# Fidelity and Mutational Spectrum of *Pfu* DNA Polymerase on a Human Mitochondrial DNA Sequence

Paulo André, Andrea Kim, Konstantin Khrapko, et al.

*Genome Res.* 1997 7: 843-852
Access the most recent version at doi:10.1101/gr.7.8.843



GENOME METHODS

# Fidelity and Mutational Spectrum of *Pfu* DNA Polymerase on a Human Mitochondrial DNA Sequence

**Paulo André, Andrea Kim, Konstantin Khrapko, and William G. Thilly**[1]

Division of Toxicology, Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The study of rare genetic changes in human tissues requires specialized techniques. Point mutations at fractions at or below $10^{-6}$ must be observed to discover even the most prominent features of the point mutational spectrum. PCR permits the increase in number of mutant copies but does so at the expense of creating many additional mutations or ''PCR noise''. Thus, each DNA sequence studied must be characterized with regard to the DNA polymerase and conditions used to avoid interpreting a PCR-generated mutation as one arising in human tissue. The thermostable DNA polymerase derived from *Pyrococcus furiosus* designated *Pfu* has the highest fidelity of any DNA thermostable polymerase studied to date, and this property recommends it for analyses of tissue mutational spectra. Here, we apply constant denaturant capillary electrophoresis (CDCE) to separate and isolate the products of DNA amplification. This new strategy permitted direct enumeration and identification of point mutations created by *Pfu* DNA polymerase in a 96-bp low melting domain of a human mitochondrial sequence despite the very low mutant fractions generated in the PCR process. This sequence, containing part of the tRNA glycine and NADH dehydrogenase subunit 3 genes, is the target of our studies of mitochondrial mutagenesis in human cells and tissues. Incorrectly synthesized sequences were separated from the wild type as mutant/wild-type heteroduplexes by sequential enrichment on CDCE. An artificially constructed mutant was used as an internal standard to permit calculation of the mutant fraction. Our study found that the average error rate (mutations per base pair duplication) of *Pfu* was $6.5 \times 10^{-7}$, and five of its more frequent mutations (hot spots) consisted of three transversions (GC → TA, AT → TA, and AT → CG), one transition (AT → GC), and one 1-bp deletion (in an AAAAAA sequence). To achieve an even higher sensitivity, the amount of *Pfu*-induced mutants must be reduced.

The ability to analyze point mutations occurring at low fractions in DNA samples is necessary for the study of somatic mutations in human tissues. For mitochondrial point mutations the more prominent hot spots arise at fractions from $\sim 2 \times 10^{-4}$ down to current limits of detection of $\sim 10^{-6}$ (Khrapko et al. 1997). Nuclear point mutant fractions as low as $10^{-7}$ may have to be determined before reproducible mutational spectra can be observed. The development of appropriate technology for this purpose has been facilitated greatly by use of PCR (Kleppe et al. 1971; Saiki et al. 1985, 1986; Chen and Thilly 1994; Khrapko et al. 1995a), but any DNA polymerase used in this technique will make errors during DNA synthesis, with the type and rate of errors varying among the specific DNA polymerases and reaction conditions used (Goodman et al. 1974; Keohavong and Thilly 1989; Eckert and Kunkel 1990, 1991; Cariello et al. 1991; Ling et al. 1991; Barnes 1992; Cha and Thilly 1993; Keohavong et al. 1993). In our laboratory we use constant denaturant capillary electrophoresis (CDCE) combined with PCR to study point mutations at fractions at or above $10^{-6}$ directly from human tissues (Khrapko et al. 1994a, 1997). The sensitivity of this procedure depends on the fidelity level of the polymerase used to catalyze the PCR. Therefore, we searched for an enzyme with high fidelity, and *Pfu*, the thermostable DNA polymerase derived from *Pyrococcus furiosus*, showed promise. Prior testing of *Pfu* reported an error rate of $1.6 \times 10^{-6}$ per base per doubling using a *lacI*-based bacterial assay (Lundberg et al. 1991), and a similar error rate of $2 \times 10^{-6}$ was reported using a p53-based biological assay (Flaman et al. 1994). However, an error rate one order of magnitude higher, $\sim 2.5 \times 10^{-5}$, was reported using

[1]Corresponding author.
E-MAIL thilly@mit.edu; FAX (617) 258-5424.

denaturing gradient gel electrophoresis (DGGE) (Cariello and Skopek 1993) and single-strand conformation polymorphism analysis (SSCP) (Brail et al. 1993). The differences in these error rates demanded clarification.

One point should be emphasized. Our study aimed only at the 96-bp fragment that we wanted to study in the human mitochondrial genome. We needed to know the hot spot mutations created by *Pfu* PCR in this sequence because they could have been mistaken for signals coming from human tissues. We hope that knowledge of the numerical distribution of unidentified mutations and the identity of the most prominent hot spots will be of value to others using this DNA polymerase in studies of rare genetic events.

We decided to measure the fidelity of *Pfu* using CDCE, which can separate DNA sequences differing by single base substitutions, small additions, or deletions. (Khrapko et al. 1994a,b). As was the case with other DNA polymerases (Keohavong and Thilly 1989; Keohavong et al. 1993), we expected the *Pfu* enzyme to generate an idiosyncratic mutational spectrum during DNA synthesis in vitro. Our study presents the first published description of these *Pfu*-induced hot spots.

By studying the accumulation of mutants as a function of DNA doublings, we discovered that *Pfu* has an error rate of $6.5 \times 10^{-7}$. This high fidelity allows for greater sensitivity in mutational spectrometry than would be possible with other polymerases. We have applied it accordingly in our study of mitochondrial mutational spectra in human cells and tissues (Khrapko et al. 1997).

