A comparison of the fidelity of copying 5-methylcytosine and cytosine at a defined DNA template site

Jiang-Cheng Shen, Steven Creighton¹, Peter A. Jones and Myron F. Goodman^{1*}

Department of Biochemistry and Comprehensive Cancer Center, University of Southern California School of Medicine, Los Angeles, CA 90033 and ¹Department of Biological Sciences, Molecular Biology Section, University Park, Los Angeles, CA 90089-1340, USA

Received June 24, 1992; Accepted August 24, 1992

ABSTRACT

5-Methylcytosine has been postulated to be an endogenous mutagen in procaryotes and eucaryotes leading to base substitution hot spots, C→T transitions, resulting from spontaneous deamination of ^mC to T. The possibility remains, however, that a second mechanism involving mispairing of ^mC with A might also contribute to base substitution mutagenesis via G→A transitions. Stimulation of the G→A mutational pathway could involve preferential misincorporation of dAMP opposite template ^mC compared to C. To investigate this possibility, we synthesized a sequence containing ^mC at a defined template location. We compared the fidelity of copying ^mC versus C and the efficiency of extending mismatched base pairs at the ^mC position using three DNA polymerases, AMV reverse transcriptase, *Drosophila* DNA polymerase α , and mutant *Escherichia coli* Klenow fragment containing no proofreading exonuclease activity. Significant differences in misinsertion and mismatch extension efficiencies were observed only for the case of AMV reverse transcriptase. AMV reverse transcriptase was observed to incorporate dAMP 4 to 5-fold more efficiently opposite ^mC than C. Favored extension of a 5-MeC·A over C·A mispair was also observed with a difference of about 3-fold. In contrast to AMV reverse transcriptase, Klenow fragment showed no significant difference when copying either ^mC or C sites or when extending mispairs involving ^mC and C. Incorporation of dAMP opposite either C or ^mC was barely detectable using pol α , although pol α has been observed to form A·C mismatches in other sequences. While we cannot completely exclude the possibility that dAMP might be incorporated opposite ^mC in preference to C, our results suggest that contributions of the G→A pathway to ^mC mutagenic hot spots are likely to be minor, lending additional support to the model invoking deamination of ^mC.

INTRODUCTION

An investigation of the mutagenic spectrum in the *Escherichia coli* *lacI* gene showed that greatly elevated mutation frequencies occurred at DNA loci containing 5-methyl cytosine (^mC) (1). It is known that an efficient repair system, involving the concerted action of uracil-N-glycosylase, endo- and exonucleases and polymerase is present in *E. coli* to remove uracil arising either by deamination of C or from incorporation of dUMP by polymerase and restore the original DNA strand (2,3). In contrast to the highly efficient elimination of uracil from DNA, Coulondre *et al.* (1) proposed that thymine, arising from deamination of ^mC, would be less likely to be repaired and would therefore lead to C→T base substitution hot spots. However, subsequent to the proposed deamination model, systems were found in *E. coli* that could repair T·G mispairs; e. g. the 'very short patch' (VSP) repair system is specific for removal of T in a T·G heteroduplex (4,5).

The ^mC deamination model has been generally accepted in the mutagenesis field. The model provides a direct logical explanation of *in vivo* mutagenic observations, and is further supported by a body of chemical evidence showing that deamination of C occurs spontaneously *in vitro* (6–8), this reaction being most pronounced in regions where DNA is melted (7–10). However, there remained the formal possibility that dAMP is incorporated more frequently opposite ^mC than C, and thus a lowered fidelity of copying ^mC might be a contributing factor to the occurrence of C→T hot spots. In this paper, we have used AMV RT, *Drosophila* DNA polymerase α , and *E. coli* Klenow fragment to measure the efficiencies of forming A·^mC and A·C mispairs and continued elongation of the mispairs by addition of a next correct nucleotide.

