VIEWPOINT

The Role of DNA Methylation in Mammalian Epigenetics

Peter A. Jones* and Daiya Takai

Genes constitute only a small proportion of the total mammalian genome, and the precise control of their expression in the presence of an overwhelming background of noncoding DNA presents a substantial problem for their regulation. Noncoding DNA, containing introns, repetitive elements, and potentially active transposable elements, requires effective mechanisms for its long-term silencing. Mammals appear to have taken advantage of the possibilities afforded by cytosine methylation to provide a heritable mechanism for altering DNA-protein interactions to assist in such silencing. Genes can be transcribed from methylation-free promoters even though adjacent transcribed and nontranscribed regions are extensively methylated. Gene promoters can be used and regulated while keeping noncoding DNA, including transposable elements, suppressed. Methylation is also used for long-term epigenetic silencing of X-linked and imprinted genes and can either increase or decrease the level of transcription, depending on whether the methylation inactivates a positive or negative regulatory element.

DNA methylation is essential for the development of mammals (*1*, *2*), but despite 25 years of work, researchers still do not know exactly why. Recent advances have led to the cloning and preliminary characterization of the three known active DNA cytosine methyltransferases (DNMT1, -3a, and -3b) (*3*, *4*) and to a greater understanding of how the methylation signal is interpreted in mammalian cells. The post-synthetic addition of methyl groups to the 5-position of cytosiness alters the appearance of the major groove of DNA to which the DNA binding proteins bind. These epigenetic "markers" on DNA can be copied after DNA synthesis, resulting in heritable changes in chromatin structure. Methylation of CpG-rich promoters is used by mammals to prevent transcriptional initiation and to ensure the silencing of genes on the inactive X chromosome, imprinted genes, and parasitic DNAs. The potential role of methylation in tissue-specific gene expression or in the regulation of CpG-poor promoters is less well established. There is also tantalizing evidence that normal chromosome structure may be affected by methylation and that human diseases, including cancer, are caused and impacted by abnormal methylation.

CpG Islands

CpG dinucleotides, the sites of almost all methylation in mammals, are underrepresented in DNA. Clusters of CpGs, called CpG islands, are often found in association with genes, most often in the promoters and first exons but also in regions more toward the 3′ end (*5*). The exact definition of a CpG island is evolving. The original suggestion by Gardiner-Garden and Frommer (*6*) of a region greater than 200 base pairs (bp) with a high-GC content and an observed/expected ratio for the occurrence of CpG > 0.6, should probably be modified to slightly higher stringency in terms of length and GC content, thus excluding a substantial number of small exonic regions and repetitive parasitic DNAs (*7*). The salient property of a CpG island is that it is unmethylated in the germline (and indeed in most somatic tissues), thus ensuring its continued existence in the face of the strong mutagenic pressure of 5-methylcytosine deamination. CpG islands often function as strong promoters and have also been proposed to function as replication origins (*8*). Even though they are generally not methylated, most investigations on the role of DNA methylation in mammals have focused on CpG islands rather than on the regions in which the majority of methylation is found.

Methylation Patterns

Prokaryotes with methylases show modification of all potential methylation sites in their genomes—why don't mammals? This question could not be addressed in detail before the invention of bisulfite sequencing (*9*); now the patterns of methyl modifications can be read with ease. Two not-necessarily-conflicting models could account for the varying patterns, (i) an exclusion of access to methylation sites by proteins bound to specific DNA regions or (ii) a methylation-targeting mechanism steered by sequence-specific binding proteins. Evidence for the first model comes from studies showing that removal of Sp1 binding sites flanking a CpG island led to its de novo methylation during development (*10*, *11*). This might suggest that the DNA methylases could not gain access to the relevant CpG island if the flanking Sp1 sites were occupied by the Sp1 transcription factors. Also, modulation of DNA binding protein affinity can directly affect the methylation state of the sites to which the protein binds (*12*, *13*). These experiments suggest that exclusion of the DNMTs could play a role in the formation of patterns such as those that occur in early development (*14*); however, there is no unequivocal evidence, as yet, that this is the dominant mechanism in vivo.

