Fidelity of DNA polymerases in DNA amplification

(polymerase chain reaction/denaturing gradient gel electrophoresis/sequencing/exon 3 of HPRT gene)

PHOUTHONE KOEHAVONG AND WILLIAM G. THILLY

Center for Environmental Health Sciences, Division of Toxicology, Whitaker College of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by Leonard S. Lerman, September 18, 1989 (received for review May 22, 1989)

ABSTRACT Denaturing gradient gel electrophoresis (DGGE) was used to separate and isolate the products of DNA amplification by polymerase chain reaction (PCR). The strategy permitted direct enumeration and identification of point mutations created by T4, modified T7, Klenow fragment of polymerase I, and Thermus aquaticus (Taq) DNA polymerases. Incorrectly synthesized sequences were separated from the wild type by DGGE as mutant/wild-type heteroduplexes and the heteroduplex fraction was used to calculate the average error rate (mutations per base pair duplication). The error rate induced in the 104-base-pair-low-temperature melting domain of exon 3 of the human hypoxanthine/guanine phosphoribosyltransferase (HPRT) gene was $3.4 \times 10^{-5}$ for modified T7, $1.3 \times 10^{-4}$ for Klenow fragment, and $2.1 \times 10^{-4}$ for Taq polymerases after a 100-fold amplification. The error rate for T4 DNA polymerase was not more than $3 \times 10^{-4}$ error per base pair duplication. The predominant mutations were sequenced and found to be transitions of G-C to A-T for T4 and modified T7 DNA polymerases, and A-T to G-C for Taq polymerase. Klenow fragment induced both possible transitions and deletions of 2 and 4 base pairs.

The ability to analyze point mutations has been greatly facilitated by the development of the polymerase chain reaction (PCR) (1–7). This technique permits a rapid synthesis of desired sequences from genomic DNA by DNA chain extension simultaneously from two opposing primers catalyzed by a DNA polymerase (1–3). Any DNA polymerase, however, will make errors during DNA synthesis with the kind and rate of errors varying among specific DNA polymerases and reaction conditions used (8–12). Indeed, estimates of the error rate (mutations per base pair duplication) in PCR have been reported to be $8 \times 10^{-2}$ using Klenow fragment (13) and $2 \times 10^{-1}$ using Thermus aquaticus DNA polymerase (Taq polymerase) (4). An error rate of $2 \times 10^{-4}$ mutation per base pair duplication would result in an accumulation of 0.8 mutation per double-strand DNA after a 100-fold amplification (20 duplications) of a 100-base-pair (bp) sequence. Such polymerase-induced mutations may not cause any serious problem during DNA sequencing if they occur uniformly along the amplified products. However, we are using PCR and denaturing gradient gel electrophoresis (DGGE) to study point mutations directly from human cell populations in which in vivo mutant fractions are $10^{-7}$ mutant per base pair, and such reported PCR error rates would create serious background problems. For this reason, we have analyzed the fidelity of a series of DNA polymerases in the PCR protocol by DGGE, which can separate DNA sequences differing by only single base substitutions, small additions, or deletions (14–20). In this study, we showed that this technique permitted us to separate polymerase-induced mutant sequences from the correctly amplified sequences.

The identities of the mutants generated by each of the four DNA polymerases were determined by sequencing predominant individual mutant sequences isolated from the denaturing gradient gels.

MATERIALS AND METHODS

Materials. T4 and Taq DNA polymerases were obtained from New England Biolabs; Klenow fragment of Escherichia coli DNA polymerase I was from Bethesda Research Laboratories; and modified T7 DNA polymerase (Sequenase) was from United States Biochemicals. 2'-Deoxynucleoside 5'-triphosphates were obtained as 100 mM solution from Pharmacia. The oligonucleotides (Synthetic Genetics, San Diego, CA) used as primers for PCR were as follows for HPRT exon 3: primer P1, 5'-CATATTTAAATATACCTCAG-3'; primer P2, 5'-TCTGATTTTTTCTCTGTAG-3'; primer P3, 5'-GAGCTGACCTTGCTGCAG-3'.