MATERIALS AND METHODS

Enzymes

AMV reverse transcriptase was purchased from U.S. Biochemical Corp.. Purified *D. melanogaster* DNA polymerase α was a generous gift of Dr. I. R. Lehman, Stanford University. The

* To whom correspondence should be addressed

a. Running Start Insertion

b. Standing Start Extension

Figure 1. Primer-template system used to measure reaction kinetics for copying 5-methylcytosine and cytosine. a, Running start insertion at cytosine and 5-methylcytosine. Primer in each reaction is first extended with 2 bases by insertion of dCMP at site 1 and site 2 under the 'saturated' concentration of dCTP. Various concentrations of dGTP (right) or dATP (wrong) are used for the insertion at site 3. The reaction velocity versus [dNTP] is determined for insertion opposite template base (C or mC) at site 3. b, Standing start extension of 5-methylcytosine and cytosine. The terminal base of the primer is either G (right) or A (wrong) opposite to the template base C or mC. The velocity of primer extension (%/min) versus [dNTP] is measured by adding various concentrations of dTTP for extending the next correct base opposite template base (A) at site 1.

intact pol α holoenzyme used in this assay was devoid of detectable 3' to 5' proofreading exonuclease activity. Klenow fragment of *E. coli* DNA polymerase I (D355A, E357A) without 3' to 5' exonuclease activity was purchased from U.S. Biochemical Corp. T4-polynucleotide kinase for 5'-end labeling of primers was purchased from Boehringer Mannheim, Germany.

Primers and templates

All the primers and templates were synthesized with trityl group on an Applied Biosystems Inc. (ABI) 381A DNA synthesizer. Purification of the tritylated primers by HPLC and polyacrylamide gel electrophoresis has been described (11). The template synthesized for the insertion and extension kinetic studies was a 30-nucleotide oligomer positioned on the wild type single-stranded M13 DNA ranging from 1797 to 1826, i.e. 5'-CTC-AGTGTACGGTACATGGGTTCTATTG-3', or a cytosine residue was replaced by a 5-methylcytosine on position 1816, i.e. 5'-CTCAGTGTTA^mCGGTACATGGGTTCTATTG-3'. The primers used in either a running start insertion or a standing start extension of the AMV reverse transcriptase are complementary to the templates, as shown in Fig. 1. The primer for the running start insertion is a 17-mer complementary to the 30-mer template, i.e. 5'-CAATAGGAACCCATGTA-3'. The primers for the standing start extension are both 19-mers

complementary to the 30-mer template with the 3'-end nucleotide matched or mismatched to the template, i.e. 5'-CAATAGGAACCCATGTACCG-3' for the matched extension and 5'-CAATAGGAACCCATGTACCA-3' for the mismatched extension.

Chemicals

Non-radioactive HPLC-purified dNTP substrates were purchased from Pharmacia LKB Biotechnology, Inc.. Radioactive [γ -³²P]ATP (4500Ci/mmol) was purchased from ICN Radiochemicals, Inc. 2-Aminopurine deoxyribonucleoside triphosphate (2-dATP) was prepared as described (12).

Reaction conditions

The procedures for the radioactive [γ -³²P]ATP end labeling of primers, primer-template annealing, and polymerase reaction have been described in detail previously (13). The reaction conditions and procedures for measuring V_{max}/K_m values of the running start insertion and standing start extension using AMV reverse transcriptase have been described elsewhere (13–15). For the AMV reverse transcriptase reaction, a typical reaction mixture (6 μ l) contains 60 mM Tris-HCl, pH 8.3, 75 mM NaCl, 8 mM MgCl₂, 0.5 mM DTT, 200 nM annealed primer-template DNA, 75 nM AMV reverse transcriptase, and various concentrations of dNTP. For the *Drosophila* Pol α reaction, the reaction mixture (6 μ l) contains 50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 40 mM NaCl, 0.1 mM EDTA, 10 mM DTT, 100 mg/ml bovine serum albumin, 4 nM Pol α , 200 nM annealed primer-template DNA, and various concentrations of dNTP. For the mutant Klenow fragment reaction, a typical reaction mixture (6 μ l) contains 50 mM Tris-HCl, pH 7.5, 7.3 mM MgCl₂, 1 mM DTT, 0.005 U mutant Klenow fragment, 200 nM annealed primer-template DNA, and various concentrations of dNTP. All of these reactions mentioned above are incubated in water bath at 37°C for desired time periods.

Reaction Time

The reaction time for each different type (running start or standing start) and different enzyme reaction was determined by doing a time course study to obtain the time range during which the rise in intensity of the extended primer is linear with time increase and less than 30% of the original primer band (I_0) was extended. Based on the time course results, AMV reverse transcriptase reaction time was chosen to be 5 min for the running start insertion and 1 min for the standing start extension on the incorporation of dNTP. The incorporation of 2-aminopurine deoxyribonucleoside triphosphate (2-dATP) generated by AMV reverse transcriptase requires 4 min for both running start insertion and standing start extension. For the mutant Klenow fragment, a 4 min reaction was chosen for the running start insertion and standing start extension, and also for the special 2-dATP incorporation in both reactions.

