

The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA

Jiang-Cheng Shen, William M. Rideout III and Peter A. Jones*

Department of Biochemistry and Molecular Biology, USC/Norris Comprehensive Cancer Center, School of Medicine, Los Angeles, CA 90033, USA

Received December 23, 1993; Revised and Accepted February 9, 1994

ABSTRACT

The modified base, 5-methylcytosine, constitutes approximately 1% of human DNA, but sites containing 5-methylcytosine account for at least 30% of all germline and somatic point mutations. A genetic assay with a sensitivity of 1 in 10^7 , based on reversion to neomycin resistance of a mutant pSV2-*neo* plasmid, was utilized to determine and compare the deamination rates of 5-methylcytosine and cytosine in double-stranded DNA for the first time. The rate constants for spontaneous hydrolytic deamination of 5-methylcytosine and cytosine in double-stranded DNA at 37°C were $5.8 \times 10^{-13} \text{ s}^{-1}$ and $2.6 \times 10^{-13} \text{ s}^{-1}$, respectively. These rates are more than sufficient to explain the observed frequency of mutation at sites containing 5-methylcytosine and emphasize the importance of hydrolytic deamination as a major source of human mutations.

INTRODUCTION

5-Methylcytosine (5-mC) is produced by site-specific DNA (cytosine-5)-methyltransferases which transfer the methyl group from S-adenosylmethionine to the C-5 position of cytosine in double-stranded DNA. Cytosine methylation sites were first shown to be mutational hot spots in the *E. coli lacI* gene (1) and have subsequently been shown to be hot spots for mutation in both human germline (2,3) and somatic cells (4,5).

In *E. coli*, spontaneous deamination of 5-mC to thymine (T) is thought to be responsible for the observed mutational hot spots since G–T mismatches are repaired less efficiently than the G–U mismatches caused by deamination of cytosine (6). G–T mispairs in *E. coli* are corrected by very short patch (VSP) mismatch repair (7–10), which specifically repairs the G–T mismatches that arise at *dcm* methylation sites (CCWGG) by excising the mismatched thymine. VSP repair is less efficient at repairing G–U mispairs than the well-characterized uracil–DNA glycosylase (UDG) repair system (11,12). Similarly, G–T mispairs in eukaryotic cells are specifically repaired back to G–C base pairs by a process initiated by a thymine glycosylase (13–15). This eukaryotic G–T repair also is presumably less efficient than the eukaryotic UDG repair system (16–19). Thus, inefficient repair

of G–T mispairs generated at CpG dinucleotides, the eukaryotic DNA methylation locus, may account for observed mutational hot spots.

The original chemical studies on hydrolytic deamination at physiological pH of 2'-deoxy-5-methylcytidine-5'-monophosphate (5-methyl-dCMP) and 2'-deoxycytidine-5'-monophosphate (dCMP) showed that 5-methyl-dCMP deaminated at a 4- to 5-fold higher rate than dCMP (20). The mechanism for deamination of cytosine in acidic and neutral solutions was proposed to be initiated by protonation at the N-3 position of cytosine; followed by an addition–elimination reaction to complete the deamination at the N-4 position (21). 5-mC was also shown to deaminate at a faster rate than C in single-stranded DNA for several temperatures between 70°C and 95°C (22). The data from these higher temperatures was extrapolated to the lower physiological temperature of 37°C in an Arrhenius plot and indicated that the deamination rates for C and 5-mC in single-stranded DNA were approximately $2.1 \times 10^{-10} \text{ s}^{-1}$ and $9.5 \times 10^{-10} \text{ s}^{-1}$, respectively, which showed a nearly 5-fold faster rate for the hydrolytic deamination of 5-mC.

