tion of ammonium and alkaloid salts.

We called the complex salts with negatively charged central atom "ate" complexes for understandable reasons (*23*). They can be compared with the "onium" complexes, which were already known, as shown in Fig. 16. Because of the inductive effect of the central atom in onium complexes, all ligands R are cationically labilized and the hydrogen atoms at the neighboring carbon atoms are proton-mobile; however, in ate complexes all ligands at the central atom are anionically labilized and the hydrogen atoms at the neighboring carbon atoms are hydride-labile. This rule explains numerous reactions. I do not have time here to discuss its importance as a heuristic principle.

Thus I come to the end of my lecture. The excursion from diyls to ylides now ends at my idyll. With this I mean the conclusion of my research work as an emeritus, which allowed me to continue my work as a chemist free from the obligations of a teacher, and finally to devote myself completely to my interest in fine arts. I want to close my talk by offering cordial thanks to my collaborators. Without them my work could not have been accomplished.

References and Notes

1. Compare G. Wittig, *Acc. Chem. Res.* **7**, 6 (1974).
2. _____ and M. Leo, *Ber. Dtsch. Chem. Ges.* **61**, 854 (1928); *ibid.* **62**, 1405 (1929).
3. _____, *ibid.* **64**, 2395 (1931); G. Wittig and H. Petri, *Justus Liebigs Ann. Chem.* **505**, 17 (1933).
4. G. Wittig and W. Schoch, *Justus Liebigs Ann. Chem.* **749**, 38 (1971).
5. G. Wittig, U. Pockels, H. Dröge, *Ber. Dtsch. Chem. Ges.* **71**, 1903 (1938).
6. G. Wittig and U. Pockels, *ibid.* **72**, 89 (1939).
7. G. Wittig, P. Davis, G. Koenig, *Chem. Ber.* **84**, 627 (1951).
8. G. Wittig, G. Pieper, G. Fuhrmann, *Ber. Dtsch. Chem. Ges.* **73**, 1193 (1940).
9. G. Wittig, *Naturwissenschaften* **30**, 696 (1942).
10. J. D. Roberts, H. E. Simmons, Jr., L. A. Carlsmith, C. W. Vaughan, *J. Am. Chem. Soc.* **75**, 3290 (1953).
11. G. Wittig and L. Pohmer, *Chem. Ber.* **89**, 1334 (1956).
12. G. Wittig and H. F. Ebel, *Justus Liebigs Ann. Chem.* **650**, 20 (1961); H. F. Ebel and R. W. Hoffmann, *ibid.* **673**, 1 (1964).
13. G. Wittig and M. Rieber, *ibid.* **562**, 187 (1949).
14. G. Wittig and K. Clauss, *ibid.* **577**, 26 (1952); *ibid.* **578**, 136 (1952).
15. G. Wittig and K. Torssell, *Acta Chem. Scand.* **7**, 1293 (1953).
16. G. Wittig and G. Felletschin, *Justus Liebigs Ann. Chem.* **555**, 133 (1944).
17. Compare A. W. Johnson, *Ylid Chemistry* (Academic Press, New York, 1966).
18. G. Wittig and M. Rieber, *Justus Liebigs Ann. Chem.* **562**, 177 (1949).
19. G. Wittig and G. Geissler, *ibid.* **580**, 44 (1953).
20. G. Wittig and U. Schöllkopf, *Chem. Ber.* **87**, 1318 (1954).
21. On a variant of synthesis of olefins with carbonyl compounds, see J. Boutagy and R. Thomas, *Chem. Rev.* **74**, 87 (1974).
22. H. Pommer, *Angew. Chem.* **89**, 437 (1977).
23. G. Wittig, *ibid.* **62**, 231 (1950); *ibid.* **70**, 65 (1958).
24. R. W. Hoffmann, *Dehydrobenzene and Cycloalkynes* (Academic Press, New York, 1967).
25. G. Wittig, *Angew. Chem.* **68**, 505 (1956); *Festschrift Arthur Stoll* (Birkhäuser, Basel, 1957).
26. _____, G. Keicher, A. Rückert, P. Raff, *Justus Liebigs Ann. Chem.* **563**, 110 (1949).

DNA Methylation and Gene Function

Aharon Razin and Arthur D. Riggs

Summary. In most higher organisms, DNA is modified after synthesis by the enzymatic conversion of many cytosine residues to 5-methylcytosine. For several years, control of gene activity by DNA methylation has been recognized as a logically attractive possibility, but experimental support has proved elusive. However, there is now reason to believe, from recent studies, that DNA methylation is a key element in the hierarchy of control mechanisms that govern vertebrate gene function and differentiation.

The methylated base 5-methylcytosine (m^5Cyt) (Fig. 1) was discovered in calf thymus DNA about 30 years ago (*1*). Since then, the occurrence of this minor base has been demonstrated in a wide variety of organisms, including all vertebrates and plants that have been studied (*2*). In mammalian DNA, 2 to 7 percent (depending on the species) of the total cytosine is converted to m^5Cyt (*3*). Methylation occurs enzymatically after DNA synthesis by methyl transfer from *S*-adenosylmethionine (SAM) to position 5 of cytosine (*4*).

In vertebrates, m^5Cyt is the only modified base yet found in DNA. The biological role of m^5Cyt in eukaryotic DNA is not known, but its ubiquity suggests some important function. In the last decade, several hypotheses suggesting a role of m^5Cyt in gene regulation have been advanced (*5-8*). However, there has been only one review on m^5Cyt in eukaryotic DNA (*9*), and this review was completed before the power of restriction enzyme analysis was fully realized. Studies of the last 2 years, mainly with restriction enzymes, have drastically changed the experimental approaches to a solution of the long-standing enigma of the function of m^5Cyt in eukaryotic DNA. The purpose of this article is to review critically the available information in this rapidly expanding field. It will be obvious that we prefer the idea that m^5Cyt is involved in gene regulation and differentiation, but its involvement in other possible functions will also be considered.

