Programming of DNA Methylation Patterns

Howard Cedar and Yehudit Bergman

Department of Developmental Biology and Cancer Research, Hebrew University Medical School, Ein Kerem, Jerusalem, Israel 91120; email: cedar@cc.huji.ac.il, yehuditb@ekmd.huji.ac.il

Abstract

DNA methylation represents a form of genome annotation that mediates gene repression by serving as a maintainable mark that can be used to reconstruct silent chromatin following each round of replication. During development, germline DNA methylation is erased in the blastocyst, and a bimodal pattern is established anew at the time of implantation when the entire genome gets methylated while CpG islands are protected. This brings about global repression and allows housekeeping genes to be expressed in all cells of the body. Postimplantation development is characterized by stage- and tissue-specific changes in methylation that ultimately mold the epigenetic patterns that define each individual cell type. This is directed by sequence information in DNA and represents a secondary event that provides long-term expression stability. Abnormal methylation changes play a role in diseases, such as cancer or fragile X syndrome, and may also occur as a function of aging or as a result of environmental influences.

Keywords

maintenance, repression, chromatin, development, reprogramming, imprinting
INTRODUCTION

Although the existence of methylated nucleotides has been known for a long time, studies of its function in vivo mostly concentrated on its role as a restriction modification mechanism in bacteria, where 6-methyladenine or 5-methylcytidine (5mC) appear at fixed nucleotide sequences throughout the genome. Nucleotide analysis of animal and plant DNA indicated that these organisms also have a fair amount of 5mC. Strikingly, when this DNA was subjected to nearest-neighbor analysis, a method that detects and quantifies the bases located 5’ adjacent to any labeled nucleotide, it was determined that almost all of this methylation is concentrated in the dinucleotide sequence CpG (1). Furthermore, in any particular cell type, only a portion of the CpGs is actually methylated (2). Taken together, these observations on the placement and distribution of methylation strongly suggested that this modification must play a different biological role in animals and plants as compared to bacteria.

PRINCIPLES OF METHYLATION

Methylation Metabolism

In order to gain some insight into the significance of this modification system, it was important to ask whether DNA methyl groups are freely metabolized throughout the genome or, alternatively, are located at fixed positions in each cell type. This was accomplished by using bacterial restriction enzymes to assay methylation at specific sites in the DNA (3). Enzymes such as HpaII (CCGG) or HhaI (GCGC), which have a CpG dinucleotide in their recognition sequences, are inhibited by methylation on the internal C residue. This unmethylated site can be cut, whereas the same site in a methylated form remains undigested. Using this assay, it was shown that different CpG residues in the genome are either highly methylated or present in an unmethylated form, strongly suggesting that methyl groups have fixed positions (4).

One of the major steps in understanding how methyl groups are organized and managed in the genome came about through the use of DNA-mediated gene transfer to stably insert foreign sequences into the endogenous genome.

5mC: 5-methylcytidine
of fibroblast cells in culture. Plasmid DNA derived from *Escherichia coli* is unmethylated at all its HpaII (CCGG) recognition sequences, but these same sites could be artificially methylated in vitro using the specific methylases found in the same *Haemophilus parainfluenza* bacteria (5). Strikingly, unmethylated DNA remained unmodified in these cells, but methylated DNA retained its original methyl groups, even after many generations of growth in culture. These studies clearly showed that methylation is not the result of ongoing transient metabolism, but rather methylation marks must be located at specific sites in the genome. Furthermore, these patterns appeared to be stably maintained through cell division (6–8).

**Maintenance of DNA Methylation Patterns**

In addition to demonstrating that methylation patterns can be faithfully maintained, these transfection experiments also revealed important principles about the mechanism for this process. Because methylation was added artificially in vitro prior to the introduction of DNA into cells, maintenance of the methylation state clearly had nothing to do with the DNA sequence. This clearly suggested that modification patterns in somatic cells do not come about by the recruitment of factors through specific local sequence motifs. Rather, it appeared that maintenance must be accomplished by some sort of autonomous mechanism that can actually read and copy the modification pattern per se at the time of replication.

The key to understanding this process evolved from studies of the enzymatic DNA methylation activity found in crude extracts from somatic cells. These experiments showed that, although completely unmodified DNA is a rather poor substrate for methylation, hemimethylated DNA (i.e., methylated on one strand) works extremely efficiently with a 100-fold better $K_m$ for this reaction (9). As previously noted, methyl moieties are located within CpG residues, and because these sites have strand symmetry, these results implied that DNA must be normally methylated on both strands. During replication, synthesis of the new strand generates a hemimethylated site that is then recognized by the maintenance methylase and therefore is methylated on the opposite strand, thereby regenerating the original dimethylated state originally present in the mother cell. In contrast, unmethylated CpG residues still appear completely unmethylated during replication and therefore do not constitute a substrate for the enzyme. In this way, the methylation pattern on the native strands serves as a template for regenerating methylation patterns during replication (10).

It is clear that the basis for this semiconservative mechanism derives from the fact that methyl groups in CpG residues are symmetrically disposed on both strands of the DNA. Strong support for this idea came from studies of plant DNA. Nearest-neighbor analysis of this DNA originally indicated that all C-containing dinucleotides are partially methylated. Upon closer examination, however, it could be seen that all of these methyl groups are actually located in CpNpG trinucleotide symmetrical sequences where N can be any of the four bases. Thus, every instance of modified C on one strand of the DNA is opposed by an equivalent methyl group two nucleotides over on the other strand (11). At the time of replication, methylation on the native DNA strand can thus serve as a template for complementing the methylation pattern on the newly synthesized strand, thus reproducing the methyl profile present in the mother cell.

