PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

DNA Methylation: Evolution of a Bacterial Immune Function into a Regulator of Gene Expression and Genome Structure in Higher Eukaryotes

T. H. Bestor

Phil. Trans. R. Soc. Lond. B 1990 **326**, 179-187
doi: 10.1098/rstb.1990.0002

Phil. Trans. R. Soc. Lond. B **326**, 179–187 (1990)
Printed in Great Britain

DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes

By T. H. Bestor

Department of Anatomy and Cellular Biology, Laboratory of Human Reproduction and Reproductive Biology, Harvard Medical School, 45 Shattuck Street, Boston, Massachusetts 02115, *U.S.A.*

The amino acid sequence of mammalian DNA methyltransferase has been deduced from the nucleotide sequence of a cloned cDNA. It appears that the mammalian enzyme arose during evolution via fusion of a prokaryotic restriction methyltransferase gene and a second gene of unknown function. Mammalian DNA methyltransferase currently comprises an N-terminal domain of about 1000 amino acids that may have a regulatory role and a C-terminal 570 amino acid domain that retains similarities to bacterial restriction methyltransferases. The sequence similarities among mammalian and bacterial DNA cytosine methyltransferases suggest a common evolutionary origin.

DNA methylation is uncommon among those eukaryotes having genomes of less than 10^8 base pairs, but nearly universal among large-genome eukaryotes. This and other considerations make it likely that sequence inactivation by DNA methylation has evolved to compensate for the expansion of the genome that has accompanied the development of higher plants and animals. As methylated sequences are usually propagated in the repressed, nuclease-insensitive state, it is likely that DNA methylation compartmentalizes the genome to facilitate gene regulation by reducing the total amount of DNA sequence that must be scanned by DNA-binding regulatory proteins.

DNA methylation is involved in immune recognition in bacteria but appears to regulate the structure and expression of the genome in complex higher eukaryotes. I suggest that the DNA-methylating system of mammals was derived from that of bacteria by way of a hypothetical intermediate that carried out selective *de novo* methylation of exogenous DNA and propagated the methylated DNA in the repressed state within its own genome. During the evolution of complex plants and animals the inactivating effects of DNA methylation spread to extraneous cellular sequences, such as highly repetitive DNA and transposable elements, and later to tissue-specific genes. Modern large-genome eukaryotes have adapted a primitive prokaryotic immune system to enable them to manage a genome that has expanded more than 1000-fold as a result of accumulation of extraneous sequences and tissue-specific genes.

1. Introduction

Most bacterial species protect themselves from bacteriophage infection by selectively degrading exogenous DNA. This process is mediated largely by type II restriction–modification (r–m) systems, in which a DNA methyltransferase methylates cellular DNA at specific residues within recognition sequences 4–8 base pairs (b.p.) in length and a sequence-specific endonuclease degrades only unmethylated DNA at those recognition sequences (Wilson 1988). There is a wide variety of sequence specificities among the type II r–m systems, and single bacterial species may contain several methyltransferase/endonuclease pairs. Most type II restriction

methyltransferases produce *N*-6-methyladenine or 5-methylcytosine (m^5C). *Escherichia coli* also has an immune mechanism termed *Mcr* that is complementary to r–m systems in that it causes the selective degradation of DNA that bears m^5C residues outside of the sequences normally methylated in *E. coli* (Raleigh *et al.* 1988). Methylated bases have other functions in bacteria, but their most important role appears to be in conferring immunity to bacteriophage infection.

