5-Methylcytosine in Eukaryotic DNA

Melanie Ehrlich and Richard Y.-H. Wang

Summary. A small portion of the cytosine residues in the DNA of higher eukaryotes as well as in that of many lower eukaryotes is methylated. The resulting 5-methylcytosine residues occur in specific sequences in the DNA, usually adjacent to guanine residues on the 3′ side. This methylation of eukaryotic DNA has been proposed to function in many ways, including control of transcription, maintenance of chromosome structure, repair of DNA, establishment of preferred sites for mutation, oncogenic transformation, and, in certain systems, protection of DNA against enzymatic degradation.

5-Methylcytosine (m^5C) (Fig. 1) appears as a minor base in the DNA of many organisms including all studied higher eukaryotes (*1*). With the exception of bacteriophage XP-12 DNA (*2, 3*), methylation occurs on cytosine (C) residues in specific DNA sequences after polymerization of the DNA (*4*). The resulting $m^5C{\cdot}G$ (G, guanine) base pairs are stronger than the corresponding C·G base pairs (*3*).

Although various unusual bases occur in some types of bacteriophage DNA (*5*), generally the only modified bases found in eukaryotic and bacterial DNA's are m^5C or 6-methyladenine (m^6A) (*1, 6, 7*). Similarities in the amount and the intragenomic distribution of m^5C residues in the DNA of higher eukaryotes and of many lower eukaryotes suggest that methylation of C residues in these diverse genomes shares some common functions. In contrast, in bacterial DNA, which has many of its m^6A or m^5C residues functioning in restriction-modification systems (*8*), the level of methylation of C residues and the types of sequences that are methylated are much more variable (*1*). Therefore, some of the functions of methylation of eukaryotic DNA may be different from those of prokaryotic DNA.

In this article we discuss the inter- and intragenomic distribution of m^5C in various eukaryotic DNA's. Also, we review evidence for DNA methylation having an important role in determining chromosome structure, spontaneous mutation rates, the direction of mismatch repair, the programmed destruction of chromosomes, DNA and RNA synthesis rates, and the regulation of transcription initiation or termination.

M. Ehrlich is an associate professor in the Department of Biochemistry at Tulane Medical School, New Orleans, Louisiana 70112. R. Wang is a Hanna Horner Research Fellow in that department.

Extent of Methylation

The only naturally occurring modified residue found in the DNA from various vertebrates and higher plants is m^5C (*1, 9*). In DNA from most higher plants m^5C constitutes 3.6 to 7.1 percent of the bases (*1*), an amount substantially higher than that in animal DNA. The m^5C in vertebrate DNA is 0.7 to 2.8 mole percent (that is, moles of m^5C per 100 bases) (*1, 9*).

The DNA of lower eukaryotes often contains m^5C as a minor base and, in addition, sometimes m^6A. The DNA's from the unicellular alga *Chlorella pyrenoidosa* (*10*), the yeast *Saccharomyces cerevisiae* (*11*), the sea urchin *Lytechinus variegatus* (*12*), and the slime mold *Physarum polycephalum* (*13*) contain m^5C (0.3 to 3.5 mole percent) as the only detected modified base. In contrast, the DNA's of several types of protozoa (*11, 14*) contain m^6A (0.2 to 2.5 mole percent) rather than m^5C. The nuclear DNA in the unicellular alga *Chlamydomonas reinhardi* contains both m^6A (0.5 mole percent) and m^5C (0.7 mole percent) (*11*) just as the DNA's in most studied bacterial species contain both modified bases (*1*). However, the DNA of cultured mosquito cells also has both m^6A and m^5C but only to the extent of 0.03 mole percent of the bases (*15*). Several types of dinoflagellates are exceptional among eukaryotes because their DNA's contain 5-hydroxymethyluracil partially replacing thymine (*14*).

The only modified base in mitochondrial and chloroplast DNA's is m^5C. In some eukaryotes, the m^5C content of these organelle DNA's is much lower than in the corresponding nuclear DNA (*16*). However, in some other eukaryotes, the m^5C content of mitochondrial and chloroplast DNA has been reported to be similar to or higher than that of the homologous nuclear DNA (*17, 18*). Recent advances in quantitation of m^5C residues in microgram amounts of unlabeled DNA will allow more systematic analyses of the m^5C content of the DNA from these organelles as well as from various nuclear DNA samples (*12, 19, 20*).

Intragenomic Distribution of m^5C Residues

Neighbors of m^5C. The m^5C residues in the DNA of higher eukaryotes are usually bordered on the 3′ side by a G residue. The nearest neighbors of m^5C residues were first studied in DNA which had been digested with deoxyribonuclease I. These digests contain mostly dinucleotides and longer oligonucleotides (*21*). More than 90 percent of the m^5C in the dinucleotide fraction from deoxyribonuclease I digests of calf thymus, HeLa cell, and sea urchin DNA's is in the sequence m^5CG (*22–24*). Higher plant DNA with its much greater percentage of methylated C residues shows a similar increase in the minor portion of m^5C-containing dinucleotides other than m^5CG (*24*). In deoxyribonuclease I digests, only a fraction of the DNA residues is converted to dinucleotides by deoxyribonuclease I, which cleaves DNA with appreciable sequence specificity (*21*), so that the distribution of m^5C among these dinucleotides is not necessarily the same as that among all the dinucleotide pairs of DNA.

Analyses of pyrimidine isostichs (oligonucleotides containing only pyrimidine residues) from the DNA of higher eukaryotes are consistent with the conclusion that G residues are usually adjacent to m^5C on one side. Pyrimidine isostichs are obtained by digestion of the DNA backbone at all purine residues (*25*). Such digestion yields, for example, monopyrimidine nucleotides (pyrimidine mononucleoside diphosphates) from py-

0036-8075/81/0619-1350$01.75/0

rimidine residues that were next to a purine on both sides (. . . Pu Py Pu . . .) and dipyrimidine nucleotides (pyrimidine dinucleoside triphosphates) from two adjacent pyrimidine residues, each of which was bordered on one side by a purine residue (. . . Pu Py Py Pu . . .). In the DNA of multicellular animals the percentage of m^5C in the monopyrimidine nucleotide fraction is approximately twice the percentage of C or of pyrimidines in this fraction (*23, 26*). The m^5C residues of di- and tripyrimidine isostichs of calf thymus, rat spleen, and wheat germ DNA are mostly at the 3′-terminus, that is, bordered in the DNA by a 3′-purine (*27*). Nonetheless, a significant percentage of m^5C residues in the DNA of higher eukaryotes have as the 3′ nearest neighbor a pyrimidine (*27–29*). All these results suggest that in many eukaryotic DNA's, a purine (G) is on the 3′ side of most of the m^5C residues (*30*) and either a pyrimidine or purine is adjacent to m^5C residues on the other (5′) side.

Subfractions of DNA and chromatin. Fragments of sheared eukaryotic DNA can be separated into fractions of different m^5C content on the basis of their ability to renature rapidly. In the DNA of higher animals and plants (*9, 31*), highly repetitive sequences, which reanneal rapidly and consist of mostly nontranscribed DNA, are relatively enriched in m^5C. For example, twice as many C residues are methylated in the highly repetitive DNA of human liver as in the unfractionated DNA (*9*). Nonetheless, m^5C is also found in lower but appreciable amounts in nonrepetitive DNA sequences, which include most of the structural genes.

