# DNA Methylation in Eukaryotes: Kinetics of Demethylation and *de Novo* Methylation During the Life Cycle

**Sarah P. Otto and Virginia Walbot**

*Department of Biological Sciences, Stanford University, Stanford, California 94305-5020*

Manuscript received April 3, 1989
Accepted for publication October 20, 1989

ABSTRACT

We present a model for the kinetics of methylation and demethylation of eukaryotic DNA; the model incorporates values for *de novo* methylation and the error rate of maintenance methylation. From the equations, an equilibrium is reached such that the proportion of sites which are newly methylated equals the proportion of sites which become demethylated in a cell generation. This equilibrium is empirically determined as the level of maintenance methylation. We then chose reasonable values for the parameters using maize and mice as model species. In general, if the genome is either hypermethylated or hypomethylated it will approach the equilibrium level of maintenance methylation asymptotically over time; events occurring just once per life cycle to suppress methylation can maintain a relatively hypomethylated state. Although the equations developed are used here as framework for evaluating events in the whole genome, they can also be used to evaluate the rates of methylation and demethylation in specific sites over time.

IN prokaryotes it is well-established that DNA modification plays many roles. For example, methylation determines recognition of self *vs.* non-self DNA through the action of methylation-sensitive restriction enzymes, is involved in the regulation of transposon activity and determines the strand specificity of mismatch repair (reviewed in MESSER and NOYER-WEIDNER 1988). Although the role of 5′ methylation of cytosine residues in eukaryotic DNA is not as well understood, methylation is intriguing because there are many examples of a negative correlation between methylation and gene expression (reviewed by CEDAR 1988). Cytosine methylation is not, however, present in all eukaryotes, thus this DNA modification cannot be universally required for gene regulation. Among those organisms with 5-methylcytosine, the methylation substrate sites also vary. In vertebrates, C residues in the dinucleotide CpG are the substrate; in higher plants C residues in both CpG and CpXpG, where X equals any nucleotide except G, are substrate sequences (GRUENBAUM *et al.* 1981).

One interesting feature of DNA methylation in eukaryotes is its *relative stability*; that is, methylation pattern is a cell-heritable phenotype, a feature similar to aspects of cell differentiation (RAZIN 1984). Cell inheritance occurs because maintenance methylation enzymes utilize a template—the methylation pattern of an original DNA strand—to impose a symmetric pattern of methylation on a newly synthesized DNA strand (Figure 1). Equally important, however, is the ease with which an existing methylation pattern can be lost. This process is termed demethylation even though the process is not enzymatic. Instead methylation patterns are lost by virtue of the failure of the maintenance system after DNA replication.

If a cell were to fail totally in maintenance methylation, the effect on methylation levels would be rapid and dramatic during subsequent cell proliferation. After one round of DNA replication in the absence of methylation, chromosomes in each daughter cell would contain a template strand carrying the original methylation pattern and an unmethylated strand. Thus, all substrate sites will either be hemi-methylated or unmethylated. Every subsequent round of DNA replication without methylation will halve the frequency of template strands with an associated increase in completely unmethylated strands. With renewed methylation, maintenance methylation can act on chromosomes containing a template strand so that most hemi-methylated sites become homomethylated. Substrate sites on completely unmethylated chromosomes will, however, become methylated much more slowly because *de novo* methylation of unmethylated DNA is infrequent. In fact, methylation is estimated to occur 10 to 200 times more frequently on hemimethylated substrates than on unmethylated ones (RAZIN 1984). This illustration of the properties of maintenance methylation applies equally well to regions of chromosomes—such as genes—which are differentially demethylated and remethylated compared to surrounding DNA.

As a consequence of the properties of the methylation system, the pattern and extent of methylation are

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

inherently dynamic. Methylation can be rapidly lost in the absence of maintenance methylation, while *de novo* methylation is only slowly imposed. Previous discussions of the dynamics of methylation have employed a qualitative approach and have been limited to modeling a few cell divisions. Often the focus has been a specific gene or transcribed sequences in a particular cell type: these sequences are typically hypomethylated compared to bulk DNA presumably because transcriptional factors prevent or slow maintenance methylation. In this paper we explore the kinetics of methylation and demethylation in the whole genome through numerical simulations which cover many cell generations. Of course, at a particular substrate site, methylation is a discrete event. At any substrate C residue, a methyl group is either present or absent; in a duplex at any particular site, the two strands are either unmethylated, hemi-methylated or fully methylated. Considering the population of all substrate sites within a cell, or the individual substrate sites within a population of cells, we can examine the proportion of methylated sites as a continuous function.

## MATERIALS AND METHODS

**Recursive model of methylation:** We take as our population the group of cytosines which can be methylated (an equivalent population consists of one site observed in many independent cells). We assume that this population is very large so that random drift may be ignored. The cell cycle with respect to replication and methylation of cytosine sites is assumed to occur as in Figure 1. Because DNA modification occurs after replication and because the newly produced daughter strands are initially unmethylated, all methylated sites are on the template strand immediately after replication. At the time of methylation, a proportion ($\alpha$) of the hemi-methylated residues becomes homomethylated by maintenance methylation. Furthermore, a proportion ($\beta$) of the unmethylated sites becomes homomethylated by *de novo* methylation. These properties can be described by the following equations.

Let $X_n$, $Y_n$, and $Z_n$ be the proportion of homomethylated, hemi-methylated, and unmethylated sites respectively in generation $n$, where census takes place after DNA methylation but before DNA replication. Note that $X_n + Y_n + Z_n = 1$.

