REVIEW ARTICLE

DNA methylation

The effect of minor bases on DNA–protein interactions

Roger L. P. ADAMS

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, U.K.

INTRODUCTION

It is paradox that DNA is a very stable molecule whose structure is faithfully maintained from generation to generation, yet with each round of replication, this structure is modified by base methylation in nearly all cells and organisms. The consequences of this modification are obviously of vital concern for all biologists.

As well as the four major bases, the DNA of most organisms contains one or more minor bases. The commonest of these is 5-methylcytosine, but *N*⁸-methyladenine and *N*⁴-methylcytosine are also found, particularly in prokaryotes. In vertebrates 3–6% of DNA cytosine is methylated, but this value decreases going down the evolutionary scale so that in many insects and single-celled eukaryotes there is no detectable 5-methylcytosine. In contrast, plants may have 30% of their DNA cytosine methylated (Adams & Burdon, 1985). The suggestion has been made that genome size is an important factor in determining the extent of cytosine methylation, and certainly *Drosophila* with a haploid value of 0.17 pg of DNA per cell has no detectable methylcytosine. The limits of detection used so far are, however, too high to say that the *Drosophila* genome is completely devoid of methylcytosine, or whether there may be, for example, one methylcytosine per gene.

Minor, modified bases are not introduced as DNA is synthesized but are produced shortly after replication, by modification of specific nucleotides already present in duplex DNA. It is often (but not exclusively) the case that methylated bases occur in pairs along the DNA duplex. This is achieved by their being present as part of a symmetrical sequence; e.g.

The distribution of methylated bases along the DNA is often far from random. For example, in vertebrates where methylcytosine occurs largely in the dinucleotide sequence mCG, the distribution of CG dinucleotides is uneven, being much less than expected in A + T-rich regions and greater in G + C-rich regions of the genome (Adams & Eason, 1984; Adams *et al.* 1987). G + C-rich regions are found in the 5' regions of many genes (Bird, 1986) which therefore have a greater potential for methylation than do other regions.

It is theoretically important to distinguish between the situation in which all possible recognition sequences in a

genome are methylated, when the modification adds no further information to that already contained in the sequence of bases in the DNA, and the alternative situation in which selective methylation of a subset of recognition sequences occurs. In the latter case, which is typical of eukaryotes where only about 70% of CG dinucleotides are methylated, there has been an increase in the information content of the DNA.

The methyl groups are transferred from *S*-adenosyl-methionine in a reaction catalysed by a DNA methyl-transferase or methylase (Adams & Burdon, 1983). On replication of DNA the parental strand retains the modified bases, but the daughter strand is initially unmethylated; the DNA is said to be hemimethylated (Fig. 1). Transfer of methyl groups to the daughter strand occurs shortly after replication (Adams, 1971; Woodcock *et al.*, 1982; Lyons & Schendel, 1984) but the lag period during which the DNA is methylated only on the parental strand provides a time window for certain reactions to occur. The most studied of these reactions is mismatch repair, but chromosome segregation and some gene expression depends on the presence of hemimethylated sites in nascent DNA (Messer & Noyer-Weidner, 1988).

This review will consider the enzymes involved in methyl transfer to DNA and the effect of methylation on DNA–protein interactions, with particular emphasis on eukaryotic systems.

DNA METHYLTRANSFERASES

There are three distinct types of DNA methyl-transferase in bacteria, and the enzymes present in eukaryotes appear different from any of these. Nonetheless, all C-5 cytosine methylases for which sequence data are available share regions containing a common amino acid sequence probably indicative of a common evolutionary origin (Som *et al.*, 1987; Wilke *et al.*, 1988; Bestor *et al.*, 1989; Adams *et al.*, 1988; Wilson, 1988; Posfai *et al.*, 1989). The conserved sequence Pro-Cys-Gln does not occur in the *N*-methylases (i.e. *N*⁴-cytosine and adenine methyltransferases), although other regions are conserved (Lauster, 1988; Posfai *et al.*, 1989), implying a close proximity to the active site (Wilson, 1988).

Prokaryotic type II methylases

The simplest enzymes are the type II DNA methyl-transferases of bacteria. These are the cognate methylases

Abbreviations used: CHO, Chinese hamster ovary; CRE, cyclic AMP-responsive element; CREB, cyclic AMP-responsive element binding protein; *dhfr*, dihydrofolate reductase gene; ERE, oestrogen-responsive element; MEL, mouse erythroleukaemia; MFI, methylation-free island; MLTF, major late transcription factor.

Fig. 1. Maintenance and 'de novo' methylation

The figure represents a stretch of vertebrate DNA with five CG dinucleotide pairs, three of which are methylated on both strands. On replication two hemimethylated duplexes are produced. Maintenance methylation (on the left) will restore the parental pattern of methylation. If, however, a second round of replication intervenes prior to methylation (right) then an unmethylated molecule is produced. A 'de novo' methylase will add methyl groups to CG dinucleotides previously unmethylated on both strands (bottom left). ● represents a methyl group.

of the restriction endonucleases which are familiar to all biochemists. They add methyl groups to A or C residues in symmetrical sequences, usually of four, five or six base pairs, and thereby prevent the action of the corresponding endonuclease. [The type IIS methylases (e.g. *FokI*-M) differ in that their recognition site is not symmetrical but can be methylated on both strands by two separate domains on the enzyme (Sugisaki *et al.*, 1989).] In addition, the *dam* and *dcm* methylases are present in most bacterial species and do not have a cognate restriction endonuclease (though nucleases with similar specificity are present in occasional strains of some species, e.g. *Escherichia coli* carrying the RII plasmid exhibits a nuclease activity that can cleave DNA at the unmethylated sequence CC^A/TGG which is the sequence recognized and methylated by the *dcm* methylase). These are small enzymes which will methylate their recognition sequence independently of the state of methylation of the complementary strand, and no regulatory proteins are known which affect their activity.

The enzymes have a region of about 50 amino acids which confers target specificity and those enzymes which show dual or triple specificity have two or three such regions arranged in tandem (Trautner *et al.*, 1988). These target recognizing domains can be rearranged using genetic engineering techniques to produce enzymes with novel combinations of specificities (Wilke *et al.*, 1988).

Prokaryotic type I methylases

The type I bacterial methylases form parts of complexes which will cleave DNA if the recognition site is not methylated on either strand, but which will methylate the recognition sequence if one strand is already methylated. Thus, they act almost exclusively on hemimethylated DNA and the methylase activity is virtually absent on unmethylated DNA (Yuan, 1981). This is also true for complexes lacking the endonuclease subunit. That methylation of unmethylated DNA can, however, occur at low frequency is attested to by the finding that an occasional, invading, viral DNA molecule escapes the growth restriction and is found in a nuclease-resistant, methylated form.

Bacteriophages T3 and T7 encode a protein, Ocr, which overcomes the restriction normally exerted on phage growth by the type I enzymes. The Ocr protein binds stoichiometrically to the enzymes and inhibits both the nuclease and the methylase activity (Kruger *et al.*, 1985, 1988).

The target recognition sequence of the type I enzymes consists of two trinucleotides separated by six, seven or eight bases whose nature is unimportant. Specificity is conferred by the specificity polypeptide which is approx. 460 amino acids long and which contains conserved regions of 30 amino acids and 70 amino acids in the

middle and C-terminus respectively (Cowan *et al.*, 1989). It is the two variable regions which confer specificity, each determining half of the split target recognition sequence. The length of the central conserved region determines the spacing between the two trinucleotides of the target; e.g. *EcoR124* methylase recognizes GAA-(N)₆RTCG and has two repeats of a four-amino-acid sequence in the central domain, whilst *EcoR124/3* recognizes GAA(N)₇RTCG and has three repeats of the same four-amino-acid sequence (Price *et al.*, 1987, 1989).

Prokaryotic type III methylases

Type III methylases resemble type II methylases in that specificity and methylase are both part of the one subunit but, *in vivo*, this is associated with a separate endonuclease subunit. What is peculiar about the target recognition sequence is that it is not symmetrical (e.g. *EcoPI-M* methylates AGACC) and so action is *always* on an unmethylated sequence produced at DNA replication; compare this with type IIS methylases.

