Journal of Cell Science, Supplement 19, 37-39 (1995)
Printed in Great Britain © The Company of Biologists Limited 1995

Studies of DNA methylation in animals

Adrian Bird, Peri Tate, Xinsheng Nan, Javier Campoy, Richard Meehan, Sally Cross, Susan Tweedie, Jillian Charlton and Donald Macleod

Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Mayfield Road, Edinburgh EH9 3JR, UK

SUMMARY

We have been studying the evolution and function of DNA methylation in vertebrate animals using three related approaches. The first is to further characterise proteins that bind to methylated DNA. Such proteins can be viewed as 'receptors' of the methyl-CpG 'ligand' that mediate downstream consequences of DNA modification. The second approach involves CpG islands. These patches of non-methylated DNA coincide with most gene promoters, but their origin and functional significance have only recently become the subject of intensive study. The third approach is to trace the evolution of DNA methylation. Genomic methylation patterns of vertebrates are strikingly different from those of invertebrates. By studying methylation in animals that diverged from common ancestors near to the invertebrate/vertebrate boundary, we will assess the possibility that changes in DNA methylation contributed causally to the evolution of the complex vertebrate lineage.

Key words: DNA methylation, methyl-CpG binding protein, CpG island, genome evolution

INTRODUCTION

The predominant methylated sequence in all animals is the self-complementary dinucleotide CpG. In vertebrates, most CpGs in the genome are methylated at the 5 position on the cytosine ring. Several biological consequences of this post-synthetic modification are known. Best understood is the methylation-associated mutagenesis that has caused the under-representation of CpG in the genome and is responsible for over one third of the point mutations that give rise to human genetic diseases (Bird, 1980; Jones et al., 1992). It is difficult to see this as a selected advantage of DNA methylation. More likely it is an unavoidable price to be paid for some other benefit of methyl-CpG. Strikingly, the invertebrates (which account for well over 95% of animal species) may not pay this price, as few, if any of their genes are methylated (see below).

The need for DNA methylation during normal mammalian development has been shown by disruption of the gene for cytosine methyltransferase (MTase) in mice (Li et al., 1992). Mutant embryos have greatly reduced levels of DNA methylation, and die in mid-gestation. In seeking an explanation for this embryonic lethal phenotype, it is tempting to focus on the well-known effects of methylation on transcription. DNA methylation has long been correlated with transcriptional repression. That it causes repression has been shown by introduction of artificially methylated constructs into cells (Vardimin et al., 1982; Stein et al., 1982), and by the use of drugs that inhibit the MTase (Jones and Taylor, 1980). A reasonable hypothesis is that embryos lacking the MTase die because the methylation-mediated repression mechanism fails.

METHYLATION-MEDIATED REPRESSION OF TRANSCRIPTION

Several parameters determine the influence of methyl-CpG on transcription. The parameters are: the location of methyl-CpGs relative to the promoter (they should be close-by; Murray and Grosveld, 1987); the local density of methyl-CpGs (the strength of repression is proportional to density of methylation; Boyes and Bird, 1992); the strength of the promoter (weak promoters are repressed by lower methylation densities than strong ones; Boyes and Bird, 1992); and the dependence of promoter function on transcription factors that are sensitive to methyl-CpG (reviewed by Tate and Bird, 1993). We have identified a protein that interacts with methylated DNA according to the density of methyl-CpGs, and have implicated this protein as a mediator of transcriptional repression (Meehan et al., 1989; Boyes and Bird, 1991). The activity is known as methyl-CpG binding protein 1 or MeCP1. Considerable effort has been expended on purification of MeCP1. As might be expected from its size (800 kDa by gel fitration), MeCP1 comprises several polypeptide chains, and dissociates upon affinity chromatography with methylated DNA, leading to loss of activity. Our belief that MeCP1 may be of central importance in understanding the mechanism of methylation-mediated repression has sustained us through the trials of its purification.