Two other types of studies indicate that m^5C is often present in higher concentrations in highly repetitive DNA. Upon isopycnic centrifugation, highly repetitive eukaryotic DNA sequences often appear as nuclear satellite peaks whose average buoyant density is different from that of bulk DNA. Some animal and plant nuclear satellite DNA's (*32, 33*) contain more methylated cytosine residues than do the corresponding bulk DNA's. These repetitive DNA sequences are also generally associated with centromeric regions of metaphase chromosomes. Antibodies to m^5C react strongly and specifically with centromeric regions of partly denatured mammalian metaphase chromosomes (*34*). These antibodies also strongly react with some isolated noncentromeric foci along mammalian metaphase chromosomes and in the arms of the Y chromosome. These immunological studies suggest that the distribution of m^5C residues along mitotic chromosomes is not uniform. However, these studies must be interpreted with caution because only the DNA sequences that are on the surface of the condensed chromosomes and have been selectively denatured can react with the antibodies (*34*).

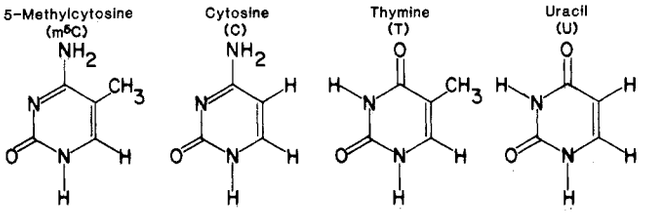

Fig. 1. Common pyrimidine bases. C and T are major bases of DNA and m^5C is a minor base of DNA. In RNA, U replaces T as a major base and m^5C is also found as one of many minor bases.

Restriction endonuclease fragments. Patterns of methylation of eukaryotic DNA have been elucidated by the use of several restriction endonucleases (*8, 35*). Information about the average distribution of m^5C residues throughout a genome or the specific location of m^5C residues within or around a specific gene can be obtained with these sequence-specific enzymes. Digestion of DNA with type II restriction endonucleases (*8*) produces fragments whose length reflects the distance between repeats of the specific cleavage sequence recognized by the enzyme. For most of these enzymes, methylation of a specific C or A residue in the DNA recognition site prevents hydrolysis at the methylated sequence (*8, 35*). The restriction endonucleases used for studies of eukaryotic DNA methylation are usually those that cleave DNA within a recognition sequence containing a CG dinucleotide (Table 1). Because CG is the main methylated dinucleotide in most examined eukaryotic DNA's (as indicated above), determination of the extent of methylation of CG-containing sequences is usually a good indicator of overall methylation of the DNA of higher organisms. However, these studies do not probe the CG sites that occur in sequences not recognized by available restriction endonucleases. Also, they generally do not address the usually minor fraction of m^5C residues in eukaryotic DNA, which do not occur in the 5′ side of G residues (*18, 24, 27–30, 36*) and which might have critical biological significance.

Table 1. DNA recognition sites of some CG-specific restriction endonucleases (*35*). All these enzymes except Msp I are inhibited when m^5CG replaces CG at these sites. Methylation of only one strand of the DNA is generally sufficient to protect against hyrolysis by restriction endonucleases (*8, 35*) although usually the CG sequence of the complementary strand is also methylated (symmetrical methylation). These sequences are all pallindromes, that is, because of 180° rotational symmetry, each sequence is the same in the 5′ to 3′ direction of one strand and in the 5′ to 3′ direction of the complementary strand. The sequence of only one strand (5′ to 3′, left to right) of the double-stranded recognition site is shown.

Endonuclease	Recognition sequence
Hpa II or Hap II	CCGG
Msp I	CCGG or Cm^5CGG*
Hha I	GCGC
Hae II	$^{G}_{A}GCGC^{C}_{T}$
Aos II	$G^{G}_{A}CG^{C}_{T}C$
Ava I	$C^{C}_{T}CG^{G}_{A}G$
Xho I	CTCGAG
Sma I	CCCGGG
Sac II	CCGCGG
Sal I	GTCGAC
Xor II	CGATCG
Sau IIIa†	GATC

*But not m^5CCGG. †Although Sau IIIa does not contain a CG sequence in its recognition site, it is inhibited when its recognition sequence is part of a larger m^5CG-containing sequence of the type $GATm^5CG$ (*37*).

This type of analysis of eukaryotic DNA methylation has been expanded to include restriction endonucleases that are inhibited by methylation of C residues in their recognition sequence, but lack a CG dinucleotide in that sequence (*18, 29*). An endonuclease with such specificity is Sau IIIa, which recognizes the sequence GATC (A, adenine; T, thymine) (*35*). Methylation of the C residues in only one strand of this sequence is sufficient to confer resistance to double-strand cleavage by this enzyme (*37*). Therefore, Sau IIIa has been used as a probe for C methylation of the GGATCG sequence in a bovine satellite DNA isolated with or without cloning in *Escherichia coli* (*37*). However, this enzyme makes the half-methylated recognition site

$$\begin{array}{c} m^5 \\ GATC \\ CTAG \end{array}$$

in one strand rather than not cleaving this sequence at all.

Figure 2 shows fragment patterns obtained on gel electrophoresis of samples of unfractionated human DNA, *Physarum polycephalum* (slime mold) DNA, and *E. coli* DNA digested with one of several CG-specific restriction endonucleases (*38*). Those enzymes with CG (and not m^5CG) in their DNA recogni-

tion sequence (Hpa II, Hha I, and Xor II) (Table 1) fragmented the bacterial DNA extensively, the human DNA only minimally, and the slime mold DNA to an intermediate extent. In contrast, Hind III, which recognizes the sequence AAGCTT (*35*), and Msp I, which recognizes the sequences CCGG and Cm^5C-GG (*39*), extensively fragmented all the above-mentioned DNA samples. Results similar to those from human DNA have been obtained with other mammalian DNA's digested with Hpa II and Msp I (*39–41*). It can be concluded from the Hpa II and Msp I digests that the sequence Cm^5CGG is generally much more frequent than CCGG in mammalian DNA's. In *E. coli* DNA, Cm^5CGG sequences are rare, and in *Physarum* DNA, Cm^5CGG sequences as well as CCGG sequences are plentiful. The results with Hha I and Xor II (Fig. 2) indicate that a similar pattern of methylation probably also applies to GCGC and CGATCG sequences throughout these genomes.

In order to probe the methylation pattern of a given portion of a eukaryotic chromosome, the total genome can be digested with a restriction endonuclease; then the DNA fragments can be resolved by electrophoresis, denatured, and directly transferred from the gel to nitrocellulose paper (*42*). Subsequent hybridization to radioactively labeled, cloned, or fractionated DNA or purified RNA followed by autoradiography allows the visualization of DNA fragments only from specific regions of the genome.

Restriction endonuclease analyses has shown that the methylated form of CG sequences predominates in vertebrate DNA, and most of the unusual nonmethylated CG sequences are interspersed among the m^5CG-containing sequences (*7, 43, 44*). In contrast, sea urchin (*Echinus esculentus*) DNA can be separated into a highly methylated fraction (approximately 40 percent of the genome) and an unmethylated fraction (approximately 60 percent of the genome) (*45*). Most of the nonrepetitive sequences are predominantly either in the unmethylated or in the methylated fraction rather than in both. Fractionation was achieved by digestion of the DNA with Hha I or Hpa II (Table 1), which yields separable low [< 4 kilobase pairs (kb)] and high (> 15 kb) molecular weight components. More than 90 percent of the total m^5C in these digests is present in the high molecular weight (Hpa II- and Hha I-resistant, and Msp I-sensitive) fraction. It therefore appears that methylation of CG sequences in this sea urchin DNA occurs mostly in specific, long tracts of the genome. The DNA from a slime mold (Fig. 2) and various nonarthropod invertebrates including several echinoderms other than *Echinus*, a coelenterate, and a mollusk (*7*) also have been shown to contain a major, unmethylated component as well as a large methylated component, as defined by sensitivity to Hpa II, Hha I, and Msp I.