Our ongoing research is aimed at further increasing the sensitivity of our techniques. Understanding the nature and origins of PCR mutants can help us to design protocols that reduce the amount of PCR noise and thus heighten the sensitivity of mutational spectra analysis.

## RESULTS

### Determination of the Homoduplex Fraction and Error Rate of *Pfu* by CDCE

Each PCR sample was run through a capillary at 9 µA current using a 5-cm water jacket at 65.0°C. The heteroduplex region was collected when it arrived at the end of the capillary into 8 µl of 0.2× TBE. Because this step achieved an ~20-fold enrichment of mutants relative to wild type, the subsequent PCR steps contributed <5% of the total mutants analyzed. This enriched heteroduplex fraction was amplified by *Pfu* DNA polymerase to create heteroduplexes. The PCR reaction was subjected to a second CDCE separation followed by another PCR. After a third CDCE purification, and after determining that the most prominent heteroduplex peaks were all represented as prominent homoduplex peaks, the collected fraction was amplified by *Pfu* to create mutant homoduplexes. This final PCR was run on CDCE with a 15-cm heated zone at a temperature of 70.6°C with a 5-µA current. Figure 1 presents the CDCE display of all the mutant homoduplexes after 14 and 34 doublings.

After 14 doublings the four most prominent peaks, labeled A, B, C, and E, were clearly visible as were perhaps some 15 other very small peaks. None of even the smallest peaks appear in the control in

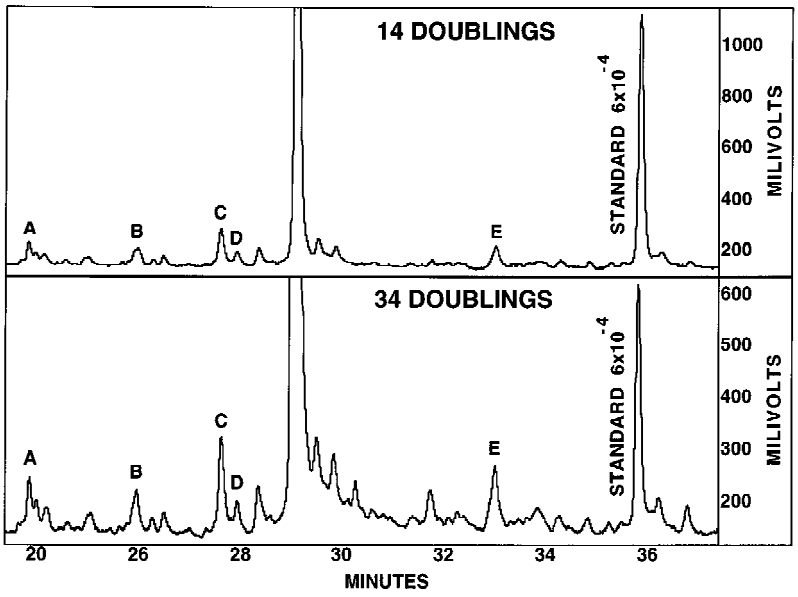


Figure 1 Analysis by CDCE of the purified homoduplex mutant fractions after 14 and 34 doublings using *Pfu* DNA polymerase. This CDCE was run using a 5% polyacrylamide gel in a 35-cm capillary with a 15-cm heated zone at a temperature of 70.6°C with a 5-µA current. Many small peaks after 14 doublings may be seen to increase significantly after 34 doublings. Of the major peaks for which the mutant sequences were determined, peak A is an AT → CG transversion at bp 10071 of human mitochondrial DNA; peak B is an AT → GC transition at bp 10108; peak C is an AT deletion occurring in a run of six consecutive AT base pairs at positions 10048–10053; peak D is an AT → TA transversion at bp 10071; and peak E is a GC → TA transversion at bp 10070.

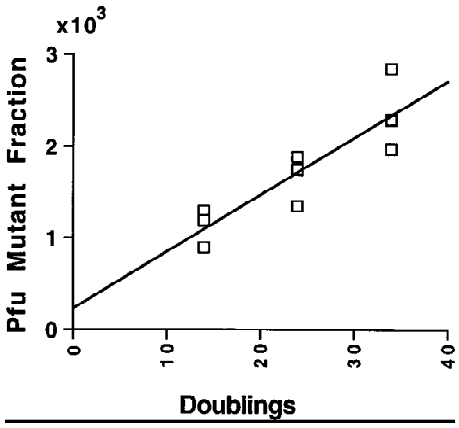


Figure 2 The relationship between the number of doublings and the mutant fraction of all mutants created by *Pfu.* The mutant fractions from three independent experiments were estimated by comparing the total area of all mutant peaks with the area of an internal standard. A linear regression analysis was applied to all data points. By dividing the estimated slope by the length of the low-temperature melting domain of our DNA fragment (96 bp), we calculated an error rate of $6.5 \pm 3.1 \times 10^{-7}$ mutants per base pair per doubling (95% confidence interval).

which DNA was amplified but not subjected to enrichment for mutants. At 34 doublings these four peaks had increased in area relative to the internal standard, and ~22 small peaks could be discerned. The 15 small peaks at 14 doublings were all found again but of larger areas after 34 doublings. We interpret this to indicate that peaks A, B, C, and E are the most probable errors produced in this sequence by *Pfu* PCR. It should be noted that there are other minor peaks not visible at 14 doublings that appear as ''nubbins'' at 34 cycles but increase in area at >50 doublings. We can see these only in the positions between the larger peaks, but it appears that there must be many of these given their frequency where they can be seen. Perhaps as many as 80 or 90 such nubbins lie across the entire output.