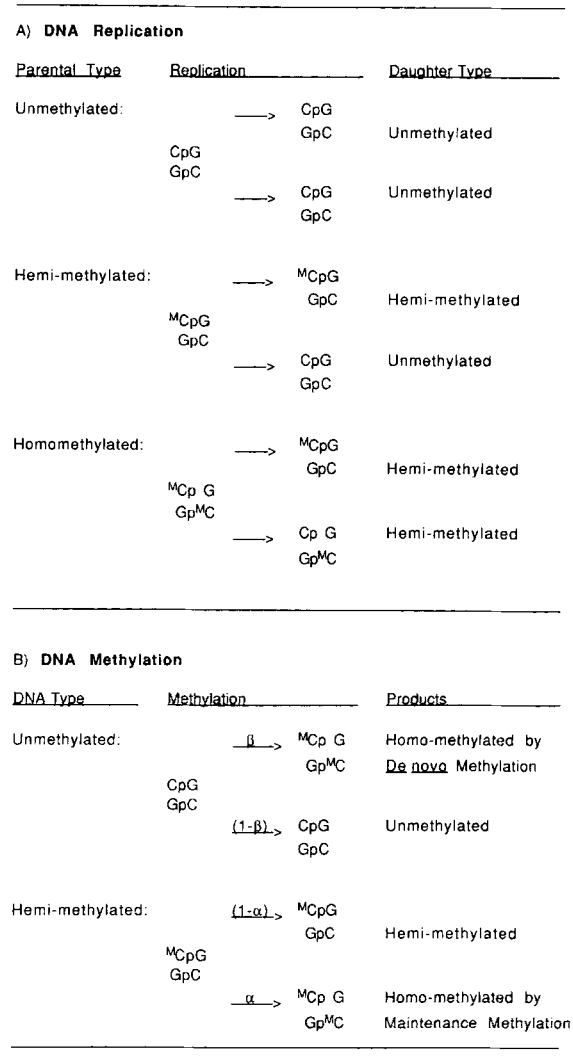



FIGURE 1.—Schematic representation of DNA replication (A) and methylation (B) at CpG substrate sites.

| Generation $n$ | Proportion of Homomethylated $X_n$ | Hemi-methylated $Y_n$ | Unmethylated $Z_n$ |
|---|---|---|---|
| After replication | 0 | $2X_n + Y_n$ | $2Z_n + Y_n$ |
| After methylation | $\alpha(2X_n + Y_n)$ $+\beta(2Z_n + Y_n)$ | $(1 - \alpha)(2X_n + Y_n)$ | $(1 - \beta)(2Z_n + Y_n)$ |

Thus in the next cell generation ($n$+1):

$$X_{n+1} = \frac{2\alpha X_n + Y_n(\alpha + \beta) + 2\beta Z_n}{2} \quad (1a)$$

$$Y_{n+1} = \frac{2(1 - \alpha)X_n + (1 - \alpha)Y_n}{2} \quad (1b)$$

$$Z_{n+1} = \frac{(1 - \beta)Y_n + 2(1 - \beta)Z_n}{2} \quad (1c)$$

where dividing by 2 normalizes the equations so that $X_{n+1} + Y_{n+1} + Z_{n+1} = 1$.

In the APPENDIX, we show that the following equilibrium point (denoted by $\hat{X}$, $\hat{Y}$, $\hat{Z}$) is reached by a population from any starting position (globally stable) under the model described by equations (1):

$$\hat{X} = \frac{\beta(1 + \alpha)}{(1 - \alpha + 2\beta)} \quad (2a)$$

$$\hat{Y} = \frac{2\beta(1 - \alpha)}{(1 - \alpha + 2\beta)} \quad (2b)$$

$$\hat{Z} = 1 - \hat{X} - \hat{Y} = \frac{(1 - \alpha)(1 - \beta)}{(1 - \alpha + 2\beta)}. \quad (2c)$$

**Estimation of cell generations during the maize life cycle:** As shown in Table 1, the maize life cycle is conveniently divided into four stages. For each of these the number of cell generations can be calculated with reasonable confidence. For the first stage, the period between the zygote and the formation of the globular embryo, the calculation of the number of cell generations is based on the observation that about 10,000 cells are present when the apical meristem is established and that this meristem forms from just a few founder cells (McDaniel and Poethig 1988; Kiesselbach 1949). Possible sources of error in this analysis and their impact on the determination of the number of cell divisions include [1] the assumption that all embryonic cells continued synchronous mitoses to this point in development, and [2] the estimate of final cell number at the time of meristem formation. Stage two involves the proliferation of the founder cells to form the apex. Two methods have been used to analyze the number of cells in the maize shoot apex: counting cells through a serially sectioned apex and clonal analysis. Both methods agree that there are about 335 cells in the apex at the time of embryo maturation (McDaniel and Poethig 1988). Of this group of cells, however, only 2 to 4 cells are fated to form each ear (Coe and Neuffer 1978; Poethig *et al.* 1986; McDaniel and Poethig 1988). What is unknown is how many cells found the apex. Consequently, the major assumption is how many cells of the globular embryo proliferate to establish the mature apex; here we have assumed that 2 to 8 cells found the apex so only 5 to 8 cell divisions are required to form the complete apex. Proliferation of the ear founding cells to make an immature ear during stage three was estimated from the recovery of total cellular DNA from postmitotic, immature cobs. DNA was extracted as previously described (Walbot and Warren 1988). Cell number was calculated by dividing the yield of DNA by the diploid genome size of 10 pg (Hake and Walbot 1980). Sources of error in this measurement include loss of DNA during sample preparation, giving as much as a twofold error in underestimating cell number, and the possibility that cells are 4C rather than 2C, a factor which would overestimate cell number by a factor of two. The fourth stage, gametogenesis, requires 3 mitotic divisions to produce the egg cell. The summed estimate of mitotic divisions in the life cycle from zygote to zygote through the megagametophyte is 47–51; we assume that analysis of the microgametophyte lineage would yield a similar result as it is known that the tassel is derived from approximately three cells in the apex (McDaniel and Poethig 1988) and the tassel yields a similar amount of DNA as the cob (V. Walbot, unpublished data).

TABLE 1

**Number of cell divisions through stages of the maize life cycle**

| Stage | Cell divisions |
|---|---|
| 1. Zygote to globular embryo | 13 |
| 2. Founder cells (2–8) to shoot apex | 5–8 |
| 3. Ear progenitor cells (2–4) to ear | 26–27 |
| 4. Gametogenesis (ear) | 3 |
| Total | 47–51 |

The number of cell generations was estimated as described in MATERIALS AND METHODS.