Ribosomal RNA genes. The ribosomal RNA–encoding genes (termed rDNA) from mammals, birds, and at least one reptile as well as those from the above-mentioned lower eukaryotes are predominantly unmethylated at CCGG sites (*7, 46*). The lack of methylation of CCGG sequences in these vertebrate rDNA's contrasts with the scarcity of unmethylated CCGG sequences elsewhere in these genomes. Unlike the above types of rDNA, the rDNA from amphibia, a fish, a dinoflagellate, and higher plants, like most of the genome, is extensively methylated at CCGG sites (*7, 14, 29*).

Methylaton of the rDNA of the toad *Xenopus laevis* has been studied in some detail. Although the reticulocyte rDNA of *X. laevis* is almost completely methylated at most of the transcribed CG-containing sequences recognized by Hha I, Hpa II, and Ava I (Table 1), there is a Hha I sequence in the 28*S* coding region which is methylated in only 40 to 70 percent of the somatic rDNA repeats (*47*). In addition, there are two clusters of undermethylated sequences that are recognized by Hpa II and Ava I in the nontranscribed spacer. It is not yet clear what determines which Hha I, Hpa II, and Ava I sites in this region are unmethylated or why only some copies of these sequences are unmethylated. All the CG-containing sequences of the rDNA can be obtained in their unmethylated form by cloning the DNA in *E. coli*. Also, these rDNA sequences are present in an unmethylated, transcriptionally active state as extrachromosomal DNA in oocytes (*48*). These extrachromosomal sequences may be unmethylated because they have been produced by amplification of the chromosomal rDNA.

DNA sequencing. Analogous methylated and unmethylated DNA sequences can be distinguished by the DNA sequencing technique of Maxam and Gilbert (*49*). If both strands of the DNA or if

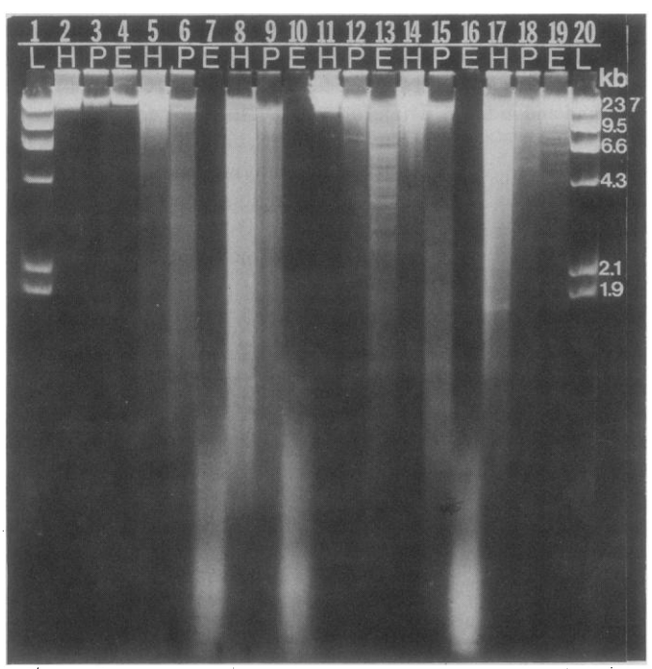

Fig. 2. Agarose gel electrophoresis of DNA digested with restriction endonucleases. DNA (2 μg) from human lymphocytes, *Physarum polycephalum* (slime mold), or *Escherichia coli* B was digested with 8 units of one of five type 2 restriction endonucleases for 14 hours at 37°C in appropriate buffers (*35*), subjected to electrophoresis, and visualized by ethidium bromide–induced fluorescence (*38*). The type of DNA used for the digest in each channel is abbreviated as follows: human, H; *Physarum*, P; *E. coli*, E; and bacteriophage λ, L (for molecular weight markers). The enzymes used in each channel were: channels 1 and 20, Hind III; channels 2 to 4, no enzyme control; channels 5 to 7, Hpa II; channels 8 to 10, Msp I; channels 11 to 13, Xor II; channels 14 to 16, Hha I; channels 17 to 19, Hind III. With the exception of Hind III, which recognizes the sequence AAGCTT, all the enzymes used have CG in their DNA recognition sequences (Table 1). All the enzymes extensively digested *E. coli* DNA, and those enzymes with four base-pair recognition sequences gave fragments with low molecular weights at the bottom of the gel, whereas those with the less frequent, six base-pair recognition sequences gave fragments of higher molecular weight near the top of the gel. Many distinct high molecular weight fragment bands can be seen in the digests of the bacterial DNA but, in the digests of the eukaryotic DNA's, fragments appear mostly as smears because of the much greater heterogeneity of DNA sequences in eukaryotic genomes than in prokaryotic genomes. Control experiments in which human DNA and λ DNA or *Physarum* DNA and λ DNA were co-digested showed that the relative resistance of human DNA and *Physarum* DNA to digestion by Xor II is due to the lack of unmodified recognition sites and not to the presence of inhibitors.

the homologous nonmethylated and methylated DNA's are sequenced, methylation of C residues at given sites in the DNA is evidenced by the absence in an electrophoregram of cleavage products terminated at the m^5C residues. In the primary sequences of uncloned *X. laevis* 5*S* DNA (*50*) and rat satellite I DNA (*51*), G was found to be the only 3′ neighbor of m^5C. All readily analyzed CG sequences in both strands of the *X. laevis* 5*S* DNA are at least partially methylated (*50*). Most of the CG sequences of the nuclear satellite DNA of *Scilla siberica*, a monocotyledonous plant, are also methylated; however, a considerable fraction of the m^5C residues in this satellite DNA have neighbors other than G on the 3′ side (*32*).

Methylation of DNA and Cellular Physiology

DNA replication. In several different mammalian and plant cell lines, methylation of DNA has been shown to be dependent on DNA synthesis (*52*). Methylation of newly synthesized DNA strands is essentially complete shortly after DNA replication in several types of cultured cells (*52*). However, there is evidence for completion of methylation of nascent DNA long after DNA synthesis (*53*). The timing of DNA methylation in relation to DNA synthesis could be critical in determining the loss of m^5C residues from previously methylated sequences. In either case, methylation probably generally involves newly synthesized (half-methylated) DNA, which is the best eukaryotic double-stranded substrate for animal DNA methylases in vitro (*54*).

Transcription. A number of studies suggest that at least for some genes in eukaryotes, methylation of certain DNA sequences is associated with inhibition of transcription. In thymidine kinase–negative cultured mouse cells genetically transformed with a cloned thymidine kinase gene (and thereby made thymidine kinase–positive), the donor gene was unmodified at CG sites (*55*). One of the thymidine kinase–negative derivatives of these transformed cells, which did not revert to a thymidine kinase–positive state, had highly methylated this donor gene and no longer expressed it. This result suggests that extensive methylation prevented transcription.

Less methylation is found in a number of transcribable eukaryotic genes and their adjacent sequences in tissues in which these genes are expressed. Restriction endonucleases were used to study methylation in the β-globin gene region in several different tissues of the chicken (*44*). This DNA region in oviduct, brain, and embryonic erythrocytes, all of which do not express this gene, was compared to the same region of adult erythrocyte DNA, in which these sequences are actively transcribed. As determined by resistance to Hpa II but not Msp I, some CCGG sequences (Table 1) in the adult β-globin gene region, are much more extensively methylated (at the internal C residue) in the DNA of cells not producing this β-globin protein than in adult erythrocyte DNA. In addition, there appear to be some CCGG sequences in this region which are invariably either extensively methylated or largely unmethylated in the DNA from all types of cells examined.