The total mutant homoduplex mutant fraction for any experiment was estimated by simply summing the areas of all mutant peaks divided by the area of the internal standard, the artificial mutant (AM). The mutant fractions from three independent experiments are shown plotted in Figure 2. This graph describes the relationship between number of doublings and the total mutant fraction created by *Pfu*. A linear regression analysis was applied to all data points. The error rate $f$ was calculated as $f = (\text{slope})/b$, where the slope was estimated using the linear regression of summed mutant fractions on number of doublings and $b$ was the length in base pairs, 96, of the low-temperature isomelting domain of our DNA sequence in which mutants could be detected. By this approach, our estimated error rate for *Pfu* DNA polymerase is $6.5 \pm 3.1 \times 10^{-7}$ mutations per base pair per doubling (95% confidence interval on the estimate of the slope).

## Sequence Analysis of Some of the Predominant *Pfu* Mutations

The four most prominent peaks (A, B, C, and E) plus the minor peak D (see Fig. 1) were collected individually by elution after CDCE. The peaks were then PCR amplified and sequenced. The most frequent mutations (hot spots) consisted of three transversions (GC → TA, mutant E, AT → TA, mutant D, and AT → CG, mutant A), one transition (AT → GC, mutant B), and a one 1-bp deletion in a AAAAAA sequence, mutant C (See Fig. 3) Two of the transversions (AT → TA, mutant D, and AT → CG, mutant A) occurred in the same position, bp 10071.

The increase in total mutant fraction from 14 to 24 to 34 doublings (see Fig. 4) is a linear function of the number of doublings. One can thus exclude the possibility that the CDCE-purified wild-type DNA contained cryptic damage that might have caused miscoding at some sites during in vitro DNA synthesis. If the noise observed had been the result of bypass of adducts present in the original DNA template, its amount would have been independent of the number of DNA duplications. However, the linear increase in the amount of noise shown here does

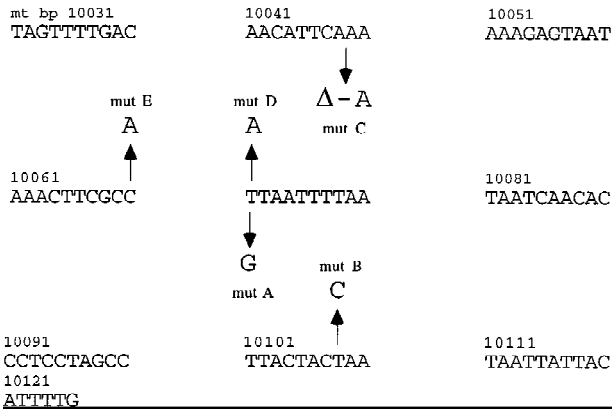


Figure 3 Summary of the kinds and positions of the mutations induced by *Pfu* DNA polymerase. The low melting domain of 96 bp in which we can detect mutations extends from the mitochondrial genome position at bp 10,031 to 10,126. The symbol Δ is used to indicate a base deletion.

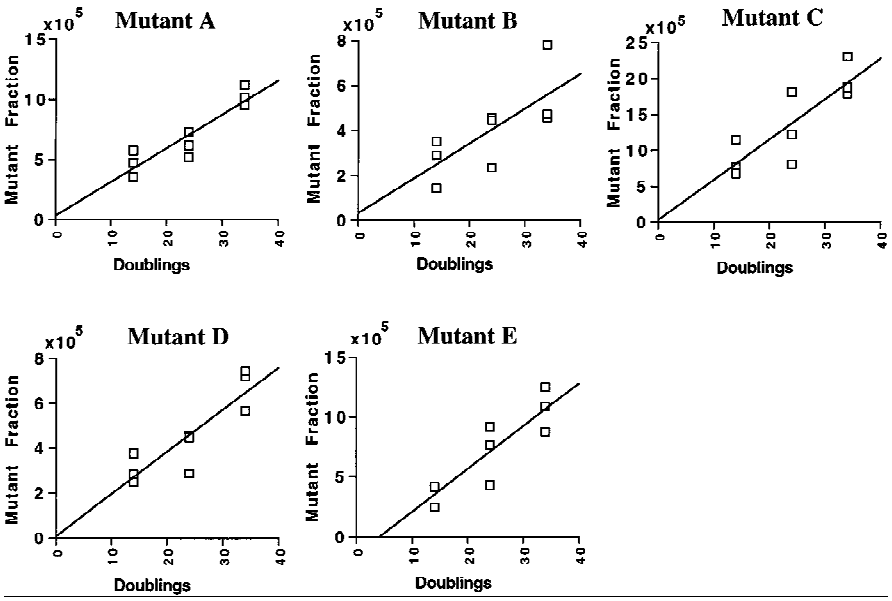


Figure 4 Plots of the mutant hot spot fractions vs. the amount of DNA doublings using data from three independent experiments. A linear regression analysis was applied to all data points. The mutation rate per doubling for each hot spot was estimated from the slope. The mutation rate for hot spot A was $2.8 \pm 1.1 \times 10^{-6}$; for hot spot B, $1.6 \pm 1.3 \times 10^{-6}$; for hot spot C, $5.6 \pm 3.3 \times 10^{-6}$; for hot spot D, $1.9 \pm 0.9 \times 10^{-6}$; and for hot spot E, $3.6 \pm 1.7 \times 10^{-6}$ (95% confidence intervals for the best estimates of the slopes).

not exclude the possibility that the source of our mutants was DNA damage created during the PCR process.

