Dedicated to the 45th anniversary of the discovery of enzymatic DNA methylation

Replicative DNA Methylation as Generator of Mutations

Alexander L. Mazin*

UDC 577.21

Translated from Molecular Biology (Moscow), Vol. 27, No. *5,* pp. 965-979, September-October, 1993.
Original article submitted February 9, 1993.

In this report, various aspects of cytosine methylation are analyzed from the standpoint of a unifying functional model. It treats this system as a special mechanism for generating the m^5C-to-T+C transition mutations in the genome. This mechanism is analyzed in detail: (i) m^5C deamination during replicative DNA methylation, (ii) repair of occurring G:T mispairs, (iii) evolvement of the new hemimethylated sites in DNA, and (iv) postreplicative methylation of these sites. According to the model proposed, the age-related loss of most m^5C residues from DNA has been shown for both animal tissues and cell lines. The *CG-to-TG+CA mutations disproportionately contribute to general mutagenesis, and may be a main cause of many human hereditary diseases and cancer. It has been concluded that DNA methylation may be the genetically programmed generator for accumulating mutations responsible for cell aging and cancer.

Key words: DNA methylation; 5-methylcytosine (m^5C, *C); m^5C deamination; m^5C-to-T+C substitutions; age-related m^5C loss; CG-mutagenesis; cell aging

CONTENTS

1. INTRODUCTION

45 years ago 5-methylcytosine (m^5C, *C) was found as an additional minor base in animal [1] and later in plant DNA [2]. At first m^5C was supposed to be incorporated by DNA polymerase like four main nucleotides. But soon it became evident that cytosine residues were methylated already in polynucleotide chains by a specific enzyme, DNA(cytosine-C5) methyltransferase (MTase, EC 2.1.1.37) [3]. This enzyme transfers the CH_3 groups from the universal donor S-adenosyl-L-methionine (AdoMet) to the C5 atom of cytosine, producing m^5C residues in *CG and *CNG sites of DNA (N=A,G,C,T) [4, 5]. Such DNA methylation system was found in most eukaryotes but far from all species [6].

Discovered before the Watson-Crick double helix, enzymatic DNA methylation still remains one of the oldest mysteries in molecular biology. This system

* A.N. Belozersky Institute of Physico-Chemical Biology, M.V. Lomonosov State University, Moscow, 119899.

0026-8933/93/2705-0593S12.50

seems to be completely studied except for the main point - its biological function in the cell.

Cytosine methylamine in DNA may affect most of genetic processes: gene expression; DNA replication and repair; transformation, transfection, transposition, and recombination; DNA restriction; mutation frequency and gene polymorphism; chromosome (de)condensation, (in)activation, and imprinting; chromosome breaks, exchanges, and aberration; cell differentiation and immortalization; and some other processes [7]. Thus evidently, DNA methylation may be related to almost all vital aspects of the genome operation. Such multifunctionality has no analogy in molecular genetics.

Most authors suppose that the function of DNA methylation can be realized by changing the ability of gene regulatory elements to bind with corresponding regulatory proteins. Thus, this modification may serve as a highly specialized epigenetic code, which can be inherited.

At the same time, experimental data are rather contradictory and often do not allow to distinguish, in principle, the cause from the consequence. For example, it is not clear whether methylation of gene regulatory elements influences their ability to be bound with protein factors controlling these genes. Or vice versa, gene activation because of protein binding takes their regulatory elements out of reach of DNA methylases [7]. The mechanism controlling the very DNA methylation is still obscure.

Moreover, it is difficult to combine the idea about the regulatory function of DNA methylation with three groups of facts. First, if DNA methylation really has vital functions, then why this system is absent from some eukaryotic species (for references see [6]). Second, it is well known that m^5C residues are "hot spots" for mutations, which can destabilize gene structure and functioning [7-9]. Third, the age-related loss of the bulk of m^5C residues from DNA is found both in organism's tissues [10] and cell cultures [11].

In the present report, I have analyzed the most important literature data together with my own results [6-19] to advance a functional model that treats enzymatic DNA methylation as a special genetically programmed mechanism for accumulating mutations with aging.

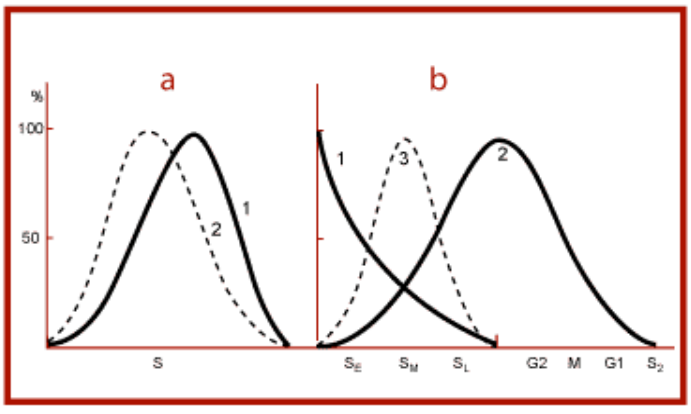

FIG. 1. Scheme of DNA methylation during the cell cycle in murine L-cells *(a)* and in the wheat first leaf *(b)*. *a)* DNA methylation *(1)* lags after DNA synthesis *(2)* [23]. *b)* Dynamics of replicative *(1)* and postreplicative *(2)* DNA methylation during DNA synthesis (*3*) and in other periods of the cell cycle [24, 42]. S_E, S_M, S_L - early, middle and late S-phases; G_2 and M (mitosis) — the phases of the same cell cycle; G_1 and S_2 – periods of the next cycle.

2. REPLICATIVE AND POSTREPLICATIVE DNA METHYLATION.

Cytosine methylation in mammalian DNA takes place in the S-period of the cell cycle (Fig. *1,a)*. Usually it peaks by the middle of the S-phase and proceeds more or less in parallel with the replicative DNA synthesis [20]. However, the main DNA methylation has a lag-period relative to DNA synthesis [21, 22] (Fig. *1)*. In L-cells, methylation is fully completed during the S-phase, and newly formed m^5C residues appear in short 5S intermediates of DNA replication before their integration into Okazaki fragments [23].

Hence, replicative DNA methylation begins just after nucleotide polymerization and likely proceeds directly in the replicative forks [23, 24]. The molecules of MTases have recently been found to be integrated into an enzymatic complex, which accomplishes DNA replication, and to be dispersed in chromatin thereafter [25]. MTase has a special targeting sequence, directing these molecules into the DNA replicative foci.

