DNA methylation *versus* gene expression

By ADRIAN P. BIRD

*MRC Mammalian Genome Unit, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, U.K.*

**TABLE OF CONTENTS**

Summary  
Introduction  
DNA methylation  
Replication of mCpG  
Methylation of active and inactive genes  
Variable levels of DNA methylation in animals  
Conclusions and speculations on the role of DNA methylation  
References

**SUMMARY**

Vertebrate DNA is methylated at a high proportion of cytosine residues in the sequence CpG, and it has been suggested that the distribution of methylated and non-methylated CpGs in a given cell type influences the pattern of gene expression in those cells. Since a DNA methylation pattern is normally transmitted faithfully to daughter cells via cell division, this idea suggests an origin for stable, clonally inherited patterns of gene expression. This article discusses some of the current evidence for a relationship between DNA methylation and gene expression. Although the evidence is incomplete, it appears already that the relationship is variable: transcription of some genes is repressed by the presence of 5-methylcytosine at certain CpGs, and may be controlled by methylation, while transcription of other genes is indifferent to methylation. In attempting to explain this variability it is helpful to adopt an evolutionary perspective.

**INTRODUCTION**

The control of gene expression is believed to be at the heart of differentiation and development. The belief is based upon the evidence supporting two assumptions: (1) that cell types differ because of qualitative and quantitative differences in the proteins which they contain; (2) that differences in protein composition are primarily due to differences in gene activity at the transcriptional level. For example, the red blood cell synthesizes globin messenger RNA which is translated to give globin protein, whereas in a skin cell the globin gene is not transcribed, but other genes, such as the keratin gene, are active. Differential transcription, then, is held to be a crucial aspect of differentiation, and the question becomes: how does differential gene expression arise?
The example of DNA, whose structure alone explains its replication and protein coding functions, leads to the idea that control of gene expression also has a simple structural basis. DNA methylation is an attractive candidate for this, as it has several features that make it a potential controlling influence upon gene expression. First, it is known that the presence or absence of methylation in bacteria can determine whether a protein (in this case a restriction endonuclease) interacts productively or unproductively with its recognition sequence on DNA. This raises the possibility that eukaryotic methylation is also involved in switching DNA–protein interactions. Secondly, once more by analogy with bacteria, the process of methylation has the properties of a replication system that is capable of perpetuating a pattern of methylation through many cell generations (see below). The idea of a heritable, but reversible, signal on the DNA, whose presence or absence does not affect genetic coding, has obvious attractions.

DNA METHYLATION

The methylated base in animals, 5-methyl cytosine (5mC), constitutes up to 5% of all cytosine (Wyatt, 1951). Early work established two points that have been reinforced by later results: first, that the act of methylation follows synthesis of the DNA polymer at DNA replication (Scarano, Iaccarino, Grippo & Winckelmans, 1965; Burdon & Adams, 1969; Sneider & Potter, 1969; Kappler, 1970); and second, that the methylated sequence is 5'XCGY, where X and Y can be any of the four nucleotides (Doscocil & Sorm, 1962; Grippo, Iaccarino, Parisi & Scarano, 1968). Not all CpGs are methylated, and it is possible to partially map the distributions of methylated and non-methylated CpGs both at specific sequences, and in the genome as a whole. This is achieved by using Type II restriction endonucleases whose recognition sequences include the sequence CpG, but which are blocked from cutting by methylation of the C (Bird & Southern, 1978).

REPLICATION OF mCpG

CpG is one of the four possible dinucleotide sequences that are self-complementary. As a result CpGs occur in pairs on opposite strands, and it is this symmetry which provides the basis for replication of the pattern of methylated and non-methylated CpG (Holliday & Pugh, 1975; Riggs, 1975). Since methylation is a postsynthetic event, there is a transient stage immediately after replication of a sequence when the parental strand carries its pattern of methylated and non-methylated CpGs, but the progeny strand is unmethylated. There is now good evidence that the methylase specifically modifies progeny CpGs whose parental complements are methylated, but not CpGs whose parental complements are non-methylated (Bird, 1978; Pollack, Stein, Razin & Cedar, 1980; Wigler, Levy & Perucho, 1981). The consequence is that patterns of CpG
DNA methylation can be retained faithfully through many cell divisions. If we imagine that the pattern of methylation encodes some information of 'epigenetic' significance, then its heritability is suggestive. It is known, for example, that during development cells can divide many times between 'determination' and the resolution of the determined state at 'differentiation'.