Ovalbumin, conalbumin, and ovomucoid gene regions in the chicken also have CG sequences that are either completely methylated or completely unmethylated in all studied tissues, whereas other such CG sequences are methylated in a tissue-specific pattern (*43*). The tissue-specificity of methylation in these genetic regions generally shows an inverse correlation between gene expression and methylation; however, the ovalbumin gene region has a CG site, more copies of which are methylated in oviduct DNA than in the DNA from other examined tissues.

Similarly, in the human γδβ-globin gene region, approximately half of the examined CG sequences are much less extensively methylated in tissues in which these sequences are expressed than in most other tissues in which these genes are essentially inactive (*36*). Also, two CCGG sites in fetal globin genes exhibit tissue-specific differences in their methylation at the 5′-terminal C (*36*), that is, in their resistance to cleavage by Msp I (Table 1). When DNA from adult tissues is used as the source of the fetal globin genes, these two sites become completely resistant to digestion by Msp I and, therefore, are present as m^5CCGG or m^5Cm^5CGG. In contrast, in fetal liver, which expresses fetal globin genes, these two sites are mostly unmethylated (CCGG). However, again the correlation between the lack of methylation at these sequences and gene expression is not complete because these sequences, as well as a number of other CG-containing sequences, are methylated only slightly or not at all in placental DNA and in DNA from two human cell lines. This undermethylation of a specific gene region may be related to our finding of a significantly lower m^5C content in human placental DNA than in the DNA of various adult human tissues (*9*). In those cases in which DNA methylation of specific sequences is lower in transcriptionally active genetic regions, there is as yet no clear pattern as to which portions of genes and their adjacent sequences are specifically undermethylated.

In the rabbit globin gene region, there are tissue-specific differences in the extent of methylation and these only partially correlate with the degree of expression of these genes (*56*). Also, the sea urchin genome has oocyte-specific and early embryo-specific histone genes with CCGG sites, most copies of which are unmethylated even in the DNA of tissues in which they are presumed to be inactive (*45*).

Viral infection and carcinogenesis. One class of DNA molecules that replicates in eukaryotic cells but may be completely or largely unmethylated is that of several types of viruses. Despite their host cell genomes containing demonstrable levels of m^5C, the DNA's of herpes saimiri virus (*57, 58*), adenoviruses (*58, 59*), polyoma virus (*60*), herpes simplex virus (*61*), and simian virus 40 (*62*), and exogenously acquired DNA copies of mouse mammary tumor virus (MMTV) RNA genome (*63*) and of an avian sarcoma RNA genome (*64*) have no detectable m^5C. The DNA of human papilloma virus type 1a has only one of four CCGG sites methylated and that sequence is methylated in only about 40 percent of the DNA molecules (*65*). In contrast to these animal viruses, many types of bacteriophage have m^5C or m^6A in their DNA genomes at levels comparable to that of their bacterial host chromosomes (*1*). Under certain conditions, DNA sequences from viruses infecting eukaryotic cells may be extensively methylated, for example, when viral episomes are formed or sometimes when viral DNA sequences are integrated into host chromosomes (*57, 59*). Extensive methylation of such integrated or episomal viral DNA is associated with relatively low levels of transcription from the viral genome or with production of little or no virus (*57, 59*). In mammary tumors, the exogenously acquired (by infection) MMTV DNA sequences are much less methylated than are the coexisting unexpressed endogenous MMTV DNA sequences. In the mouse tumor tissue, not only is there extensive undermethylation of the newly acquired MMTV DNA copies, but also the host DNA sequences appear to have more sites that are sensitive to Hpa II and approximately 10 percent less m^5C than does uninfected tissue DNA (*66*).

Less methylation (10 to 30 percent) has also been observed in cellular DNA from rat hepatomas that were induced by chemical carcinogens than in DNA from the even more rapidly dividing cells of regenerating liver (*20*). We have some indication that DNA from various malignant human tumors contains approximately 10 to 20 percent less m^5C than does DNA from normal adult tissues (*67*). A lower m^5C content in the DNA of cancer cells might be due to their cell renewal rate being higher than that of most normal tissues. However, in the study of rat hepatomas and normal or regenerating liver, most of the undermethylation of the tumor DNA could not be ascribed to differences in cell turnover (*20*). Holliday has proposed that decreased methylation of pivotal control sequences in DNA could be a common epigenetic change that triggers oncogenic transformation (*68*). If this were the case, detectable demethylation might occur in unfractionated DNA. It is also possible that oncogenesis might require demethylation of only a very small fraction of the approximately 10^8 m^5C residues per human cell (*9*).

In several oncogenically transformed hamster cell lines, the m^5C content of DNA has been found to be higher than in the nontransformed counterpart (*26, 59, 69*); however, the "nontransformed" cell line used was aneuploid and had an unusually low m^5C content (*59*). Furthermore, increases or decreases in the m^5C content of cancerous cells in vitro or in vivo might be a consequence rather than a contributing cause of oncogenic transformation.

Establishment of Patterns of DNA Methylation

DNA methylation may play some common roles in all the cells of an organism as well as specific roles in only certain kinds of cells. Most of the DNA methylation pattern might be handed down from cell to cell by the conversion of a half-methylated site

$$\begin{array}{c} {}^{m5} \\ 5'\text{-CG-}3' \\ 3'\text{-GC-}5' \end{array}$$

to a symmetrically methylated site

$$\begin{array}{c} {}^{m5} \\ 5'\text{-CG-}3' \\ 3'\text{-GC-}5' \\ {}^{m5} \end{array}$$

by a maintenance-type DNA methylase after DNA replication (*68*). Furthermore, methylation of C residues in sequences other than CG may also generally involve oligonucleotide sequences with 180° rotational symmetry, such as GC sequences (*32*) or the sequence CCGG at the 5'-terminal C (*36*). That most examined copies of CG dinucleotide sequences in various eukaryotic DNA's (*40, 47, 50, 51*) are either symmetrically methylated or symmetrically unmethylated is consistent with the existence of a maintenance-type DNA methylase. Because of the great variety of bases neighboring both methylated and unmethylated CG sequences in DNA (*28, 50, 51*), there are as yet no discernible rules for predicting which CG sequences will be unmethylated.

Tissue-specific differences in DNA methylation could arise from an initiation-type DNA methylase introducing methyl groups into symmetrically unmethylated sites during a developmental program. There is evidence that such de novo methylation occurs but only occasionally (*18, 55, 57, 59, 70*). A more common source of tissue-specific variation in the distribution of m^5C residues may be the replacement of m^5C residues with C residues during differentiation. In a number of vertebrate gene regions, sperm DNA sequences were shown to be much more highly methylated than the analogous sequences in somatic tissues (*36, 43*). The total m^5C content of rabbit sperm DNA is approximately 40 percent higher than that of rabbit liver (*71*). On the other hand, human sperm DNA has a significantly lower m^5C content, as much as 17 percent lower than does DNA from various somatic human tissues (*9*). Lower levels of m^5C in sperm DNA from humans (*9*) and certain other animals (*6*) might be due to methylation of only one strand at some of the m^5C-containing sequences (*72*) in contrast to the symmetrical methylation mentioned above in the DNA of cells other than sperm. In any event, the finding that human sperm DNA has significantly fewer of its C residues methylated than does DNA from a variety of adult human tissues that were analyzed (*9*) implies that at some stage (or stages) in human development there are substantial increases in the extent of DNA methylation.