The deamination rate of C or 5-mC in double-stranded DNA at 37°C is too low (less than 10^{-12} s^{-1}) to be determined by the chemical methods used for free nucleotides or single-stranded DNA. Frederico *et al.* (23) devised a sensitive genetic assay primarily to measure deamination rates in single-stranded DNA, but they were also able to detect deamination of C in double-stranded DNA in a plasmid incubated for 300 days at 37°C. From this data, their derived rate constant for deamination of cytosine at 37°C was $7 \times 10^{-13} \text{ s}^{-1}$. However, their assay was not suitable for measuring the rate of deamination of 5-mC. We recently modified their method and developed a genetic assay with which we demonstrated that the HpaII methyltransferase itself can catalyze cytosine deamination in the absence of S-adenosylmethionine (24). We now report the use of this sensitive assay to compare the rates of mutation, and therefore indirectly, the rates of deamination of C and 5-mC in double-stranded DNA. Interestingly, the rates we have derived for hydrolytic deamination are several orders of magnitude more than sufficient to explain the occurrence of genetic disease-causing mutational hot spots in the germline and cancer-causing mutations in somatic tissues.

*To whom correspondence should be addressed

MATERIALS AND METHODS

E. coli strains

E. coli NR8052 [$\Delta(\text{pro-lac})$, *thi*, *ara*, *trp* E9777, *ung* I], a generous gift from Dr Thomas A. Kunkel (National Institute of Environmental Health Science), was used for transformation and scoring the reversion frequencies for the mutant pSV2-*neo* plasmids. *E. coli* JM109 [*F'*(*tra* D36), *lac* I Δ (*lac* Z) M15, *pro* AB/*rec* A1, *end* A1, *gyr* A96 (Nal^r), *thi*, *hsd* R17 (*r_k*-*m_k*-), *sup* E44, *e14*⁻ (*mcr* A⁻), *rel* A1, $\Delta(\text{lac-pro AB})$] was used to produce mutant pSV2-*neo* plasmids for the study of cytosine and 5-methylcytosine deamination.

Plasmids

The mutant pSV2-*neo* plasmid, in which Neo^R activity was inactivated by the creation of a new HpaII methylase (M.HpaII) site (CCGG, a missense mutation of Leu70 to Pro70 in the *neo*^r gene), was constructed by site-directed mutagenesis as described previously (24,25). Large preparations of the unmethylated mutant pSV2-*neo* plasmid were grown in *E. coli* JM109. The methylated mutant pSV2-*neo* plasmids were obtained by M.HpaII (New England BioLabs) methylation of the mutant pSV2-*neo*.

Preparation of plasmids

The plasmid purification of pSV2-*neo* mutants was done by the alkaline lysis method as described by Maniatis *et al.* (26). Methylation of the inner C at the target CCGG site was performed by treating the plasmids with M.HpaII and 80 μ M S-adenosylmethionine in 1 \times M.HpaII buffer at 37°C overnight. The unmethylated mutant pSV2-*neo* plasmids were also incubated in the same reaction solution without M. HpaII at 37°C overnight. Both unmethylated and methylated mutant pSV2-*neo* were then purified by phenol/chloroform (1:1) extraction, ethanol precipitation, and resuspension in TE buffer (10 mM Tris-HCl, pH 8.0 and 0.1 mM EDTA).

Incubation of DNA

Either unmethylated CCGG pSV2-*neo* or methylated C^mCCGG pSV2-*neo* was incubated in reaction buffer B (100 mM KCl, 50 mM HEPES-KOH, 10 mM MgCl₂, and 1 mM EDTA, pH 7.4) (20) at a concentration of 20 μ g/ μ l in a 0.6 ml centrifuge tube and overlaid with mineral oil. Typically, the incubation volume for each time point at various temperatures was 30 μ l. The incubations were carried out in a Brinkmann circulating water bath in the dark at 80°C, 70°C or 60°C (\pm 0.2°C) or in a VWR1720 incubator at 37°C (\pm 1°C). Each time point of these incubations was immediately frozen in a -70°C freezer for later transformation.