## DISCUSSION

### *Pfu* Fidelity

The estimate of $6.5 \times 10^{-7}$ errors per base pair per duplication is somewhat lower than earlier reports by Lundberg et al. (1991) and Flaman et al. (1994), which found an error rate of ~$2 \times 10^{-6}$ using *lacI*- or *p53*-based biological assays. Short heating times and the almost anaerobic conditions of capillary PCR may contribute to the lower mutation rate that we observed. However, it is very difficult to make a direct comparison between the error rate obtained here with those reported in the literature, as other researchers use different DNA sequences, analytical methods, and PCR procedures.

Despite these limitations, we can speculate about the disparity between the $6.5 \times 10^{-7}$ error rate we measured using CDCE and the $2.5 \times 10^{-5}$ rate measured by Cariello and Skopek (1993) using DGGE, because the physical principles underlying CDCE and DGGE separations are similar (Khrapko et al. 1994a). We hypothesize that this disparity can be explained in part by our use of a mutant enrichment step that decreases the interference of a particular class of PCR-created artifacts. This class of artifacts arises when either of the DNA strands of the low melting domain in which we detect mutants differs in length from the desired PCR product by one or more nucleotides. These products may arise from incomplete synthesis, terminal transferase activity, or ''nibbling'' by an associated exonuclease activity (Hu 1993). They may also melt partially at a lower temperature than the complete fragment and migrate more slowly than wild-type DNA in the partially denaturing conditions used in DGGE or CDCE (Khrapko et al. 1997). When collected by CDCE and amplified by PCR, these artifacts are replicated as wild-type DNA as opposed to true mutants. However, if these products are measured along with the mutant heteroduplexes, one may overestimate the mutation rate. In our study we measured *Pfu*-generated mutants after three enrichments through CDCE because at that point, the amount of PCR artifacts was negligible compared to the amount of true *Pfu*-induced mutants. However, when we attempted to estimate the *Pfu* error rate without mutant enrichment, we obtained a value ~$10^{-5}$, similar to the one reported using DGGE.

Like Cariello and Skopek, Brail et al. (1993) found an error rate larger than ours (~$3 \times 10^{-5}$ vs. $6.5 \times 10^{-7}$) but used SSCP. Recognizing that SSCP is very sensitive to the different lengths of the artifacts created during PCR, Brail et al. tried to eliminate these artifacts by using a polyacrylamide gel electrophoresis (PAGE) purification step before using SSCP to measure the error rate. However, because PAGE was unable to separate out the small differences in DNA strand lengths present in some PCR artifacts, they may, nonetheless, have overestimated the error rate of *Pfu*.

To avoid an error rate overestimation when physical–chemical separation methods (DGGE,

CDCE, or SSCP) are used, we now measure error rates only after mutant fraction enrichment. This is especially important when measuring the fidelity of polymerases such as *Pfu* with a low error rate. The importance of having observed a linear increase in mutant fraction for each of the five hot spots reported as opposed to simply the sum of all non-wild-type material is emphasized by this consideration.

The fidelity of *Pfu* is significantly higher than the reported fidelities of other commonly used thermostable polymerases: *Taq* has an error rate ~$10^{-4}$, and *Vent* has an error rate ~$10^{-5}$ (Saiki et al. 1988; Tindall and Kunkel 1988; Keohavong and Thilly 1989; Flaman et al. 1994). Therefore, *Pfu* is currently the most appropriate polymerase to use in mutational analyses that demand high sensitivity.

## *Pfu* Hot Spots

The kinds of mutations created during PCR are dependent on DNA polymerases interacting with the particular DNA sequence. These mutations are distributed nonrandomly as hot spots (Keohavong and Thilly 1989; Mendelman et al. 1989; Kunkel 1990). Whereas AT → GC transitions are the predominant mutations created by *Taq* and *Vent* DNA polymerases (Keohavong and Thilly 1989; Keohavong et al. 1993), *Pfu* seems to have a different mode of operation. Of the four most prominent hot spots we found that *Pfu* PCR two transversions (GC → TA and AT → CG), one transition (AT → GC), and a one 1-bp deletion (in a AAAAAA sequence; see Figs. 1 and 3). One cannot assume that what we found in this 96-bp sequence predicts the kind of most prominent hot spots in another sequence. For application to the measurement of low point mutation fractions in other sequences as intended for human tissues, the process outlined herein will have to be repeated to differentiate signals from the most obvious PCR noise hot spots.

## Origins of *Pfu* Hot Spots

With CDCE's separation power combined with the fidelity of *Pfu* DNA polymerase, we have been able to observe the set of point mutational hot spots that arise in the mitochondrial genome of human cells and tissues (Khrapko et al. 1997). However, another order of magnitude in resolving power will be required to extend our observations to the mutational spectra of single copy nuclear genes. The sensitivity of our CDCE/hifiPCR protocol could be increased by further reducing the fraction of *Pfu*-induced mutants. Understanding the mechanisms by which these mutants are created might help us devise a strategy to further decrease PCR noise.

In the following section, we discuss six possible explanations for the origin of PCR noise: polymerase mispairing errors, strand slippage, polymerase microheterogeneity, miscoding lesions in the original DNA, heat-induced mutagenesis, and chemical modifications in the nucleotide pool. Most discussions of PCR noise focus on the first two possibilities, which refer to the idiosyncrasies of an enzyme's behavior relative to the local variations in the DNA sequence being replicated. It is conceivable that each *Pfu* hot spot studied here may have resulted from a stochastic error function distributed action of a homogeneous collection of polymerase molecules interacting with the local DNA sequence. However, we think that the phenomenon may be somewhat more complex.