To obtain end-labeled fragments, the amplification was carried out with 5'-end-labeled primers (21). Genomic DNA was isolated from exponentially growing male TK6 human lymphoblasts (22, 23).

PCR Procedure. The PCR conditions for experiments with Klenow fragment were as described (2, 3). The conditions used for T4 (21) and for modified T7 (24) DNA polymerases were similar to those described for Klenow fragment except that the concentrations of dNTP were increased to 2.15 and 2.7 mM, respectively.

For Taq polymerase amplification, the 100-μl reaction mixture was 16.6 mM (NH4)2SO4/67 mM Tris-HCl, pH 8.8 (25°C)/6.7 mM MgCl2/10 mM 2-mercaptoethanol/200 μM each dNTP/1 μM each primer. Each amplification cycle consisted of incubating the reaction mixture at 93°C for 1 min (except 3 min for the first cycle), at 53°C for 2 min, and at 70°C for 2 min. One microliter (2.5 units) of Taq polymerase was added every 10 cycles after the 2-min incubation at 53°C.

DGGE. To analyze the PCR products as mutant/wild-type heteroduplexes, amplified DNA samples (0.5–1 × 107 cpm) in a 30μl solution of 400 mM NaCl/10 mM Tris-HCl, pH 7.5/2 mM EDTA were boiled 5 min and reannealed 5 hr at 65°C. The DNA was recovered by ethanol precipitation and separated on a 12.5% polyacrylamide gel (bisacrylamide/acrylamide, 1:3.75) containing a linearly increasing gradient of denaturant from 15% (vol/vol) to 30% (vol/vol) (100% denaturant = 7 M urea/40% formamide) (17). The gel was run for 14 hr at 150 V, submerged in 60°C in TAE buffer (40 mM Tris-HCl/20 mM NaOAc/2 mM EDTA, pH 8.3), dried, and autoradiographed.

Determination of Heteroduplex Fraction. Radioactive bands separated in the denaturing gradient gel were first located by autoradiography. The wild-type band and heteroduplex region (region between the wild-type band and the origin of the gel) were excised through the autoradiogram.
superimposed on the gel. The amount of radioactivity was determined by Cerenkov counting of the dried portions of gel.

Isolation of DNA from Denaturing Gradient Gel and Sequencing Analysis. Radioactive bands were located by autoradiography and excised from the dried gel. The DNA was recovered (21) and sequenced as described in the United States Biochemicals sequencing kit (25) with the following modifications: DNA templates (1–5 × 10⁶ copies or 2–10 ng of the 204-bp DNA) were directly mixed in a 16-μl labeling reaction (0.1 μM primer P1/6.25 mM dithiothreitol/0.5 mM each dCTP, dGTP, dTTP, labeling mixture)/0.5 μl of [α-32P]dATP (3000 Ci/mmol; 1 Ci = 37 GBq)/25 mM Tris-HCl, pH 7.5/12.5 mM MgCl2/32 mM NaCl (reaction buffer)/10% dimethyl sulfoxide), boiled 3 min and cooled 2 min at room temperature. After addition of 2 μl of diluted Sequenase (1.5 units), the reaction mixtures were kept 2 min at room temperature, 3.5 μl was then taken and mixed with 2.5 μl of each of the four termination mixtures. After 2 min incubation at 37°C, 4 μl of the stop solution was added as described in the kit, and the reaction mixtures were analyzed on an 8% sequencing gel.