Vertebrate DNA also contains m^5C but there is no evidence that DNA methylation is involved in protection from virus infection. Sequence-specific *de novo* methylation and demethylation of CpG sites establish methylation patterns during gametogenesis and early embryonic development (see Monk, this symposium); these methylation patterns are transmitted to daughter cells by clonal inheritance and undergo programmed changes during cellular differentiation (Cedar 1988). The pattern of methylated sites in the vertebrate genome is heritable, tissue-specific and reversible, whereas in bacteria, methylation patterns are essentially static and are a simple function of the sequence specificities of the restriction methyltransferases present in the particular strain. In vertebrates, DNA methylation directs selective repression of DNA sequences; several lines of evidence confirm that methylated genes are not expressed and that developmentally regulated genes undergo demethylation at the time of transcriptional activation (Cedar 1988). Methylation patterns are now appreciated as an important component in the complex web of regulatory factors that control the unfolding of the development programme.

Methylation patterns in bacteria and vertebrates are very different in form and function and it was not clear that the DNA-methylating systems in these groups of organisms shared a common origin. Recently, the amino acid sequence of mammalian DNA (cytosine-5)-methyltransferase (DNA MTase) was determined and compared to those of several bacterial type II DNA cytosine methyltransferases (Bestor *et al.* 1988; Posfai *et al.* 1989). Clear evidence of homology was found; furthermore, mammalian DNA MTase appears to have been derived by gene fusion of a primitive DNA cytosine methyltransferase and a gene of unknown function, which may have evolved into a regulatory domain. The DNA-methylating systems in bacteria and higher eukaryotes appear to have had a common origin, and when genome structure and the nature of DNA methylation in some fungi and higher plants is also considered, a mechanism becomes apparent by which a cellular system that confers immunity to bacteriophage infection could have evolved into a regulator of multicellular development. Methylation-based regulatory systems have probably evolved to enable complex higher eukaryotes to regulate the function of their enormously expanded genomes. The evolution of the function of DNA methylation, and the alterations in the DNA-methylating system that have accompanied this process, will be discussed here.

2. Structure and evolution of DNA methyltransferase deduced from sequence comparisons

Accounts of the purification and characterization of mammalian DNA MTase (Bestor & Ingram 1983, 1985) and the cloning and sequencing of the DNA MTase cDNA (Bestor *et al.* 1988) have been published. DNA MTase is purified as a group of closely related polypeptides of relative molecular mass 150000–190000. These are derived by post-translational processing of a common precursor. Cell types with different methylation patterns contain very similar or identical species of DNA MTase; other data confirm that the intrinsic sequence specificity of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

DNA MTase is not sufficient to allow the enzyme to establish cell-type specific methylation patterns (Bestor *et al.* 1988). The factors that control the action of DNA MTase *in vivo* during the formation and maintenance of cell-type specific methylation patterns are not understood.

The amino acid sequence inferred from the codon order of the DNA MTase cDNA has several features of interest. First, searches of the protein sequence databases revealed similarities between the 570 C-terminal amino acids of DNA MTase and bacterial type II restriction DNA cytosine methyltransferases (Bestor *et al.* 1988). These similarities were statistically significant; results of Monte Carlo tests of the match between the C-terminal region of DNA MTase and M. *Dde* I exceeded 11 standard deviation units, where values in excess of seven denote nearly certain homology (Doolittle 1986). Comparison with the sequences of 13 other bacterial DNA cytosine methyltransferases (kindly provided by J. Posfai & R. Roberts, Cold Spring Harbor Laboratory) showed less, but still significant similarity. Matrix comparisons showed regions of homology located on the diagonal, indicating significant conservation of secondary and tertiary structure among the bacterial and mammalian proteins (Bestor *et al.* 1988). These results strongly suggest that the C-terminal of 570 amino acids of mammalian DNA MTase represents a methyltransferase domain. Detailed comparisons of sequence feature in mammalian and bacterial DNA cytosine methyltransferases have been presented elsewhere (Posfai *et al.* 1989; Lauster *et al.* 1989). The C-terminal methyltransferase domain is joined to the much larger ($\approx$1000 amino acid) N-terminal domain by an unusual sequence feature composed of a run of 13 alternating Gly and Lys residues, which are encoded in the cDNA by 39 consecutive purine residues. This hydrophilic sequence feature is likely to form a flexible hinge or link between the N- and C-terminal domains. The location of some of the relevant sequence features in DNA MTase, and some of the sequence motifs that are especially well conserved between bacterial and mammalian DNA cytosine methyltransferases, are shown in figure 1.