Eukaryotic DNA methylases

Eukaryotic DNA methylases are very large, single-subunit enzymes that preferentially methylate hemimethylated DNA (Adams *et al.*, 1986, 1990; Pfeifer & Drahovsky, 1986; Bestor *et al.*, 1989; Yesufu *et al.*, 1989). They have a much smaller recognition sequence than the bacterial enzymes, methylating cytosine in the

dinucleotide CG (animals) or CNG (plants and possibly also animals), yet, *in vivo*, not all such sites are methylated (Gruenbaum *et al.*, 1981a,b; R. L. P. Adams & H. M. I. Yesufu, unpublished work). There has been some controversy over the extent of the preference for hemimethylated DNA which, at least *in vitro*, is not nearly so marked as that seen with the type I bacterial methylases. The reason for the lack of agreement lies in the problem of obtaining a suitable substrate (Fig. 2 and Table 1).

The most commonly used substrate is made by extending a primer annealed to M13 template DNA using methyl-dCTP in place of dCTP (Fig. 2a). Although this produces a molecule in which virtually all the CG dinucleotides are hemimethylated, there are many methylcytosines in other, non-physiological sequences, and these may well interfere with enzyme activity. Nonetheless, Gruenbaum *et al.* (1982) obtained over 100-fold preference for ϕ X174 hemimethylated DNA over control, unmethylated DNA.

Another popular substrate is obtained by growing cells in the presence of azacytidine which is incorporated into nucleic acids in place of cytosine (Fig. 2b). When DNA methylase interacts with azacytosine containing DNA it forms a dead-end complex in which the enzyme is covalently linked to the azacytosine (Santi *et al.*, 1983). This depletes the cell of active enzyme (Tanaka *et al.*, 1980) and DNA subsequently replicated fails to become methylated, i.e. it is hemimethylated after one S-phase.

Fig. 2. Hemimethylated DNA

The figure shows three ways in which a hemimethylated DNA molecule can be made to study vertebrate DNA methylase: (a) using a primed single-stranded DNA molecule and methyl-dCTP to introduce a methylcytosine opposite *every* guanine; (b) incorporation *in vivo* of azacytosine immobilizes and inactivates the DNA methylase preventing further maintenance methylation; (c) DNA synthesized *in vivo* under conditions of limiting methyl donor have an increased frequency of hemimethylated sites. Key: ●, methylated CG; ○, unmethylated CG; X, azacytosine; M, methylase.

Table 1. Hemimethylated substrates for mouse DNA methylase

Methylation was measured by incorporation of radioactivity into DNA following incubation of mouse DNA methylase (18 units) for 1 h with 5 μ g of native DNA and S-adenosyl[3 H] methionine in a volume of 70 μ l. The DNA substrates were obtained from stationary phase mouse L929 cells (1); exponentially growing mouse L929 cells grown for 24 h in the absence (2) or presence of azadeoxycytidine (zdC) at 10 μ M (3) or 1 μ M (4); S-phase mouse L929 cells obtained by blockade with aminopterin and hypoxanthine for 16 h and release with thymidine into methionine-free medium for 6 h (5); and from *Micrococcus luteus* (6).

	Source of DNA	'Me' incorporated (fmol)	Relative rate
(1)	Stationary cells	25	1
(2)	Control cells	45	1.8
(3)	10 μ M-zdC cells	80	3.2
(4)	1 μ M-zdC cells	180	7.2
(5)	-Met cells	1180	47.6
(6)	<i>M. luteus</i>	3515	141

The disadvantages of using this DNA as a substrate are: (1) only a limited region may be hemimethylated as that DNA which replicated prior to the addition of azacytidine will be bifilarly methylated, and (2) the DNA contains azacytosine which will continue to bind and inactivate DNA methylase *in vitro*. Indeed, there is an optimum concentration of azacytidine to which the cells should be exposed. Initially, as the concentration of the drug is increased the DNA obtained becomes an increasingly better substrate in parallel with the extent of hemimethylation. However, further incorporation of azacytosine yields DNA which, *in vitro*, brings about an inhibition as increasing amounts of enzyme become irreversibly bound and inactivated (Jones & Taylor, 1981; Adams *et al.*, 1982).

A third, hemimethylated substrate which can be used is obtained from cells replicating in the absence of added methionine (Fig. 2c). Turnbull & Adams (1976) obtained a 130-fold preference for DNA isolated from methionine-starved S-phase cells relative to DNA from non-growing cells, and this is confirmed in Table 1. As with the azacytosine-containing DNA, the comparison is with DNA in which about 70% of the CG dinucleotides are fully methylated. The remaining 30% of CG dinucleotides are unmethylated in both the control and the test DNA and these accept 25 fmol of methyl groups in the assay, i.e. in sample 5 from the methionine-starved cells 1155 fmol were incorporated into hemimethylated sites. Making the assumption that the methionine deprivation resulted in 10% of the replicated sites remaining hemimethylated this gives a ratio of maintenance/*de novo* methylation of about 400 ($46.6/0.07$ to $1/0.3 \times 2$ as both strands can accept a methyl group).

In human DNA there are 3×10^9 base pairs, i.e. about 24×10^6 newly synthesized CGs to be methylated each S-phase. Of these, 16.8×10^6 are in hemimethylated sites and 7.2×10^6 are paired with an unmethylated CG, i.e. there are 14.4×10^6 CGs in unmethylated pairs. If the results *in vitro* can be applied *in vivo* this means there would be about 30000 'aberrant' methylations per cell in each generation. If the enzyme were in excess this figure would be expected to be even higher. Obviously, such high levels of 'aberrant' methylation may be deleterious to the cell and, *in vivo*, they may be limited by the presence of proteins binding to the unmethylated regions of DNA and preventing access of DNA methylase.

The assumption that eukaryotic DNA methylase functions primarily to maintain DNA methylation patterns by acting only on hemimethylated DNA produced during replication has been challenged. Thus, Bolden *et al.* (1986) have shown, using synthetic oligonucleotides, that certain CG-containing sequences are not acted upon by partially purified HeLa cell DNA methylase whilst other sites are methylated on one strand only. Furthermore, using an enzyme from mouse erythroleukaemia cells Ward *et al.* (1987) found similar results with a fragment of DNA from the promoter region of the mouse β -globin gene. This implies that the enzyme is acting as a site-specific modification methylase. Furthermore, Saluz *et al.* (1986) have found hemimethylated sites to exist *in vivo* in the chick vitellogenin gene. However, these observations may indicate the presence of protein factors which interact with the DNA (either *in vitro* or *in vivo*) to modulate methylase activity.

The *de novo* methylase activity of vertebrate DNA methylase might account for the inactivation of the major part of the adenoviral genome in transformed cells which arise at low frequency. It may also account for the site-specific *de novo* methylation seen when a gene is switched off in a somatic cell; and an increased amount of DNA methylase might be all that is needed to bring about the genome-wide *de novo* methylation of DNA seen in early mouse embryos (Jahner *et al.*, 1982; Lock *et al.*, 1987; Cedar, 1989; Adams *et al.*, 1989a). Methylation *de novo* may then be allowed to occur simply by removing the constraints on the enzyme and allowing it access to the DNA.

In vitro, the enzyme normally can only gain access to duplex DNA at the ends or at regions rendered single-stranded by breathing (Turnbull & Adams, 1976; Adams *et al.*, 1979; Pfeifer *et al.*, 1985), a reaction strongly inhibited by salt. An alternative *in vitro* substrate that we have used is a gapped duplex molecule. In the presence of 100 mM-NaCl, addition of methyl groups is restricted to the single-stranded region (Adams *et al.*, 1989b).

It is assumed that *in vivo* the enzyme becomes associated with the DNA at, or near to, the replicating fork, but single-stranded regions may also arise as gapped duplex molecules during the repair of damaged DNA (when methylation *de novo* of the undamaged strand may occur), or might be specifically generated by site-directed topoisomerase action as they are at the origin of rep-

Fig. 3. Variation in DNA methylase activity during the cell cycle

MEL (Friend) cells were fractionated into different phases of the cell cycle using centrifugal elutriation and the composition of the fractions analysed using a FACScan II (Becton Dickinson). Nuclei were prepared and analysed for DNA methylase activity (Adams *et al.*, 1989). Thanks are due to Drs. Birnie and Conkie of the Beatson Institute for help with the elutriation, and to Drs. Campbell and Munir Alam for help with the FACScan.

lication. Enzyme activity varies markedly during the cell cycle, being low in G₁, rising as the cells enter S-phase and falling again in G₂/M (Fig. 3).

