manifestation of epigenetic inheritance. With oligonucleotide substrates we show that mouse Dnmt1 has a 30- to 40-fold preference for hemimethylated DNA that is almost lost after addition of fully methylated oligonucleotides. Using long hemimethylated DNA substrates that carry defined methylation patterns and bisulfite analysis of the methylation reaction products, we show a 15-fold preference for hemimethylated CG sites. Dnmt1 moves along the DNA in a random walk methylating hemimethylated substrates with high processivity (>50 sites are visited on average which corresponds to linear diffusion over 6000 bp). The frequency of skipping sites is very low (<0.3%) and there is no detectable flanking sequence preference. CGCTC sites tend to terminate the processive methylation of DNA by Dnmt1. Unmethylated DNA is modified non-processively with a preference for methylation at CCGG sites. We simulate the propagation of methylation patterns using a stochastic model with the specificity of Dnmt1 observed here and conclude that either methylation of several sites is required to propagate the methylation information over several cellular generations or additional epigenetic information must be used.

INTRODUCTION

Epigenetic information is defined as heritable but not encoded in the DNA sequence (1). One form of epigenetic information in human cells is methylation of cytosine residues in DNA at the C5-position at CG sites [reviewed in (2,3)]. DNA

methylation is involved in a number of important biological processes, like control of gene expression, imprinting, development, X-chromosome inactivation, genomic integrity and protection of the genome against selfish DNA elements [reviewed in (2–7)]. On the other hand, erroneous DNA methylation causes diseases including cancer [reviewed in (1,8,9)]. The Dnmt1 DNA methyltransferase is responsible for the propagation of the DNA methylation information by maintenance of the DNA methylation pattern after DNA replication [reviewed in (10)]. The enzyme shows a preference for hemimethylated CG sites as they appear after DNA replication. Its activity, therefore, leads to the re-establishment of the same DNA methylation pattern as it was before replication. The accuracy of the transmission of the methylation information to the next cellular generation critically depends on the degree of specificity of Dnmt1 for hemimethylated sites. Surprisingly, this property has not yet been characterized sufficiently, and numbers ranging from 2- to 50-fold were published for oligonucleotide substrates, depending on the substrate, experimental test system and enzyme preparation (11–15). Moreover, many groups have reported that Dnmt1 has reduced specificity in the presence of methylated DNA (11,16–18) which would put into question the whole concept of inheritance of methylation patterns. Other important properties that determine the efficiency of this enzyme in the remethylation of DNA after replication are its processivity and whether it tends to skip hemimethylated target sites or not. Here, we have studied the accuracy, processivity and frequency of skipping sites by Dnmt1 using long hemimethylated DNA molecules that mimic physiological substrates. Our data provide a framework for the understanding of the role of this enzyme *in vivo*.

MATERIALS AND METHODS

Detailed descriptions of the experiments can be found in supplementary data. In brief, Dnmt1 was purified as described

*To whom correspondence should be addressed. Tel: +49 421 200 3247; Fax: +49 421 200 3249; Email: a.jeltsch@iu-bremen.de

(11,19). Methylation of 30mer oligonucleotide substrates containing one CG site was studied using radioactively labeled AdoMet as described (11). Two different PCR products comprising 634 and 566 bp were amplified from lambda DNA using one 5'-phosphorylated and one standard primer and processed to generate hemimethylated substrates basically as described (19,20). Briefly, they were methylated using either M.SssI or M.HpaII, then the upper strand of the DNA was digested by incubation with λ -exonuclease, which specifically removes the 5'-phosphorylated strand. After a fill in reaction using the same upper primer as before, substrates were obtained that were hemimethylated at all CG or CCG sites. The substrates were methylated by Dnmt1 in buffer containing 20 mM HEPES, pH 7.0, 1 mM EDTA, 30 mM NaCl, 7% glycerol, 25 μ g/ml BSA and 1 mM AdoMet at 37°C and methylation determined by bisulfite analysis performed as described (21–23).