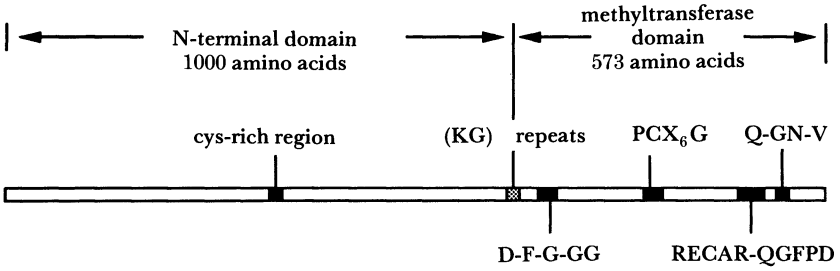

FIGURE 1. The positions of significant features of mammalian DNA MTase are shown on a linear representation of the amino acid sequence as deduced from the codon order of a cloned cDNA. The protein contains an N-terminal domain of 1000 amino acids that contains a possible metal-binding Cys-rich region; the N-terminal domain is joined to a 573 amino-acid C-terminal domain by a run of alternating lysyl and glycyl residues, shown in the figure as (KG) repeats. The C-terminal domain bears strong similarities to bacterial type II DNA cytosine methyltransferases; some of the motifs that are especially well conserved are given in the one letter amino acid code. As discussed in the text, mammalian DNA MTase appears to have arisen via gene fusion. The C-terminal region appears to represent a catalytic methyltransferase domain descended from a bacterial restriction methyltransferase, but the function of the large N-terminal domain is not known.

The N-terminal domain contains an interesting cysteine-rich region that is similar to metal-binding sites found in several regulatory proteins. The DNA MTase cysteine-rich region contains eight cysteine residues out of 37 total residues, and can be represented as having metal-binding configurations known or suspected to exist in zinc-finger proteins (Evans &

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Hollenberg 1988), in zinc-containing alcohol dehydrogenases (Eklund *et al.* 1976), or in metal-ion bridges of the type first found in the HIV TAT protein (Frankel *et al.* 1988). There is as yet no evidence that DNA MTase is a zinc metalloprotein, and the enzyme is usually purified in the presence of EDTA and dithiothreitol, two reagents that can strip zinc from protein coordination complexes. A diagram showing the positions of putative metal-binding residues in DNA MTase and alcohol dehydrogenase, HIV TAT protein and zinc-finger proteins of the Cys_2Cys_2 type are shown in figure 2.