It is now known that in animal cells the maintenance reaction is carried out by the enzyme Dnmt1 (12), which is perpetually localized to replication foci (13) and therefore is constantly available to provide CpG maintenance function. Exactly how hemimethylated DNA is recognized is not well understood (14, 15), but recent studies on Dnmt1 indicate that it does not operate alone. Rather, it probably works as part of a larger complex that includes other essential factors, such as Np95 (UHRF1) (16–18). Ultimately, these complexes constitute the biochemical basis for epigenetic
CpG island: a region that contains a high density of CpGs and is located at the promoters of many genes.

Memory by providing an enzymatic platform for copying methylation patterns in a semiconservative manner in the same way that DNA sequence itself is reproduced from generation to generation.

Methylation and Gene Repression

Once it became possible to use restriction enzymes for analyzing the DNA methylation patterns of specific endogenous genes, it became immediately obvious that this modification is correlated with gene repression. Tissue-specific genes were found to be highly methylated in most tissue samples but undermethylated in their tissue of expression (19). At the same time, housekeeping genes were shown to have a unique CpG island promoter structure, which is constitutively unmethylated in every cell (20, 21). Furthermore, early genomic studies based on analysis of total mRNA even demonstrated that active genes, in general, are undermethylated as compared to inactive DNA (22). This correlation has also been confirmed over and over again using more comprehensive assays, such as bisulfite analysis to measure endogenous gene DNA methylation (23).

Although these studies demonstrate a strong association with repression, it was still necessary to actually test whether DNA methylation has a causal effect on gene expression, and this was ultimately accomplished by employing DNA-mediated gene transfer. These experiments showed that unmethylated genes are actively transcribed when inserted into the genome, whereas the exact same sequences are repressed if the inserted DNA had been premethylated in vitro (6, 7, 24). Because the only difference between these templates is the presence of methyl groups, these studies provided convincing proof that DNA methylation itself is responsible for gene inhibition. Even though these experiments were based exclusively on the analysis of exogenous gene sequences, the implication of these results is that DNA methylation can explain why the homologous endogenous genes are completely silenced in these same exact cells, and this concept has been reinforced by subsequent transfection studies in vitro (25) and transgenic studies in vivo (26, 27), demonstrating that methylated DNA molecules are repressed.

When DNA methylation was first being explored as a mechanism of gene repression, it was commonly thought that methyl groups work by preventing the binding of key transcription factors much in the same way that bacterial restriction enzymes cannot cut when their recognition site is methylated. Although this may be true for a small number of specific regulatory factors, this is certainly not a general phenomenon. Many transcription factors do not have CpGs in their binding sites, and even when present, as is the case for Sp1 (28), DNA methylation does not necessarily inhibit their binding.

DNA Methylation and Chromatin Structure

A great deal of evidence points to the idea that the major effect of methyl groups is to model chromatin structure, and this may be carried out at many different levels. Microinjection experiments, for example, have demonstrated that a gene template inserted directly into the nucleus of cells in culture is initially unaffected by DNA methylation. Only after these substrates have had a chance to get packaged into a chromosomal structure does one begin to see the effects of this modification on transcription (29). The most convincing evidence for this idea comes from DNA-mediated cell transfection studies showing that unmethylated substrates are packaged into an open chromatin structure following their integration into the genome, whereas the exact same DNA remains completely resistant to DNase I if it is methylated (30). Because these experiments were carried out using bacterial DNA sequences that do not harbor any eukaryotic regulatory information, one can conclude that methyl moieties themselves must be responsible for generating a closed chromatin structure regardless of sequence context.

Although the precise mechanisms by which DNA methylation affects chromatin packaging...
have not been completely elucidated, many studies have concentrated on nucleosome structure. Early experiments indicated that most methylated CpG sites are concentrated within the central core of nucleosomes, as opposed to internucleosomal regions, suggesting that the positioning of this modification may have an intrinsic effect on where nucleosomes reside on the DNA (31, 32), and these results have recently been confirmed by sophisticated genome-wide analysis (33). In addition to this intrinsic effect, methylation may also play a role in regulating other factors that ultimately impinge on nucleosome displacement.

Another way that methylation mediates gene repression is through methyl-binding proteins, such as MeCP2, MBD2, and MBD3 (34). These factors can specifically recognize methyl moieties and, once bound, may model local chromatin structure, perhaps by recruiting modifying enzymes that bring about overall histone deacetylation (35, 36) or the methylation of specific lysine residues (37). Alternatively, methylation may actually prevent the binding of chromatin proteins, such as boundary-forming CTCF (38) or Cfp1 (39), which is known to be exclusively located on unmethylated CpG islands.