RESULTS AND DISCUSSION

Accuracy of Dnmt1 for hemimethylated DNA

The mouse Dnmt1 enzyme used for these studies is full-length enzyme with an additional His-tag at the N-terminal end. The enzyme was prepared as described (11,19) and is over 90% pure (Supplementary Figure 1). Initially, we have analyzed the specificity of the Dnmt1 enzyme for hemimethylated DNA using 30 bp oligonucleotide substrates containing one centrally positioned CG site. In agreement with previous results (11), we observed a high preference for hemimethylated substrates which is ~30- to 40- fold in the initial phase of the methylation reaction (Figure 1A). Next, we used hemimethylated DNA substrates to study DNA methylation by Dnmt1 on longer substrates, which is closer to a physiological situation. First, the methylation of substrate 1 (634 bp) hemimethylated at all 54 CG sites was compared with its unmethylated form under identical conditions (Supplementary Table 1). As shown in Figure 2, dense methylation was observed if the hemimethylated substrate was used (average methylation level: 58%), whereas methylation of the unmethylated substrate was only sparse (average methylation level: 4%) which indicates a 15-fold preference of Dnmt1 for the hemimethylated substrate. This ratio is identical to the results obtained with oligonucleotide substrates after 10 min of incubation (Figure 1). Thus, there are no large differences in the specificity of Dnmt1 for hemimethylated substrates if a 30mer oligonucleotide and longer DNAs are compared.

Next, we determined the methylation of the hemimethylated and unmethylated oligonucleotide substrates in the presence of an additional fully methylated oligonucleotide and observed an almost complete loss of preference for the hemimethylated target under these conditions (Figure 1B). Since in the cell unmethylated, hemimethylated and fully methylated DNA is present at the same time, this loss of specificity puts into question the whole concept of maintenance methylation. Therefore, we studied the methylation of an unmethylated and hemimethylated long substrate in competition in the same reaction mixture. Equimolar amounts of hemimethylated substrate 1 and unmethylated substrate 2 (566 bp, 44 CG sites) were mixed, incubated with Dnmt1 as described before and their respective methylation pattern was examined by bisulfite

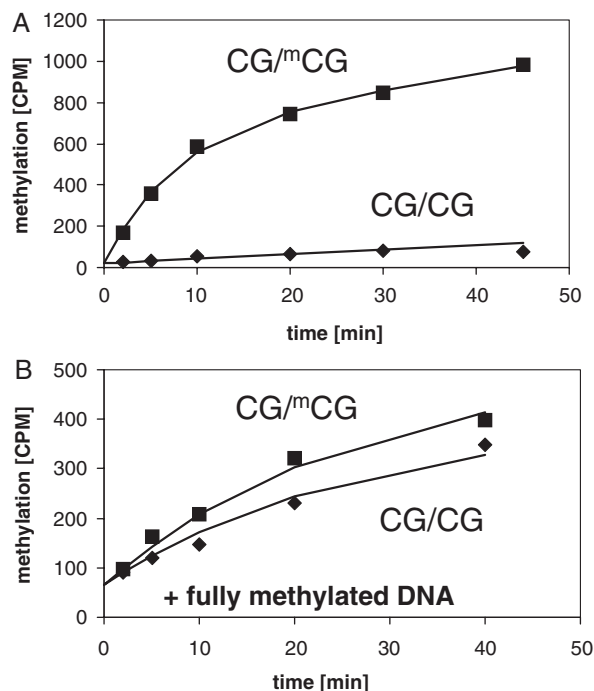


Figure 1. Methylation of unmethylated and hemimethylated oligonucleotides by Dnmt1. Methylation was analyzed by incorporation of radioactive methyl groups into the DNA. In (A) 1 μ M of hemimethylated and unmethylated oligonucleotides (CG30hm and CG30um) and 100 nM enzyme were used. In (B) 1 μ M fully methylated oligonucleotide (CG30fm) of the same sequence was added.

analysis (Figure 3A, Supplementary Table 1). We observed average levels of methylation of the hemimethylated and unmethylated substrates of 57 and 5%, which indicate that the activity and specificity of the enzyme was not changed.