In normal cells, DNA methylation may proceed for several hours even with DNA replication completed [24, 26, 27]. One maximum of DNA methylation usually coincides with the S-period and the other one with mitosis, the G2- or Gl-phase of the cell cycle [24, 28, 29] (Fig. *1,b)*. Postreplicative DNA methylation is over by the next S-phase, so that symmetrically methylated duplexes take part in DNA. In the cells of the wheat first leaf, the ratio of replicative to postreplicative methylation decreases fast from 76% at the early S-phase to 20% at the late S-phase (Fig. 1,*b*) [24].

Thus, DNA methylation proceeds in the cell cycle both during DNA synthesis and postreplicatively, i.e., in two stages. The mechanism and the functional role of such bimodal DNA methylation are still obscure.

3. ORIGIN OF MINOR THYMINE IN DNA.

In the early sixties, the deficit of CpG dinucleotides were found in vertebrate DNA [30]. For

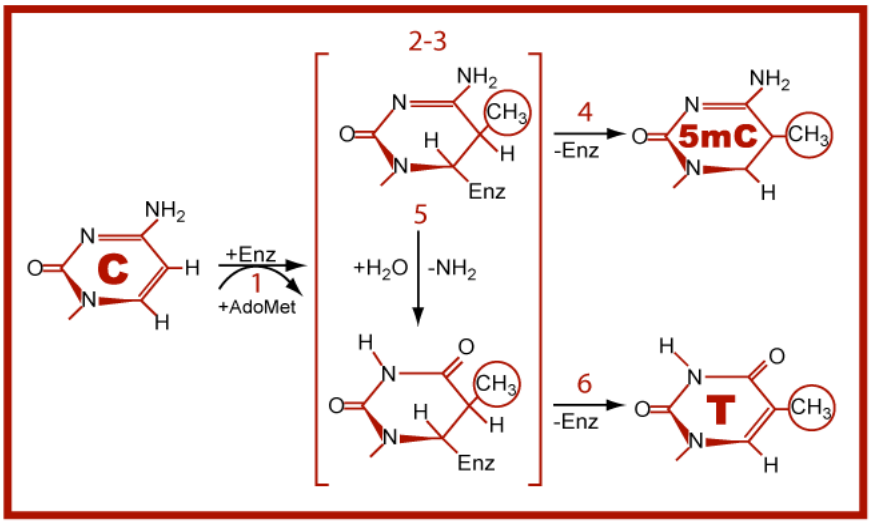

FIG. 2. Scheme of minor thymine production by cytosine methylation in DNA. *1)* DNA methylase binding to the cytosine C6 atom; *2)* break of the double C5-C6 bound with activation of the C5 atom; *3)* transfer of the methyl group from AdoMet to C5 of the 5,6-dihydrocytosine intermediate and formation of S-adenosylhomocysteine (AdoHcys); *4)* production of 5-methylcytosine after enzyme removal from the C6 atom; *5)* nucleophilic hydrolysis of the NH_4 group at C5 of the intermediate; *6)* formation of minor thymine after enzyme dissociation from the C6 atom of the complex.

instance, human DNA contains only 8 CG doublets per 1 kbp instead of 40 expected. It was found at that time that 5'-CpG-3' doublets are one of the main methylation sites in eukaryotic DNA [4].

Only by the late sixties the relation between CG methylation and the loss of these sites from DNA became clearer. A portion of radioactivity was usually found in thymine together with m^5C upon DNA methylation by AdoMet or L-methionine labeled in the methyl group [31]. This thymine labeled in the CH_3 group was termed "minor;" it could appear only as a result of m^5C deamination in DNA.

It has become evident that the deficit of CG doublets is due to a high rate of *C→T substitutions in the methylation sites of DNA. This fact was fully confirmed by the discovery that the CG deficit is accompanied by an excess of TG+CA, i.e., it reflects the accumulation of *CG→TG+CA mutations in DNA [32]. Simultaneously one of the first hypotheses has been advanced for the function of DNA methylation in consecutive inactivation of repeated regulatory elements by deamination of m^5C residues [33]. Transitions in CG sites were thereafter found to occur many times more often than in any other DNA site [8, 9, 19, 34] and to amount up to half of all point mutations [35].

Deamination of m^5C residues was long considered as a spontaneous hydrolytic reaction just after DNA methylation [31,36]. Recently minor thymine has been found to arise during DNA methylation proper, along with m^5C [15]. This has been shown through DNA methylation *in vitro* by animal, plant, and bacterial DNA methylases [15, 37]. The most surprising thing was that minor thymine and m^5C were produced by incubating DNA with only AdoMet without DNA methylase, i.e. nonenzymatical [15]. Hence both minor thymine and m^5C are the major products of the DNA methylation reaction as such.

The mechanism of evolvement of minor thymine residues during DNA methylation is shown in Fig. 2 [15, 38, 39]. The reaction of DNA methylation is initiated by nucleophilic attack of the cytosine C6 atom by the active group of DNA methylase (Fig. 2, *1),* possibly, a sulfhydryl group of a conservative cysteine residue. As a result of this attack, the C5-C6 bond is broken and a negative charge appears, which activates the previously inert C5 atom to accept the methyl group (Fig. 2,*2).* At the next stage, the CH_3 group is transferred from AdoMet to the C5 atom with the formation of the 5,6-dihydrocytosine intermediate and S-adenosylhomocysteine (Fig. 2, *3).* Dissociation of the enzyme from the complex results in restoration of the C5-C6 double bond yielding m^5C residues (Fig. 2, *4).*

Two more reactions are shown to be peculiar for 5,6-dihydropyrimidine intermediates besides electrophilic substitutions at the C5 atom [38]. One such reaction is hydrolysis of the nucleotide glycoside bond, and other one is nucleophilic deamination on C4 atom (Fig. 2, *5*). Subsequent reduction of the C5-C6 double bond after the enzyme removal gives rise to minor thymine residues (Fig. 2, 6). It has recently been found that the reaction of hydrolytic deamination is catalyzed much more actively by DNA methylase with the AdoMet deficiency [39]. Thus, minor thymine is really a product of the very reaction of DNA methylation, and MTases may be powerful generator of C→T mutations in the cell [15, 38].

4. DNA METHYLATION AS GENERATOR OF MUTATIONS.

Let us consider in more detail the results of the appearance of minor thymine in *CG sites of DNA, and then postulate the main consequences of the m^5C deamination. In the subsequent parts of the paper we will try to explain some previously obscure aspects of DNA methylation in the context of the functional model proposed.