These observations on chromatin structure provide important insight on the significance of DNA methylation as a mechanism for stable gene silencing. Although gene expression profiles are mediated by the initial interactions with transcriptional regulatory factors, it is most likely the state of gene accessibility at the level of chromatin structure that actually determines gene activity or repression. In principle then, the chromatin conformation could provide a mold for preserving long-term gene expression profiles. The problem with this idea is that these basic structural features get disrupted in the wake of DNA polymerase advancement during every round of replication, and the newly made DNA must then be repackaged into chromatin (40). There is, as yet, no well-established mechanism for reproducing chromatin profiles at the replication fork (see Reference 41). Because the underlying DNA methylation pattern is autonomously copied following replication, it is possible that this modification pattern may play a role both in the generation of active chromatin over undermethylated regions as well as in the formation of closed chromatin over methylated DNA (42). It is probably in this manner that DNA methylation mediates stable gene repression.

**DNA METHYLATION PATTERNS DURING DEVELOPMENT**

**Erasure**

Studies in tissue culture have been very helpful in characterizing basic concepts of DNA methylation metabolism and function, but these experiments do not reveal very much about the precise role of this modification during normal development in vivo. To understand this process, it was necessary to first map the dynamic pattern of DNA methylation as a function of embryogenesis and organogenesis (Figure 1). Early studies using restriction enzymes indicated that methyl groups inherited from parental gametes are largely erased in the preimplantation morula and blastula (43–45). This process of erasure appears to take place in two distinct stages, with much of the paternal genome undergoing active demethylation, which begins in the zygote (46), and further demethylation takes place during the first few early embryonic replication cycles, perhaps passively as a result of Dnmt1 relocation from the nucleus to the cytoplasm (47). Although the level of DNA methylation in the blastula is very low, the exact pattern of this modification in these preimplantation cells has not yet been accurately determined. Differentially methylated regions located at imprinting centers must have a mechanism to preserve them through this erasure stage (48, 49), and other studies indicate that additional specific regions and a number of different repeated sequences may also be partially protected from this process (50, 51).
Figure 1

The generation of DNA methylation patterns during development. Almost all methylation in the gametes is erased (yellow) in the preimplantation embryo, but imprinting centers retain methylation on one allele (black). At the time of implantation, the entire genome gets methylated (blue), but the CpG islands are protected (yellow circles). Postimplantation, pluripotency genes are de novo methylated (black). Tissue-specific genes undergo demethylation (orange in Tissue 1, pink in Tissue 2) in their cell type of expression. Imprinting centers remain differentially methylated throughout development. Somatic cell reprogramming by induced pluripotent stem cells or fusion resets the methylation pattern of somatic cells to the stage of implantation.

Generation of a Bimodal Methylation Pattern

In the next stage of embryogenesis at about the time of implantation, the entire genome gets remodified through a dramatic wave of de novo methylation (Figure 1), and genetic analyses indicate that this is mediated by Dnmt3a and -3b, which are present at high concentrations at this stage of development (52). Strikingly, this process appears to generate a bimodal pattern of methylation, with most sequences becoming methylated to high levels (>80%), while CpG island-like windows are protected and therefore remain unmethylated (23). This concept of what happens in the early embryo derives from the observation that CpG island-like windows are constitutively unmethylated in a large number of adult cell types, whereas other regions are constitutively methylated (53). This includes repeated sequences in satellite DNA that make up over 50% of all methyl moieties in the nucleus (32). Because all of these tissues are ultimately derived from the early inner cell mass at the time of implantation, it is reasonable to assume that the bimodal pattern is generated at this early stage and then maintained largely intact over succeeding cell divisions. This overall picture is also supported by the observation that mouse and human embryonic stem (ES) cells derived directly from this stage of development already show this same bimodal profile of DNA methylation (53, 54).

Mechanism of CpG Island Protection

Although the precise mechanism for CpG island protection is not known, it is quite clear that this is mediated by common sequence motifs present within the islands. This was initially deduced from transfection studies in ES cells. Unlike somatic cells, these cells can actually de novo methylate exogenously...
introduced DNA sequences and, at the same time, also have the ability to recognize and protect CpG islands (55). Experiments using these cells in vitro indicated that Sp1 motifs play a role in this protection process, and this was further confirmed in vivo using transgenic mice (56, 57). Furthermore, these elements were even able to protect a non-CpG island sequence from de novo methylation at the time of implantation after being introduced as a transgene in one-cell embryos (26).

Although these studies indicate, in principle, that cis acting sequences are involved in protecting specific CpG islands from de novo methylation at the time of implantation, the general sequence rules for this mechanism have not been deciphered. Genome-wide studies indicate that almost all unmethylated CpG islands contain transcription start sites, are marked with H3K4me3 in ES cells (58, 59), and harbor many transcription factor–binding motifs (53). This suggests a model whereby sites of RNA polymerase binding in the blastocyst may serve as a mark for preventing de novo methylation during implantation. According to this idea, generation of a bimodal DNA methylation pattern in early development essentially serves to perpetuate the factor-mediated basal transcription profile of the preimplantation embryo.

Sequence Information Generates Methylation Patterns

By comparing constitutively unmethylated and methylated DNA sequences, it has been possible to derive a well-defined accurate algorithm that can distinguish between these two types of sequences, and by means of this mathematical formula, it is possible to accurately predict the basal methylation state of any sequence window in the genome (53). These studies serve to emphasize the idea that, although DNA methylation patterns represent an epigenetic mark that is not in itself inherited from the parents, the ultimate modification profile is fully determined by underlying sequence information within the DNA itself. Once this epigenetic code is deciphered, it should be possible to predict the full dynamics of DNA methylation during development.