Studies of the second methyl-CpG binding protein, MeCP2, have recently advanced significantly. Following preliminary characterisation of MeCP2 and its gene (Lewis et al., 1992), we set out to assess the biological significance of the protein. We know that it is very abundant (over 10^6 molecules per cell) and is a tightly bound component of mammalian chromosomes. Recent studies of its localisation made use of a fusion

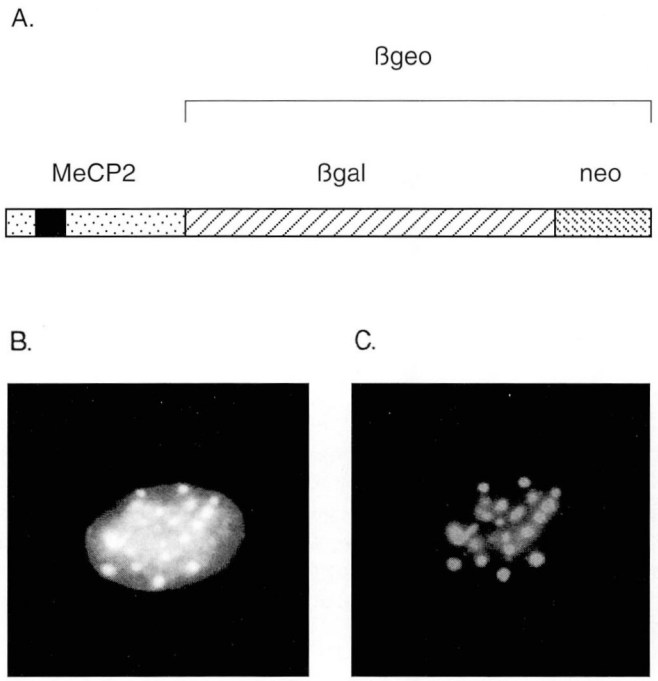

Fig. 1. Localisation of an MeCP2-lac Z fusion protein to heterochramatic foci in mouse cells. (A) Diagram of the fusion protein between MeCP2 (dotted and solid shading, left) and the *lac Z*-neomycin resistance fusion gene *β-geo* (Friedrich and Soriano, 1991). The solid box within the MeCP2 moiety represents the 'methyl-CpG binding domain', which is essential for correct localisation (X. Nan, unpublished results). (B and C) Staining of a mouse L cell nucleus with Hoescht 33258 (B) and anti-βgal antibodies (C). The cell that contained this nucleus had been transfected with a gene expressing the fusion protein diagrammed in A. The heterochromatic foci that are intensely stained by Hoechst are the primary targets of the MeCP2 fusion protein.

between the cDNA for MeCP2 and *lacZ-neoR* gene as reporter. When mouse cells are transfected with this construct, the resulting fusion protein localises preferentially to heterochromatin, thereby mimicking endogenous MeCP2 (Fig. 1). Truncations and deletions of the MeCP2 moiety have established that the 80 amino acid methyl binding domain (MBD; Nan et al., 1993) is both necessary and sufficient for localisation. More directly, it was found that the association of MeCP2 with chromosomes is dependent on methylation, as cells lacking DNA methylation cannot localise the protein efficiently (X. Nan et al., unpublished results). Thus MeCP2 is a methyl-CpG binding protein in vivo as well as in vitro, and as such may be a major mediator of the effects of DNA methylation on cells. If MeCP2 is a mediator of the effects of methylation, it should, like the MTase itself, be essential for mouse development. By disrupting the X-linked gene in embryonic stem (ES) cells, we have shown that it is indeed essential (P. H. Tate et al., unpublished results). Chimaeric embryos show developmental abnormalities whose severity depends on the proportion of mutant cells. ES cells lacking the MeCP2 gene grow normally, as do ES cells that lack the MTase. Taken together, the results tell us that our interest in MeCP2 is justified, but they do not reveal its biological function. Future work will address this problem.