To study the capacity of Dnmt1 to copy an existing methylation pattern, we have prepared substrate 2 hemimethylated at all CCG sites, but unmethylated at all other CG sites. This DNA was incubated with Dnmt1 and methylation determined as before (Figure 3B, Supplementary Table 1). We observed 61% of methylation at the hemimethylated CCG sites, while only 5% of the unmethylated CG sites had become modified, such that the preference was ~12-fold. These results demonstrate that the presence of methylated DNA does not impair specific methylation of longer hemimethylated substrates. The strong allosteric activation and loss of specificity observed with oligonucleotide substrates upon addition of methylated DNA most likely is due to the artificial experimental setup using short oligonucleotide substrates and that is unlikely to affect the maintenance activity of Dnmt1 in the cell.

Processivity of Dnmt1

We analyzed the processivity of DNA methylation by Dnmt1 using hemimethylated and unmethylated substrates. Processive methylation is defined by consecutive methylation of one DNA substrate without dissociation of the enzyme from the DNA. Processive methylation of DNA will result in a stretch of methylated sites on the DNA that is not interrupted by unmethylated sites. To determine the number of

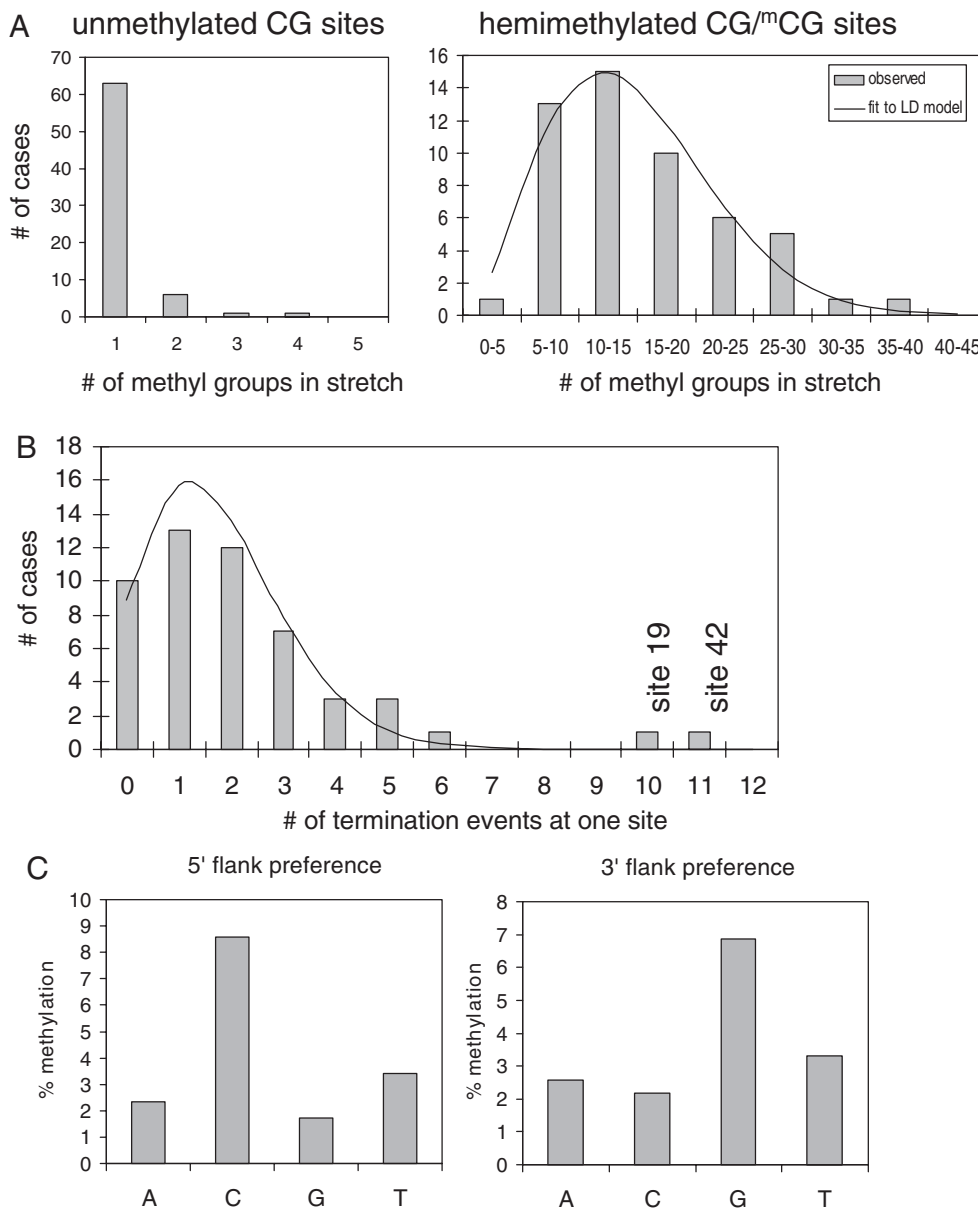


Figure 4. (A) Processivity of the methylation of the unmethylated and hemimethylated CG sites by Dnmt1. The figure displays the distribution of lengths of all stretches of methylated sites for the unmethylated (left) and hemimethylated substrate (right). The line in the right panel indicates a fit to the random walk model with a P_{dif} of 54. (B) Distribution of the number of termination events of processive methylation of the DNA at the various CG sites. The distribution is roughly fitted by a binomial distribution of the corresponding parameters as indicated by the line, if the two sites 19 and 42 are disregarded. (C) Flanking sequence preference for methylation of unmethylated DNA by Dnmt1.

somewhere in the center of the stretch of completely methylated DNA.