RESULTS

Efficiency of DNA Amplification. The initial amplification was carried out from genomic DNA using primers P1 and P2, which immediately flanked the 184-bp human HPRT exon 3 sequence (Fig. 1B) and the following four DNA polymerases: T4, modified T7, Klenow fragment of Pol I, and Taq. As shown in Fig. 2, the efficiency of amplification of the expected size 224-bp fragment varied according to the type of DNA polymerase. Efficiencies estimated during the first 20 cycles according to the equation (1 + Y)ⁿ = 10ⁿ amplification, where Y is the efficiency per cycle and n is the number of cycles (2), were 90–93% with modified T7 DNA polymerase, 88% with Taq polymerase, and ~80% with Klenow fragment. T4 DNA polymerase gave an efficiency of 56% through 30 cycles. The high efficiency and yield obtained with modified T7 DNA polymerase, which produced >4 × 10⁷-fold amplification after 30 cycles, probably resulted from its highly processive activity (25, 26). In addition to the expected size 224-bp fragment, unwanted sequences appeared, especially when using Klenow fragment (Fig. 2).

Fidelity of DNA Amplification. The HPRT exon 3 sequence is composed of an 80-bp high-temperature melting domain and a 104-bp low-temperature melting domain (Fig. 1A), which is suitable for analysis by DGGE. To compare the fidelity of DNA polymerases and to detect potential bias in amplification by differential efficiency based on DNA sequence, each DNA polymerase was used to first carry out a 10⁴-fold amplification from genomic DNA using P1 and P2, during which process the number of DNA copies amplified increases as an exponential function of cycle number. The 224-bp DNA from each amplification mixture was then gel purified and further amplified to a total 10⁶- or 10⁷-fold using primer P1 and internal primer P3 (Fig. 1B). Care was taken to ensure that the final products were at the same concentrations of DNA substrates, dNTP, and primers and were synthesized under well-behaved conditions of logarithmic increase in DNA per cycle at both 10⁶- and 10⁷-fold amplifications. The PCR products were first boiled and reannealed so that each strand of the mutant homoduplexes was hybridized to the complementary strand from either the correctly

---

Fig. 1. Melting behavior of the human HPRT exon 3 sequence and positions of the primers used for PCR. (A) Melting map for the 184-bp wild-type exon 3 sequence (20). This sequence is composed of high- (positions 220–299) and low- (positions 300–403) temperature melting domains. The melting map represents the temperature at which each base is in 50:50 equilibrium between the melted and helical state (16, 17). (B) Positions of the primers (P) used to amplify the exon 3 sequence. To compare the efficiency and fidelity of DNA amplification, three primers were used: P1 and P2 are complementary to introns (dashed lines) immediately flanking the 3’ and 5’ ends of exon 3 (boxed), respectively. Internal primer P3, adjacent to P2, is complementary to the 5’ end of exon 3. The expected size PCR products using P1 + P2 (224 bp) and P1 + P3 (204 bp) are indicated as thick lines.

Fig. 2. Electrophoretic analysis of the PCR products. Comparison of the efficiency of DNA amplification using four DNA polymerases. PCR was carried out from 5 μg of genomic DNA using end-labeled primers P1 and P2 indicated in Fig. 1. After 20, 25, and 30 cycles, 1/20th of the reaction mixture was analyzed on a 6% polyacrylamide gel (bisacrylamide/acrylamide, 2:38). The gel was stained with ethidium bromide and photographed under UV light. Lane M, marker is 250 ng of pBR322 cut by Msp I. The 224-bp fragment was excised from the gel, and the incorporated radioactivity was measured to determine the amount of DNA. To estimate the specific activity of the amplified DNA, the 30 cycle-amplified 224-bp fragment was purified from the reaction mixture, the radioactivity incorporated was measured, and the amount of DNA was determined by spectrophotometry.
amplified sequences (wild type), if present in sufficient excess, or from other mutant homoduplexes. In this manner, each mutant sequence was expected to be detected in lower denaturant concentrations of the gel as two mutant/wild-type or various mutant/mutant homoduplexes which would separate further from the wild type than the mutant homoduplex (17, 27). Fig. 3 shows the patterns of the exon 3 PCR products separated by DGGE after 10^6- and 10^8-fold amplification for each DNA polymerase. The wild-type sequence focused at 24% of denaturant concentrations, and a series of bands was observed in lower denaturant concentrations. Unique mutant bands can be distinguished with T4, Taq, and modified T7 DNA polymerases. In the pattern produced by Klenow fragment, some of the bands appeared at positions similar to those observed with either modified T7 DNA or Taq polymerases. T4, modified T7 DNA, and Taq polymerases thus appeared to induce different sets of mutations in the low-temperature melting domain of exon 3. The data also suggested that some of the mutations induced by Klenow fragment were common to those produced by modified T7 DNA or Taq polymerases.