### *Polymerase Mispairing Errors*

A hot spot mutation is the end result of preferential misincorporation and failure of proofreading activity at a specific site. Either the rate of misincorporation or the probability of exonucleolytic excision could govern the observed error rate for a given polymerase. *Pfu* replicates DNA with such high accuracy because it combines high-fidelity nucleotide incorporation with correction of mispaired nucleotides by a proofreading 3′ → 5′ exonuclease activity (Eckert and Kunkel 1991; Lundberg et al. 1991; Cha and Thilly 1993). Using a mutant polymerase lacking proofreading activity, Kroutil et al. (1996) found that the contribution of the exonuclease to fidelity decreased substantially as the length of homopolymeric runs increased. In future studies, we will use the same approach to determine how much the *Pfu* exonuclease activity contributed to the correct replication of our DNA sequence.

### *Strand Slippage*

Template–primer slippage in runs of identical bases can explain both frameshift (Kunkel 1986; Streisinger et al. 1966) and base substitution errors (Kunkel and Soni 1988). Given that both the polymerase selectivity and the exonucleolytic proofreading efficiency of *Pfu* diminishes as the number of repeats in the sequence increases (Kroutil et al. 1996) and that the DNA sequence used in our study contained many runs of 3 to 6 AT bp and one run of

3 GC bp, it is not surprising that our biggest hot spot (mutant C) was a deletion of 1-bp in a run of six consecutive AT base pairs. The mutation rate of hot spot C ($5.6 \times 10^{-6}$ per doubling (see Fig. 4) is comparable to the rate of $1.2 \times 10^{-5}$ estimated by Kroutil et al. (1996) on a run of five AT base pairs in a different DNA sequence.

### *Microheterogeneity of the* Pfu *Polymerase*

A subset of *Pfu* polymerase molecules may be particularly error-prone via microheterogeneity. That is, some enzyme molecules may contain a different amino acid sequence (Ninio 1991a,b; Cha and Thilly 1993; Slupska et al. 1996). Such microheterogeneity might arise from mistranscription, mistranslation, or a variety of rare chemical reactions with the ''normal'' polymerase. The *Pfu*-induced mutational spectra reported here may be the result of this error-prone polymerase subset within an ensemble of normal molecules with a significantly higher fidelity. If the set of polymerase molecules contained a subset of polymerase proficient-exonuclease-deficient molecules, this subset could be responsible for the observed errors of the entire set. Possibly relevant to this point is a prior observation made in our laboratory using exonuclease negative *Vent* DNA polymerase. Keohavong et al. (1993) observed that the mutational spectrum in PCR was identical except in magnitude with the exonuclease ''positive'' preparation.

### *Miscoding Lesions Present in the Original DNA*

There is always the possibility that the original DNA used in our experiments contained ''cryptic'' damaged sites that would have been miscoded during PCR (Kunkel and Alexander 1986). If one of our mutant hot spots was the result of a miscoding lesion present in our 'pure' DNA, we could predict that its fraction would be established early in the PCR process and remain unchanged with increased doublings. However, this behavior was not observed with any of our hotspots (see Fig. 4.) All mutant fractions increased from 14 to 24 to 34 doublings.

### *Heat-Induced Mutagenesis*

PCR requires multiple rounds of heat denaturation, rehybridization, and DNA synthesis, with the temperature oscillating between 50°C and 94°C. Heat treatment of DNA is known to be mutagenic, generating both transitions (Baltz et al. 1976; Ripley 1988) and transversions (Ripley 1988; Kricker and Drake 1990). In bacteriophage T4, heat-induced mutations are exclusively base-pair substitutions at GC base pairs, but these are of at least two types: GC → AT transitions, produced by the deamination of cytosine, and transversions (GC → TA) possibly resulting from misincorporation of dAMP opposite of a modified guanosine. Depurination followed by the insertion of any of three ''wrong'' bases during DNA replication has also been invoked to explain the mutagenic action of heat.

Potential contributions of each of these pathways to PCR noise are described below in turn.

#### CONTRIBUTION OF CYTOSINE DEAMINATION TO *PFU* NOISE

Hydrolytic deamination of cytosine converts it to uracil and ultimately generates GC → AT transitions after a round of DNA replication. This reaction proceeds at a rate of ~$2 \times 10^{-7}$/sec at 95°C (pH between 6 and 9) (Lindahl and Nyberg 1974). If this rate is applicable to our PCR conditions, our experiment should have generated $2 \times 10^{-6}$ deaminations per doubling, per cytosine residue present in our DNA fragment (1.5 cycles per doubling × 7 sec denaturation at 94°C × $2 \times 10^{-7}$/sec) and consequently the same amount of GC → AT transitions, assuming that dUMP had the same coding potential as dTMP and its presence was not inhibitory to DNA synthesis (Longo et al. 1990; Kunkel et al. 1991). However, none of our five predominant hot spots was a GC → AT transition. It is possible that the unsequenced minor hot spots might be such transitions, but most of them have higher melting temperatures than the wild type as homoduplexes making this possibility unlikely.

#### CONTRIBUTION OF MODIFIED GUANOSINE (G*) TO *PFU* NOISE

In another heat-promoted process in bacteriophage T4, guanosine residues undergo a glycosylic bond migration (Kricker and Drake 1990). This modified G mispairs with dAMP at a frequency ~10% (Ripley 1988), generating GC → TA transversions such as the mutant E found in our study (see Fig. 3). Experiments in which incubation time at 94°C is varied could reveal whether mutant E is a heat-induced DNA lesion. Again, these might occur among the unsequenced minor hot spots; however, the hot spots would have been expected to have melting temperatures below the wild type, whereas most minor hot spots have melting temperatures above the wild-type homoduplex.