EFFECT OF METHYLATION ON DNA STRUCTURE

Although it has been known for a long time that the presence of methylcytosine stabilizes the double helix and that methyladenine has the reverse effect, this has been elegantly demonstrated recently using electrophoresis on gels containing a gradient of denaturant (Collins & Myers, 1987). It is quite clear that even a single methylcytosine can have a marked effect on the melting of a domain of DNA. The technique allows the separation of 145 bp fragments in which neither, one, or both strands contain a single methylcytosine.

The effect of this *in vivo* is hard to predict and the effect on helix stability has not previously been proposed as the means whereby methylation might affect function. However, it is clear that AT-rich regions are important as sites of unwinding of DNA for initiation of both replication and transcription. It is precisely when present in such early melting domains that methylation of cytosine may have its most pronounced effect on stability. For example, the methylation of a 'GC box' in the promoter of a housekeeping gene may preclude the entry of an RNA polymerase molecule at the adjacent 'TATA box'. Conversely, the methylation of adenine in GATC sites

clustered around the origin of replication of *E. coli* may play a part in the unwinding of the DNA required for initiation of replication (Bramhill & Kornberg, 1988). However, methylation of an adenine in the *EcoRI* target sequence does not alter the conformation of the DNA (Frederick *et al.*, 1988).

The other major effect of methylation of cytosine is to enhance the probability that a stretch of DNA will assume a Z configuration (Behe & Felsenfeld, 1981). This is certainly the case *in vitro*, but the evidence for the existence of Z DNA *in vivo* is still controversial (Jaworski *et al.*, 1988).

EFFECT OF DNA METHYLATION ON DNA-PROTEIN INTERACTIONS

When interacting with DNA, all sequence-specific proteins form bonds with the individual bases of the target sequence (Saenger, 1983). Some of the bases in the target may be less important than others and sometimes all that is required is a consensus sequence to be present for binding to occur. In other situations the protein is completely specific in its requirements and will bind to only a single target sequence. Such conclusions have arisen from the study of very many proteins where it has been shown that alterations in a base affect binding or function of the protein. It is only to be expected that the substitution of a methylcytosine for a cytosine, or a methyladenine for an adenine, might have similar consequences.

Restriction endonucleases

These provide the clearest example where methylation of DNA prevents its cleavage by interfering with the binding and/or function of the nuclease. The type II enzymes act as dimers with one subunit cleaving each strand of the DNA and, in general, methylation of the target on one strand is sufficient to block cleavage on either strand. There are exceptions to this, however, and some enzymes, e.g. *HpaII*, will nick the unmethylated strand of a hemimethylated ³CCGG duplex (McClelland & Nelson, 1988).

The type I restriction enzymes also respond to methylation in a dramatic manner. The enzyme will bind to its target sequence and, if this is not methylated, DNA is reeled in towards the tightly bound enzyme. Soon an adjacent enzyme molecule bound to its target sequence and similarly reeling in the DNA will be encountered. At this point cleavage of the DNA occurs (Studier & Bandyopadhyay, 1988). If the DNA is hemimethylated a completely different reaction occurs, in that a methyl group is transferred by the methylase subunit and the enzyme then leaves the DNA. These controlled reactions involve complex changes in the nature of the DNA-protein complex which have been analysed by Yuan (1981) and Bickle (1982).

Mismatch repair

When a normal, but incorrect, base is introduced into DNA at replication, it is usually removed by a proof-reading mechanism before the next nucleotide is added. When this mechanism fails, the product is a duplex DNA molecule with a mismatch. A repair system which can correct the mismatch requires to discriminate between two normal bases, one of which is incorrect. A random process would accentuate the error in 50 % of cases, but

mismatch correction systems use the lack of methylation on the newly synthesized (incorrect) strand to determine which strand requires repair.

Such systems are present in both prokaryotes and eukaryotes, but they are better understood at present in the former. The *E. coli* MutS protein recognizes a mismatch and the signal translocates along the DNA using the MutL protein to activate the MutH protein which incises the DNA at the nearest unmethylated GATC. The MutH protein, therefore, distinguishes between the methylated and unmethylated target sequence.

Deamination of methylcytosine leads to a T-G mismatch in DNA at any time after replication. In *E. coli* a mismatch correction system recognizes the mismatch:

arising from deamination of one of the methylcytosines in the target sequence of the Dcm methylase (Zell & Fritz, 1987). This very short patch repair system also uses the MutL and the MutS proteins, but not the MutH protein, to help repair the mismatch.

A similar, very short patch repair system is also present in eukaryotes (Wiebauer & Jiricny, 1989) and this may be of singular importance (see below) as the initial step appears to be the removal, by a glycosylase, of the methylated pyrimidine (thymine) paired with guanine. If not repaired prior to replication then the daughter molecules will have

in place of the original

Such mutations are frequent both in prokaryotes (Coulondre *et al.*, 1978) and eukaryotes (Barker *et al.*, 1984) and are believed to be one of the reasons why the vertebrate genome is deficient in the CG dinucleotide (Salser, 1977; Bird *et al.*, 1979; Adams & Eason, 1984).

There is a third mismatch correction system which relies on nicks arising in the daughter strand at replication to discriminate between the correct and incorrect base (Lahue *et al.*, 1987) and this appears of more significance in eukaryotes (Hare & Taylor, 1985).

Initiation of replication

There are two examples where methylation plays a role in the initiation of replication.

In prokaryotes such as *E. coli* and plasmid P1 the origin of replication is characterized by the presence of multiple GATC sequences. This is the sequence methylated by the Dam methylase and replication is reduced (but not completely inhibited) in the absence of the enzyme. Replication is dependent on a transcript which starts outside the origin region and terminates at a number of sites within the origin where DNA synthesis is initiated. These sites include the GATCs, and it is possible they are recognized by an initiation protein when fully methylated or that methylation reduces the stability of the duplex, thereby allowing unwinding to occur (Barras & Marinus, 1989). As methylation is delayed by several minutes following replication this might ensure that

reinitiation does not occur. The alternative suggestion for which evidence is now accumulating is that only the hemimethylated GATC sites bind to a membrane protein involved in chromosome segregation (Ogden *et al.*, 1988). This would bring about the separation of the two daughter chromosomes in the short period following initiation of replication before they become fully methylated. In the absence of the Dam methylase the timing of replication is disturbed, but the bacteria remain viable (Bakker & Smith, 1989). However if hemimethylated or methylated plasmid DNA is introduced into such *dam*⁻ cells it fails to replicate, in contrast to unmethylated plasmids (Russell & Zinder, 1987). This is consistent with the segregation of the hemimethylated origin regions to the membrane until the time when they become fully methylated.

The other example comes from *Physarum* where the ribosomal RNA genes are present as an extra-chromosomal dimer containing four origins of replication in the non-transcribed, central spacer region. Only one origin is used in any one molecule and this origin is characterized by hypomethylation (Cooney *et al.*, 1988).

Chromatin structure

The chromatin of eukaryotic nuclei can be fractionated into regions which are highly sensitive, sensitive or insensitive to the action of nucleases *in vitro*. Genes which are active (i.e. are being transcribed or are primed for transcription) are found in the highly sensitive fraction and contain a smaller proportion of their cytosines methylated than do inactive genes. The extent of methylation is usually determined by the use of a pair of isoschizomeric restriction endonucleases. Thus, whereas *MspI* cleaves CCGG sequences whether or not the internal cytosine is methylated, *HpaII* will act only on unmethylated DNA. Cleavage of genomic DNA with the restriction enzyme, followed by electrophoresis, Southern transfer and probing with the appropriate fragment of the cloned gene, gives a pattern of bands from which the presence of methyl groups in particular CCGG sequences can be determined. The disadvantage of this method is that only about 4% of methylcytosines occur in this sequence and so analysis is at best patchy. The alternative and much superior method involves genomic sequencing and a recent modification of this method using it in combination with the polymerase chain reaction makes it simpler to use (Saluz & Jost, 1989). This approach is able to give information as to the location *in vivo* of all the methylcytosines in a gene and its flanking sequences.