4.1. Mechanism of m^5C→T+C substitutions.

The scheme in Figure 3 shows that during replicative DNA methylation (*1*) m^5C deamination *(2)* takes place,

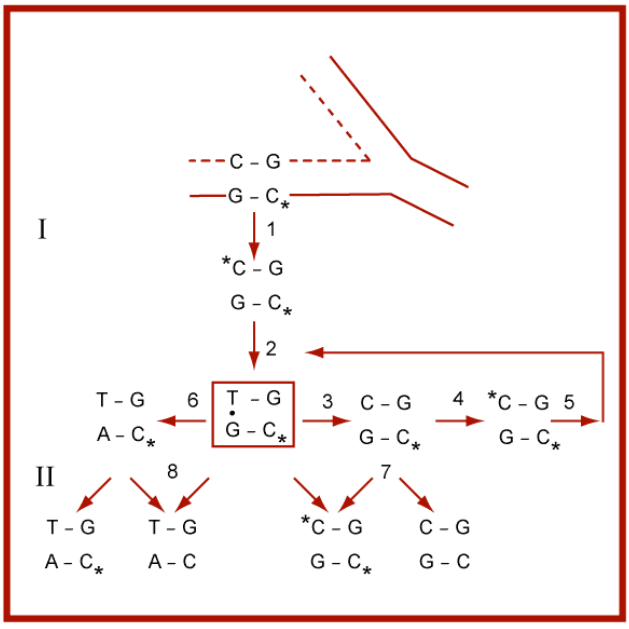

FIG. 3. Scheme of $m^5C \rightarrow T+C$ transition mutations as a result of m^5C deamination in DNA. The first (I) and the second (II) rounds of DNA replication. *1)* Replicative DNA methylation; *2)* deamination of a newly formed m^5C residue; *3*) repairing T:G mispair to the initial C:G pair; *4)* postreplicative DNA methylation; *5)* m^5C deamination; 6) incorrect repair of G:T mispair to A:T pair; *7*) production of methylated and nonmethylated CpG sites in the next round of DNA replication (II); *8)* fixation of $m^5C \rightarrow T$ mutation if T:G mispair are not or incorrectly repaired. Dotted line is the nascent DNA chain.

and minor thymine residues occur in the nascent DNA chain, i.e., $m^5C \rightarrow T$ substitutions take place. This gives rise to mismatched G:T pairs, which are usually repaired with the restoration of initial G:C pairs (*3).* As a result, new hemimethylated sites appear in DNA. These sites are a good substrate for maintenance MTases and can be methylated, but only postreplicatively (*4).*

Symmetrical methylated and nonmethylated CpG duplexes will occur in the next round of DNA replication (II), and $m^5C \rightarrow C$ substitutions will accumulate in DNA (*7*).

In case of an incorrect repair (*6*), a portion of G:T mispairs may be converted into A:T pairs by mistake. The same result will be if G:T mispairs are not repaired at all (*8)* and preserved till the next DNA replication (II). In the last two cases, $m^5C \rightarrow T$ substitutions will appear in DNA. Asymmetrical methylated *CA sites will also occur in DNA (*8*) after the next DNA replication (II).

From the model considered, it can be proposed that during postreplicative DNA methylation (*4)* a portion of m^5C residues may be deaminated again (*5*). The resulting G:T mispairs are repaired again and the cycle *(5→2→3→*4→*5)* is repeated.

A situation can readily be imagined when a part of G:T mispairs in already condensed chromatin will not be accessible for the repairing enzymes prior to the S-phase of the next cell cycle (II). In just the same way, a portion of hemimethylated sites inaccessible for MTases will not have enough time to be methylated (*4).* This gives rise to an additional amount of $m^5C \rightarrow T$ and $m^5C \rightarrow C$ substitutions in DNA.

Hence, new $*CG \rightarrow TG+CA$ and $*CG \rightarrow CG$ substitutions will appear in DNA over and over with each next DNA replication, and these substitutions will be accumulated in the genome with time. As a result, new asymmetric methylated *CA, TG/CA, and nonmethylated CG sites will appear in DNA together with symmetrical methylated *CG sites (Fig.3).

Let us draw the main conclusions from the model considered:

1. Minor thymine is formed (*2,5)* as a result of replicative *(1)* and postreplicative *(4)* DNA methylation.

2. The $m^5C \rightarrow T+C$ substitutions take place as a result of m^5C deamination (Fig. 3, *2)* and subsequent misrepairing G:T pairs *(3,6).* These substitutions give the impression that the nascent DNA chain is undermethylated compared to the parental one.

3. The number of hemimethylated sites is maximal in nsDNA (*3)* and then drops until the next S-phase because of postreplicative DNA methylation *(4).*

4. Postreplicative methylation (*4)* reflects the intensity of m^5C deamination *(2)* during replicative methylation of nsDNA *(1)* and it is "reparative" by its origin.

5. The bimodal character of DNA methylation (Figs. 1,b and Fig.3, *1,4*) in the cell cycle is a result of $*C \rightarrow T \rightarrow C \rightarrow *C$ substitutions occurring because of deamination of a portion of m^5C residues *(2).*

6. During a cell cycle (I-II), a part of hemimethylated sites may be not methylated (*4),* and this causes m^5C loss from DNA (7) increasing with age.

7. In each cell cycle, new $m^5C \rightarrow T$ mutations appear because of the absence of (*8)* or incorrect (6) repair of some G:T mispairs, and these mutations are gradually accumulated in the genome during ontogeny and phylogeny.

8. Continual accumulation of $*CG \rightarrow TG+CA$ transitions in genes can lead to a "catastrophe of errors," and cause aging and destruction or immortalization of cells.

4.2. Minor thymine is a result of DNA methylation.

Production of minor thymine residues as a result of DNA methylation by AdoMet or L-methionine labeled in methyl groups has been shown in the long list of papers [3, 15, 31, 36, 37, 40-42]. For example, the portion of minor thymine labeled *in vivo* in this way amounts in DNA of sea urchin embryos to 54—51% of m^5C [31], in HeLa to 32%, L-cells 55%, lung fibroblasts 76% [40], Novikov hepatoma cells 58% [41], chicken embryos 168% [40], and wheat seedlings 53-80% [37, 42]. Yet, the data could be

alternatively explained by metabolic reutilization of CH radioactivity *in vivo* in the mono-carbon pathway with subsequent reincorporating into dTMP during synthesis from dUMP.

However, this is completely excluded in case of DNA methylation *in vitro* by different MTases [15, 37]. Up to 30% of radioactivity from AdoMet is usually found in the CH_3 groups of minor thymine, as well as in experiments *in vivo*. It is interesting that the AdoMet deficit *in vitro* leads to more active deamination of cytosine to uracil residues by MTase [39]. If enzymatic DNA hydrolysis to nucleotides was used, the possibility of m^5C chemical deamination to thymine during DNA analysis is very small [15].