Protein-DNA Interactions

Regardless of the details, it has long been our opinion (*6*) that the essential function of m^5Cyt is to modify protein-DNA interactions. The conversion of cytosine to m^5Cyt introduces a methyl group into an exposed position in the major groove of the DNA helix (Fig. 2), and the binding to DNA of proteins such as the lac repressor, histones, and hormone receptors is known to be affected by changes in the major groove (*10*). For example, it has been shown (*11*) that changing the thymine residue at nucleotide position 13 in the lac operator to uracil or to cytosine greatly decreases the affinity of repressor for operator. Changing position 13 to m^5Cyt restores the affinity for repressor to normal. Therefore, the lac repressor only senses the presence or absence of a methyl group at position 13. Bacterial restriction enzymes have a strong affinity for unmethylated restriction sites, but have a reduced affinity for, and no activity on, methylated sites (*12-14*). Thus, it is a solid fact that m^5Cyt can profoundly affect the binding of proteins to DNA. The only question is whether

Dr. Razin is a member of the Department of Cellular Biochemistry at the Hadassah Medical School, Hebrew University, Jerusalem, Israel. Dr. Riggs is a member of the Division of Biology, City of Hope Research Institute, Duarte, California 91010.

0036-8075/80/1107-0604$01.75/0

eukaryotic cells have availed themselves of this opportunity for gene control. We think the body of evidence now accumulated suggests that they have.

Basic Assay Techniques

Rapid progress has been made in recent years in the development of highly sensitive and reliable methods for the identification and quantification of m^5Cyt in DNA. To conventional chromatographic methods (*15*) have been added more sophisticated methods such as gas chromatography (*16*), high-resolution-mass spectrometry (*17*), gas chromatography–mass spectrometry (*18*) and high-performance liquid chromatography (*19*). In addition, specific antibodies against m^5Cyt have been used to detect methylated regions in intact chromosomes (*20*). All the above-mentioned methods yielded important information, but failed to provide any clue as to the function of m^5Cyt in eukaryotic DNA. It became clear, therefore, that new methods were needed to reveal methylated sequences in specific regions of the chromosome. That need was met by the use of bacterial restriction endonucleases which cleave DNA at specific sites.

The majority (about 90 percent) of the m^5Cyt residues in eukaryotic DNA are found in the dinucleotide sequence CpG (5′-CG; C, cytosine; G, guanine) (*21*). Fortunately, several restriction enzymes include CpG in their recognition sequence (*14*). Some of these "CpG enzymes" (and cutting sites) (*22*) are the restriction endonucleases Hpa II (CCGG), Msp I (CCGG), Hha I (GCGC), Xho I (CTCGAG), Ava I (CPyCGPuG), Sal I (GTCGAC), and Sma I (CCCGGG) (T, thymine; A, adenine; Py, pyrimidine). Most of these enzymes do not cut the DNA if the CpG sequence is methylated (*23*). Thus, these enzymes can be used to probe for methylation. In practice, if the results are to be meaningful, it is necessary to have a control showing full cutting, so the pattern to be expected in the absence of methylation is known (for clarification, see Fig. 3). A control showing the full cutting pattern can be obtained if the DNA region of interest has been cloned in *Escherichia coli*, since in cloned DNA, CpG sites are unmethylated. In this way, the methylation pattern of calf thymus satellite DNA was first deduced (*24*).

A more convenient and generally applicable approach became possible when it was found that Msp I recognizes the same sequence as Hpa II (CCGG) but cuts the DNA regardless of the methylation state of the internal cytosine (*25–27*). Some of the most interesting results have come from the use of the Hpa II–Msp I enzyme pair and the Southern blotting technique (*28*). With this technique, total genomic DNA is cut with either Hpa II or Msp I. The DNA fragments then are separated according to size by agarose gel electrophoresis and transferred by blotting to a nitrocellulose sheet. The DNA fragments in or near the gene of interest are then visualized by hybridization with a labeled probe, which can be either RNA or cloned DNA. Figure 3 is an idealized autoradiogram illustrating this important technique as introduced by Waalwijk and Flavell (*29*). A comparison of the bands seen for Hpa II with those seen for Msp I shows the location of methylated CCGG sites.

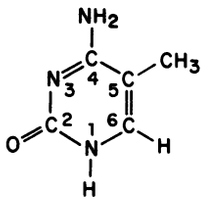

Fig. 1. Structure of 5-methylcytosine.

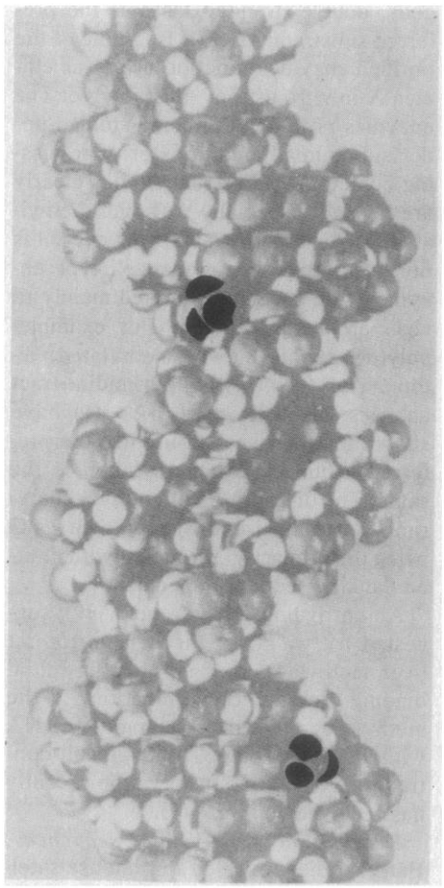

Fig. 2. Location of the methyl group of 5-methylcytosine in the major groove of B-form DNA. The methyl groups at position 5 of a pyrimidine ring are shown in black.

Residues of m^5Cyt can also be detected by the Maxam-Gilbert DNA sequencing technique (*30*) as a blank in the sequence ladder (*31*).

Basic Facts

Symmetry and heritability. Bacterial restriction and methylation enzymes recognize symmetrical sites, and both strands are symmetrically methylated (*13, 32*). Thus, these sites can exist in three states (Fig. 4): unmethylated, half-methylated, and fully methylated. After DNA replication, a half-methylated site is quickly converted to a fully methylated site. At least one bacterial methylase has been shown to act in vitro much faster ($\geq$ 100 times) on a half-methylated site than it does on an unmethylated site (*33*).

Using the above facts, Riggs (*6*) and Holliday and Pugh (*7*) pointed out that symmetrical methylation of both strands, coupled with a methylase (maintenance methylase, Fig. 4) acting only on half-methylated sites, would lead to the maintenance through DNA replication of a methylation pattern on the DNA. Methylated sites would remain methylated; unmethylated sites would remain unmethylated. Thus, there should be clonal inheritance of a methylation pattern.

