Commentary

$C^mC(a/t)GG$ methylation: A new epigenetic mark in mammalian DNA?

Matthew C. Lorincz* and Mark Groudine*†‡

*Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109; and †Department of Radiation Oncology, University of Washington, Seattle, WA 98195

Twenty-five years ago, based on the knowledge of cytosine methylation in higher organisms and the newly discovered bacterial adenine methyltransferase, Riggs (1) and Holliday and Pugh (2) independently proposed that the covalent modification of DNA by methylation might serve as a means to propagate heritable expression states in eukaryotes. In the years since, the association between cytosine methylation and transcriptional silencing in mammalian cells has become well established (3), and a number of proteins that catalyze the transfer of a methyl group to the 5-carbon of the cytosine pyrimidine ring have been cloned and characterized. These DNA methyltransferases (m5C-MTases) are encoded by a diverse family of genes found in prokaryotes as well as all four groups of eukaryotes (4). In mammals, cytosines are methylated predominantly in the context of the 5′-CpG-3′(CG) dinucleotide, and the majority of these sites are methylated. Only the short CG-rich regions known as CpG islands are methylation free in normal tissues. However, non-CG cytosine methylation has also been reported, primarily in viral or stably integrated plasmid DNA sequences. Genomic sequencing of an integrated adenovirus vector, for example, revealed methylation of cytosines in the context of 5′-CpW-3′ (CW, where W = A or T) dinucleotides (5). Subsequently, bisulfite sequencing analysis of endogenous LINE-1 retroelements (6) and stably integrated plasmid DNA (7) confirmed the presence of CW methylation in mammalian cells. In the latter study, such cytosine methylation was found predominantly in the sequence CWG. Interestingly, the methylation state of plasmid DNA premethylated at such CWG sites was faithfully maintained *in vivo* (7), suggesting that mammalian cells are capable of establishment and maintenance of cytosine methylation in CWG as well as CG sites. Taken together, these results suggested that CWG methylation might be used to stably "tag" DNA in eukaryotes. However, further evidence for the existence of such an epigenetic system has been lacking.

In this issue of PNAS, Malone *et al.* describe the presence of a subclass of CWG methylation, namely that of the internal cytosine within the sequence CCWGG (C^mCWGG), in an endogenous mammalian gene (8). The authors studied a group of primary effusion lymphoma (PEL) and myeloma cell lines that no longer express many B lineage-specific genes. To test whether the silent state of such genes is associated with DNA methylation, several cell lines in which the B cell-specific B29 gene is silent were analyzed. Whereas bisulfite genomic sequencing of the B29 promoter revealed dense methylation exclusively at CG sites in several of these lines, a significant number of CCWGG sites were methylated in the B29 promoter region in one PEL and one myeloma line.

Evidence for this class of cytosine methylation in mammalian cells comes from several previous studies. Bisulfite analysis of genomic DNA from murine erythroleukemia (MEL) cells infected with a Moloney murine leukemia virus vector revealed that several proviral CCWGG sites were methylated *de novo* shortly after infection (9). Additionally, by using the C^mCWGG-sensitive restriction enzyme *Eco*RII and Southern blotting of genomic DNA from normal peripheral blood leukocytes (PBL), Franchina *et al.* reported a high level of CCTGG methylation flanking the CpG island of the myogenic *myf-3* gene, although no CCAGG methylation was detected (10). Moreover, the study describing efficient maintenance of CWG methylation in mammalian cells involved the introduction of plasmid DNA isolated from *Escherichia coli* expressing the Dcm m5C-MTase, which methylates CCWGG sites (7). Thus, the observed maintenance of CWG methylation may have been the result of recognition of the CCWGG pentanucleotide. The observation of CCWGG methylation in several systems suggests that this site may be the target sequence of a eukaryotic epigenetic modification system.

Which enzymes are responsible for establishment and maintenance of CCWGG methylation? The diversity of target recognition sites already described for bacterial members of the m5C-MTase family (11) suggest that one or more of the mammalian m5C-MTase homologs may be capable of methylating CCWGGs. Four known or putative mammalian m5C-MTases have been identified thus far: Dnmt1, Dnmt2, Dnmt3A, and Dnmt3B. Analysis of the human genome project database reveals no additional genes in the highly conserved m5C-MTase family, suggesting that the proteins encoded by the known m5C-MTase genes are responsible for the constellation of cytosine methylation in the mammalian genome. Dnmt1, the first mammalian m5C-MTase cloned, shows a strong preference for CGs, particularly when hemi-methylated (12), suggesting that this enzyme is responsible for the maintenance of CG methylation. The closely related Dnmt3A and Dnmt3B proteins on the other hand show strong *de novo* CG MTase activity (13), apparently confirming the hypothesis that the establishment and maintenance of CG methylation are carried out by distinct m5C-MTases.