Holliday has outlined several schemes whereby DNA repair in an excision or recombination mode could lead to symmetrical losses of m^5C (*68*). DNA-damaging carcinogens could trigger DNA repair without restoration of the original levels of DNA methylation and thereby lead to transcriptional activation of genes responsible for oncogenesis (*68*). It is also possible that sequence-specific endonucleases could be programmed to initiate repair of certain DNA sequences and, in the absence of adequate DNA methylation, cause demethylation of just those sites in the genome.

Alternatively, when macromolecules involved in transcription or transcriptional control bind to specific chromatin regions, methylation of DNA in those regions may be depressed after replicative or repair-type DNA synthesis. Given the findings that two types of "high mobility group" chromatin proteins preferentially interact with actively transcribed genes in chromatin (*73*) and preferentially associate with nucleosomes containing DNA with a low m^5C content (*74*), it can be hypothesized that these or other transcription-related DNA binding proteins locally inhibit DNA methylation. Such a hypothesis would imply that gene activation causes undermethylation of DNA sequences in the vicinity. Conversely, DNA sequences may become less methylated as part of the preparation for transcriptional activation.

Evaluation of the above hypotheses for demethylation depends on clarifying the efficiency of methylation of DNA after replication or various types of repair. Also needed, is an unambiguous determination of how long newly synthesized DNA strands remain unmethylated after DNA synthesis in eukaryotic cells. Excision repair of the methylated strand of newly replicated, half-methylated DNA might lead to permanent loss of methylation from a given site. Therefore, the longer that newly replicated DNA remains half-methylated, the greater is the likelihood of the loss of maintenance methylation.

Some Possible Functions of Eukaryotic DNA Methylation

Modification-restriction. Sager and Kitchin (*75*) hypothesized that the selective inactivation or degradation of paternal chromosomes of kangaroos, coccid insects, and *Chlamydomonas* is triggered by a lack of DNA modification at certain sequences, in analogy to bacterial modification-restriction systems (*8*). Determination of chloroplast DNA base composition and of sensitivity to several restriction endonucleases and analysis of a mutant with increased DNA methylation (*18*) indicate that methylation of cytosine residues, especially in CG sequences, in the chloroplast DNA of female *Chlamydomonas* gametes protects this DNA against degradation. The mostly unmethylated chloroplast DNA of the male gamete is digested after fertilization. In contrast, the female-derived chloroplast DNA persists after fertilization and becomes even more extensively

methylated (*18*). Although restriction of certain incompletely modified or unmodified DNA's may occur in exceptional eukaryotes other than *Chlamydomonas*, unmodified CG sequences can be normal undegraded components of eukaryotic DNA (*43, 44, 47*). Unmethylated CG sites might still be hypothesized to be commonly cleaved by endonucleases in higher eukaryotes but only to yield transient breaks as a normal part of DNA metabolism such as in DNA recombination.

Mutagenesis. In the *E. coli* gene that codes for the lactose repressor protein, m^5C residues are preferred sites for spontaneous mutations (*76*). A predisposition to mutagenesis at m^5C residues might be partially the result of a conversion of some m^5C residues in DNA to T (thymine) residues (Figs. 1 and 3) on deamination. Such deamination was observed after treatment with alkali, heat at neutral *p*H, or sodium bisulfite (*77*). These studies were done with phage XP-12 DNA (m^5C, 34 mole percent; C, less than 1 mole percent) and a bacterial DNA, *Xanthomonas oryzae* DNA (C, 32 mole percent; m^5C, 0.09 mole percent), as models. It was shown that, in a physiological buffer at 95°C, m^5C residues in DNA are deaminated approximately three times faster than are C residues under the same conditions (*77*). Furthermore, when C residues undergo deamination, the resulting U residues should be relatively efficiently replaced by C residues through the ubiquitous uracil excision DNA repair system (*78*). In contrast, mismatch repair systems (*79*) should be equally likely to eliminate either the T residues arising from deamination of m^5C residues or the nonmutated G residues at the same positions in the complementary strand (Fig. 3) unless these repair systems have a bias (*79*) such as preferential excision of pyrimidine residues.

The hypothesis that m^5C residues are preferred sites for spontaneous mutation in DNA (*80*) could partially explain the underrepresentation of CG dinucleotide (methylated plus nonmethylated forms) in the total DNA and in the structural genes of higher eukaryotes (*81*), the preferential occurrence of m^5C in highly repetitive (noncoding) DNA (*9, 31*), and the high percentage of allelic variants seen at CG-containing sequences in the chicken ovalbumin gene (*43*). In vertebrate DNA, CG dinucleotides are only 20 to 40 percent as frequent as would be expected in a comparable DNA with a random distribution of bases (*82*). Because of the degeneracy of the genetic code, all CG-containing codons can be replaced by non-CG codons that code for the same amino acid. The unexpectedly low frequency of CG in DNA is correlated with a higher than predicted frequency of CA plus TG dinucleotides as well as with a relatively high percentage of C methylation (*82*). These relationships further support the hypothesis that m^5C residues have an unusually high probability of undergoing mutation to T residues. Such a proclivity for loss of m^5C by mutation provides a basis for arguing that the methylation of C residues in DNA is important to cells or else it would have been eliminated by evolution.

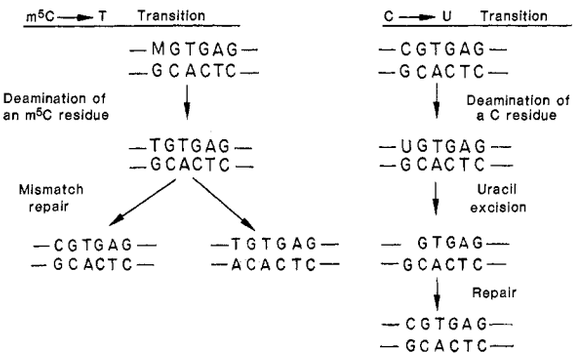

Fig. 3. A simplified scheme showing the deamination and subsequent repair of an m^5C residue or a C residue in a given sequence in DNA. Such deamination might occur spontaneously in vivo at biologically significant rates (*76, 77*). If mismatch repair removes the T residue created by deamination of an m^5C residue, a potential mutation can be reversed. On the other hand, if the nonmutated G residue of the T·G mispair is excised, then the mutation would be fixed and passed on to both progeny cells. M represents m^5C.

DNA repair. At least in bacteria, efficient methods exist for replacing a mismatched base pair created by incorporation of the wrong residues during DNA synthesis (*79*). Considerable evidence suggests that methylation of A residues, but not of C residues, in *E. coli* DNA is involved in directing this repair of mismatched bases in newly synthesized DNA sequences so that the nascent, unmethylated portion of the chromosome is preferentially repaired (*83*). Such a system for repair of misincorporated residues in DNA might also exist in the cells of higher eukaryotes where m^5C residues rather than m^6A residues might be used to asymmetrically direct repair of misincorporated bases.

Chromosome structure. The nonrandom distribution of m^5C residues along the surface of metaphase chromosomes (*34*), and throughout chromatin (*84*) and the relative abundance of m^5C in several satellite DNA's (*33*) and in highly repetitive DNA fractions (*9, 31*) suggest that enrichment of certain sequences in m^5C residues may help determine the higher order structure of chromosomes. A recent immunological study suggests that m^5C residues are specifically associated with polytenization of Diptera chromosomes (*85*). Methylation of DNA might also considerably alter the local conformation of DNA since there is evidence (*86*) that certain base sequences influence the local structure of the double helix.