## RESULTS

**Assumptions:** Several methods have been used to measure the level of methylation in eukaryotic DNA: direct determination of 5-methylcytosine content from hydrolyzed DNA (Razin, Cedar and Riggs 1984, and references therein) and calculation of methylation from the shift in $T_m$ and buoyant density of the DNA sample (for maize, see Hake and Walbot 1980). The results of these measurements are often presented as the percent of all nucleotides which are 5-methylcytosine. In order to understand the kinetics of methylation, it is more useful to measure the proportion of methylated cytosine in the population of cytosine residues which can be methylated. Given the proportion of methylated C to all cytosines, the proportion of methylated C to substrate cytosines can be calculated by dividing by the proportion of substrate cytosines to all cytosine residues. For vertebrates, although the overall level of C methylation is low, 0.7–2.8%, the percent substrate C methylation is high, ranging from 75% to 85% in adult mammalian tissues (compilation in Razin 1984). High substrate site methylation results from the relative paucity of substrate CpG sites in these genomes. Because the bias in *de novo vs.* template-based methylation has been measured in several tissues of the mouse, we will use it as a typical vertebrate, assuming an equilibrum value of 80% substrate cytosine methylation. In higher plants, in contrast to vertebrates, the overall level of methylation is high: given that about 25% of all C residues in genomes of 45–50% G+C content are methylated (Ergle and Katterman 1961), roughly 6% of all residues are 5-methylcytosine. There is only a slight bias, however, against CpG in sequenced plant DNA (Boudraa and Perrin 1987). Hence, for maize in which 6.5–7% of all substrate sites are 5-methylcytosine (Ergle and Katternman 1961; Hake and Walbot 1980) and which has a genome that is 49% G+C, the estimated level of methylation among substrate cytosines is about 65%, less than that in vertebrates.

**Comparison of the kinetics of methylation and demethylation in mouse and maize:** To understand the features of the recursion equations, we present sets of graphs which illustrate how changes in the parameters affect methylation kinetics. For mouse DNA we have assumed an 80% substrate C methylation level and for maize, 65%. We further assume that these values measure the overall equilibrium level of homomethylated DNA ($\hat{x}$). Alternatively these values might reflect the combined levels of homomethylated and hemimethylated sites ($\hat{x}$ and $\hat{y}$). This alternative assumption does not change the results to any significant extent, however, because for all biologically reasonable estimates of $\alpha$ (specifically for $\alpha$ much greater than 0.33), the proportion of homomethylated sites is much greater than the proportion of hemimethylated sites. In fact, if maintenance methylation is at least 91% effective ($\alpha > 0.91$) then the proportion of homomethylated sites is more than an order of magni-

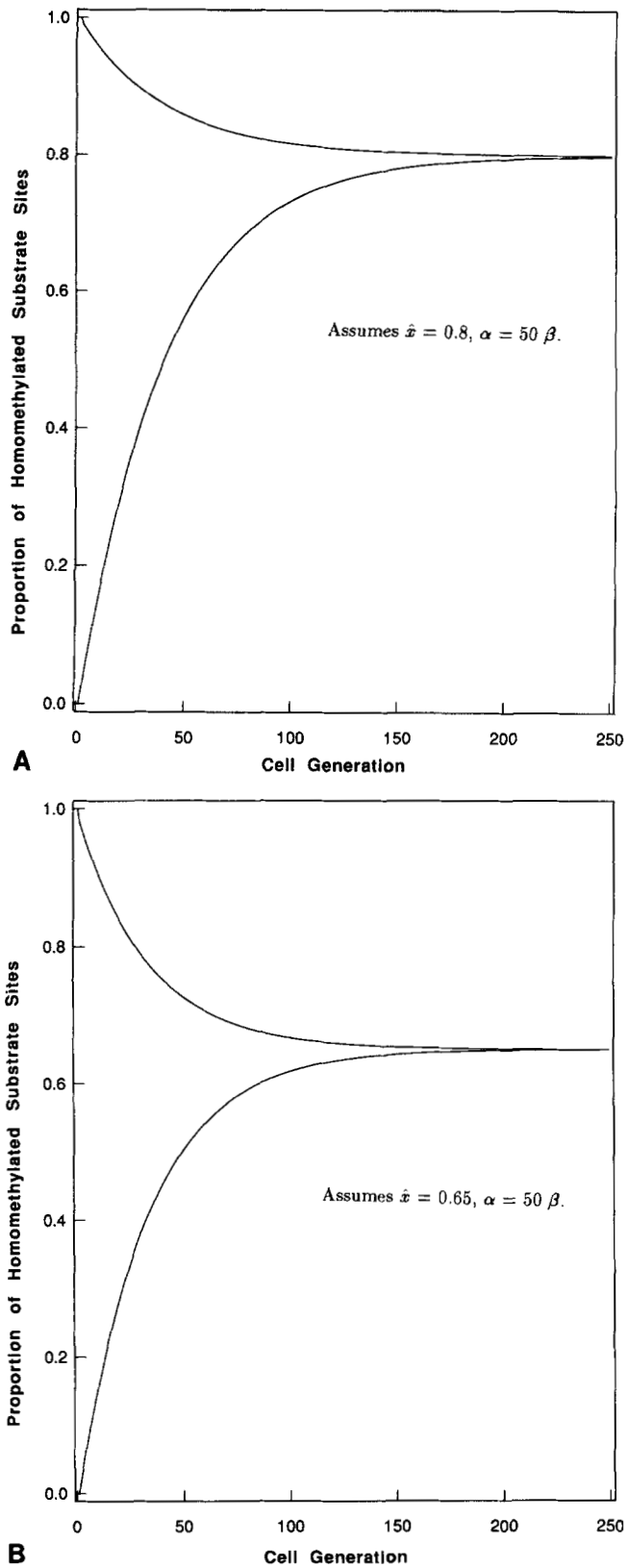



FIGURE 2.—The approach to an equilibrium level of methylation at mouse (A) and maize (B) substrate sites. For mice we assumed that ($\alpha = 50\beta$) and that the equilibrium level of methylation is 80% so that we could calculate $\alpha$ and $\beta$ from equations (2). For maize, we also assumed that ($\alpha = 50\beta$) but that the equilibrium level of methylation is only 65%. Top curves show the passive process of demethylation from a fully methylated state while bottom curves show the acquisition of methylation from a fully unmethylated state.

tude greater than the proportion of hemimethylated sites. From Equation 2a, given the equilibrium level of methylation, $\alpha$ and $\beta$ can be calculated given their relative values.