If there are physical barriers to methylation of CpG islands, there must also be mechanisms to override them under specific developmental cues. For example, de novo methylation of the promoters of X-linked genes occurs on the inactive but not the active X chromosome at about the time of implantation. The exclusion model suggests that the pattern seen in germ and somatic cells is analogous to a "footprinting" process in which everything gets methylated, except those regions (mostly CpG islands) in which access to the CpG sites is excluded by local chromatin structure, such as the Sp1 sites.

Evidence for the targeting model comes from recent work showing an association of the DNMT1 and DNMT3a enzymes with proteins including Rb, E2F1, histone deacetylases (HDACs), and the transcriptional repressor RP58, among others (*15*–*17*). However, no one has shown that interactions such as these lead to de novo methylation. Targeting is also suggested by the lack of methylation of satellite DNA in mouse or human cells deficient in DNMT3b activity, indicating that repetitive heterochromatic DNA may be a target for DNMT3b (*2*, *18*). The DNMTs can function as transcriptional suppressors in collaboration with other proteins such as HDACs in the absence of DNA methylation (*15*–*17*). Thus, it may be necessary to re-evaluate earlier experiments claiming to show that the enzymatic activities of the methyltransferases rather than the proteins themselves were responsible for the observed repression of gene expression.

The Importance of Spreading

The concept that silencing can "spread" comes from studies in the field of heterochromatinization in *Drosophila* and in X-inacti-

USC/Norris Comprehensive Cancer Center, Departments of Biochemistry and Molecular Biology, Keck School of Medicine of the University of Southern California, 1441 Eastlake Avenue, MS 8302L, Los Angeles, CA 90089–9181, USA.

*To whom correspondence should be addressed. E-mail: jones_p@ccnt.hsc.usc.edu

vation. Early experiments in the methylation field showed that de novo methylation of an integrated retroviral sequence spread into host sequences flanking the integration site (*19*). It has also been proposed that repetitive elements such as short and long interspersed nuclear elements (SINEs and LINEs) might serve as foci for de novo methylation and that methylation may spread from such "attractors" of modification (*20*, *21*). Mechanisms of spreading remain unknown but may be related to the fact that partially methylated DNA is a better substrate than unmethylated DNA for DNMT1 in vitro (*22*, *23*). Recent studies have shown that methylated DNA allosterically activates the catalytic center of DNMT1 (*24*). Perhaps methylation attracts more methylation, thus ensuring a self-perpetuating methylated state. The phenomenon of spreading may explain the propensity of CpG islands to become methylated de novo by spreading from the surrounding methylated sequences during aging and cancer while the overall level of genomic methylation is decreased (*25*). The mechanisms underlying these profound alternations in distribution of methyl groups remain unknown.

What Does Methylation Do to Mammalian Transcription?

Although it is often said that "methylation blocks gene expression," this statement is an oversimplification. Methylation changes the interactions between proteins and DNA, which leads to alterations in chromatin structure and either a decrease or an increase in the rate of transcription. The position of the methylation change relative to the transcription start site is critical to the outcome (Fig. 1). Methylation of a promoter CpG island leads to binding of methylated CpG binding proteins (MBDs) and transcription repressors including HDACs and to a block of transcription initiation (*26*, *27*). The precise roles of the MBDs as global suppressors of methylated promoters have not yet been fully defined. Knockout mice for *methyl-CpG binding protein-2* (*Mecp2*) (*28*, *29*) or *MBD2* (*30*) have not yet been shown to have increased expression of endogenous genes or parasitic DNAs. However, the proteins may well act in a redundant fashion so that deficiency of one member of the family of MBDs may be compensated for by one of the others. Resolution of these issues is likely to stimulate much further research because it is clear that we do not fully understand how the methylation signal is interpreted in the cell. Unlike its defined role in preventing transcriptional initiation, methylation does not stop transcript elongation in mammals (*31*, *32*) but it does in *Neurospora* (*33*). This property of methylation may be essential in allowing mammalian cells to prevent spurious initiation of transcription at potentially active promoters of intragenomic parasites embedded within their genes without interfering with the regulation of the host gene. Figure 1 also shows how methylation of silencer (*34*, *35*) or insulator (*36*) elements blocks the binding of the cognate binding proteins, potentially abolishing their repressive activities on gene expression. Therefore, mammalian cells have taken full advantage of altering the appearance of the major groove of DNA, which can then result in alternative effects on transcription through the modulation of the interaction with DNA binding proteins.

Does Methylation Regulate Tissue Specific Gene Expression?