**Methylation of Active and Inactive Genes**

If methylation is involved in controlling gene expression, there should be discernable differences between the methylation pattern of a gene in cells where the gene is expressed compared with cells where it is silent. Many genes that are expressed tissue specifically have now been partially analysed. Although the results are not uniform, the most frequent finding is that the gene in question is less methylated in expressing tissues than in non-expressing tissues, and is most heavily methylated in sperm (Felsenfeld & McGhee, 1982; Bird, 1984). This is an intriguing result which invites the speculation that loss of methylation at a gene is a step in activation.

The loss of methylation at an expressed gene, however, has two alternative explanations: either demethylation is a precondition for gene activation, or it is a consequence of it. Recent experiments have concentrated on this question by asking whether methylated genes can be expressed in in vivo assay systems. Unfortunately most genes must be methylated artificially, as they can only be purified by cloning, and this removes endogenous patterns of methylation. The problem has been surmounted in the case of a human globin gene, and a fully methylated gene was inserted into mouse cells by DNA-mediated gene transfer (Busslinger, Hurst & Flavell, 1983). Analysis of RNA in the transformed cells showed clearly that methylation of CpGs in a region that extends 850 bp upstream of the 5' end prevented transcription of the gamma globin gene. This suggests that loss of methylation at these sites may well be a causative factor in activation of the gene. Although the story for human globin genes is more complete than for other 'tissue specific' genes, an important gap remains. The evidence for reduced methylation at the gamma globin gene in blood cells is confined to HpaII sites (van der Ploeg & Flavell, 1980), but these are not the sites whose methylation abolishes transcription in mouse cells (Busslinger et al. 1983). We need to know, therefore, if methylation of the relevant sites changes during red cell differentiation. It is still possible, for example, that the important CpGs are unmethylated in all cell types.

The need for the latter experiment is emphasized by the recent discovery of a gene whose expression is inhibited by in vitro methylation, but which is probably unmethylated at the relevant sites in most, if not all, tissues. The gene is that for hamster adenosine phosphoribosyl transferase, and it has been studied by Stein, Razin & Cedar (1982). Methylation at HpaII sites alone is sufficient to inactivate this gene upon transformation into mouse cells, but a blot analysis of
genomic DNA from mouse sperm and liver shows that HpaII sites surrounding the 5' end are unmethylated in both tissues (Stein, Sciaky-Gallili, Razin & Cedar, 1983). Similar unmethylated domains have been observed at the 5' end of the mouse dihydrofolate reductase gene, and the chicken 2(1) collagen gene (McKeon, Ohkubo, Paston & de Crombrugghe, 1982). A reasonable conclusion is that changing levels of methylation are not a component of regulation for these genes, although in the case of the collagen gene, expression is tissue specific. Methylation, it seems, is best excluded from the 5' domains of these genes at all times.

Our own studies of DNA methylation have concentrated on the genes for ribosomal RNA, and our results are summarized below. In the amphibian *Xenopus laevis*, there are about 1000 ribosomal RNA genes per cell arranged as a head-to-tail array. Each repeating unit comprises a non-transcribed spacer that contains sequences which influence transcription, and a transcribed region (Fig. 1). In somatic cells, the genes are heavily methylated at CpG (Dawid, Brown & Reeder, 1970), but there is a region of undermethylation in the non-transcribed spacer (Fig. 1) which maps to a cluster of repeated 60 bp sequences (Bird & Southern, 1978). Copies of the 60 bp repeat have been shown to be required for

*X. laevis* rDNA

---

**Fig. 1.** A unit of *Xenopus laevis* ribosomal DNA, showing the spacer and transcribed regions. There is one site per unit for HindIII (H), and two sites per unit for EcoRI (E). The proximal region of the spacer is magnified to show the promoter (and its duplications), and the 60/81 base pair repeats, which are themselves homologous to a region of the promoter (black squares). The 60/81 repeats are hypomethylated in rDNA from somatic cells, but are fully methylated in rDNA from sperm.
DNA methylation

expression of ribosomal RNA genes in *Xenopus* embryos (Busby & Reeder, 1983). Furthermore, this sequence is highly conserved (and undermethylated) in the rDNA of *Xenopus borealis* (La Volpe, Taggart, Macleod & Bird, 1982), although the bulk of the spacer has diverged markedly in sequence between *X. laevis* and *X. borealis* (Forsheit, Davidson & Brown, 1974). Thus the undermethylated CpGs are located at sequences of regulatory significance.