Transformation of plasmid

All of the plasmids were purified by ethanol precipitation from the incubation solution and dissolved in TE buffer for electrotransformation. Electroporations of *E. coli* NR8052 competent cells with 10 ng of plasmid were carried out with a Bio-Rad Electroporator under conditions described previously (24). Typically, plasmid transformations by electroporation had a high transformation efficiency (10^9 – 10^{10} transformants/ μ g DNA) thus allowing the detection sensitivity of this genetic assay to reach at least 10^{-7} (24).

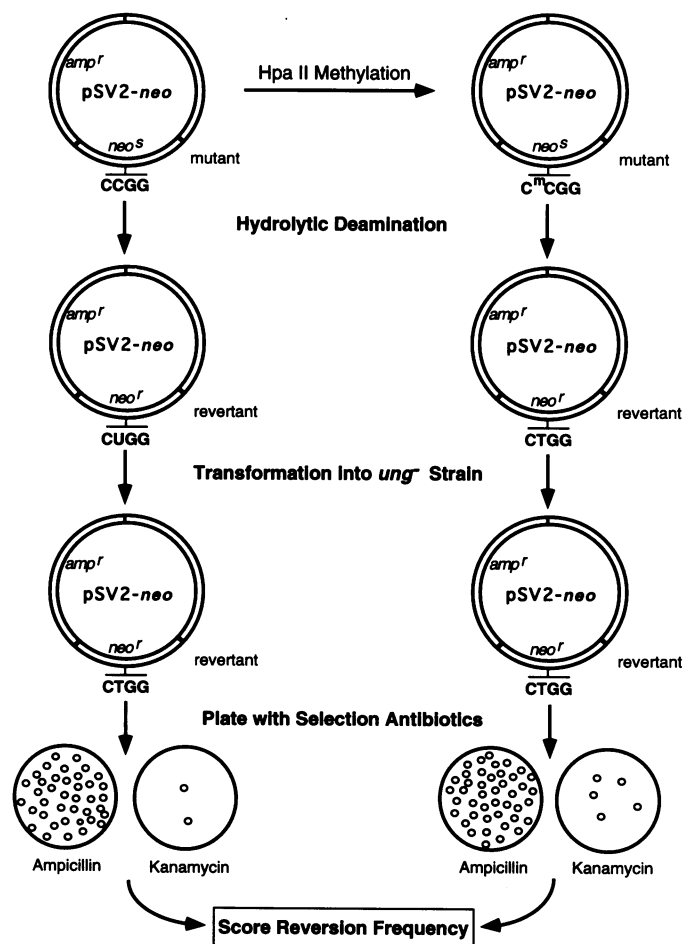


Figure 1. Genetic assay for the determination of 5-methylcytosine and cytosine deamination in double-stranded DNA. See text for rationale in detail.

Table 1. Deamination rate constants of cytosine and 5-methylcytosine in double-stranded DNA

Incubation Temperatures	Deamination Rate Constants (s^{-1}) ^a		Fold Difference
	C	5-mC	
80°C	$(2.6 \pm 0.4) \times 10^{-11}$	$(3.7 \pm 0.4) \times 10^{-11}$	1.4
70°C	$(1.1 \pm 0.1) \times 10^{-11}$	$(1.6 \pm 0.1) \times 10^{-11}$	1.5
60°C	$(4.7 \pm 0.3) \times 10^{-12}$	$(7.7 \pm 0.3) \times 10^{-12}$	1.6
37°C	2.6×10^{-13}	5.8×10^{-13}	2.2

^aDeamination rate constants, shown as mean \pm S.E., were determined from the time course plots of 80°C, 70°C and 60°C, except that the rate constants for 37°C were single determinations from incubations of 150 days each.