Gel electrophoresis, autoradiography and densitometry

After the polymerase reaction, 12 μ l of 20 mM EDTA in 95% formamide was added to each typical reaction mixture (6 μ l) to terminate the reaction. This final reaction mixture was heated at 100°C for 5 min, cooled on ice, and a 4 μ l aliquot was loaded on a 16% polyacrylamide gel containing 8 M urea. Gels (40×30×0.04 cm) are run at 2000 V for 4–5 hr to obtain the best resolution of elongated primers. Gels were dried and autoradiography were made as described (13) for the scanning by a densitometer.

Figure 2. Gel autoradiogram showing running start insertion at 5-methylcytosine versus cytosine by AMV reverse transcriptase. The primer (P) extensions with 3 bases opposite to the template are shown on the left margin of each autoradiogram. A 'saturated' concentration of dCTP (10 μM) was applied in each running start insertion for primer extension at site 1 and site 2, and allowing the extension reach the target site (site 3 with C or ^mC on template). The various concentrations of insertion bases at the target sites are shown below the primer band (P) in each lane. The AMV reverse transcriptase reactions were run for 5 min at various concentrations of dATP (misinsertion) and dGTP (correct insertion). a, Misinsertions of A opposite C versus A opposite ^mC by AMV reverse transcriptase are shown. The Michaelis-Menten plots with relative velocity, $v = I_3/I_2$ measured by densitometry, as a function of dATP concentration are shown below each autoradiogram of A • C mispair and A • ^mC mispair. b, Correct insertions of G opposite C and ^mC by AMV reverse transcriptase are shown. The first lane from left showing 0 mM of dGTP is primer itself without addition of dCTP in reaction, and the second lane showing 0 mM of dGTP is reaction with addition of 10 μM of dCTP. Both lanes are used as control in this experiment. A thin band appearing between bands 2 and 3 in each lane including the primer lane is a contaminant that occurred during primer synthesis.

Velocity determination for a running start insertion

To determine the velocity of dNTP incorporation opposite the template target site (band 3) when extending the primer by three bases, the equation $V_{23} = (I_3/I_2)(I_3 + I_2)/t$ was applied, where I_2 is the integrated intensity of band 2 ahead of the target band and t is the reaction time in min.

Velocity determination for a standing start extension

The velocity of dNTP incorporation when extending the primer (band 0) by inserting a base opposite the first template base (band 1) was determined by the terms of the integrated gel band intensities, as $V_{01} = (100I_1)/(I_0 + 0.5I_1)t$, where t is the reaction time in min.

RESULTS

We investigated two aspects of fidelity of copying DNA templates containing ^mC compared to C, the kinetics of nucleotide insertion and extension. For the insertion reaction, a 5'-³²P-labeled primer is rapidly extended by adding two correct

nucleotides, followed by either a correct (G) or incorrect (A) insertion at the template target site C or ^mC (Fig. 1a). In the reaction, dCTP is held at constant saturating concentration; steady-state kinetic measurements for nucleotide insertion opposite C or ^mC were carried out by varying the concentrations of dGTP or dATP in separate reactions (13). For the extension reaction, the primer contained either a correct (G) or incorrect (A) 3'-terminus located opposite template C or ^mC, and the rates of extending matched *versus* mismatched primer termini were measured by varying the concentration of next correct nucleotide, dTTP (Fig. 1b). For both insertion and extension measurements, a measurement of the ratio V_{max}/K_m for 'right' *versus* 'wrong' reactions suffices to determine insertion (13,14) and extension (15,16) fidelities.

Comparing misinsertion of A opposite C and ^mC

Polyacrylamide gel autoradiograms depicting ³²P-primer molecules elongated by AMV RT are shown in Fig. 2. Misinsertion of A opposite C and ^mC and correct insertion of

G opposite C and ^mC , measured as a function of $[\text{dNTP}]$, are shown in Figs. 2a and 2b respectively, in a running start reaction (see Materials and Methods). For a reaction period of 5 min, dAMP is misincorporated opposite either C or ^mC over a range of dATP concentrations from 50 μM to 3.2 mM. There does not appear to be significant substrate inhibition of AMV RT in the mM $[\text{dATP}]$ range. The relative nucleotide insertion velocity is obtained as the ratio of integrated gel band intensities at the target (I_3) and and previous (I_2) sites, where the velocity is proportional to I_3/I_2 (see Materials and Methods). The apparent second order rate constant for the insertion reaction, $V_{\max}/K_m$,