The elegant studies by Bird and Southern (*22*) and Bird (*34*) have provided experimental support for these models. Amplified oocyte ribosomal DNA (rDNA) is unmethylated, whereas unamplified somatic rDNA is highly methylated (13 percent of the total cytosine is m^5Cyt). In somatic DNA, most of the Hpa II (CCGG) sites are methylated; however, many of the repeat units in the DNA had one unmethylated site (that is, there was a methylation pattern). Bird reasoned that, if methylation were only on one strand, then hybridizing somatic rDNA with a large excess of unmethylated amplified or cloned DNA would generate many unmethylated sites susceptible to Hpa II. When this experiment was done, the result was that there was no increase in susceptibility to Hpa II. Since half-methylated sites are fully protected from cutting, this experiment suggests that methylation is symmetrical on both strands. Very few half-methylated sites exist in nonreplicating *Xenopus* rDNA; the sites are either fully methylated or unmethylated. Cedar *et al.* (*27*) recently have determined directly that, in calf thymus DNA, the internal cyto-

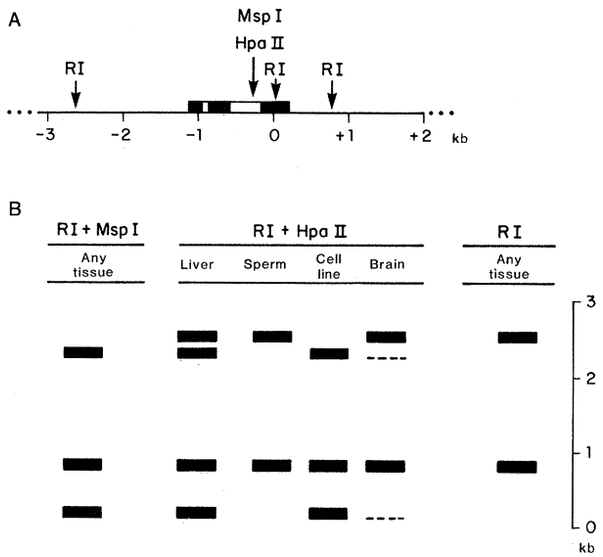

Fig. 3. The Waalwijk-Flavell experiment. (A) Map of the rabbit β-globin gene region. The globin coding regions are shown in black; introns are white. Restriction enzyme cutting sites are indicated. (B) Idealized autoradiogram obtained after treatment of unlabeled rabbit DNA from various tissues with the indicated restriction enzymes followed by agarose electrophoresis, Southern blotting, and hybridization to ^{32}P-labeled globin probe. Because of its small size, the smallest DNA fragment is not apparent in the original publication (*29*). It is shown here only for logical clarity.

sines in both strands of the Hpa II site are, in fact, methylated. During replication, of course, half-methylated sites are created, and Bird (*34*) has shown that methyl groups are added to the new strand only. Taken together, these results indicate (i) there is a methylation pattern in somatic rDNA, (ii) methylation is in both strands, and (iii) the methylation pattern is maintained through DNA replication. More recent experiments also indicate that DNA methylation is clonally inherited. pBr322 and ϕX174 DNA were methylated in vitro by means of Hpa II methylation, and inserted into mouse cells via DNA-mediated gene transfer (*35*). Whereas unmethylated DNA sequences generally remain unmethylated, the in vitro methylated DNA sequences retain their methyl moieties even after 50 generations of growth and culture (*36*). The heritability of a methylation pattern is further confirmed by experiments showing tissue specificity.

Since methylation is mainly in the symmetrical sequence CpG, Fig. 4 summarizes our current thinking. The in vivo results indicate that a maintenance methylase must exist. Simple reasoning also suggests that at some stage, de novo methylation must occur. In fact, it is known that unmethylated adenovirus DNA becomes methylated in some cell lines when integrated into the genome (*37*).

DNA methylases. There have been several studies on mammalian DNA methylases. *S*-Adenosylmethionine–dependent DNA methylases have been purified from human HeLa cells (*38*), from rat liver (*39, 40*), from hepatoma (*41*), and from mouse ascites cells (*42*). These studies have given no indication of a large number of highly specific enzymes. On the contrary, the picture that we think is emerging is that each cell type (and probably each species) has only one or two DNA methylases with limited specificity. Single- and double-stranded DNA are methylated at most (but not all) CpG (5′-CG) sequences. Denatured, nonhomologous DNA usually is the preferred substrate, but there are two reports that undermethylated, homologous, double-stranded DNA is the preferred substrate (*41, 42*). But none of the purified enzymes function nearly as efficiently in vitro as they must in vivo. The enzymes probably are being studied under suboptimal conditions or with missing cofactors. Additional studies clearly are needed, especially with half-methylated DNA's as substrates. However, it is our guess that "maintenance" type enzymes with specificity limited mainly to CpG are being studied. For example, poly(dG-dC) is slowly methylated, although not fully (*40*). By pyrimidine tract analysis, Browne *et al.* (*43, 44*) have shown that in vivo and in vitro the preferred sequence for methylation is the same. The distribution of methylated pyrimidine tracts is consistent with CpG being the primary determinant with limited flanking sequence specificity.

An intriguing early report (*39*) suggested a "walking" mechanism for rat liver methylase; that is, after an initial binding interaction (initiation?), the methylase may travel along the DNA without dissociating. Although still controversial (*42*), this observation recently has received support (*40*).

Distribution of methylated sequences. Early studies revealed that mouse satellite DNA is more than twice as methylated as main band DNA (*45*). That satellite or highly repetitious DNA tends to be highly methylated has been confirmed for a number of organisms (*19, 46, 47*). Immunofluorescence or immunoperoxidase procedures with antibodies to m^{5}Cyt show that many centromeric regions of mitotic chromosomes have a high content of m^{5}Cyt (*20, 48*). However, other studies clearly show that single-copy DNA contains considerable m^{5}Cyt. For example, rat single-copy DNA has m^{5}Cyt as 2.3 percent of total cytosine (*19*).

Two studies have indicated a nonrandom distribution of m^{5}Cyt with respect to nucleosome structure (*47, 49*). Micrococcal nuclease preferentially digests DNA between nucleosome cores. Early during the digestion with this enzyme, the DNA released contained very little m^{5}Cyt. Thus, most m^{5}Cyt residues probably are "covered" by nucleosomes.