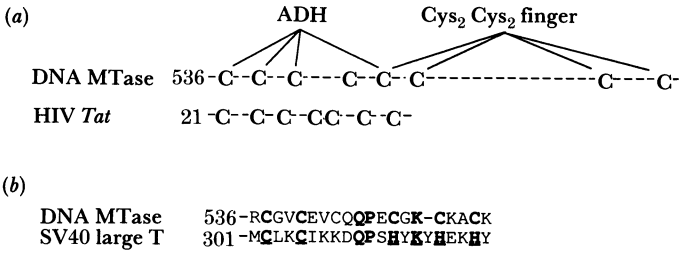

FIGURE 2. Possible coordination structures at the Cys-rich region within the N-terminal domain of mammalian DNA MTase. (*a*) Positions of cysteine residues between amino acids 546 and 586 in the amino-acid sequence of DNA MTase. As discussed in the text, the cysteine-rich region of DNA MTase can be represented as forming zinc coordination complexes of the type found in zinc-containing alcohol dehydrogenases (ADH) (Eklund *et al.* 1976, in Cys_2Cys_2 zinc fingers of the type found in steroid binding regulatory factors (Evans & Hollenberg 1988), and in metal ion bridges of the type found in HIV TAT gene product (Frankel *et al.* 1988). (*b*) Important amino acid residues identified by site-directed mutagenesis of the SV40 zinc-finger region (Loeber *et al.* 1989) has revealed a conserved pattern of essential residues in the cysteine-rich region of DNA MTase. Bold print denotes positions that are either identical in T antigen and DNA MTase or are occupied by residues capable of forming metal coordination complexes. Underlining indicates residues that are essential for T antigen function. Notice that each substitution which eliminated or greatly reduced the biological activity of SV40 large T is occupied in the DNA MTase cysteine-rich region either by the same amino acid or by cysteine instead of histidine. Both of these amino acids are known to participate in metal coordination complexes.

The Cys_2Cys_2 zinc-finger-like sequence in DNA MTase is related to a sequence in SV40 large T antigen that has recently been subjected to site-directed mutagenesis. Several residues have been found to be essential for T-antigen function (Loeber *et al.* 1989). As shown in figure 2*b*, each of the positions required for normal biological function of T antigen is occupied in the DNA MTase cysteine-rich region either by the same amino acid or by cysteine instead of histidine, both of which are known to participate in metal coordination complexes. Although the similarity between the SV40 and DNA MTase cysteine-rich sequences is not in itself sufficient to propose structural similarities between these regions, according to the most stringent criteria (Frankel & Pabo 1988), the strong conservation of important amino acid residues does suggest a common structure.

The domain organization of mammalian DNA MTase suggests that the enzyme originated by fusion of an ancestral restriction methyltransferase gene and a second gene that evolved into the N-terminal domain. Gene fusion events may be common in the evolution of vertebrate DNA-metabolizing enzymes, as mammalian DNA polymerase α (Busbee *et al.* 1987), DNA topoisomerase I (Trask & Muller 1983) and DNA ligase (Tomas Lindahl, personal communication) are all large enzymes that can be cleaved to yield enzymatically active smaller fragments, as can DNA MTase. Eukaryotic topoisomerase II has also been derived by fusion of two prokaryotic precursor genes, and the C-terminus of this enzyme has acquired sequences not present in the bacterial precursors; these new sequences contain alternating

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

blocks of acidic and basic residues (Wyckoff *et al.* 1989), a pattern also found at the N-terminus of DNA MTase. The function of the dispensable sequences acquired by eukaryotic DNA-metabolizing enzymes is not known at present, but within the environment of the nucleus these sequences may be involved in regulating the action of the catalytic portion of the molecule (as proposed for DNA MTase), in mediating interactions with other cellular factors, or they may serve to allow the catalytic portion of the molecule access to DNA in chromatin. Inspection of the sequences of the dispensable portions of other DNA metabolizing enzymes, when these become available, may reveal similarities that suggest common functions.

3. DNA methylation as a compensatory mechanism in the expanding genome

Genome size among eukaryotes varies over a range of a thousandfold or so, with fungi and protists generally at the lower end of the scale and vertebrates and higher plants at the upper end. Large genome size is not simply a function of number of genes (only a low percentage of the sequences in the vertebrate genome actually code for protein), but rather results from the accumulation of extraneous sequences over evolutionary time. It has been pointed out (Doolittle & Sapienza 1980; Orgel & Crick 1980) that the tendency of the genome to expand during evolution is natural, because the stability of the genome perpetuates chance additions. Genome expansion by gene duplication has been suggested to be an essential event in the evolution of complex organisms (Ohno 1971). (For a discussion of the evolutionary pressures on genome structure see Gould (1983).) However, an over-large genome may be deleterious to the organism because of time limitations imposed by the increased amount of sequence that must be scanned by sequence-specific regulatory proteins; genome size could easily become a limiting factor in gene regulation. Also, the time and energetic costs of duplicating the genome may tend to limit its size. Many simpler organisms have evolved mechanisms that actively oppose the expansion of the genome; these sometimes operate by preventing sequence duplication. *Saccharomyces* shows a very low frequency of illegitimate recombination and has an extremely efficient homologous recombination system that replaces endogenous genes with exogeneous copies (Jackson & Fink 1984), whereas *Neurospora* has a bizarre mechanism that selectively inactivates and alters the sequence of duplicated chromosome segments (Selker *et al.* 1987). In these and other small-genome organisms, the amount of repetitive and non-coding DNA is kept low by natural selection or by active mechanisms. Higher eukaryotes appear to lack such mechanisms; foreign DNA is not selectively eliminated, and the large genomes of higher eukaryotes are composed mostly of pseudogenes, long introns, middle repetitive DNA, duplicated genes, satellite DNA and proviral DNA.