Global Repression

De novo establishment of a bimodal methylation pattern in the early embryo has far-reaching implications for the management of gene regulation in the organism. Unlike simple biological systems where almost the entire genome is transcribed, even though only a relatively small number of genes are specifically recognized and repressed, animals have a large genome in which over 50% of the genes may be silenced in any given cell type. This type of repression pattern clearly requires a different mechanism that can carry out gene repression in a global manner without the need to recognize specific motifs on every target gene. According to this scheme, at the time of implantation, the entire genome undergoes global modification while CpG islands are protected. This insures that housekeeping genes that have CpG island promoters will be kept unmethylated and may also provide a mechanism for setting up regulatory modules, such as enhancers, in a permanent active configuration so they can bind key transcription factors at later stages of development (60).

At the same time, this wave of methylation guarantees that genes with nonisland promoters will automatically get modified and therefore be repressed in most tissues of the body (26, 27). While this relatively sparse background methylation may only have a small influence on the transcription levels of tissue-specific genes in their nonexpressing cell types, it also brings about a dramatic repression of endogenous viral sequences and foreign elements throughout the genome (61). It may also cause blanket inactivation of many cryptic promoters. Because, at this early stage, the entire genome undergoes DNA modification in a nonspecific manner, this event probably represents the only time in development when methylation itself actually serves as the primary cause of silencing. Even in this case, however, DNA methylation may
represent only one of multiple factors that can mediate long-term repression through its effects on chromatin structure (62).

**Postimplantation Methylation Changes**

Following implantation, the animal genome can undergo additional changes in methylation, but these events are all of a tissue-specific or gene-specific nature. Perhaps one of the most significant of these modifications is that involved in the silencing of pluripotency genes (**Figure 2**). Genes, such as *Oct-3/4* and *Nanog*, for example, are active in the early embryo and still maintain an unmethylated transcribed promoter at the time of implantation. Following this stage, however, these genes undergo inactivation, thereby setting the stage for embryonic differentiation (63). Using ES cells as a model system, it has been possible to learn about the mechanism of this repression process, which occurs in three steps.

With the onset of differentiation, *Oct-3/4* transcription is initially turned off through a simple repression-factor mechanism. In the second step, the histone methyltransferase G9a is specifically recruited to the promoter of these genes, thus facilitating histone deacetylation and subsequent methylation of H3K9, which then binds HP1, forming heterochromatin (64). Finally, this same G9a complex can also recruit the de novo methylases Dnmt3a and -3b, thus bringing about methylation of the pluripotency-gene promoters themselves (**Figure 2**) (65–67). Although this de novo methylation represents a secondary event, it still plays an important role in stabilizing the silent state. Indeed, experiments in vitro, using mutant ES cells, clearly show that both the transcriptional repression and heterochromatinization steps are easily reversed in culture, but once DNA methylation has occurred, it is no longer possible to return to the pluripotent state (64, 65). In a similar manner, it has been demonstrated that viruses, initially inactivated by transient mechanisms in ES cells, may then become permanently silenced following differentiation-mediated de novo methylation (68).

Another major event that occurs soon after implantation throughout all cells of the embryo is the inactivation of one X chromosome in female animals. Here too, repression takes place as a multistep process, beginning with rapid chromosome-wide changes in replication timing, gene expression, and chromatin structure (69), followed by de novo methylation of CpG island promoters (70). In this case, as well, DNA
Polycomb complex: a complex of proteins bound at specific genes where they cause repression, probably through histone modification

Polycomb complex:
a complex of proteins bound at specific genes where they cause repression, probably through histone modification

Tissue-Specific Methylation Patterns

Other alterations of the basic bimodal methylation pattern occur in a cell-type-specific manner. Many genes that are silenced throughout the organism and expressed specifically in a single tissue have non-CpG island promoters that automatically undergo de novo methylation at the time of implantation. During tissue development, these genes have to be specifically recognized by cell-type-specific factors that apparently recruit the molecules needed for demethylating their promoters (72), decondensing the overlying chromatin structures and making them accessible to the transcription machinery (Figure 1). A number of experiments in different cell types have demonstrated that this demethylation occurs in an active manner that does not require DNA replication (73, 74) and is mediated by specific *cis* acting sequences (75) and *trans* acting factors (76). Because this type of demethylation is specific and requires prior recognition, it cannot be considered the primary underlying cause of gene activation. Once demethylation has occurred, however, it is possible that this serves to stably maintain chromatin accessibility.

In a manner similar to demethylation, genes may also undergo tissue-specific de novo methylation at CpG island sequences that were originally protected at the time of implantation (77). Interestingly, these targets are not necessarily associated with promoters, and many are actually located within coding regions, where their methylation is associated with gene activation (53). Although the mechanism for this reverse effect is not known, one possibility is that these internal CpG islands carry promoters for antisense transcripts whose methylation would inhibit the production of these repressive RNA molecules. Alternatively, these regions may simply contain binding sites for transcriptional repressors. A large number of these targets have been shown to undergo de novo methylation in brain tissue, and genetic studies demonstrate that this may have profound effects on gene expression and function (78).