The undermethylated region extends only a little beyond the transcribed region and does not include the whole of the nuclease-sensitive region which may correspond to a loop of DNA (Stalder *et al.*, 1980). This undermethylation may be a cause or effect of both the chromatin structure and the gene transcription. The current thinking, however, is that an inactive gene in which the DNA is methylated will assume a nuclease-inaccessible, condensed conformation, but that this change may be reversed should the gene become transcriptionally active (see below).

On transfecting or injecting a gene into an animal cell it associates over 8 h with histones to form chromatin (Graessmann *et al.*, 1985). Methylated genes form inactive chromatin and so it is a not uncommon observation that methylation inhibits the activity of transfected DNA. To be effective this methylation does not need to be in the

control region of the gene and methylation towards the 3' end is often equally as effective as methylation at the 5' end (Busslinger *et al.*, 1983; Keshet *et al.*, 1985).

Murray & Grosveld (1987), studying the human γ -globin gene, showed that a minimum length of methylcytosine-free DNA is required for transcription. Furthermore, the gene is transcribed if all CG dinucleotides are removed from the promoter region, i.e. a methylcytosine-free island is required which can arise from the absence of CG dinucleotides or from their failure to be methylated. This is not consistent with the involvement of a sequence-specific transcription factor, but rather with the association of a methylcytosine-binding protein which might prevent transcription and seed the formation of inactive chromatin. A similar conclusion can be drawn from the study of Ben-Hattar & Jiricny (1988), which showed that a methylcytosine in any one of four sites in the promoter region of the *Herpes simplex tk* gene reduced transcription of the gene in *Xenopus* oocytes.

Further support for the proposal that the inhibitory action of methylation is mediated by the formation of inactive chromatin comes from the observation that transcription factor binding is not involved in the inhibition of expression of the adenovirus E2A gene caused by methylation of three upstream sites (Hoeveler & Doerfler, 1987; Dobrzanski *et al.*, 1988). This effect of methylation can be reversed by the protein coded for by the E1A gene of adenovirus (Langner *et al.*, 1986) or by a protein coded by frog virus 3 (Thompson *et al.*, 1986). These proteins may act at strong promoters on the adenovirus E2 gene to override the effect of methylation or may neutralize a methylcytosine-binding protein, thereby allowing ready access of all transcription factors to the gene (see below).

In this context, Saluz *et al.* (1988) have identified a protein which binds to a short region at the beginning of the chick vitellogenin gene only when this sequence is methylated, and Meehan *et al.* (1990) have partially purified a protein which binds to DNA containing 15 or more mCG dinucleotides and does not bind to unmethylated DNA. Ehrlich's group have isolated methylcytosine-binding proteins from both animals and plants (Khan *et al.*, 1988; Zhang *et al.*, 1989).

Further evidence that methylation may be acting via an effect on chromatin structure comes from a study of the time course of expression of genes injected into vertebrate cells (Buschhausen *et al.*, 1987). Little or no effect of methylation is observed at early times, but the inhibitory effect becomes apparent as the DNA is complexed into inactive chromatin.

An additional, fascinating experiment performed by Nick Hastie (personal communication) was to introduce a chromosome from the yeast *Saccharomyces pombe* (which is unmethylated) into a mouse cell where it became integrated as a less condensed, non-methylated 'continent' within the more condensed, methylated mouse chromosome. This shows that, as well as acting on individual genes, a methylcytosine-binding protein may seed the formation of large stretches of inactive chromatin. In vertebrate somatic cells, satellite DNA sequences are heavily methylated and are present in an inactive, condensed, late-replicating chromatin configuration and the same is true for one of the X-chromosomes in the cells of female mammals (see below). In *Neurospora crassa* and *Ascobolus immersus* amplified DNA also becomes heavily methylated and inactivated

in the premeiotic phase (Bull & Wootton, 1984; Selker & Stevens, 1987; Selker *et al.*, 1988; Goyon & Faugeron, 1989). This modification affects DNA undergoing duplication as well as transforming DNA introduced into the cells, and all copies of the duplicated DNA are inactivated. A similar, though less extreme, inactivation associated with methylation occurs to DNA introduced into plant cells using *Agrobacterium tumefaciens* (Ama-sino *et al.*, 1984; Peerbolte *et al.*, 1986; Matzke *et al.*, 1989). This inactivation and methylation *de novo* may be related to the pairing of DNA segments required for recombination.

The possible importance of a methylcytosine binding protein ('X') is considered further in the next section.

Transcription factor binding

The initial results showing an inverse correlation between methylation and gene expression were interpreted as indicating that methylation might be interfering with the binding of transcription factors to their target sequences in the promoter regions of genes. However, experiments involving transfecting cells with cloned genes which had been methylated *in vitro* with bacterial methylases (usually *HpaII* or *HhaI* methylase) led to a more ambiguous conclusion (see Adams & Burdon, 1985, for a consideration of several examples). This is probably a consequence of the variety of genes under investigation, each with a different, complex promoter, and the use of bacterial methylases to modify the gene.

Consider a gene whose promoter has four regions to which different transcription factors (A, B, C and D) can bind. Transcription may result if any one of the four binds, but is greater when two factors are bound. The following situations can be envisaged (Fig. 4).

Factor A is ubiquitous but binding is blocked by methylation of a CCGG sequence in the promoter.

Factor B is present only in fibroblasts and binding is blocked when a GCGA sequence in the promoter is methylated. There are no CCGGs in this region of the promoter.

Factor C is present only in hepatocytes and the binding sequence does not contain any CG dinucleotides.

Factor D is present in brain cells and its interaction is unaffected by methylation of a CCGG sequence in its recognition site.

In addition, the gene contains a number of CG dinucleotides (some in CCGGs) throughout its length which, when methylated, can bind the methylcytosine binding protein: factor X. This enhances chromatin condensation leading to a structure which is inaccessible to transcription factors. Once in this form the gene can be activated only at replication (or repair?). However,

	CCGG, GCGA, CCGG				Start CG	CCGG	GCGA	CCGG	
Binding	A	B	C	D	X	X	X	X	Expression
Most cells	-	0	0	0	-	+	-	+	-
Fibroblasts	-	+	0	0	-	+	-	+	+
Hepatocytes	-	0	+	0	-	+	-	+	+
Brain	-	0	0	+	-	+	-	+	+

Fig. 4. Transcription factor binding and *HpaII* methylation

A hypothetical gene is illustrated whose transcription can be activated by the binding of any one of four transcription factors A–D. The dots indicate the site of action of *HpaII* methylase. See text for full details. –, no binding or expression; +, binding or expression; 0, factor absent.

binding (or function) of X is prevented by transcription and may also be prevented by certain other factors which do not themselves promote transcription (i.e. determination factors).

The effect of methylation *in vitro* with *HpaII* methylase (which methylates the internal cytosine in the sequence CCGG) on expression of this gene depends on the cells transfected.

In most cells (which contain A and X) methylation will block expression and the gene will assume an inactive chromatin conformation.

In fibroblasts, hepatocytes and brain cells, tissue-specific factors will bind irrespective of *HpaII* methylation and the gene will be expressed.

Initially expression will not be as high as with the unmethylated gene (when factor A will also bind), but replication may result in demethylation of the factor A recognition sequence and enhanced expression.

Thus methylation is acting in several ways to control gene expression:

- (1) an on/off switch,
- (2) a modulator by controlling the number of factors that can bind, and
- (3) a centre from which non-transcribed genes can be packaged into inactive chromatin.