Consequently, the sites 19 and 42 represent termination sites that constitute barriers to the continuation of processive DNA methylation by Dnmt1. To understand this observation, we first investigated whether a large distance of these sites to their next neighbor sites could be due to the preferential termination of methylation. In our substrate, the CG sites are separated by up to 36 bp. Within this distance, there is no correlation detectable between the distances of CG sites to their neighbor sites and the termination frequency (data not shown), which indicates that the dwell time of Dnmt1 on the DNA is sufficiently long to allow for traveling this distance in

most cases. Next, we analyzed for statistical significance: The observed distribution of termination events at all sites (except for the 10 and 11 events at sites 19 and 42) was roughly fitted by a binomial distribution indicating no strong deviation from statistics for most sites. However, the probability of finding two sites with more than nine termination events in this distribution was very low ($P = 1.4 \times 10^{-4}$), which suggests that these sites play a special role for Dnmt1. Interestingly, both sites are flanked by the same CTC trinucleotide sequence on their 3' side. There was no sequence similarity on the 5' side of the CG sequence and there were no additional CGCTC sites in our substrate. This result suggests that CGCTC sites function as termination site for DNA methylation by Dnmt1 either by

promoting dissociation of the enzyme from the DNA or by forming a very tight interaction that prevents continuation of linear diffusion. Further experiments will show if this sequence also functions in other substrates and which mechanism is operative. It should be noticed that approaching a CGCTC site not necessarily led to termination of processive DNA methylation as indicated by the finding that several continuously methylated stretches span over the CGCTC sites.

Depending of the sequence context, CGCTC sites also function as binding site of the CTCF insulator protein (CCCTC) (27), which is involved in the control of imprinting and X-inactivation (28). One role of CTCF is to separate methylated and unmethylated parts of the DNA. On the basis of our data one could speculate that CGCTC sites could result in insulation of DNA methylation by a 2-fold effect: either by recruiting CTCF protein or in a CTCF independent way by preventing that Dnmt1 moves to the adjacent DNA. Both modes of action would result in the inhibition of spreading of DNA methylation into unmethylated regions of the DNA.