Although the precise mechanism for targeting de novo methylation in vivo is not yet known, it appears that this may be largely mediated by the polycomb complex (Figure 3). Almost all of the sites that undergo this type of modification are known polycomb targets (53), and it has been shown that the polycomb complex has an inherent ability to recruit Dnmt3a and -3b (79). It is still not known, however, which *cis* acting elements and *trans* acting factors may be involved in determining cell-type and gene-specific modification.

ABNORMAL DNA METHYLATION

De Novo Methylation in Cancer

In the same way that normal development is dependent on the proper programming of methylation, abnormal cell behavior is also associated with alterations in DNA modification patterns. A prime example of this phenomenon is cancer. All tumors that have ever been examined show changes in DNA methylation, suggesting that this may represent a basic element of cancer biology that has a significant effect on tumor pathology. Early studies indicated, for example, that cancer is characterized by widespread demethylation (80), and using more advanced technologies, it has been possible to map this effect to specific blocks covering a large portion of the genome (81). In addition, almost all tumors undergo de novo methylation at specific sites (82). It has been commonly assumed that this modification process occurs in a completely random manner, with the final pattern seen in the tumor being determined mainly by...
Targeted de novo methylation. Genes targeted by the polycomb complex, PRC2, normally have unmethylated CpG island promoters (white circles) but are repressed by virtue of the histone methylase EZH2-mediated methylation of H3K27, which then recruits the chromodomain-containing complex, PRC1, generating a form of heterochromatin. Targeted de novo methylation (red circle) can occur at specific sites during normal development or abnormally in cancer, and this probably occurs because EZH2 itself recruits Dnmt3a and Dnmt3b. In some tumors, it has been found that PRC2 may no longer remain bound to the methylated islands, leading to a situation where flexible polycomb repression is replaced with a more stable form of methylation-mediated repression.

Figure 3

According to this selection idea, DNA methylation in the tumor serves to inactivate genes that would normally be expressed in the normal tissue, but transcription analysis showed that most methylation targets are actually inactive at the beginning. It appears that many of these sites harbor polycomb, and it may be this complex that actually recruits the methylases that then bring about this abnormal modification (Figure 3) (85–87). From this point of view, de novo methylation in cancer behaves in a programmed manner, in the sense that target sites are already predetermined independent of whether these genes are active or inactive or whether they actually play some role in tumorigenesis (88). Studies on human colon cancer suggest that de novo methylation occurs very early in the process of tumor evolution and that some of the polycomb target sites are associated with genes essential for epithelial cell differentiation, thus explaining why many tumor suppressor genes are found methylated in various cancer types. Although this may represent a small part of the methylation picture, more recent results have shown that DNA methylation in cancer is actually quite extensive, with hundreds of CpG islands in the genome getting methylated in a largely biallelic manner (83).

Role of Methylation in Cancer

The overall part played by DNA methylation in cancer may best be studied by using a genetics approach in mouse model systems. Min mice carry a deletion in one allele of Apc, a key tumor suppressor gene that is deleted in almost every case of human colon cancer. These animals develop thousands of intestinal microadenomas owing to spontaneous deletion of the second Apc allele, but only about 5% of these actually develop into full-blown adenomas (89). When these mice are genetically manipulated to express lower levels of Dnmts, or when they are treated from birth with low doses of 5-azacytidine, the number of adenomas that is generated is reduced by almost 100-fold (90), whereas the number of microadenomas is unaffected (91). These observations suggest that by preventing de novo methylation after birth, it is
Fragile X syndrome: a developmental disease caused by a triplet repeat (CGG) expansion in the FMR1 gene that brings about mental retardation.

Fragile X Syndrome

Abnormal de novo methylation also plays a role in the pathogenesis of fragile X syndrome, causing the FMR1 gene, carrying an expanded triplet repeat, to become repressed at a very early stage in development. Although the precise mechanism for this process has not been elucidated, inactivation is known to be accompanied by H3K9me3 heterochromatinization (97, 98). Taken together, these findings suggest that programming of abnormal modification may involve a common mechanism (71) whereby various histone methylases mediate unscheduled DNA methylation and repression.

DE METHYLATION

Demethylation is a common event that occurs at a number of different stages during normal development. Early in preimplantation development, for example, the entire genome is subject to a wave of demethylation that apparently erases almost all of the methyl marks inherited from the parents, and a very similar process also takes place during early stages of gametogenesis (43, 44). In addition to this global form, there are also many instances of site-specific demethylation. This usually involves tissue-specific genes that become automatically methylated at the time of implantation and then undergo demethylation in their cell type of expression (19).

Active Demethylation

Extensive efforts over several decades have been invested in deciphering the biochemical mechanisms that may be involved in this demethylation process. Considering the way DNA methylation patterns are maintained throughout cell division, it was originally suggested that demethylation may occur through a passive mechanism whereby selected modified sites may not be efficiently recapitulated during replication. Although this appears to occur in an artificial manner when cells are treated with 5-azacytidine, early studies already demonstrated that demethylation in vivo largely occurs in an active manner, even in the absence of replication and cell division (73, 74).

The most obvious biochemical mechanism for demethylation would be the reverse of DNA methylation, involving direct removal of methyl groups from the 5’ position of cytosine. This type of demethylation has indeed been observed in extracts from cancer cells, where it was proposed that MBD2 may have an enzymatic activity that can combine methyl groups with H2O to generate methanol (99). Even though this type of reaction may be chemically and
energetically feasible (100), other laboratories have not yet been able to reproduce these studies (101).