This scenario can explain all the results on transfection of genes methylated by *HpaII* methylase which methylates only a limited number of CGs. The alternative method of methylating DNA has been to use methyl dCTP to produce complete methylation of all cytosines, but this leads to equally spurious interpretations in immediate expression assays as factor binding may be affected in a purely artefactual manner. The use of the eukaryotic DNA methylase is only in its infancy because the enzyme is difficult to obtain in large amounts and acts only very inefficiently. It methylates all cytosines in CG dinucleotides and probably cytosines in CAG and CTG as well (Simon *et al.*, 1983; Graessmann *et al.*, 1983; Woodcock *et al.*, 1987, 1988; Adams *et al.*, 1988). By using gapped duplex DNA molecules in the presence of 100 mM-NaCl the mouse methylase can be used as a site specific DNA methylase (Adams *et al.*, 1989b).

Perhaps the best illustration of the effect of multiple transcription factors on a complex promoter is that found by Yisraeli *et al.* (1986) for the α -actin gene. When methylated using methyl dCTP and introduced into fibroblasts, expression is blocked. However, when the methylated gene is introduced into myoblasts it is demethylated and expressed. Similar tissue-specific demethylation has been found for the IgG κ gene in B lymphocytes and for the rat insulin gene in insulinoma cells (Cedar, 1989). When a housekeeping gene (*dhfr*), methylated only at *HpaII* or *HhaI* sites (as opposed to having all its cytosine replaced by methylcytosine), is introduced into undifferentiated CHO cells it is demethylated at specific sites around the promoter (Shimada *et al.*, 1987). These sites correspond to DNAase I hypersensitive sites found in active chromatin.

No factor binding studies have been done in these systems, but methylation at a single site has been shown to interfere with the binding of (a) the major late transcription factor (MLTF) to the major late promoter of adenovirus (Watt & Molloy, 1988); (b) the E2 factor to the adenovirus E2 promoter (Kovesdi *et al.*, 1987); (c) a factor to its binding site upstream of the rat tyrosine aminotransferase gene (Becker *et al.*, 1987); and (d)

cyclic AMP-responsive element binding protein (CREB) to the CRE (Iguchi-Ariga & Schaffner, 1989). So far, this is the only direct evidence in eukaryotes that methylation can block gene expression by interfering with the binding of transcription factors.

All housekeeping genes (and some tissue-specific genes) contain a GC-rich island at their 5' end (Gardiner-Garden & Frommer, 1987) and this often contains a binding site for the transcription factor Sp1. Sp1 is ubiquitous but its binding site does not contain any sites which can be methylated by bacterial methylases. Rather the target sequence contains the sequence GCCG and binding has been studied using synthetic oligonucleotides or eukaryotic methylase and shown not to be affected by methylation (Holler *et al.*, 1988; Harrington *et al.*, 1988; Bryans & Adams, 1989; M. Bryans & R. L. P. Adams, unpublished work). GC-rich islands in active housekeeping genes are not methylated *in vivo* and are also known as methylation-free islands or MFIs. (An exception to this is a truncated *c-myc* gene which, though it is expressed, has a methylated GC-rich island covering exons 1 and 2; Bianchi *et al.*, 1989.) Sp1 is ubiquitous and, as housekeeping genes require binding for transcription to occur, it seems important that this binding is not inhibited by chance methylation events which would otherwise be lethal. The function of Sp1 (present at 5000–10000 copies per cell; Kadonaga *et al.*, 1986) and its binding site would then be, not to regulate to housekeeping genes. There are, however, a few occasions when housekeeping genes are not transcribed (e.g. on the inactive X chromosome) and in such situations the MFIs are no longer methylation free. Rather they are heavily methylated and present in inactive chromatin structures. From the point of view of Sp1 binding it is immaterial whether or not its binding site is methylated, but methylation may contribute to the formation of the inactive chromatin structure by allowing binding of the methylcytosine-binding protein X.

It is clear from these remarks that few generalizations can be made and that each system has its own peculiarities.

FUNCTION OF DNA METHYLATION IN EUKARYOTES

Maintenance

The big advantage of the DNA methylase is that it is able to maintain a given pattern of methylation throughout multiple rounds of DNA replication (Fig. 1). Thus gene expression is stable in a differentiated tissue and this stability is guaranteed by a single enzyme, rather than requiring a different system and the continuous presence of specific factors for each individual gene. However, tissues differ in their pattern of methylation and the particular patterns are established during early development and can be changed, within precise limitations, on subsequent differentiation. DNA methylation is involved in maintaining an established pattern of gene expression, but definite evidence that it performs a determining role in differentiation is not available. New methylation patterns may be imposed on a gene as a result of changes in expression, or they may be the means by which those changes are brought about. In view of the arguments put forward above, the former appears more likely and implies that differentiation occurs by the

programmed (transient) appearance of specific transcription factors which are able to override methylation controls and/or impose new patterns.

Differentiation and methylation *de novo*

Each egg receives from its parents a set of germ line DNA which is imprinted with a characteristic methylation pattern (Monk, 1988). Thus, even before any DNA replication has occurred, the maternally-derived allele of a gene may differ from the paternally-derived allele. The importance of this imprinting is not clear, though it may simply reflect the pattern of gene expression in the two germ lines. In sperm, where most genes are inactive, the level of DNA methylation is high. The exceptions to this generalization are (a) the MFIs towards the 5' ends of housekeeping genes (Bird, 1986; Groudine & Conkin, 1985) and (b) repetitive DNAs which are largely unmethylated in sperm and sometimes in oocytes (Sanford *et al.*, 1984; Adams *et al.*, 1983; Feinstein *et al.*, 1985; Ponzetto-Zimmerman & Wolgemuth, 1984); an observation commensurate with their proposed role in recombination events which occur at high frequency at meiosis.

Shortly after fertilization, the preimplantation mammalian embryo has the capacity to perform demethylation and *de novo* methylase reactions (Monk *et al.*, 1987; Silva & White, 1988). This latter potential, for what is normally only a minor reaction, is probably brought about as a result of the presence of extremely high levels of the DNA methylase (Adams *et al.*, 1989a). The methylation *de novo* is most obvious for repetitive DNA and, in females, for the GC-rich regions 5' to the genes in one of the X chromosomes (Chapman *et al.*, 1984; Sanford *et al.*, 1985, 1987; Monk, 1986). Although the methylation of the X-chromosomal DNA appears to occur after inactivation, genes on the inactive X-chromosome can be reactivated by treatment with azacytidine, showing that methylation is an important factor in maintaining the repressed state (Mohandas *et al.*, 1981; Lock *et al.*, 1986; Hansen *et al.*, 1988). Furthermore, the MIC2 gene, which keeps its MFI even on the 'inactive' X-chromosome, is active on both chromosomes (Mondello *et al.*, 1988). In addition, any DNA injected into the egg at this stage is later found methylated and in an inactive chromatin configuration (Jahner *et al.*, 1982). Such methylation *de novo* is not seen for DNA injected into the cells of the post-implantation embryo at a time when one X chromosome is already inactive and the repetitive DNA highly methylated.

In the presence of a *de novo* methylating environment, i.e. in preimplantation embryos, any regions not protected from the methylase by DNA-binding proteins would be expected to become methylated. Transcription factors will be bound in a cell-specific manner to transcribed genes and also to some genes, which although not presently transcribed, are programmed to be transcribed later in that cell lineage (Saluz *et al.*, 1986; Kolsto *et al.*, 1986). This will result in the control regions of such genes being unmethylated. It is difficult to use this argument to explain how the MFIs, which average 1 kb in length, escape methylation as the target sequences for transcription factors are usually less than 10 bp in length. The MFIs, which include the 5' flanking region and often the first exon or more of a gene (Bird, 1986), may be maintained in some unusual conformation which results in their failure to be methylated on all but one X

chromosome. In the ancestral genome consisting of AT-rich and GC-rich regions, the former may have become methylated and as a consequence of deamination of methylcytosine become even more AT-rich and deficient in the dinucleotide CG. This would leave those regions refractory to methylation as MFIs.

In addition, certain genes will be complexed with histones in inactive chromatin and this may limit the accessibility of the methylase to the body of inactive genes. Indeed, the genes on the inactive X chromosome are found to have a variable, but low, level of methylation except in their heavily methylated 5' flanking region.