A wide range of empirical measurements of the relative efficiency of mammalian methylases on hemimethylated or unmethylated subtrates exist (RAZIN 1984). $\alpha = 100\beta$ is an upper limit; $\alpha = 10\beta$ is a lower limit and reflects a relatively higher efficiency of *de novo* methylation. First, if the fidelity of maintenance methylation ($\alpha$) is set at 50 times that of *de novo* methylation ($\beta$) the approach to equilibrium is as shown in Figure 2A for mouse and Figure 2B for maize. The approach to equilibrium through either demethylation or *de novo* methylation is slightly less rapid in the mouse, with a higher equilibrium value. Second, by varying the relative values of $\alpha$ and $\beta$ between the measured limits for $\alpha$ and $\beta$, a family of curves can be generated. The shapes of the curves describe the approach to an equilibrium at $\hat{x} = 0.80$ for mouse (Figure 3A) and $\hat{x} = 0.65$ for maize (Figure 3B). The upper panels of each figure illustrate the kinetics of demethylation, and the lower panels the kinetics of *de novo* methylation.

For all values that have been measured in mammalian cells in culture (RAZIN *et al.* 1985), the most striking feature of the dynamics is that many cell generations are required to reach 50% modification status given a low or no initial modification. Note that by increasing the relative efficiency of *de novo* methylation, the 50% modification level is more quickly reached. This implies that the system more rapidly approaches equilibrium when $\alpha/\beta$ is decreased. For example, with maize, while it takes 197 generations to move from no methylation to 95% of the equilibrium value ($\hat{x} = 0.65$) for $\alpha = 100\beta$, it will only take 41 (~197/5) generations for $\alpha = 20\beta$.

**Number of cell generations in a maize life cycle:** To understand the impact that an alteration in the rate of methylation or demethylation might have on an individual, we need to compare the number of cell generations in a life cycle to the number required to reach equilibrium. As shown in Table 1, we calculate that in maize there are approximately 50-cell generations from zygote to zygote. Although these calculations rest on a number of assumptions (described in MATERIALS AND METHODS) 40-cell divisions is too few and 60 divisions is too many to accommodate what is known about the ontogeny of this plant.

Considering 50-cell generations per life cycle as reasonable, it is clear that the approach to equilibrium from a fully methylated or unmethylated state re-

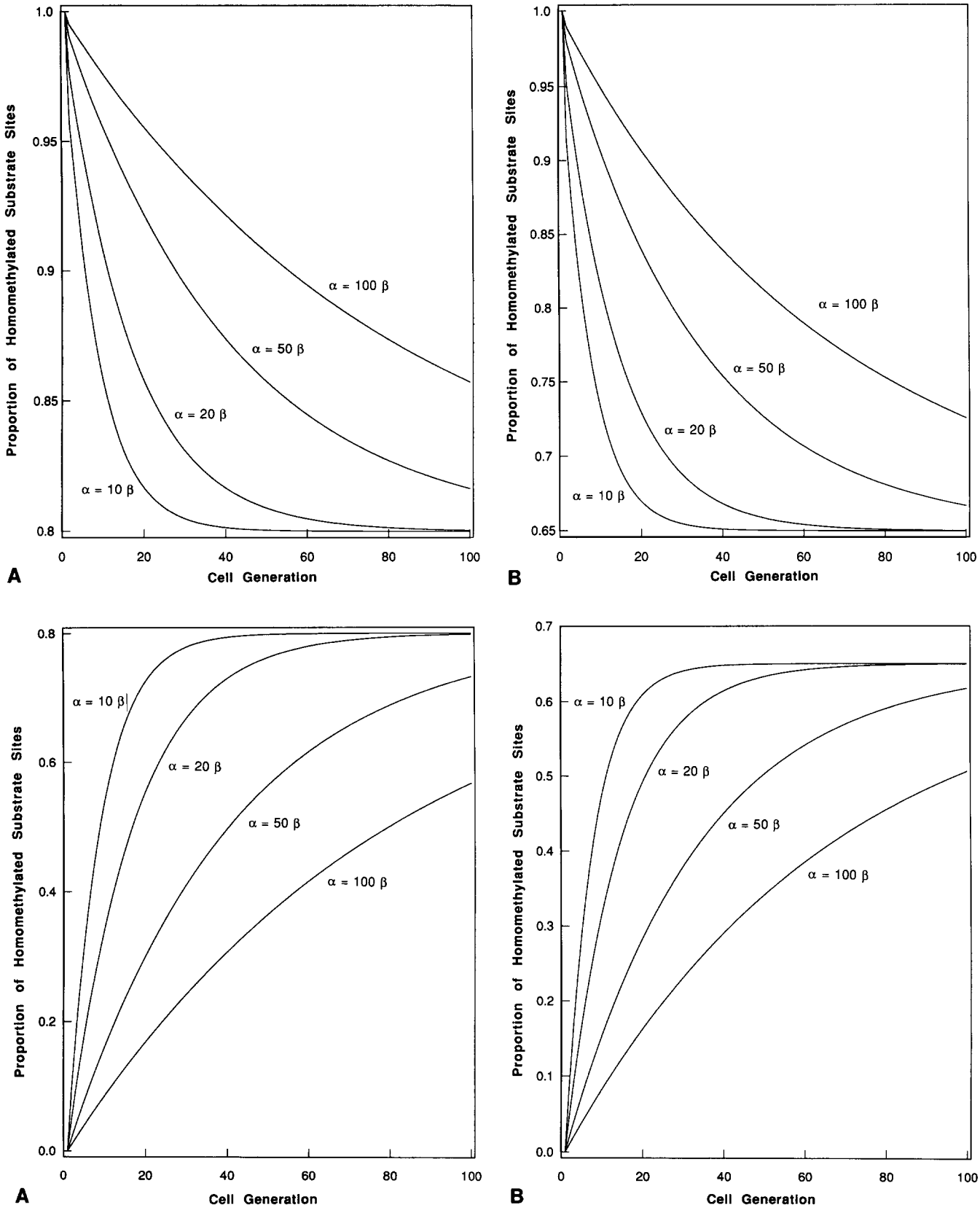
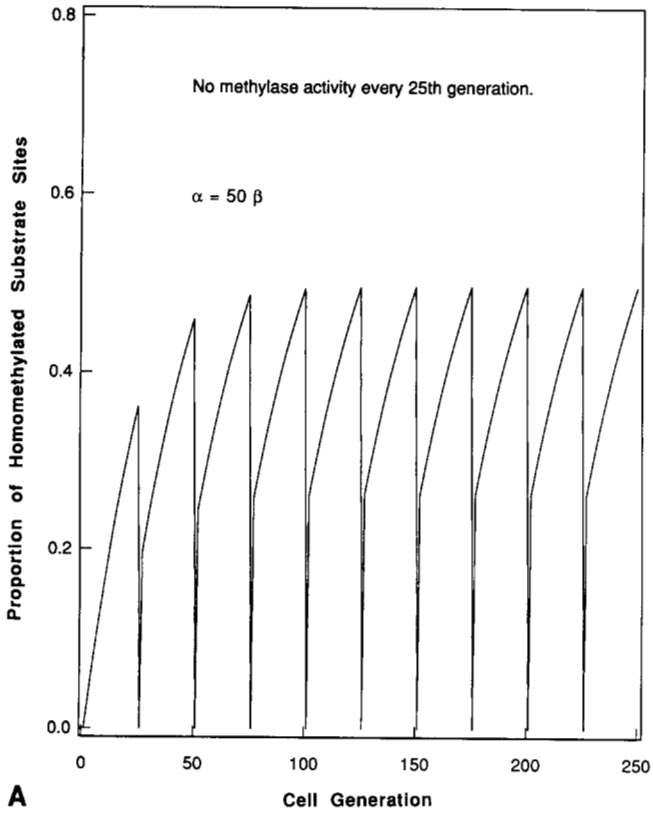