Changes in the conformation of the double helix in certain genetic regions could modulate DNA synthesis or transcription. Although the effects of 5-methylation of cytosine residues on the conformation of naturally occurring DNA are not yet known, data have recently been provided which suggest that regions of DNA that contain repetitive CG sequences might assume a specific left-handed conformation of the double helix (Z-DNA) (*86*). Since CG sequences are the main site for C methylation in eukaryotic DNA, methylation of CG sequences might be involved in altering or stabilizing regions of Z-DNA. A recent study suggests that methylation of C residues in a synthetic double-stranded polydeoxynucleotide predisposes that polymer to assuming the Z form under more physiological conditions (*87*). Localized Z-DNA structure might make m^5C residues in DNA even more readily recognized by certain DNA binding proteins. If, indeed, the sequence of DNA can affect its local helical conformation in vivo, the ability of specific DNA sequences containing m^5C residues to serve as landmarks for events involved in RNA synthesis, DNA replication, or DNA recombination would be enhanced. Even in the common B conformation of the double helix, the methyl group of m^5C is expected to be highly exposed in the major groove (*88*) and so could easily be recognized by specific eukaryotic proteins.

Methylation of DNA might also influence DNA template activity by affecting the ability of DNA to be denatured. Complete replacement of C residues in double-stranded DNA with m^5C residues markedly increases the melting tempera-

ture (T_m) of the DNA (*3*), and an increase in the T_m can even be seen on partial methylation of C residues (*48*). A higher T_m implies that separation of the two DNA strands is more difficult to achieve. Since DNA replication and transcription both require partial strand separation, the presence of m^5C residues, especially in clusters or in regions of low A·T content, could decrease the rates of DNA and RNA synthesis from m^5C-containing DNA sequences. Such decreases might have important consequences, since localized deceleration of RNA synthesis has been postulated to contribute to chain termination (*89*).

Furthermore, transcription initiation sites for *E. coli* RNA polymerases are preferentially found in the more easily denatured regions of DNA (*90*). Therefore, methylation of C residues might inhibit the initiation of DNA or RNA synthesis (*91*) in their vicinity by increasing helix stability.

Methylation of DNA could also affect DNA or RNA synthesis by influencing interactions with sequence-specific proteins. Bacterial restriction endonucleases provide models of proteins that can readily distinguish between two DNA sequences differing only in the extent of methylation of C (or A) residues (*8, 35*). Also, the interaction of the *E. coli* lactose repressor protein with *E. coli* lactose operator DNA is markedly altered by 5-methylation of a single C residue (*92*). In order to be expressed, certain eukaryote genes might need to have associated regulatory DNA sequences demethylated so that these sequences can bind specific proteins that promote transcription or not bind proteins that repress transcription.

Differentiation. Since DNA methylation might negatively control transcription of some genes, methylation of specific sequences might be necessary, although it is probably not sufficient, for committing certain cells to differentiation (*93*). Evidence for tissue-specific differences in gene methylation, inversely correlated with genetic activity was described earlier. In addition to these large differences from tissue to tissue in methylation of certain "luxury function" genes (*36, 43, 44*), differences of up to 28 percent were seen in the overall m^5C content of various human tissues (*9*). Variations in the extent of DNA methylation were also seen in the DNA of different established cell lines (*9, 94*). However, these differences might not be reflecting the state of differentiation of the cells.

Several studies of cultured cells offer evidence for a relation between decreased DNA methylation and induction of differentiation. Treatment of murine erythroleukemia cells with dimethyl sulfoxide, ethionine, or other chemicals that induce globin synthesis in these cells causes a decrease in DNA methylation (*95*). Inhibitors of differentiation prevent the decrease in methylation. Cells resistant to induction of differentiation by dimethyl sulfoxide (despite uptake of the chemical) also show no such change in methylation when treated with it. The dimethyl sulfoxide–induced decrease in m^5C content seen in sensitive cells involves a loss of less than 2 percent of the normal number of m^5C residues and was detected as a severalfold increase in the acceptance of isotopically labeled methyl groups during incubation with a crude preparation of mouse DNA methylase.

The differentiation of cultured mouse embryo cells into muscle cells also can be induced by inhibitors of DNA methylation (*96*). In this case, muscle cell formation occurs 8 to 11 cell divisions after exposure to various analogs of cytidine (*96*). Cytidine analogs that do not inhibit DNA methylation fail to cause the conversion to muscle cells. Although effects of these chemicals other than perturbation of DNA methylation may be responsible for their inducing activity, these studies combined with others described above suggest that m^5C residues in specific DNA sequences have to be replaced by C residues in order to obtain expression of certain genes that play a role in differentiation or in viral expression. Alternatively, changes in DNA methylation may be a result of other alterations in chromatin that are associated with transcriptional activity or activation.

Conclusion

There is no direct evidence for any single common function of most of the methylation of eukaryotic DNA. Methylation probably serves multiple purposes, which may include regulating the transcription of certain genes; locally enhancing the rate of mutation; increasing the stability of regions of the double helix; altering the conformation of DNA, chromatin, or chromosomes; and, in exceptional eukaryotic cells, preventing nucleolytic degradation. Given the widespread occurrence of m^5C in specific sequences of eukaryotic DNA and the fact that methylation of C residues is the only known genetically programmed, covalent modification of the DNA of higher eukaryotes, the presence of this minor base is likely to be of major importance in the functioning of DNA.