**Determination of the Heteroduplex Fraction and Error Rate for Each DNA Polymerase.** For each PCR product separated by DGGE (Fig. 3), total heteroduplex fraction was estimated as the ratio between the radioactivity in the heteroduplex region (region between the origin of the gel and the homoduplex wild type) and the sum of the radioactivity found in the heteroduplex region and the homoduplex wild type. Back-ground noise such as spontaneous depuration or photoreaction products were estimated by purifying wild-type homoduplex bands and subjecting them to handling identical to amplified DNA preparations: boiling, reannealing, and separation by DGGE. This process revealed a background unrelated to DNA amplification of ~5%. The heteroduplex fraction, HeF, of DNA in the heteroduplex region due to DNA amplification for each DNA polymerase was approximated as (counts in heteroduplex region – background counts in heteroduplex region)/(total counts in heteroduplex and wild-type homoduplex regions).

Using this value of HeF, the error rate, f, was then calculated for each DNA polymerase after 10^6- and 10^8-fold amplification as
\[ f = \frac{HeF/b}{d} \]
where \( b \) is the length of the single-strand low-temperature melting domain of \( HPRT \) exon 3 (104 nucleotides) and \( d \) is the number of DNA duplications effected (20 for 10^6 and 26.5 for 10^8-fold amplification). From this, one may also calculate the fraction of amplified DNA that contained at least one PCR-derived mutation as
\[ F(\delta^1) = 1 - e^{-bd} \]

For Taq polymerase, the observed HeF values were 43% and 53% after 10^6- and 10^8-fold amplification corresponding to the estimated error frequencies of 2.1 and 1.9 \times 10^{-4} mutation per base per duplication, respectively. For Klenow fragment, the respective values of HeF were 27 and 44% corresponding to error frequencies of 1.3 and 1.6 \times 10^{-4}. For modified T7 DNA polymerase, the values of HeF were 7% and 10%, corresponding to error frequencies of 3.4 to 3.6 \times 10^{-5}. For T4 DNA polymerase HeF values were difficult to estimate since the amount of radioactivity above background was so small after 10^8-fold amplification that no significant increase was detected in most experiments. From the experiment shown in Fig. 3, the respective estimates of HeF were 0.6% and 2.8%, corresponding to estimated error frequencies of 2.9 \times 10^{-6} and 1 \times 10^{-5}.

One of the outcomes of comparing the relative intensities of the mutant bands is the repeated observation that the ratios of the heteroduplex fractions after 10^6- and 10^8-fold amplifications for Klenow fragment (1.6) and especially for T4 DNA polymerase (4.5) are significantly higher than the ratio of the duplications 26.5/20 (1.325) expected if there was no difference in amplification efficiency based on DNA sequence. This resulted in an intense doublet with T4 DNA polymerase and an indistinct quadruplet with Klenow fragment after 10^8-fold amplification (Fig. 3) that corresponded to G+C to A+T (position 351; see Fig. 6) and two complex deletions of 2 and 4 bp (positions 400–403), respectively. Since the experiment was performed under the conditions of exponential amplification, these bands may result from a differential amplification efficiency for these particular mutant sequences using these enzymes. This indicates the need for quantitative controls for such allelic preference of the PCR products.