Flanking sequence preferences of Dnmt1 for methylation of unmethylated CG sites

In the experiments using unmethylated substrates, we observed 83 methylation events altogether. Since the number of methylation events did not suffice for a simultaneous statistical analysis with respect to next base pair on the 5' and 3' side, we studied 5' and 3' flanking preferences separately (Figure 4C). Our data show that Dnmt1 prefers to methylate DNA in a CCGG context which is methylated at ~ 10 times higher frequency than all other sides. This conclusion is also illustrated by the observation that on the unmethylated substrates $>40\%$ of all methylation events occurred at CCGG sites although these represented only 10% of all target sites. We realized that by chance we observed a flanking preference on the unmethylated DNA that is identical to the sequence used for hemimethylation in the experiment shown in Figure 3B. However, the experiments shown in Figures 2 and 3A were carried out completely independently from that shown in Figure 3B such that there was no change of a partially methylation of the unmodified substrates.

There is no preference for flanking sequences reported for the catalytic activity of Dnmt1 so far. Our finding agrees to results by Flynn *et al.* (29) who reported increased binding of Dnmt1 to GCGG and CCGC sites. In contrast, the CCGG flanking sequence is disfavored by Dnmt3a and Dnmt3b (30). The difference in the flanking sequence preferences of Dnmt1 and Dnmt3a and 3b might allow in the future to determine whether aberrant *de novo* methylation was due to malfunction of Dnmt1 or Dnmt3a and 3b.

Implications for the propagation of methylation patterns *in vivo*

DNA methylation carries information that is used for control of gene expression. One of the most fascinating questions regarding DNA methylation is to understand how methylation patterns are copied after DNA replication to achieve a stable transmission of epigenetic information to next cellular generations. This function has been attributed to Dnmt1, which can re-establish a methylation pattern after DNA replication on the daughter strands of the DNA, because it prefers

methylation at hemimethylated DNA sites. Errors in the propagation of methylation patterns are due to imperfect maintenance methylation as well as *de novo* methylation at previously unmethylated sites. In order to explore the accuracy of this mechanism we have analyzed the specificity of Dnmt1 for hemimethylated sites. Using long hemimethylated substrates, we show that Dnmt1 prefers hemimethylated sites ~ 15 -fold over unmethylated, irrespective of the presence of other sites in vicinity. This level of fidelity is compatible with *in vivo* data, which indicate that methylation is maintained with an accuracy of $\sim 95\%$ and *de novo* methylation occurs with $\sim 5\%$ (31–33), suggesting an ~ 20 -fold preference for methylation of hemimethylated sites after DNA replication. This preference of Dnmt1 sets an upper limit of accuracy for the propagation of punctual patterns of DNA methylation.

To simulate the process of the propagation of a DNA methylation pattern, we apply the model of stochastic maintenance of DNA methylation (31,33) and started with two populations of DNA molecules, one methylated at a certain site and the other not, and followed the average methylation of both populations over several generations (Figure 5). The information content of the methylation pattern is given by the ratio of the probabilities to find the site still methylated after N generations in the first population over finding it methylated in the second population after the same time (discrimination factor).