That DNA modification patterns are tissue-specific is incontrovertible (*37*) and has led to the hypothesis that variable methylation patterns are responsible for tissue-specific gene expression. However, many experiments examining relations between methylation and gene expression have been compromised in that the precise locations of the methylation changes relative to the transcription start or other regulatory regions have not been precisely defined. Also, as Yoder *et al.* (*38*) have pointed out, a large number of the sites investigated in these experiments are located in transposable elements, making the relevance of the observed changes to gene expression obscure. There have been few, if any, detailed studies showing changes in CpG-poor promoters that can be linked directly with transcriptional repression. Thus, even though there is clear evidence that the binding of transcription factors such as *myc* to their cognate sites can be blocked by methylation (*39*), detailed experiments on the potential roles of these kinds of changes in defined regulatory regions during development have not been performed. Likewise, with few exceptions, CpG islands on autosomal gene promoters do not become methylated in differentiated cells. For example, the CpG island in the muscle determination gene *MyoD1* is not methylated in nonexpressing tissues such as the brain (*40*). The potential role of DNA methylation in the control of tissue-specific gene expression, therefore, remains in question. Although specialized genes are variably methylated in differentiated cells, the link with expression remains largely unanswered.

Methylation in Human Disease

Cytosine methylation is a major contributor to the generation of disease-causing germline mutations (*41*) and somatic mutations that cause cancer (*42*). Recent work has shown that the abnormal methylation of the promoters of regulatory genes causes their silencing and is a substantial pathway to cancer development (*43*, *44*). Much excitement has been generated by findings that a rare disease [immunodeficiency, centromeric instability, facial anomalies syndrome (ICF)] (*2*, *18*) and a common kind of mental retardation in young girls (Rett syndrome) are both potentially caused by

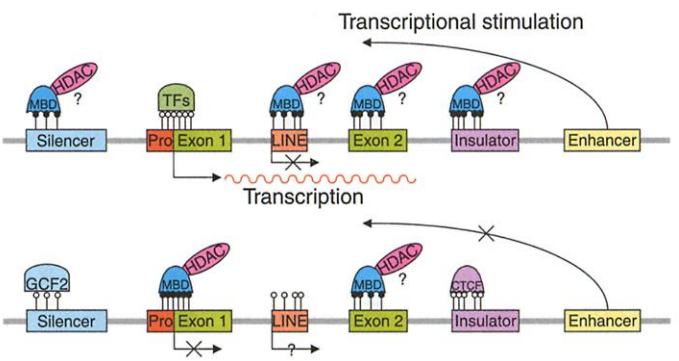

Fig. 1. How DNA methylation can influence the binding of proteins to DNA, thus altering expression of a hypothetical mammalian gene. (Top) Genes with unmethylated (open lollipops), active CpG island promoters (Pro), have transcription factors (TFs) at the transcription initiation site. Transcripts initiated here proceed through the downstream elements even though they are methylated (closed lollipops) and presumably are coated with methyl CpG binding domain proteins (MBDs) and histone deacetylases (HDACs). Note that the precise roles of the MBDs have not yet been defined—other proteins could play important roles in these processes. The enhancer is functional because the silencer and insulator are methylated and, thus, not occupied by their respective cognate proteins. Methylation here is permissive for expression. Transcripts are not initiated at the potentially active LINE element because it is methylated (*46*). Note that most autosomal genes with CpG island promoters are unmethylated independently of the level of expression; methylation of the promoter is not involved in fine control. (Bottom) A permanently silenced gene such as an imprinted gene or a gene on the inactive X chromosome. Here the promoter is methylated, leading to binding of MBDs, HDACs, other transcriptional suppressors, and chromatin compaction (*26*, *27*). The transcription factors, which normally regulate gene expression, are not able to access the promoter. The illustration also shows how lack of methylation in a silencer (*35*) or insulator (*36*) can lead to binding of the cognate proteins [e.g., GCF2 (GC binding factor 2) or CTCF (CTC binding factor), respectively], thus preventing the enhancer from functioning. Transposable elements here exemplified by a LINE element are sometimes unmethylated and, therefore, can potentially initiate transcription and transposition (*47*).

changes in the methylation machinery. ICF patients have mutations in the *DNMT3b* gene, which leads to undermethylation of satellite DNA and specific chromosomal decondensation. Rett syndrome patients have mutations in one of the MBDs (*MeCP2*) and therefore may not be able to interpret the methylation signal correctly (*45*). These two diseases suggest that methylation is not only needed to complete embryonic development but is also required for development after birth. The price for the requirements that methylation is essential in mammals seems to be a substantial increase in cancer risk.