**Identification of Mutant Sequences.** Each individual band was isolated from denaturing gradient gels, further amplified 10^6- to 10^8-fold, and separated by DGGE. In the case of bands 1 and 2 generated by T4 DNA polymerase (Fig. 4A, lane T4), this process resulted in two major homoduplexes, mutant and wild type (lanes 1a and 2a). After boiling and reannealing, this same DNA gave rise to two additional bands (lanes 1b and 2b), expected to correspond to the mutant/wild-type heteroduplexes. The fact that both bands 1 and 2 (lane T4) gave rise to precisely the same pattern of bands (lanes a and b) was consistent with the interpretation that they were heteroduplexes formed from the complementary strands of the same mutant homoduplex. Sequence analysis of both mutant homoduplexes (Fig. 4B) showed the same G+C to A+T transition at position 351. G+C to A+T transitions were found for two other less intense bands created by T4 DNA polymerase. By the same process, the most intense bands created by the other DNA polymerases were also characterized. Fig. 5 shows the

---

**Fig. 3.** Analysis by DGGE of the PCR products after 10^6- and 10^8-fold amplification using the DNA polymerases indicated. To compare the fidelity of DNA amplification, two rounds of PCR were carried out from genomic DNA template. (i) The exon 3 sequence was first amplified 10^6-fold from 5 μg of genomic DNA with endlabeled primers P1 and P2 (see Fig. 1B). The 224-bp fragment was gel purified (21) and used as template for a second round of PCR. (ii) To achieve a total 10^6- and 10^8-fold amplification by each DNA polymerase, fractions of the 224-bp fragment (5 × 10^10 and 5 × 10^8 copies) were amplified by using primers P1 and P3 an additional 10^6- and 10^8-fold, respectively. The resulting 10^6- and 10^8-fold-amplified 204-bp fragments were gel purified and 5 × 10^6 cpm aliquots were separated as heteroduplexes on a 12.5% polyacrylamide gel containing 15–30% denaturant concentrations (top to bottom).
transversion, and one contained the more visible bands. The wildtype mutant of the polymerase was isolated from the wildtype DNA and replated with a Klenow fragment. The resulting bands were observed as a 6-bp deletion. The deletions produced by Klenow fragment corresponded to a substitution of the 10-bp sequence GAGCTATTGT (positions 394–403) immediately contiguous to the 3' end of primer P1 by either a 6-bp sequence CGTCT through an 8-bp sequence CCGTTCGGC, resulting in the 4- and 2-bp deletions, respectively.

**DISCUSSION**

The combination of PCR and DGGE has permitted us to identify the kinds and positions of mutations created by several DNA polymerases and to measure their fidelities. Each DNA polymerase demonstrated a reproducible overall replication error rate and a specific pattern of bands when the amplified products were analyzed by DGGE. Individual mutant bands were isolated and sequenced to determine the nature of the mutations.

With regard to earlier reports of errors produced by Taq polymerase, the estimate of $2.1 \times 10^{-4}$ error per base per duplication is in good agreement with both the error rate of $2 \times 10^{-4}$ error per base per cycle for the amplification of a 10 mutant homoduplexes arising from Klenow fragment amplification were six G-C to A-T transitions, two A-T to G-C transitions, one 2-bp deletion, and one 4-bp deletion (Fig. 5C).

The kinds and positions of the mutations induced by the four DNA polymerases in the low-temperature melting domain of exon 3 are summarized in Fig. 6. The 35 mutations found consisted of 32 single base-pair substitutions, including only one A-T to T-A transversion produced by modified T7 DNA polymerase, and three deletions (two of 2 bp and one of 4 bp). Four of the G-C to A-T transitions and one of the A-T to G-C transitions produced by Klenow fragment were common to those observed with modified T7 DNA and Taq polymerases, respectively. All three deletions induced by both modified T7 DNA polymerase and Klenow fragment occurred at the level of the base pairs immediately adjacent to the 3' end of the primer P1. The 2-bp deletion created by modified T7 DNA polymerase may represent a by-product of a mishybridization of the last nucleotides, CA (Fig. 6, positions 404 and 405), at the 3' end of the primer with a GT at positions 402 and 403 instead of the expected GT at positions 404 and 405. Such a shorter strand would then be amplified. The deletions produced by Klenow fragment corresponded to a substitution of the 10-bp sequence GAGCTATTGT (positions 394–403) immediately contiguous to the 3' end of primer P1 by either a 6-bp sequence CGTCTT or an 8-bp sequence CGTCTTGC, resulting in the 4- and 2-bp deletions, respectively.