Determination of reversion frequency

The reversion assay used to detect C to U (or 5-mC to T) mutations was described previously in detail (24). Briefly, after electrotransformation, the cells were plated on SOB-agar plates containing 50 μ g/ml kanamycin while $1:10^4$ and $1:10^5$ dilutions of the cells were plated on SOB-agar plates containing 50 μ g/ml ampicillin. The plates were incubated overnight at 37°C to grow up the transformed bacterial colonies. The colony number scored from the ampicillin plates served as the total transformant number

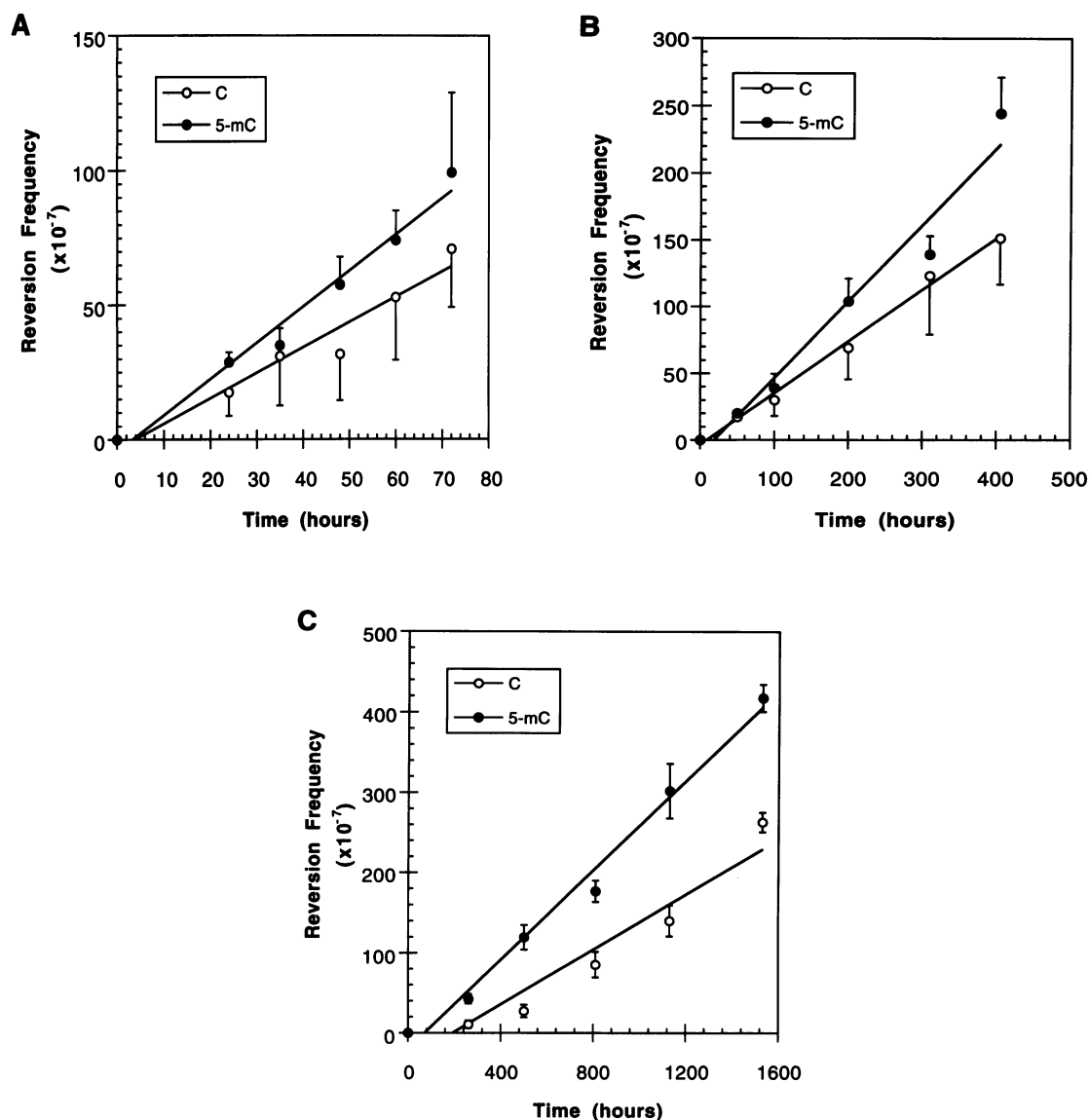


Figure 2. Deamination rates of 5-methylcytosine and cytosine in double-stranded DNA at 80°C, 70°C and 60°C. The reversion frequency at each time point in all figures was scored as described in Materials and Methods. (A) Deamination at 80°C. Results are shown as mean \pm S.D. ($n=4$) at each time point for either 5-methylcytosine (5-mC) or cytosine (C). The deamination rate constants were deduced from the slopes of each line. The correlation coefficient is 0.989 for the 5-mC line, and is 0.974 for the C line. (B) Deamination at 70°C. Results are shown as mean \pm S.D. ($n=6$); the correlation coefficients are 0.982 for the 5-mC line and 0.997 for the C line. (C) Deamination at 60°C. Results are shown as mean \pm S.D. ($n=4$); the correlation coefficients are 0.994 for the 5-mC line and 0.963 for the C line.