HISTONE H1 DOES NOT HAVE A HIGH AFFINITY FOR METHYLATED DNA

Several laboratories have proposed that the linker histone H1 binds preferentially to methylated DNA, and may therefore be involved in methylation-mediated transcriptional repression (Levine et al., 1993; Johnson et al., 1995). We have spent some time testing this idea using a variety of assays (Campoy et al., unpublished results). In our hands, no preferential affinity of H1 for methylated DNA could be detected. This was true for naked DNA and also for DNA that had been assembled into poly-nucleosomal chromatin using a *Xenopus* oocyte extract. Thus it is unlikely that histone H1 is involved in mediating the biological consequences of CpG methylation.

ORIGIN OF CpG ISLANDS

Islands of non-methylated CpG-rich DNA (CpG or HTF islands) occur at the majority of human genes. They usually cover the promoter and extend downstream into the gene for 1,000 base pairs (bp) on average (Bird, 1986). We and others have used a transgenic mouse assay to find out which parts of a CpG island determine its methylation-free status. In the case of the adenine phosphoribosyltransferase gene, retention of the island depended on the presence of sites for the transcription factor Sp1 (Macleod et al., 1994; Brandeis et al., 1994). These sites, which are required for transcription of the gene, are occupied by protein (presumably Sp1) in vivo, and surprisingly are located at the extreme 5′ edge of the island rather than in its centre (Fig. 2; Macleod et al., 1994). Two questions are raised by these findings. Firstly, how do occupied Sp1 sites at the edge of a CpG island keep 1,000 bp

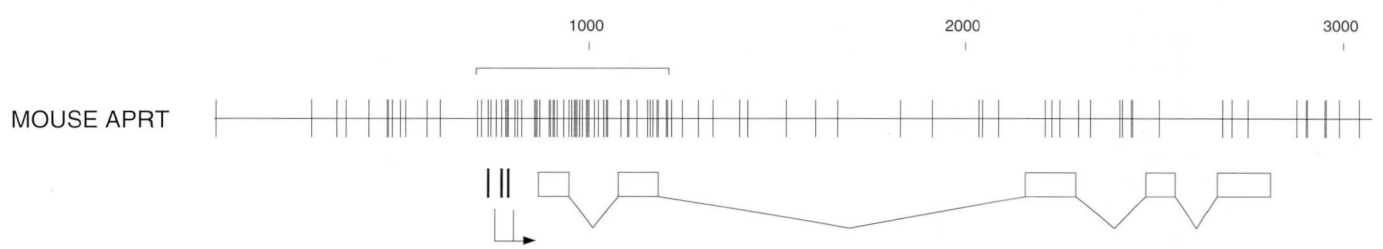

Fig. 2. Peripheral Sp1 sites are essential for maintaining the methylation-free status of the CpG island at the mouse adenine phosphoribosyltranferase gene (Macleod et al., 1994). The CpG island is denoted by the bracket. Sp1 sites are shown by the three vertical bars. Vertical crosslines on the map represent CpGs. Open boxes are exons. The two transcription starts are joined to an arrow below the diagram.

downstream free of methylation? Secondly, if transcription is necessary for the creation of CpG islands, why do many tissue-specific genes (e.g. human alpha-globin) have non-methylated islands in tissues where they are not expressed (Bird et al., 1987)? These questions will be important themes for the future.

EVOLUTION OF DNA METHYLATION PATTERNS

It has been known for some time that the extensive genomic DNA methylation seen in vertebrates is exceptional (Bird et al., 1979; Bird and Taggart, 1980). Methylation of invertebrate genomes is confined to a small fraction of the genome, and in some cases (e.g. *Drosophila melanogaster* and *Caenorhabditis elegans*) may be absent altogether. Although the data are incomplete, there is reason to believe that methylated DNA in invertebrates comprises transposable elements and other potentially damaging DNA sequences that have been detected and silenced by a mechanism involving methylation. No methylated gene has yet been reliably reported in an invertebrate, and the primary function of DNA methylation in these organisms may be to protect the genome by neutralising disruptive elements. In vertebrates, on the other hand, the genome as a whole is heavily methylated, and most genes are methylated to some extent.