**Demethylation by Repair**

An alternative mechanism for demethylation was suggested by experiments showing that 5mC may be a weak substrate for glycosylases. On this basis, it was proposed that the apyrimidinic base product of this reaction may then undergo DNA patch repair, which would result in the replacement of this damaged nucleotide with unmethylated cytosine, thereby completing the demethylation process (102). This idea was originally confirmed in studies showing nucleotide substitution both in vivo (103) and in vitro (104) and has now received wide support from a number of different cellular and biological systems showing that glycosylases (105–107) and excision-repair enzymes may indeed be essential for demethylation (108, 109).

One of the major problems with this proposed demethylation pathway is that 5mC is actually only a very weak substrate for the known glycosylases. It now appears, however, that demethylation may take place through a more complex process that involves first modifying 5mC to prepare it for glycosylation. 5-hydroxymethylcytidine (5hmC), for example, has been detected at a number of sites known to be undergoing demethylation in vivo and in vitro (110–112), suggesting that this may represent a necessary intermediate in the demethylation pathway. It has also been shown that the cytosine deaminase, Aid, may be required for demethylation during spermatogenesis (113).

Taken together, these studies suggest that demethylation may occur through a series of biochemical steps, ultimately leading to cytosine substitution by repair (Figure 4). In one possible scenario, 5mC is first hydroxylated, thereby generating 5hmC (114). This modified nucleoside may then be recognized by a deaminase, which converts it to 5-hydroxymethyluridine (5hmU), a nucleotide variant that is normally removed by specific glycosylases to generate apyrimidinic acid sites, which are subsequently repaired. In keeping with this idea, many of the enzymatic components involved in this pathway have been found together in distinct complexes, apparently making up a type of demethylation machine (107, 115). The recent discovery of other ten-eleven-translocation (Tet)-catalyzed chemical changes at the 5′ position raises the possibility that demethylation may even involve additional intermediate reaction steps (116).

Evidence for this type of multistep repair process has been derived from a number of different demethylation examples in vivo, including the global removal of methyl groups seen during gametogenesis (108), the general paternal-genome demethylation that takes

![Figure 4](image)

*Figure 4*

The demethylation pathway. Active demethylation may take place through a series of biochemical steps that modify 5-methylcytidine (5mC) to make it a recognized substrate for removal and replacement by repair with unmethylated cytosine. In the first step, a Tet enzyme brings about hydroxylation of the methyl group to form 5-hydroxymethylcytidine (5hmC). Deamination then occurs through the involvement of activation-induced deaminase (Aid) or Apobec family proteins to generate 5-hydroxymethyluridine (5hmU), which then becomes a substrate for a glycosylase (Mbd4 or Tdg). The resulting apyrimidinic acid residue is then replaced with C by means of path repair base or nucleotide excision repair (BER or NER).
Induced pluripotent stem (iPS) cells: these cells are generated by adding pluripotency factors to somatic cells, causing them to be reprogrammed.

Reprogramming

As noted above, genes required for pluripotency, such as Oct-3/4 or Nanog, undergo targeted methylation postimplantation and, in this way, prevent differentiated cells from undergoing dedifferentiation back to their pluripotent state. There is no question that the methylated state of these genes is what provides stability to the differentiated phenotype during normal development (65). Despite this layer of protection, it is still possible to reprogram somatic cells artificially either in vitro (117) or by nuclear transfer in vivo (118). How does this occur? A wide variety of studies in a number of different organisms have shown that somatic cell nuclei transplanted into primed denuded oocytes undergo reprogramming to totipotency and then serve as the genetic source for generating an entire organism, including extraembryonic tissues (118). It appears that this is made possible by the global demethylation that takes place in the preimplantation embryo, which first enables trophoderm differentiation by turning on the initially silenced Elf5 gene (119), and then activates key early embryonic genes, such as Oct-3/4 and Nanog (120, 121).

Reprogramming of somatic nuclei in vitro can be carried out in two different ways. The most efficient method is to fuse somatic cells with ES cells. When this is done, Oct-3/4, Nanog, and other pluripotency CpG island-like genes undergo rapid demethylation, thus setting in motion the regulatory network that defines the early ES cell phenotype (122).

Previous experiments have shown that mouse ES cells have an activity capable of demethylating a wide variety of CpG islands in a reaction that is targeted by specific cis acting sequences, such as Sp1 (56, 57), and this probably includes the regulatory regions of Oct-3/4 and other pluripotency genes (123). A similar process may occur in induced pluripotent stem (iPS) cells, where the addition of key stem-cell transcription factors apparently serve to turn on a set of endogenous ES master genes, including the factors needed to carry out CpG island-specific demethylation (Figure 4). It is interesting that, although demethylation of CpG islands generates a pluripotency phenotype, it is not sufficient for totipotency. This may be because genes like Elf5 that are required for trophoblast formation (119) in the preimplantation embryo have a non-CpG island promoter, which remains methylated in ES cells.