How does the large-genome cell manage to correctly control its own genes, which are embedded in a much larger collection of extraneous sequences? I suggest that DNA methylation of the vertebrate type has evolved to compensate for the potentially deleterious effects of genome expansion in complex eukaryotes by compartmentalizing the genome into an unmethylated fraction that is available for interactions with diffusible regulatory factors and a larger methylated fraction that is propagated in the repressed state. This idea is consistent with what is known of the nature of methylation patterns in different groups of organisms.

Organisms such as *Drosophila*, which go through complex developmental processes without the aid of detectable DNA modification, may be able to do so because their genomes are

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

$< 5\%$ of the size of the typical mammalian genome and $< 1\%$ of the size of some plant and amphibian genomes. In addition, most invertebrates, such as *Drosophila*, are small animals with short life cycles where tissue determination takes place at an early stage in development. Self-renewing stem cell populations that can retain the ability to undergo determination for years or decades, such as those in the intestinal crypts and bone marrow of vertebrates, are unknown in insects. In organisms such as *Drosophila* the stability that a regulatory system based on heritable DNA modification can provide may not be of primary importance. In small-genome organisms, protein–DNA interactions appear to be sufficient to ensure proper regulation of the developmental programme; in large genome animals, such as vertebrate animals and higher plants, this may not be true. Although some members of all major phylogenetic groups have m^5C in their DNA (suggesting that methylated genomes are ancestral) methylated DNA is uncommon among eukaryotes with small genomes but nearly universal among large-genome eukaryotes. Furthermore, those invertebrates that do contain methylated DNA, such as echinoderms and ascidians, have not been reported to have methylated genes; instead, m^5C is restricted to the non-transcribed simple repeat sequences which comprise satellite DNA in these organisms (Bird & Taggart 1980). The inactive, methylated compartment of the genome comprises primarily non-coding repetitive sequences, and there does not seem to be any exchange of sequences between methylated and unmethylated compartments. In higher plants and vertebrates, active genes are mostly found in a relatively unmethylated compartment of the genome, whereas inactive genes and extraneous sequences are mostly methylated and tend to be propagated in the nuclease-inaccessible repressed state (Cedar 1988). In vertebrates, activation of tissue-specific genes is associated with their transfer to the unmethylated compartment. Bird (1986) has shown that the putative promoter regions of housekeeping genes tend to be constitutively unmethylated and has argued persuasively that these genes, which are not subject to developmental regulation, are in fact not regulated by DNA methylation. However, this argument does not hold for tissue-specific genes in vertebrate animals, many of which undergo demethylation and are transferred to the unmethylated compartment at the time of activation. The presence of constitutive genes in the unmethylated compartment of the genome supports the idea that DNA methylation has evolved to reduce the effective size of the genome in large-genome plants and animals.

4. The changing role of DNA methylation in evolution

As discussed previously, the dominant function of cytosine methylation in bacteria is immunity from phage infection, whereas in vertebrates methylation acts primarily as a long-term bulk repressor of transcription. In both cases, the effect is to limit exposure of DNA sequences to the transcriptional machinery. Examination of the roles of methylation patterns in other taxa suggests a route by which the bacterial DNA-methylating system may have evolved into the vertebrate system.