FIGURE 3.—The approach to an equilibrium level of methylation at substrate sites in mouse (A) [$\hat{x} = 0.80$] and maize (B) [$\hat{x} = 0.65$]. For both we varied the ratio of $\alpha$ to $\beta$, using ($\alpha = 10\beta$), ($\alpha = 20\beta$), ($\alpha = 50\beta$), and ($\alpha = 100\beta$). Top panels show demethylation while bottom panels show methylation.



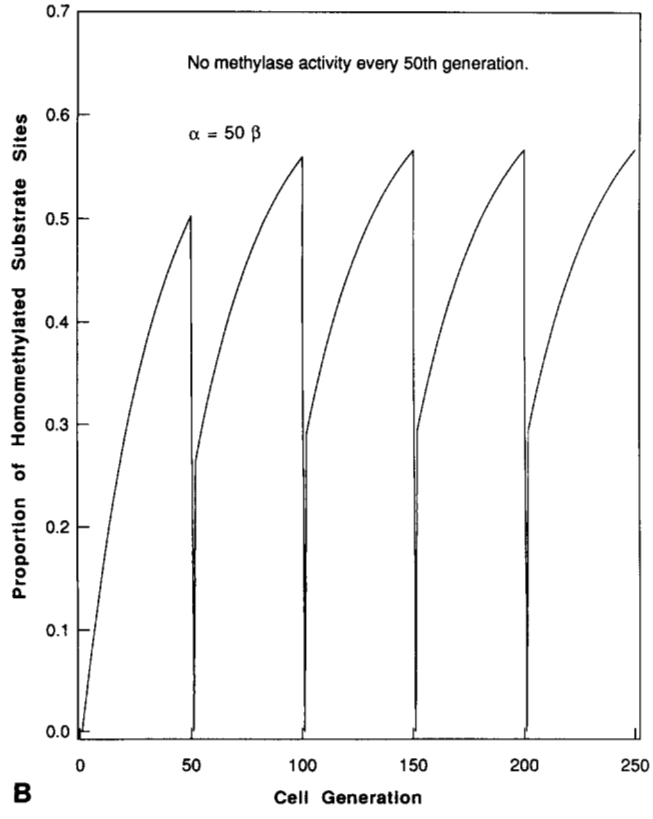




FIGURE 4.—The approach to an equilibrium level of methylation at substrate sites in mouse (A) [$\hat{x} = 0.80$] and maize (B) [$\hat{x} = 0.65$] when there is periodic failure of methylase activity. For both we let ($\alpha = 50\beta$). Note that in the first generation all substrate sites are unmethylated. Subsequently, there is a total absence of new methylation once in every life cycle which lasts approximately 50 cell generations in maize and 25 cell generations in mice.

quires several plant life cycles in the case where the rate of *de novo* methylation is low compared to the fidelity of maintenance methylation (Figure 3B, see $\alpha = 100\beta$ curve). When the rates of demethylation and *de novo* methylation are more similar, about one life cycle is required to near the equilibrium value (Figure 3B, see $\alpha = 10\beta$ curve).

Returning to Figure 2B in which $\alpha = 50\beta$, if there is an interruption in methylation in just one of every 50 cell generations, so that cells are produced in which the newly replicated DNA undergoes neither maintenance nor *de novo* methylation, the approach to equilibrium is disrupted (Figure 4B). Note that in this generation, there are no homomethylated sites. In subsequent cell generations with *de novo* and template-based methylation reimposed, the approach to equilibrium will follow the usual kinetics. In this example, an oscillation is established between very low and subequilibrium levels of methylation ($x < 0.57$) with an imposed periodicity of 50 cell generations. Thus even occasional failures of maintenance methylation (here once in 50 cell generations) serve to keep the population away from the true equilibrium ($\hat{x} = 0.65$).