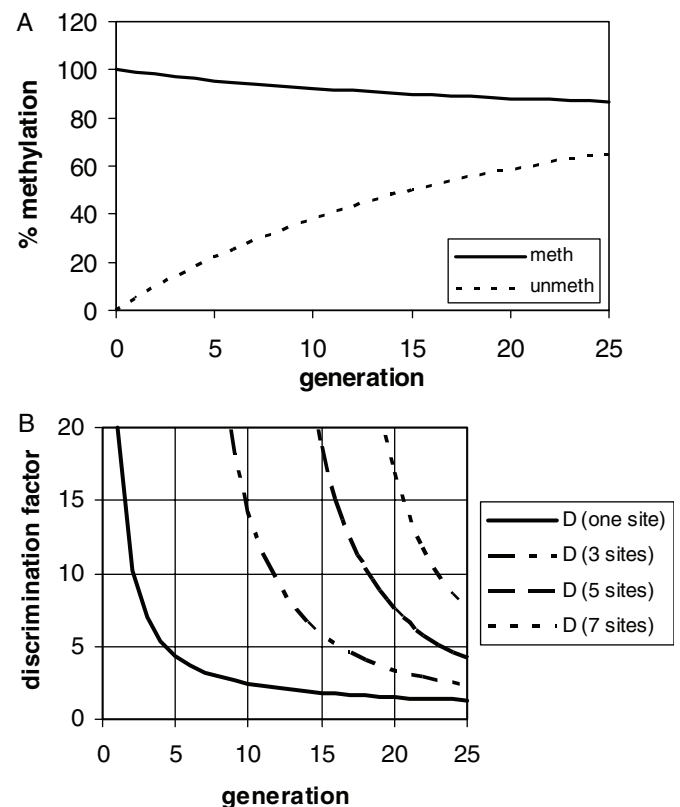


Figure 5. Simulation of the average methylation level of an unmethylated and hemimethylated population of DNA molecules during several generations. For the simulation an average level of maintenance methylation of 95% and a 20-fold preference for hemimethylated DNA was used. In (A) the average methylation levels of the hemimethylated and unmethylated populations at one CG site are shown, in (B) the discrimination factor (defined as the ratio of the average methylation levels of both populations) is shown if one, three, five or seven CG sites are concerned.

If we assume a discrimination factor of 10 is required for transmission of the epigenetic information, we observe that the signal is lost already after three generations, if a single CG site is considered (Figure 5). However, if the biological system makes use of the combined reading of the methylation state of several CG sites, our simulation suggests that combination of 5 sites would allow to propagate a signal over 18 cellular generations and 7 sites suffice to propagate the signal over 23 generations which is already close to the typical somatic lifetime of many cells. This observation could explain why methylation of more than one CG site is required for strong gene inhibition and targeting of methyl cytosine binding proteins to methylated DNA. On the other hand, in the light of these data, it is not surprising that loss of methylation information occurs during aging and development. In the cell the specificity of maintenance methylation could be increased by local differences in the rate of methylation of unmodified CG which could be mediated by other chromatin modifications like histone 3 at lysine 9 methylation. Such mechanism could make use of Dnmt1 or one of the Dnmt3 enzymes and combine the effects of the various epigenetic systems into self enforcing epigenetic circuits whose state can be transmitted stably.

ACKNOWLEDGEMENTS

This work has been supported by the DFG priority program Epigenetics (JE 252/4) and the BMBF NGFN2 (SMP Epigenetics, 01GR0497) BioFuture, and BioChance programs. Funding to pay for the open access publication charges for this article was provided by International University Bremen.

Conflict of interest statement. None declared.