The transition from the predominantly non-methylated genome of invertebrates to the predominantly methylated genome of vertebrates appears to occur within the chordates (A. Bird, S. Tweedie and J. Charlton, unpublished results). Could this dramatic change have facilitated the evolutionary development of the complex vertebrate lineage? It has been suggested that the total number of genes in vertebrates is considerably higher than in invertebrates (50,000-100,000 versus 10,000-25,000; Bird, 1995). On the strength of this and other data, it was proposed that the increased gene number (and therefore complexity) of vertebrates is due to improved methods of reducing transcriptional noise (that is, transcription of non-genic DNA or of genes that are inappropriate for the cell type concerned). The theory has the virtue that it makes some testable predictions and that it might explain a major macroevolutionary change. Its disadvantage is that it is rather speculative, going some way beyond the available data. Whether or not the noise reduction idea is relevant to the evolutionary origin of vertebrates, the transition in methylation patterns deserves further study for the light that it may shed on the biology of DNA methylation generally.

We thank Joan Davidson and Aileen Greig for technical assistance. This work was funded by grants from The Wellcome Trust, Imperial Cancer Research Fund and The Howard Hughes Medical Institute.

REFERENCES

Bird, A. P., Taggart, M. H. and Smith, B. A. (1979). Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* **17**, 889-901.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucl. Acids Res.* **8**, 1499-1594.

Bird, A. P. and Taggart, M. H. (1980). Variable patterns of total DNA and rDNA methylation in animals. *Nucl. Acids Res.* **8**, 1485-1497.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213.

Bird, A. P., Taggart, M. H., Nicholls, R. D. and Higgs, D. R. (1987). Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *EMBO J.* **6**, 999 -1004.

Bird, A. P. (1995). Gene number, noise reduction and biological complexity. *Trends Genet.* **11**, 94-100.

Boyes, J. and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123-1134.

Boyes, J. and Bird, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* **11**, 327-333.

Brandeis, M., Frank, D., Keshet, I., Siegried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**, 435-438.

Friedrich, G. and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513-1523.

Johnson, C., Goddard, J. and Adams, R. (1995). The effect of histone H1 and DNA methylation on transcription. *Biochem. J.* **305**, 791-798.

Jones, P. A. and Taylor, S. M. (1980). Cellular differentiation, cytidine analogues and DNA methylation. *Cell* **20**, 85-93.

Jones, P. A., Rideout, W. M., Shen, J.-C., Spruck, C. H. and Tsai, Y. C. (1992). Methylation, mutation and cancer. *BioEssays* **14**, 33-36.

Levine, A., Yeivin, A., Ben-Asher, E., Aloni, Y. and Razin, A. (1993). Histone H1-mediated inhibition of transcription initiation of methylated templates in vitro. *J. Biol. Chem.* **268**, 21754-21759.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992). Purification, sequence and cellular localisation of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905-914.

Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926.

MacLeod, D., Charlton, J., Mullins, J. and Bird, A. P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* **8**, 2282-2292.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. and Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499-507.

Murray, E. J. and Grosveld, F. (1987). Site specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation mediated suppression. *EMBO J.* **6**, 2329-2335.

Nan, X., Meehan, R. R. and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucl. Acids Res.* **21**, 4886-4892.

Stein, R., Razin, A. and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphorybosyl transferase gene inhibits its expression in mouse L cells. *Proc. Nat. Acad. Sci. USA* **79**, 4418-3422.

Tate, P. H. and Bird, A. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Biol.* **3**, 226-231.

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M. and Doerfler, W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proc. Nat. Acad. Sci. USA* **79**, 1073-1077.