The study of DNA methylation in mammals has been stimulated by the identification of the key enzymes that methylate DNA and their interactions with DNA and DNA binding proteins, as well as by the link between methylation and chromatin structure. Perhaps the major function of methylases in mammals is in the long-term silencing of noncoding DNA in the genome, which contains a very substantial portion of repetitive elements. Lack of methylation in promoters of essential genes allows them to be potentially active and to be regulated by other processes. On the other hand, we still do not know whether the specific patterns seen in the genes of differentiated cell types are involved in transcriptional control or whether they simply reflect altered chromatin states.

The DNA methylation field is currently in a state of high activity as the links between stable epigenetic states, chromatin structure, and heterochromatinization begin to become clearer. Disruption of these links leads to substantial disease states, including chromosomal integrity, mental retardation, and cancer. Understanding how epigenetic states are established and maintained and developing strategies to modify them therapeutically is, therefore, likely to be an area of intense future research.

References and Notes

1. E. Li, T. H. Bestor, R. Jaenisch, *Cell* **69**, 915 (1992).
2. M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **99**, 247 (1999).
3. T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, *J. Mol. Biol.* **203**, 971 (1988).
4. M. Okano, S. Xie, E. Li, *Nature Genet.* **19**, 219 (1998).
5. F. Larsen, G. Gundersen, R. Lopez, H. Prydz, *Genomics* **13**, 1095 (1992).
6. M. Gardiner-Garden, M. Frommer, *J. Mol. Biol.* **196**, 261 (1987).
7. D. Takai, P. A. Jones, unpublished data.
8. S. Delgado, M. Gomez, A. Bird, F. Antequera, *EMBO J.* **17**, 2426 (1998).
9. G. Grigg, S. Clark, *Bioessays* **16**, 431 (1994).
10. D. Macleod, J. Charlton, J. Mullins, A. P. Bird, *Genes Dev.* **8**, 2282 (1994).
11. M. Brandeis *et al.*, *Nature* **371**, 435 (1994).
12. I. G. Lin, T. J. Tomzynski, Q. Ou, C. L. Hsieh, *Mol. Cell. Biol.* **20**, 2343 (2000).
13. L. Han, I. G. Lin, C. L. Hsieh, *Mol. Cell. Biol.* **21**, 3416 (2001).
14. W. Reik, W. Dean, J. Walter, *Science* **293**, 1089 (2001).
15. K. D. Robertson *et al.*, *Nature Genet.* **25**, 338 (2000).
16. M. R. Rountree, K. E. Bachman, S. B. Baylin, *Nature Genet.* **25**, 269 (2000).
17. F. Fuks, W. A. Burgers, N. Godin, M. Kasai, T. Kouzarides, *EMBO J.* **20**, 2536 (2001).
18. G. L. Xu *et al.*, *Nature* **402**, 187 (1999).
19. D. Jahner, R. Jaenisch, *Nature* **315**, 594 (1985).
20. A. N. Magewu, P. A. Jones, *Mol. Cell. Biol.* **14**, 4225 (1994).
21. M. S. Turker, *Semin. Cancer Biol.* **9**, 329 (1999).
22. J. K. Christman, G. Sheikhnejad, C. J. Marasco, J. R. Sufrin, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7347 (1995).
23. D. Carotti, S. Funiciello, F. Palitti, R. Strom, *Biochemistry* **37**, 1101 (1998).
24. M. Fatemi, A. Hermann, S. Pradhan, A. Jeltsch, *J. Mol. Biol.* **309**, 1189 (2001).
25. J. P. Issa *et al.*, *Nature Genet.* **7**, 536 (1994).
26. P. L. Jones *et al.*, *Nature Genet.* **19**, 187 (1998).
27. X. Nan *et al.*, *Nature* **393**, 386 (1998).
28. J. Guy, B. Hendrich, M. Holmes, J. E. Martin, A. Bird, *Nature Genet.* **27**, 322 (2001).
29. R. Z. Chen, S. Akbarian, M. Tudor, R. Jaenisch, *Nature Genet.* **27**, 327 (2001).
30. B. Hendrich, J. Guy, B. Ramsahoye, V. A. Wilson, A. Bird, *Genes Dev.* **15**, 710 (2001).
31. P. A. Jones, *Trends Genet.* **15**, 34 (1999).
32. M. L. Gonzalgo *et al.*, *Cancer Res.* **58**, 1245 (1998).
33. M. R. Rountree, E. U. Selker, *Genes Dev.* **11**, 2383 (1997).
34. M. Constancia, W. Dean, S. Lopes, T. Moore, G. Kelsey, W. Reik, *Nature Genet.* **26**, 203 (2000).
35. S. Eden *et al.*, *EMBO J.* **20**, 3518 (2001).
36. A. T. Hark *et al.*, *Nature* **405**, 486 (2000).
37. A. Yeivin, A. Razin, in *DNA Methylation: Molecular Biology and Biological Significance*, J. P. Jost, H. P. Saluz, Eds. (Birkhauser Verlag, Basel, Switzerland, 1993), vol. 64, pp. 523–568.
38. J. A. Yoder, C. P. Walsh, T. H. Bestor, *Trends Genet.* **13**, 335 (1997).
39. G. C. Prendergast, E. B. Ziff, *Science* **251**, 186 (1991).
40. P. A. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6117 (1990).
41. D. N. Cooper, H. Youssoufian, *Hum. Genet.* **78**, 151 (1988).
42. W. M. Rideout III, G. A. Coetzee, A. F. Olumi, P. A. Jones, *Science* **249**, 1288 (1990).
43. S. B. Baylin, J. G. Herman, J. R. Graff, P. M. Vertino, J. P. Issa, *Adv. Cancer Res.* **72**, 141 (1998).
44. P. A. Jones, P. W. Laird, *Nature Genet.* **21**, 163 (1999).
45. R. E. Amir *et al.*, *Nature Genet.* **23**, 185 (1999).
46. R. J. W. O'Neill, M. J. O'Neill, J. A. Marshall Graves, *Nature* **393**, 68 (1998).
47. R. W. Hendriks, H. Hinds, Z. Y. Chen, I. W. Craig, *Genomics* **14**, 598 (1992).