### CONTRIBUTION OF DEPURINATION TO *PFU* NOISE

The sugar-base glycosyl bond of purine residues is particularly susceptible to hydrolysis. At pH 7.4 and at 95°C, depurination occurs at a rate ~$4 \times 10^{-7}$/sec (Lindahl and Nyberg 1972; Lindahl and Karlstrom 1973). Because the reaction occurs faster at lower pH, the depurination rate should have been higher in our PCR buffer (pH 6 at 94°C). An expected $4 \times 10^{-6}$ apurinic lesion per doubling should have been generated per purine residue present in our DNA fragment (1.5 cycles per doubling × 7 sec denaturation at 94°C × $4 \times 10^{-7}$/sec). It has been shown previously that enzymes with exonuclease activity are less efficient at replicating past apurinic sites and thus may not give rise to mutations (Kunkel et al. 1983; Loeb 1985). If in some cases *Pfu* misreplicated past an apurinic site, transversions or transitions would be expected. It is possible that depurination could explain the presence of mutants A (AT → CG), B (AT → GC), D (AT → TA), and E (GC → TA) in our study.

#### *Chemical Modifications in the Nucleotide Pool*

Spontaneous oxidation of dGTP forms 8-hydroxy-dGTP (8-oxo-7,8-dihydro-2′-dGTP), which is inserted opposite dA and dC residues of template DNA with almost equal efficiency (Maki and Sekiguchi 1992; Michaels and Miller 1992). Once inserted, this modified nucleotide could continue its miscoding propensity in a subsequent cycle. The GC → TA (hot spot E) and AT → CG (hot spot A) transversions found in our study could have resulted from this process. The spontaneous oxidation of guanine could also occur in the DNA itself (Michaels and Miller 1992). (The spontaneous deamination of dCTP in the nucleotide pool is not expected to be mutagenic because incorporating dUTP in DNA synthesis is equivalent to incorporating dTTP.)

In sum, our five *Pfu* hot spots are compatible with any of several possible mutagenic mechanisms including polymerase mispairing errors, strand slippage, polymerase microheterogeneity, heat-induced mutagenesis (depurination or modified guanosine residues), or the spontaneous formation of 8-hydroxyguanine. Further study is necessary to quantify the contribution of each of these pathways to *Pfu* noise. However, heat-induced cytosine deamination or miscoding lesions present in the original DNA are unlikely to have contributed significantly to *Pfu* noise represented by the four predominant hot spots.

## CONCLUSION

Our observations are perforce limited to the 96-bp target studied, and our specific observations should not be used to generalize expectations for other DNA sequences. But the process we describe herein is essential for the accurate assessment of mutational spectra in human tissues where low mutant fractions are encountered and for which PCR DNA amplification is essential. Knowing the amounts and kinds of *Pfu* polymerase-induced mutations within our particular DNA sequence was essential for our studies of mitochondrial mutations in human tissues (Khrapko et al. 1997). Additional research is needed to discover the mechanistic basis of *Pfu* hot spots. Knowledge of the actual mechanism(s) may permit the design of protocols that decrease PCR noise and increase the sensitivity of mutational spectrometry.

## MATERIALS AND METHODS

### Equipment

The CDCE instrumentation used was similar to that described previously (Hanekamp et al. 1996; Khrapko et al. 1994b), with few modifications. Electrophoresis was performed in 75 μm interior diameter, 350 μm outer diameter precoated capillaries filled with 5% linear polyacrylamide in TBE (89 mM Tris, 89 mM boric acid, and 1 mM EDTA at pH 8.3). The polyacrylamide medium in the capillary was replaced prior to each run. PCR products were diluted 10 times and loaded into the capillary by applying 2 μA current for 15 sec. Under these conditions, the capillary was loaded with ~$10^8$ copies of amplified DNA fragments. A portion of the capillary was heated by a water jacket connected to a constant temperature circulating water bath. The length of the jacket was 5 cm for the enrichment of heteroduplexes and 15 cm for high-resolution CDCE separation of mutant homoduplexes. To detect DNA, the capillary was illuminated by a 515-nm argon laser, and emitted light was collected at a right angle by a microscope objective. This light was directed through two filters, 540-nm bandpass and 530-nm long pass, into a photomultiplier. The signal from the photomultiplier was amplified by a current preamplifier and recorded by an MP 100 data acquisition system (Biopac Systems).

### PCR

PCR was performed inside closed glass capillaries using an Air Thermo Cycler (Idaho Technologies). Native *Pfu* DNA polymerase was obtained from Stratagene (La Jolla, CA). The 2′-deoxynucleoside-5′-triphosphates were purchased as 100 μM solutions from Pharmacia (Piscataway, NJ). The primers CW7 (5′-ACCGTTAACTTCCAATTAC) and 5′-fluorescein-labeled J3 (5′-ATGGAGAAAGGGACGCGGGC) were obtained from Synthetic Genetics (San Diego, CA). Each 10 μl PCR reaction mixture contained 20 mM Tris, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 2 mM $MgCl_2$, 100 μM each dNTP, 110 μg/ml of

nuclease free bovine serum albumin, 0.2 μM each primer, and 0.04 U/μl of *Pfu* DNA polymerase. The reaction mixture pH was 8.5 at 25°C but at 94°C the pH should be ~6. (The pH of Tris buffers decreases at higher temperatures: ~0.035 pH units/°C (Brail et al. 1993).) Each cycle of the PCR reactions consisted of 7 sec at 94°C for DNA template denaturation, 15 sec at 50°C for template–primer hybridization, and 5 sec at 72°C for DNA polymerization. After cycling, the PCR mixture was incubated at 72°C for 2 min and at 45°C for 30 min.