because the *amp^r* gene in the mutant pSV2-*neo* plasmid is not altered. The colony number scored from the kanamycin plates was the revertant number since the *neo^r* gene in the mutant pSV2-*neo* is inactivated, thus, the reversion frequency for each incubation can be determined by dividing the number of revertant colonies by the total number of transformed colonies.

RESULTS

Measurement of hydrolytic deamination of 5-mC and C in double-stranded DNA by a genetic assay

Our genetic assay (Fig. 1) was developed to measure and directly compare mutations induced at a specific site by deamination of 5-mC or C in double-stranded DNA. As described elsewhere in detail (24) and outlined in Fig. 1, the assay utilizes a mutant

pSV2-*neo* plasmid in which the neomycin resistance activity is inactivated by the creation of a new M.HpaII site (5'-CCGG-3') in the *neo^r* coding region resulting in a missense mutation. This design allowed us to convert the C residue to a 5-mC residue (CCGG versus C^mCGG) by treating the plasmid with M.HpaII (Fig. 1) thus allowing a direct comparison of the rates of deamination for C and 5-mC. Methylation of the inner C at the specific CCGG sequence by M.HpaII was confirmed by HpaII restriction endonuclease digestion as more than 99% methylated (result not shown). The mutant plasmid contains a T to C transition which changes codon 70 from CTGG for leucine to CCGG for proline. Codon 70 reverts to the wild-type CTGG sequence only when the inner C (or 5-mC) of the HpaII site on the sense strand deaminates to U and subsequently becomes T after DNA replication (or when 5-mC deaminates directly to T).