It has been reported recently (*50*) that long stretches of sea urchin DNA are unmethylated, and an earlier autoradiographic study (*51*) indicated that this also may be the case for mammalian DNA.

Tissue specificity. Waalwijk and Flavell (*29*) first provided convincing evidence that there are tissue-specific methylation patterns (Fig. 3). Rabbit DNA's from different tissues were digested with Eco RI and either Hpa II (CCGG, sensitive to m^{5}Cyt) or Msp I (CCGG, insensitive to m^{5}Cyt) restriction enzymes and then analyzed by agarose gel electrophoresis, Southern blotting (*28*), and visualization of globin containing restriction fragments by hybridization with a labeled rabbit β-globin probe (cloned rabbit globin DNA). The restriction fragment patterns obtained with Hpa II (but not with Msp I) clearly were tissue-specific, indicating tissue specificity in methylation at a site within the β-globin gene intron (intervening sequence). At this site, brain and sperm DNA were 80 and 100 percent methylated, respectively, whereas DNA from a rabbit cell line was completely unmethylated, and DNA from liver was about 50 percent methylated. This type of result has been confirmed for chicken globin genes (*52*), human globin (*23*), ovalbumin (*53*), and for several cell lines, with viral DNA hybridization probes (*54*). Thus, when specific sites are observed, tissue specificity is readily apparent.

In contrast, if one measures the average or overall level of methylation in vertebrates, only small changes of questionable significance are seen (*47, 55*). There have been several reports where tissue-specific differences in total m^{5}Cyt are claimed (*56, 57*), but these changes generally were less than 10 percent. There has been one widely quoted report that

bull sperm DNA has only half the m^{5}Cyt content of somatic cells (*3*). However, we think these measurements should now be made again with more modern methods, with due consideration being given to the high satellite content of bovine DNA. Restriction analyses indicate that rabbit and mouse sperm DNA is highly methylated (*29, 53*) and that sea urchin sperm DNA has the same m^{5}Cyt content as somatic DNA (*58*).

Restriction enzymes for sequences containing CpG have been used to determine the average methylation of mouse embryo DNA from the two-cell to blastocyst stages. Thus, we now know that CCGG and GCGC sites are methylated to similar extents (70 to 80 percent) in embryos and in all tissues of the adult mouse (*55*). (It may be of interest to note that although tissue specificity was not observed, species specificity in average methylation of GCGC sites was observed: 50 percent in rabbit, 75 percent in mouse.) It is clear that the specific changes in the methylation pattern that occur during differentiation take place against a high background so that the average total methylation changes less than 10 percent. It is also worth noting that restriction enzyme analyses probe only about 10 percent of the total methylated sites.

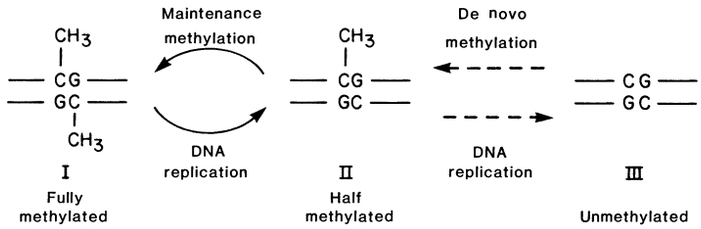

Fig. 4. Three possible states of methylation: fully methylated, half-methylated, and unmethylated. It is probable that de novo methylation occurs rarely; most DNA methylation is maintenance-type methylation following DNA replication.

Some Possible Functions

Restriction and modification. Protection from eukaryotic restriction enzymes is the most obvious function for m^{5}Cyt. Sager and Kitchin (*8*) have, in fact, proposed a model for differentiation based on restriction and modification. However, in spite of numerous (unpublished) efforts, sequence-specific endonucleases in mammalian cells have not been detected. Moreover, the methylation pattern of eukaryotic DNA argues against this possibility. In contrast to those in *E. coli* DNA (*59*), specific methylation sites in mammals are not completely methylated; for example, 50 percent of Hpa II sites are unmethylated in rabbit DNA (*26*). Also, mammalian cells will take up unmethylated DNA and maintain it intact (*35*). Nevertheless, one should still keep an open mind, because preferential cleavage of SV40 DNA near the replication origin by an extract of green monkey testis cells has been reported (*60*), and a SAM-dependent, nonspecific endonuclease in hamster kidney fibroblasts has been detected (*61*).

In chlamydomonas, it now seems certain that methylation is involved in the maternal inheritance of chloroplast DNA (*62*). Chloroplast DNA from mating type plus (Mt$^+$) gametes is highly methylated, whereas chloroplast DNA from Mt$^-$ gametes is only slightly methylated. After zygote formation, the Mt$^-$ chloroplast DNA is degraded. Thus, in this system, restriction-modification clearly exists.

DNA replication. In the phage ϕX174, DNA methylation at one specific site is known to be involved in replication; blocking methylation blocks DNA synthesis (*63*). The replication of the *E. coli* chromosome also stops after one round of synthesis in the absence of methylation (*64*). The sequence of the replication origin of *E. coli* has been published (*65*), and at the origin the sequence GATC, which is always methylated under normal growth conditions (*59*), occurs ten times more frequently than expected. Together these observations suggest a possible role for DNA methylation in prokaryotic DNA replication.

Eukaryotic cells, since they have so many bidirectional replication origins (*66*), must orchestrate carefully their complex DNA synthesis process. Taylor (*9*) has proposed a rather detailed model for the control of replication by DNA methylation. It is clearly too early to judge the validity of this model. For example, because of conflicting results, it is still not known with certainty when, after DNA synthesis, methylation is completed (*56, 67*).

Recombination and mutation. Recent results obtained with prokaryotes suggest that this possibility should be considered seriously. Methylase-deficient *E. coli* mutants have been isolated (*68*) that have reduced methylation of GATC sequences (*69*). These mutants show increased spontaneous mutagenesis, increased spontaneous induction of lambda prophage, increased sensitivity to mutagens and ultraviolet light, and hyperrecombination phenotype (*70*). Glickman *et al*. (*71*) also have found that, when a lambda heteroduplex DNA molecule with a single base mismatch and only one strand methylated is used for transfection, the methylated strand is preferentially used as a template during in vivo mismatch correction.