Table 1. The apparent $V_{\max}/K_m$ values and nucleotide misinsertion ratios (f) for the running start insertions by AMV reverse transcriptase

Primer/Template ^a	$V_{\max}/K_m$ ^b	Misinsertion Ratio (f) ^c
	%/min · mM	
5'-----CCG 5'-----CC↓ 3'-----GGC----- G	3.2 ± 1.1	1
5'-----CC↓ 3'-----GG ^m C----- A	1.3 ± 0.1	1
5'-----CC↓ 3'-----GGC----- A	1.8 (±0.2) × 10 ⁻⁴	5.8 (±1.2) × 10 ⁻⁵
5'-----CC↓ 3'-----GG ^m C-----	3.4 (±0.1) × 10 ⁻⁴	2.6 (±0.2) × 10 ⁻⁴

^aThe nucleotide base above the downward arrow is the one used to measure the insertion velocity as a function of $[\text{dNTP}]$. In this running start reaction, an additional dCTP is provided so that polymerase can extend primer by two nucleotides to reach the insertion target site.

^b $V_{\max}/K_m$ is obtained by a least squares fit of the kinetic data to the linear region of the Michaelis-Menten curve. Values shown with ± S.D. are average values obtained from two separate experiments.

^cEvaluated from $V_{\max}/K_m$ ratios by Equation 1.

is obtained by a least squares fit of the kinetic data to the linear region of the Michaelis-Menten curve (Fig. 2a). The ratio of wrong to right insertions is given by,

$$f_{\text{ins}} = (V_{\max}/K_m)_w / (V_{\max}/K_m)_r \quad (1)$$

The insertion fidelity, $F_{\text{ins}} = 1/f_{\text{ins}}$.

The rate of misinserting A opposite ^mC appears to be about 2-fold greater than A opposite C, while correct insertion of G opposite C appears to be roughly 2.5-fold greater than G opposite ^mC (Table 1). The misinsertion ratio to form A · ^mC mispairs is $f_{\text{ins}}(\text{A} \cdot ^m\text{C}) = 2.6 \times 10^{-4}$, and to form A · C mispairs, $f_{\text{ins}}(\text{A} \cdot \text{C}) = 5.8 \times 10^{-5}$ (Table 1). Thus, using the primer-templates shown in Fig. 1a, AMV RT appears to favor A · ^mC mispairs with about a 4.5-fold higher efficiency than A · C mispairs.

Comparing extension of A opposite C and ^mC

In a second set of experiments, we measured the relative efficiencies of extending the two mispairs. For the extension data, we annealed a 5'-³²P-labeled primer molecule containing either a preformed mismatched base, A, or correct match, G, at its 3'-terminus located opposite C (Fig. 1b). The kinetics of extending the correctly matched G · C and G · ^mC or mismatched A · C or A · ^mC base pairs were measured by varying the concentration of the next correct nucleotide, dTTP (Fig. 1b), and measuring the percent primer extended by addition of a single nucleotide for a standing start reaction (see Materials and Methods). Extensions of mismatched A · C, A · ^mC and correctly matched G · C, G · ^mC base pairs as a function of $[\text{dTTP}]$ are shown in Figs. 3a and 3b respectively. Apparent $V_{\max}/K_m$ values required to obtain primer extension ratios were obtained by a least squares fit to the slope of linear portion of the Michaelis-Menten curve, v (i.e., percent of primer extended) versus $[\text{dTTP}]$.

The relative efficiency for extending mispairs, f° , is given by the ratio of $V_{\max}/K_m$ values comparing mispaired to correctly paired termini,

Figure 3. Gel autoradiogram showing standing start extension of 5-methylcytosine versus cytosine by AMV reverse transcriptase. The terminal base and the extended base of primer (P) pairing with the opposite bases of the template are shown on the left margin of each autoradiogram, and various concentrations of dTTP used for extension from target sites are shown below the primer band (P) in each lane. The AMV reverse transcriptase reactions were run for 1 min at various concentrations of dTTP. a, Extension of A · C mispair and A · ^mC mispair with the next correct base T are shown. b, Extension of G · C and G · ^mC pairs with the next correct base T are shown.

$$f^{\circ}\text{ext} = (V_{\max}/K_m)_{\text{mis}}/(V_{\max}/K_m)_{\text{corr}} \quad (2)$$

In contrast to the nucleotide misinsertion ratio, f_{ins} , which does not depend on the polymerase-DNA equilibrium binding constant, and is independent of the absolute concentration of 'right' and 'wrong' [dNTP] substrates (13), the ratio of mismatch to correct match extension efficiencies is dependent on both K_D and [dNTP] concentration (15). $f^{\circ}\text{ext}$ is defined as the intrinsic

Table 2. The apparent $V_{\max}/K_m$ values and relative efficiencies of extension (f°) for the standing start extensions by AMV reverse transcriptase

Primer/Template ^a	$V_{\max}/K_m$ ^b	Relative Efficiency of Extension (f°) ^c
	%/min·mM	
5'-----T 3'-----CA----- T	0.17 ± 0.09	1
5'-----G↓ 3'----- ^m CA----- T	0.31 ± 0.14	1
5'-----A↓ 3'-----CA----- T	$3.4 (\pm 0.7) \times 10^{-4}$	$2.0 (\pm 0.7) \times 10^{-3}$
5'-----A↓ 3'----- ^m CA----- T	$9.9 (\pm 2.5) \times 10^{-4}$	$3.2 (\pm 1.1) \times 10^{-3}$

^aNucleotide base (T) above the downward arrow is the next correct base for primer extension. The terminal base of primer is either matched or mismatched to the template as shown in each primer/template site.

^b $V_{\max}/K_m$ is obtained by a least squares fit of the kinetic data to the linear region of the Michaelis-Menten curve. Values shown with ± S.D. are average values obtained from 2 (matched extension) or 3 (mismatched extension) separate experiments.

^cEvaluated from $V_{\max}/K_m$ ratios by Equation 2.

mismatch/correct match extension efficiency, which measures the direct competition for polymerase to extend a terminal mispair in the presence of a correct pair when both are present at equimolar concentrations, and when the polymerase binds with equal affinities to each (15). We have recently shown that K_D values are similar for AMV RT binding to all natural matched and mismatched base pairs (17). One can also use equation 2 to compare rates of extending *any* primer-template termini. For the two mismatched termini, the extension efficiency for A·^mC mispairs is about 3-fold higher than A·C, while the efficiency to extend correctly matched termini is about 2-fold higher for G·^mC compared to G·C (Table 2).

Insertion and extension using *Drosophila* pol α and *E.coli* Klenow fragment

The fidelity of nucleotide insertion and primer extension depends on surrounding base sequence (14) and on the identity of the polymerase. Compared to the relative 'ease' with which AMV RT can insert A opposite template C and ^mC, *Drosophila* pol α has considerable difficulty carrying out this reaction on this DNA sequence. Insertion of dAMP opposite C and ^mC for times ranging from 30 sec to 32 min is shown in Fig. 4a. Extension of the ³²P-labeled primer to the C and ^mC template target sites can just begin to be detected following an 8 min incubation period. Even after 32 min, the formation of A·C and A·^mC base pairs is less than observed for a 5 min incubation using AMV RT. Pol α had no difficulty making the two correct insertions, C opposite G, to reach the target C or ^mC target site (Fig. 4a). Insertions of G opposite C or ^mC occurs (Fig. 4b), suggesting that the difficulty in catalyzing A·C and A·^mC mismatches was not the result of having an inactive or unstable pol α. We note that pol α has greater difficulty inserting G opposite ^mC than opposite C (Fig. 4b).

Figure 4. Gel autoradiogram showing time course analysis of running start insertion at 5-methylcytosine versus cytosine by *Drosophila* DNA polymerase α. A 'saturated' concentration of dCTP (10 μM) was introduced to each reaction to allow primer extension to reach the target site (C or ^mC on the template). The extended bases of primer (P) opposite template are shown on the left margin of each gel autoradiogram. Different time points of reactions are shown below the primer band (P) in each lane. a, Misinsertions of A opposite C and A opposite ^mC by *Drosophila* Pol α are shown. A 'saturated' concentration of dATP (1 mM) was added in both reactions for the insertion of dAMP at the target sites (C or ^mC) of different templates. A·C and A·^mC mispairs are barely seen on the autoradiograms after an 8 min reaction. b, Correct insertion of G opposite C and G opposite ^mC by *Drosophila* Pol α are shown. A 'saturated' concentration of dGTP (10 μM) was added in both reactions for the insertion of dGMP at the target sites (C or ^mC on the template). The correct insertions of G·C and G·^mC appear less than 1 min of reaction time as seen on the autoradiograms.

We also measured the kinetics of misinserting dAMP opposite C and m C using *E. coli* Klenow fragment, lacking detectable 3' to 5' exonuclease activity. The ratios of $V_{\max}/K_m$ were 7.7×10^{-3} and 8.3×10^{-3} for insertion of A opposite C and m C respectively, which is about 20- to 40- fold larger than observed for AMV RT. Similar to the case of AMV RT which exhibited an approximate 2-fold higher insertion efficiency for dAMP misinsertion opposite m C compared to C, there appears to be little discernible bias favoring misinsertion opposite m C by Klenow fragment, at most 1.1-fold. These differences in misinsertion efficiencies were so small that it did not seem relevant to calculate f_{ins} (Eq. 1), which requires measuring the efficiencies of insertion dGMP opposite C and m C. We did, however, determine that there were no significant differences in the efficiencies to extend correctly matched G·C and G· m C primer termini (Fig. 1b; data not shown).

The m C template sites giving rise to elevated levels of spontaneous mutations were also found to be hot spots for 2-aminopurine (2-AP) mutagenesis (1,18,19). Therefore, using the substrate 2-aminopurine deoxyribonucleoside triphosphate (2-dAPTP), we measured rates of incorporation and extension of 2-AP at template C and m C with the same templates used to measure incorporation and extension of A. The $V_{\max}/K_m$ values for insertion of 2-AP opposite C and m C are $(3.2 \pm 0.1) \times 10^{-3}$ and $(6.4 \pm 2.5) \times 10^{-3}$ respectively. In accord with its documented mutagenic activity, 2-AP is inserted 20-fold more efficiently opposite both C and m C than is A.

DISCUSSION

An analysis of the mutational spectrum of the *lacI* gene in *E. coli* revealed that high mutation rates occurred at m C DNA template loci (1,20). Subsequently, anomalously high mutation rates have been correlated with m C sites in procaryotic and eucaryotic organisms (21–24). The widely accepted model to account for these observations was originally proposed to explain the *lacI* mutational data (20). J. Miller and colleagues (20) proposed that m C→T mutagenic hot spots could be accounted for by deamination of m C to form T. Among the various possibilities leading to high mutation rates are: (i) heteroduplex T·G mispaired intermediates may be less likely to be repaired than their U·G counterparts (resulting from deamination of C) because the pathway for removal of uracil in DNA, via uracil N-glycosylase (2,25), is highly efficient (26,27); (ii) deamination reactions are known to occur spontaneously in vitro, and m C is deaminated more rapidly than C (8).

There are postreplication biochemical pathways available to eliminate T·G mispairs in DNA (28–30). Methylation directed mismatch repair is used by the cell to correct all types of mismatches, with the possible exception of C·C, by a mechanism involving long patch repair on the unmethylated nascent DNA strand (31). A short patch, methylation independent repair pathway, specific for excision of a T moiety in a T·G mispair, is also present in the cell (4,5). The wild-type level of T glycosylase coded for by the short patch repair *vsr* gene (32,33), leaves intact a significant number of T·G mismatches, as can be deduced by the appearance of m C driven hot spots (1). The short patch repair pathway is believed to be saturable, and, as M. Lieb has recently shown, an increase in T glycosylase levels by introduction of the *vsr* gene on a plasmid, is able to reduce significantly, but **not eliminate**, the m C hot spot (34).

In view of the inability of *E. coli* to reduce m C-induced mutagenesis to 'spontaneous' levels occurring at normal C sites, even in the presence of increased amounts of T glycosylase, it is of interest to consider mechanisms, in conjunction with deamination of m C, that might also contribute to mutagenic hot spots occurring at m C loci. One such mechanism would involve low fidelity DNA replication of template m C. Specifically, enhanced C→T error rates would result if dAMP were incorporated more efficiently opposite m C than C. The simplest experimental test would be to use an *in vitro* model system to measure the relative efficiency of incorporating dAMP opposite template m C and C.

DNA molecules containing either m C or C at a unique template location were copied with three non-proofreading polymerases, AMV RT, Klenow fragment (containing a mutated 3' to 5' exonuclease), and *Drosophila* DNA polymerase α . AMV RT showed a modest (4 to 5-fold) increase in dAMP misinsertion opposite template m C relative to C (Table 1). The efficiency of forming A· m C mispairs compared to G·C correct pairs is, $f_{\text{ins}} = 2.6 \times 10^{-4}$, and for A·C mispairs, $f_{\text{ins}} = 5.8 \times 10^{-5}$ (Eq. 1, Table 1).

Using AMV RT, misinsertion of dAMP opposite m C or C could easily be observed within a 5 min incubation period. Although correct nucleotide insertions by *Drosophila* pol α were readily detected in 30 sec reactions (Figs. 4a, 4b), pol α required roughly 8 min to misincorporate measurable levels of dAMP opposite m C and C at similar concentrations (1 mM) of dATP (Fig. 4a). It has been well-documented that different polymerases may exhibit widely different nucleotide misinsertion specificities at a given template site (14) and conversely, misinsertion rates for a single enzyme may be influenced significantly by surrounding base sequences (14). It is useful to note that the inability of pol α to insert dAMP opposite m C provides strong evidence for the integrity of the m C, because if it had undergone deamination to form T then pol α would have been able to incorporate dAMP readily at that site. Small differences favoring insertion of dAMP opposite m C compared to C were also observed for the exonuclease deficient Klenow fragment, but these differences were even less pronounced than those for AMV RT.

A factor that could influence the persistence of A· m C and A·C mispairs is the relative ease of extending each mispaired terminus by addition of downstream correct nucleotides. The likelihood that a base mispair will undergo error correction by a proofreading exonuclease should diminish as the efficiency of addition of the next correct nucleotide is increased (11). Since the presence of the methyl group provides additional base stacking interactions to stabilize (increase melting temperatures) for correct base pairs in duplex DNA (35), a similar effect might take place at the growing fork to increase the rate of nucleotide addition onto m C compared to C. AMV RT appears to extend both correct and incorrect base pairs involving m C more efficiently than C; there is about a 2-fold difference favoring extension of G· m C pairs over G·C pairs, and about a 3-fold enhancement of A· m C over A·C extension (Table 2).

We also measured insertion and extension opposite template C and m C using the mutagenic nucleotide analogue 2-aminopurine deoxyribonucleoside triphosphate (2-dAPTP) as a substrate for AMV RT and Klenow fragment. The rationale for this measurement is that hot spots for 2-AP mutagenesis were also found to occur at m C sites on DNA (1). We found that

incorporation of 2-AP opposite template C occurred about 20-fold more efficiently than incorporation of A opposite C, in agreement with previous measurements on formation of 2-AP·C mispairs *in vitro* (36,37) and heteroduplex-heterozygotes *in vivo* (38). However, 2-AP was incorporated only about twice as efficiently opposite ^mC than C, and extension from ^mC was also about twice as rapid as extension from C. These small differences for 2-AP misinsertion and extension at template C sites were similar in magnitude to those found for A.

The experiments described in this paper are an attempt to determine whether mutational hot spots occurring at ^mC sites on DNA in procaryotes (21,22) and eucaryotes (23,24) might involve mispairing properties of ^mC with A (or 2-aminopurine) as a component contributing to the excess mutagenesis. We observed a small (less than 5-fold) enhancement in the efficiencies to form and extend A·^mC over A·C mispairs. It is possible that the presence of replication accessory proteins, including proofreading exonucleases, might result in greater differences favoring misincorporation of dAMP opposite ^mC compared to C, and it is also possible that differences in mispairing efficiencies may be more pronounced in other sequence contexts. Therefore, while the limited subset of data that we describe here cannot be used to rule out of that A·^mC mispairs form preferentially and are extended more efficiently than A·C mispairs, these data offer no basis for challenging the prevailing model (1) ascribing deamination as the primary cause of mutagenic hot spots at ^mC residues.

ACKNOWLEDGMENTS

Support for this work is from grants GM21422, GM42554, and CA49758 from the National Institutes of Health.

REFERENCES

1. Coulondre, C., Miller, J. H., Farabaugh, P. J. and Gilbert, W. (1978) *Nature*, **274**, 775–780.
2. Lindahl, T. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 3649–3653.
3. Kornberg, A. and Baker, T. (1992) *DNA Replication*, 2nd Ed., W. H. Freeman Company, New York, p. 501 and p. 783.
4. Lieb, M. (1983) *Mol. Gen. Genet.*, **191**, 118–125.
5. Jones, M., Wagner, R. and Radman, M. (1987) *J. Mol. Biol.*, **193**, 155–159.
6. Shapiro, R. and Klein, R. S. (1966) *Biochemistry*, **5**, 2385–2362.
7. Lindahl, T. and Nyberg, B. (1974) *Biochemistry*, **13**, 3405–3410.
8. Wang, R. Y., Kuo, K. C., Gehrke, C. W., Huang, L. -H. and Ehrlich, M. (1982) *Biochim. Biophys. Acta*, **697**, 371–377.
9. Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C. and Gehrke, C. W. (1986) *Biosci. Rep.*, **6**, 387–393.
10. Frederico, L. A., Kunkel, T. A. and Shaw, B. R. (1990) *Biochemistry*, **29**, 2532–2537.
11. Clayton, L. K., Goodman, M. F., Branscomb, E. W. and Galas, D. J. (1979) *J. Biol. Chem.*, **254**, 1902–1912.
12. Singer, B., Chavez, F., Spengler, S. J., Kusmierek, J. T., Mendelman, L. and Goodman, M. F. (1989) *Biochemistry*, **28**, 1478–1483.
13. Boosalis, M. S., Petruska, J. and Goodman, M. F. (1987) *J. Biol. Chem.*, **262**, 14689–14696.
14. Mendelman, L. V., Boosalis, M. S., Petruska, J. and Goodman, M. F. (1989) *J. Biol. Chem.*, **264**, 14415–14423.
15. Mendelman, L. V., Petruska, J. and Goodman, M. F. (1990) *J. Biol. Chem.*, **265**, 2338–2346.
16. Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. S., Cheoung, C. and Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6252–6256.
17. Creighton, S., Huang, M. -M., Cai, H., Arnheim, N. and Goodman, M. F. (1992) *J. Biol. Chem.*, **267**, 2633–2639.
18. Miller, J. H., Ganem, D., Lu, P. and Schimitz, A. (1977) *J. Mol. Biol.*, **109**, 275–302.
19. Coulondre, C. and Miller, J. H. (1977) *J. Mol. Biol.*, **117**, 577–606.
20. Duncan, B. K. and Miller, J. H. (1980) *Nature*, **287**, 560–561.
21. Ehrlich, M., Zhang, X.-Y. and Inamdar, N. M. (1990) *Mut. Res.*, **238**, 277–286.
22. Yatagai, F. and Glickman, B. W. (1990) *Mut. Res.*, **243**, 21–28.
23. Rideout, W. M. (III), Coetzee, G. A., Olumi, A. F. and Jones, P. A. (1990) *Science*, **249**, 1288–1290.
24. Jones, P. A., Rideout, W. M. (III), Shen, J. -C., Spruck, C. H. and Tsai, Y. C. (1992) *BioEssays*, **14**, 33–36.
25. Lindahl, T., Ljungquist, S., Siebert, W., Nyberg, B. and Sperens, B. (1977) *J. Biol. Chem.*, **252**, 3286–3294.
26. Lindahl, T. (1979) *Progr. Nucleic Acids Res., Mol. Biol.*, **22**, 135–192.
27. Shenoy, S., Ehrlich, K. C. and Ehrlich, M. (1987) *J. Mol. Biol.*, **197**, 617–626.
28. Brown, T. C. and Jiricny, J. (1987) *Cell*, **50**, 945–950.
29. Brown, T. C. and Jiricny, J. (1988) *Cell*, **54**, 705–711.
30. Wiebauer, K. and Jiricny, J. (1989) *Nature*, **339**, 234–236.
31. Modrich, P. (1989) *J. Biol. Chem.*, **264**, 6597–6600.
32. Sohail, A., Lieb, M., Dar, M. and Bhagwat, A. S. (1990) *J. Bacteriol.*, **172**, 4214–4221.
33. Hennecke, F., Kolmar, H., Brundl, K. and Fritz, H. -J. (1991) *Nature*, **353**, 776–778.
34. Lieb, M. (1991) *Genetics*, **128**, 23–27.
35. Murchie, A. I. H. and Lilley, D. M. J. (1989) *J. Mol. Biol.*, **205**, 593–602.
36. Watanabe, S. M. and Goodman, M. F. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2864–2868.
37. Watanabe, S. M. and Goodman, M. F. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6429–6433.
38. Hopkins, R. and Goodman, M. F. (1979) *J. Mol. Biol.*, **135**, 1–22.