For the mouse, estimation of the number of cell divisions in the life cycle is more difficult, because the lineage analysis is incomplete, however, MAYNARD SMITH (1989) has estimated the number to be 20–30. We have evaluated simulations of periodic interruptions in methylation (Figure 4A) based on data from mouse cells in which $\hat{x} = 0.8$ and $\alpha = 50\beta$ (cited in RAZIN 1984). Here, 36% substrate methylation is reached within 25 cell generations from a point of no homomethylation or hemi-methylation. Even starting from a point with some hemi-methylation but no homomethylation (subsequent troughs in Figure 4A), at most 49.7% substrate methylation is observed. Hence, interruption in methylation once in every 25 cell divisions would create an oscillation that lasts approximately one life cycle and which keeps the level of substrate methylation below 62.1% of its equilibrium value ($\hat{x} = 0.80$).

## DISCUSSION

The global level of DNA methylation in a genome depends on the behavior of the maintenance methylation system. As far as is known, the global level within a particular genome at a particular stage in the life cycle is fairly constant. That is, in mammals both methylation pattern and extent are heritable (SILVA and WHITE 1988). This is true even though methylation can be different in genomes transmitted through sperm and egg (SWAIN, STEWART and LEDER 1987; REIK *et al.* 1987; SAPIENZA *et al.* 1987). The absolute

level of 5-methylcytosine, however, differs widely among those eukaryotes with 5-methylcytosine modification. In higher plants, methylation of 25% of the total cytosine residues is common, and for maize this means that about 65% of the substrate sites CpG and CpXpG are methylated. In mammals, a lower fraction of the total cytosine residues are methylated. Nevertheless, because of the paucity of CpG dinucleotides, 75–85% of the substrate C residues are methylated (RAZIN 1984). Thus the level of maintenance methylation is relatively higher in mammalian DNA. The higher level of methylation observed in mammals can only be maintained by a more efficient system of methylation than is present in higher plants.

Equations have been developed to explore the kinetics of methylation and demethylation required to achieve an equilibrium; measured values for the bias in template-based methylation compared to *de novo* methylation were used to generate families of curves describing methylation kinetics. Starting from unmethylated DNA about 22 cell generations ($\alpha = 10\beta$) to 200 cell generations ($\alpha = 100\beta$) are required to be within 95% of the equilibrium in an organism such as maize which attains 65% substrate C methylation. In contrast, in a mammal with an equilibrium value of 80% methylation, from 25 cell generations ($\alpha = 10\beta$) to 240 cell generations ($\alpha = 100\beta$) are required to reach 95% of this value.

An important conclusion from these projections is that gain and loss of methylation at particular sites is expected and that this process should reach an equilibrium. The probability that an unmethylated site will become methylated depends on the capacity for *de novo* methylation and is generally low. The probability that loss of methylation will occur depends on the error rate of template-dependent, maintenance methylation which is also generally low. Conversely the probability that a hemi-methylated site will become unmethylated is about equal to the probability that it will become homomethylated; each of which is about 1/2. In any generation just before DNA replication, however, few sites are hemi-methylated and therefore unstable with regard to methylation status. The frequency of methylation in the population of all sites in the DNA of a single cell or the extent of methylation of a particular site in a population of cells should approach equilibrium. Hence sequences which consistently fail to approach equilibrium are probably controlled by different rates of maintenance and *de novo* methylation. Our model can be used to study such specific sequences as long as sufficient data are available. For example, given a curve which shows the change in methylation over time at a specific site in many cells, all the parameters in the model can be estimated and compared with those used in the global genome analysis reported here. Deviations from the expected rate of return to equilibrium would signify specific regulation of the sequences under study.

The usual explanation for differential methylation of active *vs.* inactive genes is that transcriptional factors or other proteins bound to DNA block methylation; this hypothesis has not been critically tested. Indeed, the role DNA methylation plays in gene expression is still controversial. Demethylation of sites within the coding region and the 5′ flanking region of genes is often correlated with induced or tissue-specific expression (CEDAR 1988), but whether this is a cause or consequence of expression has been difficult to establish. Experimentally, methylation is implicated as a determining factor by studies in which methylated genes were introduced into mammalian cells and were expressed at lower levels than unmethylated controls (YISRAELI *et al.* 1988). On the other hand, transcription of some genes has been shown to cease prior to an increase in methylation (ENVER *et al.* 1987; LOCK, TAKAGI and MARTIN 1987) suggesting that the higher methylation level is a consequence rather than a cause of the change in transcription. If protein factors are involved in altering methylation, then the affinity of individual proteins for methylated *vs.* unmethylated binding sites should also be important in setting the methylation status. One transcriptional factor with a lower affinity for methylated substrate DNA has been detected (WATT and MOLLOY 1988) as has one factor in which the methylation status of its substrate site is unimportant (HÖLLER *et al.* 1988). Clearly more data are required to evaluate the impact of bound proteins on the kinetics of DNA methylation and the consequences this modification has on the ability of proteins to interact with their DNA substrates.

During development there may be complex regulation of the rates of methylation, as evidenced by tissue-specific patterns of gene methylation. A simplistic mode of regulation, for example, would be to have normal methylation rates during each cell generation of a life cycle except one in which maintenance methylation is absent. Such an event could be natural or induced, but would certainly keep the overall level of methylation below equilibrium levels. Given the slow kinetics of methylation in maize with $\alpha = 50\beta$, we calculate that failure of methylation at a particular site during just one cell generation in every life cycle of 50 cell generations would be sufficient to keep methylation at that site (in a population of cells) below 57% during the life cycle. Methylation is often measured by the sensitivity of DNA to restriction enzyme digestion; if a particular site is methylated in 57% of the sample, this can be readily detected. When methylation is lower, in the range of 10%, however, the presence of the methylated form might go undetected.