In *E. coli*, methylated sites can be mutation "hot spots" (*72*). A major hot spot in the lac repressor gene is the sequence CC(A/T)GG which, in wild-type *E. coli* is always methylated (*59*). Coulondre *et al*. (*72*) suggest that deamination of cytosine produces uracil, which is recognized and preferentially removed, whereas deamination of m^{5}Cyt generates thymine, which is not preferentially corrected.

Scarano (*5*) pointed out that deamination of m^{5}Cyt at specific sites would lead to a heritable change in the DNA (GC to AT transition) that could influence differentiation. Although this is an interesting possibility, there has been little experimental support for this idea (*73*).

It has been known for some time that the doublet CpG is rare in eukaryotic DNA, and Salser *et al*. (*74*) have pointed out that if methylated CpG sites were mutation hot spots, evolution would tend to eliminate them except where there was positive selection. This model suggests that noncoding regions will show a scarcity of CpG sequences relative to coding sequences. This idea should be testable as more sequences appear, but note that satellite DNA's tend to be rich in methylated CpG. A recent analysis of nearest neighbor dinucleotide frequencies and the level of DNA methylation supports the idea that m^{5}Cyt tends to mutate to T (*75*).

Sneider *et al*. (*76*) have proposed a meiotic recombination model which predicts half-methylated sites in sperm and oocyte DNA. By restriction analysis, the DNA of sperm is, indeed, methylated differently from somatic DNA (*29, 53, 76*).

Chromosome folding, packing, and sorting. The complex secondary and tertiary folding of eukaryotic chromosomes probably involves specific protein-DNA interactions (*77*) which could be influenced by the distribution of m^{5}Cyt in the DNA (*49*). The high methylation of centromeric regions suggests a possible role in mitosis and chromosome sorting.

None of the possible biological functions mentioned above have yet received any experimental support from eukaryotic studies. In sharp contrast, the idea that methylation is involved in gene regulation and differentiation has been strongly supported. Several predictions made by the differentiation models have been verified. (i) Specific methylation patterns exist; (ii) methylation is symmetrical in both strands; (iii) methylation patterns are clonally heritable; and (iv) methylation patterns are tissue-specific. This last-mentioned deserves emphasis. If methylation is not involved in gene regulation and differentiation, then why are there tissue-specific methylation patterns? The case for an important role of m^5Cyt in the control of gene activity is even stronger than presented so far, as will be seen below.

Correlation of gene activity with undermethylation. McGhee and Ginder (*52*) first reported that certain specific methylation sites (Hpa II, CCGG) in the region of the chicken β-globin genes are less methylated in erythrocytes and reticulocytes than in oviduct tissue. Corresponding results have been obtained by Mandel and Chambon (*53*), who studied the chicken ovalbumin gene and found a correlation between undermethylation and gene activity. They observed three classes of Hpa II and Hha I sites, namely, nonmethylated in all tissues, fully methylated in all tissues, and variably methylated with respect to tissue type. The variable sites are less methylated in active tissue. Similar conclusions have been reached from studies on rabbit and human globin genes (*23, 78*). Some, but not all, sites change their methylation level; active genes are always less methylated.

A similar story is emerging from viral studies (*37, 54, 79*). For some time, it has been known that adenovirus virion DNA is essentially unmethylated, whereas chromosomally integrated and nonexpressed DNA is methylated (*37*). It also has been found (*54*) that the herpes Saimiri-transformed cell line 1670, which is not producing detectable virus, contains many episomal copies of highly methylated DNA. Cell lines actively producing herpes Saimiri virus have unmethylated viral DNA.

Another study (*80*) has probed for methylation of integrated retrovirus DNA, and a correlation was observed between undermethylation of specific sites and expression of viral gene sequences. Finally, studies have been done where thymidine kinase minus (Tk^-) mouse L cells were transformed to Tk^+ with the thymidine kinase gene of herpes simplex (*35*). Even when transformation is done with a *tk* gene fully methylated in vitro at Hpa II sites, most Tk^+ transformed cells have only unmethylated herpes *tk* genes (*36*). Tk^- revertants obtained from Tk^+ transformed cell lines are usually unstable, frequently converting back to Tk^+. In one intriguing instance, an unusually stable Tk^- revertant was found to have a highly methylated *tk* gene (*81*).

Active chromatin regions are known to be more sensitive to deoxyribonuclease I (*82*). A recent report (*83*) indicates that undermethylated DNA in the region of the chicken ovalbumin gene is also sensitive to this enzyme. For both deoxyribonuclease I sensitivity (*82*) and undermethylation (*83*), it is important to note that continued transcription is not necessary after the gene has been set in the active configuration. For example, in the case of chicken erythrocytes, although the globin gene is not being transcribed, it still is undermethylated (*52*) and sensitive to deoxyribonuclease I (*82*).

For all of the variably methylated sites that have been observed to date, there is no obvious reason why methylation should critically affect transcription. For example, the Hpa II site in the rabbit globin gene is in an intron. As always with regard to correlations, one should be cautious in drawing conclusions. In particular, the above studies give no information as to cause or effect; undermethylation could be just a trivial result of transcription inhibiting methylation. However, the results to be described next suggest that methylated sites in DNA do, in fact, determine the differentiated state of the cell.

Ethionine and 5-azacytidine. Ethionine, a methionine analog that inhibits most methyltransferases, is an effective inducer of globin gene expression in Friend erythroleukemia cells. In spite of the fact that ethionine is toxic and affects many cellular processes, Christman *et al.* (*84*) suggested that induction of globin production is caused by undermethylation of the DNA. Although this experiment is difficult to interpret unambiguously, nonetheless it was the first indication that undermethylation of DNA can affect gene activity.