Methylation During Gametogenesis

Erasure of Methylation Patterns

In addition to the DNA methylation changes that occur during early embryonic development, the process of gametogenesis is also characterized by an elaborate program of modification dynamics. Germ line cells probably emerge from the postimplantation embryo carrying the full bimodal methylation pattern and migrate along the genital ridge to ultimately take part in forming the gamete (124). These primitive germ cells undergo global demethylation at about 10–12 days post coitum (44). One of the important consequences of this event is that it facilitates the removal of all methyl sites associated with imprinted genes, and it is this erasure that allows reestablishment of new parent-specific methyl imprints at late stages of gametogenesis, either in the oocytes or sperm (125).

Following this global demethylation, the entire genome gets remethylated, while CpG islands are protected in a manner reminiscent of the process that occurs during implantation.
in the early embryo (44). It thus appears that the cycle of erasure followed by reestablishment of a new methylation pattern is a basic aspect of epigenetic regulation. Both in man and in mouse, there are a large number of genes that emerge from spermatogenesis in an unmethylated form, even though they are fully methylated in all somatic cells, and many of these indeed have a testes-specific pattern of expression (53, 58). This profile of undermethylation may be generated by one of two possible mechanisms. Some genes clearly get methylated as part of the remethylation process and then undergo tissue-specific demethylation later in spermatogenesis (126). It is also possible, however, that some genes are actually protected from de novo methylation during early gametogenesis and then remain this way during the formation of a haploid genome in sperm. Recent evidence indicates that CpG islands also undergo specific methylation changes during oogenesis (51).

Imprinting

One of the important events that occurs during gametogenesis is the establishment of imprinting. Over 100 imprinted genes have been identified in mammals, and these are all clustered at well-defined loci in the genome, where cis-acting imprinting centers regulate their allele-specific expression. DNA methylation plays a major role in orchestrating this process. At most imprinted loci, the center undergoes de novo methylation during late oogenesis but remains unmodified during spermatogenesis. However, there are also a few loci where the imprinting center actually becomes methylated postmitotically in the testis, while remaining unmethylated in the oocytes. These parental-specific methylation profiles are generated by a combination of paternal or maternal gametic factors that either promote or prevent de novo methylation (127).

Once formed, differential methyl imprints are preserved in an allele-specific form throughout early embryogenesis and then serve as an epigenetic mark that either turns on or turns off the imprinting center’s regulatory activity controlling all of the genes in the entire locus. It is not yet clear why these methylation patterns do not get erased in the preimplantation embryo (49) and do not lose their allele specificity during the wave of de novo methylation at the time of implantation, but it is possible that other maintenance mechanisms, such as replication timing (128) or histone modification (129), may play a supporting role. Following implantation, however, there is no question that allele-specific methylation is carried on through every replication cycle by the maintenance methylase, Dnmt1 (130).

NONPROGRAMMED INFLUENCES ON DNA METHYLATION

It has already been demonstrated that many CpG islands are subject to creeping de novo methylation as a function of aging in a variety of different cell types (131), suggesting that, in addition to built-in developmental events, the basic pattern of modification may also be influenced by nonprogrammed changes during the lifetime of the organism. The major question, however, is whether methylation patterns can also be influenced by environmental cues, either during embryogenesis or even afterward in the newborn or adult organism. A number of pioneering studies on this topic appear to indicate that this may indeed be possible.

Using a mouse model that contains a retrotransposon integrated upstream of the agouti gene, it was shown that offspring from each litter have variable coat color, ranging from dark brown to yellow, even though all the animals are isogenic, and molecular analysis indicated that this probably occurs because the agouti gene promoter is methylated to different degrees in each individual (132). This methylation variability appears to occur stochastically and probably takes place at about the time of implantation. The degree of methylation appears to be under the control of a number of different modifier genes located throughout the genome (133, 134), but the coat color in the offspring
can also be modulated by exposing pregnant mothers to ethanol (135) or to a methyl-rich diet (136). In general, there appear to be many natural sequences in the genome that are also subject to variable stochastic methylation during development, and it is clear that this effect can generate trait variability in isogenic mice (137). Similar stochastic effects may also be responsible for causing tumor variability (81).

These results on a mouse model confirm the idea that DNA methylation patterns can be influenced by environment, either during gestation or perhaps even after birth. Nonetheless, these changes do not seem to be fully inherited through the germ line into the next generation (136). In the case of the agouti trait, for example, all parent mice, regardless of coat color, still generate approximately the same range of variable offspring. This is consistent with the idea that methylation patterns are largely erased between generations first during gametogenesis, and then again in early embryogenesis (44). This picture is very different from the trans-generational epigenetic effects observed in plants, and this is undoubtedly caused by the lack of embryonic erasure in this organism (138).

### SUMMARY POINTS

1. DNA methylation patterns are erased during preimplantation and then re-established throughout development via sequence information in the DNA.
2. Once established, DNA methylation patterns can be maintained autonomously through many cell divisions.
3. DNA methylation inhibits gene expression by affecting chromatin structure.
4. Changes in methylation during postimplantation development are usually secondary to factor-mediated gene activation or repression, but this subsequent methylation pattern provides long-term stability.
5. Active demethylation takes place through a multistep biochemical pathway that involves hydroxylation, deamination, glycosylation, and subsequent repair.
6. Somatic cell reprogramming involves resetting the methylation pattern to that of an early implantation embryo.
7. Abnormal methylation in cell lines and cancer takes place through a programmed process that involves the recruitment of de novo methylases by the polycomb complex.