References and Notes

1. R. H. Hall, *The Modified Nucleosides in Nucleic Acids* (Columbia Univ. Press, New York, 1970), pp. 281–294; H. S. Shapiro, in *Handbook of Biochemistry*, H. A. Sober, Ed. (CRC Press, Cleveland, ed. 2, 1970), pp. H-89–H-103.
2. T.-Y. Feng, J. Tu, T.-T. Kuo, *Eur. J. Biochem.* **87**, 29 (1978).
3. M. Ehrlich, K. Ehrlich, J. A. Mayo, *Biochim. Biophys. Acta* **395**, 109 (1975).
4. D. Simon, F. Grunert, U. Achen, H. P. Doring, H. Kroger, *Nucleic Acids Res.* **5**, 2153 (1978).
5. C. K. Mathews, in *Comprehensive Virology*, H. Fraenkel-Conrat and R. W. Wagner, Eds. (Plenum, New York, 1977), vol. 7, p. 179; R. A. J. Warren, *Annu. Rev. Microbiol.* **34**, 137 (1980).
6. B. F. Vanyushin, A. L. Mazin, V. K. Vasilyev, A. N. Belozersky, *Biochim. Biophys. Acta* **299**, 397 (1973); B. F. Vanyushin, S. G. Tkacheva, A. N. Belozersky, *Nature (London)* **225**, 948 (1970).
7. A. P. Bird and M. H. Taggert, *Nucleic Acids Res.* **8**, 1485 (1980).
8. W. Arber, *Science* **205**, 361 (1979); H. O. Smith, *ibid.*, p. 455.
9. L.-H. Huang, M. A. Gama-Sosa, C. W. Gehrke, K. C. Kuo, M. Ehrlich, in preparation.
10. M. Bayen and J. Dalmon, *Biochim. Biophys. Acta* **432**, 273 (1976).
11. S. Hattman, C. Kenny, L. Berger, K. Pratt, *J. Bacteriol.* **135**, 1156 (1978).
12. J. M. Pollock, M. Swihart, J. H. Taylor, *Nucleic Acids Res.* **5**, 4855 (1978).
13. H. H. Evans, T. E. Evans, S. Littman, *J. Mol. Biol.* **74**, 563 (1973).
14. P. M. M. Rae and R. E. Steele, *Biosystems* **10**, 37 (1978); *Nucleic Acids Res.* **8**, 4709 (1980).
15. R. L. P. Adams, E. L. McKay, L. M. Craig, R. H. Burdon, *Biochim. Biophys. Acta* **563**, 72 (1979).
16. M. M. K. Nass, *J. Mol. Biol.* **80**, 155 (1973); D. S. Ray and P. C. Hanawalt, *ibid.* **9**, 812 (1964); K. K. Tewari and S. G. Wildman, *Science* **153**, 1269 (1966); I. B. Dawid, *ibid.* **184**, 80 (1974); G. S. P. Groot and A. M. Kroon, *Biochim. Biophys. Acta* **564**, 355 (1979); D. J. Cummings, A. Tait, J. M. Goddard, *ibid.* **374**, 1 (1974).
17. H. H. Evans and T. E. Evans, *J. Biol. Chem.* **245**, 6436 (1970); B. F. Vanyushin and M. D. Kirnos, *Mol. Cell. Biochem.* **14**, 31 (1977).
18. W. G. Burton, C. T. Grabowy, R. Sager, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1390 (1979); H-D. Royer and R. Sager, *ibid.*, p. 5794; R. Sager, C. Grabowy, H. Sano, *Cell* **24**, 41 (1981).
19. J. Deutsch, A. Razin, J. Sedat, *Anal. Biochem.* **72**, 586 (1976); J. Singer, R. H. Stellwagen, J. Roberts-Ems, A. D. Riggs, *J. Biol. Chem.* **252**, 5509 (1977); H. Sano, H.-D. Royer, R. Sager, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3581 (1980); M. Ehrlich and K. Ehrlich, *J. Chromatog. Sci.* **17**, 531 (1979); K. C. Kuo, R. A. McCune, C. W. Gehrke, R. Midgett, M. Ehrlich, *Nucleic Acids Res.* **8**, 4763 (1980).
20. J. N. Lapeyre and F. F. Becker, *Biochem. Biophys. Res. Commun.* **87**, 698 (1979).
21. M. Laskowski, Sr., in *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, ed. 3, 1971), vol. 4, p. 289.
22. P. H. Roy and A. Weissbach, *Nucleic Acids Res.* **2**, 1669 (1975).
23. P. Grippo, M. Iaccarino, E. Parisi, E. Scarano, *J. Mol. Biol.* **36**, 195 (1968).
24. R. L. Sinsheimer, *J. Biol. Chem.* **208**, 445 (1954); *ibid.* **215**, 579 (1955).
25. K. Burton, *Methods Enzymol.* **12A**, 222 (1967).
26. M. J. Browne and R. H. Burdon, *Nucleic Acids Res.* **4**, 1025 (1977).
27. J. Doskocil and F. Sorm, *Biochim. Biophys. Acta* **55**, 953 (1962).
28. R. Salomon and A. M. Kaye, *ibid.* **204**, 340 (1970); T. W. Sneider, *J. Biol. Chem.* **247**, 2872 (1972).
29. W. L. Gerlach and J. R. Bedbrook, *Nucleic Acids Res.* **7**, 1869 (1979).
30. In the DNA of some lower eukaryotes and of some types of cultural cells, m^5C residues may not occur predominantly in m^5CG sequences, for example, in *C. reinhardi* (*18*) and possibly in insects (*15*) and Novikoff rat hepatoma cells [T. W. Sneider, *Nucleic Acids Res.* **8**, 3829 (1980)].
31. D. Drahovsky, T. L. J. Boehm, W. Kreis, *Biochim. Biophys. Acta* **563**, 28 (1979); V. A. Guseinov and B. F. Vanyushin, *ibid.* **395**, 229 (1975).
32. B. Deumling, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 338 (1981).
33. R. Salomon, A. M. Kaye, M. Hertzberg, *J. Mol. Biol.* **43**, 581 (1969); R. Gantt, F. Montes De Oca, V. J. Evans, *In Vitro* **8**, 288 (1973); F. Hatch and J. Mazrimas, *Nucleic Acids Res.* **1**, 559 (1974).
34. R. R. Schreck, B. F. Erlanger, O. J. Miller, *Exp. Cell Res.* **108**, 403 (1977); B. W. Lubit, T. D.