REFERENCES

1. Egger, G., Liang, G., Aparicio, A. and Jones, P.A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, **429**, 457–463.
2. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
3. Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBiochem.*, **3**, 274–293.
4. Jones, P.A. and Takai, D. (2001) The role of DNA methylation in mammalian epigenetics. *Science*, **293**, 1068–1070.
5. Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Rev. Genet.*, **3**, 662–673.
6. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.*, **33**, 245–254.
7. Reik, W., Dean, W. and Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science*, **293**, 1089–1093.
8. Feinberg, A.P. and Tycko, B. (2004) The history of cancer epigenetics. *Nature Rev. Cancer*, **4**, 143–153.
9. Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nature Rev. Genet.*, **3**, 415–428.
10. Hermann, A., Gowher, H. and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol. Life Sci.*, **61**, 2571–2587.
11. Fatemi, M., Hermann, A., Pradhan, S. and Jeltsch, A. (2001) The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J. Mol. Biol.*, **309**, 1189–1199.
12. Tollefsbol, T.O. and Hutchison, C.A., III (1995) Mammalian DNA (cytosine-5)-methyltransferase expressed in *Escherichia coli*, purified and characterized. *J. Biol. Chem.*, **270**, 18543–18550.
13. Tollefsbol, T.O. and Hutchison, C.A., III (1997) Control of methylation spreading in synthetic DNA sequences by the murine DNA methyltransferase. *J. Mol. Biol.*, **269**, 494–504.
14. Flynn, J., Glickman, J.F. and Reich, N.O. (1996) Murine DNA cytosine-C5 methyltransferase: pre-steady- and steady-state kinetic analysis with regulatory DNA sequences. *Biochemistry*, **35**, 7308–7315.
15. Pradhan, S., Bacolla, A., Wells, R.D. and Roberts, R.J. (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of *de novo* and maintenance methylation. *J. Biol. Chem.*, **274**, 33002–33010.
16. Fatemi, M., Hermann, A., Gowher, H. and Jeltsch, A. (2002) Dnmt3a and Dnmt1 functionally cooperate during *de novo* methylation of DNA. *Eur. J. Biochem.*, **269**, 4981–4984.
17. Christman, J.K., Sheikhejad, G., Marasco, C.J. and Sufrin, J.R. (1995) 5-Methyl-2'-deoxycytidine in single-stranded DNA can act in *cis* to signal *de novo* DNA methylation. *Proc. Natl Acad. Sci. USA*, **92**, 7347–7351.
18. Bacolla, A., Pradhan, S., Roberts, R.J. and Wells, R.D. (1999) Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna. *J. Biol. Chem.*, **274**, 33011–33019.
19. Hermann, A., Goyal, R. and Jeltsch, A. (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.*, **279**, 48350–48359.
20. Roth, M. and Jeltsch, A. (2000) Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biol. Chem.*, **381**, 269–272.
21. Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.*, **22**, 2990–2997.
22. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (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.
23. Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R. and Clark, S.J. (2002) Identification and resolution of artifacts in bisulfite sequencing. *Methods*, **27**, 101–107.
24. Vilkaitis, G., Suetake, I., Klimasauskas, S. and Tajima, S. (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J. Biol. Chem.*, **280**, 64–72.
25. Ehbrecht, H.J., Pingoud, A., Urbanke, C., Maass, G. and Gualerzi, C. (1985) Linear diffusion of restriction endonucleases on DNA. *J. Biol. Chem.*, **260**, 6160–6166.
26. Jeltsch, A. and Pingoud, A. (1998) Kinetic characterization of linear diffusion of the restriction endonuclease EcoRV on DNA. *Biochemistry*, **37**, 2160–2169.
27. Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J. and Lobanenko, V.V. (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol*, **16**, 2802–2813.
28. Reik, W. and Lewis, A. (2005) Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nature Rev. Genet.*, **6**, 403–410.
29. Flynn, J., Azzam, R. and Reich, N. (1998) DNA binding discrimination of the murine DNA cytosine-C5 methyltransferase. *J. Mol. Biol.*, **279**, 101–116.
30. Handa, V. and Jeltsch, A. (2005) Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J. Mol. Biol.*, **348**, 1103–1112.
31. Pfeifer, G.P., Steigerwald, S.D., Hansen, R.S., Gartler, S.M. and Riggs, A.D. (1990) Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl Acad. Sci. USA*, **87**, 8252–8256.
32. Laird, C.D., Pleasant, N.D., Clark, A.D., Sneed, J.L., Hassan, K.M., Manley, N.C., Vary, J.C., Jr, Morgan, T., Hansen, R.S. and Stoger, R. (2004) Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc. Natl Acad. Sci. USA*, **101**, 204–209.
33. Riggs, A.D. and Xiong, Z. (2004) Methylation and epigenetic fidelity. *Proc. Natl Acad. Sci. USA*, **101**, 4–5.