### FUTURE ISSUES

1. What is the mechanism involved in setting up methylation patterns? How is local sequence information translated in epigenetic information?
2. How does methylation play a role in lineage determination during development?
3. What are the roles of environment and aging on DNA methylation? What are methylation’s effects on long-term physiology and disease susceptibility?

### DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.
ACKNOWLEDGMENTS

This work was supported by research grants from the Israel Academy of Sciences (H.C., Y.B.), the National Institutes of Health (Y.B.), the Israel Cancer Research Foundation (H.C., Y.B.), the European Community 5th Framework Quality of Life Program (Y.B.), and the European Research Council (H.C.).

LITERATURE CITED

52. Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–57
Contents

Preface
Preface and Dedication to Christian R.H. Raetz
   JoAnne Stubbe ................................................................. xi

Prefatories
A Mitochondrial Odyssey
   Walter Neupert .............................................................. 1

The Fires of Life
   Gottfried Schatz ............................................................ 34

Chromatin, Epigenetics, and Transcription Theme
Introduction to Theme “Chromatin, Epigenetics, and Transcription”
   Joan W. Conaway .......................................................... 61

The COMPASS Family of Histone H3K4 Methylases:
   Mechanisms of Regulation in Development
   and Disease Pathogenesis
   Ali Shilatifard ............................................................... 65

Programming of DNA Methylation Patterns
   Howard Cedar and Yehudit Bergman ................................ 97

RNA Polymerase II Elongation Control
   Qiang Zhou, Tiandao Li, and David H. Price ....................... 119

Genome Regulation by Long Noncoding RNAs
   John L. Rinn and Howard Y. Chang .................................. 145

Protein Tagging Theme
The Ubiquitin System, an Immense Realm
   Alexander Varshavsky .................................................. 167

Ubiquitin and Proteasomes in Transcription
   Fuqiang Geng, Sabine Wenzel, and William P. Tansey ............ 177
The Ubiquitin Code
David Komander and Michael Rape .......................................................... 203

Ubiquitin and Membrane Protein Turnover: From Cradle to Grave
Jason A. MacGurn, Pi-Chiang Hsu, and Scott D. Emr ........................................ 231

The N-End Rule Pathway
Takafumi Tasaki, Shashikanth M. Sriram, Kyong Soo Park, and Yong Tae Kwon ... 261

Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions
Koraljka Husnjak and Ivan Dikic ............................................................... 291

Ubiquitin-Like Proteins
Annemartbe G. van der Veen and Hidde L. Ploegh ........................................ 323

Recent Advances in Biochemistry

Toward the Single-Hour High-Quality Genome
Patrik L. Ståhl and Joakim Lundeberg .......................................................... 359

Mass Spectrometry-Based Proteomics and Network Biology
Ariel Bensimon, Albert J.R. Heck, and Ruedi Aebersold .................................. 379

Membrane Fission: The Biogenesis of Transport Carriers
Felix Campelo and Vivek Malhotra ........................................................... 407

Emerging Paradigms for Complex Iron-Sulfur Cofactor Assembly and Insertion
John W. Peters and Joan B. Broderick ......................................................... 429

Structural Perspective of Peptidoglycan Biosynthesis and Assembly
Andrew L. Lovering, Susan S. Safadi, and Natalie C.J. Strynadka ...................... 451

Discovery, Biosynthesis, and Engineering of Lantipeptides
Patrick J. Knerr and Wilfred A. van der Donk .............................................. 479

Regulation of Glucose Transporter Translocation in Health and Diabetes
Jonathan S. Bogan ....................................................................................... 507

Structure and Regulation of Soluble Guanylate Cyclase
Emily R. Derbyshire and Michael A. Marletta ............................................... 533

The MPS1 Family of Protein Kinases
Xuedong Liu and Mark Winey ......................................................................... 561

The Structural Basis for Control of Eukaryotic Protein Kinases
Jane A. Endicott, Martin E.M. Noble, and Louise N. Johnson ......................... 587
Measurements and Implications of the Membrane Dipole Potential
Liguo Wang ................................................................. 615

GTPase Networks in Membrane Traffic
Emi Mizuno-Yamasaki, Felix Rivera-Molina, and Peter Novick .......................... 637

Roles for Actin Assembly in Endocytosis
Olivia L. Mooren, Brian J. Galletta, and John A. Cooper .................................. 661

Lipid Droplets and Cellular Lipid Metabolism
Tobias C. Walther and Robert V. Farese Jr. ......................................................... 687

Adipogenesis: From Stem Cell to Adipocyte
Qi Qun Tang and M. Daniel Lane ................................................................. 715

Pluripotency and Nuclear Reprogramming
Marion Dejosez and Thomas P. Zwaka ......................................................... 737

Endoplasmic Reticulum Stress and Type 2 Diabetes
Sung Hoon Back and Randal J. Kaufman ......................................................... 767

Structure Unifies the Viral Universe
Nicola G.A. Abrescia, Dennis H. Bamford, Jonathan M. Grimes, and David I. Stuart ......................................................... 795

Indexes
Cumulative Index of Contributing Authors, Volumes 77–81 ....................... 823
Cumulative Index of Chapter Titles, Volumes 77–81 ................................. 827

Errata
An online log of corrections to Annual Review of Biochemistry articles may be found at http://biochem.annualreviews.org/errata.shtml