One route of evolutionary descent, out of many that are conceivable, is shown in figure 3. Neither the figure or the following discussion are meant to imply that animals are descendants of higher plants nor that fungi are the ancestors of either; the organisms mentioned are those that may have retained vestiges of the types of methylation pattern that were present in intermediates during the evolution of bacterial restriction–modification systems into the large-genome eukaryotic DNA methylation system.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

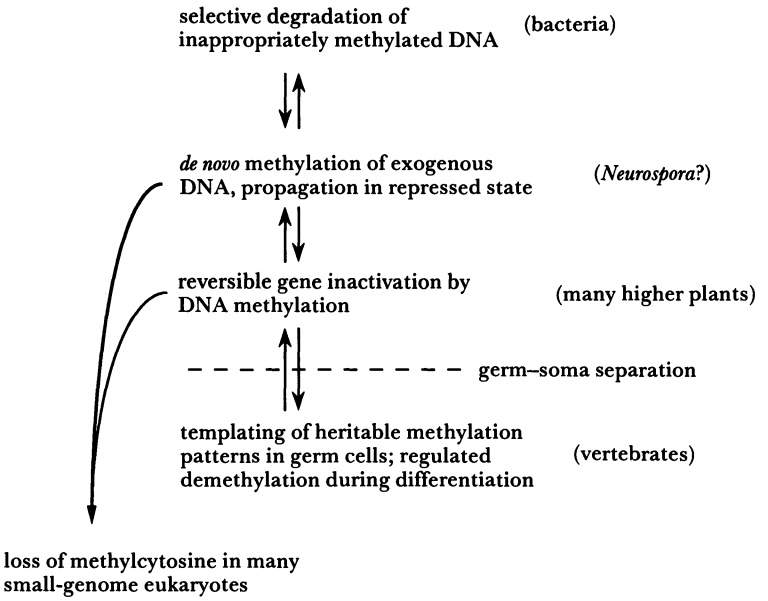

FIGURE 3. Highly schematic depiction of one possible route of evolution of DNA methylation from an immune mechanism in prokaryotes to a regulator of gene expression and genome structure in large-genome plants and vertebrate animals. The figure is not meant to imply that vertebrate animals are descendants of plants or that fungi were the direct ancestors of either.

It is postulated that early microbial ancestors of higher eukaryotes possessed phage immunity systems of the restriction–modification type. In the precursors of the eukaryotic lineage this system became modified to perform *de novo* methylation of incoming DNA, so that the methylated sequences were propagated in the genome in the repressed state. This is not inconsistent with what is known of methylation patterns in extant bacteria; some modern mycoplasma (Razin & Razin 1980) and myxobacteria (Yee & Inouye 1982) are methylated at all or most CpG dinucleotides. It has been reported that fruiting body formation in *Myxococcus* is associated with sequence-specific demethylation of sites within the genome of differentiating cells (Yee & Inouye 1982). This observation suggests that gene control by cytosine methylation may have preceded the appearance of eukaryotes. A prokaryote that represses differentiation-specific genes by *de novo* DNA methylation may have attained immunity to phage infection by a similar mechanism. Only two activities would be required; a DNA MTase that selectively methylates exogenous DNA and a general repressor of transcription that is specific for methylated DNA. Type I restriction–modification systems operate in a similar manner, because they make a decision as to whether incoming DNA will be subjected to restriction or modification. In addition, some type I methyltransferases are strongly stimulated by hemimethylated substrates, while known type II methyltransferases are not (Yuan & Hamilton 1984). The *E. coli Mcr* system is also similar to the hypothetical *de novo* methylation/repression system in that it selectively inactivates methylated DNA (Raleigh *et al.* 1988). In short, the nature of methylation systems in extant bacteria is not too different from that postulated for the ancestor of modern vertebrate animals. This *de novo* methylation/inactivation system would cause the retention of phage genes; this may be of adaptive significance to the host, as occasional activation of limited regions of the phage genome (by accidental failure of maintenance methylation) may make potentially useful new

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

gene functions available to the host cell without the riks of cell mortality that would result from full activation of the phage genome.