More definitive results have been obtained by Taylor and Jones (*85*) from studies with 5-azacytidine. 5-Azacytidine is a cytosine analog with a nitrogen atom replacing the carbon atom at position 5 of the pyrimidine ring and thus cannot accept a methyl group. Taylor and Jones treated the mouse cell line $10T^1/_2$ with 5-azacytidine for a short time, removed the analog, and observed that several generations later, foci of differentiated cells (muscle, chondrocytes, and adipocytes)

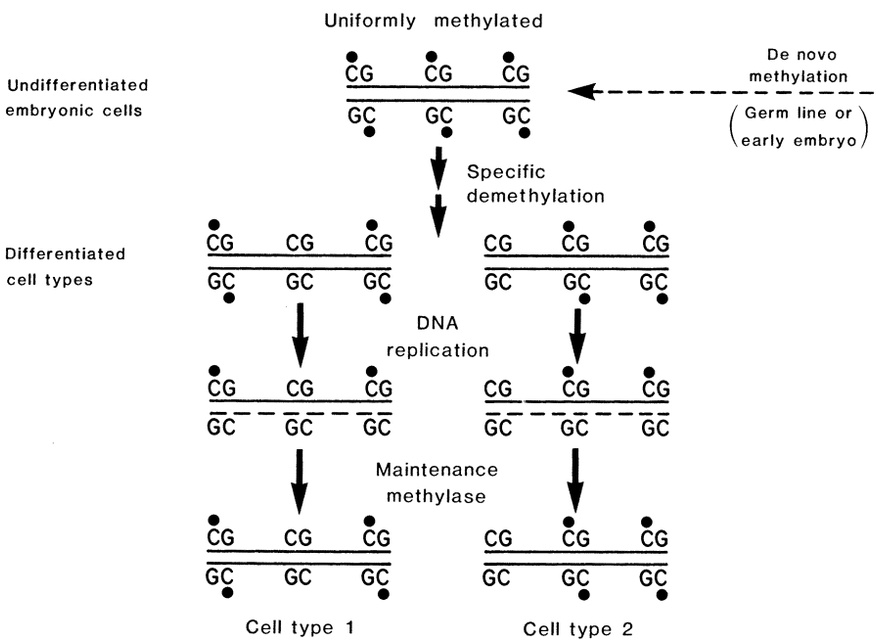

Fig. 5. A demethylation model for the establishment and maintenance of a differentiated state. Inhibition of methylation by sequence-specific proteins during DNA replication leads to demethylation and the establishment of specific methylation patterns (closed circles represent methyl groups) in various cell types. The maintenance methylase system ensures heritability of the methylation pattern.

were apparent. Mutagenesis was considered, but none of the mutagens tested caused differentiation of this cell line. After the above observations were published, it was found (*86*) that 5-azacytidine causes undermethylation of DNA. Since the continued presence of 5-azacytidine is not necessary (short treatment during DNA synthesis is sufficient), this analog is causing a clonally heritable change in gene activity.

Demethylation models. A possibility considered likely in earlier models (*6, 7*) was that at some point, perhaps in the early embryo, DNA would be undermethylated or nonmethylated, providing a clean slate upon which to lay down a methylation pattern during differentiation. However, now it is known that the DNA in mammalian sperm and early embryos is highly methylated (*29, 53, 55*) and undermethylation is correlated with gene activity.

Therefore, several investigators (*23, 55, 84, 86, 87*) have suggested that the specific event often must be demethylation. One extreme model (*55*) is illustrated schematically in Fig. 5. In the early undifferentiated state, the DNA is postulated to be fully or "uniformly" methylated in that all sites that ever will be are methylated. During development, sequence-specific proteins would inhibit methylation during DNA replication, leading to methylation patterns specific for each tissue. The necessary involvement of DNA replication (critical mitoses) in differentiation has long been postulated (*88*) and also suggested by the recent 5-azacytidine experiments (*85*). Once the specific demethylation events occur, the differentiated methylation pattern would be inherited clonally as a result of the maintenance methylase system, provided that de novo methylation is slow. Starting from a uniform, fully methylated ground state, and proceeding unidirectionally by demethylation is logically attractive and also is consistent with the studies showing sperm and embryonic DNA to be highly methylated (*23, 29, 53, 55, 58*).

However, at this time it would be premature to focus on one model because it is obvious that there are many possible additions and variations. For example, restriction enzymes could be added and specific demethylation could influence gene deletion, translocation, or inversion (*8, 89*). We should also point out that some limited de novo methylation probably does occur in somatic cells (*35*), and thus there may be "shuffling," that is, addition as well as removal, of methyl groups during differentiation.

Conclusion

Vertebrate gene regulation clearly results from an interplay between DNA and chromosomal proteins: nucleosomal proteins, nonhistone proteins, HMG proteins (*90*), and specific regulator proteins (repressors, activators). Protein modification by acetylation may be one of the control mechanisms (*91*). But the point we wish to make here is that DNA modification by methylation probably is also an important part of this complex hierarchy of controls. Perhaps methylation is utilized primarily for stable memory of a differentiated state through DNA replication, a process that might otherwise disrupt delicate protein-DNA and protein-protein interactions. This possibility is suggested by the observation that some organisms in which differentiated cells do not continue replication may not utilize m^5Cyt DNAmodification (*27, 92*).