The DNA template used in this study was a 206-bp human mitochondrial fragment containing part of the tRNA glycine and part of the NADH dehydrogenase subunit 3 (Khrapko et al. 1994b). The sequence in which mutations can be detected is the 96-bp low-temperature isomelting domain on the 5′ end of the 206-bp fragment. Total DNA was isolated from human lymphoblastoid cells, line TK6 (Skopek et al. 1978), using *Pfu* polymerase and primers CW7 and J3 defined above. To measure the fidelity of DNA amplification, a CDCE-purified wild-type sequence was mixed with an artificial mutant (GC → AT at bp 10040) at a final fraction of $6 \times 10^{-4}$ (Khrapko et al. 1994a,b). During the first PCR, *Pfu* was used to amplify $6 \times 10^7$ copies of wild type to a total of $10^{12}$ (14 doublings). For samples subjected to 24 and 34 DNA doublings, a series of dilutions followed by PCR amplifications were performed. The efficiency of *Pfu* PCR under the conditions used is ~66%. This is somewhat lower than can be achieved for other thermostable DNA polymerases, but the higher fidelity more than makes up for the need for a few extra cycles. We use the term ''doubling'' to indicate the log to the base 2 of the fold increase in copy number in any PCR process. In the exponential phase of PCR, $N = N_o\ (1 + Y)^n$, where $N$ is the number of copies observed after n cycles, $N_o$ is the initial number of copies and $Y$ is the yield or efficiency per cycle. Thus, the number of doublings, $D = \log_2 (N/N_o) = \log_2 (1 + Y)^n = n \log_2 (1 + Y)$.

### Formation of Wild-Type/Mutant Heteroduplexes

The *Pfu* mutant fraction was enriched by collecting wild-type/mutant heteroduplexes separated from the wild-type peak. It is important to understand that by virtue of the 96-bp target being a DNA isomelting domain, any and all mutant/wild-type heteroduplexes will have significantly lower equilibrium melting temperatures. As a result, all heteroduplexes arising from PCR mutants will be well separated from wild-type homoduplex in this initial enrichment step. Heteroduplexes were created during PCR by subjecting the sample to a sufficient number of PCR cycles to exhaust the primers. When primers are exhausted, new full-length templates are no longer synthesized; instead, the full-length target sequence strands melt and reanneal with each other. Mutant strands are forced by mass action to form heteroduplexes with wild-type strands present in excess. Because mismatches destabilize the double helix, the mutants are separated more easily from the wild type in heteroduplex than in homoduplex form. All heteroduplexes can be collected together while avoiding the wild-type peak.

The CDCE-purified *Pfu* mutants were analyzed as homoduplexes in a subsequent run. If a mutant homoduplex has a different melting temperature from the wild-type homoduplex it elutes before or after the wild-type peak on CDCE. Most but not all mutants of a 100-bp sequence have significantly different melting temperatures from a wild-type homoduplex and can thus be conveniently separated. In the work described herein the five most prominent *Pfu* hotspots all separated well from the wild-type homoduplex, making it convenient to display the direct CDCE separations for the homoduplex forms. Readers should note that this is not always possible requiring analysis of heteroduplex peaks. Homoduplexes were created by stopping PCR when the molar amount of unused primers still exceeds that of the PCR product.

## ACKNOWLEDGMENTS

We are grateful to Hilary Coller, Aoy Tomita, Luisa Marcelino, and Jillian Dickert for their critical reading and careful editing of this manuscript. We also wish to thank Wei-Ming Zheng for providing reference material. This work was supported, in part, by the grants from the National Institute for Environmental Health Sciences: PO1-ESO7168, Mutagenic Effects of Air-Borne Toxicants; P42-ESO4675, Superfund Hazardous Substances Basic Research Program; and PO1-ESO3926, Genetics and Toxicology; and a grant from the U.S. Department of Energy, Office of Health and Environmental Research (DE-FGO2-86ER60448), Comparative Mutagenesis of Human Cells In Vitro and In Vivo.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked ''advertisement'' in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

Baltz, R.H., P.M. Bingham, and J.W. Drake. 1976. Heat mutagenesis in bacteriophage T4: The transition pathway. *Proc. Natl. Acad. Sci.* 73: 1269–1273.

Barnes, W.M. 1992. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* 112: 29–35.

Brail, L., E. Fan, D.B. Levin, and D.M. Logan. 1993. Improved polymerase fidelity in PCR-SSCPA. *Mutat. Res.* 303: 171–175.

Cariello, N.F. and T.R. Skopek. 1993. Mutational analysis using denaturing gradient gel electrophoresis and PCR. *Mutat. Res.* 288: 103–112.

Cariello, N.F., J.A. Swenberg, and T.R. Skopek. 1991. Fidelity of *Thermococcus litoralis* DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* 19: 4193–4198.

Cha, R.S. and W.G. Thilly. 1993. Specificity, efficiency, and fidelity of PCR. *PCR Methods Applic.* 3: S18–29.

———. 1994. Mutational spectrum of chromium(VI) in human cells. *Mutat. Res.* 323: 21–27.

Eckert, K.A. and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* 18: 3739–3744.

———. 1991. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Applic.* 1: 17–24.

Flaman, J.M., T. Frebourg, V. Moreau, F. Charbonnier, C. Martin, C. Ishioka, S.H. Friend, and R. Iggo. 1994. A rapid PCR fidelity assay. *Nucleic Acids Res.* 22: 3259–3260.

Goodman, M.F., W.C. Gore, N. Muzyczka, and M.J. Bessman. 1974. Studies on the biochemical basis of spontaneous mutation. III. Rate model for DNA polymerase-effected nucleotide misincorporation. *J. Mol. Biol.* 88: 243–235.