Recently however, several groups have reported that Dnmt3A also methylates non-CG cytosines. As determined by nearest neighbor analysis and bisulfite sequencing, ^mCW represent ≈15% of the total number of methylated cytosines in embryonic stem (ES) cells (14), which express Dnmt3A at high levels. Expression of dnmt3A in *Drosophila*, which lack a *Dnmt3* homolog and show no cytosine methylation in adult tissues, revealed that Dnmt3A is capable of methylating CW dinucleotides (14). Whereas the non-CG methylated sites showed no clear flanking consensus sequence *in vivo*, the internal cytosine of CCAGG is efficiently methylated by Dnmt3A *in vitro* (15). Analysis of CCWGG methylation in dnmt3A$^{-/-}$ (16)

See companion article on page 10404.

‡To whom reprint requests should be addressed. E-mail: markg@fhcrc.org.

www.pnas.org/cgi/doi/10.1073/pnas.201392598

vs. normal ES cells will be informative in this regard.

In contrast to *Dnmt1*, *Dnmt3A*, and *Dnmt3B*, the function of the widely expressed *Dnmt2* gene remains unclear. The Dnmt2 protein shows neither CG MTase activity *in vivo* (17), nor general cytosine MTase activity *in vitro* (18). Phylogenetic analysis of this putative MTase reveals that *Dnmt2* is closely related to the *Drosophila* homolog *dDnmt2* and the *Schizosaccharomyces pombe* gene *pmt1* (19, 20). Pmt1 is catalytically inert because of the insertion of a Ser residue between the Pro-Cys motif found at the active site of related DNA MTases. However, when this amino acid is deleted, Pmt1 binds to CCWGGs and methylates the internal cytosine residue (21). *Drosophila* dDnmt2 is expressed primarily during embryogenesis (22), the period during which cytosine methylation is detected (23). As with Pmt1, CGs do not seem to be the preferred target sites. Rather, dDnmt2 preferentially methylates CW sites, although an expanded consensus sequence has yet to be defined. Given that mammalian Dnmt2 does not methylate CGs (17), experiments designed to test whether CCWGG is a substrate for Dnmt2 *in vivo* are clearly warranted. The availability of Dnmt2-deficient ES cells (17) should greatly facilitate such studies.

Interestingly, Malone *et al.* (8) found that the B29 promoter typically contains either C^{m}CWGGs or ^{m}CGs, but not both types of cytosine methylation together. An inverse correlation between CG and CCWGG methylation was also observed in the proviral (9) and *myf-3* studies (10): in contrast to normal PBL, in which CCTGG but no CG methylation was detected, non-Hodgkin's lymphoma cells show a significant level of CG methylation but no CCTGG methylation of the *myf-3* gene (10). Taken together, these results indicate that chromatin marked by C^{m}CWGG methylation might be refractory to CG methylation and *vice versa*. Precedent for an interplay between two methylation types can be found in *Arabidopsis*, where CG and CpNpG methylation are carried out by the dnmt1 homolog MET1, and the plant-specific CHROMOMETHYLASE3, respectively (24). Whereas inhibition of MET1 activity leads to the hypomethylation of CGs as expected, hypermethylation of CpNpGs in some regions of the genome is also observed. Thus, the absence of CG methylation seems to stimulate CpNpG methylation in wild-type plants.

Although the majority of CGs in the mammalian genome are methylated, CpG islands associated with promoters are normally unmethylated. The mutually exclusive nature of CG and CCWGG methylation may shed light on this paradox: might CpG islands be "protected" from CG methylation by the presence of CCWGG methylation? If so, then the aberrant methylation of CpG islands often observed in transformed cells might result from the loss of CCWGG methylation. The association of CCWGG methylation with inactive promoters would seem to contradict this model; however, the influence of CCWGG methylation on promoters associated with CpG islands has yet to be addressed. Determination of the methylation state of CCWGGs within CpG islands in normal, and transformed cells may lead to a better understanding of the etiology of CpG island methylation.

A fundamental question raised by Malone *et al.* (8) is that of the function of CCWGG methylation. The correlation between CCWGG methylation and the absence of B29 expression prompted the authors to test directly whether C^{m}CWGG methylation is sufficient to repress expression (8). Transient transfection of a B29 promoter-luciferase reporter construct methylated exclusively at CCWGG sites yielded a 50% reduction in reporter activity compared with unmethylated controls (8). Taken together with the retroviral (9) and *myf-3* gene studies (10), which also showed a correlation between CCWGG methylation and transcriptional repression, these results suggest that C^{m}CWGGs might mark genetic elements for transcriptional silencing.