Demethylation

The changes observed in the early embryo involve methylation *de novo*, but it is clear that this must be balanced by a demethylation of DNA during gametogenesis or early embryogenesis to produce the undermethylated GC-rich islands, X chromosomes and repetitive DNA. Demethylation similar to that occurring in the early embryo can be studied *in vitro* using the teratocarcinoma, F9 cells. Methylated housekeeping genes introduced into these cells are demethylated in the 5' regions, but tissue-specific genes are not demethylated (Cedar, 1989). However, as tissue-specific genes are undermethylated in the tissue of expression their demethylation must occur during differentiation, and has been observed on inducing differentiation in a number of tissues (Young & Tilghman, 1984; Bestor *et al.*, 1984; Razin *et al.*, 1986).

Demethylation can occur passively by a failure to methylate newly replicated DNA (Fig. 1). After one division the DNA will be hemimethylated and after two divisions half the genes will be unmethylated. This is a slow process and does not seem sufficient to explain some of the findings of rapid demethylation such as that seen on inducing the vitellogenin gene by oestrogen in the chick liver (Wilks *et al.*, 1984) or the globin gene in stimulated erythroleukaemia cells (Razin *et al.*, 1986).

From the work of Saluz *et al.* (1986) it is clear that a hemimethylated site exists in the uninduced vitellogenin gene at the glucocorticoid receptor binding site. This site may be maintained by determination factors which protect one strand of the DNA from methylation and which would then allow a much faster response to oestrogen. Other sites in the overlapping oestrogen-responsive element (ERE) become hemimethylated within 1 day of oestrogen administration. Furthermore, in addition to the oestrogen receptor, a protein NHP1 (which has been purified from HeLa cells; Hughes *et al.*, 1990) binds to an AGCG sequence in the ERE and makes nicks in the DNA near its binding site. This ubiquitous protein shows a preference for methylated DNA. It may be analogous to the MutH protein in allowing access of a repair system which, in this case, replaces the methylcytosine with cytosine. Hypersensitive sites exist at the promoter regions in the chromatin of many genes and may allow rapid demethylation of such regions.

In the mouse erythroleukaemia (MEL) cell the demethylation observed is genome wide. The proportion of CG dinucleotides methylated falls transiently from over 80% to as low as 40% at 12–20 h after stimulation with hexamethylene bisacetamide. Razin *et al.* (1986) postulate that an enzyme system is activated which replaces methylcytosine with cytosine without breaking the phospho-

diester backbone of the DNA and a similar mechanism is proposed for the hypomethylation associated with the Epstein-Barr virus lytic cycle (Szyf *et al.*, 1985). The evidence for such a transglycosylation mechanism revolves around the incorporation of [5-³H]deoxycytidine into normal density DNA when MEL cells were labelled in the presence of bromodeoxyuridine. Such incorporation was observed for only a limited period corresponding to the time of reduced methylation and the incorporated radioactivity was lost (*as the cytosine is methylated*) during a chase. This result is also consistent with a very short patch repair process which has been shown to recognize and remove a methylated pyrimidine paired with guanine (Wiebauer & Jiricny, 1989), but no similar incorporation of deoxyadenosine was seen apparently ruling out this mechanism. Using cells prelabelled with [¹⁴C]deoxycytidine (which labels thymine, cytosine and methylcytosine) we have found no diminution in the proportion of prelabelled cytosines methylated when MEL cells are stimulated with either dimethyl sulphoxide or hexamethylene bisacetamide. This result is compatible with a very short patch repair (in which all pyrimidines in the patch are replaced), but not with a base exchange mechanism of demethylation.

Aging and cancer

Cell lines growing in culture exhibit small irregular changes in their pattern of DNA methylation but the overall level remains fairly constant (Wigler *et al.*, 1981; Reis & Goldstein, 1982). This is not so for primary cells, which steadily lose 5-methylcytosine as their proliferative potential declines (Wilson & Jones, 1983). The rate of fall is high with mouse cells but much less with human cells, a change which correlates with the rate of transformation of these cells to a malignant phenotype (Holliday, 1987). In turn, this correlates with the finding that small, short-lived animals have a far higher incidence of cancer than long-lived animals (Holliday, 1989; Wilson *et al.*, 1987).

The presence of a low 5-methylcytosine content is not necessarily harmful. BHK21 fibroblasts have only 2% of their cytosines methylated and the T11 lymphoblastoid cell line only 1.6% (Macnab *et al.*, 1988; Lindsay *et al.*, 1989). It is probably important where the remaining methylcytosines are situated, and Michalowsky & Jones (1989) found some sites persistently methylated in three clones having only 0.65% of their cytosines methylated.

When a transformed cell arises spontaneously from a primary cell line this quickly replaces the original, aging population to produce a cell line with a stable level of DNA methylation. One can speculate that the initial population is aging as a result of the failure of the methylase to maintain certain genes inactive. As inappropriate genes are activated, they disturb the cellular metabolism leading to reduced growth rates and eventually to cell death. Every so often a cell will arise in which a proto-oncogene has been activated by demethylation. This cell may now grow independently of growth factors normally provided in the culture medium and may fail to show the growth limitations imposed by a requirement to be attached to a substratum, i.e. it is a transformed cell which will quickly outgrow the surrounding cells.

The transformation process may involve the activation of two or more genes in a random manner, i.e. it is a multistep process, and is akin to the process of tumorigenesis which occurs *in vivo*. Agents which increase the rate of demethylation would be expected to increase the

rate at which transformation occurs and both azacytidine and transformation with herpes simplex virus may act in this manner (Hsiao *et al.*, 1985; Macnab *et al.*, 1988).

In contrast, DNA methylation may enhance the chance that a tumour-suppressing gene might be inactivated. Indeed many mutant cells are believed to arise by a mutation on one allele of a functionally heterozygous cell in which the other allele is already inactivated by methylation. Thus patients with spontaneous Wilm's tumour or retinoblastoma/osteosarcoma show the loss of one allele (the maternal) while the paternal allele is retained in an inactive, imprinted, methylated state (Holliday, 1987; Toguchida *et al.*, 1989; Reik & Surani, 1989).

SUMMARY

DNA methylation is found almost ubiquitously in nature and the methyltransferases show evidence of a common evolutionary origin. It will be a fascinating study in protein evolution to follow the ways in which the structures of the various enzymes have developed.

Although methylation may have a direct effect on DNA structure the evidence for the importance of this *in vivo* is accumulating only slowly. In contrast, there is now abundant evidence that methylation of DNA affects DNA-protein interactions and so may have a function in all processes in which such interactions occur. The binding of nucleases is affected in the processes of mismatch repair, DNA restriction and possibly demethylation during differentiation in vertebrates. The binding of transcription factors is affected by DNA methylation and the association of DNA with packaging and segregation proteins may play a part in the control of transcription and replication. The interplay of these effects makes DNA methylation a complex but rewarding area for study.

Perhaps we should no longer refer to methylcytosine and methyladenine as *minor* bases, but rather as *key* bases which help regulate the functions of DNA.

I would like to acknowledge the Medical Research Council and the Cancer Research Campaign which have funded my research in recent years, and Professor M. D. Houslay and the University of Glasgow for the provision of facilities. I would also like to thank my graduate students for valuable discussions, and Mrs Angela Rinaldi for invaluable technical assistance.