Hanekamp, J.S., P. Andre, H.A. Coller, X.-C. Li, W.G. Thilly, and K. Khrapko. 1996. CDCE—Constant denaturant capillary electrophoresis for detection and enrichment of sequence detection variants. In *Laborotory protocols for mutation detection* (ed. U. Landegren), pp. 38–41. Oxford University Press, New York, NY.

Hu, G. 1993. DNA polymerase-catalyzed addition of non-templated extra nucleotides to the 3′ end of DNA fragments. *DNA Cell Biol.* 8: 763–770.

Keohavong, P. and W.G. Thilly. 1989. Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci.* 86: 9253–9257.

Keohavong, P., L. Ling, C. Dias, and W.G. Thilly. 1993. Predominant mutations induced by the *Thermococcus litoralis*, vent DNA polymerase during DNA amplification in vitro. *PCR Methods Applic.* 2: 288–292.

Khrapko, K., P. Andre, R. Cha, G. Hu, and W.G. Thilly. 1994a. Mutational spectrometry: Means and ends. *Prog. Nucleic Acid Res. Mol. Biol.* 49: 285–312.

Khrapko, K., J.S. Hanekamp, W.G. Thilly, A. Belenkii, F. Foret, and B.L. Karger. 1994b. Constant denaturant capillary electrophoresis (CDCE): A high resolution approach to mutational analysis. *Nucleic Acids Res.* 22: 364–369.

Khrapko, K., H. Coller, P. Andre, X.-C. Li, F. Foret, A. Belensky, B.L. Karger, and W.G. Thilly. 1997. Mutational spectrometry without phenotypic selection: Human mitochondrial DNA. *Nucleic Acids Res.* 25: 685–693 .

Kleppe, K., E. Ohtsuka, R. Kleppe, I. Molineux, and H.G. Khorana. 1971. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* 56: 341–361.

Kricker, M.C. and J.W. Drake. 1990. Heat mutagenesis in bacteriophage T4: Another walk down the transversion pathway. *J. Bacteriol.* 172: 3037–3039.

Kroutil, L.C., K. Register, K. Bebenek, and T.A. Kunkel. 1996. Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry* 35: 1046–1053.

Kunkel, T.A. 1986. Frameshift mutagenesis by eucaryotic DNA polymerases in vitro. *J. Biol. Chem.* 261: 13581–13587.

———. 1990. Misalignment-mediated DNA synthesis errors. *Biochemistry* 29: 8003–8011.

Kunkel, T.A. and P.S. Alexander. 1986. The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. *J. Biol. Chem.* 261: 160–166.

Kunkel, T.A. and A. Soni. 1988. Mutagenesis by transient misalignment. *J. Biol. Chem.* 263: 14784–14789.

Kunkel, T.A., R.M. Schaaper and L.A. Loeb. 1983. Depurination-induced infidelity of deoxyribonucleic acid synthesis with purified deoxyribonucleic acid replication proteins in vitro. *Biochemistry* 22: 2378–2384.

Kunkel, T.A., K. Bebenek, and J. McClary. 1991. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* 204: 125–139.

Lindahl, T. and O. Karlstrom. 1973. Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. *Biochemistry* 12: 5151–5154.

Lindahl, T. and B. Nyberg. 1972. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11: 3610–3618.

———. 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* 13: 3405–3410.

Ling, L.L., P. Keohavong, C. Dias, and W.G. Thilly. 1991. Optimization of the polymerase chain reaction with regard to fidelity: Modified T7, *Taq*, and *Vent* DNA polymerases. *PCR Methods Applic.* 1: 63–69.

Loeb, L.A. 1985. Apurinic sites as mutagenic intermediates. *Cell* 40: 483–484.

Longo, M.C., M.S. Berninger, and J.L. Hartley. 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* 93: 125–128.

Lundberg, K.S., D.D. Shoemaker, M.W. Adams, J.M. Short, J.A. Sorge, and E.J. Mathur. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* 108: 1–6.

Maki, H. and M. Sekiguchi. 1992. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* 355: 273–275.

Mendelman, L.V., M.S. Boosalis, J. Petruska, and M.F. Goodman. 1989. Nearest neighbor influences on DNA polymerase insertion fidelity. *J. Biol. Chem.* 264: 14415–14423.

Michaels, M.L. and J.H. Miller. 1992. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* 174: 6321–6325.

Ninio, J. 1991a. Connections between translation, transcription and replication error-rates. *Biochimie* 73: 1517–1523.

———. 1991b. Transient mutators: A semiquantitative

analysis of the influence of translation and transcription errors on mutation rates. *Genetics* 129: 957–962.

Ripley, L.S. 1988. Estimation of in-vivo miscoding rates. *J. Mol. Biol.* 202: 17–34.

Saiki, R.K., S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350–1354.

Saiki, R.K., T.L. Bugawan, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1986. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324: 163–166.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487–491.

Skopek, T.R., H.L. Liber, B.W. Penman, and W.G. Thilly. 1978. Isolation of a human lymphoblastoid line heterozygous at the thymidine kinase locus: Possibility for a rapid human cell mutation assay. *Biochem. Biophys. Res. Commun.* 84: 411–416.

Slupska, M.M., C. Baikalov, R. Lloyd, and J.H. Miller. 1996. Mutator tRNAs are encoded by the *Escherichia coli* mutator genes mutA and mutC: A novel pathway for mutagenesis. *Proc. Natl. Acad. Sci.* 93: 4380–4385.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 31: 77–84.

Tindall, K.R. and T.A. Kunkel. 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* 27: 6008–6013.

*Received September 10, 1996; accepted in revised form June 24, 1997.*