References and Notes

1. R. D. Hotchkiss, *J. Biol. Chem.* **175**, 315 (1948).
2. R. H. Hall, *The Modified Nucleosides in Nucleic Acids* (Columbia Univ. Press, New York, 1971); H. A. Sober, *Handbook of Biochemistry* (Chemical Rubber Company, Cleveland, Ohio, 1970).
3. B. F. Vanyushin, S. G. Tkacheva, A. N. Belozersky, *Nature (London)* **225**, 948 (1970).
4. B. Sheid, P. R. Srinivasan, E. Borek, *Biochemistry* **7**, 280 (1968); T. W. Sneider and V. R. Potter, *J. Mol. Biol.* **42**, 271 (1969); R. H. Burdon and R. L. P. Adams, *Biochim. Biophys. Acta* **174**, 322 (1969); F. Kalousek and N. R. Morris, *J. Biol. Chem.* **244**, 1157 (1969).
5. E. Scarano, *Adv. Cytopharmacol.* **1**, 13 (1971).
6. A. D. Riggs, *Cytogenet. Cell Genet.* **14**, 9 (1975).
7. R. Holliday and J. E. Pugh, *Science* **187**, 226 (1975).
8. R. Sager and R. Kitchin, *ibid.* **189**, 426 (1975).
9. J. H. Taylor, *Molecular Genetics*, part 3, *Chromosome Structure*, J. H. Taylor, Ed. (Academic Press, New York, 1979).
10. S. Lin and A. D. Riggs, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2574 (1972); S. Lin, D. Lin, A. D. Riggs, *Nucleic Acids Res.* **3**, 2183 (1976); J. Kallos, T. M. Fasy, V. P. Hollander, M. D. Bick, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4896 (1978).
11. E. F. Fisher and M. H. Caruthers, *Nucleic Acids Res.* **7**, 401 (1979).
12. R. Yuan and M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **65**, 357 (1970).
13. H. O. Smith, *Science* **205**, 455 (1979).
14. R. J. Roberts, *Gene* **4**, 183 (1978).
15. K. Randerath and E. Randerath, *Methods Enzymol.* **12**, 323 (1968); A. Razin, J. W. Sedat, R. L. Sinsheimer, *J. Mol. Biol.* **53**, 251 (1970).
16. A. Razin and J. W. Sedat, *Anal. Biochem.* **77**, 370 (1977).
17. J. Deutsch, A. Razin, J. Sedat, *ibid.* **72**, 586 (1976). A low-resolution mass spectrometric method also has been developed [J. L. Weibers, *Nucleic Acids Res.* **3**, 2959 (1976)]. However, it should be noted that when low-resolution mass spectrometry is used, some artifacts might occur resulting from molecular ion fragments [such as the (M-1) ion of thymine] identical in mass value with m^5Cyt.
18. J. Singer, W. C. Schnute, Jr., J. E. Shively, C. W. Todd, A. D. Riggs, *Anal. Biochem.* **94**, 297 (1979).
19. J. Singer, R. H. Stellwagen, J. Roberts-Ems, A. D. Riggs, *J. Biol. Chem.* **252**, 5509 (1977).
20. O. J. Miller, W. Schnedl, J. Allen, B. F. Erlanger, *Nature (London)* **251**, 636 (1974); B. W. Lubit, R. R. Schreck, O. J. Miller, B. F. Erlanger, *Exp. Cell Res.* **89**, 426 (1974); B. W. Lubit, T. D. Pham, O. J. Miller, B. F. Erlanger, *Cell* **9**, 503 (1976).
21. J. Doscocil and F. Sorm, *Biochim. Biophys. Acta* **55**, 953 (1962); P. Grippo, M. Iaccarino, E. Parisi, E. Scarano, *J. Mol. Biol.* **36**, 195 (1968).
22. A. P. Bird and E. M. Southern, *J. Mol. Biol.* **118**, 27 (1978).
23. L. H. T. Van der Ploeg and R. A. Flavell, *Cell* **19**, 947 (1980).
24. F. Gautier, H. Bünemann, L. Grotjahn, *Eur. J. Biochem.* **80**, 175 (1977).
25. C. Waalwijk and R. A. Flavell, *Nucleic Acids Res.* **5**, 3231 (1978).
26. J. Singer, J. Roberts-Ems, A. D. Riggs, *Science* **203**, 1019 (1979).
27. H. Cedar, A. Solage, G. Glaser, A. Razin, *Nucleic Acids Res.* **6**, 2125 (1979).
28. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
29. C. Waalwijk and R. A. Flavell, *Nucleic Acids Res.* **5**, 4631 (1978).
30. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
31. H. Ohmori, J. Tomizawa, A. M. Maxam, *Nucleic Acids Res.* **5**, 1479 (1978).
32. R. J. Roberts, *CRC Crit. Rev. Biochem.* **4**, 123 (1976).
33. G. F. Vovis, K. Horiuchi, N. D. Zinder, *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 3810 (1974).
34. A. P. Bird, *J. Mol. Biol.* **118**, 49 (1978).
35. M. Wigler *et al.*, *Cell* **16**, 777 (1979).
36. Y. Pollack, R. Stein, A. Razin, H. Cedar, *Proc. Natl. Acad. Sci. U.S.A.*, in press; R. Stein, Y. Pollack, A. Razin, H. Cedar, personal communication.
37. D. Sutter, M. Westphal, W. Doerfler, *Cell* **14**, 569 (1978); D. Sutter and W. Doerfler, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 253 (1980); L. Vardiman, R. Neumann, I. Kuhlmann, D. Sutter, W. Doerfler, *Nucleic Acids Res.* **8**, 2461 (1980).
38. P. Y. Roy and A. Weissbach, *Nucleic Acids Res.* **2**, 1669 (1975).
39. D. Drahovsky and N. R. Morris, *J. Mol. Biol.* **57**, 475 (1971); *ibid.* **61**, 343 (1971).
40. D. Simon, F. Grunert, U. V. Acken, H. P. Doring, H. Kröger, *Nucleic Acids Res.* **5**, 2153 (1978).
41. T. W. Sneider, W. M. Teague, L. M. Regachevsky, *ibid.* **2**, 1685 (1975).
42. J. F. Turnbull and R. L. P. Adams, *ibid.* **3**, 677 (1976).
43. M. J. Browne, J. F. Turnbull, E. L. McKay, R. L. P. Adams, R. H. Burdon, *ibid.* **4**, 1039 (1977).
44. M. J. Browne and R. H. Burdon, *ibid.*, p. 1025.
45. R. Solomon, A. M. Kay, M. Hertzberg, *J. Mol. Biol.* **43**, 581 (1969).
46. K. Harbers, B. Harbers, J. H. Spencer, *Biochem. Biophys. Res. Commun.* **66**, 738 (1975).
47. A. Solage and H. Cedar, *Biochemistry* **17**, 2934 (1978).
48. W. Schneal, V. G. Den, R. Tantravahi, D. A. Miller, B. F. Erlanger, O. H. Miller, *Chromosoma* **52**, 59 (1975).
49. A. Razin and H. Cedar, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2725 (1977).
50. A. P. Bird, M. H. Taggart, B. A. Smith, *Cell* **17**, 889 (1979).
51. H. Molitor, D. Drahovsky, A. Wacker, *Biochim. Biophys. Acta* **432**, 28 (1976).
52. J. D. McGhee and G. D. Ginder, *Nature (London)* **280**, 419 (1979).
53. J. L. Mandel and P. Chambon, *Nucleic Acids Res.* **7**, 2081 (1979).
54. R. C. Desrosiers, C. Mulder, B. Fleckenstein, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3839 (1979).
55. J. Singer, J. Roberts-Ems, F. W. Luthardt, A. D. Riggs, *Nucleic Acids Res.* **7**, 2369 (1979).
56. J. W. Kappler, *J. Cell. Physiol.* **78**, 33 (1971).
57. B. F. Vanyushin, A. L. Razin, V. K. Vasilyev, A. N. Belozersky, *Biochim. Biophys. Acta* **299**, 397 (1973).
58. J. M. Pollock, M. Swihart, J. H. Taylor, *Nucleic Acids Res.* **5**, 4855 (1978).
59. There are two sites [CC(A/T)GG and GATC], methylated by the *E. coli* methylases coded by the genes *mec* and *dam*. These sites are fully methylated in both chromosomal and episomal DNA [A. Razin, S. Urieli, Y. Pollack, Y. Gruenbaum, G. Glaser, *Nucleic Acids Res.* **8**, 1783 (1980)].
60. F. L. Brown, P. R. Musich, J. J. Maio, *Nucleic Acids Res.* **5**, 1093 (1978).
61. R. H. Burdon, A. C. B. Cato, I. S. Valladsen, in FEBS 12th Meeting, Dresden, 1978, *Gene Function*, S. Rosenthal *et al.*, Eds. (Pergamon, New York, 1979), vol. 51, p. 159.
62. W. G. Burton, C. T. Grabowy, R. Sager, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1390 (1979); H. Royer and R. Sager, *ibid.*, p. 5794.
63. A. Razin, in *The Single-Stranded DNA Phages*, D. T. Denhardt, D. Dressler, D. S. Ray, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1978).
64. D. Billen, *J. Mol. Biol.* **31**, 477 (1968); C. Lark, *ibid.*, p. 389.
65. K. Sugimoto *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 575 (1979); M. Meijer *et al.*, *ibid.*, p. 580.