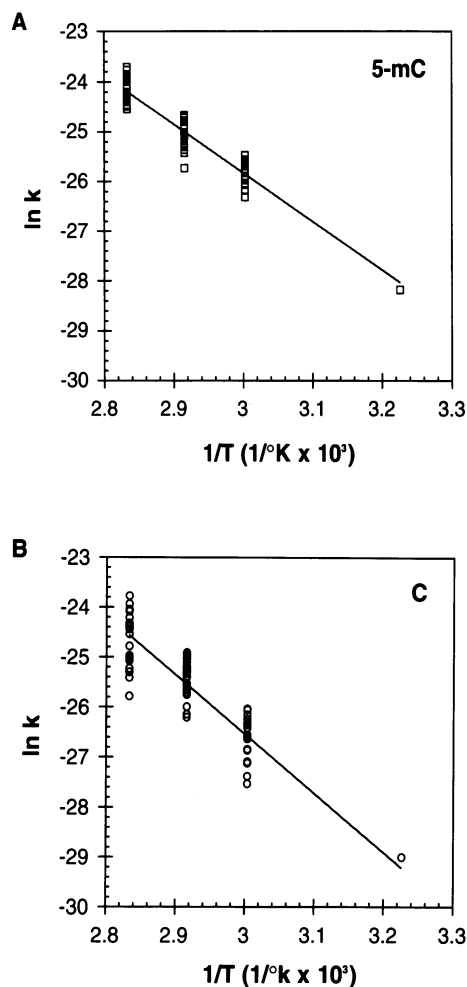


Figure 3. Arrhenius plot for determination of the activation energy for 5-mC and C deamination. Rate constants (k) are derived from several independent experiments at each time point at various temperatures. The slope was determined by linear regression for the calculation of activation energy. (A) 5-mC deamination. The activation energy for 5-mC deamination is (19.5 ± 0.8) kcal/mole. The correlation coefficient is 0.952. (B) C deamination. The activation energy for C deamination is (23.5 ± 1.6) kcal/mole. The correlation coefficient is 0.883.

Deamination at the inner C of this HpaII site on the anti-sense strand results in a sense 5'-CCAG-3' sequence and does not change the mutant amino acid proline. Likewise, a C to T transition in the outer C of the CCGG site on the sense strand gives rise to a serine codon, TCGG, which does not restore Neo^R and consequently does not affect the specificity of the assay.

Both unmethylated (C) and methylated (5-mC) mutant pSV2-*neo* were incubated in the same reaction buffer (buffer B) at various temperatures and the extent of deamination of C and 5-mC were determined at various times by electroporation of the plasmids into the indicator bacteria *ung*⁻ *E. coli* (Fig. 1). These bacteria lack the capacity to initiate uracil excision repair (27) because they lack uracil-DNA glycosylase as demonstrated by less than 0.1% residual UDG activity compared with the wild-type strain. Additionally, they will not correct G-T mismatches because the *dam* methylation-directed long patch repair system does not function on fully *dam* methylated DNA (methylated adenine on both strands of GATC sequences) (28,29). Our

reporter plasmids were fully methylated at all *dam* methylation sites since they were grown in a *dam*⁺ strain (JM109) and digestion with the *dam* methylation sensitive enzyme MboI showed no cutting of the purified plasmids (data not shown). Thus, the assay using the indicator bacteria accurately measures the rates of *in vitro* hydrolytic deamination for 5-mC and C.

The frequencies of hydrolytic deamination for 5-mC and C in double-stranded DNA were determined by incubating unmethylated and HpaII methylated plasmids for various times at 80°C, 70°C and 60°C (Fig. 2A-C). The deamination rate constants of 5-mC or C at each temperature were derived from the slopes of the plots and are summarized in Table 1. Table 1 also shows results obtained from a single time point for DNA incubated at 37°C for 150 days. The reversion frequencies for 5-mC and C at 37°C for 150 days are 7.5×10^{-6} and 3.3×10^{-6} , respectively; the calculated rate constants and fold-difference between 5-mC and C are also shown in Table 1. 5-mC deaminated at a 1.4- to 2.2-fold higher rate than C at all temperatures studied (Table 1).

The activation energies for 5-mC and C deamination in double-stranded DNA were derived from the deamination rate constants for 5-mC and C obtained at various temperatures since deamination rates are temperature-dependent (Fig. 3A and Fig. 3B). The slopes were determined by linear regression and gave activation energies for 5-mC and C deamination in double-stranded DNA of (19.5 ± 0.8) kcal/mole and (23.5 ± 1.6) kcal/mole, respectively.