REFERENCES

- Adams, R. L. P. (1971) *Biochim. Biophys. Acta* **254**, 205-212
- Adams, R. L. P. & Burdon, R. H. (1983) in *Enzymes of Nucleic Acid Synthesis and Processing* (Jacob, S. T., ed.), vol. 1, pp. 119-144, CRC Press, Boca Raton, FL
- Adams, R. L. P. & Burdon, R. H. (1985) *Molecular Biology of DNA Methylation*, Springer Verlag, New York
- Adams, R. L. P. & Eason, R. (1984) *Nucleic Acids Res.* **12**, 5869-5877
- Adams, R. L. P., McKay, E. L., Craig, L. M. & Burdon, R. H. (1979) *Biochim. Biophys. Acta* **561**, 345-357
- Adams, R. L. P., Fulton, J. & Kirk, D. (1982) *Biochim. Biophys. Acta* **697**, 286-294
- Adams, R. L. P., Burdon, R. H. & Fulton, J. (1983) *Biochem. Biophys. Res. Commun.* **113**, 695-702

- Adams, R. L. P., Gardiner, K., Rinaldi, A., Bryans, M., McGarvey, M. & Burdon, R. H. (1986) *Biochim. Biophys. Acta* **868**, 9–16
- Adams, R. L. P., Davis, T., Rinaldi, A. & Eason, R. (1987) *Eur. J. Biochem.* **165**, 107–115
- Adams, R. L. P., Rinaldi, A., McGarvey, M., Bryans, M. & Ball, K. (1988) *Gene* **74**, 125–128
- Adams, R. L. P., Rinaldi, A. & Monk, M. (1989a) *J. Cell. Biochem. Suppl.* **13D**, 220
- Adams, R. L. P., Bryans, M., Goddard, J. P., Rinaldi, A., Smart, A. & Yesufu, H. M. I. (1989b) *Philos. Trans. R. Soc. London Ser. B*, in the press
- Amasino, R. M., Powell, A. L. T. & Gordon, M. P. (1984) *Mol. Gen. Genet.* **197**, 437–446
- Bakker, A. & Smith, D. W. (1989) *J. Cell. Biochem. Suppl.* **13D**, 115
- Barker, D., Schafer, M. & White, R. (1984) *Cell* **36**, 131–138
- Barras, F. & Marinus, M. G. (1989) *Trends Genet.* **5**, 139–143
- Becker, P. B., Ruppert, S. & Schutz, G. (1987) *Cell* **51**, 435–443
- Behe, M. & Felsenfeld, G. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1619–1623
- Ben-Hattar, J. & Jiricny, J. (1988) *Gene* **65**, 219–227
- Bestor, T. H., Hellewell, S. B. & Ingram, V. M. (1984) *Mol. Cell. Biol.* **4**, 1800–1806
- Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. (1989) *J. Mol. Biol.* **203**, 971–983
- Bianchi, N. O., Bianchi, M. S., Alitalo, K. & de la Chapelle, A. (1989) *Biochim. Biophys. Acta* **1007**, 350–355
- Bickle, T. A. (1982) *Cold Spring Harbor Monogr. Ser.* **14**, 85–108
- Bird, A. P. (1986) *Nature (London)* **321**, 209–213
- Bird, A. P., Taggart, M. H. & Smith, B. A. (1979) *Cell* **17**, 889–901
- Bolden, A. H., Nalin, C. M., Ward, C. A., Poonian, M. S. & Weissbach, A. (1986) *Mol. Cell. Biol.* **6**, 1135–1140
- Bramhill, D. & Kornberg, A. (1988) *Cell* **54**, 915–918
- Bryans, M. & Adams, R. L. P. (1989) *J. Cell. Biochem. Suppl.* **13D**, 221
- Bull, J. H. & Wootton, J. C. (1984) *Nature (London)* **310**, 701–704
- Buschhausen, G., Wittig, B., Graessmann, M. & Graessmann, A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1177–1181
- Busslinger, M., Hurst, J. & Flavell, R. A. (1983) *Cell* **34**, 197–206
- Cedar, H. (1989) *Philos. Trans. R. Soc. London Ser. B*, in the press
- Chapman, V., Forrester, L., Sanford, J., Hastie, N. & Rossant, J. (1984) *Nature (London)* **307**, 284–286
- Collins, M. & Myers, R. M. (1987) *J. Mol. Biol.* **198**, 737–744
- Cooney, C. A., Eykholt, R. L. & Bradbury, E. M. (1988) *J. Mol. Biol.* **204**, 889–901
- Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. (1978) *Nature (London)* **274**, 775–780
- Cowan, G. M., Gann, A. A. F. & Murray, N. E. (1989) *Cell* **56**, 103–109
- Dobrzanski, P., Hoeveler, A. & Doerfler, W. (1988) *J. Virol.* **62**, 3941–3946
- Feinstein, S. I., Racaniello, V. R., Ehrlich, M., Gehrke, C. W., Miller, D. A. & Miller, O. J. (1985) *Nucleic Acids Res.* **13**, 3969–3978
- Frederick, C. A., Quigley, G. J., van der Mare, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. (1988) *J. Biol. Chem.* **263**, 17872–17879
- Gardiner-Garden, M. & Frommer, M. (1987) *J. Mol. Biol.* **196**, 261–282
- Goyon, C. & Faugeron, G. (1989) *Mol. Cell. Biol.* **9**, 2818–2827
- Graessmann, M., Graessmann, A., Wagner, H., Werner, E. & Simon, D. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6470–6474
- Graessmann, A., Bumke-Vogt, C., Buschhausen, G., Bauer, M. & Graessmann, M. (1985) *FEBS Lett.* **179**, 41–45
- Groudine, M. & Conkin, K. F. (1985) *Science* **228**, 1061–1068
- Gruenbaum, Y., Naveh-Man, T., Cedar, H. & Razin, A. (1981a) *Nature (London)* **292**, 860–862
- Gruenbaum, Y., Stein, R., Cedar, H. & Razin, A. (1981b) *FEBS Lett.* **124**, 67–71
- Gruenbaum, Y., Cedar, H. & Razin, A. (1982) *Nature (London)* **295**, 620–622
- Hansen, R. S., Ellis, N. A. & Gartler, S. M. (1988) *Mol. Cell. Biol.* **8**, 4692–4699
- Harrington, M. A., Jones, P. A., Imagawa, M. & Karin, M. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2066–2070
- Hare, J. T. & Taylor, J. H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7350–7354
- Hoeveler, A. & Doerfler, W. (1987) *DNA* **6**, 449–460
- Holler, M., Westin, G., Jiricny, J. & Schaffner, W. (1988) *Genes Dev.* **2**, 1127–1135
- Holliday, R. (1987) *Science* **238**, 163–170
- Holliday, R. (1989) *Nature (London)* **337**, 311
- Hsiao, W.-L. W., Gattoni-Celli, S. & Weinstein, I. B. (1985) *Mol. Cell. Biol.* **5**, 1800–1803
- Hughes, M. J., Liang, H., Jiricny, J. & Jost, J.-P. (1990) *Biochemistry*, in the press
- Iguchi-Arigo, M. M. & Schaffner, W. (1989) *Genes Dev.* **3**, 612–619
- Jahner, D., Stuhlmann, H., Stewart, C. L., Harbers, K., Lohler, J., Simon, I. & Jaenisch, R. (1982) *Nature (London)* **298**, 623–628
- Jaworski, A., Zacharias, W., Hseh, W.-T., Blaho, J. A., Larsen, J. E. & Wells, R. D. (1988) *Gene* **74**, 215–220
- Jones, P. A. & Taylor, S. M. (1981) *Nucleic Acids Res.* **9**, 2933–2947
- Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23
- Keshet, I., Yisraeli, J. & Cedar, H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2560–2564
- Khan, R., Zhang, X.-Y., Supakar, P. C., Ehrlich, K. C. & Ehrlich, M. (1988) *J. Biol. Chem.* **263**, 14374–14383
- Kolsto, A.-B., Kollias, G., Giguere, V., Isobe, K.-I., Prydz, H. & Grosveld, F. (1986) *Nucleic Acids Res.* **14**, 9667–9678
- Kovesdi, I., Reichel, R. & Nevins, J. R. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2180–2184
- Kruger, D. H., Schroeder, C., Reuter, M., Bogdarina, I. G., Buryanov, Y. I. & Bickle, T. A. (1985) *Eur. J. Biochem.* **150**, 323–330
- Kruger, D. H., Schroeder, C., Reuter, M., Bickle, T. A., Bogdarina, I. G. & Buryanov, Y. I. (1988) *Gene* **74**, 85–87
- Lahue, R. S., Su, S.-S. & Modrich, P. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1482–1486
- Langner, K.-D., Weyer, U. & Doerfler, W. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1598–1602
- Lauster, R. (1988) *Gene* **74**, 243
- Lindsay, S., Adams, R. L. P. & Monk, M. (1989) *Hum. Genet.* **81**, 252–256
- Lock, L. F., Melton, D. W., Caskey, C. T. & Martin, G. R. (1986) *Mol. Cell. Biol.* **6**, 914–924
- Lock, L. F., Takagi, N. & Martin, G. R. (1987) *Cell* **48**, 39–46
- Lyons, S. M. & Schendel, P. F. (1984) *J. Bacteriol.* **159**, 421–423
- Macnab, J. C. M., Adams, R. L. P., Rinaldi, A., Orr, A. & Clark, L. (1988) *Mol. Cell. Biol.* **8**, 1443–1448
- Matzke, M. A., Primig, M., Trnovsky, J. & Matzke, A. J. M. (1989) *EMBO J.* **8**, 643–649
- McClelland, M. & Nelson, M. (1988) *Gene* **74**, 291–304
- Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, F. L. & Bird, A. P. (1989) *Cell* **58**, 499–507
- Messer, W. & Noyer-Weidner, M. (1988) *Cell* **54**, 735–737
- Michalowsky, L. A. & Jones, P. A. (1989) *Mol. Cell. Biol.* **9**, 885–892