**FIG. 4.** Identification of mutant sequences by DGGE. (A) Example of isolation of individual mutant bands for sequencing. Bands 1 and 2 generated by T4 DNA polymerase were isolated from the gel and amplified an additional 102-fold by using T4 DNA polymerase and end-labeled primers P1 and P3. The amplified DNA was gel purified and 5 × 10^4 cpm were analyzed by DGGE without boiling and reannealing (lanes 1 and 2a) or after boiling and reannealing (lanes 1b and 2b). The positions of the homoduplexes (Homod.) and heteroduplexes (Het.) are indicated. The heteroduplexes seen as minor bands in lanes 1a and 2a were probably formed from the homoduplex mutant and wild type during the last cycles of boiling and reannealing of the PCR process. (B) Sequencing analysis of the mutant DNA. Both mutant homoduplexes separated from the wild type by DGGE as shown in A were isolated from the gel and sequenced in parallel with the wild-type DNA. Both contained the identical mutation, a C to T transition, at position 351 in exon 3 (see Fig. 6).

**FIG. 5.** Analysis by DGGE of individual mutant sequences produced by modified T7, Taq, and Klenow fragment of Pol I DNA polymerases. Visible heteroduplex bands as shown in Fig. 3 were isolated from the gel and the types of mutations were identified as in Fig. 4. From each mutant homoduplex characterized, 5 × 10^4 cpm were boiled and reannealed with wild-type DNA to generate heteroduplexes and were then separated by DGGE. Ten mutants for modified T7 (A), 12 mutants for Taq (B), and 10 mutants for Klenow fragment of Pol I (C) DNA polymerases were identified. The positions of the wild-type band (WT) and each of the mutant homoduplexes are indicated by an arrow. The positions and the types of mutations found for each mutant are indicated at the top and bottom of each lane.
human HLA-BFβ sequence (4) and the error rate of 2.8 × 10⁻⁴ substitution per nucleotide synthesized reported for a single round of DNA synthesis using the phage M13mp2 (29). Our observations of 12 different A-T to G-C transitions with Taq polymerase are also consistent with the predominant A-T to G-C transitions determined in both studies (4, 29) and in a Taq-amplified 798-bp fragment of human APO-B gene (30) by clone-by-clone sequencing of the amplified products. Such close agreement supports the validity of the protocol described here as a tool to measure the fidelity of DNA polymerases in vitro.

The fidelities determined here were based on the conditions optimized with regard to maximum efficiencies of amplification for each particular DNA polymerase. The reaction conditions (temperature, dNTP, and salt concentrations) and the DNA templates used will undoubtedly influence the error rates and perhaps the kinds of mutations. However, similarities observed for the error rates and also predominant mutations for Taq polymerase using four different templates indicate that the fidelity of this enzyme remains relatively constant during DNA synthesis for these templates. The protocol described should also facilitate studies of the effects of experimental conditions and auxiliary factors on the fidelity of DNA polymerases in vitro to achieve high fidelity. Knowledge of amounts and kinds of polymerase-induced mutations within the particular DNA sequence are essential for our studies and may be important to others’ work involving use of the PCR.

We gratefully acknowledge Drs. L. S. Lerman and M. F. Goodman for helpful suggestions during the course of this work, Dr. N. F. Cariello for critically reading the manuscript, and L. Cote for expert technical assistance. This work was supported by grants from the National Institute of Environmental Health Sciences (Grants 1-P42-ES04675, S-P01-ES00597, and 1-P50-ES03926-05) and Office of Helath and Environmental Research, U. S. Department of Energy (Grant DE-FGO2-86-ER60448).