48. P.A.J. and D.T. are supported by grant R01 CA82422 from the NCI.

REVIEW

DNA Methylation and Epigenetic Inheritance in Plants and Filamentous Fungi

Robert A. Martienssen* and Vincent Colot*†

Plants and filamentous fungi share with mammals enzymes responsible for DNA methylation. In these organisms, DNA methylation is associated with gene silencing and transposon control. However, plants and fungi differ from mammals in the genomic distribution, sequence specificity, and heritability of methylation. We consider the role that transposons play in establishing methylation patterns and the epigenetic consequences of their perturbation.

Epigenetic changes, so-called "epimutations," occur because nucleotide sequence is not the only form of genetic information in the cell: Chromosomal proteins and DNA methylation can also be inherited, with important phenotypic consequences (*1*). In plants and filamentous fungi, genomic methylation is restricted mostly to transposons and other repeats (*2*, *3*). In mammals, by contrast, coding sequences are methylated as well, except for the so-called CpG islands that often encompass the first exons of genes (*4*). This difference likely reflects the colonization of mammalian introns by transposons, and the possibility of methylation spreading into flanking exons (*3*). Of course, the human genome contains far less exonic DNA (<2%) than transposons (>45%) (*5*), which thus contribute more to the level of cytosine methylation overall (*4*). Naïvely, when transposons lose methylation they become activated, while when genes gain extra methylation, they become silenced.

Some of the first epimutations were observed by B. McClintock, who noted that transposons underwent cycles of inactivity in maize. These "changes in phase" are associated with changes in DNA methylation and are distinct from "changes in state," which are usually sequence rearrangements (*6*). Both defective and intact transposons can also be modified epigenetically ("preset") after exposure to an active transposon (*6*).

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

*To whom correspondence should be addressed. E-mail: martiens@cshl.org (R.A.M); colot@evry.inra.fr (V.C.).

†Permanent address: Unité de Recherches en Génomique Végétale (INRA-CNRS), 2 rue Gaston Crémieux, CP 5708, 91057 Evry Cedex, France.