Moreover, minor thymine has been found after *in vitro* methylation by DNA(adenine)methylase and even without any MTase [15]. In this case nonenzymatic DNA methylation by AdoMet takes place with formation of labeled thymine and m^5C residues at ratio of 9:1 [15]. Earlier it has been found that m^7G, O^6mG, and m^3A residues also occur in DNA as a result of nonenzymatic methylation by AdoMet [43]. Judging by the kinetics of enzymatic and nonenzymatic DNA methylation, minor thymine, as well as m^5C, are produced just at the moment of DNA methylation [15]. Both reactions (transmethylation of 5,6-dihydrocytosine intermediates and their subsequent deamination) proceed almost simultaneously and are catalyzed by the same MTase [15, 38, 39] (Fig. 2).

The origin of minor thymine in DNA was obscure for a long time. It is supposed to arise by spontaneous m^5C deamination just after DNA methylation [36]. The rate of this reaction *in vitro,* taking into account the efficiency of G:T repair in mammalian cells, is estimated at $1.66x10^{-16}$ sec^{-1} or $4.9x10^{-2}$ per genome per day [44]. During nonenzymatic DNA methylation the rate of minor thymine production is several orders higher and equal to 10^4 residues per genome per day [15]. As a result of $m^5C \rightarrow T+C$ substitutions, the rate of m^5C loss from DNA during *in vivo* and *in vitro* cell aging has approximately the same order [10, 11].

Thus, minor thymine production can be considered as a result of both spontaneous m^5C deamination in DNA [36] and DNA methylation reaction as such [15, 39] (Fig. 2). There is much direct and indirect experimental evidence in favor of this conception. By incubating rat bone marrow cells with [2-^{14}C]orotic acid for a long time, the minor thymine content in DNA is increased by 42%, while the m^5C content drops by 46.5% [45]. The amount of labeled cytosine in DNA does not change, which may point to a constant pyrimidine precursor pool in these cells. On the contrary, as a result of pulse labeling (2-5 min) the thymine content in nsDNA decreases by 33%, while the m^5C and cytidine contents increase by 21 and 12%, respectively [46], and this may be the effect of G:T repair.

It has been found that during the incubation of cut wheat shoots with L-[methyl-^{14}C]methionine, 43% of the label is incorporated into thymine residues and 40.4% into m^5C [42]. The radioactivity in m^5C becomes almost three times lower during the S-phase. The most interesting, different phytohormones suppressed in parallel the label incorporation from [$^{14}CH_3$]methionine into m^5C and into thymine residues of DNA are approximately in the same proportion [42].

These facts suggest that minor thymine is produced right at the moment of DNA methylation, and both reactions, transmethylation and deamination, are catalyzed by the same MTase (Figs. 2, 3). Therefore a lag period is observed between DNA synthesis and methylation in normal cells (Fig. 1,b).

4.3. Is m^5C loss a result of cytosine methylation?

Figure 3 shows that m^5C deamination *(2)* and subsequent G:T repair *(3,6)* lead to $m^5C \rightarrow T+C$ transitions in the daughter chain, i.e., to the m^5C loss from nsDNA. If this is the case, the level of DNA methylation in the newly formed chain should be considerably lower than that in the parental one.

In murine L-cells, the newly formed DNA chains already at the Okazaki fragment stage (5S) are undermethylated by a factor of 1.6-2.4 as compared with the parental ones [23, 47]. In suspension-cultured tobacco cells, cytosine residues in the newly synthesized (4-6S) fragments are methylated 2.4-fold lower than in legated (8S) fragments and in parental DNA [48]. And finally, in growing wheat shoots the methylation level [$ML=m^5Cx100(m^5C+C)$] of Okazaki fragments is 3-4-fold lower than in total DNA [49]. Thus, the newly replicated DNA chain has been found to contain less than half of all m^5C residues of the parental chain.

Undermethylation of the nascent DNA chain is usually explained by a lag period between methylation and DNA synthesis [21, 22] (Fig. 1). Newly replicated DNA duplexes are supposed to be incompletely methylated by the time of their isolation from the cells.

Hence, there are two alternative explanations for the same fact of nsDNA undermethylation. One of them suggests that this undermetylation exists from the very beginning of DNA replication. The other attributes the m^5C loss to active deamination of the residues (Fig. 2) already after symmetrical methylation of nsDNA during replication (Fig. 3). The correct understanding of the nature of the DNA methylation reaction depends on the choice between one and the other mechanism.

In early seventies, it was found that during the S-phase of synchronized hamster CHO cells the rate of DNA methylation is first maximal and then decreases relatively to the rate of DNA synthesis [50]. In rat regenerating liver and bone marrow cells, the nsDNA pulse-labeled by orotic acid turned out to be hypermethylated in relation to total DNA [45, 46]. The level of DNA methylation was observed to decrease fast after long exposure with the label. It was found that soon after DNA replication the symmetry of chain methylation was fully preserved in the most of

methylated sites [51], but after the end of replication the daughter DNA chain turned out to be considerably undermethylated in comparison with the parental one.

Compared to legated DNA, hypermethylation of the newly formed Okazaki fragments was found in L-cells incubated with labeled thymine and L-methiomine as a result of the inhibition of DNA replication by arabofuranosylcytosine [23]. The ratio of radioactivities in m^5C/T for DNA fragments of <5S, 5-18S, and >18S was equal to 1.2, 0.54, and 0.53, respectively.

Thus, it has been found that short replicative intermediates (<5S) are symmetrically methylated during their synthesis. At further stages of DNA replication no methylation took place, and label incorporation into m^5C dropped sharply [23]. If DNA legation was blocked by hy-droxyurea or any other way, no subsequent increase in the methylation level of Okazaki fragments was observed. The authors supposed that the dynamics observed might be explained by partial enzymatic demethylation of hypermethylated Okazaki fragments during the subsequent stages of DNA replication [23].

Two sets of data may be indirect evidence of the intensity of m^5C loss due to replicative methylation of nsDNA. On the one hand, up to 50% of m^5C radioactivity are found in minor thymine residues of nsDNA (Part 4.2). On the other hand, the nsDNA chains contain 40-75% m^5C less than the parental one [23, 24,47-49]. The dynamics of replicative and postreplicative DNA methylation in the S-phase of the cell cycle confirms this fact (Fig. 1,*b*).

All this seems to give proof that the m^5C loss from the daughter DNA chains is secondary. It results from active deamination of the portion of m^5C intermediates already during replicative DNA methylation (Figs. 1-3). Paradoxically, m^5C is lost from the daughter DNA chain just because of cytosine methylation in DNA.

4.4. Hemimethylated sites in DNA.

According to the model proposed, new hemimethylated sites arise in DNA after the G:T repair (Fig. 3, *3),* and the number of these sites should be maximal in nsDNA. These sites are known to be a very good substrate for maintenance MTases [3], and their amount should gradually decrease as a result of postreplicative DNA methylation (Fig. 3, *4)*. The symmetrical DNA chain methylation should be completely restored by the beginning of the next DNA replication round.