In the ancestors of modern higher eukaryotes, DNA inactivation by methylation came to encompass cellular sequences as well. As described previously, the driving force in this development was the pressure to reduce the effective size of the genome. At early stages only highly repetitive sequences were affected, as is the case in modern echinoderms and ascidians. As the genome continued to accumulate tissue-specific genes and extraneous sequences during the evolution of ever more complex organisms, DNA methylation came to be involved in the repression of tissue-specific genes; in modern eukaryotes with large genomes DNA methylation appears to be the basic mechanism that ensures the inactivity of such sequences.

The limited amount of available evidence suggests that vertebrates and higher plants are unique because their methylation patterns are templated or imprinted onto the genome during gametogenesis and in the earliest stages of embryonic development (Monk 1987). These methylation patterns are subject to clonal inheritance during cell growth and undergo programmed alterations during cellular differentiation. Although the DNA MTase of higher plants has been reported to act on unmethylated and hemimethylated substrates with equal efficiency (Theiss *et al.* 1987), all or most of the m^5C in higher plant DNA is at the self-complementary sequences CpG and CpNpG and the potential for clonal transmission of methylation patterns through hemimethylated intermediates is obvious. Whereas both vertebrates and higher plants appear to have adapted DNA methylation as a means of reducing the effective size of the genome to aid target acquisition by regulatory proteins, it is not yet known if plants use DNA methylation to control the expression of tissue- or stage-specific genes; inactivating effects of DNA methylation have to date only been demonstrated for transposable elements and transfected DNA (Matzke *et al.* 1989). It is striking that the two groups of organisms with very large genomes (vertebrates and higher plants) should have evolved such similar DNA-methylating systems despite the ancient divergence of their ancestors and their very different natural histories.

The arguments presented here imply that transcription regulation by DNA methylation arose to compensate for the deleterious effects of the expanding genome and is a relatively recent development; it is to be expected that this regulatory system, which appears to be derived from a bacterial immune function, should differ in fundamental ways from regulatory systems based on sequence-specific protein–DNA interactions that have a far longer evolutionary history. Paradigms derived from the study of bacterial gene regulation may not always be appropriate to higher eukaryotes, where gene control may involve not only repressor and activator proteins, but also interaction of such diffusible factors with heritable methylation patterns in DNA.

I thank Eric Selker and Tomas Lindahl for helpful discussions and Adrian Bird for discussions. Part of the work done here was carried out while I was a postdoctoral fellow of the Leukemia Society of America in the laboratory of Vernon M. Ingram at the Massachusetts Institute of Technology, and was supported by the National Institutes of Health.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

References

Bestor, T. H. & Ingram, V. M. 1983 Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. natn. Acad. Sci. U.S.A.*, **80**, 5559–5563.

Bestor, T. & Ingram, V. 1985 Growth-dependent expression of multiple species of DNA methyltransferase in murine erythroleukemia cells. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2674–2678.

Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. 1988 Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cell. The carboxyl terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. molec. Biol.* **202**, 971–983.

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

Bird, A. P. & Taggart, M. H. 1980 Variable patterns of total DNA and rDNA methylation in animals. *Nucl. Acids Res.* **8**, 1485–1496.

Busbee, D. L., Sylvia, V. L., Cernosek, S. J. Z. & Norman, J. O. 1987 Lability of DNA polymerase α correlated with decreased DNA synthesis and increased age in human cells. *J. natn. Canc. Inst.* **79**, 1231–1239.

Cedar, H. 1988 DNA methylation and gene activity. *Cell* **53**, 3–5.

Doolittle, R. F. 1986 *Of URFS and ORFS: a primer on how to analyze derived amino acid sequences*. New York: University Science Books.