Unlike somatic tissues, the rDNA of sperm is heavily methylated in this and other regions of the spacer (Bird, Taggart & Macleod, 1981), and direct sequencing shows that all 19 CpGs in the ‘promoter region’ are methylated on both strands (Macleod & Bird, 1983). The transition from the fully methylated spacer of sperm to the undermethylated spacer of somatic cells take place during the

![Diagram of DNA methylation in rDNA spacer during early development of *Xenopus laevis*.](image)

Fig. 2. Loss of methylation in the rDNA spacer during early development of *Xenopus laevis*. DNA from each source was digested with HindIII (−) or HindIII plus HpaII (+). The DNA was then fractionated on an agarose gel, blotted, and hybridized to labelled rDNA. HindIII alone gives rDNA fragments of repeat unit length (see Fig. 1), which are of variable size due to spacer length heterogeneity. The methyl-sensitive enzyme HpaII has no effect on sperm rDNA, but gives rise to a cluster of extra fragments (8–10 kb) in blastula DNA. By the heartbeat stage (about 40 h after fertilization), most repeat units are cut by HpaII. The 8–10 kb fragments arise due to cleavage in the undermethylated regions of the spacer as shown in the diagram. (Figure adapted from Macleod & Bird, 1982, with permission.)
Fig. 3. Transcription in oocyte nuclei of highly methylated rDNA from *X. laevis* sperm. Purified sperm rDNA was circularized and injected into oocyte nuclei of *X. borealis*. As a control, cloned (i.e. non-methylated) rDNA that had been modified by deletion (a gift from Dr Tom Moss, Portsmouth) was also injected with the purified sperm rDNA. The deletion gave a transcript that was distinguishable in size (61 nucleotides) from that of normal rDNA (93 nucleotides) by the S1-nuclease assay (see diagram). The lanes show S1-nuclease assays of RNA from oocytes injected with (a) cloned rDNA alone; (b) equal amounts of sperm rDNA circles and cloned rDNA; (c) four parts sperm rDNA to one part cloned rDNA. The results show that sperm rDNA is transcribed as efficiently as cloned rDNA in the oocyte nucleus. The double band given by cloned rDNA is an artifact of the deletion (Moss, 1982). For technical details see Macleod & Bird (1983) from which this figure is adapted, with permission.
first day of embryonic development (Fig. 2; Bird et al. 1981). It is during this period that nucleoli appear in embryonic cells and the rRNA genes are transcriptionally activated (Brown & Littna, 1964). Thus the loss of spacer methylation and the onset of rRNA synthesis are roughly coincident.

If loss of methylation is a prerequisite for transcription of the rRNA genes, then sperm rDNA should not be transcribed upon injection into the nucleus of a *Xenopus* oocyte. Fortunately this can be directly tested, as a significant advantage of the *Xenopus* system is that rDNA can be purified without recourse to cloning, and therefore has its endogenous methylation pattern intact. When purified sperm rDNA was tested for transcription in oocytes we were surprised to find that sperm rDNA and non-methylated rDNA were transcribed with equal efficiency (Fig. 3; Macleod & Bird, 1983). DNA replication does not occur in the oocyte, and no loss of methylation was detected. The conclusion is that a high degree of methylation does not interfere with rDNA transcription in oocytes. Therefore loss of methylation in the spacer is more likely to be a consequence of transcription, rather than a cause.

**VARIABLE LEVELS OF DNA METHYLATION IN ANIMALS**

Before attempting to draw general conclusions about these observations on methylation of specific genes, it is important to take a broader view of DNA methylation in animals. Most important is an appreciation of the wide variation in levels of methylation. This is dramatically demonstrated in Fig. 4. DNA from *Xenopus laevis* and other vertebrates is poorly cleaved by the methyl-sensitive endonuclease HpaII, whereas *Drosophila* DNA is cleaved identically by HpaII and MspI (an isoschizomer of HpaII that is insensitive to CpG methylation). This and other results illustrate the general conclusion that there is little or no 5mC in insects (Urieli-Shoval, Gruenbaum, Sedat & Razin, 1982), while vertebrate DNA is heavily methylated. Between these extremes, most invertebrate genomes contain a DNA fraction that is methylated (usually 5–35 %) and a larger fraction that lacks detectable 5mC (Bird & Taggart, 1980). The wide variation in levels of 5mC among animals argues against the idea that gene expression is always controlled by DNA methylation. This argument is reinforced by the absence, so far, of any clear example of a methylated gene among the invertebrates. From this point of view, it is vertebrates, with their methylated genes, that are unusual.