In the nascent DNA chain, the m^5C content is really 2-4 times lower than in the parental one (Part 4.3). This means that 50-75% of all methylated sites in nsDNA contain m^5C only in one of the chains. In mouse L-cells, methylation of these sites proceeds very fast during DNA replication [23] (Fig. *1,a)*.The methylation level of Okazaki fragments (≤5S) is 2.7%, and after their legation into >5S fragments the ML increases to 4.2%, which differs slightly from that in total DNA of the L-cells, 4.4%. In suspension culture of tobacco cells the ML of short (4~6S) fragments of nsDNA amounts to 17%, and for longer ones (>8S) as well as for parental DNA it is 40.2% [48]. A different situation is found in the wheat first leaf cells [49]. Okazaki fragments (ML=7.0%) continue losing the m^5C content after their legation into 8S intermediates (ML=6.4%). Then at the stage of 12S fragments a slight increase in the ML is observed (7.2%). Gradually it grows in >12S fragments to 7.4% and reaches the ML of mature DNA (25-27%) only by the S-phase of the next cell cycle [49] (Fig. 1,*b)*.

Thus, nsDNA really contains the maximal number of hemimethylated sites. That is why such DNA of wheat shoots has a tenfold ability to accept CH_3 groups during its methylation by homologous MTase compared to animal or bacterial DNAs [52]. Sequencing of the kininogen promoter showed that these sites amount to 23.5% of all symmetrically methylated sites in human liver DNA, 57.1% in sperm, and 41.9% in HeLa cells [53]. Hence, hemimethylated sites are almost always observed in DNA.

4.5 What is the "reparative" DNA methylation?

Up to 3/4 of all m^5C residues are produced after DNA synthesis, i.e., as a result of postreplicative DNA methylation (Part 4.4). Therefore m^5C deamination (Fig. 3, *2),* G:T repair *(3),* appearance of hemimethylated sites *(3),* and their modification *(4),* i.e., *C→T→C→*C substitutions, should be also observed in nonreplicating DNA.

Loss of 9% of m^5C residues from DNA was found in F9 cells of mouse embryo carcinoma just after induction of differentiation by retinoic acid [54]. In the same way, Frend's erythroleukemia cells (FEL) induced with dimethylsulfoxide (DMSO) lose 4% of m^5C from DNA. The drop in the m^5C content in DNA was accompanied by a 2-5-fold increase in the number of hemimethylated sites; it was rather transient and reversible, ending already in several hours with DNA methylation. The kinetics of the demethylation did not coincide with the replicative m^5C loss, which is usually observed in the absence or inhibition of maintenance MTases [54].

Further development of this research clearly showed that reversible DNA demethylation during FEL cell differentiation is related to the m^5C substitutions by newly incorporated cytosine residues [55]. In this experiment, nsDNA (H-L) was separated from the parental DNA (L-L) after cell incubation with 5-bromodeoxycytosine and d[5-^{3}H]cytidine. Active incorporation of d[5-^{3}H]C into L-L duplexes was found only after cell induction and took place simultaneously with DNA demethylation. The effect observed could not be explained by the usual reparative synthesis, because there was no d[6-^{3}H]adenine incorporation into L-L DNA. The d[5-^{3}H]C label was rather short-lived and was actively removed from L-L DNA in 12-18 h after cell induction. The most interesting, m^5C→C substitutions took place during *de novo* DNA methylation and were found in *CG sites of nonreplicated DNA in the Gl-phase of

the cell cycle [55] (Fig. *1,b)*.

The authors concluded that direct enzymatic elimination of CH_3 group from m^5C owing to the break of the C-C bond has no biochemical precedents and therefore is *a priori* improbable [55]. The suggestion about selective excision repair of m^5C-containing sites in DNA also was not confirmed experimentally yet. Therefore the authors postulated the existence of an unknown transglycosylase reaction with direct m^5C elimination from DNA and substitution of cytosine residues, i.e., $m^5C \rightarrow C$.

Another attempt to elucidate the nature of reversible nonreplicative DNA demethylation has been undertaken recently [56]. Frend erythroleukemia cells were preincubated with d[2-^{14}C]cytidine for 3 days and then induced to differentiate with DMSO. As in the previous studies, a transitory and reversible drop by 7% in the [2-^{14}C] m^5C content in DNA was found. It peaked between the 15th and 21st hour after induction, followed by rapid restitution of the initial m^5C content. The period of the most intensive m^5C loss fully coincided with an almost 5-fold increase in the number of hemimethylated sites in DNA [56]. It is interesting that in the control 17% of the [2-^{14}C]C label was also incorporated into thymine residues of FEL cells. Nineteen hours after cell induction with DMSO, the C/T radioactivity ratio decreased by 22% in comparison with control; this corresponded to a 39% increase of labeled thymine in DNA [56].

The results of these experiments [54-56] are in perfect accord with the model considered (Fig. 3) and may reflect the *C→T→C→*C substitutions proceeding in nonreplicated DNA. The very idea of m^5C deamination and subsequent G:T repair gives a clue to understanding the surprising selectivity of $m^5C \rightarrow C$ substitutions observed in these papers. Subsequent methylation of such cytosine residues may therefore be considered as "reparative" DNA methylation.

The mutagenic effect of hydrolytic m^5C deamination is usually compensated for by a special mismatch repair system operating in the cell. First, a special protein recognizes the G:T mispairs, then minor thymine residues are removed by DNA glycosylase and the resulting apyrimidine gap is filled with a cytosine residue by DNA polymerase B [57]. Thus, m^5C is replaced first by thymine and then by cytosine newly incorporated into DNA, i.e., $m^5C \rightarrow T \rightarrow C$. The occurrent hemimethylated sites (Fig. 3, *3)* are modified by maintenance MTase *(4),* and thereafter the cycle (5→2→*5)* may be repeated again several times till the next DNA replication. Thus, the postreplicative DNA methylation is reparative in essence and reflects selective modification of the methylated sites proper after repair.

4.6. Postreplicative DNA methylation compensates for m^5C loss, but not completely.

The rate of replicative DNA methylation is evidently much higher than that of postreplicative one. Probably the reason is that in newly synthesized chromatin the methylation sites are more accessible and MTases are integrated into replicative foci [25]. Obviously that is why nsDNA is fully methylated already during replication (Part 4.3). At the same time, the high minor thymine content as well as the low m^5C content in nsDNA (especially in Okazaki fragments) clearly testifies a portion of newly formed m^5C is eliminated from DNA during replicative methylation (Parts 4.3 and 4.4). This m^5C loss is compensated by postreplicative DNA methylation (Fig. 3).