Pham, O. J. Miller, B. F. Erlanger, *Cell* **9**, 503 (1976); W. Schnedl, B. F. Erlanger, O. J. Miller, *Hum. Genet.* **31**, 21 (1976).
35. R. J. Roberts, *Nucleic Acids Res.* **9**, r75 (1981).
36. L. H. T. van der Ploeg and R. A. Flavell, *Cell* **19**, 947 (1980).
37. R. E. Streeck, *Gene* **12**, 267 (1980).
38. R. Y.-H. Wang, J. G. Shedlarski, M. B. Farber, D. Kuebbing, M. Ehrlich, *Biochim. Biophys. Acta* **606**, 371 (1980); J. R. Jeter, Jr., M. A. Gama-Sosa, C. W. Gehrke, K. C. Kuo, R. Y.-H. Wang, M. Ehrlich, in preparation. For digestion of *P. polycepalum* DNA, also see P. A. Whittaker, A. McLachlin, N. Hardman, *Nucleic Acid Res.* **9**, 801 (1981).
39. C. Waalwijk and R. A. Flavell, *Nucleic Acids Res.* **5**, 3231 (1978).
40. H. Cedar, A. Solage, G. Glaser, A. Razin, *ibid.* **6**, 2125 (1979).
41. J. Singer, J. Roberts-Ems, A. D. Riggs, *Science* **203**, 1019 (1979).
42. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
43. J. L. Mandel and P. Chambon, *Nucleic Acid Res.* **7**, 2081 (1979); M. T. Kuo, J. L. Mandel, P. Chambon, *ibid.*, p. 2105.
44. J. D. McGhee and G. D. Ginder, *Nature (London)* **280**, 419 (1979).
45. A. P. Bird, M. H. Taggart, B. A. Smith, *Cell* **17**, 889 (1979).
46. J. G. Reilly, R. Braun, C. A. Thomas, *FEBS Lett.* **116**, 181 (1980).
47. A. P. Bird and E. M. Southern, *J. Mol. Biol.* **118**, 27 (1978); A. P. Bird, *ibid.*, p. 49. See P. Boseley, T. Moss, M. Machler, R. Portmann, M. Birnstiel [*Cell* **17**, 19 (1979)] for an analysis of the sequence of part of this region.
48. I. B. Dawid, D. D. Brown, R. H. Reeder, *J. Mol. Biol.* **51**, 341 (1970).
49. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977); H. Ohmori, J. Tomizawa, A. M. Maxam, *Nucleic Acids Res.* **5**, 1479 (1978).
50. J. R. Miller, E. M. Cartwright, G. G. Brownlee, N. V. Fedoroff, D. D. Brown, *Cell* **13**, 717 (1978).
51. M. Pech, T. Igo-Kemens, H. G. Zachau, *Nucleic Acids Res.* **7**, 417 (1979).
52. J. W. Kappler, *J. Cell. Physiol.* **75**, 21 (1970); Y. Hotta and N. Hecht, *Biochim. Biophys. Acta* **238**, 50 (1971); D. Drahovsky and A. Wacker, *Naturwissenschaften* **62**, 189 (1975); M. H. Schneiderman and D. Billen, *Biochim. Biophys. Acta* **308**, 352 (1973).
53. J. H. Taylor, *Molecular Genetics*, part 3, *Chromosome Structure* (Academic Press, New York, 1979), pp. 89–115; B. Bugler, O. Bertraux, R. Valencia, *J. Cell. Physiol.* **103**, 149 (1980); but several of the studies discussed in these articles involved inhibitors of DNA synthesis, which might disturb the normal time course of DNA methylation.
54. T. W. Sneider, W. M. Teague, L. M. Rogachevsley, *Nucleic Acids Res.* **2**, 1685 (1975); R. L. P. Adams, E. L. McKay, L. M. Craig, R. H. Burdon, *Biochim. Biophys. Acta* **561**, 345 (1979).
55. A. Pellicer, D. Robins, B. Wold, R. Sweet, J. Jackson, I. Lowy, J. M. Roberts, G. K. Sim, S. Silverstein, R. Axel, *Science* **209**, 1414 (1980); also, see M. Wigler, D. Levy, M. Perucho, *Cell* **24**, 33 (1981).
56. C. Waalwijk and R. A. Flavell, *Nucleic Acids Res.* **5**, 4631 (1978); C. K. J. Shen and T. Maniatis, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6634 (1980).
57. R. C. Desrosiers, C. Mulder, B. Fleckenstein, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3839 (1979).
58. U. v. Acken, D. Simon, F. Grunert, H.-P. Doring, H. Kroger, *Virology* **99**, 52 (1979).
59. L. Vardimon, R. Neumann, I. Kuhlman, D. Sutter, W. Doerfler, *Nucleic Acids Res.* **8**, 2461 (1980); U. Gunthert, M. Schweiger, M. Stupp, W. Doerfler, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3923 (1976). The m^5C content of DNA from adenoviruses types 2 and 12 is < 0.02 mole percent.
60. A. M. Kaye and E. Winocour, *J. Mol. Biol.* **24**, 75 (1967).
61. M. Low, J. Hay, H. M. Keir, *ibid.* **46**, 205 (1969).
62. J. P. Ford, M. Coca-Prados, M.-T. Hsu, *J. Biol. Chem.* **255**, 7544 (1980).
63. J. C. Cohen, *Cell* **19**, 653 (1980).
64. R. V. Guntaka, P. Y. Rao, S. A. Mitsialis, R. Katz, *J. Virol.* **34**, 569 (1980).
65. D. Danos, M. Katinka, M. Yaniv, *Eur. J. Biochem.* **109**, 457 (1980).
66. J. C. Cohen, T. Breznik, C. W. Gehrke, M. Ehrlich, *ICN-UCLA Symp. Mol. Cell. Biol.* **18**, 401 (1980).
67. M. Gama-Sosa, C. W. Gehrke, K. C. Kuo, M. Ehrlich, unpublished results.
68. R. Holliday, *Br. J. Cancer* **40**, 513 (1979).
69. E. D. Rubery and A. A. Newton, *Biochim. Biophys. Acta* **324**, 24 (1973).
70. D. Drahovsky, S. Kaul, T. L. J. Boehm, A. Wacker, *ibid.* **607**, 201 (1980)
71. T. W. Sneider, J. Pondysh, C. W. Gehrke, M. A. Gama-Sosa, M. Ehrlich, unpublished results.
72. J. Kaput and T. W. Sneider, *Nucleic Acid Res.* **7**, 2303 (1979).
73. S. Weisbrod, M. Groudine, H. Weintraub, *Cell* **19**, 289 (1980).
74. S. Weisbrod and H. Weintraub, *ibid.* **23**, 391 (1981).
75. R. Sager and R. Kitchin, *Science* **189**, 426 (1975).
76. C. Coulondre, J. H. Miller, P. J. Farabaugh, W. Gilbert, *Nature (London)* **274**, 775 (1978); B. K. Duncan and J. H. Miller, *ibid.* **287**, 560 (1980).
77. R. Y.-H. Wang, C. W. Gehrke, K. C. Kuo, M. Ehrlich, in preparation; R. Y.-H. Wang, C. W. Gehrke, M. Ehrlich, *Nucleic Acids Res.* **8**, 4777 (1980).
78. T. Lindahl, *Prog. Nucleic Acid Res. Mol. Biol.* **22**, 135 (1979).
79. C. M. Radding, *Annu. Rev. Biochem.* **47**, 847 (1978); B. W. Glickman and M. Radman, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1063 (1980).
80. W. Salser, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 985 (1977).
81. G. J. Russell, P. M. B. Walker, R. A. Elton, J. H. Subak-Sharpe, *J. Mol. Biol.* **108**, 1 (1976).
82. A. P. Bird, *Nucleic Acids Res.* **8**, 1499 (1980).
83. M. G. Marinus and N. R. Morris, *J. Bacteriol.* **114**, 1143 (1973); M. Radman, G. Villiani, S. Boiteux, A. R. Kinsella, B. W. Glickman, S. Spadari, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 937 (1979).
84. A. Solage and H. Cedar, *Biochemistry* **17**, 2934 (1978).
85. E. A. Eastman, G. M. Goodman, B. F. Erlanger, O. J. Miller, *Chromasoma* **79**, 225 (1980).
86. A. H.-J. Wang, *Nature (London)* **282**, 680 (1979); S. Arnott, R. Chandrasekaran, D. L. Birdsall, A. G. W. Leslie, R. L. Ratliff, *ibid.* **283**, 743 (1980).
87. M. Behe and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1619 (1981).
88. N. C. Seeman, J. M. Rosenberg, A. Rich, *ibid.* **73**, 804 (1976); C. J. Alden and S. H. Kim, *J. Mol. Biol.* **132**, 411 (1979).
89. D. Pribnow, in *Biological Regulation and Development*, R. F. Goldberger, Ed. (Plenum, New York, 1979), vol. 1, p. 219.
90. H. J. Vollenweider, M. Fiandt, W. Szybalski, *Science* **205**, 508 (1979).
91. H. Sano and R. Sager, *Eur. J. Biochem.* **105**, 471 (1980).
92. E. F. Fisher and M. H. Caruthers, *Nucleic Acids Res.* **7**, 401 (1979).
93. For a review of arguments in favor of DNA methylation negatively controlling transcription during differentiation, see A. Razin and A. D. Riggs [*Science* **210**, 604 (1980)].
94. J. W. Kappler, *J. Cell. Physiol.* **78**, 33 (1971).
95. J. K. Christman, N. Weich, B. Schoenbrun, N. Schneiderman, G. Acs, *J. Cell Biol.* **86**, 366 (1980).
96. P. A. Jones and S. M. Taylor, *Cell.* **20**, 85 (1980).
97. Supported by NIH grants CA-19942 and GM-26986.

Allocating Petroleum Products During Oil Supply Disruptions

Roger H. Bezdek and William B. Taylor, Jr.

For the foreseeable future, the United States will be heavily dependent on imported oil. Political instability in several oil-exporting regions makes future disruptions possible and contingency plans must be made for such events (*1*). One consequence that must be planned for is that the demand for petroleum products at predisruption prices will exceed the available supply. Whether the market or the government allocates oil products, the United States will suffer severe costs from a significant curtailment of oil imports. In the event of a severe shortfall of long duration, government intervention of some sort may be required, and allocaton plans to moderate the effects of this shortfall must be evaluated. In this article we analyze four specific petroleum allocation options: (i) oil price and allocation controls, (ii) coupon gasoline rationing, (iii) a variable gasoline tax and rebate, and (iv) no oil price controls with partial rebates (*2, 3*).

Impacts of Oil Supply Disruptions

As in previous oil shortfalls, unless the government reimposes price controls on domestic crude oil, prices will rise with the world price, and this increase will be passed on to consumers in the form of higher prices for petroleum products. For large disruptions—an unprecedented 20 percent reduction in the U.S. supply of petroleum, for example—the size of such a price increase cannot be estimated precisely. However, based on reason-

Roger Bezdek is special assistant on energy in the Office of the Assistant Secretary for Economic Policy, Department of the Treasury, Washington, D.C. 20220. William Taylor is a member of the staff of Senator Bill Bradley, Dirksen Senate Office Building, Washington, D.C. 20510. This analysis was conducted while Mr. Taylor was with the U.S. Department of Energy.

0036-8075/81/0619-1357$01.50/0