Doolittle, W. F. & Sapienza, C. 1980 Selfish genes, the phenotype paradigm, and genome evolution. *Nature, Lond.* **284**, 601–603.

Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiew, T., Soderburg, B., Tapia, O. & Branden, C. 1976 Structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. molec. Biol.* **102**, 27–42.

Evans, R. M. & Hollenberg, S. M. 1988 Zinc fingers: guilt by association. *Cell* **52**, 1–3.

Frankel, A. D., Bredt, D. S. & Pabo, C. O. 1988 TAT protein from human immunodeficiency virus forms a metal-linked dimer. *Science, Wash.* **240**, 70–73.

Frankel, A. D. & Pabo, C. O. 1988 Fingering too many proteins. *Cell* **53**, 675.

Gould, S. J. 1983 *Hen's teeth and horse's toes*. New York: W. W. Norton.

Gruenbaum, Y., Stein, R., Cedar, H. & Razin, A. 1982 Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature, Lond.* **295**, 620–621.

Jackson, J. A. & Fink, G. R. 1984 Meiotic recombination between duplicated genetic elements in *Saccharomyces cerevisiae*. *Genetics* **109**, 303–332.

Lauster, R., Trautner, T. A. & Noyer-Weidner, M. 1989 Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. molec Biol.* **206**, 305–312.

Loeber, G., Parsons, R. & Tegtmeyer, P. 1989 The zinc-finger region of simian virus large T antigen. *J. Virol.* **63**, 94–100.

Matzke, M. A., Primig, M., Trnovsky, J. & Matzke, A. J. M. 1989 Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**, 643–649.

Monk, M. 1987 Memories of mother and father. *Nature, Lond.* **328**, 203–204.

Ohno, S. 1971 *Evolution by gene duplication*. New York: Springer-Verlag.

Orgel, L. E. & Crick, F. H. C. 1980 Selfish DNA: the ultimate parasite. *Nature, Lond.* **284**, 604–607.

Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. 1989 Predictive motifs derived from cytosine methyltransferases. *Nucl. Acids Res.* **17**, 2421–2430.

Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., Reith, A. D., Rigby, P. W. I., Elhai, J. & Hanahan, D. 1988 *Mcr*A and *Mcr*B restriction phenotypes of some *E. coli* strains and implications for gene cloning. *Nucl. Acids Res.* **16**, 1563–1575.

Razin, A. & Razin, S. 1980 Methylated bases in mycoplasma DNA. *Nucl. Acids Res.* **8**, 1383–1390.

Selker, E. U., Canbareri, E. B., Jensen, B. C. & Haack, K. R. 1987 Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell* **51**, 741–752.

Theiss, G., Schleicher, R., Schimpff-Weiland, G. & Follman, H. 1987 DNA methylation in wheat. Purification and properties of DNA methyltransferase. *Eur. J. Biochem.* **167**, 89–96.

Trask, D. K. & Muller, M. T. 1983 Biochemical characterization of topoisomerase I purified from avian erythrocytes. *Nucl. Acids Res.* **11**, 2779–2800.

Wilson, G. G. 1988 Type II restriction–modification systems. *Trends Genet.* **4**, 314–318.

Wyckoff, E., Natalie, D., Nolan, J. M., Lee, M. & Hsieh, T. 1989 Structure of the *Drosophila* DNA topoisomerase II gene. Nucleotide sequence and homology among topoisomerases II. *J. molec. Biol.* **205**, 1–13.

Yee, T. & Inouye, M. 1982 Two-dimensional DNA electrophoresis applied to the study of DNA methylation and the analysis of genome size in *Myxococcus xanthis*. *J. molec. Biol.* **154**, 181–196.

Yuan, R. & Hamilton, D. L. 1984 Type I restriction/modification systems. In *DNA methylation: biochemistry and biological significance* (ed. A. Razin, A. Riggs & H. Cedar). New York: Springer-Verlag.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES