Cytosine Deamination in Mismatched Base Pairs†

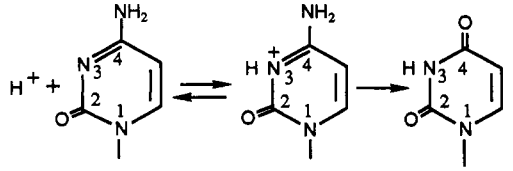
Lisa Ann Frederico,‡,§ Thomas A. Kunkel,‖ and Barbara Ramsay Shaw*,‡

Department of Chemistry, P. M. Gross Chemical Laboratory, Duke University, Durham, North Carolina 27708-0346, and Laboratory of Molecular Genetics, National Institute of Environmental Health Science, Research Triangle Park, North Carolina 27709

Received December 23, 1992; Revised Manuscript Received March 29, 1993

ABSTRACT: The rate of deamination of cytosine in mismatched base pairs has been determined. Incubation of M13mp2 nicked heteroduplex DNA molecules containing T·C or C·C mispairs in the *lacZ* α-complementation gene results in deamination of cytosine to uracil, producing T·U or C·U mispairs. Strands which have undergone deamination at the target site to produce uracil will yield dark blue plaque revertants, while all other strands yield faint blue or colorless plaque phenotypes upon transfection of an *ung*⁻ α-complementation *Escherichia coli* host strain. Rate constants were calculated from the reversion frequencies for several different heteroduplexes incubated at either 60 or 37 °C. For the 60 °C incubations, the hydrolytic deamination rate constants for mispairs in three different local sequence environments ranged from 8×10^{-10} to 40×10^{-10} s^{-1}. For incubations at 37 °C, the rate constants were between 0.4×10^{-10} and 1.3×10^{-10} sec^{-1}. At both temperatures and for all mispairs, these rate constants are significantly greater than deamination rate constants in properly matched Watson–Crick G·C base pairs and are similar to those constants determined for cytosine deamination in single-stranded DNA. Since deamination most likely occurs via a single-stranded intermediate, the data suggest that, at 37 °C, the T·C and C·C mispairs exhibit from 20% to 100% single-stranded character. We conclude that cytosine residues involved in a mispair in DNA are 1–2 orders of magnitude more prone to deaminate to uracil than are cytosines in double-stranded DNA. These studies underscore the importance of Watson–Crick base-pairing interactions in maintaining the genetic integrity of DNA.

Deamination of cytosine (**1**) to uracil (**2**) occurs over 100 times faster in single-stranded (ss)[1] than in double-stranded (ds) DNA (Lindahl & Nyberg, 1974; Frederico *et al.*, 1990).



CYTOSINE 1 URACIL 2

The ds DNA is protected by its secondary structure from deamination. Duplex formation reduces the basicity of cytosine, making it less available for protonation (Sowers *et al.*, 1987). For cytosine monomers in neutral and acidic solutions, deamination proceeds via an intermediate protonated at the N-3 position of cytosine (Shapiro & Klein, 1966; Garrett & Tsau, 1972). In double-stranded DNA, the N-3 of cytosine in a Watson–Crick base pair is involved in a hydrogen bond with the N-1 proton of guanine and thus is less available for protonation. By contrast, in single-stranded DNA, the N-3 of cytosine would be available for protonation. Once protonated, the cytosine could deaminate. Hence, a Watson–Crick-type pairing should make protonation of cytosine more difficult and thereby limit its deamination. In addition to protonation, steric factors might also play a major role in the kinetics of deamination. The rate-determining step in the addition–elimination pathway of deamination involves attack by water or hydroxyl at the C-4 of the cytosine (Shapiro, 1981) to give a tetrahedral intermediate, which should be less likely to form in the more rigid environment of the duplex. The hypothesis that Watson–Crick pairing limits deamination can be tested by examining the reactivity of cytosines residing in different types of pairing environments.

Here, we compare the rate of cytosine deamination in matched and C·C and C·T mismatched base pairs and relate the finding to rates of cytosine deamination in single-stranded DNA. DNA molecules that are identical except for a single base pair can be prepared with recombinant DNA technology. By placing various base mismatches at a site, the effect of base-pairing interactions on deamination can be probed. It should be possible to determine whether or not Watson–Crick base-pairing interactions are predominantly responsible for the 100-fold slower rate observed for cytosine deamination at 37 °C in double-stranded DNA compared to that in single-stranded DNA (Frederico *et al.*, 1990).

We have recently described a sensitive genetic assay that allows determination of the rate of cytosine deamination at a single cytosine base in a 7196-bp M13mp2 phage DNA molecule (Frederico *et al.*, 1990). With this reversion assay, deamination rate constants of cytosine in single- and double-stranded DNA can be measured under conditions where it had previously been impossible to assess rates of deamination, *i.e.*, at physiologically relevant temperatures and pH. Half-lives as long as 30 000 years can be determined to within an accuracy of 2–3-fold. Additionally, *site-specific* rate constants for deamination at a single cytosine residue can be obtained with this genetic assay, in contrast to average rate constants obtained from chemical methods. In this paper, we describe

† This work was supported by a Faculty Research Award to B.R.S. from the American Cancer Society (FRA 301) and by NIH (CA 44709).

* Corresponding author.

‡ Duke University.

§ Current address: Department of Chemistry, The Citadel, Charleston, SC 29409.

‖ NIEHS.

[1] Abbreviations: DTT, dithiothreitol; ndDNA, nicked double-stranded M13mp2 hybrid (heteroduplex); RF, replicative form; ds, double-stranded; ss, single-stranded; TE, 10 mM Tris and 0.1 mM EDTA, pH 7.4; A, adenine; C, cytosine; G, guanine; T, thymine; U, uracil.

0006-2960/93/0432-6523$04.00/0

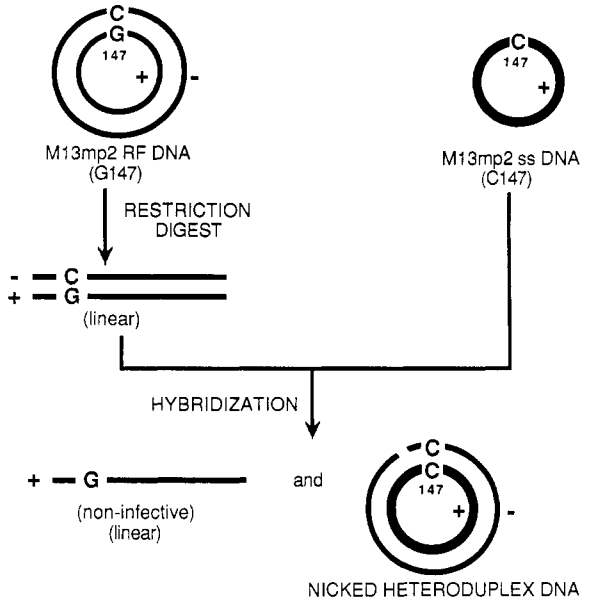


FIGURE 1: Mismatched heteroduplex preparation. RF DNA is digested to completion by a single site endonuclease (*Ava*II) as determined by agarose gel analysis. Hybridization conditions are optimized to minimize excess circular single-stranded DNA. The hybridization is carried out by heating RF linear DNA in low ionic strength TE buffer at 70 °C for 5 min. Circular single-stranded DNA is added and the sample is placed on ice. SSC (20×) is added to a final concentration of 2×, after which the sample is incubated at 60 °C. The nicked heteroduplex illustrated here is constructed from G147 mutant duplex DNA for the (−) strand and single-stranded C147 mutant for the (+) strand circular DNA.

the application of the M13mp2 reversion assay to investigate the deamination rate of a specific cytosine when it is present in either a C·C or a T·C mispair. We report that C residues in mismatched base pairs are deaminated at 10–100 times the rate of C in the normal C·G base pair, rates which approach the values for single-stranded DNA, implying a large degree of single-stranded character to C·C and C·T mismatches.

EXPERIMENTAL PROCEDURES

Escherichia coli cell strains of the *ung*⁻ and *ung*⁺ phenotype were the same as those described in Frederico *et al.* (1990). The M13 mutants (C–11, G–11, G87, T87, C147, T147, and G147) used to prepare the heteroduplex molecules are described by the convention defined in Frederico *et al.*, 1990.[2] Uracil glycosylase was a gift from Dr. Bruce Duncan.

Preparation of Nicked Heteroduplex (ndDNA) Molecules.[3] *Optimizing Hybridization While Minimizing Deamination and Excess Circular ss DNA.* Single- and double-stranded DNA from M13mp2 mutants was isolated as described in Frederico *et al.* (1990). Construction of the nicked duplex DNA (ndDNA) is outlined in Figure 1 and generally followed Kunkel and Soni (1988). *Ava*II (International Biotechnologies, Inc.) was used to generate the linear duplex, producing a single cut at site −264 of the *lacZ* gene. The appropriate M13mp2 ds DNA was diluted to 1 μg/μL in medium-salt restriction buffer [10 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM dithiothreitol (DTT)] and digested with the restriction endonuclease *Ava*II (International Biotechnologies, Inc.) at a concentration of 200 ng/μL DNA and 1 unit of enzyme/μg of DNA in a total volume of 500 μL. The digested sample was ethanol-precipitated at −20 °C with 2 volumes of ice-cold ethanol and 0.1 volume 3 M sodium acetate, pelleted, and resuspended in 10 mM Tris and 0.1 mM EDTA, pH 7.4, to a concentration of 1 μg/μL.

Common hybridization procedures used to prepare a heteroduplex molecule often involve incubations at high temperatures (95–100 °C) for 10–30 min, followed by annealing at 65 °C for several hours or overnight [*e.g.*, Kramer *et al.* (1984)]. It can be calculated from Frederico *et al.* (1990) that use of such a procedure involving a 10-min incubation of a single-stranded DNA at 100 °C would result in the deamination of 1 out of 5000 cytosines (0.02%) at a single site, which would yield an unacceptable background for the M13mp2 reversion assay. Therefore, the initial step in our protocol involved determining the optimum hybridization conditions that would result in the least damage to the DNA. Hybridization is affected by a number of conditions including DNA concentration, salt concentration, and most important, temperature. On the basis of these considerations, the following protocol was devised for hybridizing the linear double-stranded DNA to the circular single-stranded DNA.

The full-length, linear, double-stranded molecule (1 μg/μL in 0.1× TE buffer) at a concentration of 100 ng/μL was heated at 70 °C for 5 min in an Eppendorf tube to denature the DNA. [The low salt concentration allows the DNA to denature at a lower temperature (Marmur & Doty, 1962), which in turn reduces the rate of heat-induced deamination.] Immediately upon removal of the sample from the 70 °C water bath, circular single-stranded DNA (0.25 μg/μL in TE buffer) was added at a 5:1 molar ratio of double- to single-stranded DNA, and the mixture was placed on ice. (Addition of the single-stranded DNA at this step minimizes the length of time the single-stranded DNA mutagenesis target is subjected to high temperature and thus minimizes the heat damage to the molecule.) Soon thereafter, 1/9 volume of 20× SSC (3.0 M NaCl and 0.3 M sodium citrate) was added to give a final concentration of 2× SSC. The sample was renatured at 60 °C for 5 min and then placed on ice. Following ethanol precipitation and resuspension in sample buffer B (0.1 M KCl, 0.05 M Hepes-KOH, 0.01 M $MgCl_2$, and 0.001 M EDTA, pH 7.4), 300 ng of the sample was analyzed by electrophoresis on 0.8% agarose gels with ethidium bromide to determine the yield of ndDNA and ensure that the single-stranded circular DNA had annealed to form a duplex. Agarose gel analysis indicated that the yields of the ndDNA obtained using this procedure were similar to yields obtained with procedures involving higher temperatures.

The appropriate molar ratio that resulted in no detectable circular single-stranded DNA was determined by several test hybridizations. We monitored the disappearance of ss DNA on a 0.8% agarose gel in the presence of ethidium bromide. Typically, hybridizations with a 5:1 molar ratio of double-stranded linears to circular single-stranded DNA resulted in no detectable ss DNA visible on the gel. Once prepared, the hybrid does not denature due to the high salt concentration in sample buffer B and handling of the sample at temperatures of 60 °C or less.

[2] The mutant base letter designation on the (+) strand is followed by the number of the site of the mutation in the *lacZ* gene, with +1 being the start of transcription. The mutant has only a single base change unless otherwise mentioned; *i.e.*, all other bases are the same as the wild-type sequence.

[3] Nicked double-stranded (ndDNA) hybrids constructed are referred to by the mispair rather than by the mutants used for their preparation. The designation for the mispairs is from left to right: the base in the mispair in the (−) strand followed by a center dot, the base in the mispair in the (+) strand, and the site of the mismatch. For example a thymine–cytosine mispair with thymine in the (−) strand and cytosine in the (+) strand at site 147 would be referred to as T·C147. This molecule is prepared from hybridization of closed circular single-stranded C147 mutant DNA and linear double-stranded A147 mutant DNA.

Heat Treatment. Aliquots (100 μL) of the ndDNA (in sample buffer B at a concentration of 25 ng/μL) were sealed in 200-μL capillary tubes and incubated at either 37 or 60 °C (±1 °C) in a water bath. After incubation, the samples were removed from the capillary tubes and stored at −70 °C.

Transfection with Competent Hanahan Cells. After the appropriate incubation time, all the samples were analyzed by transfection of competent cells prepared by the method of Hanahan (1983). By this method, the transfection efficiency for double-stranded DNA is ~100-fold greater than for circular single-stranded DNA [Table II of Kunkel and Soni (1988)]. Thus, there is a negligible contribution of any circular single-stranded DNA that remains after formation of the ndDNA to the measured reversion frequency.

Scoring Revertants, Calculating Reversion Frequency, and Determining DNA Sequences. Plating and scoring mutants in *ung*⁻ and *ung*⁺ cells followed methods described in Frederico *et al.* (1990). Revertant plaques, which appear as dark blue plaques against a colorless or faint blue background, were picked from the plate, diluted in 50 mM sodium borate buffer, and replated to confirm revertant phenotypes. Sequencing of DNA in randomly selected revertants was carried out with the chain terminator method (Sanger *et al.*, 1977). The number of revertant plaques was divided by the total plaques to calculate the reversion frequency (F), which describes the population of deaminated molecules.

RESULTS

Design of the Experiment, Construction of Heteroduplexes, and Method of Transfection. The rate of cytosine deamination in mismatched base pairs is determined by incubating M13mp2 heteroduplex DNA molecules containing T·C or C·C mispairs in the *lacZ* α-complementation gene.[4] Deamination produces T·U or C·U mispairs which can be scored and quantified by the method of Frederico *et al.* (1990). The uracil-containing strand yields blue plaque revertants, while the complementary strand yields faint blue or colorless plaque phenotypes upon transfection of an *ung*⁻ α-complementation *E. coli* host strain. Rate constants are calculated from the reversion frequencies for several different heteroduplexes incubated at either 60 or 37 °C for varying lengths of time.

Nicked heteroduplexes (ndDNA hybrids) with a single base mismatch involving cytosine were constructed by hybridizing DNA from two different M13mp2 bacteriophage mutants as shown in Figure 1. The (−) strand was derived from a linear double-stranded mutant and the (+) target DNA strand was derived from another closed circular single-stranded DNA mutant that differed in sequence at only one base. Because cytosine in single-stranded DNA deaminates about 100 times faster than in double-stranded DNA (Frederico *et al.*, 1990) and a small amount of single-stranded DNA could in principle contribute measurably to the reversion frequency, four precautions were taken during the sample preparation, incubation, and transfection of the mismatched hybrids to minimize background deamination in the target strand and to limit expression from the single-stranded DNA: (1) Linear duplex DNA was used in excess over ss DNA for preparation of the hybrids. Gel electrophoresis of test hybridizations showed that, under our experimental conditions with a ratio of ≥5:1 (linear duplex DNA to circular single-stranded DNA), no circular single-stranded DNA was visible in an agarose gel containing ethidium bromide (data not shown). (2) To minimize background deamination in the preparation of the ndDNA, temperatures were maintained as low as possible for the briefest amount of time. Denaturation of the ds linear DNA in 0.1× TE buffer was carried out at 70 °C for only 5 min. The circular ss M13 (+)-strand target was never heated above 65 °C and was annealed at 60 °C for only 5 min. (3) Buffers and temperatures were chosen such that, once hybrids were formed, they would remain double-stranded. Duplex denaturation would not occur if denaturation temperatures were lower than 65 °C; therefore, the hybrids, once prepared, should not denature or switch partners at the 37 and 60 °C incubation temperatures in our experiments. The ndDNA hybrids were further stabilized after preparation because the incubations were carried out in buffer B, a higher salt than in the buffer used for the denaturations; the higher salt would further increase the T_m of the hybrids and make it unlikely that they would denature or switch partners. (4) Cells made competent by the procedure described by Hanahan (1983) were used for all transfections to ensure that deamination of any remaining single-stranded (unhybridized) DNA did not contribute to the reversion frequency. The efficiency of transfection of circular single-stranded DNA is ~100-fold lower than that for double-stranded DNA [see legend to Table II in Kunkel and Soni (1988)]. Also, linear DNA remaining in the hybrid preparation is not infective in the Hanahan transfection (data not shown) and therefore the residual (+)-strand linear DNA does not have to be removed prior to transfection.

Table I: Phenotype of DNA Molecules Used in This Study

site	base in (+) strand	plaque phenotype[a]
site 147	A	1 (faint blue)
	T	4 (dark blue)
	G	0 (colorless)
	C	1 (faint blue)
site 87	A	1 (faint blue)
	T	4 (dark blue)
	G	unknown
	C	1 (faint blue)
site −11	A	4 (dark blue)
	T	0 (colorless)
	G	0 (colorless)
	C	0 (colorless)

[a] Color convention from Kunkel and Soni (1988).

Choice of Target: Distinguishing Deamination by Phenotype. Three sites (Table I) were located where deamination in a mismatched heteroduplex could be detected phenotypically as a dark blue plaque among colorless or faint blue plaques. For example, at site 147 we could prepare either a C·T or C·C heteroduplex in which the component (−) and (+) strands would produce only 0- or 1-plaque phenotypes. (See Figure 2.) Only a change from C to U in the (+) strand yielded a wild-type blue phenotype (in boxes, Figure 2). Thus, a deamination event at cytosine 147 in the (+) strand, producing a 4-plaque phenotype, would be easily discernable from the 0- and 1-phenotypes produced. Further, and very important for this study, all other resulting base changes also produced a faint blue or colorless phenotype. For example, repair of the T·C mispair (Figure 2, bottom) as directed by the (−) strand would produce an A in the (+) strand. This correction event would produce a 1-phenotype and thus would not be observable in our assay. Likewise, repair from the (+) strand would result in G in the (−) strand, also yielding a 1-plaque phenotype. In this way, as shown in Figure 2, cytosine deamination in a T·C as well as a C·C base mispair at site 147 can be distinguished from all other processes, including mismatch repair. A similar argument, based on plaque

[4] See Kunkel (1984) for a description of the gene and its sequence.

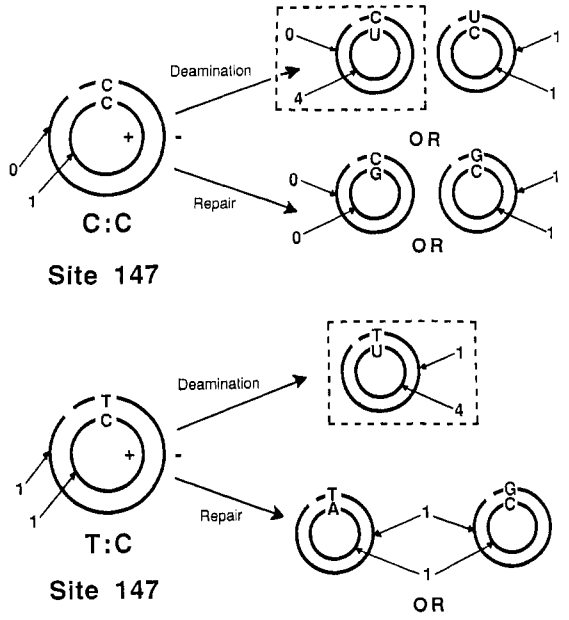


FIGURE 2: C·C and T·C mispairs at site 147 in the *lacZα* structural gene. These mispairs allow the phenotypic distinction of mismatch repair from deamination by the resulting phenotypes. The boxes represent the blue-plaque phenotypes that permit detection of deamination events. Numbers refer to the phenotypes in Table I. A logical experiment would be to examine the effect of a purine–pyrimidine mismatch on deamination. However, note that this assay does not allow the rate of deamination of an A·C mismatch to be examined, because deamination cannot be phenotypically distinguished from mismatch repair. Deamination of cytosine produces an A·U base pair (which codes like A·T), and mismatch repair of an A·C mispair produces an A·T or G·C base pair.

phenotype described in Table I, can be made for the appropriate mispairs at positions 87 or −11 (not shown). Thus, by judicious choice of target, it was possible to prepare heteroduplexes with mismatches such that (−)-strand expression or rectification events in C·C or T·C base mispairs could be discriminated from deamination events.

Mismatch-Induced Deamination in a C·C Mispair at Site 147. In order to compare deamination at a C·C mismatch with that at a G·C perfectly matched duplex, two ndDNA molecules were used: (1) the C·C147 mismatch ndDNA, prepared as shown in Figure 1 from C147 (+) circular single-stranded DNA and G147 linear double-stranded DNA and (2) similarly, the G·C147 perfectly matched ndDNA, prepared from C147 (+) circular single-stranded DNA and C147 linear double-stranded DNA.

Transfection of the C·C147 ndDNA produced two phenotypes: a 0-plaque phenotype from the (−) strand of the G147 linear duplex DNA and a 1-plaque phenotype from the (+) strand composed of C147 single-stranded DNA (Figure 2, top). Because the phenotypes produced by each strand were distinguishable, the contribution to expression of the phenotype from each strand of the C·C147 nDNA could be determined. The percent expression of the (+) C147 strand (obtained by dividing the number of faint blue 1-plaque phenotypes by the number of total 0- and 1-plaque phenotypes and multiplying by 100) upon transfection into *ung*⁻ cells was found to be 21% ± 4% in 20 transfections of several different ndDNA hybridizations. This number is consistent with results using other M13 ndDNA molecules (Kramer *et al.*, 1984; T. A. Kunkel, unpublished data). A (−)-strand bias exists due to the asymmetry of the rolling circle replication of M13 (Kornberg & Baker, 1991). This analysis was necessary for accurately quantifying the reversion frequency, which here represented the number of dark blue plaques arising from deamination of cytosine in the target strand divided by the total number of plaques originating from that strand.

Table II: Mismatch-Induced Deamination in a C·C Mispair at 60 °C

DNA sample[a]	total[b] plaques (10^4)	revertants[c]	reversion frequency (10^{-5})
C·G (G147, ds)	25	3	1
G·C (C147, ds)	31	0	≤0.3
C147, ss	12	2	0.3
	ung⁻ Cells		
C·C, 0 days	8.1	5	6
C·C, 4.7 days	6.5	64	98
G·C, 0 days	18.6	0	≤0.5
G·C, 4.7 days	11.9	1	0.8
	ung⁺ Cells		
C·C, 0 days	4.2	2	5[e]
C·C, 4.7 days	7.2	2	3
	ung⁻ Cells, U-Glycosylase[d]		
C·C, 4.7 days	3.3	1	3

[a] A nicked heteroduplex with a C·C147 mispair (prepared from G147 linear double-stranded DNA and C147 circular single-stranded DNA) was incubated in buffer B at 60 °C and transfected into either *ung*⁻ or *ung*⁺ Hanahan cells. As a control, a nicked duplex DNA with a G·C base pair (prepared from C147 linear double-stranded DNA and C147 single-stranded DNA) was treated in a similar manner. [b] The total number of plaques indicates the number of plaques arising from the (+) strand, determined either by counting the number of 1-plaque phenotypes (in the C·C147 ndDNA experiments) or by multiplying the total 0- and 1-plaques by 0.21 (in the G·C147 ndDNA control experiments) to account for strand expression as described in the text. [c] The revertants have a 4-plaque (dark blue) phenotype. [d] U-glycosylase indicates that uracil glycoslyase treatment of the DNA molecule [according to Frederico *et al.* (1990)] was performed after the incubation but before transfection. [e] The higher reversion frequency at $t = 0$ for the C·C mispair as compared to the G·C duplex in the unincubated control was due to the higher background reversion frequency in the G147 double-stranded (RF) preparation than in the C147 (RF) preparation.

The C·C147 ndDNA molecule was next incubated at 60 °C to determine the effect of a mismatch on cytosine deamination. As a control, the analogous ndDNA (G·C147) with a Watson–Crick G·C base pair at site 147 was used. For the C·C mispair experiment, counting faint blue plaques enabled us to accurately determine the total plaques originating from the (+) strand. The results are shown in Table II. Upon heating the C·C mismatched heteroduplex at 60 °C for 4.7 days, there was a 16-fold increase in the reversion frequency in the *ung*⁻ cell strain over that of the unincubated control.

Several control experiments were performed. First, a G·C ndDNA at site 147, which had been constructed identically to that of the mismatched C·C ndDNA except for the base pair at site 147, failed to show an increase in reversion frequency after incubation at 60 °C. At this site, the G·C plaques reported in Table II are taken to be 20% of the total plaques observed since it is not possible to discriminate phenotypes of the two strands. Second, to confirm that the revertant plaques arose from deamination mutagenesis, a transfection was also carried out in *ung*⁺ Hanahan cells with the heat-treated mismatched ndDNA. The reversion frequency decreased to unincubated control levels (Table II), indicative of mutagenic events proceeding through a uracil intermediate. Third, the mismatched heat-treated heteroduplex was treated *in vitro* with purified uracil glycosylase prior to *ung*⁻ Hanahan cell transfection. Again, the reversion frequency of the heat-treated mismatched ndDNA returned to the reversion frequency of the unincubated control samples. Fourth, single-stranded C147 DNA was subjected to the same hybridization and incubation conditions as the hybrid duplexes. Results of transfection of this DNA yielded a total of less

than 200 plaques. Thus, if all the single-stranded DNA used to prepare the C147·C147 mismatched hybrid had remained single-stranded, it would have produced only 200 plaques, a negligible amount compared to the 6.5×10^4 total plaques that were observed for the mismatch. Further, during all our experiments at different sites, 700 plaques were observed collectively from all the heat-treated ss DNA samples, yet no revertants were observed. Additionally, the C·T mismatch at site −11 does not contain a cytosine in the single-strand DNA, yet it still is found to have a 60 °C rate constant that is similar to those for the other mismatches (see below). These experiments thus ensured that any residual, unhybridized single-stranded DNA was not contributing revertants to the reversion frequency and that the results indeed reflected deamination within the mismatch.

The results in Table II indicate that deamination mutagenesis can be observed in a mispair and that the rate of deamination is accelerated compared to the G·C base pair.

Comparison of Deamination in C·C and T·C Mismatches at Site 147 and 60 °C. The extent of deamination of cytosine in a C·C mismatch at site 147 was compared with its deamination in a T·C mismatch at the same site. The T·C mismatch at site 147 does not permit discrimination of the phenotype between strands, since both strands produce the same faint blue phenotype (Figure 2, bottom). Therefore, instead of directly counting the number of plaques produced by the (+) C147 strand, the total number of plaques was determined and 20% of this amount was designated to be from the (+) strand, based on the relative expression of strands in the C·C mispair at 147. The data for incubations at 60 °C are shown in Table III. After 4.7 days, the increase in the reversion frequency for the T·C mismatch at site 147 was 17-fold above that of the unincubated controls. The reversion frequency for the T·C mismatch at site 147 was only 35% of the value determined for the C·C mismatch at the same site and conditions.

Mismatch-Induced Deamination at Sites 87 and −11. In addition to site 147, two other sites in the *lacZ* target were available to test deamination mutagenesis at a cytosine mispair. (See Table I.) Site 87 was used to examine the T·C mispair and site −11 was used to study both the C·C and C·T mispairs. The construction of ndDNAs at sites −11 and 87 is analogous to that in Figure 1. In each case only a change from a C to U, as expected from deamination mutagenesis, can be observed as a 4-plaque phenotype. At site −11 the mismatch was prepared in the opposite orientation; *i.e.*, the cytosine which was monitored for deamination was in the (−) strand instead of the (+) strand of the duplex. This change had the advantage that 80% of the expression should occur with the target cytosine (−) strand, instead of the 20% expression for the other ndDNA molecules used here.

At site 87 a significant increase in reversion frequency was observed for a 10-day incubation of the T·C mispair at 60 °C (Table III). This increase was 40-fold greater than the untreated mismatched heteroduplex control and 13-fold above the properly matched heat-treated control. Similar results were found at the third site, −11 in the *lacZ* gene control region. Again, the reversion frequencies of the heat-treated C·C and C·T mismatches were 13- and 15-fold, respectively, above the levels of the untreated mismatches. The C·C mismatch at site −11 showed a greater than 10-fold higher reversion frequency than the normal C·G base pair, while the C·T mismatch showed a 4-fold increase in the reversion frequency over the G·C base pair.

Table III: Rate Constants (60 °C) for Three Types of Matched and Mismatched Base Pairs at Three Sites in M13mp2 Heteroduplexes

base pair[a]	incubation time (days)	total plaques[b] (10^4)	revertants[c]	reversion frequency (10^{-5})	60 °C rate constant[d] (10^{-11} s^{-1})
		Site 147 (GC C CC)[e]			
G·C	0	19	0	≤0.5	
	4.7	12	1	0.8	(≤1)
T·C	0	19	3	2	
	4.7	19	65	34	80
C·C	0	8.1	5	6	
	4.8	6.5	64	98	200
		Site 87 (AC C CC)			
G·C	0	12	7	6	
	10	1.9	3	16	(≤10)
T·C	0	3.8	2	5	
	10	4.2	82	200	200
		Site −11 (CA C AC)			
C·G	0	53	47	8.9	
	5	28	41	15	(≤10)
C·T	0	58	25	4.3	
	5	33	216	66	100
C·C	0	12	35	29	
	10	8.3	324	390	400
SS DNA[f]					350[f]

[a] The ndDNA heteroduplexes were prepared as shown in Figure 1 with the appropriate mutants and then incubated at 60 °C in buffer B for the indicated time. [b] The total number of plaques indicates the number of plaques from the (+) strand, which for $C{\cdot}C_{147}$ were counted directly but for $T{\cdot}C_{147}$ and $G{\cdot}C_{147}$ and at site 87 were taken to be 21% of the total plaques. At site −11 the target C is in the (−) strand, which is taken to have 80% expression. [c] The reversion frequencies are for *ung*⁻ Hanahan cell transfections of the ndDNA samples. All revertants has a 4-plaque phenotype. [d] The deamination rate constants for the ndDNA molecules were calculated by subtracting the reversion frequency determined before incubation from the reversion frequency obtained after incubation to obtain F', a corrected reversion frequency, and then using the equation $k = [-\ln (1 - F')]/t$ from Frederico *et al.* (1990). Rate constants are presented to 1 significant figure. (Numbers in parentheses are only approximate rate constants, derived from small differences in reversion frequencies measured at only one time interval. Caution should be exercised in assigning significance to small differences.) [e] The cytosine target at which deamination can be scored is underlined; the two bases on either side are also shown. [f] For comparison, the value for the single-stranded deamination rate constant at 60 °C was taken from Table III of Frederico *et al.* (1990) calculated by using the 28 kcal/mol activation energy for deamination of C in single-stranded DNA.

Sequencing Confirms Deamination. Sequencing 8–19 revertants at each mismatch site showed that all revertants had the C to T change as would be expected in deamination mutagenesis. No other mutations were found.

Calculations of Rate Constants for Deamination. The rate constants k were calculated from the reversion frequencies given in Table III by subtracting the reversion frequency of the unincubated sample from the reversion frequency of the heat-treated sample to obtain a corrected frequency F' and using

$$k = \frac{-\ln (1 - F')}{t} \quad (1)$$

[according to Frederico *et al.*, (1990)]. As shown in Table III, mismatched cytosines deaminate 10–200-fold faster than perfectly matched cytosines, and cytosines in C·C mismatches appeared to deaminate 3-fold faster than those in T·C mismatches at 60 °C. For the matched hybrids, the small number of revertants and the 2–3-fold differential in reversion frequencies between the incubated and zero-time samples do not allow us to derive accurate rate constants. By contrast, the reversion frequencies for the mispaired hybrids are all 1–2 orders of magnitude above those of the matched hybrids (Table

Table IV: Rate Constants (37 °C) at Site 147 for Three Types of Mismatched Base Pairs

base pair	incubation time (days)	total plaques (10^4)	revertants	reversion frequency (10^{-5})	37 °C rate constant[a] (10^{-12} s^{-1})
G·C	0	19	0	≤0.5	
	158	16	11	6.8	<5
C·C	0	8.1	5	6	
	67	19	53	27	40
	158	4.2	43	102	70
T·C	0	19	3	2	
	52	12	69	59	130
ss DNA, 37 °C					100[b]

[a] The G:C_{147}, C:C_{147} and T:C_{147} nicked heteroduplexes at site 147 were prepared with the appropriate DNAs and then incubated at 37 °C in buffer B. The reversion frequencies were determined in *ung*⁻ Hanahan cells transfected with each heat-treated ndDNA. The total number of plaques indicates the number of plaques from the (+) strand. The plaques from the (+) strand of the G·C and T·C were taken to be 21% of the total plaques as described in the text. The revertants have a 4-plaque phenotype. [b] This value is the 37 °C deamination rate constant for cytosine in single-stranded DNA from Frederico *et al.* (1990).

III), and accurate rate constants can be calculated. At 60 °C, the C·C and T·C mismatches in the nicked duplex molecules deaminate with rate constants from 80×10^{-11} to 400×10^{-11} s^{-1}. By comparison, the approximate deamination rate constants at 60 °C for the matched G·C ndDNA molecules ranged from 1×10^{-11} to 10×10^{-11} s^{-1}, which lie in the range expected for double-stranded DNA (see Discussion).

Mismatch-Induced Deamination at 37 °C. To examine deamination under more physiologically relevant conditions, as well as to confirm the 60 °C results, heat incubations of the same G·C, C·C, and T·C nicked heteroduplexes at site 147 were also performed at 37 °C (Table IV). Since deamination is slowed by lower temperatures, the ndDNA molecules had to be incubated at 37 °C for several months in order to achieve reasonable increases in reversion frequencies. As shown in Table IV, a significant increase in the reversion frequency of the C·C mispair at site 147 was seen at 67 days as well as at 158 days. The reversion frequency of the T·C mispair also increased with treatment at 37 °C for 52 days. In contrast, the G·C molecule showed only a small increase in the reversion frequency after 158 days at 37 °C; this value was close to that of the unincubated C·C and T·C molecules. The 37 °C rate constants of the mispairs at site 147 were found to be 8–26 times greater than that of the normal G·C duplex and in the range of single-stranded DNA. Thus, these 37 °C results were in agreement with the results of experiments conducted at 60 °C, confirming that cytosine in a mispair deaminates considerably faster than in a Watson-Crick G·C base pair.

DISCUSSION

Cytosine Deamination Is Accelerated in a Mismatched Base Pair. In this study we have investigated the effect of mispairing on cytosine deamination in order to test the hypothesis that deamination of cytosine in a mismatched base pair would be enhanced considerably compared to that of a cytosine in a Watson–Crick base pair. We have examined the tendency of a cytosine to deaminate when it is present in a C·C or C·T mispair and compared the rate of deamination with that in a perfectly matched C·G base pair as well as with a cytosine in single-stranded DNA. At 37 °C, the deamination rate constants obtained for a mispaired cytosine (Table IV) vary from 0.4×10^{-10} to 1.3×10^{-10} s^{-1} at 37 °C, depending on the type of mismatch (at site 147), and are significantly higher than for properly-matched double-stranded DNA (0.05×10^{-10} s^{-1} at 37 °C for the ndDNA). Deamination in mispairs is enhanced 8–26-fold and the rates approach, or are about equal to, that of single-stranded DNA (1×10^{-10} s^{-1}) at 37 °C.

Deamination at neutral pH is thought to proceed via an intermediate in which the N-3 of cytosine is protonated (Shapiro, 1981). In an unpaired state, as in single-stranded DNA, the cytosine would be readily accessible for protonation, whereas in the properly paired state it would be inaccessible except during the times that the base pair would breathe. Since deamination most likely occurs via a single-stranded intermediate, it would appear that the C in the C·C and C·T mispairs has from 20% to 100% single-strand character. Alternatively, 20%–100% of the time (depending on the mismatch) the mismatched cytosine may find itself in an unpaired state and thus available for protonation and subsequent deamination. This study would suggest that double-stranded DNA deaminates more slowly than single-stranded DNA at least in part because of hydrogen-bonding interactions between the nucleic acid bases.

Effect of Temperature on Deamination. At a higher temperature (60 °C), the trends are the same as at 37 °C, although the rate constants are greater, as expected. Deamination is accelerated by 10–200 fold in C·C and C·T mispairs relative to that in C·G pairs (Table III). Again, the mismatched rate constants [$(8–40) \times 10^{-10}$ s^{-1}] all approach the 60 °C value (35×10^{-10} s^{-1}) obtained from our previous measurements (Frederico *et al.*, 1990) of cytosine deamination in single-stranded DNA.

Reliability of the Assay. Deamination rate constants have been determined on the basis of the assumption that, in our system, mismatch repair of deaminated base pairs and deamination in residual single-stranded DNA are inconsequential events and therefore contribute little to the reversion frequency. We feel these are reasonable assumptions since our choice of mutational targets allowed us to phenotypically distinguish mismatch repair events from deamination events. Deamination resulted in dark blue revertants; mismatch repair gave colorless or faint blue revertants. Both sequencing and glycosylase treatment confirmed the deamination mutagenesis. Conditions were selected to give hybrid formation with minimal deamination and to minimize the contribution of single-stranded DNA to the reversion frequency. Salt concentration and temperature were chosen to keep the hybrid in the duplex form. Further, any circular single-stranded DNA target was very inefficiently transfected in Hanahan cells relative to ndDNA hybrids, and linear DNA was biologically inactive. Thus, revertants scored could arise only from deamination of mispaired cytosine in the ndDNA hybrids. Finally, the availability of two or three types of mismatched heteroduplexes at each site allowed us to compare deamination of three perfectly matched hybrids with mispaired hybrids in three different sequence contexts.

For each of three sites, the mismatched hybrid deaminated with a rate constant close to that of single-stranded DNA, while the perfectly matched DNA was up to 2 orders of magnitude more resistant to deamination. Thus, there is a striking difference in propensity for deamination among matched and mismatched cytosines. Comparison of deamination rates between a C·C and a T·C mispair is more difficult. For sites 147 and −11, the C·C mispair deaminated 2–4-fold faster than the T·C mispair at 60 °C; however, the order was reversed at 37 °C. The differences may not be significant

since rates differing by 2–3-fold are typical of the variation in the reversion frequencies for this type of experiment (Kunkel *et al.*, 1983; Frederico *et al.*, 1990).

Structure of Mismatched Base Pairs. Both C·C and T·C mispairs investigated in this study are pyrimidine–pyrimidine mispairs, which have been described as the type of mispair most distorting to the helix [reviewed by Modrich (1987)]. At room temperature, the most stable mismatches contain guanine; the least stable contain cytosine (C·T, C·A, and C·C), with C·C being the least stable of all (Aboul-ela *et al.*, 1985; Werntges *et al.*, 1986). Pyrimidine–pyrimidine mispairs in DNA do not conform to the standard DNA helix geometry and in some polynucleotides have been suggested to be slipped outside in an "open" state, rather than stacked inside the DNA helix (Werntges *et al.*, 1986; Topal & Fresco, 1976). Helix destabilization occurs at the site of a mismatch in double-stranded oligonucleotides, as evidenced by the drop in melting temperature (T_m) for a series of dodecamers with different mismatched bases. Pyrimidine–pyrimidine mismatches destabilize the helix more than purine–pyrimidine mismatches; *e.g.*, replacing two normal C·G base pairs with two C·T mismatches dropped the melting temperature (T_m) of a duplex dodecamer by 32 °C from that (72 °C) of the normal C·G base pair (Patel *et al.*, 1984). Such destabilization is also reflected by the increased rate of deamination in the C·T and T·C mismatches observed here.

Mismatched base pairs have different effects on perturbations at neighboring residues and on rates of helix opening. Whereas a G·T or G·A base pair affects only the directly adjacent base pair, an A·C or T·C base mispair can affect the neighboring three base pairs in the oligonucleotide (Patel *et al.*, 1984). The proton exchange rates (an indirect measure of the base pair opening lifetimes; Geuron *et al.*, 1987; Leroy *et al.*, 1988) of imino protons that are 1–3 base pairs removed from the T·C mismatch were found to be almost 10-fold less than that of the parent helix (Patel *et al.*, 1984). The T·C mismatch had a dramatic effect on the deduced rate of helix opening (Patel *et al.*, 1984), increasing it by 10-fold compared to the properly paired C·G base pair. This increase is the same order of magnitude as the increase in deamination rate that we observe here. If each time the helix opens there is a small but finite chance that cytosine will deaminate, then as the rate of helix opening increases, the rate of deamination would also increase, in agreement with our findings.

The presence of a mismatch affects not only the thermal and kinetic stability but also the dimensions of the helix and its ability to be recognized by DNA repair systems. The juxtaposition of a pyrimidine opposite a pyrimidine serves to narrow and distort the helix. The fact that pyrimidine–pyrimidine mispairs result in a greater distortion of the DNA helix than the other mispairs may be related to their biological properties. The C·C and T·C mispairs are poorly repaired by mismatch repair, a system usually used to correct polymerase errors (Kramer *et al.*, 1984; Modrich, 1987, 1991).

Thus, the pyrimidine–pyrimidine mispairs destabilize and distort the DNA helix and lead to an "open" type of base pair. We predict that this distortion could have consequences for mutagenesis. The longer a mispair goes unrepaired, the greater the chance for its deamination and subsequent fixation of this mutation. As shown here, the rates of cytosine deamination in C·C and T·C mismatches are accelerated dramatically compared to that in a perfectly matched C·G base pair. Thus, the data from our experiments are a further confirmation that mispairs destabilize DNA. Furthermore, they show that mispairs are more prone than normal pairs to undergo spontaneous hydrolytic deamination.

Deamination in Matched Heteroduplexes. The approximate deamination rate constants for the perfectly matched G·C ndDNA (Tables III and IV) are consistent with estimates derived from the literature. Cytosine in native *E. coli* DNA was estimated by Lindahl and Nyberg (1974) to deaminate with an (initial) average rate constant of $<1 \times 10^{-10}$ s^{-1} at 70 °C; by extrapolation using their calculated activation energy, the 60 °C value would be on the order of $<3 \times 10^{-11}$ s^{-1}. This is consistent with the value of 2×10^{-11} s^{-1} for site-specific deamination of cytosine in ds M13mp2 DNA obtained by extrapolating 37 °C data from Frederico *et al.*, (1990) to 60 °C. The "less than or equal to" values in Table III for matched hybrids are consistent with these earlier observations. Likewise, at the lower temperature of 37 °C, the value ($<5 \times 10^{-12}$ s^{-1}) in Table IV for the nicked heteroduplex is in the same range as that (7×10^{-13} s^{-1}) found by Frederico *et al.* (1990) for double-stranded M13mp2 DNA.

Effect of Neighboring Bases on Cytosine Deamination Mutagenesis in Mismatches. For the mismatched T·C and C·C heteroduplexes, the 2–3-fold site-to-site variations in deamination rate constants were below, or at, normal variation levels for this type of assay. Among a limited number of sites assayed, neighboring bases thus appear not to affect deamination rates in mismatches within the limits of the assay, *i.e.*, no more than a 3-fold variation at any one temperature.

Conclusion. Our studies on deamination of mispaired cytosines show that a cystosine residue in a C·C or C·T mispair in DNA is 10–100-fold more prone to deaminate than cytosine in a perfectly matched double-stranded DNA. Cytosine appears to derive its protection from spontaneous deamination by the local environment offered by the Watson–Crick base pairing with guanine and not necessarily by the double-stranded nature of the entire molecule. Even though only one base pair was mismatched (in an otherwise perfectly matched 7196 base pair duplex M13mp2 DNA), its rate of deamination increased substantially at both 37 and 60 °C when compared to a control having a complementary base pair at the same site.

The genetic assay developed is sensitive enough to reveal mutations and quantitate rates of reactions that occur with half-lives of 10 000 years, with an estimated accuracy to within a factor of 2 or 3 (Frederico *et al.*, 1990). Here, the assay was modified to enable us to discriminate between cytosine deamination and mismatch repair events. We estimate the accuracy to be about the same. The deamination rate constants obtained for mismatched cytosine in the nicked duplex DNA molecules were on the order of $(8–40) \times 10^{-10}$ s^{-1} at 60 °C and $(4–13) \times 10^{-10}$ s^{-1} at 37 °C. Although these are some of the slowest reactions ever to be accurately measured, they are 10–100-fold faster than the same reaction in double-stranded DNA.

ACKNOWLEDGMENT

We thank Ms. Shereen Barry, Mr. Ted Gonzalez, Dr. Richard Moyer, Dr. Paul Modrich, and Dr. Ken Tindall for comments on the manuscript.

REFERENCES

Aboul-ela, F., Koh, D., Martin, F. H., & Tinoco, I., Jr. (1985) *Nucleic Acids Res. 13*, 4811–4825.

Frederico, L. A., Kunkel, T. A., and Shaw, B. Ramsay (1990) *Biochemistry 29*, 2532–2537.

Garrett, E. R., and Tsau, J. (1972) *J. Pharm. Sci. 61*, 1052–1061.
Gueron, M., Kochoyan, M., and Leroy, J.-L. (1987) *Nature 328*, 89–92.
Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–580.
Kornberg, A., and Baker, T. (1991) *DNA Replication*, 2nd ed., W. H. Freeman and Co., San Francisco, CA.
Kramer, B., Kramer, W., and Fritz, H. G. (1984) *Cell 38*, 879–887.
Kunkel, T. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1494–1498.
Kunkel, T. A., and Soni, A. (1988) *J. Biol. Chem. 263*, 4450–4459.
Kunkel, T. A., Schaaper, R. M., and Loeb, L. A. (1983) *Biochemistry 22*, 2378–2384.
Leroy, J. L., Kochoyan, J., Huynh-Dinh, T., and Gueron, M. (1988) *J. Mol. Biol. 200*, 223–238.
Lindahl, T., and Nyberg, B. (1974) *Biochemistry 13*, 3405–3410.
Marmur, J., and Doty, P. (1962) *J. Mol. Biol. 5*, 109–118.
Modrich, P. (1987) *Annu. Rev. Biochem. 56*, 435–466.
Modrich, P. (1991) *Annu. Rev. Genet. 25*, 229–53.
Patel, D. J., Kozlowski, S. A., Ikuta, S., and Itakura, K. (1984) *Fed. Proc. 43*, 2663–2670.
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5476.
Shapiro, R. (1981) in *Chromosome Damage and Repair* (Seberg, E., and Kleppe, K., Eds.) pp 3–18, Plenum Press, New York.
Shapiro, R., and Klein, R. S. (1966) *Biochemistry 5*, 2358–2362.
Sowers, L. C., Shaw, B. R., Veigl, M. L., and Sedwick, W. D. (1987) *Mutat. Res. 177*, 201–218.
Topal, M. D., and Fresco, J. R. (1976) *Nature 263*, 285–289.
Werntges, H., Steger, G., Riesner, D., and Fritz, H.-J. (1986) *Nucleic Acids Res. 14*, 3773–3790.