- Mohandas, T., Sparkes, R. S. & Shapiro, L. J. (1981) *Science* **211**, 393–396
- Mondello, C., Goodfellow, P. J. & Goodfellow, P. N. (1988) *Nucleic Acids Res.* **16**, 6813–6824
- Monk, M. (1986) *Bioessays* **4**, 204–208
- Monk, M. (1988) *Genes Dev.* **2**, 921–925
- Monk, M., Boubelik, M. & Lehnert, S. (1987) *Development* **99**, 371–382
- Murray, E. J. & Grosveld, F. (1987) *EMBO J.* **6**, 2329–2335
- Ogden, G. B., Pratt, M. J. & Schaechter, M. (1988) *Cell* **54**, 127–135
- Peerbolte, R., Leenhouts, K., Hooykaas-van Slogteren, G. M. S., Wullems, G. J. & Schileroort, R. A. (1986) *Plant Mol. Biol.* **7**, 285–299
- Pfeifer, G. P. & Drahovsky, D. (1986) *Biochim. Biophys. Acta* **868**, 238–242
- Pfeifer, G. P., Speiss, E. B., Grunwald, S., Boehm, T. L. J. & Drahovsky, D. (1985) *EMBO J.* **4**, 2879–2884
- Ponzetto-Zimmerman, C. & Wolgemuth, D. J. (1984) *Nucleic Acids Res.* **12**, 2807–2822
- Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. R. (1989) *Nucleic Acids Res.* **17**, 2421–2435
- Price, C., Shepherd, J. C. W. & Bickle, T. A. (1987) *EMBO J.* **6**, 1493–1497
- Price, C., Lingner, J., Bickle, T. A., Firman, K. & Glover, S. W. (1989) *J. Mol. Biol.* **205**, 115–125
- Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D. & Cantoni, G. L. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2827–2831
- Reik, W. & Surani, M. A. (1989) *Nature (London)* **338**, 112–113
- Reis, R. J. S. & Goldstein, S. (1982) *Nucleic Acids Res.* **10**, 4293–4304
- Russell, D. W. & Zinder, N. D. (1987) *Cell* **50**, 1071–1079
- Saenger, W. (1983) *Principles of Nucleic Acid Structures*, Springer Verlag, New York
- Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 985–1002
- Saluz, H. P. & Jost, J. P. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2602–2606
- Saluz, H. P., Jiricny, J. & Jost, J. P. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7167–7171
- Saluz, H. P., Feavers, I. M., Jiricny, J. & Jost, J. P. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6697–6700
- Sanford, J., Forrester, L., Chapman, V., Chandley, A. & Hastie, N. (1984) *Nucleic Acids Res.* **12**, 2823–2836
- Sanford, J. P., Chapman, V. M. & Rossant, J. (1985) *Trends Genet.* **1**, 89–93
- Sanford, J. P., Clark, H. J., Chapman, V. M. & Rossant, J. (1987) *Genes Dev.* **1**, 1039–1046
- Santi, D. V., Garrett, C. E. & Barr, P. J. (1983) *Cell* **33**, 9–10
- Selker, E. U. & Stevens, J. N. (1987) *Mol. Cell. Biol.* **7**, 1032–1038
- Selker, E. U., Camberesi, E. B., Garrett, P. W., Haack, K. R., Jensen, B. C. & Schabtach, E. (1988) *Gene* **74**, 109–111
- Shimada, T., Inokuchi, K. & Nienhuis, A. W. (1987) *Mol. Cell. Biol.* **7**, 2830–2837
- Silva, A. J. & White, R. (1988) *Cell* **54**, 145–152
- Simon, D., Stuhlmann, H., Jahner, D., Wagner, H., Werner, E. & Jaenisch, R. (1983) *Nature (London)* **304**, 275–277
- Som, S., Bhagwat, A. S. & Friedman, S. (1987) *Nucleic Acids Res.* **15**, 313–332
- Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. & Weintraub, H. (1980) *Cell* **20**, 451–460
- Studier, F. W. & Bandyopadhyay, P. K. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4677–4681
- Sugisaki, H., Kito, K. & Takanami, M. (1989) *J. Biol. Chem.* **264**, 5757–5761
- Szyf, M., Eliasson, L., Mann, V., Klein, G. & Razin, A. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8090–8094
- Tanaka, M., Hibasami, H., Nagai, J. & Ikeda, T. (1980) *Aus. J. Exp. Biol. Med. Sci.* **58**, 391–396
- Thompson, J. P., Granoff, A. & Willis, D. B. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7688–7692
- Toguchida, J., Ishizaki, K., Sasaki, M. S., Nakamura, Y., Ikenaga, M., Kato, M., Sugimoto, M., Kotoura, Y. & Yamamura, T. (1989) *Nature (London)* **338**, 156–158
- Trautner, T. A., Balganes, T., Wilke, K., Noyer-Weidner, M., Rauhut, E., Lauster, R., Behrens, B. & Pawlek, B. (1988) *Gene* **74**, 267
- Turnbull, J. & Adams, R. L. P. (1976) *Nucleic Acids Res.* **3**, 677–695
- Ward, C., Bolden, A., Nalin, C. M. & Weissbach, A. (1987) *J. Biol. Chem.* **262**, 11057–11063
- Watt, F. & Molloy, P. L. (1988) *Genes Dev.* **2**, 1136–1143
- Wiebauer, K. & Jiricny, J. (1989) *Nature (London)* **339**, 234–236
- Wigler, M., Levy, D. & Perucho, M. (1981) *Cell* **24**, 33–40
- Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. & Trautner, T. A. (1988) *EMBO J.* **7**, 2601–2609
- Wilks, A., Seldran, M. & Jost, J.-P. (1984) *Nucleic Acids Res.* **12**, 1163–1177
- Wilson, G. G. (1988) *Trends Genet.* **4**, 314–318
- Wilson, V. L. & Jones, P. A. (1983) *Science* **220**, 1055–1057
- Wilson, V. L., Smith, R. A., Ma, S. & Cutler, R. G. (1987) *J. Biol. Chem.* **262**, 9948–9951
- Woodcock, D. M., Adams, J. K. & Cooper, I. A. (1982) *Biochim. Biophys. Acta* **696**, 15–22
- Woodcock, D. M., Crowther, P. J. & Diver, W. P. (1987) *Biochem. Biophys. Res. Commun.* **145**, 888–894
- Woodcock, D. M., Crowther, P. J., Jefferson, S. & Diver, W. P. (1988) *Gene* **74**, 151–152
- Yesufu, H. M. I., Kuo, Y. M. & Adams, R. L. P. (1989) *Biochem. Soc. Trans.* **16**, 757–758
- Yisraeli, J., Adelstein, R. S., Melloul, D., Nudel, U., Yaffe, D. & Cedar, H. (1986) *Cell* **46**, 409–416
- Young, P. R. & Tilghman, S. M. (1984) *Mol. Cell. Biol.* **4**, 898–907
- Yuan, R. (1981) *Annu. Rev. Biochem.* **50**, 285–315
- Zell, R. & Fritz, H.-J. (1987) *EMBO J.* **6**, 1809–1815
- Zhang, D., Ehrlich, K. C., Supakar, P. & Ehrlich, M. (1989) *Mol. Cell. Biol.* **9**, 1351–1356