During the S-phase in wheat first-leaf cells, the ratio of postreplicative to replicative DNA methylation rises from 20% at the beginning of the S-period to 76% at the end and amounts to 100% after completion of DNA synthesis [42] (Fig. *1,b)*. Apparently it reflects an almost 4-fold increase in the number of hemimethylated sites formed during replicative DNA methylation (Fig. 3). These sites are the main substrate for postreplicative DNA methylation [42].

However another situation takes place during postreplicative DNA methylation in condensed chromatin when hemimethylated sites in nucleosomes and linkers may be spatially inaccessible for DNA methylases, being bound with different nonhistone proteins. This is especially true for condensed heterochromatin regions; just where m^5C-rich repeats and satellite DNAs are concentrated [16]. It is quite probable that at the first stage the MTases bind with special proteins recognizing m^5C residues in one of the DNA chains [58], and then they modify cytosine residues in the opposite chain. In any case, symmetrical methylation of the chains should be complete by the next round of DNA replication (Fig. *1,b)*.

It is seen from the model (Fig. 3) that a portion of the newly formed m^5C residues may be deaminated again during postreplicative methylation (5). Therefore the G:T repair with the formation of hemimethylated sites (Fig. *3, 3)* and their subsequent modification *(4)* should take place also in nonreplicated DNA of condensed chromatin. As a subsiding wave, *CG→TG→CG→*CG substitutions may continue (Fig. 3, *2→5→2)* till the next S-phase (Fig. *1,b*), and some m^5C residues may undergo several such cycles during one cell division.

A situation is possible wherein a portion of G:T mispairs has no enough time to be repaired (Fig. 3, *8)* and begins the next replication (II) together with a portion of hemimethylated sites (7). As a result, *CG→TG+CA transitions would be fixed and fully nonmethylated duplexes would appear in DNA. In this way, new $m^5C \rightarrow T+C$ substitutions would be produced over and over with each cell division, and accumulating in the genome, they should result in a marked m^5C loss from DNA with time.

5. GENOME LOSES MOST OF m^5C OVER LIFE SPAN.

The phenomenon of age-related m^5C loss was first

discovered in our laboratory in salmon tissues, which lose up to 30% of m^5C during spawning [59], and then in rats [60]. Later the fact was reproduced by many laboratories in experiments with different tissues of mouse, rat, hamster, cow, and humans (for references see [10]). Yet heretofore it has been unknown what portion of total m^5C is lost from the genome over ontogeny and how this mechanism operates.

The fastest m^5C loss from rat liver DNA takes place during embryonal development and in the first 15 days after birth (I period) [10]. In the subsequent 45 days (II period) this rate slows down, and remains minimal and constant over 25 months of adult rat life (III period). The DNA hypomethylation rate can be calculated as the decrease in the total m^5C content in the genome over the corresponding age interval. It is convenient to express this value as percentage of relative m^5C loss per day.

It is easy to determine that in the I period the rate of rat DNA hypomethylation is equal to 1.28%, in the II period to 0.33%, and in the III period to 0.028% of daily m^5C loss [10]. Knowing this rate and the duration of the ontogeny periods, the total m^5C loss from the genome can be calculated over the whole life span. Thus, during the embryonic period (23 days) and the first 15 days of postnatal development the rat liver DNA loses 48.6% (1.28x38) of all m^5C. During the subsequent 45 days the genome loses 14.9% (0.33x45) and then during the maximal life span (about 4 years for rats) 39.2% (0.028x1400) of m^5C residues may be lost. Summing up the total loss of m^5C from DNA, we obtain 102.7%. It means that the rat genome may lose all m^5C residues over the life span [10]. In reality, rat genome loses only 61.7% of m^5C over the period experimentally studied [10].

The experimental data of different authors can be organized so as to calculate the average rate of the m^5C loss from animal DNA during different periods of their ontogeny (for references see [10]). Summing up the m^5C loss over the entire ontogeny, it is easy to determine that the mouse genome may lose 93.7%, rat 99%, and cow 91% of total m^5C residues. Hence, the genome may really lose the most of, if not all, m^5C residues over the life span [10].

The age-related loss of the bulk of m^5C from DNA of an organism's tissues is a new and rather unexpected issue, which may be a convicting argument for the model considered above (Fig. 3). The mechanism of such m^5C loss is most likely associated with the accumulation of $m^5C \rightarrow T+C$ substitutions continually with each cell cycle (Part 4).

The age-related m^5C loss from DNA also marks the cells aging *in vitro* [11]. Knowing the DNA hypomethylation rate, it is easy to calculate the number of cell population doublings (PD) till exhausting all of m^5C residues from the genome of different cell lines. The experimentally obtained values of the Hayflick limit have been found to coincide or to be somewhat lower than the number of PD until all m^5C is lost from the genome of the corresponding cell cultures (for references see [11]). Thus, the situation is, in principle, alike for both *in vivo* and *in vitro* aging: the genome may lose the bulk of m^5C residues during the life span [10, 11].

A constant rate of the m^5C loss from DNA is a feature of both cell aging in culture and in adult animal tissues. The loss most of m^5C from DNA may reflect the accumulation of a critical number of $m^5C \rightarrow T$ mutations in genes that is no longer compatible with normal cell proliferation and the organism's life.

6. ACCUMULATION OF $m^5C \rightarrow T$ MUTATIONS IN GENOME.

Figure 3 shows that mismatched G:T pairs occur in DNA as a result of m^5C deamination *(2)*. The efficiency of the G:T repair system is rather high in mammalian cells, and that is why only 1-2% of these mispairs are not corrected at all in DNA (Fig. 3,*8*) [57]. However, about 8% of the G:T mismatches are converted not into G:C but into A:T pairs by mistake (Fig. 3, 6). This means that up to 10% of all deaminated m^5C may be preserved in DNA as minor thymine residues, which generate $CG \rightarrow TG+CA$ mutations.

The $m^5C \rightarrow T$ transitions are usually irreversible, because the probability of reverse substitutions is very small [8, 15]. Thus, they are a unique example of directional mutations, which may be continually accumulated in the genome of somatic cells with aging. The genome of germ-line cells is only partly protected from methylation [61-63], and that is why such mutations might be accumulated in phylogeny. Let us see what trace they leave in DNA after millions of years of evolution.

A likely number of $CG \rightarrow TG+CA$ mutations, which have ever taken place in the CG sites may be calculated by the analysis of frequencies of CG, TG, and CA dinucleotides in sequenced genes [7-9, 16-19]. In this way it is somehow possible to look into the remote past and to retrieve information about "fossil" methylation of different genes throughout the whole period of their evolution.