The stimulus for this project was our previous report that there is an increase in methylation associated with the loss of activity of the *Mutator* transposable elements of maize (CHANDLER and WALBOT 1986). In

this case, transposase is hypothesized to be the trans-acting protein which allows the numerous *Mu* elements in the genome to be maintained in a hypomethylated state (WALBOT *et al.* 1988). When *Mutator* activity is lost, the level of methylation in *Mu* elements is similar to bulk DNA, that is, the system reaches equilibrium (WALBOT *et al.* 1988). Viewed at the plant level, loss of *Mutator* activity is measured by loss of somatic instability at a mutable receptor allele; this process typically takes place in several steps extending over several life cycles. In an active line, densely spotted kernels created by the excision of a *Mu* element from a reporter allele in the anthocyanin pigment pathway are found; as the line loses activity first fewer spots per kernel are noted and then fewer spotted kernels than expected from segregation of the reporter allele are found. These observations indicating that activity changes can require several plant generations support the simulations reported here which demonstrate that multiple life cycles are required to reach methylation equilibrium for $\alpha > 20\beta$. In addition to *Mutator*, differential methylation of other transposable elements in maize is similarly correlated with their activity status (SCHWARTZ and DENNIS 1986; CHOMET, WESSLER and DELLAPORTA 1987; BANKS, MASSON and FEDEROFF 1988). For example, active *Suppressor-mutator* elements are hypomethylated; when activity changes through intermediate activity states, element remethylation and activity loss occur over several plant generations (FEDOROFF and BANKS 1988). Genetic evidence suggests that activation of cryptic *Spm* elements also requires several generations (FEDOROFF 1989). An interesting prediction of our models, is that factors which suppress methylation as rarely as once per maize life cycle could significantly affect the methylation status of maize transposable elements and such changes could be important in setting the activity state.

In mammals which have a lower percent of overall cytosine methylation, but a higher percentage of methylated substrate sites, the approach to equilibrium is slightly slower for the same ratio of $\alpha:\beta$. Rare events in a life cycle or in a stem cell lineage could also produce hypomethylation of particular sites in a population of cells. The level of genome methylation in mammals is thought to remain near the equilibrium value in most cell types, the major exception being hypermethylated sperm DNA (SANFORD *et al.* 1987). Nevertheless, methylation patterns within or near genes are often tissue or cell-type specific. Unmethylated CpG-rich regions are often found near the 5′ terminus of both vertebrate (BIRD 1986) and higher plant (ANTEQUERA and BIRD 1988) genes. Recall that the inherent error rate of the methylation is low, mainly because a template strand is utilized (Figure 1). Thus, loss of methylation at particular sites during a few cell divisions must occur from programmed, local failures in the methylation process. One possibility is that proteins bound to DNA will block maintenance methylation. The specificity of the methylation pattern would therefore not be random but rather would depend on the specificity of DNA:protein interaction at individual sites in the genome.

We thank MARC FELDMAN and ANNE BRITT for useful comments on the manuscript. Research support was provided by National Institutes of Health grant GM32422 to V.W. S.P.O. was a recipient of an undergraduate research award from the Genetics Society of America.

## LITERATURE CITED

ANTEQUERA, F., and A. P. BIRD, 1988 Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J. **7:** 2295–2299.

BANKS, J. A., P. MASSON and N. FEDOROFF, 1988 Molecular mechanisms in the developmental regulation of the maize *Suppressor-mutator* transposable element. Genes Dev. **2:** 1364–1380.

BIRD, A. P., 1986 CpG islands and the function of DNA methylation. Nature **321:** 209–213.

BOUDRAA, M., and P. PERRIN, 1987 CpG and TpA frequencies in the plant system. Nucleic Acids Res. **15:** 5729–5737.

CEDAR, H., 1988 DNA methylation and cell activity. Cell **53:** 3–4.

CHANDLER, V. L., and V. WALBOT, 1986 DNA modification of a transposable element of maize correlates with loss of activity. Proc. Natl. Acad. Sci. USA **83:** 1767–1771.

CHOMET, P. S., S. WESSLER and S. L. DELLAPORTA, 1987 Inactivation of the maize transposable element *Activator* (*Ac*) associated with DNA modification. EMBO J. **6:** 295–302.

COE, E. H., and M. G. NEUFFER, 1978 Embryo cells and their destinies in the corn plant, pp. 113–129 in *The Clonal Basis of Development*, edited by S. SUBTELNY and I. M. SUSSEX. Academic Press, New York.

ENVER, T., J. ZHANG, T. PAPAYANNOPOULOU and G. STAMATOYANNOPOULOS, 1987 DNA methylation: a secondary event in globin gene switching? Genes Dev. **2:** 698–706.

ERGLE, D. R., and F. R. H. KATTERMAN, 1961 Deoxyribonucleic acid of cotton. Plant Physiol. **36:** 811–815.

FEDOROFF, N., 1989 The heritable activation of *cryptic Suppressor-mutator* elements by an active element. Genetics **121:** 591–608.

FEDOROFF, N. V., and J. A. BANKS, 1988 Is the *Suppressor-mutator* element controlled by a basic developmental regulatory mechanism? Genetics **120:** 559–577.

GRUENBAUM, Y., T. NAVEH-MANY, H. CEDAR and A. RAZIN, 1981 Sequence specificity of methylation in higher plant DNA. Nature **292:** 860–862.

HAKE, S., and V. WALBOT, 1980 The genome of *Zea mays*, its organization and homology to related grasses. Chromosoma **79:** 251–270.

HÖLLER, M., G. WESTIN, J. JIRICNY and W. SCHAFFNER, 1988 Spl transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev. **2:** 1127–1135.

KIESSELBACH, T. A., 1949 The structure and reproduction of corn. Nebr. Agric. Exp. Stn. Res. Bull. **161:** 1–96.

LOCK, L. F., N. TAKAGI and G. R. MARTIN, 1987 Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell **48:** 39–46.

MAYNARD SMITH, J., 1989 *Evolutionary Genetics*. Oxford University Press, Oxford.