**CONCLUSIONS AND SPECULATIONS ON THE ROLE OF DNA METHYLATION**

A reasonable conclusion from the data described above is that DNA methylation is unlikely to provide a coherent, unified mechanism for controlling gene expression. Many organisms lack methylation of genes, and methylation in those organisms that have it appears not always to affect transcription. The
Fig. 4. The wide range in levels of CpG methylation in animals as detected by HpaII (H) and MspI (M). Both enzymes cleave the unmodified sequence CCGG, but HpaII will not cleave CmCGG whereas MspI does cleave this sequence. *Drosophila melanogaster* (fruit fly) DNA gives indistinguishable fragment patterns indicating that most CCGG sites are unmethylated. *Echinus esculentus* (sea urchin) DNA is partially resistant to HpaII due to methylated and unmethylated fractions in the genome. *Xenopus laevis* (clawed frog) DNA appears largely undigested by HpaII indicating that most sites are methylated at CpG. DNA from each animal was digested and fractionated on a 1-2% agarose gel. DNA was stained with ethidium bromide and photographed under short-wave ultra-violet light. Numbers refer to fragment sizes in kilobase pairs. Modified from Bird & Taggart (1980) with permission.

comparative data invite us to go further, as control of gene expression is evident in all multicellular animals, but methylated genes have only been found in vertebrates. It is likely, therefore, that gene control mechanisms evolved in the absence of DNA methylation.

Logic of this kind appears to relegate DNA methylation to a minor regulatory role, but this does not follow. There is no need to believe that mechanisms of gene control have remained static during evolution. On the contrary, it seems likely that as multicellular forms gained in complexity, new avenues of control have been explored. Although at present all explanations of the role of DNA
methylation are speculative, I offer one possibility which can account for the observations, and which accommodates the possibility that methylation is an important regulatory factor for some genes. The starting assumption is that methylation of genes in the invertebrates somehow interferes with transcription, and is for that reason excluded from transcribed portions of the genome. During evolution of the vertebrates, methylation has spread throughout the genome and has led for the first time to methylation of genes. This has brought forth mechanisms that allow genes to tolerate and, in some cases, exploit methylation for the control of gene expression.

We can imagine three ways in which genes might cope with methylation, and it is surprising to find that likely examples of each response have already been observed. First, regulatory sequences might be maintained permanently in an unmethylated condition by rendering them refractory to methylation. How this may be achieved is not clear (e.g. DNA sequence, chromatin structure, timing of replication during S phase), but already there are three genes whose 5' domains are unmethylated in expressing and non-expressing tissues alike (see above). It seems likely that many more genes of this kind will emerge in the future. A second response to methylation is for the gene to cope by tolerating methylated bases at its regulatory sequences. The ribosomal RNA genes of *Xenopus* may be an example. Last, but not least, is the opportunistic response of genes that are kept suppressed by methylation in inappropriate tissues, but which can be ‘released’ by demethylation of relevant sites in tissues where the gene is to be expressed. It is the latter role for methylation that has hitherto attracted most interest, and it is to this category that the globin genes may belong.

Although a demethylation step is usually envisaged as a prerequisite for transcription of certain genes, there is another way in which methylation might be involved. Activation of transcription may take place with a gene still in the methylated state, with the process of activation leading later to loss of methylation. The effect of this delayed demethylation might then be to relax control of the gene. In this case demethylation is not seen as a mechanism to facilitate initial activation of the gene, but to simplify continued expression once activation has occurred, perhaps by bringing the gene under more ‘everyday’ control mechanisms.

While the above speculations fit the data, the evidence for them is circumstantial because some of the basic questions about methylation remain unanswered. What is the role of methylation in invertebrates? How often does *de novo* methylation occur? Why are some sequences methylated while others in the same genome remain unmethylated? Answers to these questions are vital if we are to place methylation in its evolutionary or its functional context. Now that the simple unified theories have had to be discarded, we realise that the study of DNA methylation is still in its infancy.
REFERENCES