The number of $CG \rightarrow TG+CA$ transitions has been found to be the higher the higher the genome methylation level is in different species, and vice versa [8, 34], For instance, the frequencies of CG, TG, and CA duplets in the DNA of drosophila, yeasts, and other species with undetectable amount of m^5C in their genomes [6] usually do not differ from stochastic [8, 17]. On the contrary, vertebrates exhibit a marked deficiency of CG (-30) and an excess of TG+CA (+29) per 1 kbp of DNA. Thus, about 30 $CG \rightarrow TG+CA$ substitutions per each kbp might take place in their genomes [8, 34]. This is indicative of the loss of 3 mole % of m^5C from the CG sites alone.

Analysis of more than 1000 vertebrate sequences (EMBL and GenBank) demonstrated that the frequency of $CG \rightarrow TG+CA$ substitutions in some genes does not differ from stochastic, while in other genes it may run as high as 60-100 and more substitutions per

1 kbp [8]. As a result of these mutations, the CG sites are almost completely lost from 20% of repeated sequences and genes studied (M^+_o genes). These genes have lost their ability to be methylated, though have been deeply modified in their cytosine residues in the past. Most of these genes have already finished evolution and become pseudogenes [8].

About 8% of the vertebrate genes studied apparently have never been heavy methylated (M^- genes). These genes are tRNA, long and short rRNA, snRNA U1-U6, H3 and H4 histone genes, genes of many enzymes, and also 5'-regulatory regions of many genes (CpG islands) [8, 9]. Thus, the most vital and evolutionarily conserved genes of proteins and RNA are party or completely protected from methylation.

Interestingly, "fossil" methylation turned out to contribute disproportionately to accelerating the divergence of some histone genes, snRNAs, and also of 5S and 5.8S rRNA and other gene families [8, 9]. In snRNA pseudo-genes, up to 30-40% of total point substitutions studied are the CG→TG+CA transitions [9], which also amount to 30% of point mutations in other mammalian pseudogenes [64]. This is due to the fact that the methylated CG sites mutate many folds more frequently than any other DNA doublet [8, 9, 35, 65]. However, this analysis ignores a part of total m^5C→T transitions, because they may take place in palindrome *CNG sites of DNA [5]. Therefore, more than half all point mutations accumulated over the evolution may occur as a result of gene methylation.

The analysis of the mutations causing a number of human hereditary diseases showed that CG→TG+CA transitions might amount up to half of total point substitutions [65]. Diseases caused by the deficiency of some enzymes in the organism are often related to this kind of mutations. For instance, in the gene of a-antitrypsin such mutations amount to 21% of all point substitutions, in that of adenine deaminase to 29%, glucose-6-phosphate dehydrogenase 86%, b-globin (thalassemia) 25%, immunoglobulin 33%, phenylalanine hydroxylase (phenylketonuria) 36% (for references see [10]), p53 suppressor (leukemia and other cancers) 33% of total point mutations [66].

In this respect, the mutations in genes of blood coagulation factors VIII and IX, which cause hemophilia, are studied in more detail. In patients with hemophilia B, 47% of 216 mutations found in the defective factor IX genes are exactly related to CG→TG+CA transitions [67]. Peculiarly, in many of the patients such mutations are not inherited but occur *de novo* during 1-2 generations. The mutations arise much more often in male germ line than in female one, which may be due to different extents of genome methylation in these cells [61-63]. Therefore direct contribution of gene methylation to the general mutagenesis may be especially high.

7. CONCLUSION.

The biological function of DNA methylation has been investigated over the past 45 years. For the most part, the studies concerned the relations between this enzymatic modification and regulation of different genetic processes in the cell [7]. It is noteworthy that no general patterns for this function have been established despite all attempts.

The situation becomes understandable assuming that the m^5C→T+C transition mutations are essentially stochastic in different genes. The driving pattern here is that the age-related m^5C loss from DNA reflects the accumulation of m^5C→T transitions over the genome. Thus the model proposed postulates a new biological function of DNA methylation as genetically programmed powerful generator for accumulation of mutations in genome during cell aging.

ACKNOWLEDGEMENTS

The work was partly supported by Russian State Programs Frontiers in Genetics and Human Genome.

REFERENCES

1. R. D. Hotchkiss, *J. Biol. Chem.,* 175, 315-332 (1948).
2. G. R. Wyatt, *Biochem. J.,* 48, 584-590 (1951).
3. R. L. P. Adams, *Biochem. J., 265,* 303-320 (1990).
4. J. Doskocil and F. Shorm, *Biochim. Biophys. Acta,* 55, 953-959 (1962).
5. D. M. Woodcock, P. J. Crowther, and W. P. Diver, *Biochem. Biophys. Res. Communs.,* 145, 888-894 (1987).
6. A. L. Mazin, L. M. Muchovatova, N. G. Shuppe, and B. F. Vanyushin, *Dokl. Akad. Nauk USSR, 276,* 760-762 (1984).
7. A. L. Mazin, *Mol. Biol.(Moscow), 26,* 244-263 (1992).
8. A. L. Mazin and B. F. Vanyushin, *Mol. Biol.,* 21, 543-551, 552-562 (1987).
9. A. L. Mazin and B. F. Vanyushin, *Mol. Biol., 21,* 678-687, 1099-1109(1987).
10. A. L. Mazin, *Mol. Biol.,* 27, 160-173 (1993).
11. A. L. Mazin, *Mol. Biol., 27,* 895-907 (1993).
12. A. L. Mazin, V. K. Vasilyev, B. F. Vanyushin, and A. N. Belozersky, *Dokl. Akad. Nauk USSR,* 210, 963-966 (1973).
13. B. F. Vanyushin, A. L. Mazin, V. K. Vasilyev, and A. N. Belozersky, *Biochim. Biophys. Acta,* 299, 397-403 (1973).
14. A. L. Mazin and G. E. Sulimova, *Dokl. Akad. Nauk USSR,* 211, 985-988 (1973).
15. A. L. Mazin, O. A. Gimadutdinov, S. A. Turkin, N. N. Burzeva, and B. F. Vanyushin, *Mol. Biol., 19,* 903-914 (1985).
16. N. I. Shengeliya, M. V. Japiashvili, A. L. Mazin, and T. G. Beridze, *Mol. Biol.,* 20, 798-810 (1986).
17. A. L. Mazin and B. F. Vanyushin, *Mol. Biol, 22,* 1399-1404 (1988).
18. A. L. Mazin, L. M. Boiko, O. A. Ogarkova, and B. F. Vanyushin, *Mol. Biol.,* 22, 1688-1696 (1988).
19. A. L. Mazin and B. F. Vanyushin, *Mol. Biol., 24,* 23-43 (1990).
20. R. L. P. Adams, *Biochim. Biophys. Acta,* 254, 205-212 (1971).
21. Y. Gruenbaum, M. Szif, H. Cedar, and A. Razin, *Proc. Natl. Acad. Sci. USA,* 80, 4919-4921 (1983).
22. T. L. Kautiainer and P. A. Jones, *Biochemistry, 24,* 5575-5581 (1985).
23. G. I. Kiryanov, L. V. Isaeva, M. D. Kimos, N. I. Ganicheva, and B. F. Vanyushin, *Biokhimiya (Moscow), 47,* 153-161 (1982).
24. M. D. Kirnos, N. I. Alexandrushkina, and B. F. Vanyushin, *Biokhimiya, 54,* 313-322 (1989).
25. H. Leonardt, A. M. Page, H. Weiner, and T. H. Bestor, Cell, 71, 865-873 (1992).