How might such repression be mediated? In the case of the CG dinucleotide, two general mechanisms of repression have been described in vertebrates: a "direct" mechanism (25), whereby methylation of CGs within a transcription factor binding site inhibits factor binding, and an "indirect" mechanism (3), which involves the binding of one of the methyl-DNA binding domain (MBD) family of proteins, such as MeCP2, to ^{m}CGs. These MBD proteins interact with corepressor complexes, such as Sin3A (3), which mediates repression at least in part by histone deacetylation. To determine whether the binding of transcription factors known to interact with the B29 promoter region is perturbed by C^{m}CWGG methylation, electrophoretic mobility shift assays were conducted with B29 dsDNA oligonucleotide probes (8). Whereas several protein complexes were detected with the unmethylated probe, a distinct set of reduced mobility bands were detected when the same probe was methylated at CCWGG sites. Importantly, cold unmethylated competitor did not inhibit this unique banding pattern, suggesting that the complement of transcription factors normally bound to the B29 promoter may be displaced by an as yet uncharacterized complex that associates with this region in a CCWGG methylation-dependent manner. The inability of a CG methylated oligonucleotide to compete with this novel complex suggests that it does not include members of the MBD protein family. Clearly, identification of the uncharacterized C^{m}CWGG-binding proteins will be essential to understanding the mechanism of C^{m}CWGG-mediated repression.

The absence of m5C-MTase genes in representative species of protists, fungi, and animals suggests that alternative mechanisms of genomic marking can substitute for both CG and CCWGG methylation in some eukaryotes. Nevertheless, in mammals, CG methylation is believed to play an important role in genomic imprinting (26), X chromosome inactivation (1), and repression of parasitic sequence elements (27). Whereas C^{m}CWGGs also seem to mark genes for transcriptional repression, little is known about the biological role for which this form of methylation is put to use. Chromosomal processes other than transcription, such as replication or repair, should also be considered. Given that most studies of DNA methylation in mammals have focused on the CG dinucleotide, neither the prevalence nor the genomic distribution of CCWGG methylation has been systematically addressed. The development of techniques designed to study genome-wide patterns of methylation, coupled with the use of genetic mutants of each of the known mammalian m5C-MTases, should help to resolve some of the unanswered questions concerning CCWGG methylation.

We thank Steven Henikoff, En Li, and Dirk Schübeler for insightful discussions of the DNA methylation field.

1. Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14,** 9–25.
2. Holliday, R. & Pugh, J. E. (1975) *Science* **187,** 226–232.
3. Bird, A. P. & Wolffe, A. P. (1999) *Cell* **99,** 451–454.
4. Colot, V. & Rossignol, J. L. (1999) *Bioessays* **21,** 402–411.
5. Toth, M., Muller, U. & Doerfler, W. (1990) *J. Mol. Biol.* **214,** 673–683.
6. Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P. & Warren, W. D. (1997) *J. Biol. Chem.* **272,** 7810–7816.
7. Clark, S. J., Harrison, J. & Frommer, M. (1995) *Nat. Genet.* **10,** 20–27.
8. Malone, C. S., Miner, M. D., Doerr, J. R., Jackson, J. P., Jacobsen, S. E., Wall, R. & Teitell, M. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10404–10409. (First Published August 14, 2001; 10.1073/pnas.181206898)
9. Lorincz, M. C., Schubeler, D., Goeke, S. C., Walters, M., Groudine, M. & Martin, D. I. (2000) *Mol. Cell. Biol.* **20,** 842–850.
10. Franchina, M. & Kay, P. H. (2000) *DNA Cell Biol.* **19,** 521–526.

11. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J. & Wilson, G. G. (1994) *Nucleic Acids Res.* **22,** 1–10.
12. Bestor, T. H. & Ingram, V. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5559–5563.
13. Okano, M., Xie, S. & Li, E. (1998) *Nat. Genet.* **19,** 219–220.
14. Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. & Jaenisch, R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5237–5242.
15. Gowher, H. & Jeltsch, A. (2001) *J. Mol. Biol.* **309,** 1201–1208.
16. Okano, M., Bell, D. W., Haber, D. A. & Li, E. (1999) *Cell* **99,** 247–257.
17. Okano, M., Xie, S. & Li, E. (1998) *Nucleic Acids Res.* **26,** 2536–2540.
18. Van den Wyngaert, I., Sprengel, J., Kass, S. U. & Luyten, W. H. (1998) *FEBS Lett.* **426,** 283–289.
19. Tweedie, S., Ng, H. H., Barlow, A. L., Turner, B. M., Hendrich, B. & Bird, A. (1999) *Nat. Genet.* **23,** 389–390.
20. Yoder, J. A. & Bestor, T. H. (1998) *Hum. Mol. Genet.* **7,** 279–284.
21. Pinarbasi, E., Elliott, J. & Hornby, D. P. (1996) *J. Mol. Biol.* **257,** 804–813.
22. Lyko, F., Whittaker, A. J., Orr-Weaver, T. L. & Jaenisch, R. (2000) *Mech. Dev.* **95,** 215–217.
23. Lyko, F., Ramsahoye, B. H. & Jaenisch, R. (2000) *Nature (London)* **408,** 538–540.
24. Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S. & Jacobsen, S. E. (2001) *Science* **292,** 2077–2080.
25. Watt, F. & Molloy, P. L. (1988) *Genes Dev.* **2,** 1136–1143.
26. Ferguson, S. A., Sasaki, H., Cattanach, B. M. & Surani, M. A. (1993) *Nature (London)* **362,** 751–755.
27. Yoder, J. A., Walsh, C. P. & Bestor, T. H. (1997) *Trends Genet.* **13,** 335–340.