MCDANIEL, C. N., and R. S. POETHIG, 1988 Cell-lineage patterns in the shoot apical meristem of the germinating maize embryo. Planta **175:** 13–22.

MESSER, W., and M. NOYER-WEIDNER, 1988 Timing and targeting: the biological functions of Dam methylation in *E. coli*. Cell **54:** 735–737.

POETHIG, R. S., COE, E. H., JR. and M. M. JOHRI, 1986 Cell lineage patterns in maize embryogenesis: a clonal analysis. Dev. Biol. **117:** 581–594.

RAZIN, A., 1984 DNA methylation patterns: Formation and biological functions, pp. 127–146 in *DNA methylation, Biochemistry and Biological Significance*, edited by A. RAZIN, H. CEDAR and A. D. RIGGS. Springer Verlag, New York.

RAZIN, A., H. CEDAR and A. D. RIGGS, 1984 Introduction and general overview, pp. 1–10 in *DNA Methylation, Biochemistry and Biological Significance*, edited by A. RAZIN, H. CEDAR and A. D. RIGGS. Springer Verlag, New York.

REIK, W., A. COLLICK, M. L. NORRIS, S. C. BARTON and M. A. SURANI, 1987 Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature **328:** 248–251.

SANFORD, J. P., H. J. CLARK, V. M. CHAPMAN and J. ROSSANT, 1987 Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev. **1:** 1039–1046.

SAPIENZA, C., A. C. PETERSON, J. ROSSANT and R. BALLING, 1987 Degree of methylation of transgenes is dependent on the gamete of origin. Nature **328:** 251–254.

SCHWARTZ, D., and E. A. DENNIS, 1986 Transposase activity of the *Ac* controlling element in maize is regulated by its degree of methylation. Mol. Gen. Genet. **205:** 476–482.

SILVA, A. J., and R. WHITE, 1988 Inheritance of allelic blueprints for methylation patterns. Cell **54:** 145–152.

SWAIN, J. L., T. A. STEWART and P. LEDER, 1987 Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50:** 719–727.

WALBOT, V., A. BRITT, K. LUEHRSEN, M. MCLAUGHLIN and C. A. WARREN, 1988 Regulation of *Mutator* activities in maize, pp. 121–135 in *Plant Transposable Elements*, edited by O. E. NELSON, JR. Plenum Press, New York.

WALBOT, V., and C. WARREN, 1988 Regulation of *Mu* element copy number in maize lines with an active or inactive *Mutator* transposable element system. Mol. Gen. Genet. **211:** 27–34.

WATT, F., and P. L. MOLLOY, 1988 Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. **2:** 1136–1143.

YISRAELI, J., D. FRANK, A. RAZIN and H. CEDAR, 1988 Effect of *in vitro* DNA methylation on β-globin gene expression. Proc. Natl. Acad. Sci. USA **85:** 4638–4642.

Communicating editor: S. D. TANKSLEY

## APPENDIX

### Equilibrium of the equations describing methylation patterns

Given that $Z_n = 1 - X_n - Y_n$, we can rewrite the system of equations (1) in matrix form:

$$\begin{pmatrix} X_{n+1} \\ Y_{n+1} \end{pmatrix} = \begin{pmatrix} \alpha - \beta & \dfrac{\alpha - \beta}{2} \\ 1 - \alpha & \dfrac{(1 - \alpha)}{2} \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} + \begin{pmatrix} \beta \\ 0 \end{pmatrix}$$

which is symbolically denoted as $\underline{u}_{n+1} = \mathbf{A}\underline{u}_n + b$. The characteristic polynomial of the matrix $\mathbf{A}$ is $\lambda^2 - \tfrac{1}{2}(1 + \alpha - 2\beta)\lambda = 0$. The two eigenvalues of $\mathbf{A}$ are thus $\lambda_1 = 0$ and $\lambda_2 = \tfrac{1}{2}(1 + \alpha - 2\beta)$. For $0 \leq \alpha, \beta \leq 1$ except $\alpha = 1$ and $\beta = 0$, both eigenvalues are strictly less than unity in absolute value. Thus, with this system of equations, $\underline{u}$ converges globally to $(\mathbf{I} - \mathbf{A})^{-1}b$. In other words, an equilibrium is reached by the population so that $X_{n+1} = X_n = \hat{X}$, $Y_{n+1} = Y_n = \hat{Y}$, and $Z_{n+1} = Z_n = \hat{Z}$ where

$$\begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = \begin{pmatrix} 1 - \alpha + \beta & \dfrac{\beta - \alpha}{2} \\ \alpha - 1 & \dfrac{1 + \alpha}{2} \end{pmatrix}^{-1} \begin{pmatrix} \beta \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{1 + \alpha}{1 - \alpha + 2\beta} & \dfrac{\alpha - \beta}{1 - \alpha + 2\beta} \\ \dfrac{2(1 - \alpha)}{1 - \alpha + 2\beta} & \dfrac{2(1 - \alpha + \beta)}{1 - \alpha + 2\beta} \end{pmatrix} \begin{pmatrix} \beta \\ 0 \end{pmatrix}.$$

Hence

$$\hat{X} = \frac{\beta(1 + \alpha)}{(1 - \alpha + 2\beta)} \tag{2a}$$

$$\hat{Y} = \frac{2\beta(1 - \alpha)}{(1 - \alpha + 2\beta)} \tag{2b}$$

so that

$$\hat{Z} = 1 - \hat{X} - \hat{Y} = \frac{(1 - \alpha)(1 - \beta)}{(1 - \alpha + 2\beta)}. \tag{2c}$$

This equilibrium is reached under all conditions except for the trivial case when $\alpha = 1$ and $\beta = 0$. In this special case, maintenance methylation is perfectly faithful and *de novo* methylation does not occur. Here the equilibrium composition depends on the initial state $(X_0, Y_0, Z_0)$ of the population:

$$\hat{X} = X_0 + \tfrac{1}{2}Y_0 \qquad \hat{Y} = 0 \qquad \hat{Z} = \tfrac{1}{2}Y_0 + Z_0.$$