DISCUSSION

This paper reports the rates of deamination of 5-mC and C at the same locus in double-stranded DNA for the first time. The deamination of 5-mC and C in double-stranded DNA at 37°C and pH 7.4 proceeded at rate constants of $5.8 \times 10^{-13} \text{ s}^{-1}$ (a half life of about 38,000 years) and $2.6 \times 10^{-13} \text{ s}^{-1}$ (a half-life of about 85,000 years), respectively. The rate constant of C deamination measured by our assay differs from the previous result of Frederico *et al.* ($7 \times 10^{-13} \text{ s}^{-1}$) by a factor of less than three, a minor difference which may be due to experimental variation. Our data showing a 2- to 3-fold greater rate of deamination at 5-mC than C in double-stranded DNA are consistent with an earlier report on deamination in single-stranded DNA in which results from high temperature incubations were extrapolated to 37°C yielding a calculated 4- to 5-fold greater deamination rate for 5-mC than C (22).

The lower activation energy for the deamination of 5-mC (20 kcal/mole) relative to C (24 kcal/mole) in double-stranded DNA are consistent with prior results in single-stranded DNA, in which again 5-mC has a lower activation energy than C [24 kcal/mole for 5-mC and 27 kcal/mole for C (22)]. However, the lower activation energy for the deamination of both bases in double-stranded compared to single-stranded DNA was surprising since, lower activation energies generally coincide with faster rate constants. An enthalpy-entropy compensation (30-32) for the organization of base-stacking between single- and double-helical DNA might explain this difference.

Koeberl *et al.* (33) have estimated a mutation rate of 3.7×10^{-8} per base per generation for 5-mC transitions in human DNA. Assuming a generation time of 20 years, this translates into a mutation rate of about $5.8 \times 10^{-17} \text{ s}^{-1}$ at each CpG site after DNA repair. While it is difficult to be certain about the accuracy of this number, it is clearly considerably smaller

(about 10^4 -fold) than the rate of hydrolytic deamination we have determined for 5-mC. The data support the point made recently by Lindahl (34) and Smith (35) that spontaneous decay of DNA may be a major factor in mutagenesis. In this regard it should be noted that CpG sites are also hot spots for mutation in tumor suppressor genes in somatic cells (4,5). Hydrolytic deamination may therefore contribute significantly to the generation of cancer.

The two fold difference in the rate of hydrolytic deamination of 5-mC relative to C is not by itself sufficient to explain the occurrence of mutational hot spots at 5-mC since the CpG dinucleotide is about 12- to 42-fold more mutable than other sequences in cells (2,3,33,36). Because the spontaneous deamination rate of 5-mC or C is several orders of magnitude more than sufficient to account for the observed germline mutation frequency, the DNA repair systems in living cells must be efficient enough to correct the natural decay of DNA. The data support the idea that the increased mutability of 5-mC relative to C is not only due to the increased rate of deamination but also due to the fact that the G-U mismatch repair system (37) is more efficient than G-T repair (13). In addition, methyltransferase-catalyzed cytosine deamination (24) may also contribute to mutational hot spots at CpG. When the concentration of S-adenosylmethionine is limiting (24), cytosine deamination catalyzed by the HpaII methyltransferase occurs at 10^4 -fold higher rate than the spontaneous deamination of cytosine. Thus the enzyme-induced deamination at CpGs may increase the burden for the G-U mismatch repair system.

In summary, the spontaneous deamination rate of 5-mC in double-stranded DNA, determined for the first time in this study, is vastly in excess of that needed to explain the frequent occurrence of mutational hot spots at CpG dinucleotides in vertebrates. The data emphasize the importance of DNA repair in protecting cells from spontaneous mutagenesis, and that the leakage in repair efficiency seems to be a major factor contributing to mutagenesis at 5-mC.