26. G. Theiss and H. Follmann, *Biochem. Biophys. Res. Communs.,* 94, 291-297 (1980).
27. D. M. Woodcock, T. K. Adams, and I. A. Cooper, *Biochim. Biophys. Acta,* 696, 15-22 (1982).
28. R. Valencia and 0. Bertaux, *Cell Int. CNRS,* 240, 69-75 (1974).
29. B. Bugler, B. B. Bertaux, and R. Valencia, *J. Cell. Physiol.,* 103. 149-157 (1980).
30. J. Josse, A. D. Kaiser, and A. Komberg, *J. Biol. Chem.,* 236, 864-875 (1961).
31. P. Grippo, M. laccarino, E. Parisi, and E. Scarano, *J. Mol. Biol.,* 35, 195-208 (1968).
32. W. Salser, *Cold. Spring Harbor Symp., 42,* 985-1003 (1977).
33. R. Holliday, *Fed. Proc., 34,* 51-55 (1975).
34. A. P. Bird, *Nucl. Acids Res.,* 8, 1499-1504 (1980).
35. D. N. Cooper and S. Gerber-Huber, *Cell Differ.,* 817,199-205 (1985).
36. M. Ehrlich, K. F. Norris, R. Y.-H. Wang, K. C. Kuo, and C. W. Gehrke, *Biosci. Repts., 6,* 387-393 (1986).
37. G. Theiss, R. Schleicher, G. Schimpff-Weiland, and H. Foll mann, *Eur. J. Biochem.,* 167, 89-96 (1987).
38. K. M. Ivanetich and D. V. Santi, *Progr. Nucl. Acids Mol. Biol., 42,* 127-156 (1992).
39. J.-C. Shen, W. M. Rideout III, and P. A. Jones, *Cell,* 71, 1073-1080 (1992).
40. L. A. Gulp, E. Dore, and G. M. Brown, *Arch. Bioch. Bioph.,* 136, 73-79 (1970).
41. T. W. Sneider and V. R. Potter, *J. Mol. Biol., 42,* 271-284 (1969).
42. M. D. Kirnos, N. A. Artyuchovskaya, N. I. Alexandrushkina, V. V. Ashapkin, and B. F. Vanyushin, *Biokhimiya,* 51. 1875-1855 (1986).
43. B. Rydberg and T. Lindahl, *EMBO J.,* 1, 211-216 (1982).
44. D. N. Cooper and M. Krawczak, *Human Genet.,* 83, 181-188 (1989).
45. N. A. Fedorov and N. A. Koreshkova, *Dokl. Akad. Nauk USSR,* 209, 231-234 (1973).
46. N. A. Fedorov, I. N. Ovgartchuk, B. C. Borisov, and B. F. Vanyushin, *Dokl. Akad. Nauk USSR,* 236, 1256-1259 (1977).
47. R H. Grafstrom, R. Yuan, and D. L. Hamilton, *Nucl. Acids Res.,* 13, 2827-2842 (1985).
48. E. A. Bashkite, M. D. Kirnos, G. I. Kiryanov, N. I. Alexan drushkina, and B. F. Vanyushin, *Biokhimiya,* 45, 1448-1456 (1980).
49. M. D. Kimos, L. I. Kutueva, N. I. Ganicheva, and B. F. Vanyushin, *Biokhimiya,* 49, 1357-1366 (1984).
50. T. W. Sneider, *J. Mol. Biol.,* 79, 731-734 (1973).
51. D. M. Woodcock, D. L. Simmons, P. J. Crowther, I. A. Cooper, K. J. Trainor, and A. A. Morley, *Exptl. Cell Res.,* 166, 103-112 (1986).
52. T. I. Vlasova, Z. N. Demidenko, M. D. Kirnos, and B. F. Vanyushin, *Biokhimiya,* 57, 1049-1058 (1992).
53. M. Frommer, L. E. McDonald, D. S. Millar, et al, *Proc. Natl. Acad. Sci. USA,* 89, 1827-1831 (1992).
54. T. Y. Bestor, S. B. Hellewell, and V. M. Ingram, *Mol. Cell. Biol., 4,* 1800-1806 (1984).
55. A. Razin, M. Szif, J. Kafri, M. Roll, et al., *Proc. Natl. Acad. Sci. USA,* 83, 2827-2831 (1986).
56. R. L. P. Adams, A. Hanley, and A. Rinaldi, *FEES Lett.,* 269, 29-31 (1990).
57. T. C. Brown and J. Jiricny, *Cell,* 50, 945-950 (1987).
58. N. M. Inamdar, K. C. Ehrlich, and M. Ehrlich, *Plant. Mol. Biol.,* 17, 111-123(1991).
59. G. D. Berdishev, G. K. Korotaev, G. V. Boyarskich, and B. F. Vanyushin, *Biokhimiya,* 32, 988-993 (1967).
60. B. F. Vanyushin, L. T. Nemirovsky, V. V. Klemenko, V. K. Vasylyev, and A. N. Belozersky, *Gerontologia,* 19, 138-152 (1973).
61. S. I. Feinstein, V. R. Racaniello, M. Ehrlich, etal., *Nucl. Acids Res.,* 13, 3969-3978 (1985).
62. J. Kaput and T. W. Sneider, *Nucl. Acids Res.,* 7, 2303-2322 (1979).
63. J. Sanford, L. Forrester, V. Chapman, and A. Chandley, *Nucl. Acids Res., 12,* 2823-2836 (1984).
64. W. Bains and J. Bains, *Mutat. Res.,* 179, 65-74 (1987).
65. D. N. Cooper and H. Youssoufian, *Human. Genet.,* 78, 151-155 (1988).
66. W. M. Rideout III, G. A. Coetzee, A. P. Olumi, and P. A. Jones, *Science, 249,* 1288-1290 (1990).
67. F. Giannelli, K. A. High, et al., *Nucl. Acids Res.,* 18, 4053-4059 (1990).