66. J. A. Huberman and A. D. Riggs, *J. Mol. Biol.* **32**, 327 (1968).
67. R. L. P. Adams, *Biochim. Biophys. Acta* **254**, 205 (1971); H. H. Evans, T. E. Evans, S. Littman, *J. Mol. Biol.* **74**, 563 (1973).
68. M. G. Marinus and N. R. Morris, *J. Bacteriol.* **114**, 1143 (1973).
69. S. Lacks and B. Greenberg, *ibid.*, p. 153.
70. M. G. Marinus and N. R. Morris, *Mutat. Res.* **28**, 15 (1975).
71. B. Glickman, P. Van den Elsen, M. Radman, *Mol. Gen. Genet.* **163**, 307 (1978).
72. C. Coulondre, J. H. Miller, P. J. Farabough, W. Gilbert, *Nature (London)* **274**, 775 (1978).
73. T. W. Sneider, *J. Mol. Biol.* **79**, 731 (1973).
74. W. A. Salser, I. Cummings, A. Liu, J. Strommer, J. Padayatty, P. Clarke, in *Cellular and Molecular Switching*, G. Stamatoyannopoulos and A. W. Nienhius, Eds. (Grune & Stratton, New York, 1979), p. 621.
75. A. P. Bird, *Nucleic Acids Res.* **8**, 149 (1980).
76. T. W. Sneider, J. Kaput, D. Neiman, B. Westmoreland, in *Transmethylation*, E. Usdin, R. T. Borchardt, C. R. Creveling, Eds. (Elsevier/North-Holland, New York, 1979), p. 473; J. Kaput and T. W. Sneider, *Nucleic Acids Res.* **7**, 2303 (1979).
77. D. E. Comings and A. D. Riggs, *Nature (London)* **233**, 48 (1971).
78. C. Shen and T. Mariatis, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1787 (1980).
79. J. C. Cohen, *Cell* **19**, 653 (1980).
80. J. Mullins, personal communication.
81. A. Pellicer *et al.*, *Science* **209**, 1414 (1980).
82. H. Weintraub and M. Groudine, *ibid.* **193**, 848 (1976).
83. M. T. Kuo, J. L. Mandel, P. Chambon, *Nucleic Acids Res.* **7**, 2105 (1979).
84. J. K. Christman, P. Price, L. Pedriman, G. Acs, *Eur. J. Biochem.* **81**, 53 (1977).
85. S. M. Taylor and P. A. Jones, *Cell* **17**, 771 (1979).
86. P. A. Jones and A. M. Taylor, *ibid.* **20**, 85 (1980).
87. C. B. Olson, *Speculations Sci. Technol.* **2**, 365 (1979).
88. H. Holtzer, *Q. Rev. Biophys.* **8**, 523 (1975).
89. D. G. Dickinson and R. F. Baker, *Science* **205**, 816 (1979).
90. E. W. Jones, G. H. Goodwin, J. M. Walker, C. Sanclus, *Ciba Found. Symp.* **28**, 95 (1975); S. Weisbad and H. Weintraub, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 630 (1979); B. Gazit, A. Panet, H. Cedar, *ibid.*, in press.
91. J. R. Davie, E. Peter, M. Candido, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3574 (1978).
92. P. M. M. Rae and R. E. Steele, *Nucleic Acids Res.* **6**, 2987 (1979). There is preliminary evidence that *Drosophila* may utilize 6-methyladenine for DNA modification (J. Mullins, personal communication).
93. Supported by NIH grants GM20483 and GM25825.

AAAS–Newcomb Cleveland Prize
To Be Awarded for an Article or a Report Published in *Science*

The AAAS-Newcomb Cleveland Prize is awarded annually to the author of an outstanding paper published in *Science* from August through July. This competition year starts with the 1 August 1980 issue of *Science* and ends with that of 31 July 1981. The value of the prize is $5000; the winner also receives a bronze medal.

Reports and Articles that include original research data, theories, or syntheses and are fundamental contributions to basic knowledge or technical achievements of far-reaching consequence are eligible for consideration for the prize. The paper must be a first-time publication of the author's own work. Reference to pertinent earlier work by the author may be included to give perspective.

Throughout the year, readers are invited to nominate papers appearing in the Reports or Articles sections. Nominations must by typed, and the following information provided: the title of the paper, issue in which it was published, author's name, and a brief statement of justification for nomination. Nominations should be submitted to AAAS-Newcomb Cleveland Prize, AAAS, 1515 Massachusetts Avenue, NW, Washington, D.C. 20005. Final selection will rest with a panel of distinguished scientists appointed by the Board of Directors.

The award will be presented at a session of the annual meeting. In cases of multiple authorship, the prize will be divided equally between or among the authors.

Deadline for nominations: postmarked 15 August 1981

Nomination Form
AAAS–Newcomb Cleveland Prize

AUTHOR: _____

TITLE: _____

_____

DATE PUBLISHED: _____

JUSTIFICATION: _____