REFERENCES

1. Coulondre, C., Miller, J. H., Farabaugh, P. J. and Gilbert, W. (1978) *Nature*, **274**, 775-780.
2. Cooper, D. N. and Youssoufian, H. (1988) *Hum. Genet.*, **78**, 151-155.
3. Sommer, S. S. (1992) *FASEB J.*, **6**, 2767-2774.
4. Rideout, W. M., III, Coetzee, G. A., Olumi, A. F. and Jones, P. A. (1990) *Science*, **249**, 1288-1290.
5. Jones, P. A., Rideout, W. M., III, Shen, J. -C., Spruck, C. H. and Tsai, Y. C. (1992) *BioEssays*, **14**, 33-36.
6. Duncan, B. K. and Miller, J. H. (1980) *Nature*, **287**, 560-561.
7. Lieb, M. (1983) *Mol. Gen. Genet.*, **191**, 118-125.
8. Lieb, M. (1987) *J. Bacteriol.*, **169**, 5241-5246.
9. Jones, M., Wagner, R. and Radman, M. (1987) *J. Mol. Biol.*, **194**, 155-159.
10. Lieb, M. (1991) *Genetics*, **128**, 23-27.
11. Lindahl, T., Ljungquist, S., Siebert, W., Nyberg, B. and Sperens, B. (1977) *J. Biol. Chem.*, **252**, 3286-3294.
12. Lindahl, T. (1979) *Progr. Nucleic Acids Res. Mol. Biol.*, **22**, 135-192.
13. Brown, T. C. and Jiricny, J. (1987) *Cell*, **50**, 945-950.
14. Brown, T. C. and Jiricny, J. (1988) *Cell*, **54**, 705-711.
15. Wiebauer, K. and Jiricny, J. (1989) *Nature*, **339**, 234-236.
16. Caradonna, S. J. and Cheng, Y. C. (1980) *J. Biol. Chem.*, **255**, 2293-2300.
17. Crosby, B., Prakash, L., Davies, H. and Hinkle, D. C. (1981) *Nucleic Acids Res.*, **9**, 5797-5809.
18. Seal, G., Arenaz, P. and Sirover, M. A. (1987) *Biochim. Biophys. Acta*, **925**, 226-233.
19. Morgan, A. R. and Chlebik, J. (1989) *J. Biol. Chem.*, **264**, 9911-9914.
20. Lindahl, T. and Nyberg, B. (1974) *Biochemistry*, **13**, 3405-3410.
21. Shapiro, R. and Klein, R. S. (1966) *Biochemistry*, **5**, 2358-2362.
22. Ehrlich, M., Norris, K. F., Wang, R. Y. -H., Kuo, K. C. and Gehrke, C. W. (1986) *Biosci. Rep.*, **6**, 387-393.
23. Frederico, L. A., Kunkel, T. A. and Shaw, B. R. (1990) *Biochemistry*, **29**, 2532-2537.
24. Shen, J. -C., Rideout, W. M., III and Jones, P. A. (1992) *Cell*, **71**, 1073-1080.
25. Southern, P. J. and Berg, P. (1982) *J. Mol. Appl. Genet.*, **1**, 327-341.
26. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
27. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488-492.
28. Pukkila, P. J., Peterson, J., Herman, G., Modrich, P. and Meselson, M. (1983) *Genetics*, **104**, 571-582.
29. Radman, M. and Wagner, R. (1986) *Ann. Rev. Genet.*, **20**, 523-538.
30. Breslauer, K. J., Remeta, D. P., Chou, W. -Y., Ferrante, R., Curry, J., Zaunczkowski, D., Snyder, J. G. and Marky, L. A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8922-8926.
31. Searle, M. S. and Williams, D. H. (1992) *J. Am. Chem. Soc.*, **114**, 10690-10697.
32. Searle, M. S. and Williams, D. H. (1993) *Nucleic Acids Res.*, **21**, 2051-2056.
33. Koeberl, D. D., Bottema, C. D. K., Ketterling, R. P., Bridge, P. J., Lillicrap, D. P. and Sommer, S. S. (1990) *Am. J. Hum. Genet.*, **47**, 202-217.
34. Lindahl, T. (1993) *Nature*, **362**, 709-715.
35. Smith, K. C. (1992) *Mut. Res.*, **277**, 139-162.
36. Sved, J. and Bird, A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4692-4696.
37. Lindahl, T. (1982) *Ann. Rev. Biochem.*, **51**, 61-87.